

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2025)11-3564-19

论文引用格式: Peng Y and Shao Y F. 2025. Lightweight pavement defect detection algorithm combining partial convolution and inception depthwise convolution. Journal of Image and Graphics, 30(11):3564-3582(彭毅, 邵宇飞. 2025. 结合部分卷积与初始深度卷积的轻量级路面缺陷检测算法. 中国图象图形学报, 30(11):3564-3582)[DOI:10.11834/jig.240727]

结合部分卷积与初始深度卷积的轻量级路面缺陷检测算法

彭毅, 邵宇飞*

辽宁工程技术大学软件学院, 葫芦岛 125000

摘要: 目的 路面缺陷检测是道路维护和管理的重要环节, 人工智能技术能够极大地提升路面缺陷检测的性能和效率。为解决当前路面缺陷检测算法难以满足在计算资源受限设备上实时检测的问题, 本文基于YOLOv8 (you only look once version 8) 目标检测框架, 结合部分卷积 (partial convolution) 与初始深度卷积 (inception depthwise convolution) 机制提出一种轻量级的路面缺陷检测算法 YOLOv8n-PIVI。方法 该算法在主干网络引入 PartialBlock 降低模型参数量, 同时引入 IDBlock (inception depthwise block) 丰富网络的特征提取能力并进一步降低计算复杂度; 在特征融合网络引入 VanillaBlock 减少网络参数的同时优化网络特征融合能力; 引入 ID-Detect (inception depthwise detect) 检测头, 极大降低模型的计算复杂度并优化检测头对不同尺度特征的解析能力。结果 在 Pothole Dataset 数据集上的实验结果显示, 本文算法的 mAP50 (mean average precision at 50% intersection over union) 达到 0.55, 较基线算法提升 3.5 个百分点, 帧速率 (frames per second, FPS) 达到 243, 较基线算法提升 43, 并且模型参数量和计算量仅为基线的 67% 和 72%, 模型参数文件由 6.3 MB 降低至 4.3 MB, 内存占用较原模型降低了近 1/3。在与目标检测领域经典算法的对比实验中, 本文算法在同参数量级模型中检测精度、计算复杂度及 FPS 较对比算法都有一定优势。此外, 在 RDD2022 (road damage detector 2022)-China-MotorBike 数据集和 VOC2012 (visual object classes 2012) 数据集上的实验结果表明, 本文算法具有良好的适应性和鲁棒性, 能够适应不同的目标检测任务。结论 综合实验结果表明, 本文算法能够以轻量级的计算资源需求, 实现更高的目标检测性能, 并且具有良好的适应性和鲁棒性, 有助于降低路面缺陷检测算法的使用限制, 拓宽路面缺陷检测系统的适用场景。

关键词: 缺陷检测; 路面缺陷检测; YOLOv8; 轻量级模型; 部分卷积; 初始深度卷积

Lightweight pavement defect detection algorithm combining partial convolution and inception depthwise convolution

Peng Yi, Shao Yufei*

School of Software, Liaoning Technical University, Huludao 125000, China

Abstract: Objective The acceleration of global urbanization and the rapid development of intelligent transportation has highlighted the importance of road construction and maintenance in urban management. With the rapid growth of urban

收稿日期: 2024-12-06; 修回日期: 2025-03-06; 预印本日期: 2025-03-13

* 通信作者: 邵宇飞 yfshao@alum.imr.ac.cn

基金项目: 国家自然科学基金项目 (62173171)

Supported by: National Natural Science Foundation of China (62173171)

population and the increase in the number of motor vehicles, road facilities are facing great pressure. Road surface defects, such as cracks and potholes, are prone to occur under the action of multiple factors such as frequent vehicle crushing and climate change. These defects not only affect driving comfort but also potentially pose serious traffic safety risks. Pavement defect detection is an important part of road maintenance and management and is related to traffic safety and road service life. Traditional pavement defect detection methods have disadvantages such as low efficiency, poor security, strong subjectivity, and insufficient accuracy. The development of computer technology and artificial intelligence has encouraged researchers to apply machine vision and deep learning technology to pavement defect detection, which greatly improved the performance and efficiency of pavement defect detection. However, due to the large amount of parameters and calculation, the traditional deep learning model has difficulty meeting the real-time detection requirements on embedded devices and edge devices with limited hardware resources. This complexity limits the promotion and application of deep learning technology in the field of road defect detection. This study aims to reduce the computational resource requirements and optimize the target detection performance of the pavement defect detection algorithm. Based on the YOLOv8 (you only look once version 8) object detection framework, along with partial convolution and inception depthwise convolution mechanism, a lightweight pavement defect detection algorithm called YOLOv8n-PIVI is proposed.

Method First, the PartialBlock module and IDBlock (inception depthwise block) module are introduced into the backbone network to reduce the number of parameters of the backbone network and improve the processing capability of the network for features of different scales. While introducing the lightweight advantage of partial convolution, the PartialBlock module also supplements the 1×1 convolution to extract the overall information and arrange features. This convolution not only makes up for the lack of feature analysis capability of partial convolution to a certain extent but also helps partial convolution learn a more effective spatial feature extraction method. The IDBlock module inherits the advantages of partial lightweight and deeper feature analysis of inception depthwise convolution, and it performs a more comprehensive analysis of features at a smaller cost through the traditional 1×1 convolution. In addition, PartialBlock module and IDBlock module contain a MaxPool module in the tail to expand the receptive field of network features and reduce the number of network parameters. The organic combination of the PartialBlock module and the IDBlock module helps reduce the number of backbone network parameters and optimize the feature extraction capability of the backbone network. Second, inspired by the excellent feature parsing capability and the applicability in resource-constrained scenarios of VanillaNet, we introduce VanillaBlock, which is the core module of VanillaNet, into the feature fusion network of YOLOv8n. This incorporation aims to optimize the feature fusion capability of the model and reduce the computational burden of the model. Finally, the inception depthwise convolution module with low overhead and multi-scale feature parsing capability is introduced to replace the two series 3×3 convolutions in the original detection head by combining the 1×1 convolution with low overhead. This way reduces the computational cost of the detection head and improves its parsing capability of different scale features. A more lightweight and efficient ID-Detect (inception depthwise detect) is proposed to make the detection head lighter and further improve the performance of the detection head.

Result The comparison and ablation experiments are conducted on the Pothole Dataset, which consists of 1 784 still images taken by vehicle-mounted cameras and other mobile devices. The dataset is divided into a training set (1 265 images), a validation set (401 images), and a test set (118 images) with an 11:3:1 ratio. To verify the effectiveness of the improved model in optimizing the pavement defect detection effect, the improved model is compared with some classical algorithms in the field of object detection, including Centernet, RetinaNet, Faster R-CNN (faster region-based convolutional neural networks), SSD (single shot multibox detector), Efficientdet, and YOLOv8. Experimental results show that the Precision and Recall of the optimization algorithm proposed in this study are improved to a certain extent compared with the baseline, and the mAP50 (mean average precision at 50% intersection over union) reaches 0.55. It is 3.5 percentage points higher than 0.515 of the baseline algorithm. The FPS (frames per second) reaches 243, which is 43 higher than the baseline algorithm. The parameter number of the model is reduced from 3.01×10^6 to 2.01×10^6 , which is only 67% of the baseline. The calculation amount is reduced from 8.1 GFLOPs (giga floating point of operations per second) to 5.9 GFLOPs, which is only 72% of the baseline. The model parameter file is reduced from 6.3 MB to 4.3 MB, and the memory occupation is decreased by nearly one-third compared with the original model. Compared with the classical algorithms in the field of object detection, the proposed algorithm has certain advantages in detection accuracy, computa-

tional complexity, and FPS compared with other algorithms in the model of the same parameter magnitude. In the ablation experiment, the PartialBlock module, IDBlock module, VanillaBlock module, and ID-Detect module introduced in this study are individually added to the baseline model and removed from the final model. The effectiveness of the modules introduced in this study in improving the performance of the baseline algorithm for pavement defect detection is verified. In the subsequent comparative analysis, the improvement of the optimization algorithm proposed in this study compared with the baseline algorithm in pavement defect detection performance is shown in a more detailed and intuitive way through the comparison of experimental result index parameters, the comparison of training process index changes, and the comparison of pavement pothole detection effects. Subsequently, we also design experiments to roughly simulate the performance of the pavement defect detection algorithm proposed in this study in the actual situation, and the shortcomings of the algorithm are analyzed. In addition, we compare the proposed algorithm and the baseline algorithm on the RDD2022 (road damage detector 2022–China-MotorBike) dataset and VOC2012 (visual object classes 2012) dataset to analyze the robustness of the algorithm. Experimental results show that the proposed algorithm can adapt to the multi-type road defect detection tasks in China on the RDD2022–China-MotorBike dataset and the multi-class daily object detection tasks on the VOC2012 dataset. Meanwhile, it shows better object detection performance than the baseline model. Therefore, the proposed algorithm can adapt to different object detection tasks, has good adaptability and robustness, and is expected to be applicable to highly complex and diverse road defect detection scenarios. **Conclusion** The comprehensive experimental results show that the optimized algorithm YOLOv8n-PIVI proposed in this study can achieve high object detection performance with lightweight computing resource requirements. The model parameter file size is as low as 4.3 MB. The algorithm also has good adaptability and robustness, which help reduce the application limitation of road defect detection algorithms and broaden the applicable scene of pavement defect detection systems.

Key words: defect detection; pavement defect detection; YOLOv8; lightweight model; partial convolution; inception depthwise convolution

0 引言

随着全球城市化进程的加速推进以及智能化交通的快速发展,道路建设和维护成为城镇管理的重要组成部分。城镇人口的快速增长和机动车辆的增多,使得道路设施面临巨大的压力。路面缺陷,如裂缝、坑洼等,在车辆频繁碾压和气候变化(如温度变化、雨水侵蚀)等多重因素作用下极易产生,这些缺陷不仅影响驾驶舒适度,还潜藏着严重的交通安全隐患。

路面缺陷检测是道路维护和管理的重要环节,关系到行车安全和道路使用寿命。传统的路面缺陷检测方法主要依靠人工巡查,这种方式存在效率低下、安全性差、主观性强以及精度不足等弊端(Zakeri等,2017)。随着计算机技术和人工智能的发展,研究者们逐渐将机器视觉和深度学习的技术应用于路面缺陷检测(Gopalakrishnan,2018)。早期的基于图像处理的路面缺陷检测方法,通常包括图像预处理、特征提取和缺陷识别等步骤。这些方法在简单背景下(如均匀光照条件)能够取得较好的效果,但在复

杂环境下(如有阴影、杂物遮挡等)往往失效严重(Spencer等,2019)。近年来,深度学习技术在计算机视觉领域的成功应用,为路面缺陷检测提供了新思路 and 有效解决方案。深度学习模型,尤其是卷积神经网络(convolutional neural network, CNN)(Liang等,2021)、区域提议网络(regional proposal network, RPN)(Cai等,2018)和注意力机制(Carion等,2020;李颖等,2023)等,显示出强大的特征提取和目标检测能力,有助于快速、准确地识别路面缺陷。然而,传统的深度学习模型(如ResNet(residual network)(He等,2016)、Inception(Szegedy等,2015)等)虽然在图像分类和目标检测任务上表现出色,但由于计算量大,通常只能在桌面或服务器级别的图形处理器(graphics processing unit, GPU)上运行,无法满足硬件资源受限的嵌入式设备及边缘设备的实时检测需求。为此,研究者开始探索轻量级(Wan等,2022;曾怡峰等,2023;李朝辉等,2023;王红霞等,2023;卓文琦等,2023)的检测算法,这类算法在保证检测精度的前提下,大幅减少计算量和模型大小,使其能够在低计算资源设备上实现实时检测。

目前主流的目标检测算法可按是否存在候选区

域分为两阶段目标检测算法和一阶段目标检测算法两大类(Weng 和 Zhang, 2023)。两阶段的目标检测算法首先生成候选区域,然后对这些区域进行分类和边界框回归,其特点是精度较高而实时性较差,其中典型的算法包括 R-CNN(region-based convolutional neural network)系列(R-CNN(Girshick 等, 2014)、Fast R-CNN(Girshick, 2015)、Faster R-CNN(Ren 等, 2017)、Mask R-CNN(He 等, 2020))和 R-FCN(region-based fully convolutional network)(Dai 等, 2016)等;一阶段目标检测算法则不单独生成候选区域,而直接在网络中提取特征来预测物体的位置和类别,其特点是推理速度快但损失了部分精度,其中典型的算法包括 SSD(single shot multibox detector)(Liu 等, 2016)和 YOLO(you only look once)系列(YOLOv5、YOLOv7、YOLOv8 等)(Redmon 等, 2016; Cong 等, 2023; Terven 等, 2023; 赖杰 等, 2024; 周涛 等, 2024; 肖振久 等, 2024)。鉴于实时目标检测模型对于推理速度及轻量化的需求,研究者更多选择推理速度更快的一阶段的目标检测算法作为基准并对其相关优化开展进一步的研究。

针对复杂场景下道路病害检测模型占用内存空间大、计算复杂度高和检测速度难以满足实时目标检测要求等问题,胥铁峰等人(2024)在 YOLOv8 中使用可变形卷积和通道剪枝等技术,在提升模型检测精度的同时降低了模型的参数量和计算量,实现了模型轻量化。Zhang 等人(2023)用坐标注意力和 SPD-Conv(space to depth convolution)模块优化了 YOLOv5 的特征提取网络并对其多尺度检测头的结构进行改进,实现了模型的轻量化和模型检测能力的提升。王春梅和刘欢(2024)通过在 YOLOv8 骨干网络引入轻量级网络 VanillaNet 并在特征提取网络中引入 SPD(space-to-depth)模块,使模型在保持高检测精度的同时所需的计算资源更少。梁礼明等人(2024)基于 YOLOv8 采用轻量级的 EfficientViT(efficient vision transformer)网络作为骨干网络结合 HWD(haar wavelet downsampling)构建了一种轻量级钢材缺陷检测算法。陈建瑜等人(2023)通过在 YOLOv5 中引入 CBAM(convolutional block attention module)注意力机制并使用 K-means 算法和 1-IoU(1-intersection over union)优化锚框,实现了对路面缺陷更高精度的快速检测。顾宙瑜等人(2023)提出的轻量化网络通过 HEB(hybrid enhancement block)和

SPP(spatial pyramid pooling)模块的改进,显著提升实时监控中的目标检测能力。魏峰等人(2024)以轻量级网络 ShuffleNetv2 替换 YOLOv5 主干网络并引入 SENet(squeeze-and-excitation networks)注意力机制,提出一种适用于移动端的轻量级目标检测模型。Qin 等人(2022)在骨干网络中引入 EfficientNet,通过 MFF(multi-layer feature fusion)对多层特征进行融合,提出一种轻量级缺陷检测算法。Wang 等人(2023)将 DCAM(dilated convolution aggregation module)作为特征增强网络的特征学习结构,并用 DCM(dilated convolution and channel attention fusion module)和 DSM(dilated convolution and spatial attention fusion module)作为检测头中回归和分类任务的特征增强模块,降低了模型的计算量和复杂度,并一定程度上提升模型的检测精度。为了提高道路缺陷检测精度,章东平等人(2024)在 Transformer 的基础上,通过构建动态可变形卷积模块以及引入可变形注意力机制和轻量化模块,降低了模型的计算量和参数量并优化了模型的检测效果。针对现有道路缺陷检测模型特征提取不完善以及检测设备计算资源的限制, Du 和 Jiao(2022)在 YOLOv5 的基础上结合了 BiFPN(bidirectional feature pyramid network)和 PANet(path aggregation network),不仅增强了模型的特征提取能力,还大幅减少了内存占用,极大地提高在嵌入式设备上的实用性。

在路面缺陷检测中,日常复杂多变情况下的检测场景对模型的检测能力及稳定性提出挑战,而模型的轻量化有助于扩展检测算法的应用场景以及降低路面缺陷检测系统的计算资源需求及应用成本, YOLOv8 系列算法中最小的 YOLOv8n 模型参数量为 3.1×10^6 , 模型计算量为 8.9 GFLOPs(giga floating point of operations per second),这样的计算资源需求较边缘设备而言仍处于较高水平,并且 YOLOv8n 在不增加模型参数的情况下仍难以满足日常复杂环境中对于路面缺陷检测的精度要求。

基于上述研究及模型现存的局限与面临的挑战,本文以 YOLOv8n 为基础提出优化算法 YOLOv8n-PIVI,该算法于主干网络引入 PartialBlock 模块及 IDBlock(inception depthwise block)模块,降低模型参数量的同时强化模型的特征提取能力;于颈部网络引入轻量级 VanillaBlock 模块优化模型的特征融合能力,提出 ID-Detect(inception depthwise

detect)检测头降低模型的参数量和计算量,并强化头部网络的多尺度特征解析能力。

综上所述,本文的主要贡献如下:1)基于YOLOv8n提出一种轻量级的路面缺陷检测算法YOLOv8n-PIVI,降低路面缺陷检测模型参数量与计算量的同时提高模型检测精度及帧速率(frames per second, FPS),使其能够满足更广泛使用场景的需求。2)在特征提取网络中引入PartialBlock模块及IDBlock模块,通过两者的有机组合,在减少网络参数的同时增强网络的特征提取能力。3)在特征融合网络中引入轻量级VanillaBlock模块,促进网络以更少的参数实现更高效的特征融合。4)提出ID-Detect检测头,在优化检测头结构使其更轻量化的同时,增强其对不同尺度特征的解析能力。

1 算法描述

1.1 基线算法:YOLOv8n

YOLO系列算法是当下比较流行的一阶段目标检测算法,YOLOv8是YOLO系列的一个版本,由Ultralytics公司在2023年发布,它继承了YOLO系列的高效实时处理能力,并在速度和准确性上较之前版本有了显著的提升。YOLOv8的网络结构如图1

所示,图中虚线框内为网络的局部结构,实线框为网络中部分模块的具体结构,箭头表示数据流动,后续结构图采用相同表示方式。网络的主要结构可分为主干网络(backbone)/特征提取网络、颈部网络(neck)/特征融合网络、头部网络(head)/预测输出层3个部分。主干网络主要负责提取输入图像的特征,输入的图像通过网络转化为便于后续环节使用的多层特征图;颈部网络负责将主干网络提取的特征进行融合,并生成适合头部网络输入的高质量特征图;头部网络负责将颈部网络输出的特征进一步处理,将特征图转换为目标检测、分类和关键点检测任务所需要的具体信息,进而生成最终的检测输出,包括边界框坐标、类别和置信度分数。

目前,YOLOv8有YOLOv8n、YOLOv8s、YOLOv8l、YOLOv8m和YOLOv8x共5个子类。这5个子类在网络结构上是相同的,但特征图通道数(width)与网络深度(depth)有所不同。其中,YOLOv8n拥有最小的参数量和计算量,其检测精度也相应最低,而YOLOv8x的参数量和计算量最大且检测精度最高。鉴于本文研究对于轻量化的需求,选择参数量和计算量较低的YOLOv8n作为基线算法。

1.2 本文所提出算法:YOLOv8n-PIVI

针对YOLOv8n存在的模型参数量、计算量难以

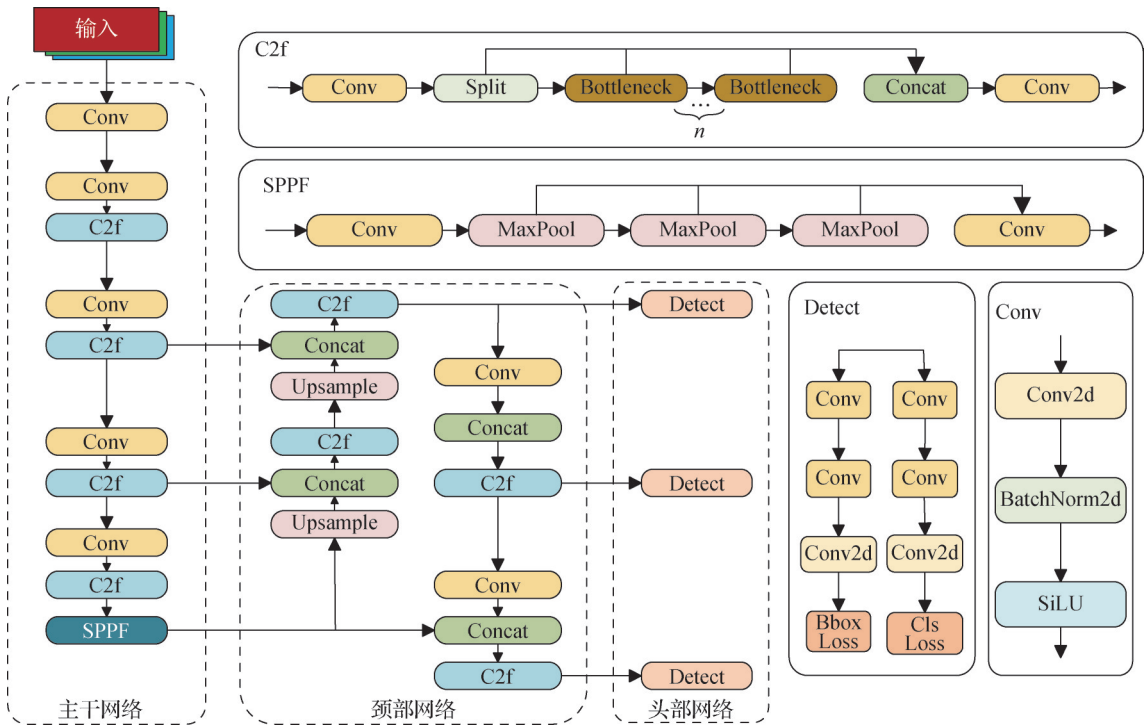


图1 YOLOv8结构图

Fig. 1 Diagram of YOLOv8 structure

适配算力有限设备以及特征提取与解析能力不足导致的检测精度不足的问题,本文提出优化后的轻量级算法 YOLOv8n-PIV1,模型结构如图2所示,虚线网络内的深橙色模块为本文相较于基准模型(图1)做出修改的模块,实线框内为部分修改模块的具体结构。该算法在YOLOv8n的基础上进行优化,首先,在主干网络中引入 PartialBlock 模块,减少模型的参数

量并优化特征提取效率,并引入 IDBlock 模块在降低参数量与计算量的同时提高模型对不同尺度特征的处理能力;其次,在特征融合网络中引入 VanillaBlock 模块,使模型轻量化的同时优化模型的特征融合能力;最后,提出轻量高效的 ID-Detect 模块优化检测头,在使模型更轻量化的同时进一步提升模型的检测性能。

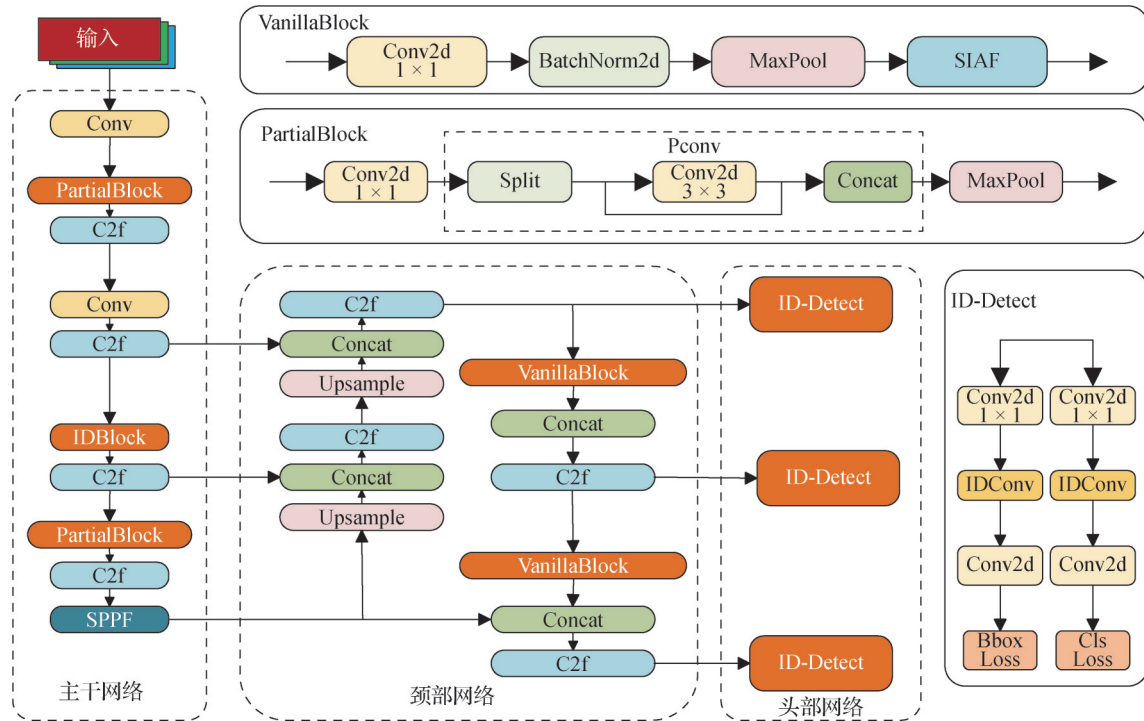


图2 YOLOv8n-PIV1结构图

Fig. 2 Diagram of YOLOv8n-PIV1 structure

1.2.1 PartialBlock 模块

为了减轻模型的计算资源需求,提升模型特征提取能力,本文在基线 YOLOv8n 模型的特征提取网络中引入部分卷积(partial convolution, PConv)(Chen 等, 2023b),其示意图如图3所示,并在其基础上构建 PartialBlock 模块,结构如图4所示。PartialBlock 模块由1个 1×1 Conv2d、1个 PConv 模块以及1个 MaxPool 最大池化块构成,首先通过 1×1 Conv2d 以较少计算资源保留更多特征信息并完成通道特征变换;接着通过 PConv 仅对部分输入通道进行卷积提取空间特征,从而减少冗余计算和内存访问,并且 1×1 Conv2d 和 PConv 的结合有助于让模型学习识别和选择对目标检测更有效的通道特征进行空间特征提取,从而优化模型的特征提取能力;最后在 PConv 模块后进行 MaxPool 最大池化,增大像素感受

野并减少网络的参数量(Boureau 等, 2010)。

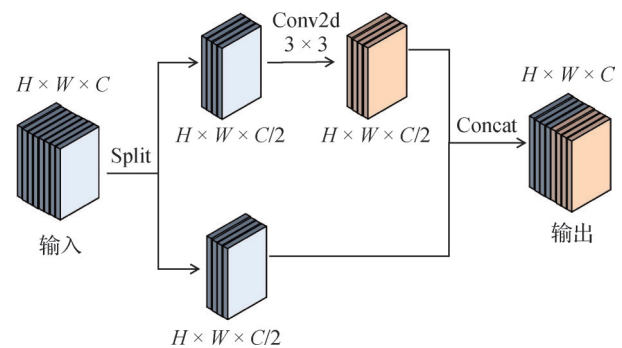


图3 PConv 示意图

Fig. 3 Schematic diagram of PConv

假设 h 和 w 为卷积核的高和宽, C_1 和 C_2 为输入通道数和输出通道数,省略偏置项,可以计算出传统卷积和 PConv 卷积(div 为 PConv 中划分输入通道的参数,用以调整部分卷积的程度,当 $div = n$ 时,代表模块

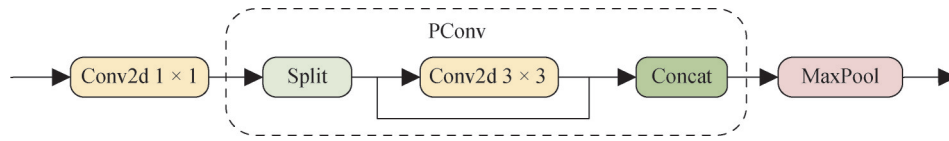


图4 PartialBlock 结构图

Fig. 4 Diagram of PartialBlock structure

仅在 $1/n$ 输入通道上应用常规卷积,而其余通道保持不变)的参数量,具体为

$$Params_{Conv} = C_1 \times h \times w \times C_2 = hwC_1C_2 \quad (1)$$

$$Params_{PConv} = \frac{C_1}{n} \times h \times w \times \frac{C_2}{n} = \frac{Params_{Conv}}{n^2} \quad (2)$$

式中, $Params_{Conv}$ 和 $Params_{PConv}$ 分别为传统卷积的参数量与 PConv 的参数量。当 $div = n$ 时, PConv 的参数量相较于传统卷积而言缩小了 n^2 倍,体现了 PConv 相较于传统卷积在参数量上的优势。

本文构建的 PartialBlock 模块在引入 PConv 轻量化优势的基础上,补充 1×1 Conv2d 对整体信息进行提取及特征编排,在一定程度上弥补 PConv 特征解析能力不足的同时协助 PConv 学习更有效的空间特征提取方式,并通过 MaxPool 扩大网络特征感受野以及减少网络参数量。相较于传统卷积, PartialBlock 能以更少的参数量完成更高效的特征提取。PartialBlock 相较于 YOLOv8n 模型中的 Conv 在模块参数量上的比较为

$$\frac{Params_{PartialBlock}}{Params_{Conv}} = \frac{C_1 \times 1^2 \times C_2 + \frac{C_2}{2} \times 3^2 \times \frac{C_2}{2}}{C_1 \times 3^2 \times C_2} \quad (3)$$

式中, $Params_{Conv}$ 为 YOLOv8n 中 Conv 卷积块的参数量, $Params_{PartialBlock}$ 为本文构建的 PartialBlock 模块的参数量, C_1 和 C_2 分别对应输入通道数和输出通道数。当 $C_2 = 2C_1$ 时, $\frac{Params_{PartialBlock}}{Params_{Conv}} = \frac{11}{18}$, PartialBlock 相较于 Conv 在参数量上有一定优势。

1.2.2 IDBlock 模块

受 InceptionNeXt(Yu 等, 2024) 启发, 从中引入模块复杂度低、具有更大感受野且有助于提取更丰富特征的初始深度卷积(inception depthwise convolution, IDConv), 以增强模型主干网络对不同尺度特征的提取能力并提高模型的计算效率。IDConv 通过将昂贵的大核深度卷积分解为 4 个并行的分支并对这些分支沿着通道维度进行深度可分离卷积完成对特征的解析, 如图 5 所示, 该操作结合了

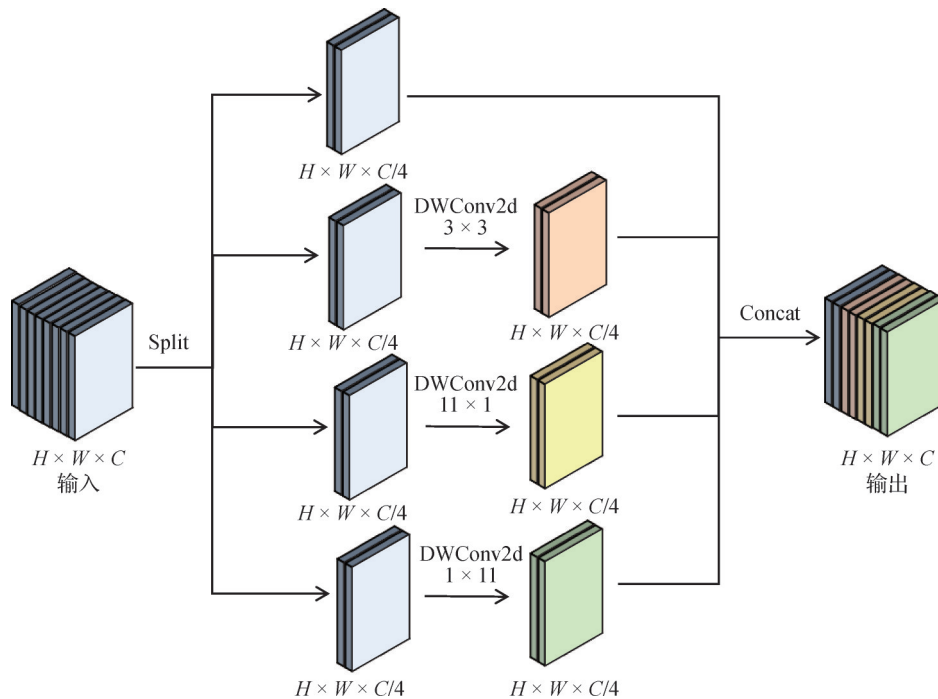


图5 IDConv 示意图

Fig. 5 Schematic diagram of IDConv

Inception(Szegedy等,2016)模块的多尺度卷积特性和深度可分离卷积(depthwise separable convolution, DSC)(Liu等,2022)的计算效率优势,通过这种方式,IDConv模块能够在保持或提高网络性能的同时,实现更紧凑和高效的模型设计。结合了深度可分离卷积,IDConv在计算资源需求上相较于传统卷积有巨大优势。为简单起见,假设输入和输出通道相同,省略偏置项, k 、 C 、 H 和 W 分别表示卷积核大小、通道数、特征的高度和宽度,不同类型卷积的复杂度计算如表1所示(Yu等,2024)。由表中计算可知,传统卷积和深度卷积的参数量和计算量与卷积核大小 k^2 成正比,相比之下,IDConv的参数量和计算量则与 k 呈线性关系;此外,传统卷积的参数量与计算量和 C^2 成正比,而深度卷积和IDConv的参数量和计算量则与 C 呈线性关系,这进一步凸显了IDConv在计算资源需求上相较于传统卷积存在的巨大优势。

本文在IDConv的基础上构建了IDBlock模块,其结构如图6所示。IDBlock(inception depthwise block)

表1 不同类型卷积的复杂度

Table 1 Complexity of different types of convolutions

卷积类型	参数量	计算量/FLOPs
conventional conv.	k^2C^2	$2k^2C^2HW$
depthwise conv.	k^2C	$2k^2CHW$
inception dep. conv.	$(2k+9)C/8$	$(2k+9)CHW/4$

模块由1个 1×1 Conv2d、1个IDConv以及1个Max-Pool最大池化块构成,首先通过 1×1 Conv2d完成初步特征提取和通道转换,接着IDConv对特征进行多尺度的深度解析, 1×1 Conv2d和IDConv的结合有助于强化模型对特征通道的编排及解析,进而强化模型的特征解析能力,最后经过MaxPool完成下采样。

本文构建的IDBlock在引入IDConv轻量化及对特征更深度解析优势的基础上,结合传统的 1×1 Conv2d以较小的计算负担协助IDConv增强模型的通道特征解析能力,并通过MaxPool减少参数的同时扩大特征提取感受野,有助于优化模型的特征解析能力并减少模型的计算资源需求。

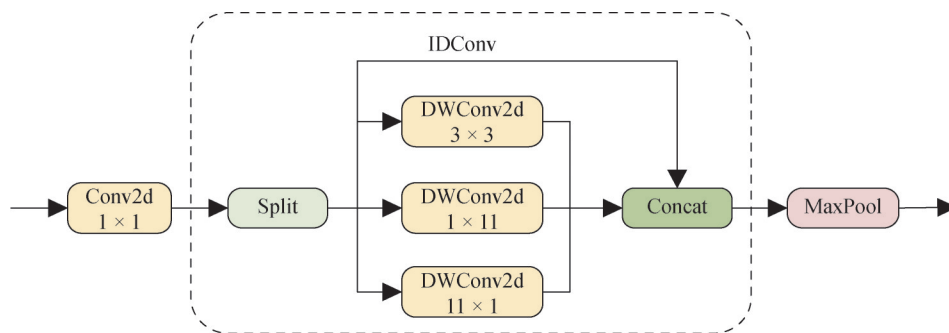


图6 IDBlock结构图

Fig. 6 Diagram of IDBlock structure

1.2.3 VanillaBlock 模块

特征融合网络对不同尺度的缺陷检测性能有着十分重要的影响,为了提升模型对不同尺度特征的检测能力,并进一步减轻模型计算资源需求,从VanillaNet(Chen等,2023a)中引入轻量高效的VanillaBlock模块对基线模型的特征融合网络进行优化。

VanillaNet是一种以极简主义和效率为理念设计的神经网络架构,其结构简单且层数较少,主干在每个阶段只使用一层卷积,避免了如深度架构和自注意力模块等复杂的网络结构,十分适用于资源有限的环境。VanillaNet网络示意图如图7所示(以VanillaNet-6为例),网络主要可分为Stem、Stage、Cls

3个部分,Stem将输入的图像通过4倍下采样转换为下一阶段输入特征图。之后在Stage阶段通过多个Block模块对特征进行解析,Block模块结构如图8所示,主要由 1×1 Conv2d、BN(batch normalization)、Pooling及SIAF(series informed activation function)(Chen等,2023a)构成。 1×1 Conv2d能够使用尽可能小的计算成本对特征进行解析并完成通道转换;卷积和步长为2的Pooling结合使得特征经过每个阶段的Block尺寸变为上层的一半而通道数扩充为上层的两倍;SIAF则是为浅层网络专门设计的激活函数,其能够在不显著增加推理延迟的情况下改善网络的非线性表达能力,特别适合用于资源受限的环境。

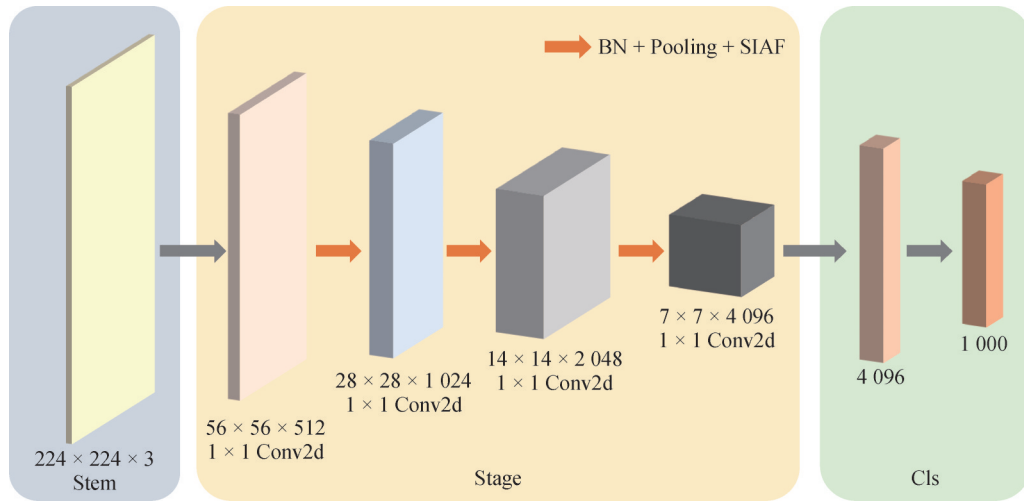


图7 VanillaNet示意图

Fig. 7 Schematic diagram of VanillaNet

境和对推理速度要求较高的应用场景, SIAF与BN的结合使用可以进一步优化训练过程和提升网络性能。最后通过在Cls阶段对经过前两个阶段处理的特征进行解析并输出分类结果。

VanillaNet通过极简的浅层结构,实现了较好的网络性能,其核心的Block模块在网络中发挥着重要的作用。鉴于VanillaBlock优秀的特征解析能力及对于资源受限情景的适用性,本文将VanillaBlock模块引入YOLOv8n的特征融合网络以优化模型的特征融合能力并减轻模型的计算负担。

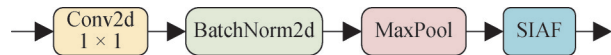


图8 VanillaBlock结构图

Fig. 8 Diagram of the VanillaBlock structure

1.2.4 ID-Detect 模块

基线YOLOv8n的检测头承担着对经过主干网络提取及颈部网络融合的复杂特征进行深度解析的重要任务,基线模型的Detect结构如图9所示,通过两个直接串联的 3×3 Conv模块对经过多次多维融合的特征进一步解析。

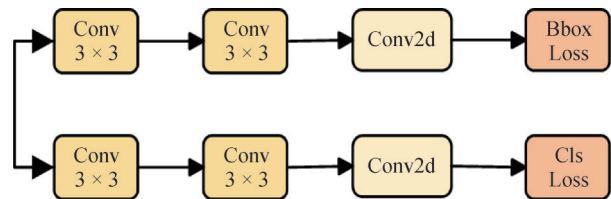


图9 Detect结构图

Fig. 9 Diagram of Detect structure

通过多次实验及对基线YOLOv8n模型结构的分析,学者发现基线YOLOv8n的Detect存在参数量占用较大和结构简单的问题。为降低检测头的计算成本并提升其特征解析能力,本文通过引入开销较小且具有多尺度特征解析能力的IDConv(Yu等, 2024)模块结合具有较小开销的 1×1 Conv2d替换原来的两个串联的 3×3 Conv模块,提出改良后的ID-Detect模块, ID-Detect模块的结构如图10所示,其中左侧深色部分为本文做出修改模块。

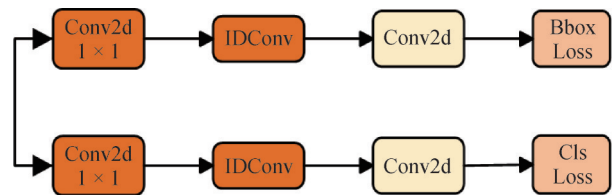


图10 ID-Detect结构图

Fig. 10 Diagram of ID-Detect structure

假设输入和输出通道相同,省略偏置项。 k 、 C 、 H 和 W 分别表示卷积核大小、通道数、特征的高度和宽度。模块改动前后的计算量差异为

$$\frac{FPLOs_{1 \times 1}}{FPLOs_{3 \times 3}} = \frac{2 \times C \times 1^2 \times C \times H \times W}{2 \times C \times 3^2 \times C \times H \times W} = \frac{1}{9} \quad (4)$$

$$\frac{FPLOs_{IDConv}}{FPLOs_{3 \times 3}} = \frac{(2k + 9)CHW/2}{2 \times 9C^2HW} = \frac{2k + 9}{36C} \quad (5)$$

式中, $\frac{FPLOs_{1 \times 1}}{FPLOs_{3 \times 3}}$ 为 1×1 Conv2d和 3×3 Conv的计算量之比, $\frac{FPLOs_{IDConv}}{FPLOs_{3 \times 3}}$ 为IDConv模块和 3×3 Conv的计算量之比。

可以看出,当 $C > k$ 时,改进之后检测头

的计算量较改进之前有明显的减少。后续实验进一步表明引入新的卷积结构替换原 Detect 中两个串联 3×3 Conv 模块后,改良的 ID-Detect 在大幅降低模型参数和计算量的同时,还增强了模型的综合检测性能。

2 实验结果与分析

2.1 实验设置

2.1.1 实验环境

本文实验所用实验配置和实验关键参数如表2和表3所示。

表2 实验配置

Table 2 Experimental setup

名称	参数
CPU	Intel(R) Xeon(R) CPU E5-2673 v4
GPU	NVIDIA GeForce RTX 3060 12 G
语言	Python3.8.0
操作系统	Ubuntu 20.04.6 LTS
深度学习框架	PyTorch2.4.0 + CUDA12.1.1

2.1.2 实验数据集介绍

本实验选用来自南非 Steellenbosch 大学电子系公开的 Pothole Dataset(Nienaber等,2015a,b)数据集进行实验和验证,该数据集针对日常情况下的路面

表3 实验参数

Table 3 Experimental parameters

名称	取值
Epochs	150
初始学习率	0.01
最终学习率	0.000 1
优化器	SGD
权重衰减	0.000 5
Batch size	8
图像尺寸	640×640
Mosaic	10

的坑洼进行识别,主要由两类数据构成,其中主要的一部分是由车载摄像头在不同情景下拍摄的路面坑洼图像构成,囊括了强光、阴影以及遮蔽等复杂情况,能够模拟大多数日常行车中的路面检测情景,部分样例如图11所示。另外,为了增强该数据集的丰富性及多样性,还掺杂了另一部分其他移动设备拍摄的不同场景下的路面坑洼图像,这有助于增强训练模型在不同坑洼检测任务中的稳定性及鲁棒性,部分样例如图12所示。该数据集共1784幅静态图像,将数据集按11:3:1的比例划分为训练集、验证集以及测试集,最终得到训练集1265幅图像、验证集401幅图像、测试集118幅图像,并且训练集、验证集、测试集中的两类数据都经过均匀分配。

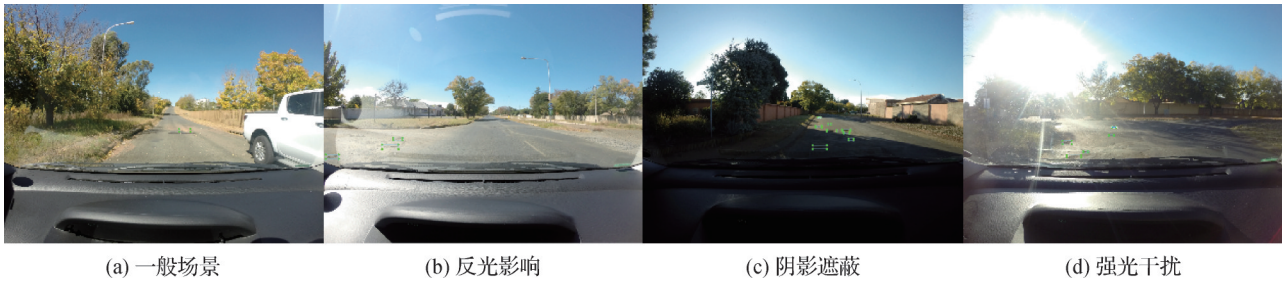


图11 车载摄像头拍摄的不同场景坑洞图例

Fig. 11 Pothole legends of different scenes captured by the on-board camera

((a) general scenario; (b) effect of reflection; (c) shadowing by shadow; (d) strong light interference)

2.1.3 评价指标

本文所用评价指标包括精确度(precision)、召回率(recall)、平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)、参数量(parameters, Params)、浮点运算次数(GFLOPs)以及

帧速率(FPS)。其中精确度、召回率、平均精度以及平均精度均值能够综合反映模型的检测精度,FPS能够衡量模型的检测速度,而模型参数量和GFLOPs能够一定程度体现模型的运行开销,以上这些评价指标能共同反映模型的综合性能。

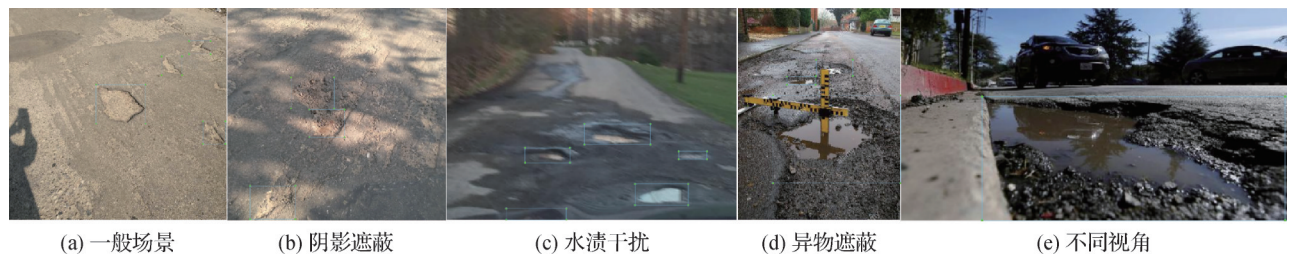


图 12 其他移动设备拍摄的不同场景坑洞图例

Fig. 12 Pothole legends of different scenes captured by other mobile devices ((a) general scenario; (b) shadowing by shadow; (c) water disturbance; (d) occlusion by foreign objects; (e) different perspectives)

2.2 对比实验

为验证本文算法对于优化路面缺陷检测效果的有效性,在 Pothole Dataset 数据集上,将本文算法与目标检测领域的一些经典算法进行比较,包括 Centernet(Zhou 等, 2019; Duan 等, 2019)、RetinaNet(Lin 等, 2017)、Faster R-CNN、SSD、Efficientdet(Tan 等, 2020)、YOLOv8,实验结果如表 4 所示。

通过实验结果可以看出,本文提出的优化算法在同参数量级的比较之间,模型检测精度、模型参数量、模型计算量及 FPS 均达到了最优。YOLOv8x-ours 相较于 YOLOv8x 参数量降低了近 16 M,是 YOLOv8x 的 77%,计算量降低了近 32 GFLOPs, FPS 提升了 5 帧/s, mAP50 由 0.596 上升至 0.623,提高近 3 个百分点;在中等参数量级的模型中,YOLOv8m-

ours 在与 Centernet-resnet50、RetinaNet-resnet50、Faster R-CNN-resnet50 及 SSD-VGG 的比较中, YOLOv8m-ours 的综合性能处于最优水平,在模型参数量、模型计算量方面 YOLOv8m-ours 模型较其他类模型有较大的优势,并且 YOLOv8m-ours 相较于 YOLOv8m 参数量由 25.84 M 降到了 19.49 M,降低了近 1/5,计算量降低了近 11.5 GFLOPs, FPS 提升了 7 帧/s, mAP50 上升至 0.596,达到 YOLOv8x 的水平;在轻量级模型对比中, YOLOv8n-ours 较 SSD-mobilenetv2 及 Efficientdet-d0, 在模型检测精度、模型计算量及 FPS 上有较大优势,并且与基线模型 YOLOv8n 相比, YOLOv8m-ours 将 mAP50 提高 3.5 个百分点,模型参数减少了 1×10^6 , 减少了近 1/3,模型的计算量降低了 2.2 GFLOPs, FPS 提升了 43 帧/s,综合性能有较大的提升。

表 4 对比实验结果
Table 4 Comparing experimental results

模型	精确度	召回率	mAP50	Params/M	FLOPs	FPS/(帧/s)
Centernet-resnet50	0.830	0.362	0.525	32.664	438.606 G	49
RetinaNet-resnet50	0.814	0.223	0.300	36.330	653.970 G	33
Faster R-CNN-resnet50	0.381	0.339	0.301	28.275	1.896 T	13
SSD-VGG	0.843	0.325	0.579	23.745	1.094 T	29
SSD-mobilenetv2	0.907	0.090	0.196	3.675	24.416 G	73
Efficientdet-d0	0.854	0.190	0.296	3.828	29.604 G	15
YOLOv8x	0.729	0.524	0.596	68.124	257.4 G	49
YOLOv8m	0.741	0.550	0.590	25.840	78.7 G	96
YOLOv8n	0.659	0.482	0.515	3.010	8.1 G	200
YOLOv8x-ours	0.758	0.541	0.623	52.502	225.7 G	54
YOLOv8m-ours	0.716	0.549	0.596	19.490	67.2 G	103
YOLOv8n-ours	0.700	0.498	0.550	2.010	5.9 G	243

注:加粗字体表示各列最优结果。

综上所述,本文提出的优化算法能够以更少的计算资源实现更优异的路面缺陷检测性能,有助于促进模型在计算资源有限设备上的部署及实时检测,有望应用于更轻量级道路缺陷检测系统的构建当中。

2.3 消融实验

为验证本文各项改进对于提升模型路面缺陷检测性能的有效性,在 Pothole Dataset 数据集上设置消融实验。首先,通过在基线模型(YOLOv8)上单独引进每个改进模块来验证其对于基线模型的有效性;其次,通过在本文改良模型(YOLOv8-PIVI)上单独剔除每个改进模块来验证其对于本文改良模型的有效性。实验结果如表5所示。

由表5数据可知,PartialBlock 模块的加入使基线模型参数量由 3.01×10^6 降低至 2.89×10^6 ,精度从原来的 0.515 提升至 0.526,但计算量由 8.1 GFLOPs 上升至 8.8 GFLOPs,表明 PartialBlock 模块的引入对参数量的降低以及检测精度的提升有积极作用,但会导致计算复杂度的提升;IDBlock 模块的加入,使基线模型参数量由 3.01×10^6 降低至 2.94×10^6 ,计算量由 8.1 GFLOPs 降至 8.0 GFLOPs,并且精度由 0.515 提升至 0.524,表明 IDBlock 模块的引入能够在降低参数量和计算量的同时提高模型检测精度;当 PartialBlock 模块和 IDBlock 模块同时引入时,模型的参数由 3.01×10^6 降至 2.82×10^6 ,计算量由

8.1 GFLOPs 升至 8.6 GFLOPs,而模型精度由 0.515 提升至 0.523。通过对比 PartialBlock 和 IDBlock 单独引入的数据可以发现,两者共同引入对参数量和计算量的影响能够叠加,使计算量进一步降低的同时保持与单独引入相近的检测精度,体现了骨干网络优化的有效性;VanillaBlock 模块加入后,基线模型参数量由 3.01×10^6 降低至 2.87×10^6 ,精度由 0.515 提升至 0.516,而计算量几乎不变,表明 VanillaBlock 模块的引入能够在保持计算量几乎不变的情况下降低模型参数量并提升模型检测精度;将 PartialBlock 模块、IDBlock 模块以及 VanillaBlock 模块共同引入,模型参数量由基线的 3.01×10^6 降至 2.69×10^6 ,计算量由 8.1 GFLOPs 上升至 8.6 GFLOPs,而精度由 0.515 提升至 0.529,相较于骨干网络的单独改进,在引入 VanillaBlock 模块后,模型在计算量几乎不变的情况下实现了参数量和检测精度的进一步优化;在单独引入 ID-Detect 模块后,模型的参数量由原来的 3.01×10^6 降至 2.32×10^6 ,计算量从 8.1 GFLOPs 降至 5.4 GFLOPs,而精度由 0.515 提升至 0.525,相较于前面描述的模块而言,ID-Detect 模块的引入能够在较大幅度降低基线模型参数量与计算量的同时提升基线模型的检测性能,一定程度上证明了本文提出的 ID-Detect 模块的优越性;当 PartialBlock 模块、IDBlock 模块、VanillaBlock 模块以及 ID-Detect 模块共同加入基线模型后,

表 5 消融实验结果
Table 5 Results of ablation experiments

YOLOv8n	PartialBlock	IDBlock	VanillaBlock	ID-Detect	精确度	召回率	mAP50	Params/M	GFLOPs
√					0.659	0.482	0.515	3.01	8.1
√	√				0.623	0.507	0.526	2.89	8.8
√		√			0.649	0.484	0.524	2.94	8.0
√			√		0.630	0.489	0.516	2.87	8.1
√				√	0.646	0.490	0.525	2.32	5.4
√	√	√			0.674	0.499	0.523	2.82	8.6
√		√	√	√	0.668	0.480	0.524	2.12	5.2
√	√		√	√	0.693	0.485	0.535	2.07	6.1
√	√	√		√	0.654	0.501	0.523	2.14	5.9
√	√	√	√		0.654	0.504	0.529	2.69	8.6
√	√	√	√	√	0.700	0.498	0.550	2.01	5.9

注:加粗字体表示各列最优结果。“√”表示使用对应模块。

模型的参数量从 3.01×10^6 降到了 2.01×10^6 , 减少了近 1/3, 计算量从 8.1 GFLOPs 降低至 5.9 GFLOPs, 降低了 2.2 GFLOPs, 检测精度从 0.515 提高 3.5 个百分点, 达到了 0.550, 并且精确度和召回率较基线都有一定程度提升, 表明本文提出的改良模型在原基线模型的基础上实现了轻量化及提高模型检测性能两个方面的综合提升, 体现了本文提出的各改良模块的有效性 & 改良模型的优越性。此外, 从表中实验数据可以看出, 将本文引入的 4 个改良模块中的任何一个从最终模型中剔除都会导致模

型综合性能在一定程度上降低, 再次验证了本文中引入的各改良模块的有效性。

2.4 对比分析

为了进一步验证本文改良模型的优化效果, 将基线模型和改良模型的指标数据进行综合比较, 如表 6 所示。可以看出, 训练后改良模型的精确度、召回率、mAP50、mAP50-95、Params、GFLOPs、FPS 指标均优于基线模型, 并且模型参数文件大小由 6.3 MB 降低至 4.3 MB, 内存占用较原模型降低了近 1/3, 体现出改良模型较好的综合性能。

表 6 基线模型与本文模型实验结果比较

Table 6 Comparison of experimental results between baseline model and ours

模型	精确度	召回率	mAP50	mAP50-95	Params/M	GFLOPs	FPS/(帧/s)	模型大小/MB
YOLOv8n	0.659	0.482	0.515	0.261	3.01	8.1	200	6.3
YOLOv8n-ours	0.700	0.498	0.550	0.265	2.01	5.9	243	4.3

注: 加粗字体表示各列最优结果。

训练过程中的指标变化曲线如图 13 所示, 可以看出, 在训练过程中改良模型的精确度、召回率、mAP50 和 mAP50-95 参数总体上均优于基线模型, 进一步体现了改良模型良好的优化效果。

为了更直观展示本文改良模型相对于基线模型的优化效果, 将模型在 Pothole Dataset 验证集中的车载摄像头以及其他移动设备所拍摄图像的识别效果进行对比, 如图 14 和图 15 所示, 可以看出, 本文模型在一定程度上优化了基准模型路面缺陷检测中的漏检、误检情况, 强化了模型的小目标识别能力, 并且在识别形状较复杂的缺陷时能够以更精准的识别框以及更高的置信度识别目标缺陷。

综上所述, 本文改良模型较基准模型能够以更轻量级的计算资源需求完成对不同情景下的路面坑洞检测, 并展现出良好的综合检测性能, 体现了本文所提优化算法的有效性 & 优越性。

2.5 模拟实际检测场景

为了验证本文算法在实际情况下的适用性, 设计模拟实际路面坑洞检测情景如下: 选取 RDD2022 (road damage detector 2022) 数据集 (Arya 等, 2024) 中国地区部分包含路面坑洞缺陷的图样, 对使用本文提出的优化算法在 Pothole Dataset 数据集下训练所得路面坑洞检测模型进行测试, 检测结果如图 16 所示。

从图 16 可以看出, 在 Pothole Dataset 数据集中训练所得的路面坑洞缺陷检测模型对其他地区不同场景下的路面坑洞缺陷仍保有一定的检测能力, 表明模型在训练中习得的路面坑洞缺陷相关特征具有一定的泛化能力, 在一定程度上能够适用于训练数据集以外其他场景的路面坑洞缺陷检测。然而, 由于模型训练所用数据集图像和检测图像的地区背景存在较大差异, 模型在训练时未能学习到其他地区背景的相关图像特征, 因此错误地将我国城市路面上的井盖识别成了坑洞缺陷。这表明在实际应用中, 需要因地制宜地运用具有当地路况特征的相关图像数据对模型进行训练, 才能得到适应当地路况的路面缺陷检测模型。

综上所述, 本文的改良模型在训练过程中学习到了具有一定泛化能力的道路坑洞缺陷特征, 能够检测出其他地区或不同设备拍摄图像中的路面坑洞缺陷, 但检测结果存在一定程度的误检或漏检等问题。在实际应用中, 仍需要具有相关地区及相关设备拍摄图像特征的数据集进一步训练, 才能使模型具备适应当地路况特征的路面缺陷检测能力。

2.6 不同数据集鲁棒性验证

为了验证算法的鲁棒性, 将本文提出的优化算法与基线算法在 RDD2022-China-MotorBike (Arya 等, 2024) 数据集和 VOC2012 (visual object classes

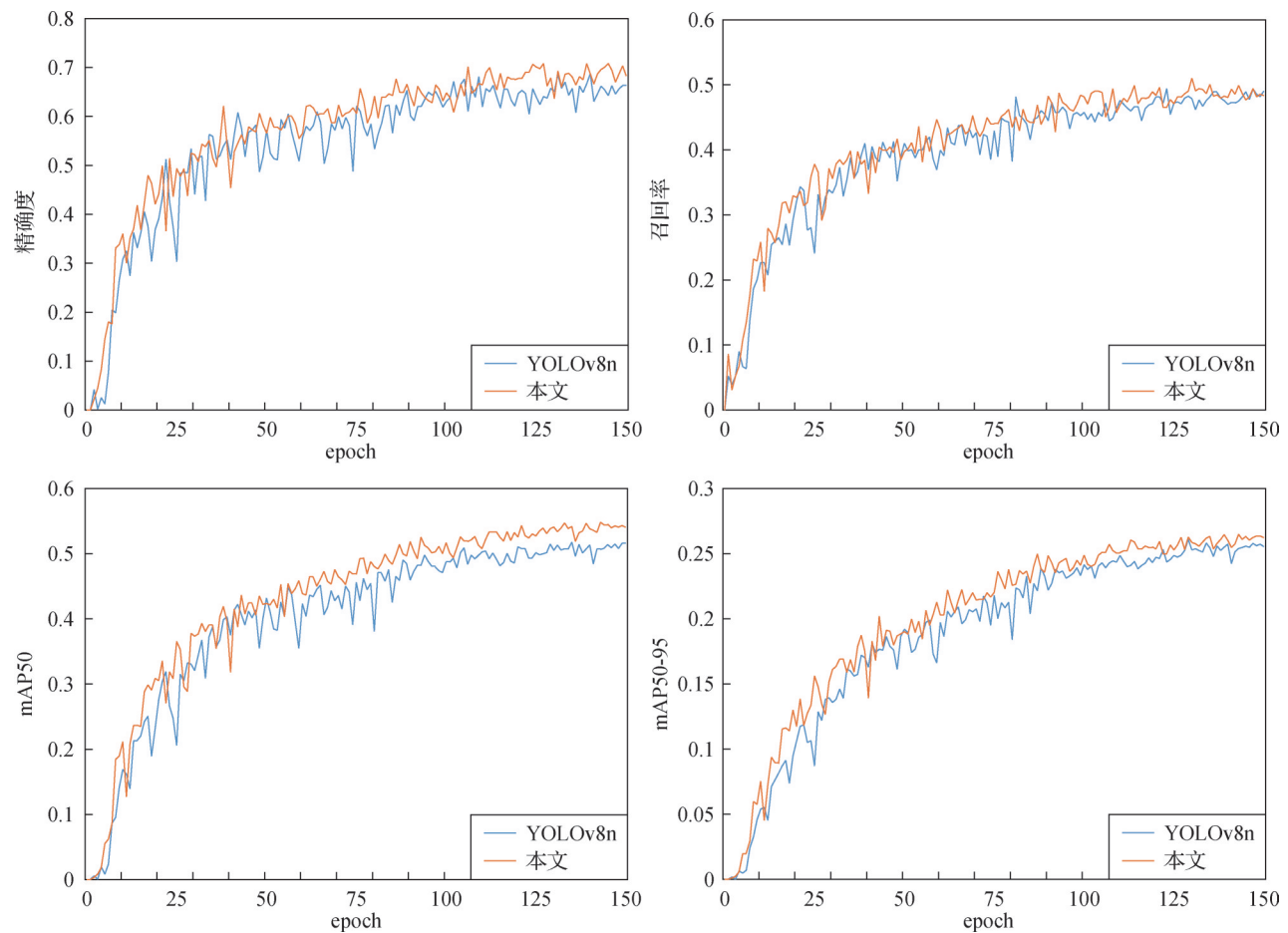


图 13 基线模型和本文模型各评价指标训练曲线

Fig. 13 Training curves for each evaluation index of the baseline model and the proposed model

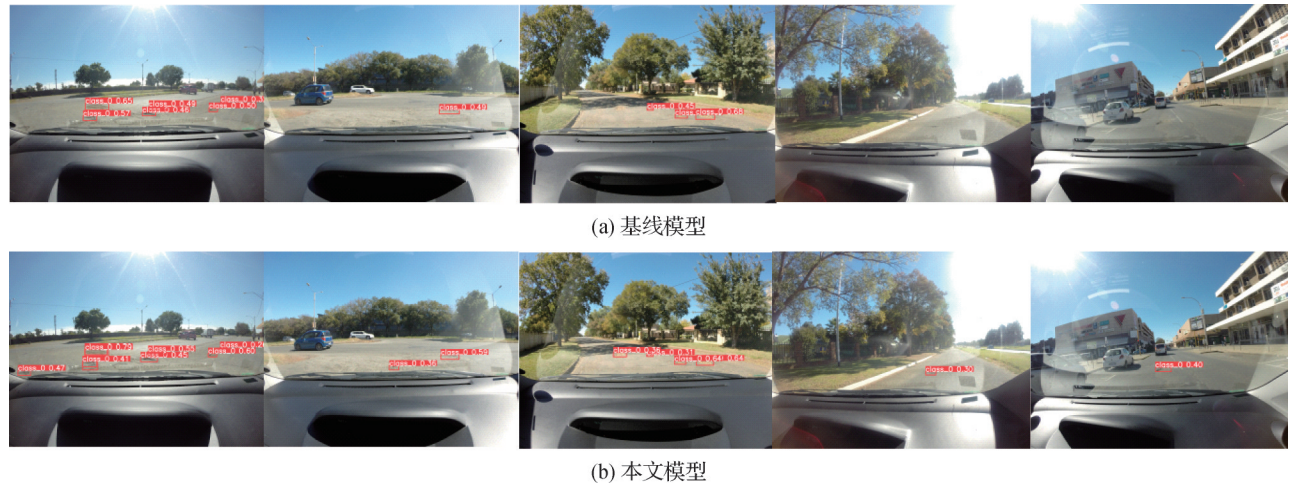


图 14 基线模型和本文模型对车载摄像头图像的路面坑洞检测效果

Fig. 14 Pothole detection effect of the baseline model and the proposed model on on-board camera images
((a) baseline model; (b) ours)

2012)数据集上进行实验。

RDD2022数据集是一个多国道路损伤图像数据集,用于自动检测道路损伤。该数据集包含来自日本、印度、捷克共和国、挪威、美国和中国道路图像,共 47 420 幅。为了验证算法在我国道路的适用性,选取中国地区数据集 RDD2022-China-MotorBike

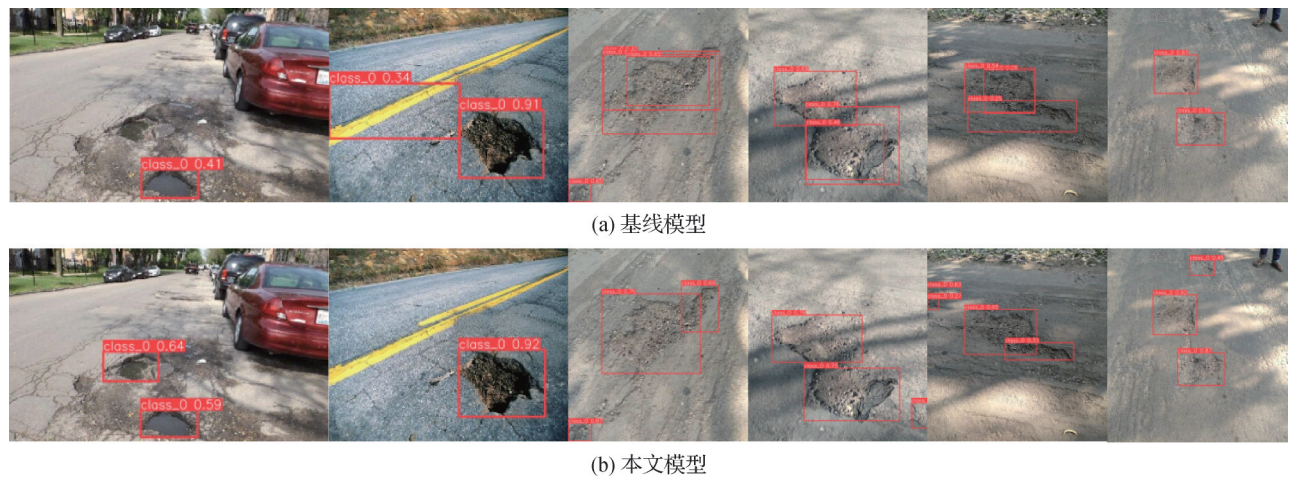


图 15 基线模型和本文模型对其他移动设备图像的路面坑洞检测效果
Fig. 15 Pothole detection effect of the baseline model and the proposed model on images from other mobile devices
((a) baseline model; (b) ours)



图 16 模拟实际场景的路面坑洞检测效果
Fig. 16 The pavement pothole detection effect of simulating the actual scene
((a) images captured by a smartphone attached to a motorcycle; (b) images captured by a camera mounted on a drone)

进行对比实验,该数据集由摩托车上装置智能手机拍摄得到,含纵向裂纹、横向裂纹、网状裂纹、坑洞及修复区域 5 类特征,一共 1 977 幅图像,部分图例如图 17 所示。本文将该数据集按照 8:1:1 的比例划分为训练集、验证集、测试集进行实验,各集合中的数据通过随机程序抽取,为了缓解随机抽取中可能存在的不同类型特征图样富集或稀疏的问题,进行 3 次不同的随机数据集划分及实验,并计算 3 次实验数据的均值进行比较,实验结果如表 7 所示。从表中实验数据可以看出,本文提出的优化算法在 RDD2022-China-MotorBike 数据集上的多次实验中,对横向裂纹、纵向裂纹及修复区域的检测较基线算法有一定优势,其中对横向裂纹及修复区域的检测

精度提升较为明显。值得注意的是,在第 3 组实验中,基线模型表现较为突出,其对坑洞的检测精度高达 0.993,较同组中本文模型高出 7.3 个百分点,并且 FPS 达到了 333 帧/s,但即便如此,该组实验中的 mAP50 依然未能胜过本文模型,并且在多组实验中本文提出算法较基线算法在 mAP50、模型参数量及计算量上均取得了较优值,模型参数量降低了近 1/3,模型计算量降低了近 1/4。实验结果表明,本文提出的优化算法相较于基线算法在针对我国的多类道路缺陷检测中的综合检测性能存在优势,并实现了模型的轻量化。

此外,本文还在 VOC2012 数据集上进行了实验,VOC2012 数据集是计算机视觉领域一个广泛使

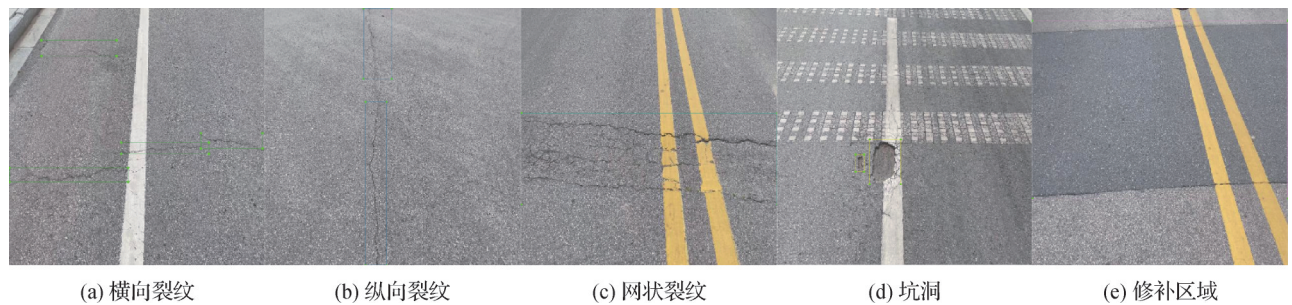


图 17 RDD2022-China-MotorBike 数据集部分图例

Fig. 17 Part of legend of RDD2022-China-MotorBike dataset

((a) transverse cracks; (b) longitudinal cracks; (c) network cracks; (d) potholes; (e) repair areas)

表 7 RDD2022-China-MotorBike 数据集实验结果对比

Table 7 Comparison of experimental results on RDD2022-China-MotorBike dataset

模型	AP					mAP50	Params/M	GFLOPs	FPS/ (帧/s)
	纵向裂纹	横向裂纹	网状裂纹	坑洞	修复区域				
YOLOv8n(1)	0.875	0.921	0.959	0.934	0.939	0.926	3.006	8.1	204
本文(1)	0.878	0.926	0.940	0.944	0.966	0.931	2.013	5.9	285
YOLOv8n(2)	0.894	0.888	0.968	0.992	0.971	0.943	3.006	8.1	158
本文(2)	0.907	0.905	0.980	0.995	0.979	0.953	2.013	5.9	270
YOLOv8n(3)	0.897	0.794	0.937	0.993	0.995	0.923	3.006	8.1	333
本文(3)	0.908	0.857	0.945	0.920	0.995	0.925	2.013	5.9	217
YOLOv8n(A)	0.888	0.867	0.954	0.973	0.968	0.930	3.006	8.1	231
本文(A)	0.897	0.896	0.955	0.953	0.980	0.936	2.013	5.9	257

注:模型后数字序号表示不同组数据集划分中的实验,(A)为 3 组实验数据均值,加粗字体为同组实验数据中的最优结果。

用的标准数据集,主要用于图像分类、目标检测、语义分割和实例分割等任务。该数据集涵盖 20 种常见的物体类别,包括人、动物(如猫、狗、鸟等)、交通工具(如车、船、飞机等)以及家具(如椅子、桌子、沙发等),共 17 125 幅图像,本文将该数据集按照 8:1:1 的比例划分为训练集、验证集、测试集进行实验,各集合中的数据通过随机程序抽取,为了缓解随机抽取中可能存在的不同类型特征图样富集或稀疏的问题,进行 3 次不同的随机数据集划分及实验,并计算 3 次实验数据的均值进行比较,实验结果如表 8 所示。从表 8 的实验数据可以看出,本文提出的优化算法在 VOC2012 数据集上的多次实验中,其 mAP50、模型参数量、模型计算量及 FPS 相较于基线算法均有一定优势,平均而言,mAP50 提升 0.3 个百分点,模型的参数量降低近 1/3,模型的计算量降低近 1/4,FPS 提升近 80 帧/s。实验结果表明,本文提出的优化算法相较于基线算法对于多类日常物体目

标检测中的综合检测性能存在优势,并实现了模型的轻量化。

表 8 VOC2012 数据集实验结果对比

Table 8 Comparison of experimental results on VOC2012 dataset

模型	精确度	召回率	mAP50	Params/M	GFLOPs	FPS/ (帧/s)
YOLOv8n(1)	0.699	0.615	0.666	3.009	8.1	238
本文(1)	0.715	0.606	0.672	2.016	5.9	400
YOLOv8n(2)	0.709	0.570	0.646	3.009	8.1	370
本文(2)	0.697	0.582	0.648	2.016	5.9	416
YOLOv8n(3)	0.679	0.565	0.638	3.009	8.1	357
本文(3)	0.657	0.588	0.639	2.016	5.9	384
YOLOv8n(A)	0.695	0.583	0.650	3.009	8.1	321
本文(A)	0.689	0.592	0.653	2.016	5.9	400

注:模型后数字序号表示不同组数据集划分中的实验,(A)为 3 组实验数据均值,加粗字体为同组实验数据中最优结果。

综上所述,本文所提出的优化算法在RDD2022-China-Motor数据集和VOC2012数据集上仍能实现不错的检测效果并呈现出较基线算法更优的综合检测性能,表明本文所提出的优化算法有较好的图像特征提取能力及特征解析能力,具有良好的适应性和鲁棒性,能够适应不同的目标检测任务,有望适用于更复杂多样的道路缺陷检测场景。

3 结 论

针对当前路面缺陷检测模型计算成本较高,难以应用于资源受限设备的问题,本文在YOLOv8n的基础上,结合部分卷积与初始深度卷积机制提出一种轻量级路面缺陷检测算法YOLOv8n-PIVI。该算法首先在主干网络中引入PartialBlock模块和IDBlock模块,降低模型参数量的同时提高模型对不同维度特征的处理能力;其次在特征融合网络中引入VanillaBlock模块,使模型轻量化的同时优化模型的特征融合能力;最后用轻量高效的ID-Detect模块优化检测头。该算法在Pothole Dataset路面坑洼检测数据集中,mAP50较基线算法有3.5个百分点的提升,而模型的参数量和计算量仅为基线的67%和72%,模型参数文件由6.3 MB降低至4.3 MB,内存占用较原模型降低了近1/3。与目前主流的目标检测算法相比,本文算法在同参数量级的比较中,综合检测性能处于较优水平,能够在更小的参数量和计算量的情况下实现更高的检测精度及检测效率,并且本文提出的最终模型YOLOv8-PIVI在对比实验中以最少的参数量和计算量实现了中上水平的检测精度及最高的FPS,模型参数文件大小仅4.3 MB,体现了本文所提轻量级算法的优越性。此外,在RDD2022-China-Motor数据集和VOC2012数据集上的实验进一步证明了本文算法具有良好的适应性和鲁棒性,能够适应不同的目标检测任务。

综合实验表明,本文算法能够以轻量级的计算资源需求实现较高帧率的路面缺陷检测,并且模型具有良好的鲁棒性及适应性,适用于计算资源受限的实时路面缺陷检测场景。但模型的检测精度、参数量和计算量仍有改进空间,在未来的研究工作中,将对模型检测性能和计算资源需求进一步优化和平衡,实现更高效益的高精度检测。

参考文献(References)

- Arya D, Maeda H, Ghosh S K, Toshniwal D and Sekimoto Y. 2024. RDD2022: a multi-national image dataset for automatic road damage detection. *Geoscience Data Journal*, 11(4): 846-862 [DOI: 10.1002/gdj3.260]
- Boureau Y L, Ponce J and LeCun Y. 2010. A theoretical analysis of feature pooling in visual recognition//*Proceedings of the 27th International Conference on Machine Learning*. Haifa, Israel: Omnipress: 111-118
- Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 6154-6162 [DOI: 10.1109/cvpr.2018.00644]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]
- Chen H T, Wang Y H, Guo J Y and Tao D C. 2023a. VanillaNet: the power of minimalism in deep learning//*Proceedings of the 37th International Conference on Neural Information Processing Systems*. New Orleans, USA: Curran Associates Inc.: #308
- Chen J R, Kao S H, He H, Zhuo W P, Wen S, Lee C H and Chan S H G. 2023b. Run, don't walk: chasing higher FLOPS for faster neural networks//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 12021-12031 [DOI: 10.1109/CVPR52729.2023.01157]
- Chen J Y, Zou C L, Wang S H, Xia L and Chen Z. 2023. Research on rapid detection method of pavement defects by improving YOLOv5. *Electronic Measurement Technology*, 46(10): 129-135 (陈建瑜, 邹春龙, 王生怀, 夏力, 陈哲. 2023. 改进YOLOv5的路面缺陷快速检测方法研究. *电子测量技术*, 46(10): 129-135) [DOI: 10.19651/j.cnki.emt.2211195]
- Cong X H, Li S X, Chen F K, Liu C and Meng Y. 2023. A review of YOLO object detection algorithms based on deep learning. *Frontiers in Computing and Intelligent Systems*, 4(2): 17-20 [DOI: 10.54097/fcis.v4i2.9730]
- Dai J F, Li Y, He K M and Sun J. 2016. R-FCN: object detection via region-based fully convolutional networks//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona, Spain: Curran Associates Inc.: 379-387
- Du F J and Jiao S J. 2022. Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. *Sensors*, 22(9): #3537 [DOI: 10.3390/s22093537]
- Duan K W, Bai S, Xie L X, Qi H G, Huang Q M and Tian Q. 2019. CenterNet: keypoint triplets for object detection//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*

- (ICCV). Seoul, Korea (South): IEEE: 6568-6577 [DOI: 10.1109/iccv.2019.00667]
- Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/iccv.2015.169]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/cvpr.2014.81]
- Gopalakrishnan K. 2018. Deep learning in data-driven pavement image analysis and automated distress detection: a review. *Data*, 3(3): #28 [DOI: 10.3390/data3030028]
- Gu Z Y, Yu Y C and Zhe T T. 2023. Rapider-YOLOX: lightweight object detection network with high precision. *Chinese Journal of Intelligent Science and Technology*, 5(1): 92-103 (顾宙瑜, 於跃成, 者甜甜. 2023. Rapider-YOLOX: 高效的轻量级目标检测网络. *智能科学与技术学报*, 5(1): 92-103) [DOI: 10.11959/j.issn.2096-6652.202303]
- He K M, Gkioxari G, Dollar P and Girshick R. 2020. Mask R-CNN//Proceedings of 2020 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/cvpr.2016.90]
- Lai J, Peng R H, Sun D X and Huang J. 2024. Detection of camouflage targets based on attention mechanism and multi-detection layer structure. *Journal of Image and Graphics*, 29(1): 134-146 (赖杰, 彭锐晖, 孙殿星, 黄杰. 2024. 融合注意力机制与多检测层结构的伪装目标检测. *中国图象图形学报*, 29(1): 134-146) [DOI: 10.11834/jig.221189]
- Li Y, Bian S, Wang C T and Lu W. 2023. CNN and transformer-coordinated deepfake detection. *Journal of Image and Graphics*, 28(3): 804-819 (李颖, 边山, 王春桃, 卢伟. 2023. CNN 结合 Transformer 的深度伪造高效检测. *中国图象图形学报*, 28(3): 804-819) [DOI: 10.11834/jig.220519]
- Li Z H, An J T, Jia H Y and Fang Y. 2023. Lightweight object detection model in remote sensing image by combining rotation box and attention mechanism. *Journal of Image and Graphics*, 28(9): 2706-2718 (李朝辉, 安金堂, 贾红雨, 方艳. 2023. 结合旋转框和注意力机制的轻量遥感图像检测模型. *中国图象图形学报*, 28(9): 2706-2718) [DOI: 10.11834/jig.220839]
- Liang F T, Zhou Y, Chen X, Liu F Z, Zhang C H and Wu X J. 2021. Review of target detection technology based on deep learning//Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence. Sanya, China: Association for Computing Machinery: 132-135 [DOI: 10.1145/3448218.3448234]
- Liang L M, Long P W, Lu B H, Ou Y Y and Zeng L. 2024. EHH-YOLOv8s: a lightweight algorithm for strip surface defect detection. *Journal of Beijing University of Aeronautics and Astronautics*, 1-15 (梁礼明, 龙鹏威, 卢宝贺, 欧燕燕, 曾璐. 2024. EHH-YOLOv8s: 一种轻量级的带钢表面缺陷检测算法. *北京航空航天大学学报*, 1-15) [DOI: 10.13700/j.bh.1001-5965.2024.0426]
- Lin T Y, Goyal P, Girshick R, He K M and Dollar P. 2017. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/iccv.2017.324]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu Z, Mao H Z, Wu C Y, Feichtenhofer C, Darrell T and Xie S N. 2022. A ConvNet for the 2020s//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11966-11976 [DOI: 10.1109/cvpr52688.2022.01167]
- Nienaber S, Booyens M J and Kroon R. 2015a. Detecting potholes using simple image processing techniques and real-world footage//Proceedings of Cognitive Informatics and Soft Computing. Pretoria, South Africa: [s. n.]: 893-902 [DOI: 10.13140/RG.2.1.3121.8408]
- Nienaber S, Kroon R S and Booyens M J. 2015b. A comparison of low-cost monocular vision techniques for pothole distance estimation//Proceedings of 2015 IEEE Symposium Series on Computational Intelligence. Cape Town, South Africa: IEEE: 419-426 [DOI: 10.1109/SSCI.2015.69]
- Qin R B, Chen N J and Huang Y H. 2022. EDDNet: an efficient and accurate defect detection network for the industrial edge environment//Proceedings of the 22nd IEEE International Conference on Software Quality, Reliability and Security (QRS). Guangzhou, China: IEEE: 854-863 [DOI: 10.1109/QRS57517.2022.00090]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/cvpr.2016.91]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/tpami.2016.2577031]
- Spencer B F, Hoskere V and Narazaki Y. 2019. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2): 199-222 [DOI: 10.1016/j.eng.2018.11.030]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 1-9 [DOI:

- 10.1109/cvpr.2015.7298594]
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. 2016. Rethinking the inception architecture for computer vision//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2818-2826 [DOI: 10.1109/cvpr.2016.308]
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/cvpr42600.2020.01079]
- Terven J, Córdova-Esparza D M and Romero-González J A. 2023. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4): 1680-1716 [DOI: 10.3390/make5040083]
- Wan F, Sun C, He H Y, Lei G B, Xu L and Xiao T. 2022. YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s. EURASIP Journal on Advances in Signal Processing, 2022(1): #98 [DOI: 10.1186/s13634-022-00931-x]
- Wang C M and Liu H. 2024. YOLOv8-VSC: lightweight algorithm for strip surface defect detection. Journal of Frontiers of Computer Science and Technology, 18(1): 151-160 (王春梅, 刘欢. 2024. YOLOv8-VSC: 一种轻量级的带钢表面缺陷检测算法. 计算机科学与探索, 18(1): 151-160) [DOI: 10.3778/j.issn.1673-9418.2308060]
- Wang H X, Zhang Y S, Song B, Chen D S and Yang Y. 2023. NAS-FPNLite object detection method fused with cross stage connection and inverted residual. Journal of Image and Graphics, 28(4): 1004-1018 (王红霞, 张永善, 宋邦, 陈德山, 杨益. 2023. 融合跨阶段连接与倒残差的NAS-FPNLite目标检测方法. 中国图象图形学报, 28(4): 1004-1018) [DOI: 10.11834/jig.211099]
- Wang J, Xu P Q, Li L J and Zhang F. 2023. DAssd-Net: a lightweight steel surface defect detection model based on multi-branch dilated convolution aggregation and multi-domain perception detection head. Sensors, 23(12): #5488 [DOI: 10.3390/s23125488]
- Wei F, Zhou J P, Tan X, Lin J, Tian L and Wang H. 2024. Lightweight YOLOv5 detection algorithm for low-altitude micro UAV. Journal of Optoelectronics·Laser, 35(6): 641-649 (魏峰, 周建平, 谭翔, 林静, 田莉, 王虎. 2024. 针对低空微小型无人机的轻量型YOLOv5检测算法. 光电子·激光, 35(6): 641-649) [DOI: 10.16136/j.joel.2024.06.0741]
- Weng C K and Zhang L X. 2023. A survey of object detection models based on deep learning. Computer Science and Technology, 2(2): #93 [DOI: 10.57237/j.cst.2023.02.006]
- Xiao Z J, Lin B H and Qu H C. 2024. SAR ship detection with multi-mechanism fusion. Journal of Image and Graphics, 29(2): 545-558 (肖振久, 林瀚翰, 曲海成. 2024. 融合多重机制的SAR舰船检测. 中国图象图形学报, 29(2): 545-558) [DOI: 10.11834/jig.230166]
- Xu T F, Huang H, Zhang H M and Niu X F. 2024. Lightweight road damage detection method based on improved YOLOv8. Computer Engineering and Applications, 60(14): 175-186 (胥铁峰, 黄河, 张红民, 牛晓富. 2024. 基于改进YOLOv8的轻量化道路病害检测方法. 计算机工程与应用, 60(14): 175-186) [DOI: 10.3778/j.issn.1002-8331.2402-0243]
- Yu W H, Zhou P, Yan S C and Wang X C. 2024. InceptionNeXt: when inception meets ConvNeXt//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 5672-5683 [DOI: 10.1109/CVPR52733.2024.00542]
- Zakeri H, Nejad F M and Fahimifar A. 2017. Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. Archives of Computational Methods in Engineering, 24(4): 935-977 [DOI: 10.1007/s11831-016-9194-z]
- Zeng Y F, Yao X, Hua F, Wang P P and Gu M. 2023. A lightweight network-involute and decoupled for panoramic wisdom tooth detection. Journal of Image and Graphics, 28(8): 2491-2504 (曾怡峰, 姚潇, 华飞, 王佩佩, 顾敏. 2023. 面向全景智齿检测的内卷解耦轻量化网络. 中国图象图形学报, 28(8): 2491-2504) [DOI: 10.11834/jig.220377]
- Zhang D P, He S J, Wei Y Y, Xu Y C, Hu H M and Huang W J. 2024. Lightweight road defect detection method based on dynamic deformable convolution. Journal of Computer-Aided Design and Computer, 1-12 (章东平, 何数技, 魏杨悦, 徐云超, 胡海苗, 黄文君. 2024. 基于动态可变形卷积的轻量化道路缺陷检测方法. 计算机辅助设计与图形学学报, 1-12) [DOI: 10.3724/SP. J. 1089.2024-00332]
- Zhang J, Wan G Y, Jiang M, Lu G F, Tao X W and Huang Z Y. 2023. Small object detection in UAV image based on improved YOLOv5. Systems Science and Control Engineering, 11(1): #2247082 [DOI: 10.1080/21642583.2023.2247082]
- Zhou T, Ye X Y, Zhao Y N, Lu H L and Liu F Z. 2024. Cross-modal attention YOLOv5 PET/CT lung cancer detection. Journal of Image and Graphics, 29(4): 1070-1084 (周涛, 叶鑫宇, 赵雅楠, 陆惠玲, 刘凤珍. 2024. 跨模态注意力YOLOv5的PET/CT肺部肿瘤检测. 中国图象图形学报, 29(4): 1070-1084) [DOI: 10.11834/jig.230169]
- Zhou X Y, Wang D Q and Krähenbühl P. 2019. Objects as points [EB/OL]. [2024-12-06]. <https://arxiv.org/pdf/1904.07850.pdf>
- Zhuo W Q, Li D Z, Wang W and Dong J. 2023. Data-free model compression for light-weight DeepFake detection. Journal of Image and Graphics, 28(3): 820-835 (卓文琦, 李东泽, 王伟, 董晶. 2023. 面向轻量级深度伪造检测的无数据模型压缩. 中国图象图形学报, 28(3): 820-835) [DOI: 10.11834/jig.220559]

作者简介

彭毅,男,硕士研究生,主要研究方向为目标检测和缺陷检测。E-mail:1192744976@qq.com

邵宇飞,通信作者,男,副教授,主要研究方向为智能数据处理。E-mail:yfshao@alum.imr.ac.cn