

基于改进 YOLOv8 的桥梁小目标裂缝检测

李金沛^{1, 2, 3}, 孟晓林^{4, 5}, 胡亮亮^{1, 2, 3}, 鲍 艳^{1, 2, 3}, 赵诗雨^{1, 2, 3}

(1. 北京工业大学 建筑工程学院, 北京 100124, 中国;

2. 北京工业大学 桥梁工程安全与韧性全国重点实验室, 北京 100124, 中国;

3. 北京工业大学 城市与工程安全减灾教育部重点实验室, 北京 100124, 中国;

4. 东南大学 仪器科学与工程学院, 南京 211189, 中国; 5. 帝国理工大学 工程学部, 伦敦 SW7 2AZ, 英国)

摘 要: 随着中国桥梁大规模进入运维阶段, 传统的桥梁检测方法通常依赖人工巡检和目视检查, 无人机的桥梁表面图像数据采集能力较强, 因此被逐步应用于桥梁表面病害检测。针对交通管控下无人机拍摄桥梁图像尺寸较大和小目标裂缝较多, 目前检测算法提取小目标裂缝特征信息困难且存在识别漏检、误检等问题, 该文提出了一种基于改进 YOLOv8 的桥梁小目标裂缝检测算法。首先, 用高效视觉变换器(efficient vision transformer, EfficientViT)替换 YOLOv8 的主干网络, 以减少大量冗余参数, 增强图像的局部特征提取能力; 其次, 引入大型选择性核网络(large selective kernel network, LSKNet), 融合 C2f 模块注意力机制, 选择性调整卷积核尺寸, 降低运行参数, 减少计算复杂度; 最后, 增加双向特征金字塔网络(bidirectional feature pyramid network, BiFPN), 融合 P2 小目标检测头, 增加不同尺度特征图之间的信息耦合和小目标裂缝检测框, 实现小目标裂缝精准识别与提取。依托无人机采集的某桥梁裂缝数据集进行小目标裂缝检测验证, 与 YOLOv8 模型相比, 该算法模型的图像识别准确率提高了 3.7%, 召回率提高了 3.5%, F1 分数提高了 3.5%, 平均精度均值 50 (mAP50)提高了 3.9%, 平均精度均值 50 到 95(mAP50-95)提高了 7.4%。该文所提的基于改进 YOLOv8 的桥梁小目标裂缝检测算法提高了桥梁结构小目标裂缝的识别准确率, 为广泛应用于桥梁结构的表面健康监测提供了可行的解决方案。

关键词: 桥梁结构健康监测; 无人机; 小目标裂缝检测; YOLOv8; 图像处理

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-0054(2025)07-1260-12

DOI: 10.16511/j.cnki.qhdxxb.2025.26.023

截至 2023 年末, 中国累计建设了 107.93 万座公路桥梁, 其中包括 18.79 万座大桥及特大桥, 中国在桥梁数量、设计理论、建造技术和建设装备方面均达到国际领先水平^[1-2]。随着桥梁使用年限的增加, 通过维护提高在役桥梁的安全性、适用性和耐久性, 以及桥梁的防灾减灾和结构韧性能力, 已成为桥梁工程师面临的主要挑战之一^[3]。在外部荷载及内部因素等耦合作用下, 受力关联影响导致桥梁在不同位置会出现裂缝, 因此裂缝可以作为评价在役桥梁安全性的重要指标^[4]。对在役桥梁进行周期性裂缝监测, 并制定科学的评估和养护决策, 已成为当今世界范围内的研究热点。

在传统桥梁现场检测中, 一般借助桥梁检测车、搭接脚手架等设备平台, 使用裂缝尺、裂缝显

微镜、游标卡尺等测量仪器, 采用人工目视测量的手段开展裂缝检测^[5]。由于传统裂缝检测方法存在费时费力、依赖主观性、检测效率低、安全隐患大、病害漏检率高, 以及复杂桥梁结构难以进行检测的问题, 因此难以满足实际工程的时效性和准确性需求。随着无人机技术和计算机视觉的发展, 利用无人机采集桥梁表面数据和结合计算机视觉算法进行裂缝检测识别已经逐步应用于裂缝检测领域, 为桥梁裂缝检测智能化、高效化、精确化提供了一条可行的技术路线^[6-7]。

与传统图像数字处理相比, 深度学习算法能自动适应复杂裂缝, 避免了手工设计特征提取器, 是计算机视觉的核心^[8]。裂缝识别定位算法分为 2 类: 基于区域卷积神经网络(region-based

收稿日期: 2024-08-13

基金项目: 国家自然科学基金面上项目(51829801, 52378385)

作者简介: 李金沛(1999—), 男, 硕士研究生。

通信作者: 孟晓林, 教授, E-mail: xiaolin_meng@seu.edu.cn

convolutional neural network, R-CNN)的两阶段算法,如 Fast R-CNN、Faster R-CNN 和 Mask R-CNN 等,通过区域网络生成目标候选区域,进行裂缝的分类和边界框回归,输出裂缝的类别和准确位置^[9];基于你只看一次(You Only Look Once, YOLO)的单阶段算法,如 YOLOv1 至 YOLOv8 等版本,通过放弃区域提取步骤,将检测任务作为回归问题解决,直接获得图像中裂缝的位置和类别信息,提高裂缝识别效率,已广泛应用于裂缝识别领域^[10]。Yu 等^[11]针对无人机无法实时监测桥梁裂缝的问题,提出了融合焦点损失和剪枝模型(focal loss and pruning model, FLPM)损失函数的 YOLOv4-FPM 深度学习模型,可在提高模型精度的同时降低模型大小和参数量,不仅能实时处理图像,还能识别不同尺寸的桥梁裂缝。Kumar 等^[12]提出一种基于边缘算法的 YOLOv3 模型,并部署在六轴无人机上,解决了超高层建筑损伤的实时识别问题,实现了超高层建筑外表面损伤的实时健康监测。Yang 等^[13]用 W-PAN 网络替换 YOLOv5 主干网络,通过 GhostC3 和 ShuffleConv 模块优化原有网络的 C3 和 Conv 模块,提高了沥青路面非变形病害检测的精度和速度。Teng 等^[14]基于 YOLOv3 深度学习算法,通过引入迁移学习模型和数据增强技术,提高模型的特征提取性能和鲁棒性,实现了桥梁表面裂缝和钢筋裸露在不同图像分辨率下的精确高效识别。Wu 等^[15]提出了一种改进的 YOLOv8 深度学习模型,该模型结合了轻量级主干网络和高效的原型掩膜分支,降低了模型的复杂度,并保持模型的精度不变,完成了对大坝和桥梁结构裂缝的实时像素级检测。有关学者已对 YOLO 算法进行了改进,结合无人机数据采集,满足了算法轻量化、准确率提升、复杂环境裂缝识别、实时处理的需求。无人机安全距离的限制和裂缝过小导致桥梁裂缝检测过程中的小目标特征信息提取困难,进而导致小目标裂缝被误检漏检,这为桥梁运营带来巨大的安全隐患。因此,需要一种算法对桥梁中的小目标裂缝进行准确识别。

本文在 YOLOv8 算法的基础上提出一种基于多尺度融合的 EfficientViT-LSKNet-BiFPN-YOLOv8 算法。首先,该算法用高效视觉变换器(efficient vision transformer, EfficientViT)替换 YOLOv8 主干网络,增强算法提取图像局部特征的能力;其次,引入大型选择性核网络(large selective kernel

network, LSKNet)注意力机制,融合 C2f 模块,降低运行参数,减少计算复杂度;最后,利用双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)融合 P2 小目标检测层,增加不同尺度特征图之间的信息耦合和小目标裂缝的检测头。该算法实现了对小目标特征信息的提取,提高了小目标裂缝检测的准确率。

1 裂缝智能检测算法

1.1 YOLOv8 算法

YOLOv8 是一种将物体检测和图像分割集成于一体的一阶段目标检测算法,由 Ultralytics 公司于 2023 年 1 月 10 日发布,精度、速度、易用性和通用性等性能较好,是目前目标检测领域的先进算法^[16],其网络结构如图 1 所示。

YOLOv8 在 Backbone 部分使用了更强大的视觉主干网络,并使用梯度流更丰富的 C2f 替换原本的 C3 模块,以增强卷积神经网络的特征融合能力和推理速率^[16];Neck 采用了简洁高效的 FPNLite,能较好地融合不同层次的语义和位置信息^[16];Head 使用了新的解耦头,将目标检测头从 Anchor-Based 变为 Anchor-Free,同时将分类头和检测头分离^[16];在数据增强部分,YOLOv8 借鉴了 YOLOX 的方法,在最后 10 个训练过程中关闭 Mosaic 增强数据集,有助于提高模型的泛化能力和精度。YOLOv8 的数据预处理依然采用 YOLOv5 的策略,同时采用 Mixup、random_perspective 和 HSV 这 3 种数据增强手段;YOLOv8 改变了以往的交并比(intersection over union, IoU)匹配策略或者单边比例的分配方式,利用 Task-Aligned Assigner 动态分配和匹配样本^[17];此外,YOLOv8 目前有 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x 等 5 个版本,可供相关研究人员在不同条件下使用。

1.2 改进网络优化

1.2.1 EfficientViT 网络结构

EfficientViT 是 Cai 等^[18]于 2023 年提出的一种在运算能力有限的设备上部署深度学习模型的轻量级多尺度线性主干网络,其结构如图 2 所示。EfficientViT 是一种高效的视觉 Transformer 模型,与传统的 Transformer 模型相比,视觉转换器

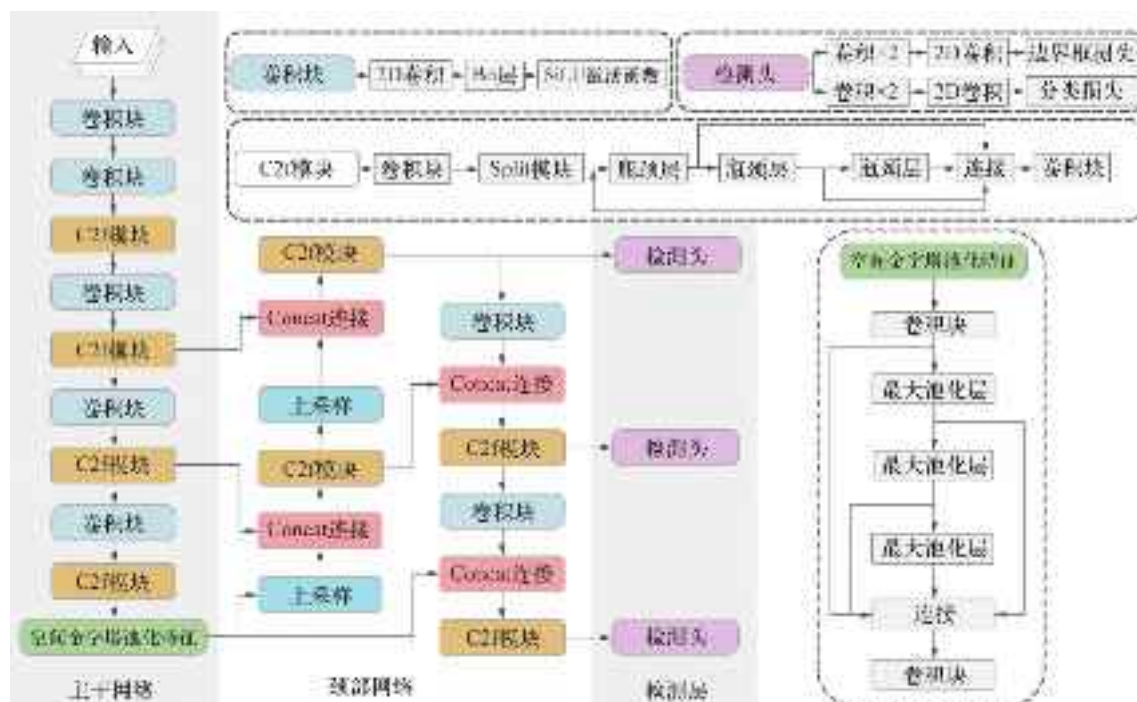


图1 YOLOv8 的网络结构

(vision transformer, ViT)模型将 Transformer 模型的自注意力机制应用于计算机视觉任务,将任意大小和比例的图像划分成网格,并视为输入值,不必像卷积神经网络一样,必须输入固定尺寸的图像^[19];在构架融合方面, EfficientViT 采用 backbone-head 结构,并在 backbone 结构后期插入了轻量级多尺度注意力模块,用于捕获上下文信息,提供良好的上下文感知能力;在注意力机制方面, EfficientViT 引入了基于修正线性单元 (rectified linear unit, ReLU) 的全局注意力机制,利用矩阵的关联性和 ReLU 全局注意力,将计算复杂度从二次关系降至线性关系,以减少计算冗余;在特征表示方面, EfficientViT 用可学习的线性层生成特征图代替传统的图像编码过程,并与多尺度小卷积核和 ReLU 全局注意力结合,实现了多尺度特征学习;在主干网络方面, EfficientViT 采用具有深度可分离卷积的移动翻转瓶颈卷积及其变种分布移位卷积等轻量级的卷积操作,可减少参数数量,并保持可观的特征学习能力。

1.2.2 LSKNet 注意力机制

2023 年, Li 等^[20]提出针对遥感图像目标检测任务的轻量级深度神经网络 LSKNet,旨在通过调整大感受视野,适应遥感图像中不同的物体,其注意力机制结构如图 3 所示。LSKNet 基于自注意力

机制和卷积核选择技术设计,可动态调整感受视野大小,并可针对性地捕获和建模遥感图像中目标周围复杂多样的背景信息。LSKNet 采用大型可调卷积核,大大减少了网络参数,降低了计算复杂度,提高了在移动终端等资源受限环境下的实际应用的计算效率。LSKNet 通过动态调整特征提取主干网络中的卷积核尺寸,并融合暗通道先验的传统去雾算法,提出了一种局部敏感关键点绝缘子图像去雾算法,显著提高了绝缘子目标检测的精确度。将 LSKNet 融合 C2f 注意力机制模块,弥补因非结构化环境造成的精度损失,选择性调整卷积核尺寸,降低运行参数,并减少计算复杂度,提高 YOLOv8 算法注意力机制的运行效率。

1.2.3 BiFPN 融合 P2 小目标检测层

YOLOv8 原始模型包含 3 个检测层: P5 特征层 20×20 , 用于检测尺寸大于 32×32 像素的目标,对应大型目标物体; P4 特征层 40×40 , 用于检测尺寸大于 16×16 像素的目标,对应中型目标物体; P3 特征层 80×80 , 用于检测尺寸大于 8×8 像素的目标,对应小型目标物体。为进一步提升模型对极小目标的检测能力,本研究引入 P2 特征层 160×160 , 专门用于检测尺寸大于 4×4 像素的小目标裂缝,可提高小目标裂缝的检测精确率。为实现 P2 特征层与其他检测头之间的高效多尺度特征融合,

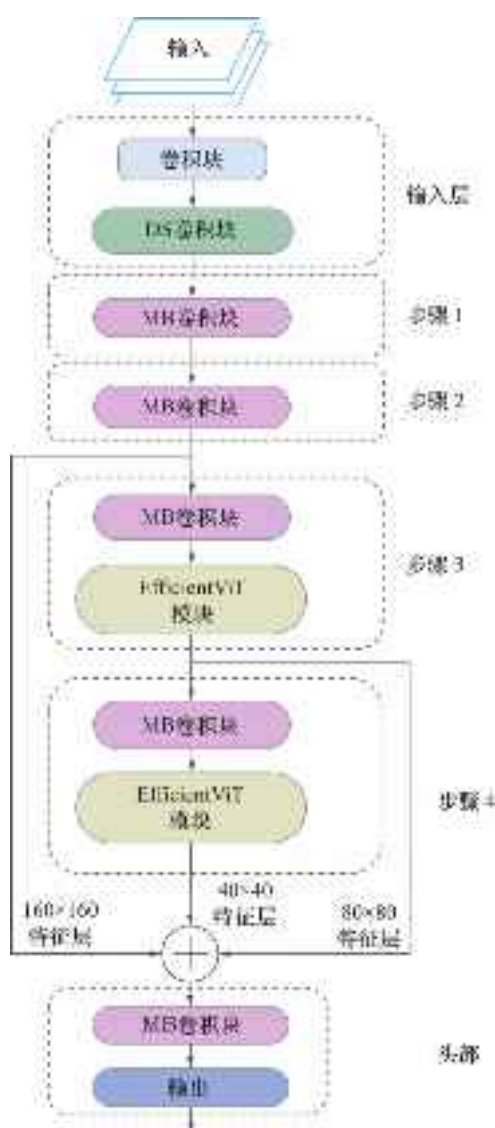


图2 EfficientViT 的网络结构

引入 BiFPN 融合 P2 小目标检测层,通过跨层级的双向特征融合,使得语义低级特征能获得更广泛的感受野,以提升识别小目标的能力^[21];通过引入多层特征融合机制,高级语义特征也获得了充分的细节特征信息,有助于识别大目标的边界细节,促进了特征的上下文传播,提高了整体对象检测的精确度,具体结构如图4所示。

1.3 改进 YOLOv8 算法

一方面,考虑安全因素,无人机无法近距离拍摄桥梁裂缝;另一方面,桥梁裂缝尺寸较小且分布分散,导致获取的桥梁裂缝图像数据存在裂缝相对面积占比低且分布不均匀等问题。针对上述问题,原始 YOLOv8 模型存在漏检误检率高、检测精度低和检测头尺寸过大等缺陷。为解决上述问题,针

对小目标裂缝数据的特点,本研究对 YOLOv8 算法的网络结构进行改进,具体改进措施如下,改进的 YOLOv8 网络结构如图5所示。

1) 用 EfficientViT 替换 YOLOv8 的主干网络,通过自注意力机制和卷积核分组稀疏化,减少冗余参数。

2) 保留并增强关键特征通道,提升对图像局部细节的特征提取能力,提高网络检测效率。

3) 在 YOLOv8 颈部结构中采用 LSKNet 注意力机制融合 C2f 模块,将标准 2D 卷积核分解为水平和垂直 2 个 1D 卷积核并级联,通过动态调整卷积核尺寸和稀疏化处理,降低模型的计算复杂度和内存占用。

4) 在 YOLOv8 颈部结构中引入 BiFPN 融合 P2 小目标检测层,实现不同尺度特征图之间的双向特征融合,降低上采样和下采样过程中的信息损失,增加 160×160 的 P2 检测头,以精准识别小目标裂缝。

2 网络训练与检测试验

2.1 数据集

目前,目标检测领域的公开数据集主要有 PASCAL VOC、MS COCO 和 Image Net 等^[22]。本研究选取的裂缝检测数据集来自 Roboflow 中的公开数据集,包括南洋理工大学土木工程系公开裂缝数据集(https://universe.roboflow.com/ntu-department-of-civil-engineering/building_crack_obj)、new-workspace-hk52d 公开裂缝数据集(https://universe.roboflow.com/new-workspace-hk52d/crack_1000_train)和 fine 公开裂缝数据集(<https://universe.roboflow.com/fine-aqmsj/construction-site-jwnnh>)。数据集共包含 7 426 张图像,其中:裂缝训练集图像 6 426 张,裂缝测试集图像 1 000 张。在裂缝训练集中选取 1 396 张图像作为裂缝验证集。本研究选取的裂缝模型训练数据集包含不同尺寸和类型的裂缝,以及同一图像中多种裂缝的情况,以提高模型的适用性,数据集的相关裂缝类型如图6所示。小目标裂缝不仅占比过小,而且颜色较浅,因此难以识别。本研究选取的数据集具有多样性特征,可提高模型的泛化能力,增强模型的鲁棒性,并使模型在多场景下都有较强的实际应用能力,从而准确检测识别桥梁的小目标裂缝。本研究使用 LabelImg 标注数据集中的裂缝信息,将标注完成的图片另存为 XML 文件^[23]。由于 YOLOv8 使用

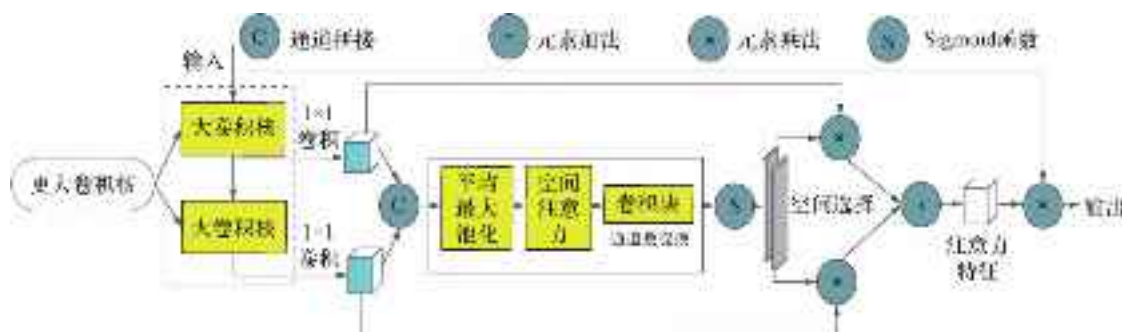


图3 LSKNet 注意力机制结构

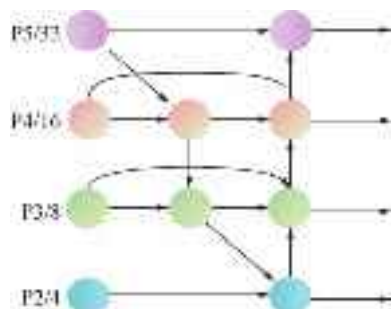


图4 BiFPN融合P2小目标检测层结构

TXT 格式的标注, 因此须利用脚本从 XML 文件中提取每个 ITEM 的类别信息和 4 个坐标点信息, 并将其保存在与图像命名一致的 TXT 文件中。标注完成的图像按照训练数据集、验证数据集和测试数据集分开保存。

2.2 试验平台及模型

本研究的模型训练平台为一高性能台式计算机, 具体参数如表 1 所示。为控制变量, 本研究的每次试验均使用相同的计算机配置。

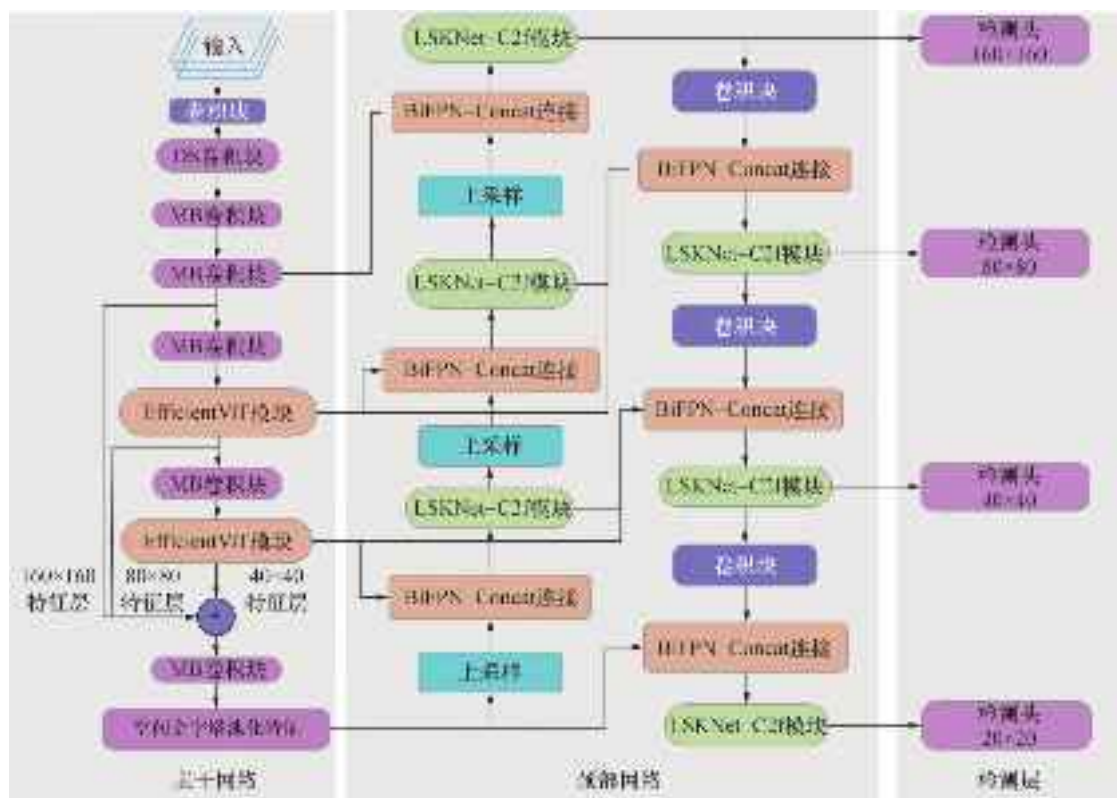


图5 改进的YOLOv8网络结构

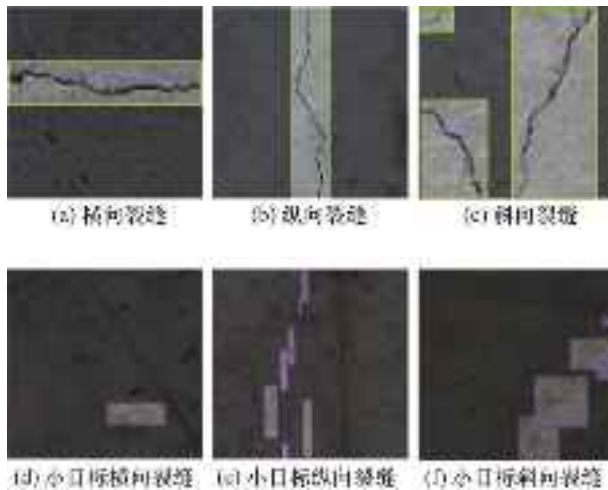


图6 数据集裂缝类型

表1 训练平台信息

计算机配置	参数信息
CPU	13th Gen Intel(R)Core(TM)i9-13900K
CPU 内存	128 GB
GPU	NVIDIA GeForce RTX 4090
GPU 内存	64 GB
操作系统	Windows10
Python 版本	3.8.18
Pytorch 版本	2.0.1
CUDA 版本	11.7
cuDNN 版本	8.9.5

2.3 模型训练

为确保研究的严谨性和模型训练结果的可比性,本研究的所有模型设置相同的训练参数,如表2所示,这些参数基于训练平台、模型训练效果和实际经验的综合考虑设置,以确保模型训练的稳定性和效率。上述参数在训练全过程中保持不变,以最大限度地控制训练条件的一致性,从而使不同模型的性能对比结果具有可靠性和参考价值。

表2 模型训练参数

名称	信息
训练批量大小	16 张图像
初始学习率	0.01
最终学习率	0.01
最大训练轮数	200
图像交并比	0.5
动量参数	0.8

2.4 评价指标

混淆矩阵是一种衡量分类模型性能的工具,广泛应用于图像识别领域的各类评价指标计算。本研究基于混淆矩阵计算了图像识别的多项评价指标,包括召回率 R 、准确率 P 、 F_1 分数、平均精度均值 50(mAP50)和平均精度均值 50 到 95(mAP50-95)。这些指标能综合评估模型在小目标裂缝检测中的表现。

3 分析与讨论

为验证本文提出的改进 YOLOv8 算法的有效性,本研究进行了注意力机制 LSKNet 与 C2f 模块的融合对比试验、主干网络 EfficientViT 与融合 BiFPN-P2 的小目标改进对比试验和综合改进对比试验。

3.1 LSKNet 与 C2f 模块的融合对比试验

为验证 LSKNet 注意力机制在不同位置改进的有效性,本试验将 LSKNet 融合 C2f 模块以不同的方式加入 YOLOv8 模型中的 4 个位置(Neck、Backbone、Backbone 输出和 Backbone 输出与 Neck 组合),并分别命名为 YOLOv8-LSKNet-Neck、YOLOv8-LSKNet-Backbone、YOLOv8-LSKNet-Backbone 输出和 YOLOv8-LSKNet-Backbone 输出-Neck。在控制相关变量并保持模型训练条件一致的情况下,对 4 种模型依次进行训练。 P 、 R 和 mAP50-95 随迭代次数的变化曲线如图 7 所示,模型最终的 R 、mAP50、mAP50-95、训练耗时和大小如表 3 所示。

训练结果表明,将注意力机制 LSKNet 融合 C2f 模块添加到 YOLOv8 算法的不同位置均可有效提高 YOLOv8 算法的相关性能。但添加注意力机制会导致模型大小、复杂度和构建时间增加。分析模型训练的迭代曲线可知,YOLOv8-LSKNet-Neck 与 YOLOv8-LSKNet-Backbone 输出-Neck 的 R 相似。由 P 和 R 评价结果可知,YOLOv8-LSKNet-Backbone 输出-Neck 模型在裂缝识别方面最有效。在其他模型的训练迭代曲线中,YOLOv8-LSKNet-Backbone 输出-Neck 均表现出优异的性能。与 YOLOv8 相比,YOLOv8-LSKNet-Backbone 输出-Neck 的提升最大, R 提高了 4.6%,mAP50 提高了 3.9%,mAP50-95 提高了 7.8%,表明其可提升裂缝识别的准确率。

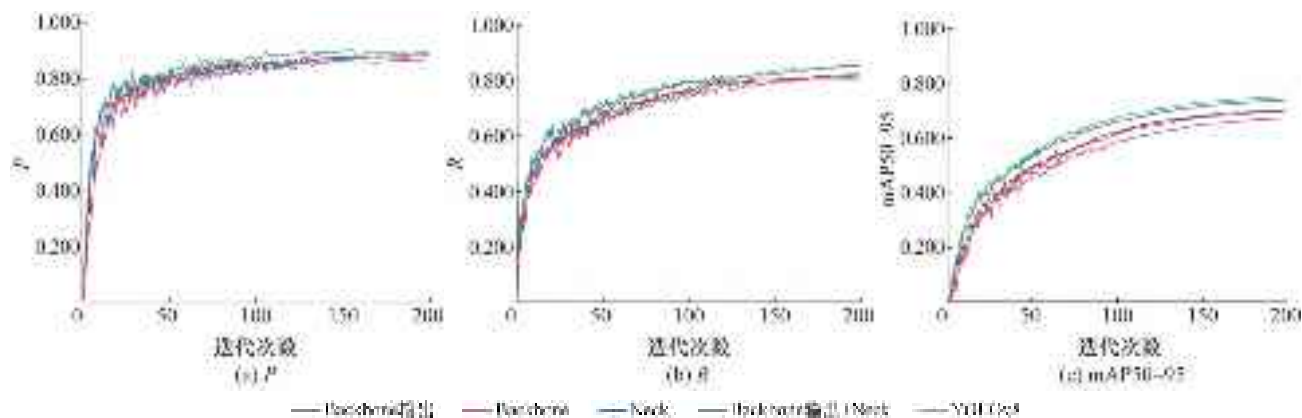


图 7 注意力机制改进对比试验

表 3 注意力机制改进对比试验最终结果

模型名称	R	mAP50	mAP50-95	训练耗时/h	模型大小/MB
YOLOv8	0.809	0.858	0.674	1.750	6.3
YOLOv8-LSKNet-Neck	0.852	0.885	0.739	2.100	6.7
YOLOv8-LSKNet-Backbone	0.823	0.871	0.699	1.959	6.6
YOLOv8-LSKNet-Backbone 输出	0.820	0.877	0.701	2.555	10.9
YOLOv8-LSKNet-Backbone 输出-Neck	0.855	0.897	0.752	2.035	7.0

3.2 EfficientViT 与融合 BiFPN-P2 的小目标改进对比试验

为验证 EfficientViT 替换原始 YOLOv8 主干网络和融合 BiFPN-P2 小目标的有效性和可行性,本试验对原始 YOLOv8 主干网络和颈部结构进行了改进,并进行了 3 个试验:用 EfficientViT 替换 YOLOv8 算法的主干网络、用 BiFPN-P2 改进

YOLOv8 算法的颈部结构、融合 EfficientViT 和 BiFPN-P2 对 YOLOv8 算法进行改进。这 3 种改进模型分别命名为 YOLOv8-EfficientViT、YOLOv8-BiFPN-P2 和 YOLOv8-EfficientViT-BiFPN-P2。在相同的模型训练条件下控制相关变量,并对 3 种改进模型进行训练。 P 、 R 和 $mAP50-95$ 随迭代次数的变化曲线如图 8 所示,模型最终的 R 、 $mAP50$ 、 $mAP50-95$ 、训练耗时和大小如表 4 所示。

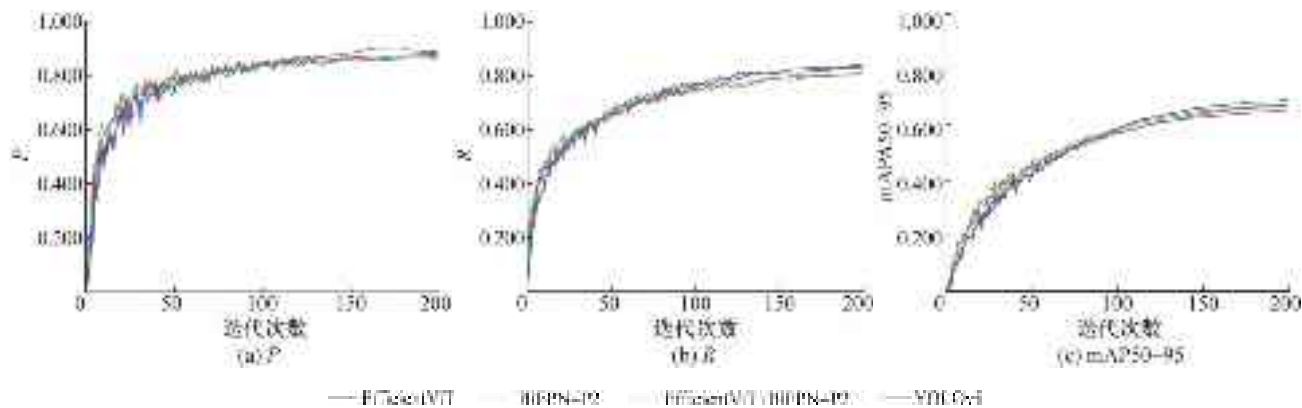


图 8 主干网络和颈部结构的改进对比试验

表 4 主干网络和颈部结构的改进对比试验的最终结果

模型名称	R	mAP50	mAP50-95	训练耗时/h	模型大小/MB
YOLOv8	0.809	0.858	0.674	1.750	6.3
YOLOv8-BiFPN-P2	0.809	0.852	0.653	2.929	4.9
YOLOv8-EfficientViT	0.841	0.883	0.689	2.669	17.9
YOLOv8-EfficientViT-BiFPN-P2	0.829	0.874	0.670	4.328	16.2

由各个模型的训练结果可知, YOLOv8-EfficientViT 在各项评价指标上的总体提升效果最明显。但由于算法改进, 模型参数增加, 因此与原始的 YOLOv8 算法相比, 所有模型的总体训练时间都变长, YOLOv8-BiFPN-P2 甚至出现了负优化。与 YOLOv8-BiFPN-P2 相比, YOLOv8-EfficientViT-BiFPN-P2 在性能上有所提升, 表明 YOLOv8 融合 EfficientViT 和 BiFPN-P2 可以提高小目标算法的相关性能, 但不及 YOLOv8 仅融合 EfficientViT。最终结果表明, 与 YOLOv8 算法相比, YOLOv8-EfficientViT 的 R 提升了 3.2%, $mAP50$ 提升了 2.5%, $mAP50-95$ 提升了 1.5%。综上所述, YOLOv8-EfficientViT 和 YOLOv8-EfficientViT-BiFPN-P2 在性能上均有提升, 因此能更准确地识别小目标裂缝。

3.3 综合改进对比试验

为验证本文改进 YOLOv8 模型的有效性, 本试验用 EfficientViT 替换 YOLOv8 主干网络, 并在基于颈部结构融合 BiFPN-P2 的基础上, 引入 LSKNet 注意力机制, 分别在改进算法的 Neck、Neck 与 Backbone 输出位置进行试验, 并分别命名为 YOLOv8 改进-Neck、YOLOv8 改进-Neck-Backbone 输出。在 3.2 节中, YOLOv8-EfficientViT 表现出色, 因此将 LSKNet 注意力机制引入, 也进行 2 组试验, 分别命名为 EfficientViT-Neck、EfficientViT-Neck-Backbone 输出。保持有关变量不变, 按照相同的模型训练条件对 2 种改进模型进行训练, 并将其与 YOLOv8、YOLOv8-EfficientViT 和 YOLOv8-EfficientViT-BiFPN-P2 进行比较。 P 、 R 和 $mAP50-95$ 随迭代次数的变化曲线如图 9 所示, 模型最终的 P 、 R 、 F_1 、 $mAP50$ 、 $mAP50-95$ 、训练耗时和大小如表 5 所示。

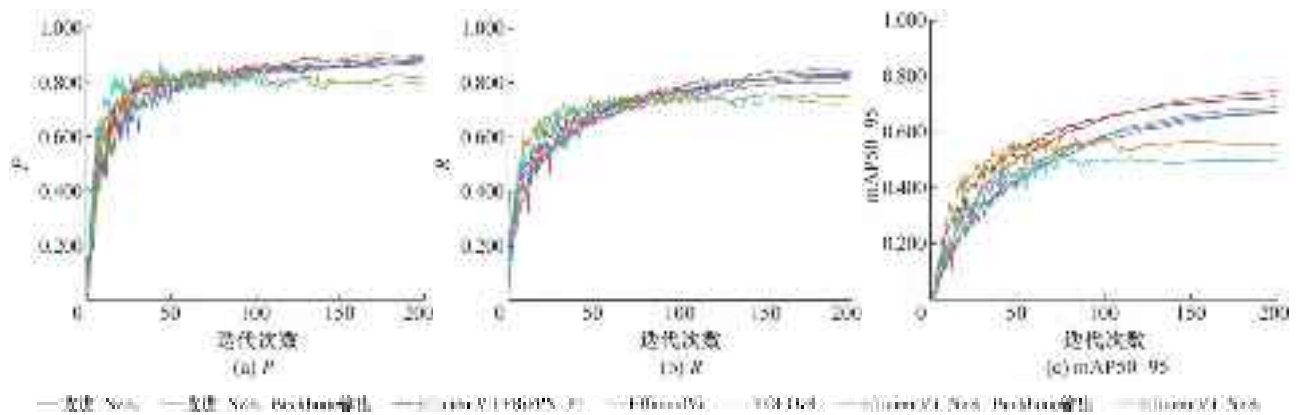


图 9 主干网络和颈部结构的改进最终对比试验

表 5 主干网络和颈部结构的改进对比试验的最终结果

模型名称	P	R	F_1	$mAP50$	$mAP50-95$	耗时/h	模型大小/MB
YOLOv8	0.860	0.809	0.834	0.858	0.674	1.750	6.3
EfficientVit+BiFPN-P2	0.878	0.829	0.853	0.874	0.670	4.328	16.2
EfficientVit	0.884	0.841	0.862	0.883	0.689	2.669	17.9
EfficientVit-Neck	0.799	0.727	0.757	0.800	0.516	2.625	17.9
EfficientVit-Neck-Backbone 输出	0.809	0.756	0.781	0.804	0.583	2.709	18.3
改进-Neck	0.870	0.826	0.847	0.880	0.728	5.022	17.5
改进-Neck-Backbone 输出	0.897	0.844	0.869	0.897	0.748	5.702	17.7

对比不同改进模型可知, YOLOv8 改进-Neck-Backbone 输出模型在各项评价指标上提升明显, 但算法复杂度增加会导致模型的构建时间和大小均不如原始的 YOLOv8 算法。与 YOLOv8 算法相比, 将主干网络替换为 EfficientViT, 融合 BiFPN-P2

结构进行小目标检测, 在 Neck 和 Backbone 输出部分引入 LSKNet 融合 C2f 模块注意力机制, 是提升效果最显著的改进方式, 其中: P 提升了 3.7%, R 提升了 3.5%, F_1 提升了 3.5%, $mAP50$ 提升了 3.9%, $mAP50-95$ 提升了 7.4%。由上述试验可知,

与 YOLOv8 算法相比,用 EfficientViT 网络替换 YOLOv8 主干网络,在颈部结构引入 LSKNet 融合 C2f 模块注意力机制,并增加 BiFPN 融合 P2 小目标检测头的模型性能提升较大,可准确识别小目标裂缝,能应用于实际应用中。

4 案例验证

本文以陕西省宝鸡市某斜拉桥为例,进行研究模型测试,如图 10 所示,旨在验证 YOLOv8 改进模型在实际工程应用中的有效性。



图 10 陕西省宝鸡市某斜拉桥

本文利用 DJI Mini 2 消费级轻型无人机采集桥梁整体图像。由于 DJI Mini 2 无人机属于轻型无人机,因此能随时随地采集桥梁数据。飞行路线采用“弓”字形,从桥底向桥顶拍摄。依据 DJI Mini 2 提供的精确定位信息,确保无人机移动路径与桥梁表面保持平行,并将镜头垂直于航线方向,图像采集参数如表 6 所示。最终获取了 2 224 张分辨率为 $4\,000\times 2\,250$ 像素的桥梁原始图像,这些图像既包含桥梁裂缝的图像,又包含无裂缝的图像。由于桥梁状况良好,因此,经过人工复核,仅 222 张图像包含裂缝,这些图像被用于制作实际验证样本数据集。

表 6 图像采集参数

设备/飞行参数	说明
无人机	DJI Mini 2
影像传感器	11.04 mm CMOS 像素 1 200 万
卫星导航系统	GPS+GLONASS+GALILEO
定焦镜头	独立镜头
焦距/mm	24
分辨率	4 000×2 250 像素
镜头光圈	f/2.8
拍摄时间/min	120
拍摄距离/m	3
重叠率	50%

本研究分别利用改进的 YOLOv8 模型和原始的 YOLOv8 模型对采集的桥梁图像进行实际裂缝预测,部分预测结果如图 11 所示。通过对比相同图像,本研究发现,与原始的 YOLOv8 算法相比,改进的 YOLOv8 算法更能准确地识别桥梁裂缝,尤其是像素占比较小的裂缝目标,避免了漏检情况。本研究还使用改进的 YOLOv8 模型和原始的 YOLOv8 模型对实际采集的桥梁裂缝图像进行了验证,部分验证结果如表 7 所示。由图 11 可知,本文所提改进 YOLOv8 算法在近景和远景图像中均表现出色,能准确识别桥梁裂缝。在实际桥梁应用中,与原始的 YOLOv8 算法相比,改进的 YOLOv8 算法的 R 提高了 3.5%, mAP_{50} 提高了 0.4%, mAP_{50-95} 提高了 5.1%,而且图像处理耗时缩短了 0.1 ms。改进 YOLOv8 算法能有效、快速、准确地检测桥梁裂缝,并能验证消费级无人机在桥梁裂缝监测方面的可行性。

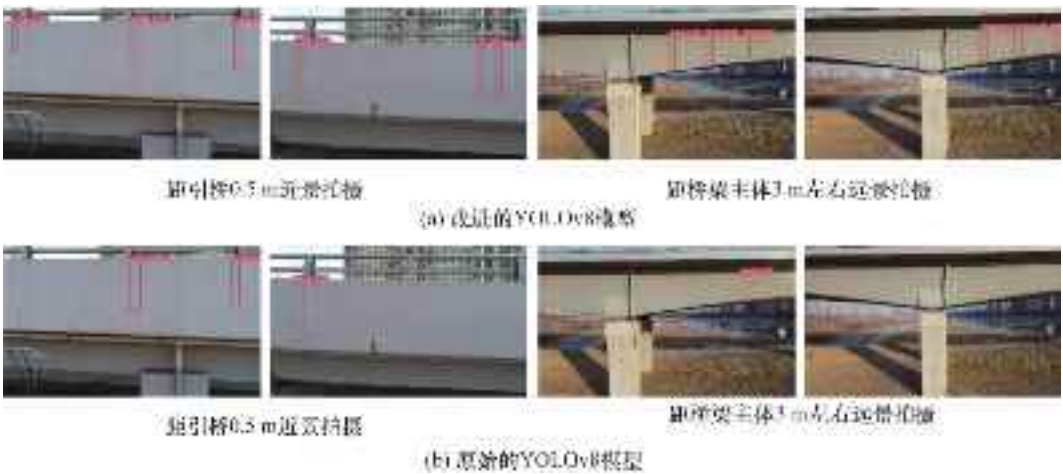


图 11 桥梁裂缝识别对比图

表 7 模型验证结果

模型名称	R	mAP50	mAP50-95	预处理/ms	后处理/ms
EfficientViT-LSKNet-BiFPN-YOLOv8	0.743	0.782	0.568	0.1	0.4
YOLOv8	0.708	0.778	0.517	0.2	0.4

5 结 论

本文基于 YOLOv8 算法提出一种改进的桥梁裂缝小目标检测算法。首先,用 EfficientViT 网络替换 YOLOv8 的主干网络,并用深度可分离卷积优化特征融合网络,以减少大量冗余参数,并增强图像局部特征的提取能力;其次,通过引入 LSKNet 融合 C2f 模块注意力机制,弥补非结构化环境造成的精度损失,并通过选择性调整卷积核尺寸,降低运行参数,减少计算复杂度;最后,增加 BiFPN 融合 P2 小目标检测头、不同尺度特征图之间的信息耦合和小目标裂缝检测框,以提高小目标裂缝特征信息提取效果。与 YOLOv8 算法相比,本文算法的准确率提升了 3.7%,召回率提升了 3.5%, F_1 提升了 3.5%,mAP50 提升了 3.9%,mAP50-95 提升了 7.4%,小目标裂缝的识别准确性得到较好提升。

通过无人机采集桥梁表观数据,可更高效、快捷地获取桥梁表观信息,与图像识别算法相结合,能解决桥梁健康监测中桥梁表观病害检测不足的问题。本文利用消费级无人机获取桥梁图像,既能降低桥梁的监测成本,又能满足图像的处理需求。

本文改进算法虽然能准确识别桥梁小目标裂缝,但模型参数的增加导致模型复杂度提升,从而使模型的训练耗时和大小大于原始 YOLOv8 算法。未来,本研究团队将继续深入研究小目标裂缝算法,以更加轻量化、高效的方式呈现模型内部的特征表示,解决算法误检和漏检等相关问题,将图像识别算法更好地应用于智能移动设备的使用场景中,使桥梁表观病害检测更便捷。

参考文献 (References)

[1] 《中国公路学报》编辑部. 中国桥梁工程学术研究综述·2021 [J]. 中国公路学报, 2021, 34(2): 1-97.
China Journal of Highway and Transport. Review on China's bridge engineering research: 2021 [J]. China Journal of Highway and Transport, 2021, 34(2): 1-97. (in Chinese)
[2] 交通运输部. 2023 年交通运输行业发展统计公报 [N]. 中国交通报, 2024-06-18(2).

Ministry of Transport. 2023 statistical bulletin on the development of the transport industry [N]. China Transportation News, 2024-06-18(2).
[3] KUŠTER MARIĆ M, MANDIĆ IVANKOVIĆ A, SRBIĆ M, et al. Assessment of performance indicators of a large-span reinforced concrete arch bridge in a multi-hazard environment [J]. Buildings, 2022, 12(7): 1046.
[4] 岳清瑞, 徐刚, 刘晓刚. 桥梁裂缝智能识别与监测方法研究 [J]. 中国公路学报, 2024, 37(2): 16-28.
YUE Q R, XU G, LIU X G. Crack intelligent recognition and bridge monitoring methods [J]. China Journal of Highway and Transport, 2024, 37(2): 16-28. (in Chinese)
[5] 刘宇飞, 冯楚乔, 陈伟乐, 等. 基于机器视觉法的桥梁表观病害检测研究综述 [J]. 中国公路学报, 2024, 37(2): 1-15.
LIU Y F, FENG C Q, CHEN W L, et al. Review of bridge apparent defect inspection based on machine vision [J]. China Journal of Highway and Transport, 2024, 37(2): 1-15. (in Chinese)
[6] 余加勇, 李锋, 薛现凯, 等. 基于无人机及 Mask R-CNN 的桥梁结构裂缝智能识别 [J]. 中国公路学报, 2021, 34(12): 80-90.
YU J Y, LI F, XUE X K, et al. Intelligent identification of bridge structural cracks based on unmanned aerial vehicle and Mask R-CNN [J]. China Journal of Highway and Transport, 2021, 34(12): 80-90. (in Chinese)
[7] 余加勇, 刘宝麟, 尹东, 等. 基于 YOLOv5 和 U-Net3+ 的桥梁裂缝智能识别与测量 [J]. 湖南大学学报(自然科学版), 2023, 50(5): 65-73.
YU J Y, LIU B L, YIN D, et al. Intelligent identification and measurement of bridge cracks based on YOLOv5 and U-Net3+ [J]. Journal of Hunan University (Natural Sciences), 2023, 50(5): 65-73. (in Chinese)
[8] SHI Y, CUI L M, QI Z Q, et al. Automatic road crack detection using random structured forests [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(12): 3434-3445.
[9] REN S 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, 2017, 39(6): 1137-1149.

- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: unified, real-time object detection [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016.
- [11] YU Z W, SHEN Y G, SHEN C K. A real-time detection approach for bridge cracks based on YOLOv4-FPM [J]. Automation in Construction, 2021, 122: 103514.
- [12] KUMAR P, BATCHU S, SWAMY S N, et al. Real-time concrete damage detection using deep learning for high rise structures [J]. IEEE Access, 2021, 9: 112312–112331.
- [13] YANG Z, LI L, LUO W T. PDNet: Improved YOLOv5 nondeformable disease detection network for asphalt pavement [J]. Computational Intelligence and Neuroscience, 2022, 2022: 5133543.
- [14] TENG S, LIU Z C, LI X D. Improved YOLOv3-based bridge surface defect detection by combining high- and low-resolution feature images [J]. Buildings, 2022, 12(8): 1225.
- [15] WU Y, HAN Q B, JIN Q L, et al. LCA-YOLOv8-Seg: An improved lightweight YOLOv8-Seg for real-time pixel-level crack detection of dams and bridges [J]. Applied Sciences, 2023, 13(19): 10583.
- [16] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8 [Z/OL]. (2023-05-17) [2024-03-15]. <https://arxiv.org/abs/2305.09972>.
- [17] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: Task-aligned one-stage object detection [C]// Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021.
- [18] CAI H, LI J Y, HU M Y, et al. EfficientViT: Lightweight multi-scale attention for high-resolution dense prediction [C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023.
- [19] Alexey D. An image is worth 16×16 words: Transformers for image recognition at scale [Z/OL]. (2020-10-22) [2024-03-15]. <https://arxiv.org/abs/2010.11929>.
- [20] LI Y X, HOU Q B, ZHENG Z H, et al. Large selective kernel network for remote sensing object detection [C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023.
- [21] ZHANG H L, DU Q F, QI Q Y, et al. A recursive attention-enhanced bidirectional feature pyramid network for small object detection [J]. Multimedia Tools and Applications, 2023, 82(9): 13999–14018.
- [22] CHEN L W, YANG J J. A lightweight YOLOv5-based model with feature fusion and dilation convolution for image segmentation [J]. Mathematics, 2023, 11(16): 3538.
- [23] PATIL S, DAVE G, URMODE S. A comparative analysis of YOLOv8 and YOLOv5 for nut thread classification-deep learning approach [J]. International Journal for Research in Applied Science & Engineering Technology, 2024, 12(2): 176–182.

Bridge small target crack detection based on improved YOLOv8

LI Jinpei^{1, 2, 3}, MENG Xiaolin^{4, 5}, HU Liangliang^{1, 2, 3}, BAO Yan^{1, 2, 3}, ZHAO Shiyu^{1, 2, 3}

(1. College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China;

2. National Key Laboratory of Bridge Safety and Resilience, Beijing University of Technology, Beijing 100124, China;

3. The Key Laboratory of Urban Security and Disaster Engineering of the Ministry of Education, Beijing University of Technology, Beijing 100124, China;

4. School of Instrument Science and Engineering, Southeast University, Nanjing 211189, China;

5. Faculty of Engineering, Imperial College London, London SW7 2AZ, UK)

Abstract: [Objective] The structural integrity of bridges is a critical concern as infrastructure ages, necessitating the development of reliable methods for detecting potential failures. Among these, the identification of small target cracks is particularly important, as these cracks often grow undetected until they result in severe damage. Traditional inspection methods, such as manual visual inspections, are hindered by their labor-intensive nature and susceptibility to human error, often resulting in the oversight of small but significant defects. Recent advancements in computer vision and deep learning technologies offer new opportunities to improve the accuracy and efficiency of bridge inspections. This study introduces an innovative approach for detecting small target cracks in bridge structures by employing an enhanced version of the You Only

Look Once (YOLOv8) object detection model, a widely recognized algorithm known for its rapid processing capabilities and high detection accuracy. The enhanced YOLOv8 model is tailored to detect small-scale cracks on bridge surfaces that may not be easily identifiable by traditional inspection methods or earlier versions of computer vision models. [Methods] The proposed algorithm modifies the standard YOLOv8 model to address the specific challenges associated with detecting small cracks on bridge surfaces. A key modification is the integration of efficient vision transformer (EfficientViT) into the backbone of the YOLOv8 model. EfficientViT is an advanced transformer-based architecture that reduces redundant parameters and optimizes the extraction of local features from high-resolution images, enabling more precise detection of subtle crack features. This enhancement is crucial, as small cracks often exhibit low contrast against their background and may be easily overlooked by less sophisticated models. In addition to EfficientViT, the proposed algorithm also incorporates large selective kernel network (LSKNet) within the C2f module of YOLOv8. LSKNet employs a dynamic kernel selection mechanism that allows the model to adaptively adjust the size of the convolutional kernels based on the input features, making it highly suitable for detecting cracks of varying sizes, orientations, and morphological characteristics. This adaptability ensures that the model can detect small cracks, regardless of their form. Furthermore, the model uses bidirectional feature pyramid network (BiFPN) to merge feature maps at different scales. Traditional models struggle with detecting small targets due to the loss of critical information during downsampling operations. BiFPN mitigates this issue by preserving high-resolution feature maps across multiple layers, enhancing the model's ability to detect small cracks that would otherwise be missed. The combined effect of these modifications improves the accuracy of small target crack detection while maintaining computational efficiency. [Results] The effectiveness of the proposed model was validated using a dataset of crack images from a specific bridge, captured by unmanned aerial vehicles (UAVs). UAVs provided detailed images from areas that were often difficult or dangerous to access using traditional inspection methods. The experimental results demonstrated that the enhanced YOLOv8 model significantly outperformed the original version in terms of key performance metrics. Specifically, the modified model achieved improvements of 3.7%, 3.5%, 3.5%, 3.9%, and 7.4% in terms of the detection precision, recall, F_1 score, mAP50, and mAP50-95, respectively. These results indicated a substantial improvement in the model's ability to detect small cracks that often had low contrast and irregular shapes, which were typical characteristics of cracks on bridge surfaces. Furthermore, compared to conventional methods, the proposed model was able to detect cracks with higher precision and fewer false positives, making it a promising tool for improving the efficiency of bridge inspections. [Conclusions] In conclusion, the improved YOLOv8 algorithm introduced in this study represents a significant advancement in the detection of small target cracks in bridge structures. The modifications made to the original YOLOv8 model, including the integration of EfficientViT, LSKNet, and BiFPN, result in a more accurate and computationally efficient model for crack detection. This approach offers a practical and scalable solution for the widespread application of bridge health monitoring, particularly in areas that are difficult to inspect using traditional methods. By leveraging advanced surface data processing techniques, this research contributes to the development of modern methods for assessing the health of bridge structures, ultimately helping to ensure the safety and longevity of infrastructure systems.

Key words: bridge structure health monitoring; unmanned aerial vehicle; small target crack detection; YOLOv8; image processing

(责任编辑 金延秋)