

英红,覃云涛,刘曦民,等. 基于改进YOLOv8的沥青路面深度图像病害检测算法[J]. 湖南科技大学学报(自然科学版), 2025, 40(1):88-101. doi:10.13582/j.cnki.1672-9102.2025.01.012

YING H, QIN Y T, LIU X M, et al. Asphalt Pavement Depth Image Disease Detection Algorithm Based on Improved YOLOv8 [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2025, 40(1):88-101. doi:10.13582/j.cnki.1672-9102.2025.01.012

基于改进YOLOv8的沥青路面深度图像 病害检测算法

英红*,覃云涛,刘曦民,朱嘉丽,陈蔚铭

(桂林电子科技大学 建筑与交通工程学院,广西 桂林 541004)

摘要:为解决沥青路面病害检测人工成本高、检测精确度与效率低下的问题,提出一种面向多尺度目标沥青路面深度图像的轻量化道路病害检测模型RBDN-YOLO-E.通过无人机快速扫描沥青路面并建立三维点云模型得到路面深度图像,将2496张深度图像建立沥青路面病害检测数据集,并按7:2:1的比例划分为训练集、验证集以及测试集.在网络主干的C2f中引入重塑空间注意力卷积模块RFACnv得到C2f_RFACnv模块以关注感受野空间特征,为图像每个区域的特征提取提供了定制化的关注度;在网络颈部的C2f中引入可变形卷积DCNv3得到C2f_DCNv3模块以提升模型对目标形变的建模能力,降低输出通道数和模型计算成本,提高识别的精确度;将网络中SPPF模块替换为SPPELAN模块的空间金字塔池化结构以生成更多尺度,提高特征表示能力.试验结果表明:RBDN-YOLO-E模型相对于YOLOv8n模型的mAP50、F1值、精确率、召回率和推理时间分别提升了1.70%、2.00%、4.29%、2.00%、3.40%,而模型计算量、参数量和模型大小分别降低了0.9 GFLOPs,0.73 M,1.4 M.通过结合无人机与改进模型RBDN-YOLO-E,可以更安全、快速和准确地提取沥青路面病害,有效解决传统沥青路面病害的检测精度与成本问题,提高对沥青路面病害的检测效率.

关键词:病害检测;轻量化;无人机;深度图像;YOLOv8n

中图分类号:TP391.4;U418.6 **文献标志码:**A **文章编号:**1672-9102(2025)01-0088-14

Asphalt Pavement Depth Image Disease Detection Algorithm Based on Improved YOLOv8

YING Hong, QIN Yuntao, LIU Ximin, ZHU Jiali, CHEN Weiming

(School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In order to solve the problems of high labor cost, low detection accuracy and low efficiency of asphalt pavement disease detection, a lightweight road disease detection model RBDN-YOLO-E for multi-scale target asphalt pavement depth image is proposed. The pavement depth image is obtained by scanning the asphalt pavement quickly and establishing a three-dimensional point cloud model. The 2496 depth images are established into a data set of asphalt pavement disease detection, and divided into training set, verification set and test set according to the ratio of 7:2:1. The receptive field attention convolution module RFACnv is introduced into the C2f of the network backbone to obtain the C2f_RFACnv module to focus on the spatial

收稿日期:2024-09-10

基金项目:国家自然科学基金资助项目(51668012;51968011)

*通信作者,E-mail:452659071@qq.com

features of the receptive field, which provides customized attention for the feature extraction of each region of the image. The deformable convolution DCNv3 is introduced into the C2f of the network neck to obtain the C2f_DCNv3 module to improve the modeling ability of the model to the target deformation, reduce the number of output channels and the computational cost of the model, and improve the accuracy of recognition. The SPPF module in the network is replaced by the spatial pyramid pooling structure of the SPPELAN module to generate more scales and improve the feature representation ability. The experimental results show that the mAP50, F1 value, precision rate, recall rate and inference time of the RBDN-YOLO-E model are increased by 1.70%, 2.00%, 4.29%, 2.00% and 3.40 % respectively compared with the YOLOv8n model, while the model calculation amount, parameter quantity and model size are reduced by 0.9 GFLOPs, 0.73 M and 1.4 M respectively. By combining the UAV and the improved model RBDN-YOLO-E, the asphalt pavement disease can be extracted more safely, quickly and accurately, which can effectively solve the problem of traditional asphalt pavement disease detection accuracy and cost, and improve the efficiency of asphalt pavement disease detection.

Keywords: disease detection; lightweight; UAV; depth image; YOLOv8n

道路网络作为推动国家经济繁荣和改善人民生活的基础,其重要性不言而喻。道路病害是影响道路安全的重要因素,开发一种高效、精确且便捷的道路病害检测方法对确保道路安全具有重要意义。

目前,国内外学者已经对沥青路面病害做出了大量研究,基于图像识别检测的算法已成为研究趋势。两阶段算法是以 R-CNN^[1-4] 系列为代表。孙朝云等^[5]通过调整 Faster R-CNN 的特征提取层和候选框比例以增强路面病害检测的准确性;牛慧余等^[6]则借助图像预处理与 Mask R-CNN 骨干网络的优化,增强了对裂缝特征的提取能力,从而提升检测效率。以 SSD^[7] 和 YOLO^[8-13] 系列为代表的单阶段算法,在保持较高检测精度的同时,也面临运行速度慢和内存消耗大的问题。高晶等^[14]在 YOLOv5s 中嵌入 SimAM 无参数注意力机制和 Res2NetBlock 残差块得到 PD-YOLOv5s,并结合 SPD-GSConv 模块,在减少计算成本的同时,显著提高了路面病害检测精度,展现出优越的工程应用价值;WU 等^[15]针对移动设备优化了 YOLOv5,推出了 YOLO-LWNet 模型,显著降低了计算复杂度;王海群等^[16]则在 YOLOv8 中整合 CNX2f 特征模块、RepConv 和 DBB 重参数化模块,以及 SPPF-Avg 模块,改善了多尺度特征提取能力,增强了对路面病害的识别精度;卢小平等^[17]利用无人机拍摄道路图像,使用迭代二值化方法检测出道路裂缝,在复杂的场景下有较高的鲁棒性,且提高了检测效率;贺泳超等^[18]在 YOLOv7 中引入轻量级卷积、Slim-neck 架构及动态稀疏注意力模块,得到的 TLC-YOLO 算法在隧道衬砌裂缝检测中展现出优于 YOLOv7 的准确率,有效提升了裂缝智能识别的性能;胡鹏等^[19]提出一种多尺度特征增强的路面病害检测模型,通过混合空洞卷积、多路径特征融合、锚点框优化、面积惩罚项及高效注意力机制,显著提升了检测精度与效率,并在边缘设备上实现快速部署,具备较高的工程应用价值。尽管这些研究在特定道路环境的单一或有限目标检测上表现优异,但面对复杂背景下多尺度的多目标检测任务,仍需进一步提升检测性能。YOLOv8 标志着目标检测技术的重要进步,作为 YOLO 系列的迭代算法,在实时检测方面展现出优异的性能。为了克服实际道路环境中目标多样性、尺度变化大以及复杂背景的干扰,减少错检和漏检,因此对 YOLOv8 进行改进,旨在通过创新模块和算法优化,更有效地应对这些挑战。

YOLOv8 模型以其检测速度快、精度高的优势在目标检测领域得到了广泛应用,但对于复杂场景下的病害,目标检测容易出现漏检和误检问题,并且将算法应用于实际病害检测时需要考虑计算成本问题,若计算成本过高,会导致在使用此算法检测时,对硬件的要求较高,即欠缺实际应用能力。针对上述问题,同时考虑模型的计算量、参数量以及模型大小,本文使用 YOLOv8n 作为基线模型并针对基线模型的网络结构进行改进与设计,得到新的沥青路面病害深度图像检测模型 RBDN-YOLO-E。结果表明:改进算法保持计算量较低的同时,提高了识别精度、检测实时性与实际应用能力。

1 沥青路面深度图像病害检测算法

1.1 重塑空间注意力卷积模块 C2f_RFAConv

近年来,关于空间的注意力机制已被广泛应用于改进并提升卷积神经网络,但并没有解决卷积核参数

共享问题,只关注了空间特征,如协调注意力机制(CA)^[20]、卷积块注意力模块(CBAM)^[21].而感受野注意力(Receptive-Field Attention)^[22]本质上解决了卷积核参数共享问题,把焦点从空间特征转移到感受野空间特征上.这种注意力机制关注了感受野空间特征,为大尺寸卷积核提供了有效的注意力权重与特征,以覆盖更大的感受野.这种感受野注意力卷积操作代表了一种替代标准卷积操作的新方法,这种方法可以有效地理解并处理深度图像中的病害细节区域.该注意力机制基本不会增加参数量与计算成本,同时提升性能参数.在沥青路面深度图像中,病害特征以黑色展现在图像上,然而部分病害深度是渐进式加深的,导致路面与病害是以灰色作为过渡色呈现在图片上,增加了图像的复杂性.为了让模型更精确地识别病害与病害的种类,在YOLOv8n网络Backbone的C2f中引入重塑空间注意力卷积RFACnv(Receptive-Field Attention Convolution),得到C2f_RFACnv模块,以更有效地处理图像中的细节和复杂的背景.图1为C2f_RFACnv模块的结构示意图.

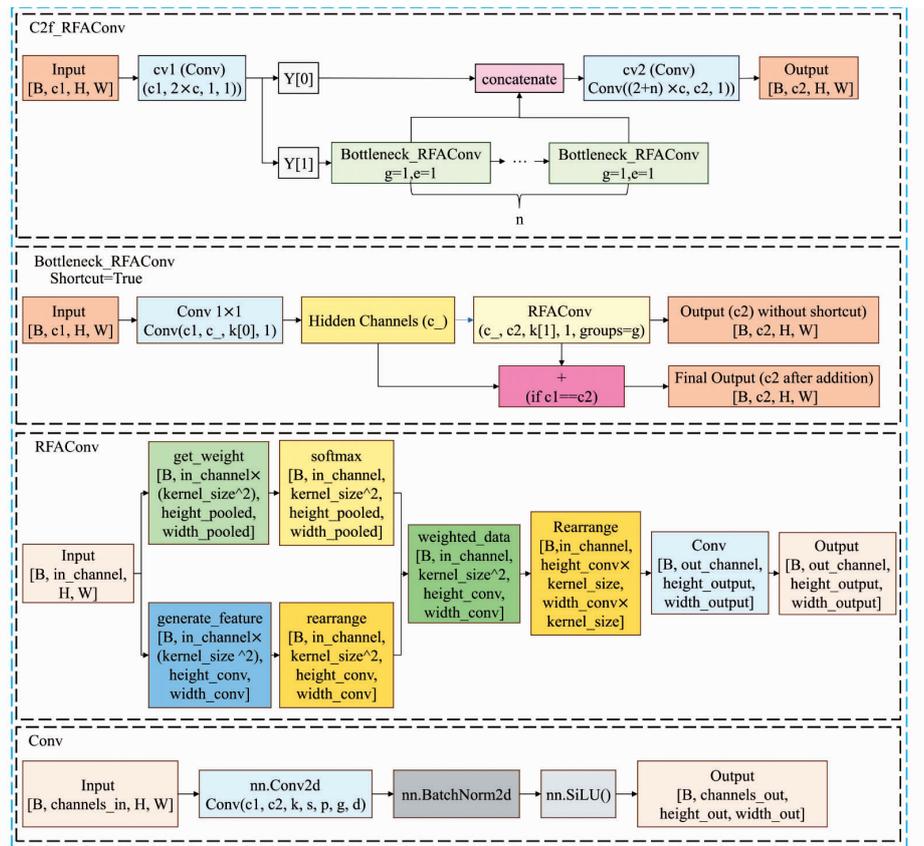


图 1 C2f_RFACnv 结构

RFACnv 模块的主要思想是将空间注意力机制与卷积操作相结合,从而提高卷积神经网络的性能.这种方法的核心在于优化卷积核的工作方式,特别是在处理感受野内的空间特征时.以下是 RFACnv 模块的几个关键思想:

1) RFACnv 模块中的感受野空间特征被用来指导注意力机制,能够为不同区域和不同尺寸的感受野提供不同的处理.感受野空间特征是指在卷积神经网络中,卷积层能“看到”输入数据的局部区域,每个卷积操作的输出是基于输入数据的一个局部感受野.这个感受野定义了卷积核可以接触到的输入数据的大小和范围.图 2 展示了一个 3x3 的卷积操作,其特征是通过将卷积核与同样大小的感受野滑块相乘后求和而得.输入图像 X 上的每一个 3x3 的区域都被一个 3x3 的卷积核 K 处理.每个感受野内的元素 $X_{i,j}$ (其中 i 和 j 表示在感受野内的位置)都与卷积核 K 所对应位置的权重 $K_{i,j}$ 相乘,所得乘积会被求和得到一个新的特征值 F .这个过程在整个输入图像上滑动进行,以生成新的特征图.这种标准的卷积操作强调了局部连接和权重共享的概念.

2) RFACnv 模块通过引入注意力机制以解决参数共享问题,允许网络为每个感受野生成特定的权重.图 3 展示了一个卷积操作的过程.

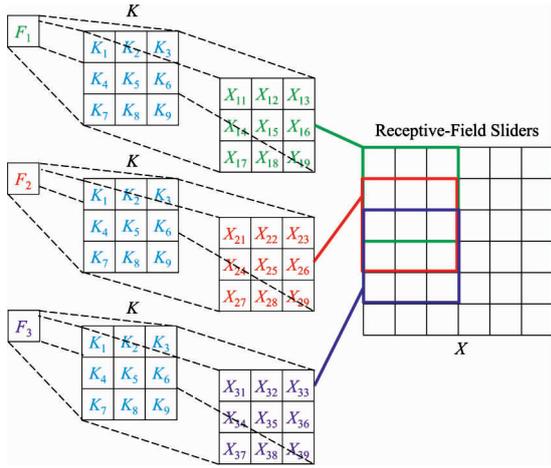


图 2 3×3 卷积操作

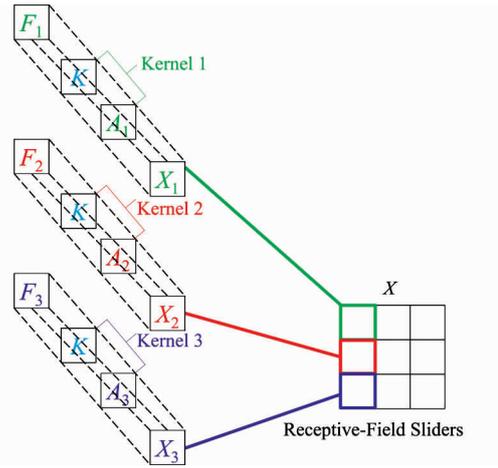


图 3 1×1 卷积操作

其中卷积核参数 K_i 通过将注意力权重 A_i 与卷积核参数 K 相乘得到,使得每个感受野滑块的卷积操作都有一个独特的卷积核参数,这些参数是将通用的卷积核参数与特定位置的注意力权重相结合来获得的.这个过程将注意力机制与卷积核相结合,为每个感受野位置产生一个定制化的卷积核.图 3 中的 Kernel 1, Kernel 2 和 Kernel 3 分别是通过将通用卷积核参数 K 与对应的注意力权重 A_1, A_2 和 A_3 相乘得到的.因此卷积核可以根据每个感受野内的不同特征动态调整其参数,而不是对所有区域做相同处理.

3) RFACConv 模块通过利用感受野注意力机制来动态调整卷积核的权重,从而为每个区域的特征提取提供了定制化的关注度.即使是大尺寸卷积核,也能够更加有效地捕捉和处理重要的空间特征,且不会对不重要的信息分配过多的计算资源.而感受野滑块内的特征输出是将输入特征 X 与对应的注意力权重 A 以及卷积核 K 的权重进行逐元素乘法运算后得到,式(1)为不同感受野滑块内的特征输出公式.

$$\begin{aligned}
 F_1 &= X_{11}A_{11}K_1 + X_{12}A_{12}K_2 + X_{13}A_{13}K_3 + \dots + X_{19}A_{19}K_9; \\
 F_2 &= X_{21}A_{21}K_1 + X_{22}A_{22}K_2 + X_{23}A_{23}K_3 + \dots + X_{29}A_{29}K_9; \\
 &\dots \\
 F_N &= X_{N1}A_{N1}K_1 + X_{N2}A_{N2}K_2 + X_{N3}A_{N3}K_3 + \dots + X_{N9}A_{N9}K_9.
 \end{aligned}
 \tag{1}$$

式中: F_N 为不同感受野滑块内的特征输出; X 为输入特征; A 为注意力权重; K 为卷积核.

图 4 为 RFACConv 卷积模块操作,此操作描述了感受野滑块中特征的重叠,该现象常出现于标准卷积操作中.特征的重叠导致了注意力权重的共享问题,意味着不同的感受野可能会对相同的输入特征使用相同的注意力权重,并且强调了每个感受野滑块内的卷积操作的参数不应完全共享,而是应该根据每个特定区域内的特征和相应的注意力权重进行调整.该调整允许网络对每个局部区域进行更加精细的处理,能够更好地捕捉和响应输入数据的特征,而不是简单地对整个图像应用相同的权重.该方法能够提升网络对特征的理解和表示,从而改善最终的学习和预测性.RFACConv 卷积模块通过其独特的设计,显著增强了模型的代表能力,使其能够更为精确地适应并表达输入数据中的细节特征,特别是在面对复杂多变的图像内容时.该卷积模块的参数调整策略,为优化卷积神经网络的性能与泛化能力开辟了新的道路.

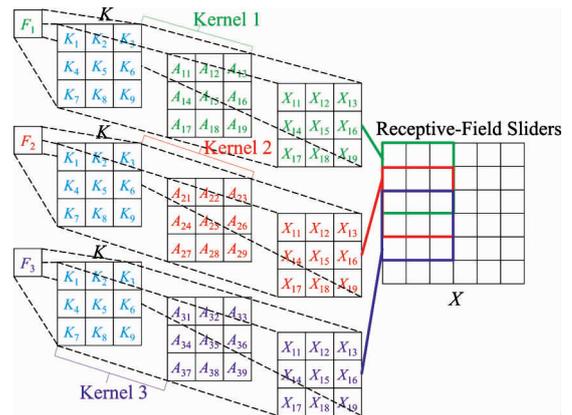


图 4 RFACConv 卷积模块操作

1.2 特征提取模块 C2f_DCNv3

可变形卷积 DCNv3(Deformable Convolutional Networks v3)^[23] 模块结构如图 5 所示.该卷积通过动态调整其采样位置,以适应不同尺度目标的特征提取需求,从而显著提升模型对多尺度目标的感知能力,尤其是在处理复杂多变的物体检测任务时,展现出更高的实用性和准确性.

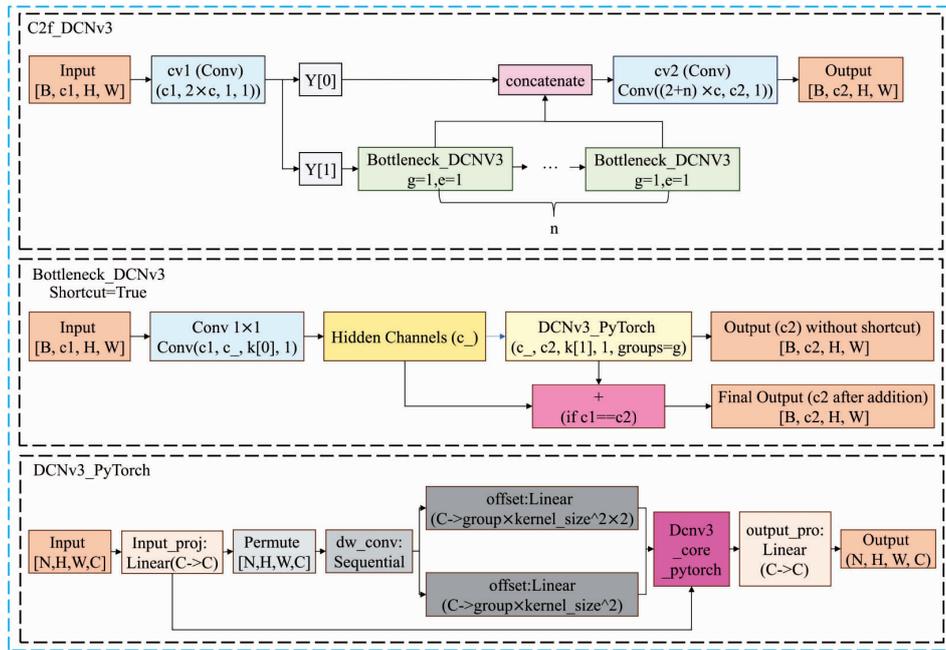


图 5 C2f_DCNv3 结构

图 6 展示了标准卷积和可变形卷积中的采样位置.在标准卷积(图 6(a))中,采样位置(绿色点)按照规则的网格形式排列.而在可变形卷积(图 6(b))中,采样位置(蓝色点)是通过引入偏移量进行变形的,并用增强的偏移量(浅蓝色箭头)进行表示,即在可变形卷积中,不再局限于规则的网格位置,而是可以根据需要在输入特征图上自由地进行采样.通过引入可变形卷积,可以推广各种变换,如尺度变换、(异向)长宽比和旋转等变换.图 6(c)和图 6(d)中进行了特殊情况的展示,即可变形卷积能够更灵活地适应不同类型的变换,从而增强了模型对目标形变的建模能力.

图 7 为可变形卷积的三维示意图,其中左侧的是输入特征,右侧的是输出特征,卷积核大小为 3×3,将输入特征中 3×3 区域映射为输出特征中的 1×1,如何选择 3×3 区域是其关键所在,传统的卷积是规则的形状,可变形卷积是在其中加入一个偏移量,对于每个点分别计算,改变 3×3 区域中每个点的选取,提取一些可能具有更丰富特征的点,从而提高检测效果.

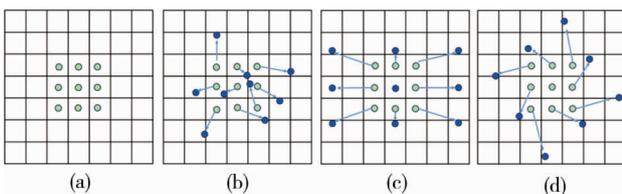


图 6 标准卷积与可变形卷积采样位置

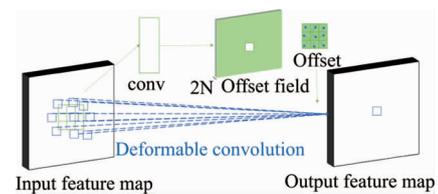


图 7 可变形卷积三维示意图

标准卷积(规则采样)在进行卷积操作时按照规则的网格位置进行采样,而可变形卷积通过引入偏移量来实现非规则采样,从而在形状变换(尺度、长宽比、旋转等)方面具有更强的泛化能力.可变形卷积的采样位置可以根据深度图像的背景与各种病害自适应地调整,从而在不同尺度中更准确地捕捉病害特征与隔离背景.因此在 YOLOv8 网络 Neck 的 C2f 中引入 DCNv3 得到 C2f_DCNv3 模块,为了应对资源受限的环境与本次试验中小规模的数据集,选择降低输出通道数,使用更少的权重参数,从而降低模块的计算量与参数量却不改变模型的准确率.

1.3 空间金字塔池化结构 SPPELAN

原 YOLOv8n 的 SPPF 部分是一种快速的空间金字塔池化层,用于从单一尺度的输入中提取多尺度的特征.由于 SPPF 模块的特征融合主要依赖于最终的卷积层,导致在特征融合时的细致程度不高.

基于路面深度图像中尺度变化较大,为了更好地识别多尺度目标与更细腻的特征融合,因此在 SPPF 中引入高效的层聚合网络结构 ELAN(Efficient Layer Aggregation Networks)^[24]得到 SPPELAN 模块,结构如

图 8 所示,通过一系列的卷积和最大池化操作来生成和融合更多尺度的特征,结合了卷积层和空间金字塔池化(Spatial Pyramid Pooling, SPP)层来增强特征表示.该模块首先通过一个卷积层 cv1,从输入通道 c1 转换为输出通道 c3,而后输出的结果分别执行相同的最大池化操作(核大小为 5,步长为 1).将所有最大池化操作的输出以及 cv1 的输出在通道维度上进行拼接,拼接后的结果通过卷积层 cv5 进行处理,从 $4 \times c3$ 通道转换为 c2 通道,最后输出处理后的特征图.

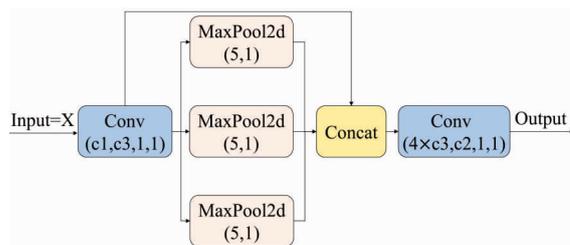


图 8 SPPELAN 结构

SPPELAN 模块相比 SPPF 模块优势在于其使用了多层最大池化操作(每个池化层的核大小为 5),这种多层次的池化操作能够更细致地捕捉多尺度的特征.在 SPPELAN 中,每个池化层的输出都会叠加到最终的特征向量中,这意味着它能够捕捉到更丰富的尺度信息.通过提取多尺度的特征,在不同的尺度上应用最大池化操作,可以增强网络对尺度变化的鲁棒性,帮助网络更好地适应不同尺度的裂缝目标,从而提高检测的准确性.通过结合卷积和最大池化操作,SPPELAN 模块可以生成更丰富的特征表示.这些特征表示不仅包含局部细节信息,还包含全局上下文信息,有助于网络更好地理解图像内容.与直接在不同尺度上应用卷积操作相比,SPPELAN 模块通过最大池化操作来生成多尺度特征,不仅保持了计算效率,还进一步提升了识别准确率.这一特点对于构建高效且性能卓越的网络结构至关重要.

1.4 改进算法理论分析与整体结构

带有深度信息的沥青路面病害种类包括横向裂缝、纵向裂缝、斜向裂缝、网状裂缝、路面沉降以及坑槽等,由于路面纹理与杂物的干扰,导致深度图像背景复杂.并且沥青路面深度图像所涉及的病害种类繁多,特征尺度多变,不能较好地处理多尺度与复杂背景的问题.

在深度学习中,感受野是卷积神经网络中的一个重要概念,它决定了网络中的每个神经元能够观察到的输入图像区域的大小.对于沥青路面深度图像病害检测任务而言,病害特征往往分布在不同的空间尺度上.因此,合理设计感受野对于提高检测精度至关重要.在网络主干的 C2f 中引入 RFACConv,使得网络能够根据输入图像的不同区域自动调整感受野的大小和形状.这种创新自适应机制有助于模型在处理复杂多变的沥青路面病害时,更加灵活地捕捉病害特征,从而提高检测的准确性和鲁棒性.通过动态调整感受野,RFACConv 使得网络能够更好地适应不同尺度和形状 of 病害特征,增强了模型对于复杂场景的处理能力.

沥青路面的病害往往伴随着形状和大小的变化,如裂缝的宽度、长度以及坑槽的深度等.传统的卷积神经网络在处理形变目标时存在困难,它们通常假设目标的形状和大小是固定的.DCNv3 的引入,为模型提供了一种处理形变目标的有效方法.DCNv3 的创新在于通过引入可学习的偏移量,使卷积核能够自适应地调整其采样位置,从而更有效地捕捉形变目标的特征.这种形变感知机制使得模型在处理复杂形变病害时,能够保持较高的准确性.在网络颈部的 C2f 中引入 DCNv3,网络能够学习到更加丰富的特征表示,这些特征表示对于识别沥青路面深度图像中的病害至关重要.

特征表示能力是深度学习模型性能的关键因素之一.对于沥青路面深度图像病害检测而言,有效的特征表示能够显著提高模型的检测精度和泛化能力.将网络中的 SPPF 模块替换为 SPPELAN 模块,为模型提供了一种更加高效和准确的特征表示方法.SPPELAN 模块在保持 SPP 多尺度特性的基础上,引入了 ELAN 模块来增强局部注意力.通过计算特征图之间的相关性,ELAN 模块为不同区域分配不同的权重,从而实现局部注意力增强,使得模型能够更加关注病害区域的细节特征,提高识别的准确性,同时 SPPELAN 模块还采用了多任务学习的方式,进一步提高了模型的性能和泛化能力.

从整体上看,改进模型 RBDN-YOLO-E 提高了识别沥青路面深度图像病害的性能,在处理复杂多变的沥青路面病害时,能够保持较高的准确性和鲁棒性.引入 RFACConv 模块使得网络能够更加灵活地捕捉病害特征;引入 DCNv3 模块增强了模型对于形变目标的处理能力;SPPELAN 模块替换 SPPF 模块,则进一步提高了模型的特征表示能力和检测精度.改进过后的 RBDN-YOLO-E 网络模型结构如图 9 所示,红色框中为修改模型的位置.

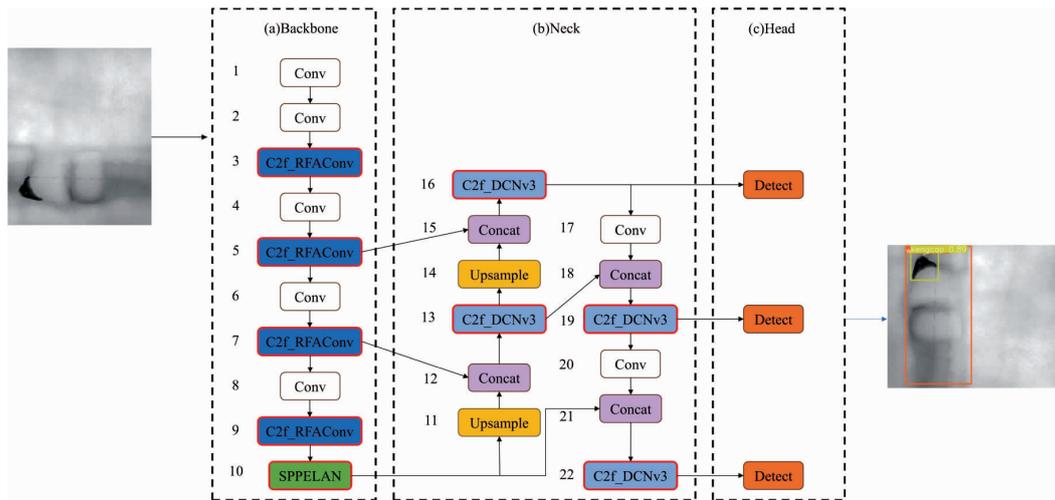


图9 RBDN-YOLO-E 网络结构

2 试验设计和结果分析

2.1 数据集介绍与处理

目前,对于沥青路面的研究普遍依赖于使用摄像机直接拍摄路面进行数据采集.由于道路上车流量大,这种方法不仅危险系数高、采集难度大,而且采集图像时间长,需要多人协作,对采集设备的要求也相当高.此外,缺乏可用的公开数据集也限制了研究的进展.为此,本文提出了采用无人机建立三维点云模型并导出深度图像的方法来建立试验所需的目标检测数据集.无人机采集图像的方式具有高效性,可以显著减少人员直接在道路上的危险作业,并且从三维点云模型中导出的深度图像仅包含路面深度信息,从而有效地排除了路面上杂乱信息的干扰.

本试验主要采用自制沥青道路目标数据集进行验证.此次采集采用大疆 Air 2S 无人机,配备 1 英寸 CMOS 传感器,像素为 2 000 万,焦距为 22 mm,拍摄分辨率为 5 472×3 648 像素,单像素尺寸为 2.4 μm,对 5 km 的沥青路面进行数据采集,采集到的图像以 JPG 格式保存.根据《低空数字航摄与数据处理规范》^[25],道路绘制规范对实景模型建立的第一等级误差,要求将影像的地面分辨率 GSD 控制在 40 mm 以下.为了清晰呈现路面细节(如裂缝、标志线等),地面分辨率通常需要达到厘米级甚至更精细的级别.具有深度信息的沥青路面病害,其中路面裂缝对于精度要求是最高的,根据沥青路面裂缝宽度分类,对路面造成影响的裂缝宽度为 3 mm 以上,即要求地面分辨率最高不超过 3 mm.规范中,相对航高的计算公式如式(2)所示.

$$H = \frac{f \times \text{GSD}}{a} \quad (2)$$

式中: H 为相对航高, m; f 为相机镜头的焦距, mm; GSD 为影像的地面分辨率, mm; a 为单像素尺寸, μm.

依据式(2)得到相对航高不高于 27.5 m,由于建立路面点云模型时既需要所拍摄图像略大于道路宽度,又需要影像的地面分辨率控制在 3 mm 之内,同时经过实地勘探与试飞,将航高 H 设定为 6 m,计算得出影像的地面分辨率为 0.65 mm,满足病害检测条件要求.

将采集后的影像通过 ContextCapture Center 软件,建立路面三维点云模型,示例如图 10 所示.将三维点云模型导入 CloudCompare 软件中,通过平面拟合确定基准路面,将 z 轴的高程信息投影至 xy 平面,并使用灰度信息表示路面高程,即低于基准路面将会呈现出深色,如图 11 所示为投影后的三维点云模型示例.使用 CloudCompare 软件,设置 xy 平面为基准面,将路面横向划分为 3 个区域,每个区域输出分辨率为 480×480 像素的 JPG 深度图像,如图 12 所示为其中一块区域输出的深度图像.5 km 沥青路面共输出 2 496 张深度图像以构成自制数据集.

该自制数据集由横向裂缝(heng)、纵向裂缝(zong)、斜向裂缝(xie)、网状裂缝(wang)、坑槽(kengcao)和路面沉降(chen)等 6 种常见的路面病害组成,即数据集共有 6 个种类标签,若呈现窄而长的深色区域,则为线型裂缝,再依据方向判断其为横向裂缝、纵向裂缝与斜向裂缝;若呈现网状的深色区域,则判断为网状裂缝;若呈现边界颜色剧烈变化的黑色区域,则为坑槽;若呈现边界颜色过渡自然的深色区域,则判断为路面沉

降.数据集按照 7 : 2 : 1 的比例随机将其划分为训练集、验证集和测试集.各类别实例数量如表 1 所示.



图 10 沥青路面三维点云模型示例

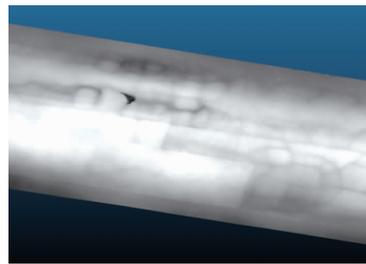


图 11 投影后的三维点云模型示例



图 12 深度图像示例

表 1 各病害数据分布情况

实例名称	标签名称	训练集标签数量/个	验证集标签数量/个	测试集标签数量/个
横向裂缝	heng	610	279	144
纵向裂缝	zong	750	349	89
斜向裂缝	xie	105	301	43
网状裂缝	wang	887	116	69
坑槽	kengcao	333	103	63
路面沉降	chen	253	47	18

2.2 试验设置

为了保证本文试验的公平性与合理性,所有对比试验以及消融试验均在相同环境与同样试验配置参数下进行.本文试验环境:基于 Ubuntu18.04 操作系统,英伟达 GeForce RTX3060Ti 显卡,Python3.9.7、CUDA11.1 作为编译环境,以 PyTorch1.12.1 作为网络训练框架,CPU 采用 Intel® CoreI5-14600KF CPU@2.6 GHz.训练参数:输入图像分辨率采用 480×480,并以随机梯度下降(SGD)优化器对模型进行优化,总的训练轮数(epoch)设置为 250,批次大小(batch)设置为 16,工作线程数(works)设为 8,初始学习率为 0.001,其他参数采用默认值.

2.3 评价指标

为了全面评估本文所提出算法的有效性,以 YOLOv8n 为基线模型,采用多个关键指标进行量化分析对比.这些指标包括多类别问题的平均精度均值(mAP)、精确度(Precision)、每秒的浮点运算数(GFLOPs)、模型参数量(Params)、F1 值(F1-Score)、召回率(Recall)、推理时间(Inference Time)以及模型大小.其中,mAP, Precision 和 F_1 分数是衡量模型检测精度提升效果的重要指标,mAP50 是在置信度阈值为 50%时的平均精度均值;推理时间是指深度学习模型从输入到输出的时间,即模型从接收输入数据到生成输出结果所消耗的时间,推理时间通常以毫秒为单位表示,除了直接计算时间外,还可以用每秒处理的帧数(FPS)来衡量模型的推理速度;GFLOPs、模型大小和模型参数量等指标则用于衡量模型的轻量化程度,在资源受限的环境中,一个轻量化的模型不仅能够降低硬件性能要求,还能减少能耗和存储空间的使用.

2.4 试验分析

2.4.1 C2f_RFACConv 模块有效性分析

C2f_RFACConv 模块允许网络利用感受野注意力机制为每个感受野生成特定的权重与动态调整卷积核的权重,从而为每个区域的特征提取提供了定制化的关注度.为了评估 C2f_RFACConv 模块在 YOLOv8n 模型中的效果及其在不同网络位置的性能表现,设计了详细的有效性分析试验.试验分为 3 种配置:C2f_RFACConv-Backbone(仅在主干网络中替换 C2f)、C2f_RFACConv-Neck(仅在颈部网络中替换 C2f)、C2f_RFACConv-All(在所有 C2f 中替换).

试验结果由表 2 可知:在 Neck 或模型所有 C2f 中引入 C2f_RFACConv 模块,与基线模型 YOLOv8n 相比较,尽管参数量、模型大小和计算量得到了控制,但 mAP50 却出现了下降.然而,当仅在 Backbone 中引入 C2f_RFACConv 模块时,与基线模型 YOLOv8n 相比较,不仅参数量减少 0.86 M、模型大小减少 1.7 M、计算量降低 0.8 GFLOPs,而且 mAP50 提高 0.1%.这一结果表明,在 Backbone 的 C2f 中引入 RFACConv 模块,能够在降低模型复杂度的同时,显著提升算法的检测准确性.Backbone 作为目标检测网络的核心组件,负责提

取图像中的基础特征,该特征对于后续的检测和识别至关重要.在 Backbone 的 C2f 中加入 RFACConv,意味着在特征提取的初始阶段引入感受野注意力机制,能够使网络在特征提取的初始阶段关注图像中的关键区域,提取到更加丰富和准确的特征.相比之下,虽然在颈部或所有 C2f 中引入 RFACConv 也能提高特征质量,但由于 Backbone 中的特征对整个网络的影响更为突出,因此在 Backbone 中引入 C2f_RFACConv 模块能够显著提高特征质量,进而提升模型的检测性能.

表2 C2f_RFACConv 模块有效性试验

模型	mAP50/%	Params/M	模型大小/M	GFLOPs
YOLOv8n	85.6	3.01	6.2	8.2
C2f_RFACConv-Backbone	85.7	2.15	4.5	7.4
C2f_RFACConv-Neck	85.1	2.81	4.5	7.7
C2f_RFACConv-All	85.3	2.17	4.6	7.5

2.4.2 C2f_DCNv3 模块有效性分析

DCNv3 是基于可变形卷积的改进版本,它可以根据输入特征的形状和位置动态地调整卷积核,从而更好地捕捉特征之间的空间关系.为了评估 C2f_DCNv3 模块在 YOLOv8n 模型中的效果及其在不同网络位置的性能表现,设计了详细的有效性分析试验.试验分为 3 种配置:C2f_DCNv3-Backbone(仅在主干网络中替换 C2f)、C2f_DCNv3-Neck(仅在颈部网络中替换 C2f)、C2f_DCNv3-All(在所有 C2f 中替换).

由表 3 中数据可知:3 种引入方式均降低了参数量、模型大小以及计算量.但是与基线模型 YOLOv8n 相比较,只有在 Neck 中引入 C2f_DCNv3 模块时,mAP50 提高 0.2%.这是因为一方面在目标检测网络中,Neck 部分通常负责连接 Backbone 和 Head,并对 Backbone 提取的特征进行进一步的处理和融合.由于 Neck 部分接收到的特征经过了 Backbone 的初步处理,因此它们通常包含较为丰富的语义信息.在 Neck 的 C2f 中加入 DCNv3,可以利用其动态调整卷积核的能力,对特征进行更加精细的处理和融合,从而提高网络的检测性能.另一方面,在 Neck 部分进行特征融合时,由于融合了多个不同尺度的特征,因此需要一种机制来确保这些特征能够有效地结合在一起.DCNv3 的可变形卷积机制可以为这种特征融合提供更加灵活和准确的手段,从而使得融合后的特征更加有利于目标检测任务的进行.如果在所有的 C2f 或 Backbone 中都加入 DCNv3,可能会增加网络的复杂度,导致过拟合的风险增加.而在 Neck 的 C2f 中单独加入 DCNv3,则可以在不增加过多参数的情况下提高网络的性能,从而降低了过拟合的风险

表3 C2f_DCNv3 模块有效性试验

模型	mAP50/%	Params/M	模型大小/M	GFLOPs
YOLOv8n	85.6	3.01	6.2	8.2
C2f_DCNv3-Backbone	84.3	1.97	4.2	6.8
C2f_DCNv3-Neck	85.8	2.08	4.4	6.9
C2f_DCNv3-All	85.2	1.94	4.1	6.7

2.4.3 SPPELAN 有效性分析

根据 ELAN 设计出的 SPPELAN 模块,用来替换基线模型 YOLOv8n 中的 SPPF 模块.为了验证其有效性,选取了当前较新且较热门的 AIFI^[26]以及 FocalModulation^[27]模块替换 SPPF 模块,与 SPPELAN 模块在相同数据集上进行对比试验.试验分为 3 种配置:YOLOv8-AIFI(SPPF 模块替换为 AIFI 模块)、YOLOv8-FocalModulation(SPPF 模块替换为 FocalModulation 模块)以及 YOLOv8-SPPELAN(SPPF 模块替换为 SPPELAN 模块).

试验结果如表 4 所示,可以看出替换了 AIFI 模块后,相较于基线模型 YOLOv8n,参数量、模型大小以及计算量有大幅度下降,mAP50 却随之下降了 1.5%.在替换了 SPPELAN 与 FocalModulation 模块之后,相较于基线模型 YOLOv8n,参数量、模型大小以及计算量均有大幅度提高,但替换了 SPPELAN 模块的模型,mAP50 得到了 1.1%的提升.这一结果表明,替换了 SPPELAN 模块的模型在极少量提高计算量的同时,模型准确率得到显著的提升.这得益于 SPPELAN 模块通过一系列多层最大池化操作,不仅提升了特征提取的尺度多样性,同时也确保了计算的高效性.这种设计使得网络能够捕获到多个尺度的特征信息,并在不同尺度上应用最大池化操作,从而显著增强了网络对尺度变化的适应性.这种适应性使得网络能够更有效

地处理不同尺寸的裂缝目标,进而提高了检测的精确性.

表 4 SPPELAN 模块有效性试验

模型	mAP50/%	Params/M	模型大小/M	GFLOPs
YOLOv8n	85.6	3.01	6.2	8.2
YOLOv8-AIFI	84.1	2.94	6.1	8.1
YOLOv8-FocalModulation	85	3.12	6.5	8.3
YOLOv8-SPPELAN	86.7	3.18	6.6	8.3

2.4.4 消融试验结果分析

为验证本文提出的改进模型在道路病害检测中的优势与对算法性能的影响,以 YOLOv8n 模型作为基线模型,将主干中的 C2f 替换为 C2f-RFAConv 模块得到 RB-YOLO 模型,将颈部中的 C2f 替换为 C2f_DCNv3 模块得到 DN-YOLO 模型,将 SPPF 替换为 SPPELAN 模块得到 YOLO-E 模型,RBDN-YOLO-E 为改进后针对道路病害深度图像检测的模型.

以上提及的模型均在相同数据集上试验,如表 5 所示.在基线模型 Backbone 的 C2f 中引入 RFAConv 模块,其关注了感受野空间特征,能动态调整卷积核的权重,更有效地处理图像中的细节和复杂模式,与基线模型 YOLOv8n 相比,在略微牺牲 FPS 的同时,少量提高了 mAP50 与 Precision,且模型计算量减少 9.76%,模型参数量减少 28.57%,模型大小减少 27.42%.在基线模型 Neck 的 C2f 中引入 DCNv3 模块,其能够根据输入特征的具体形状和位置来实时调整卷积核结构,更好地捕捉特征之间的空间关系,与基线模型 YOLOv8n 相比,FPS 虽有明显的下降,但是 mAP50 提高 0.2%,Precision 提高 2.87%,模型计算量减少 14.63%,模型参数量减少 30.9%,模型大小减少 29.03%,显著地降低了模型的计算成本.将基线模型 YOLOv8n 中的 SPPF 模块替换为 SPPELAN 模块,加强了特征提取的尺度多样性,模型 mAP50 和 Precision 显著提升 1.1%和 0.93%,FPS 提升 23.61%的同时,计算量、参数量以及模型大小都略微升高 1.22%,5.65%和 6.45%.融合重塑空间注意力卷积的基础上再融合可变形卷积,在略微降低 9.72%的 FPS 的情况下,mAP50 提高 0.6%,Precision 提高 0.45%,计算量降低 12.2%,参数量降低 29.57%,模型大小降低 27.42%.最后在 RBDN-YOLO 的基础上将 SPPF 模块替换为 SPPELAN 模块,得到 RBDN-YOLO-E 模型,相比较 RBDN-YOLO 模型,极少地牺牲 1.37%的计算量、7.02%的参数量以及 6.25%的模型大小,mAP50,Precision 与 FPS 显著地提高 1.1%、和 12.56%.相较于基线模型 YOLOv8n,mAP50、Precision 与 FPS 提高 1.7%,3.84%和 3.14%,计算量降低 10.98%,参数量降低 24.25%,模型大小降低 22.58%,改进过后的模型有效提高了精度的同时,也使模型得到了轻量化.

表 5 消融试验结果对比

模型	C2f_RFAConv	C2f_DCNv3	SPPELAN	mAP50/%	Precision/%	GFLOPs	Params/M	模型大小/M	FPS
YOLOv8n				85.6	81.48	8.2	3.01	6.2	216
RB-YOLO	√			85.7	82.23	7.4	2.15	4.5	194
DN-YOLO		√		85.8	84.35	7.0	2.08	4.4	176
YOLO-E			√	86.7	82.41	8.3	3.18	6.6	267
RBDN-YOLO	√	√		86.2	81.93	7.2	2.12	4.5	195
RBDN-YOLO-E	√	√	√	87.3	85.77	7.3	2.28	4.8	223

2.4.5 与不同算法结果对比

为了进一步验证本文所提模型的精度和轻量化程度的优势,本文对所提出的 RBDN-YOLO-E 模型与 FasterNet^[28],ResNet18^[29],GhostNetV2^[30],YOLOv5n,YOLOv6n,YOLOv8n,YOLOv9t 以及 YOLOv10n 等几种近些年性能最优,应用最广泛的目标检测网络进行对比试验,所有试验均在相同的硬件设备上测试,且均使用自制病害数据集进行训练,以保证试验的公平性与可靠性.在实际检测中,深度图像需经建模后导出,所以认为改进模型的准确率比实时性更重要,同时也需要兼顾轻量化.为了对以上模型进行量化评估,采用标准化加权求和方法计算评估分数 S.标准化加权求和公式如式(3)所示.

$$S = \sum_{i=1}^{i=7} w_i x_i. \quad (3)$$

式中: $x_1 \sim x_7$ 分别为 mAP50、F1、Recall、计算量倒数、参数量倒数、模型大小倒数、推理时间倒数; $w_1 \sim w_2$ 分

别为7个指标所对应的权重,本文分别取值0.40,0.15,0.15,0.10,0.10,0.05,0.05.

表6为试验结果与计算出的评估分数,其中粗体代表性能最佳指标.与以上8种模型相比较,本文所提出的模型RBDN-YOLO-E在mAP50、F1以及Recall这3个指标上均表现最优,在GFLOPs、Params、模型大小以及推理时间等4项轻量化指标上表现最佳.ResNet18模型展现出最小的计算量、参数量与模型体积,YOLOv6n模型展现出最低的推理时间,但以上两种模型mAP50较低.改进过后的RBDN-YOLO-E模型的mAP50、F1以及Recall均超越了其他8种目标检测网络模型,并且很好地控制了模型计算量,取得了最高的评估分数S,为66.09%.综上,本文所提出的RBDN-YOLO-E模型对多尺度下路面病害深度图像目标检测的效果更好,在提高精度与模型轻量化的同时,降低了推理时间,证明了此模型的实用性较高.

表6 不同模型性能比较

模型	mAP50/%	F1/%	Recall/%	推理时间/ms	GFLOPs	Params/M	模型大小/M	S/%
FasterNet	83.8	76	75.1	5.13	37	15.19	30.7	58.90
ResNet18	81.2	75	77.5	4.46	6.2	2.11	4.4	62.72
GhostNetV2	86.0	81	79.1	5.29	8.7	6.34	13.2	62.62
YOLOv5n	82.7	77	79.5	4.78	7.2	2.51	5.3	62.97
YOLOv6n	78.3	74	73.0	4.24	11.9	4.24	8.7	58.32
YOLOv8n	85.6	80	78.2	4.63	8.2	3.01	6.2	63.82
YOLOv9t	74.9	68	69.0	4.71	10.7	2.62	17.1	55.77
YOLOv10n	73.6	67	65.6	4.82	8.4	2.71	5.49	55.35
RBDN-YOLO-E	87.3	82	80.2	4.48	7.3	2.28	4.8	66.09

2.4.6 可视化对比试验分析

为直观展示改进模型在道路目标检测任务中的有效性,在自制数据集外选取了4张复杂的沥青道路病害深度图像作为测试数据.将本文改进模型RBDN-YOLO-E与FasterNet, ResNet18, GhostNetV2, YOLOv8n进行沥青道路目标可视化对比试验,各算法检测结果如图13所示.

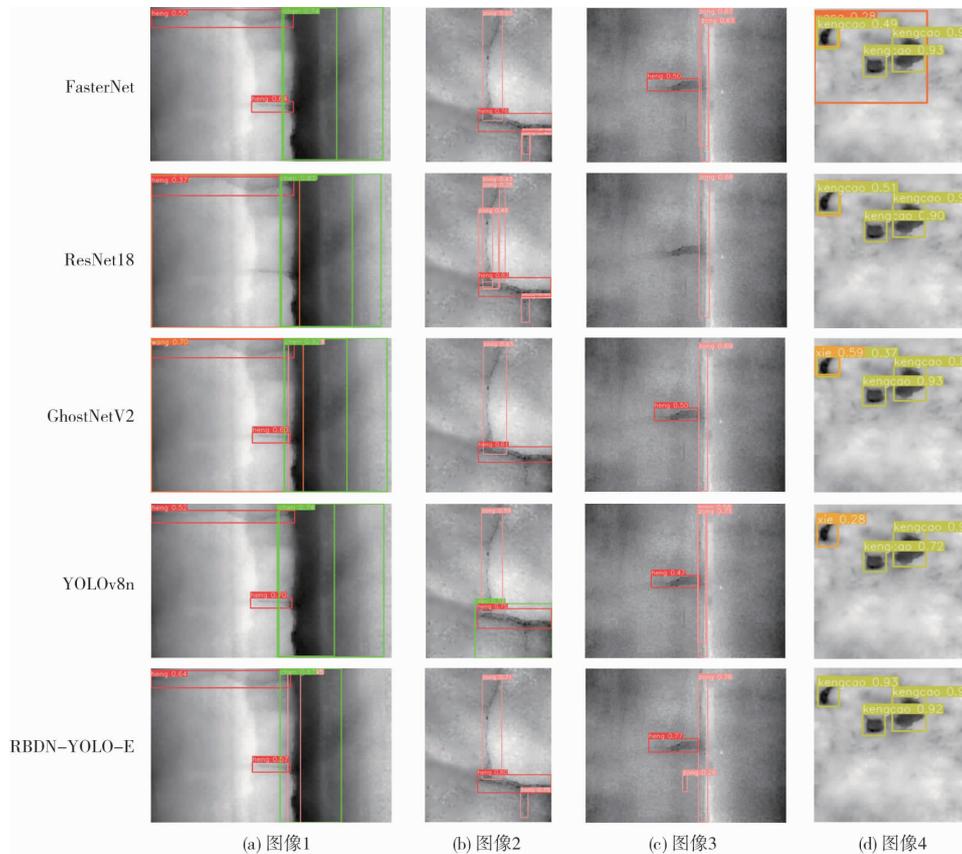


图13 检测效果对比

图13(a)组对比图像中,图像存在车道线干扰,且病害较多,造成图像复杂,增加了检测难度.FasterNet, ResNet18, YOLOv8n模型均存在漏检问题.由于车道线与裂缝相交,ResNet18与GhostNetV2模型均存在将其

误判为网状裂缝。FasterNet, ResNet18, GhostNetV2, YOLOv8n 模型均存在重复多检情况。而 RBDN-YOLO-E 模型具有更强的特征提取能力和判别能力,可以成功检测出所有病害且无误判与多检情况。

图 13(b)组对比图像中,图像分辨率较低且存在较小的裂缝,导致裂缝特征信息不明显, FasterNet, ResNet18, GhostNetV2, YOLOv8n 模型均有漏检与多检情况,而 RBDN-YOLO-E 模型使用了 SPPELAN 模块,有效加强了在分辨率较低情况下的小目标识别能力。

图 13(c)组对比图像中,裂缝的深度较浅,造成裂缝与背景之间的颜色对比度较小,且图像中含有细小裂缝。FasterNet, ResNet18, GhostNetV2, YOLOv8n 模型均存在漏检细小裂缝, FasterNet 与 YOLOv8n 模型均存在多检。由于 RBDN-YOLO-E 模型具有强大的区分目标与背景能力的可变形卷积 DCNv3,能成功检测出图像中所有病害的位置。

图 13(d)组对比图像中,路面不平整导致图像背景复杂,且 3 个坑槽的分布较为集中,导致 FasterNet 模型错误识别为网状裂缝, ResNet18, GhostNetV2, YOLOv8n 模型均将坑槽错误识别为斜向裂缝。RBDN-YOLO-E 模型在准确检测出所有坑槽的同时并未出现误检与多检,说明改进后的模型具有更强的背景分辨能力,处理较为集中的病害时更加细致,误检率更低。

可视化对比结果表明,本文改进模型 RBDN-YOLO-E 表现出了更高的准确性与鲁棒性,为实际应用提供了更高的价值。

为直观验证改进模型的泛化能力,选择与改进模型应用场景相关的路面重建数据集(Road Surface Reconstruction Dataset, RSRD)^[31]作为测试数据,该数据集采用激光雷达、双目摄像头等设备实车采集,其中包括沥青路面点云模型与沥青路面实拍图像。通过激光雷达采集的沥青路面点云模型,经 CloudCompare 软件转换为深度图像后,选择其中 2 张复杂的深度图像作为部分测试数据;通过双目摄像头采集的沥青路面图像,使用 ContextCapture Center 软件建立沥青路面点云模型,并在 CloudCompare 软件中转换为深度图像,选择其中两张复杂的深度图像作为另一部分测试数据。将本文改进模型 RBDN-YOLO-E 与 FasterNet, ResNet18, GhostNetV2, YOLOv8n 进行沥青道路目标可视化对比试验,各算法检测结果如图 14 所示。

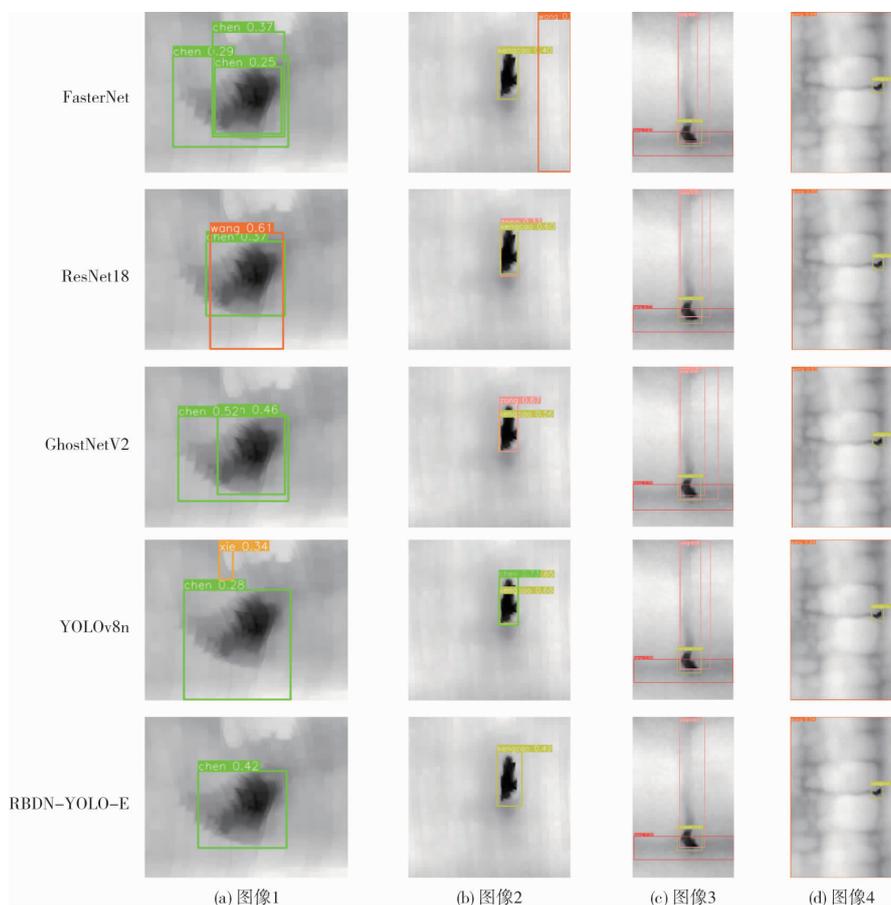


图 14 在 RSRD 数据集中各算法检测效果对比

图 14(a)、图 14(b)以及图 14(c)组对比图像中, FasterNet, ResNet18, GhostNetV2 与 YOLOv8n 均存在重复多检和误判情况, 而本文改进模型 RBDN-YOLO-E 成功检测出所有病害且无误判与多检情况. 图 14(d)组对比图像中, 所有算法均正确检测出所有病害且无误判与多检情况.

可视化对比结果表明, 本文改进模型 RBDN-YOLO-E 在公共数据集中仍然保持了较高的准确率, 表现出了更强的泛化能力.

3 结论

1) 针对沥青路面病害特征分布在不同空间尺度上的问题, 在 YOLOv8n 模型网络主干的 C2f 中引入重塑空间注意力卷积 RFACConv, 设计出 C2f_RFACConv 模块, 为每个区域的特征提取提供了定制化的关注度.

2) 针对沥青路面的病害伴随着形状和大小变化的问题, 在 YOLOv8n 模型网络颈部的 C2f 中引入可变形卷积 DCNv3, 并降低输出通道数以降低模型计算成本, 设计出 C2f_DCNv3 模块, 以捕捉特征与空间之间的关系, 精确识别病害目标并准确分离出背景.

3) 将 YOLOv8n 模型中的 SPPF 模块替换为 SPPELAN 模块以生成更多尺度, 为图像不同区域分配不同的权重, 实现局部注意力增强, 加强特征提取的尺度多样性.

4) 试验结果表明, 相比原 YOLOv8n 模型, 改进后的 RBDN-YOLO-E 模型对于多尺度的沥青路面深度图像病害检测性能更加优越, 泛化能力更强, mAP50 与 Precision 分别提高了 1.7% 和 4.29%, 计算量、参数量和模型大小分别降低了 0.90 GFLOPs、0.73 M、1.40 M, 推理时间降低了 0.15 ms. 因此该模型具有精度更高、模型更轻量化且推理时间更短的优势, 为道路病害的自动化检测提供了更加可靠的技术支持.

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. ACM, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015: 1440-1448.
- [3] 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.
- [4] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 2980-2988.
- [5] 孙朝云, 裴莉莉, 李伟, 等. 基于改进 Faster R-CNN 的路面灌封裂缝检测方法[J]. 华南理工大学学报(自然科学版), 2020, 48(2): 84-93.
- [6] 牛慧余, 包腾飞, 李扬涛, 等. 基于改进 Mask R-CNN 的混凝土坝裂缝像素级检测方法[J]. 水利水电科技进展, 2023, 43(1): 87-92.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]. Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [8] FARHADI A, REDMON J. Yolov3: An incremental improvement[C]//Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection[J]. ArXiv e-Prints, 2020; arXiv: 2004.10934.
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 779-788.
- [11] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arxiv preprint arxiv: 2209.02976, 2022.
- [12] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023: 7464-7475.

- [13] TALAAT F M, ZAINELDIN H. An improved fire detection approach based on YOLO-v8 for smart cities [J]. *Neural Computing and Applications*, 2023, 35(28): 20939–20954.
- [14] 高晶, 姚金杰, 刘鹏杰, 等. 基于改进YOLOv5s的路面病害检测[J]. *计算机系统应用*, 2024, 33(9): 253–260.
- [15] WU C G, YE M, ZHANG J L, et al. YOLO-LWNet: a lightweight road damage object detection network for mobile terminal devices[J]. *Sensors*, 2023, 23(6): 3268.
- [16] 王海群, 王炳楠, 葛超. 重参数化YOLOv8路面病害检测算法[J]. *计算机工程与应用*, 2024, 60(5): 191–199.
- [17] 卢小平, 张航, 张冬梅, 等. 一种基于无人机影像的迭代二值化道路裂缝检测方法[J]. *河南理工大学学报(自然科学版)*, 2019, 38(6): 55–59.
- [18] 贺泳超, 陈秋南, 程家杰. 基于改进YOLOv7算法的隧道衬砌裂缝智能识别[J]. *湖南科技大学学报(自然科学版)*, 2024, 39(2): 35–43.
- [19] 胡鹏, 夏晓华, 钟预全, 等. 采用多尺度特征增强的路面病害检测模型[J/OL]//*西安交通大学学报*, 1–12[2024-09-11].
- [20] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 13708–13717.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018: 3–19.
- [22] ZHANG X, LIU C, YANG D G, et al. RFAConv: innovating spatial attention and standard convolutional operation [EB/OL]. 2023: 2304.03198. <https://arxiv.org/abs/2304.03198v6>.
- [23] WANG W H, DAI J F, CHEN Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023: 14408–14419.
- [24] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information [C]// *Computer Vision-ECCV 2024*. Cham: Springer Nature Switzerland, 2025: 1–21.
- [25] 国家市场监督管理总局, 国家标准化管理委员会. 低空数字航摄与数据处理规范: GB/T 39612—2020 [S]. 北京: 中国标准出版社, 2020.
- [26] ZHAO Y A, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2024: 16965–16974.
- [27] YANG J, LI C, DAI X, et al. Focal modulation networks [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 4203–4217.
- [28] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023: 12021–12031.
- [29] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 770–778.
- [30] TANG Y, HAN K, GUO J, et al. GhostNetv2: Enhance cheap operation with long-range attention [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 9969–9982.
- [31] ZHAO T, XU C, DING M, et al. RSRD: A road surface reconstruction dataset and benchmark for safe and comfortable autonomous driving [J]. *arxiv preprint arxiv: 2310.02262*, 2023.