E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP18;TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2026)10-3704-18

论文引用格式:Wang Y S, Feng J L and Fang M. 2026. Multi-stage spatial-aware network for road damage detection. Journal of Image and Graphics, 31(10):3704-3721(汪裕盛, 冯俊龙, 方明. 2026. 多阶段空间感知道路损伤检测网络. 中国图象图形学报, 31(10):3704-3721)[DOI:10.11834/jig.250422]

多阶段空间感知道路损伤检测网络

汪裕盛[1],冯俊龙[2],方明[2,3*]

1. 长春理工大学计算机科学技术学院,长春　130022; 2. 长春理工大学人工智能学院,长春　130022;
3. 长春理工大学中山研究院,中山　528437

摘　要:**目的**　道路损伤检测是道路基础设施智能养护与交通管理的核心环节。传统检测方法易受环境干扰且泛化能力不足,而当前深度学习方法多聚焦于目标提取能力等单维度优化,忽视损伤目标存在细长拓扑结构、形态不规则性及低对比度等复杂物理特性,无法实现高效检测。为解决上述问题,提出了多阶段空间感知与细节增强道路损伤检测模型。**方法**　设计空间感知与细节增强的混合注意力模块,通过全局方向感知与细节强化的协同机制,构建长程空间依赖关系,同时显著提升对弱纹理、边缘模糊损伤的细节表征能力。构建跨尺度特征交叉融合模块,优化网络颈部架构以实现跨尺度特征的异构级联融合,有效平衡局部空间细节与全局语义信息的协同表达。此外,改进的C3K2(CSP three-convolution kernel)模块嵌入坐标感知卷积,通过空间信息增强有效优化高维特征的空间耦合建模效能。**结果**　在RDD2022(road damage dataset)基准数据集上的系统实验表明,本文模型有效识别各类道路损伤,在保持142 帧/s实时推理速度的同时,mAP@0.5(mean average precision)、mAP@0.5:0.95和F1-Score较现有最优方法分别提高了1.9%、4.8%和1.8%,其中mAP@0.5达到87.7%,消融实验验证了各模块的贡献度。跨数据集测试与泛化测试进一步证实该模型具备优异的检测鲁棒性与工程适用性。**结论**　本研究构建的多级协同优化框架,通过空间感知强化与细节特征增强的机制耦合,为道路基础设施智能养护提供了具有显著工程价值的技术解决方案。论文代码开源于https://www.scidb.cn/s/RZBnIz。

关键词:道路损伤检测;空间感知;跨尺度特征融合;YOLO(you only look once);注意力机制;实时目标检测

Multi-stage spatial-aware network for road damage detection

Wang Yusheng[1], Feng Junlong[2], Fang Ming[2,3*]

1. School of Computer Science and Technology, Changchun University of Science and Technology, Changchun　130022, China;
2. School of Artificial Intelligence, Changchun University of Science and Technology, Changchun　130022, China;
3. Zhongshan Institute, Changchun University of Science and Technology, Zhongshan　528437, China

Abstract: **Objective** Road damage detection is a fundamental component of the intelligent maintenance and management of transportation infrastructure. As global transportation systems continue to modernize and road networks expand rapidly, pavements inevitably deteriorate because of a range of external and internal factors. Traditional inspection methods, including manual surveys and existing automated approaches, still face major challenges in scalability, efficiency, and accuracy, which limits their suitability for large-scale deployment. There is therefore an urgent need for a high-performance, accurate, and robust framework for road damage detection and maintenance. In practical scenarios, road-surface damage often

收稿日期:2025-09-22;**修回日期**:2026-03-06;**预印本日期**:2026-03-13

*通信作者:方明　fangming@cust.edu.cn

基金项目:中山市科技局引进科研创新团队项目(CXTD2023005)

Supported by:Zhongshan Science and Technology Bureau Introduces Research and Innovation Team Project(CXTD2023005)

exhibits elongated geometries with high aspect ratios, low texture, low contrast, and strong heterogeneity. These characteristics pose fundamental challenges to existing algorithms, which tend to focus on local texture or semantic cues while overlooking the intrinsic physical properties of road damage. In particular, such models fail to capture the directional feature coupling inherent in the continuous spatial distribution of cracks and the edge attenuation of blurred potholes. The lack of direction-aware features and precise positional cues weakens long-range dependency modeling and reduces responsiveness to targets with indistinct boundaries. In addition, the hierarchical convolution and downsampling operations in convolutional neural networks (CNNs) progressively degrade cross-scale features, causing the loss of critical detail from high-resolution shallow feature maps. **Method** To address these limitations, this study proposes a novel multi-stage spatial-aware detection network (MSANet) built on the YOLOv11 framework. The design follows a progressive feature enhancement-fusion-optimization strategy tailored to the physical characteristics of elongated cracks and blurred-edge potholes. In the direction-sensitive feature enhancement stage, a hybrid Manhattan self-attention and coordinate attention (MCA) mechanism is introduced. Manhattan self-attention models the spatial continuity of crack axes, while coordinate attention improves the localization of blurred-edge defects. Together, these components create a dual-driven enhancement scheme that strengthens both directionality and detail representation, enabling the model to capture long-range spatial dependencies more precisely. To mitigate feature degradation caused by repeated downsampling, a cross-scale feature concat (CFC) module is incorporated into the neck. By cascading feature maps across resolutions, the CFC module balances semantic completeness at larger scales with detail richness at smaller scales, thereby compensating for the loss of fine-grained information. Finally, in the detection head, a coordinate-aware convolution kernel (C3K2_CAC) is embedded within the C3K2 block. This module jointly optimizes spatial position and channel response through direction-aware spatial attention and dynamic channel reweighting, producing anisotropic representation enhancement in high-dimensional feature space. While maintaining computational efficiency, it significantly improves edge clarity, detection robustness, and global discriminative capability. During training, experiments were conducted on a standardized platform using Ubuntu 20.04 and PyTorch. YOLOv11s served as the baseline for comparative and ablation studies. Across all experiments, the same hyperparameters were used: a batch size of 32, a learning rate decayed from 0.01 to 0.000 1, an image resolution of 640 × 640, stochastic gradient descent (SGD) as the optimizer, and a training schedule of 300 epochs. All other settings followed the default network configuration to ensure fairness and reproducibility. **Result** Experiments were conducted on the publicly available RDD2022 Chinese subset, which contains multiple road-damage categories, including longitudinal cracks (D10), transverse cracks (D10), alligator cracks (D20), potholes (D40), and repaired regions. Performance was evaluated using seven standard metrics: precision (P), recall (R), mAP@0.5, mAP@0.5:0.95, F1-score, frames per second (FPS), and GFLOPs. The results show that MSANet significantly outperforms existing state-of-the-art detectors. Compared with mainstream models, MSANet achieves an mAP@0.5 of 0.877 and a comprehensive performance score of 0.845 (Recall: 0.811, Precision: 0.883, F1-score: 0.845), surpassing YOLOv8s (0.824) and RT-DETR (0.801). Its mAP@0.5:0.95 improves by 4.8% over the baseline, confirming the robustness of the proposed design. In fine-grained damage detection, MSANet further demonstrates strong adaptability, achieving a 4.15% increase in mAP@0.5 (to 0.878), a 3.18% increase in mAP@0.5:0.95 (to 0.649), and a 2.69% gain in F1-score (to 0.840). Importantly, the model maintains a high inference speed of 142 frame/s, representing only a 6% reduction from the baseline while still substantially outperforming real-time detectors such as RT-DETR. This balance between accuracy and efficiency supports practical deployment in real engineering settings, where rapid decision-making and reliable predictions are both essential. Ablation studies further validate the contribution of each proposed module. The MCA module improves performance without significant computational overhead. The CFC module effectively fuses low-level detail with high-level semantics across network stages, introducing cross-scale contextual information that yields stable gains of 0.5% in mAP@0.5 and 2.3% in mAP@0.5:0.95. The C3K2_CAC module introduces only a modest increase in computational cost while enhancing the spatial awareness of neck features and supplying the detection head with more discriminative representations. Detection heatmaps further confirm that MSANet provides better focus on target regions, stronger suppression of background noise and irrelevant activations, and more complete coverage of damage areas. To assess generalization and deployment potential, experiments were also performed on additional public datasets and field-collected road images, where MSANet

achieved consistently promising results, demonstrating strong robustness and scalability across diverse environments. **Conclusion** Efficient and accurate road damage detection is crucial for the intelligent maintenance of transportation infrastructure. By combining spatial-awareness enhancement with detail preservation, the proposed MSANet framework addresses long-standing challenges in directional modeling and fine-grained feature retention. The model achieves substantial improvements in detection accuracy while preserving real-time inference capability, offering both theoretical novelty and engineering practicality. MSANet therefore represents a promising solution for intelligent road infrastructure management and provides useful insights for the design of next-generation detection frameworks that seek to balance efficiency, accuracy, and scalability. With its strong adaptability to both benchmark datasets and real-world imagery, MSANet also lays a foundation for future integration into large-scale intelligent transportation systems and smart-city applications. The code for this paper will be available at https://www.scidb.cn/s/RZBnIz.

Key words: road damage detection; spatial perception; cross-scale feature fusion; you only look once (YOLO); attention mechanism; real-time object detection

0 引 言

随着全球交通现代化的推进,各国路网规模显著扩大并持续增长,同时由于气候变化、车辆超载及材料老化等多重因素,路面在使用过程中易产生裂缝、坑洞等损伤(de Abreu等,2022)。若不能及时发现与修复,将加速路面结构性能的衰减,导致养护成本显著上升,并进一步降低道路通行效率。传统人工巡检在小范围内具有一定灵活性,能够直观识别路面病害,但其检测效率低且缺乏统一的标准化评估流程,难以满足大规模路网精细化管理的需求。近年来,基于地面穿透雷达、红外成像及激光轮廓仪等多传感器的自动化检测技术逐渐应用于路面状态评估,在检测效率和精度方面具有明显优势,但其设备成本高、系统复杂且运维难度大等问题,仍制约了在不同地区的大范围推广应用(Ragnoli等,2018)。

现有路面损伤检测方法可分为两大类:基于传统视觉算法的方法与基于深度学习的方法。前者主要依赖图像处理、边缘检测及纹理分析等手段,对裂缝、坑洞等损伤进行识别与评估。该类方法实现相对简单、计算开销较小,但在复杂光照、噪声干扰及非结构化路面等场景下鲁棒性不足,难以满足大范围、全天候道路检测与管理的需求。

深度学习方法又可分为单阶段检测与两阶段检测。两阶段检测算法(如R-CNN(regions with convolutional neural network)、Faster R-CNN)(Girshick等,2014;Ren等,2017)通常先生成候选区域,再逐步精细化分类与回归,具备更高的定位精度,但此类方法处理流程复杂、计算量大,难以满足实时性检测场景的要求;单阶段检测算法(如SSD(single shot multibox detector)、YOLO(you only look once)系列)(Liu等,2016;Redmon等,2016)则将特征提取与目标分类、定位统一在同一网络中。该类算法具有检测速度快、推理效率高等优势,但其单一设计在复杂道路场景中仍面临细节丢失、对细长裂缝等目标建模能力不足等问题。

在道路损伤检测领域,以裂缝类为主导的路面损伤由于具有高长宽比结构、低纹理特征、低对比度及几何形体的高度异质性等特质,严重制约检测算法性能的提升。现有研究多借助注意力机制(如CBAM(convolutional block attention module)、LSKA(large separable kernel attention)模块)(Woo等,2018;Lau等,2024)增强损伤区域的显著性响应,但此类方法过度侧重于局部纹理或语义信息的聚合(Zhang等,2022),对路面损伤在连续空间分布下所呈现的方向性特征耦合及边缘响应空间衰减等本质物理属性刻画不足,方向感知特征缺失与位置信息表征模糊,削弱了模型对目标长程依赖关系的建模能力和边缘模糊目标的响应能力。

此外,在卷积神经网络层级卷积与下采样过程中,浅层高分辨率特征图中蕴含的关键细节信息(如微小裂缝纹理、细长损伤边缘轮廓)出现了显著尺度衰减。针对上述方向感知不足与细节信息弱化的问题,本文基于YOLOv11架构提出多阶段空间感知与细节增强道路损伤检测模型(multi-stage spatial-aware road damage detection network,MSANet)。本文的主要创新点及贡献如下:

1)提出方向感知混合注意力模块(MaSA and CA attention,MCA)。该模块通过方向特征增强与细

节纹理强化的协同作用，提升了空间结构建模与精细纹理感知能力，实现沿裂缝等损伤的方向自适应特征强化。该机制显著提升模型对低纹理、边缘模糊损伤的长程空间建模与细节表征能力，有效解决方向感知特征缺失问题。

2）设计跨尺度特征融合模块（cross-scale feature connection，CFC）。针对道路损伤检测中跨尺度目标特征融合困难和下采样过程中的细节信息丢失问题，采用异构连接架构与双池化特征对齐融合策略，通过非对称跳跃连接，实现高层语义特征与低层细节特征的跨尺度高效对齐与互补，弥补下采样造成的细节缺失，从而增强了模型对损伤目标的细节融合能力。

3）构建空间信息强化的 C3K2 模块（C3K2_CAC）。通过在 C3K2 模块中引入坐标注意力卷积以突出关键区域的通道响应，并增加融合残差路径保障原始信息传递，从机制上显著强化了空间结构与通道特征的协同表达，实现高维特征空间的表征增强，在维持高效计算结构的同时，有效提升路面损伤的边缘清晰度与全局检测精度。

在道路损伤检测数据集 RDD2022（road damage dataset 2022）中，在维持 142 帧/s 的高实时推理速度的同时，mAP@0.5（mean average precision）、mAP@0.5∶0.95 和 F1-Score 较 YOLOv11s 基准模型分别提高了 1.9%、4.8% 和 1.8%，mAP@0.5 达到 0.877。

1　相关工作

1.1　YOLO 网络模型用于道路损伤领域

YOLO（you only look once）系列模型凭借其端到端、高效且具备实时性等优势，已经成为道路损伤检测领域的主流目标检测框架。近年来，针对路面损伤目标尺度小、背景干扰复杂以及移动端部署受限等问题，研究者围绕各版本 YOLO 模型开展了多维度改进，有效提升了检测精度与工程适用性。Li 和 Li（2024）在 YOLOv8 框架基础上引入卷积块注意力机制与选择性核网络，增强了通道—空间联合特征建模能力与多尺度特征提取能力，实现了对道路缺陷的高精度识别与定位。Ma 等人（2022）提出路面裂缝生成对抗网络（pavement crack generative adversarial network，PCGAN）以及基于 YOLO-MF 的裂缝检测与跟踪系统，构建了完整的路面裂缝自动智能检测与跟踪方案。彭毅和邵宇飞（2025）提出基于 YOLOv8n 的优化算法 YOLOv8n-PIVI，降低模型参数量与计算量，提高检测精度和速度。作为 YOLO 系列的最新演化，YOLOv11（Khanam 和 Hussain，2024b）在 YOLOv8（Swathi 和 Challa，2024）的基础上，通过引入 C3K2 模块和并行空间注意力机制（C2PSA）等多阶段特征增强模块，实现了检测精度与推理速度的更优平衡。在多个道路损伤检测的基准数据集上，YOLOv11 在检测精度与实时性能方面均优于现有主流模型。

1.2　注意力机制在道路损伤检测领域的应用

注意力机制通过对特征分配不同权重以突出关键区域与显著特征，已成为应对道路场景中复杂背景干扰和目标形态多样性的重要技术路径。模块化注意力机制通常在预定义的特征维度上进行分组与加权，具有参数量较小、易于解释和便于嵌入现有网络结构等优点，但对长程依赖关系的建模能力有限；自注意力机制基于 Q-K-V 相似度动态学习元素间依赖，可实现全局上下文建模，但存在计算开销大、可解释性相对不足等问题。近年来，大量研究从空间—通道注意力基础模块出发，逐步融合 Transformer 全局建模、噪声抑制与多尺度特征融合等创新机制，推动检测精度与效率的协同提升。Jiang 等人（2023）提出基于 Transformer 的视觉中心显式建模框架，通过捕捉长距离依赖关系与关键特征聚合机制，在无人机（unmanned aerial vehicle，UAV）平台的道路裂缝检测任务中验证了其较高的识别准确性。张志华等人（2022）提出在骨干网络中融入空洞卷积和两种注意力机制的方法，将其中的轻量型注意力机制与残差模块结合为残差注意力模块，有效解决复杂背景下道路裂缝难检测的问题。Chu 等人（2022）构建了一种融合注意力机制的多尺度特征融合网络 TCN（ting-crack-net），利用改进的残差网络提取微小裂缝的局部特征，并在网络中集成双注意力模块，有效提升了细小裂缝与复杂背景的分离能力。许正森等人（2023）提出一种基于 Transformer 架构的深度神经网络道路裂缝检测模型 CTNet（crack Transformer network），实现强噪声背景下的道路裂缝检测。相关研究结果表明，引入注意力机制能够显著增强道路损伤检测中的特征选择性与空间定位能力，为高精度检测提供了有力支撑。

1.3 道路损伤检测领域的挑战和解决方案

受不同国家和地区路面材料、施工工艺及光照、天气等环境条件差异的影响，道路损伤在外观形态与纹理特征上呈现高度多样性；细长条形裂缝等目标表现的纹理往往与周围背景相似，导致模型易出现漏检和误检。同时，在边缘计算设备与移动终端上，实时性要求与有限计算资源之间的矛盾依然突出。为推动道路损伤检测技术的发展，自2018年起，IEEE大数据杯连续举办世界道路损伤检测挑战赛，并先后发布RDD2020（Arya等，2020）、RDD2022（Arya等，2022）等全球多源道路损伤数据集，为研究者提供了统一的评价标准和丰富的标注样本。在该赛事上，ShiYuSeaView团队（2022）采用YOLO系列模型进行密集预测，并结合以Swin Transformer为骨干的Faster R-CNN扩展感受野，实现了对复杂场景道路损伤的高精度检测。其他参赛队伍则在高分辨率图像处理、注意力机制设计以及实时推理效率优化等方面提出了多种创新方案，为道路损伤检测技术的持续演进与工程落地提供了重要参考和探索方向。

2 方 法

2.1 总体架构

本研究针对道路损伤检测场景中裂缝类细长目标（高长宽比、低纹理对比度）与坑洞类边缘模糊目标（几何异质性、弱边界特征）等经典难题，提出一种基于YOLOv11架构的多阶段空间感知与细节增强网络（MSANet）。网络总体结构如图1所示。本文网络模型通过3个关键创新模块实现对复杂道路损伤的精准检测。在特征增强阶段引入混合注意力模块（MCA），融合全局方向感知和局部细节强化机制，实现空间结构建模与精细纹理感知的协同增强，显著提升细长损伤的响应强度和细节表征能力。同时，在特征融合阶段重新架构了跨尺度特征交叉融合模块（CFC），通过非对称层级特征的跳跃连接与级联整合，兼顾深层特征的语义完整性与浅层特征的细节丰富性，有效缓解多次下采样带来的细节信息丢失问题。进一步，在检测头前的C3K2（CSP three-convolution kernel 2）模块中引入坐标感知卷积核（coordinate attention convolution，CAConv），通过坐标注意力生成空间敏感权重矩阵与残差融合，强化高维特征的方向感知能力与通道表达能力，实现跨层特征融合与空间位置建模的有效结合。整体而言，MSANet通过多阶段的空间感知增强与细节保持机制，提升了对复杂道路损伤目标的检测鲁棒性与定位精度。

2.2 方向感知与细节增强注意力模块MCA

在细长道路损伤目标检测任务（如裂缝检测）中，通过坐标注意力机制（coordinate attention，CA）（Hou等，2021）将空间坐标信息嵌入通道注意力，利用其方向感知优势，对特征进行方向建模和空间位置编码，为下游任务提供具有方向感知的上下文信息，原理如图2所示。

具体而言，CA将传统通道注意力中的二维全局池化拆解为沿水平方向和垂直方向的两个一维特征聚合过程，输入特征$\boldsymbol{X} \in \mathbf{R}^{C \times H \times W}$，CA分支沿水平与垂直方向对$\boldsymbol{X}$进行全局平均池化，分别获得一维向量$\boldsymbol{f}_{\mathrm{h}}$和$\boldsymbol{f}_{\mathrm{w}}$，具体为

$$\boldsymbol{f}_{\mathrm{h}}(c,h)=\frac{1}{W}\sum_{w=1}^{W}X(c,h,w) \tag{1}$$

$$\boldsymbol{f}_{\mathrm{w}}(c,w)=\frac{1}{H}\sum_{h=1}^{H}X(c,h,w) \tag{2}$$

将$\boldsymbol{f}_{\mathrm{h}}$和$\boldsymbol{f}_{w}$在空间特征维度拼接之后，通过1×1卷积降维后得到紧凑表示$\boldsymbol{f}$。随后，利用两组平行的1×1卷积将$\boldsymbol{f}$映射生成沿两个方向的注意力图$\boldsymbol{M}_{\mathrm{h}} \in \mathbf{R}^{C \times H \times 1}$和$\boldsymbol{M}_{\mathrm{w}} \in \mathbf{R}^{C \times 1 \times W}$，并经过sigmoid激活得到权重系数，两个方向的注意力图为每个通道的各空间提供了权重矩阵，并与原始特征逐元素相乘，从而完成对输入特征的空间和通道联合重标定，得到CA增强后的输出$\boldsymbol{X}_{\mathrm{CA}}$，具体为

$$\boldsymbol{X}_{\mathrm{CA}}=\boldsymbol{X}_{\mathrm{CA}} \odot \sigma(\boldsymbol{M}_{h}) \odot \sigma(\boldsymbol{M}_{w}) \tag{3}$$

然而，由于CA将二维空间信息压缩为水平方向和垂直方向的全局描述，它在捕捉细粒度空间细节方面仍显不足，对局部结构和纹理的精细刻画能力有限。

为了平衡空间感知与局部细节建模，引入了基于曼哈顿距离的空间自注意力机制（Manhattan self-attention，MaSA）（Fan等，2024），通过在注意力计算中融入曼哈顿距离作为衰减权重，使模型更加聚焦于局部邻域，精细建模局部细节特征，有效弥补了CA在细节和邻域关系建模上的不足，并对CA提供的全局方向信息形成有益补充。

具体而言，MaSA引入基于空间位置距离的衰减

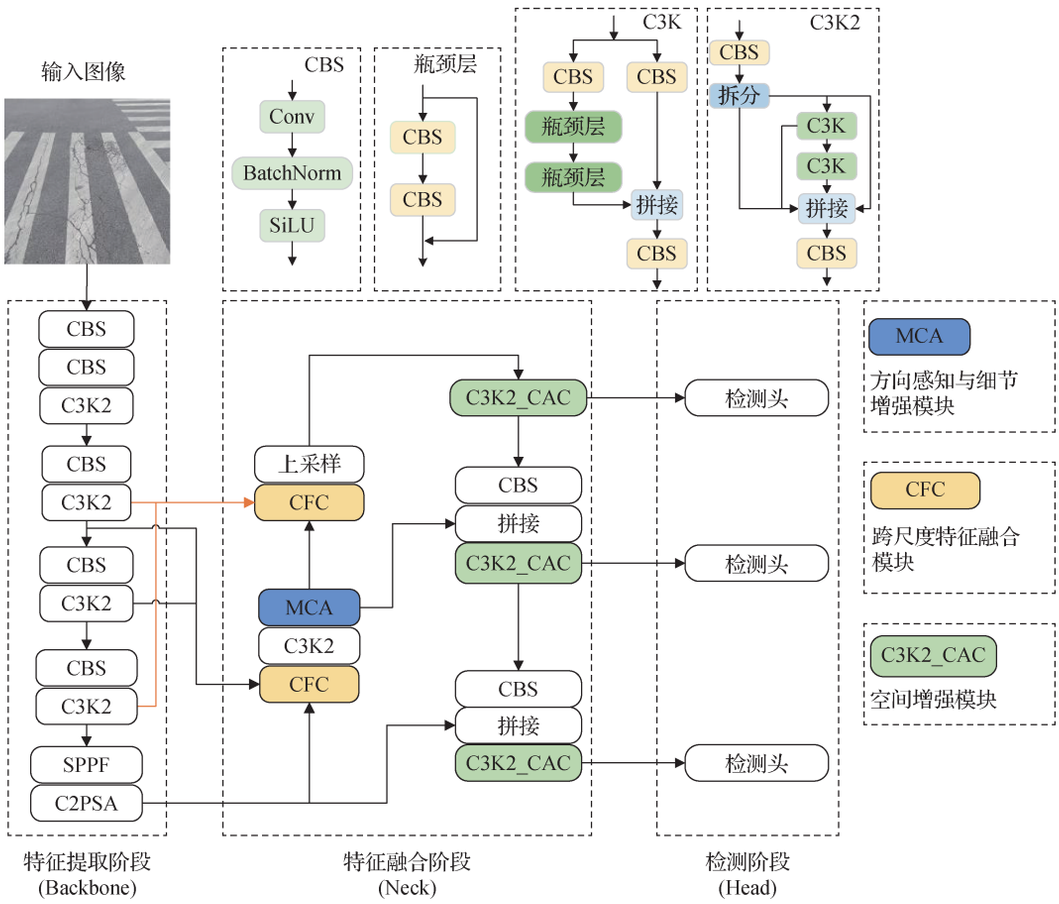


图 1　MSANet 模型总体结构图

Fig. 1　Overall architecture diagram of the MSANet model

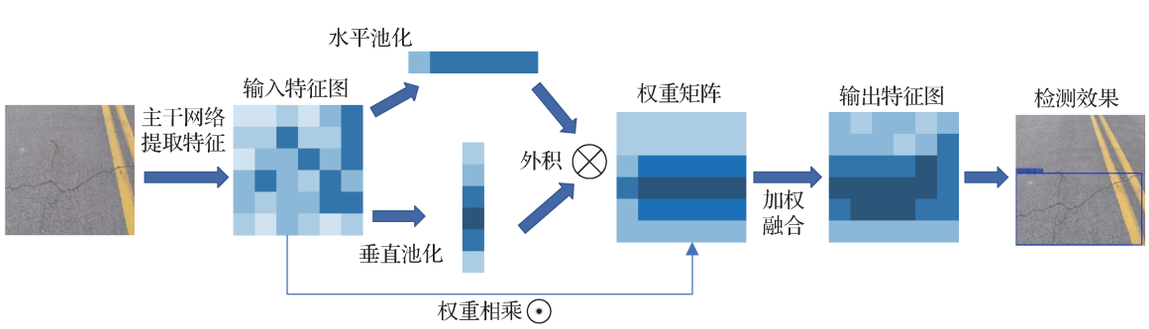


图 2　CA 坐标注意力机制原理

Fig. 2　Principle of the CA coordinate attention mechanism

矩阵 $\boldsymbol{D}$，以引导自注意力权重向邻域聚焦。定义二维曼哈顿距离衰减矩阵 $\boldsymbol{D}$ 为

$$\boldsymbol{D}_{nm} = \gamma^{|x_n - x_m| + |y_n - y_m|} \tag{4}$$

式中，(x_n, y_n) 和 (x_m, y_m) 分别表示位置 n 和 m 的坐标，其中 $\gamma \in (0, 1)$ 为基于注意力头分布设计的衰减系数，用于区分不同注意力头的有效感受野尺度。该权重矩阵体现了明确的空间先验：当两个位置距离较远时，其注意力权重将按照曼哈顿距离加速衰减。对远距离位置赋予较小的指数权重，从而在自注意力计算中削弱长距离依赖。

基于此，利用输入特征 $\boldsymbol{X} \in \mathbf{R}^{C \times H \times W}$，经过独立线性变换得到的查询、键、值的注意力矩阵 $(\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_Q, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_K, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_V)$，将上述空间注意力矩阵与空间衰减矩阵 $\boldsymbol{D}$ 沿水平和垂直两个方向解耦（分别记为 $Attn_H$ 和 $Attn_W$）并作用于原始特征矩阵 $\boldsymbol{V}$，使注意力分别聚焦于行和列方向的局部关系，实现定向的空间特征增强的同时，降低自注意力机制的计

算复杂度。具体为

$$Attn_H = \mathrm{softmax}\left(\boldsymbol{Q}_H\boldsymbol{K}_H^{\mathrm{T}}\right)\odot\boldsymbol{D}^H \tag{5}$$

$$Attn_W = \mathrm{softmax}\left(\boldsymbol{Q}_W\boldsymbol{K}_W^{\mathrm{T}}\right)\odot\boldsymbol{D}^W \tag{6}$$

$$MaSA(\boldsymbol{X}) = Attn_H\left(Attn_W\boldsymbol{V}\right)^{\mathrm{T}} \tag{7}$$

式中，$\boldsymbol{D}^H$为$\gamma^{|y_n-y_m|}$，$\boldsymbol{D}^W$为$\gamma^{|x_n-x_m|}$，$\boldsymbol{Q}_H$，$\boldsymbol{K}_H$分别是$\boldsymbol{Q}$，$\boldsymbol{K}$在垂直方向的分解，$\boldsymbol{Q}_W$，$\boldsymbol{K}_W$分别是$\boldsymbol{Q}$，$\boldsymbol{K}$在水平方向的分解，其中$MaSA(\boldsymbol{X})$的输出为$\boldsymbol{X}_{\mathrm{MaSA}}$。

最后，为了高效融合这两种互补的注意力信息，本文设计了混合注意力机制（MCA），具体结构如图3所示，采用先线性融合再卷积整合的策略，将CA和MaSA分支的输出按系数α加权并行融合，并通过1×1深度可分离卷积（DWConv）进一步整合和精炼融合特征信息，形成最终输出特征$\boldsymbol{F}_{\mathrm{out}}\in\mathbf{R}^{C\times H\times W}$，具体为

$$\boldsymbol{F}_{\mathrm{out}} = DWConv_{1\times1}\left(\alpha\boldsymbol{X}_{\mathrm{CA}} + (1-\alpha)\boldsymbol{X}_{\mathrm{MaSA}}\right) \tag{8}$$

式中，$\boldsymbol{X}_{\mathrm{CA}}$为来自CA注意力分支的输出特征；$\boldsymbol{X}_{\mathrm{MaSA}}$为来自曼哈顿自注意力机制的输出特征。

MCA模块在并行协同框架下实现了全局感知与局部强化的互补建模。模块利用方向编码的注意力捕获长距离的空间依存关系，并结合距离衰减策略突出邻域内的细节特征，从而赋予模型对全局方向和局部空间细节均具有敏锐的感知能力。

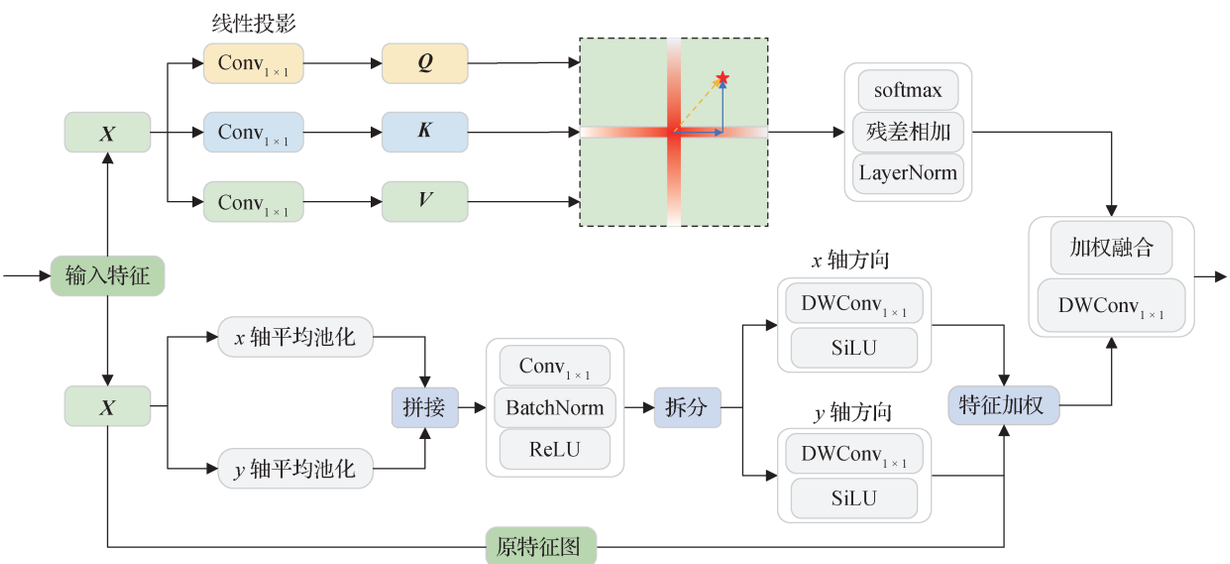


图3　方向感知与细节增强模块（MCA）结构图

Fig. 3　Structure diagram of the directional awareness and detail enhancement module (MCA)

2.3　跨尺度特征融合模块CFC

在YOLO的骨干网络中，连续卷积与下采样虽然能提取高层语义信息，却不可避免地削弱细粒度纹理与边缘特征（道路裂缝等细微结构），从而限制模型对目标的精细表征能力。同时，现有的特征金字塔结构（如FPN（feature pyramid network）/PANet（path aggregation network））虽然通过相邻层级之间的对称连接提升了跨尺度特征的语义一致性，但其特征交互局限于相邻尺度（如主干网络中的B3层，将信息融合至P3，随后传达至N3层），难以实现跨尺度的信息整合。具体结构如图4（a）和图4（b）所示。为缓解上述问题，本文设计了一种非对称跨尺度特征融合模块（CFC），以替代传统的简单拼接操作，具体模块结构如图5所示。CFC模块通过选取来自多个非相邻阶段的特征进行跳跃式融合，突破了仅限于相邻层的连接方式，有效实现语义层次间的深度互补与空间细节的跨尺度增强，如图4（c）所示。具体过程如下：

首先，对浅层特征$\boldsymbol{F}_L$进行全局平均池化和最大池化，并在通道维度上拼接，有效保留细长裂缝的边缘细节，生成既保留统计分布又突出局部激活的下采样表示$\boldsymbol{F}_L'$，具体为

$$\boldsymbol{F}_L' = Concat\left(AvgPool\left(\boldsymbol{F}_L\right), MaxPool\left(\boldsymbol{F}_L\right)\right) \tag{9}$$

随后将深层特征$\boldsymbol{F}_S$经双线性插值上采样至中层空间分辨率$\boldsymbol{F}_S'$，与中层特征共同拼接；最后，通过1×1聚合卷积对三尺度特征进行融合与通道

整理，保证了裂缝区域在特征图上的连贯描绘，提高了裂缝形态的完整性，强化其互补性。具体过程为

$$\boldsymbol{F}_S' = UpSample_{\text{Bilinear}}(\boldsymbol{F}_S) \tag{10}$$

$$\boldsymbol{F}_{\text{out}} = Conv_{1\times1}\left(Concat\left(\boldsymbol{F}_L', \boldsymbol{F}_M, \boldsymbol{F}_S'\right)\right) \tag{11}$$

式中，$\boldsymbol{F}_L \in \mathbf{R}^{C\times H\times W}$具有高空间分辨率和细节纹理，$\boldsymbol{F}_M \in \mathbf{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$保留部分空间信息并包含区域结构，$\boldsymbol{F}_S \in \mathbf{R}^{C\times \frac{H}{4}\times \frac{W}{4}}$则具有丰富的全局语义信息。其中$\boldsymbol{F}_L' \in \mathbf{R}^{2C\times \frac{H}{2}\times \frac{W}{2}}$，$\boldsymbol{F}_S' \in \mathbf{R}^{2C\times \frac{H}{2}\times \frac{W}{2}}$，$\boldsymbol{F}_{\text{out}} \in \mathbf{R}^{6C\times \frac{H}{2}\times \frac{W}{2}}$。

CFC 模块在不显著增加计算量的前提下，通过非对称的跨尺度特征融合机制，使高层语义信息不经中间层逐级传递即可与浅层细节特征交互，有效增强了上下文信息的高效传递与结构互补性。该设计显著提升了模型对跨尺度目标的感知能力与检测鲁棒性。

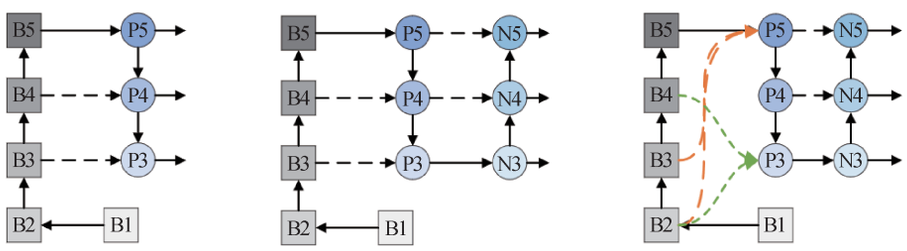


图 4　不同特征融合网络结构对比

Fig. 4　Comparison of different feature fusion network structures ((a) FPN; (b) PANet; (c) CFC)

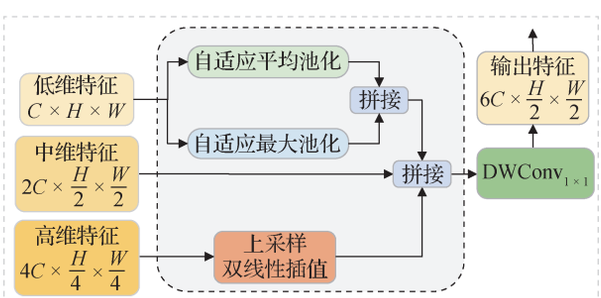


图 5　跨尺度特征融合（CFC）模块结构图

Fig. 5　Structure diagram of the cross-scale feature fusion (CFC) module

2.4　空间增强模块 C3K2_CAC

在 YOLOv11 检测框架中，检测头前的网络尾部采用的 C3K2 模块作为 CSP（cross stage partial）结构的一种轻量残差变体，具有良好的特征融合效果和较大的感受野，用于承接主干网络上采样后的高维聚合特征。然而，该模块本身缺乏全局感知与空间敏感性，难以为检测头提供明确的位置引导。对于检测细长结构化损伤（如裂缝类）检测任务而言，空间细节捕捉和方向信息感知尤为关键。尽管网络的主干和颈部已经集成了空间感知机制，但当特征传递至尾部时，空间信息可能被逐步弱化。因此，本文在 C3K2 模块中引入了坐标注意力卷积模块 CAConv（Zhang 等，2023），构建了 C3K2_CAC 复合模块。具体模块结构如图 6 所示。

具体而言，CAConv 模块首先分别沿特征图的水平和垂直方向进行池化，提取空间方向编码。随后，通过共享卷积将这两个方向的空间编码进行融合，生成联合空间注意力图（即空间敏感权重矩阵），用于对输入特征进行方向与空间维度的联合重标定，在此基础上，模块采用残差连接形式在增强空间感知能力的同时，有效保留原始特征信息，从而提升输出特征的空间表达完整性与定位引导能力。

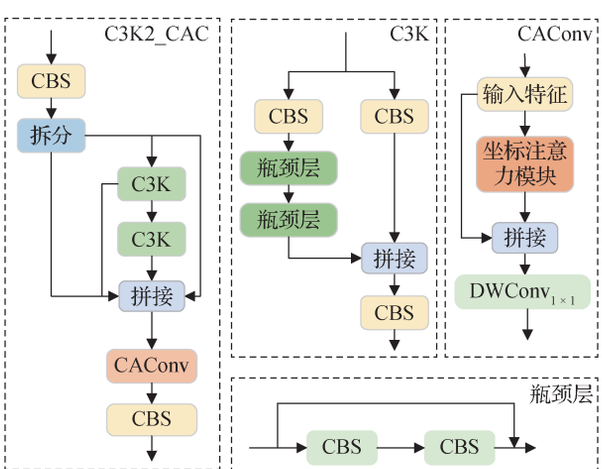


图 6　空间增强模块结构图（C3K2_CAC）

Fig. 6　Spatial enhancement module structure diagram (C3K2_CAC)

首先输入特征 $\boldsymbol{X} \in \mathbf{R}^{C \times H \times W}$ 经过卷积—批归一化—激活的特征提取模块CBS(conv + BN + SiLU)初步处理后,通过1 × 1卷积调整通道,并划分为主支特征 $\boldsymbol{X}_m$ 和旁路特征 $\boldsymbol{X}_b$,随后主支通过两层C3K模块增强特征表达能力,并与旁路分支拼接形成融合特征,具体为

$$\left[\boldsymbol{X}_m, \boldsymbol{X}_b\right] = Conv_{1\times1}\left(C3K\left(\boldsymbol{X}\right)\right) \tag{12}$$

$$\boldsymbol{X}_{\text{cat}} = Concat\left(C3K\left(C3K\left(\boldsymbol{X}_m\right)\right), \boldsymbol{X}_b\right) \tag{13}$$

式中,$C3K$ 表示基于残差连接与Bottleneck模块构建的方向感知特征提取单元。

接着,为进一步挖掘通道与空间交互信息,引入坐标注意力机制CAConv,通过垂直与水平方向池化与共享卷积获得两方向注意力图,按元素加权作用于原始拼接特征,并采用残差连接保留基础结构信息,具体为

$$\boldsymbol{X}_{\text{CA}} = \boldsymbol{X}_{\text{cat}} + \boldsymbol{X}_{\text{cat}} \odot \sigma\left(CA\left(\boldsymbol{X}_{\text{cat}}\right)\right) \tag{14}$$

式中,$\odot$ 表示逐元素乘法,CA 表示坐标注意力编码,σ 为激活函数。最后通过一个1 × 1卷积对融合特征进行通道压缩与结构整合,形成输出,具体为

$$\boldsymbol{X}_{\text{out}} = Conv_{1\times1}\left(\boldsymbol{X}_{\text{CA}}\right) \tag{15}$$

式中,输出特征 $\boldsymbol{X}_{\text{out}} \in \mathbf{R}^{C \times H \times W}$。该结构在保持高效计算的前提下,实现了特征的局部细节强化与空间依赖建模。

融合后的C3K2_CAC模块充分发挥了两种机制的互补优势。在全局特征表达方面,C3K2提供了宽广的感受野和高效的残差信息流,使模型能够抓取整体上下文;同时,CAConv引入的方向性空间注意力为特征赋予了对局部结构和方向模式的选择性增强。两者的结合使模型能够针对道路损伤的结构化特征进行自适应调整,为后续检测层提供了更加具有判别性和鲁棒性的特征表达。

3 实验及分析

3.1 实验设置

为了验证所提出方法的有效性与实验的客观性,本研究搭建了一个实验平台,操作系统为Ubuntu 20.04,使用PyTorch作为深度学习框架。YOLOv11s网络模型作为基线模型,进行对比和消融实验。在所有实验方案中,应用了一致的超参数,具体实验环境配置和超参数设置在表1中详细说明。

表1 实验设置与超参数详情

Table 1 Details of experimental settings and hyperparameters

实验设置	超参数
操作系统	Ubuntu 20.04
深度学习框架	PyTorch2.2.2
编程语言	Python3.10
CPU	Intel i7-7900X CPU
GPU	NVIDIA TITAN V
运行内存	64 GB
学习率	0.01～0.000 1
图像尺寸	640 × 640像素
优化器	SGD
批处理大小	32
随机种子	42
权重衰减	0.000 5
训练轮次	300
其他超参数	默认

3.2 实验数据集

本研究使用开源数据集RDD2022,该数据集由IEEE大数据杯举办的基于群智感知的道路损伤检测挑战赛发布,提供了来自不同国家的路面损伤图像和科学标注。为了验证效果,本研究选取了来自中国区的图像,其中图像由无人机和车载摄像头拍摄完成,涵盖了多种道路缺陷:纵向裂缝(D00)、横向裂缝(D10)、网状裂缝(D20)、坑洞(D40)和修复等。其中,修复是道路工程中特指为修复既有损伤而进行的人工维护作业后留下的痕迹,具有特定的识别特征和工程意义。本文对数据进行清洗和筛选,最终保留5种类型的道路损伤(示例图像如图7所示),其具体比例和数据集划分在表2中展示。数据集标签尺寸及分布如图8所示,从中可以进一步反映出道路损伤目标细长拓扑结构、高长宽比的物理特性。

3.3 评价指标

为了量化实验模型检测结果,实验中使用了精确率(precision,P)、召回率(recall,R)、mAP@0.5、mAP@0.5:0.95、F1-Score、检测速率(帧/s)和浮点运算量(floating point operations,FLOPs)这7类评价指标对模型的性能进行客观评估。

研究中通过多维度指标全面衡量模型在检测精

表 2　数据集类别与划分

Table 2　Categories and classification of datasets

名称	危险等级代码	训练集数量	验证集数量
纵向裂缝	D00	3 686	416
横向裂缝	D10	2 106	251
网状裂缝	D20	851	82
坑洞	D40	289	32
修补区域	/	945	100

注：“/”表示修补区域无危险等级代码，标签中使用 repair 标记。



图 7　数据集不同缺陷图像示例

Fig. 7　Examples of different defect images in the dataset ((a) transverse cracks; (b) longitudinal cracks; (c) repair areas; (d) pits; (e) reticular cracks)

度、定位能力、泛化性能及推理效率等方面的综合表现，确保评估结果的客观性与实用性。

3.4　对比实验

如表 3 所示，本研究提出的 MSANet 模型在道路损伤检测任务中展现出显著优势。在最新目标检测模型对比中，MSANet 以 0.877 的 mAP@0.5 和 0.845 的综合性能（Recall：0.811、Precision：0.883、F1-Score：0.845）超越主流模型（YOLOv8s 的 0.824、RT-DETR 的 0.801），其 mAP@0.5：0.95 指标较基准模型提升 4.8%，比 mAP@0.5 产生了更明显的增幅，反映出模型在高质量预测方面的显著提升。相较于最新的 YOLOv13s 模型，MSANet 在 mAP@0.5，mAP@0.5：0.95 和 F1-Score 上分别取得了 3.4%，5.5% 和 2.1% 的显著优势，验证了多阶段空间感知与细节增强网络结构有效提升了模型的检测精度。

如表 4 所示，在各类道路损伤检测的细粒度任务中，MSANet 进一步体现了场景目标适应性。mAP@0.5 提升了 4.15% 至 0.878，mAP@0.5：0.95 提升了 3.18% 至 0.649，F1-Score 提升了 2.69% 至 0.840；裂缝类检测（包括纵向裂缝、横向裂缝和网状裂缝）：mAP@0.5 提升了 2.70%，F1-Score 提升了 2.22%；坑洞检测（D40）：mAP@0.5 提升了 7.13%，F1-Score 提升了 7.78%。同时，模型保持了 142 帧/s 高推理速度，相较于基线仅损失 6% 的推理速度，显著优于实时检测器 RT-DETR，充分满足实际工程对效率与精度的双重要求。

检测热力图直观表明了改进后模型的优势，如图 9 所示。本文模型展现出更高的目标区域聚焦

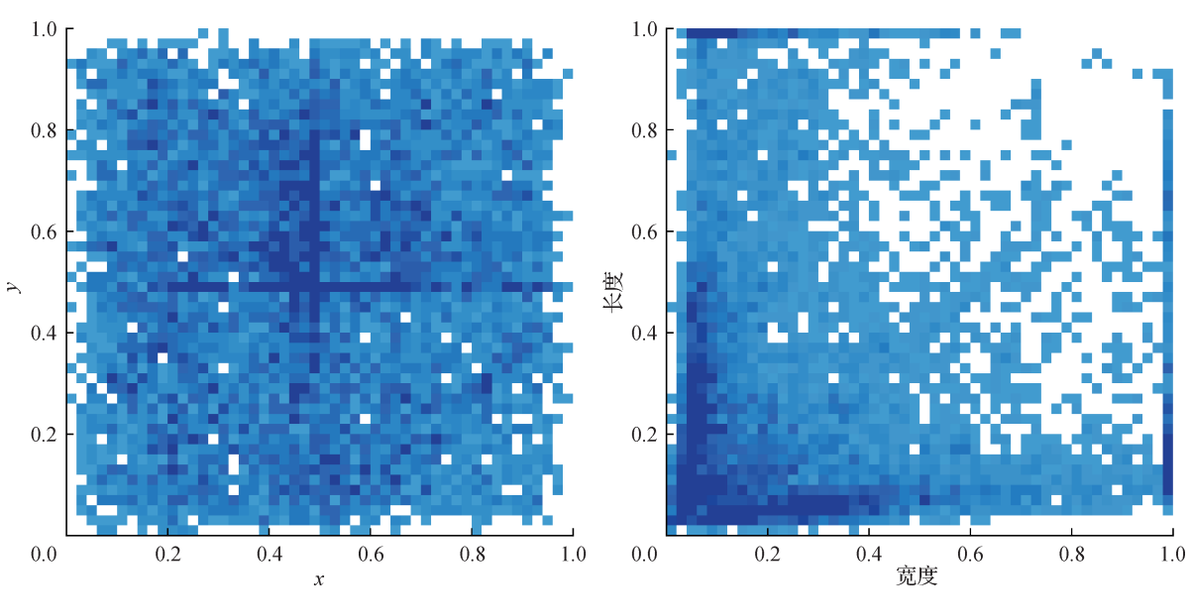


图 8　数据集中目标标签位置分布与长宽比分布

Fig. 8　Distribution of target label positions and aspect ratios in the dataset

表3 不同模型对比实验结果

Table 3 Comparative experimental results of different models

方法	P	R	mAP@0.5	mAP@0.5:0.95	F1-Score	FLOPs/G	检测速度(帧/s)
Faster R-CNN(Ren等,2017)	0.701	0.783	0.732	0.453	0.740	940.7	11
YOLOv5s(Khanam和Hussain,2024a)	0.867	0.800	0.860	0.603	0.832	24.1	307
YOLOv6s(Li等,2022)	0.862	0.760	0.823	0.561	0.808	44.2	278
YOLOv8s(Swathi和Challa,2024)	0.835	0.813	0.861	0.599	0.824	28.4	273
YOLOv8-Ghost(Han等,2020)	0.848	0.803	0.855	0.600	0.825	**16.3**	**341**
YOLOv8-RTDETR	0.823	0.743	0.790	0.489	0.781	32.4	77
YOLOv9s(Wang等,2024b)	0.841	0.770	0.843	0.578	0.804	27.4	188
YOLOv10s(Wang等,2024a)	0.872	0.752	0.845	0.602	0.807	24.8	158
YOLOv11s(Khanam和Hussain,2024b)	0.870	0.795	0.861	0.601	0.830	21.7	151
YOLOv11s-RTDETR	0.742	0.747	0.770	0.488	0.744	30.4	134
YOLOv12s(Tian等,2025)	0.846	0.823	0.857	0.603	0.834	21.2	122
YOLOv13s(Lei等,2025)	0.875	0.787	0.848	0.597	0.828	21.3	113
RT-DETR(Zhao等,2024)	0.830	0.773	0.816	0.531	0.801	108.0	45
D-FINE(Peng等,2025)	0.834	0.782	0.844	0.599	0.807	72.6	71
MSANet(本文)	0.883	0.811	**0.877**	**0.630**	**0.845**	32.0	142
相对提升率/%			**+1.9**	**+4.8**	**+1.8**		

注:加粗字体表示该指标下获得的最佳数据和相较于基线YOLOv11s模型的相对提升百分比。

表4 细粒度任务对比实验结果

Table 4 Comparative experimental results on fine-grained tasks

模型	mAP@0.5				mAP@0.5:0.95				F1-Score			
	全部类别	裂缝类	坑洞	修补	全部类别	裂缝类	坑洞	修补	全部类别	裂缝类	坑洞	修补
基准模型	0.843	0.853	0.827	0.850	0.629	0.578	0.521	0.595	0.818	0.811	0.810	0.845
MSANet(本文)	0.878	0.876	0.886	0.873	0.649	0.601	0.547	0.610	0.840	0.829	0.873	0.862
相对提升率/%	**+4.15**	**+2.70**	**+7.13**	**+2.71**	**+3.18**	**+3.98**	**+4.99**	**+2.52**	**+2.69**	**+2.22**	**+7.78**	**+2.01**

注:加粗字体表示本文模型较于基线YOLOv11s的相对提升百分比。

度,背景噪声与无关区域的伪激活被有效抑制,对目标的覆盖更为完整。实验结果表明,MSANet通过特征融合机制与注意力协同优化,在道路损伤检测中实现了精度与速度的平衡。

以上结果共同验证了MSANet模型特有的方向感知和细节边缘增强对道路损伤检测瓶颈的突破性。

3.5 消融实验

为验证所提模块的有效性,实验阶段分别对混合注意力(MCA)、跨尺度特征融合(CFC)及C3K2_CAC模块进行了系统性消融实验。实验结果如表5所示。MCA模块在未显著增加计算成本的前提下,有效提升了模型性能。虽然在mAP@0.5指标上有所下降,但其在更严格衡量检测精度的mAP@0.5:0.95指标上带来了2.2%的显著提升,表明MCA模块更侧重于增强模型在高交并比(intersection-over-union,IoU)阈值条件下的检测能力。CFC模块通过融合主干网络不同阶段的低层细节信息与高层语义特征,为网络颈部提供了跨尺度上下文信息,其引入使模型在mAP@0.5和mAP@0.5:0.95上分别取得了0.5%和2.3%的稳定提升。

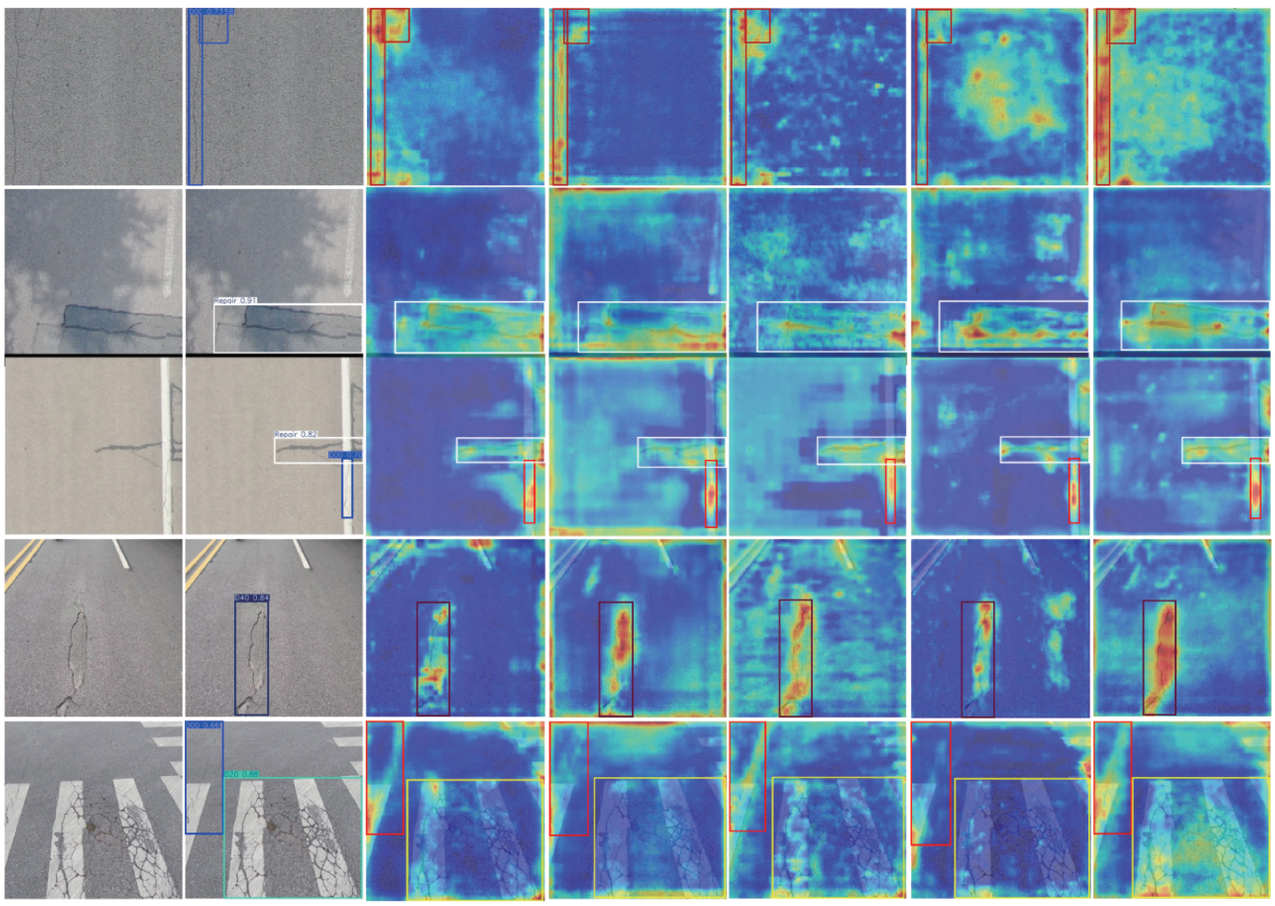


图 9　不同模型热力图比较

Fig. 9　Heatmap comparison of different models

表 5　模块消融实验结果

Table 5　Ablation study results of different modules

MCA	CFC	C3K2_CAC	mAP@0.5	mAP@0.5:0.95	F1-Score	FLOPs/G
–	–	–	0.861	0.601	0.830	**21.3**
√	–	–	0.858	0.614	0.831	22.6
–	√	–	0.865	0.615	0.835	22.6
–	–	√	0.867	0.608	0.832	29.4
√	√	–	0.873	0.617	0.837	30.4
√	–	√	0.857	0.601	0.829	28.3
–	√	√	0.875	0.614	0.841	34.8
√	√	√	**0.877**	**0.630**	**0.845**	31.4

注：加粗字体表示各列最优结果。“√”表示使用了该模块，“–”表示未使用。

C3K2_CAC 模块引入一定计算开销为代价，进一步增强了 Neck 输出特征的空间感知能力，为检测头提供了更具判别性的高级特征。将 CFC 与 C3K2_CAC 模块结合使用，产生了显著的协同效应，模型综合性能得到进一步跃升。F1-Score 和 mAP@0.5 分别大幅提高了 1.3% 和 1.6%。集成上述模块后，改进后的模型在核心检测指标上实现了全面提升：mAP@0.5:0.95 提升了 4.8%，F1-Score 提升了 1.8%，并保

持142 帧/s的推理速度。

3.6 混合注意力机制(MCA)消融实验

为探索曼哈顿自注意力(MaSA)与坐标注意力(CA)的有效融合方式,本研究系统比较了门控卷积融合(MaSA connect CA by GBC, MGC)、串行连接(MaSA to CA, MTC; CA to MaSA, CTM)及权重融合(MaSA connect CA by weight, MWC) 3种策略,具体网络结构如图10所示。其中,门控卷积融合是使用门控计算单元对两个分支进行加权组合;串行连接是将MaSA模块作为CA模块的输入或将CA模块作为MaSA模块的输入;权重连接是设置固定权重对两个分支进行线性加权融合。

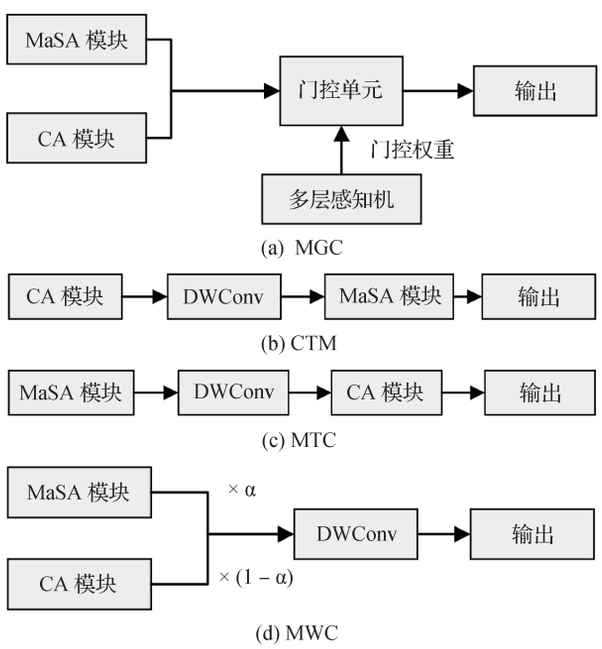


图10 不同融合策略网络结构图

Fig. 10 Network structure diagrams of different fusion strategies ((a) MGC; (b) CTM; (c) MTC; (d) MWC)

实验分析表明:MaSA模块主要聚焦于增强特征图中的局部细节与纹理特征表征能力;CA模块则有效捕获目标全局结构信息与方向性上下文特征,两者不同程度上提升了模型的检测性能。在对比的融合方法中,加权融合策略展现出最佳性能。实验结果如表6所示。其中,CTM网络结构取得最差提升,分析其原因可能是局部注意力机制破坏了全局感知信息。通过细致调整融合权重,当MWC权重α = 0.6时,模型性能达到最优。相较于基准模型,该融合方案在mAP@0.5和mAP@0.5:0.95指标上分别显著提升了1.9%和4.2%,表明MaSA的细节信息与CA的全局感知信息实现了最佳融合。这种协同作用显著提升了模型的综合表征能力,驱动了检测性能的实质性进步。

3.7 跨尺度特征融合(CFC)模块消融实验

为确定跨尺度特征融合(CFC)模块在改进网络框架中的最优部署策略,系统性地对其部署位置及数量进行了消融实验,实验结果如表7所示。实验结果表明,CFC模块的部署位置对模型性能(精度指标)和参数量均产生显著影响。在特征提取阶段,浅层特征(B_3)分辨率高、感受野较小,可捕获丰富的空间细节;中层特征(B_4)兼具一定空间分辨率与区域性结构信息;深层特征(B_5)则依赖大感受野编码全局语义,但其对微小纹理的响应较弱。通过综合权衡计算开销与检测精度,最终确定了CFC模块的最优部署方案为12层与15层。该方案有效优化了特征信息流的传递效率,并促进了跨层级语义信息的有效恢复。

3.8 模型跨数据集与泛化性实验

为了验证模型的泛化性能与未来部署的实用性,在多个公开道路损伤数据集进行跨数据集实验,并在实地拍摄图像上进行泛化性实验。

RDD2022全球数据集是由IEEE大数据杯发布的多源数据集,其中包含8类缺陷,涵盖多种拍摄形式。本文对RDD2022全球数据集进行清洗与筛选,保留了含有横向裂缝、纵向裂缝、网状裂缝和坑洞4种缺陷类别的约30 000幅图像,并在本文的网络中进行训练和推理。本文模型在此数据集中与其他最佳通用目标检测网络进行对比实验。实验结果如表8所示。

沥青路面典型病害样本数据集(China road damage detection, CNRDD)选取中国境内G303路段采集道路损坏数据。采集到的单幅图像中道路损伤密度较高,在该数据集上进行道路损伤检测更具挑战性。本文对数据进行筛选和清洗,保留了裂缝、纵向裂缝、横向裂缝、坑槽和修补,并在本文的网络中进行训练和推理。实验结果如表9所示。

UAV-PDD2023 (pavement distress detection based on uav images)数据集是由无人机在30 m高度拍摄,包含横向裂缝、纵向裂缝、网状裂缝、斜裂缝、修复区域和坑洼,该数据集由尺寸为2 592 × 1 944像素的图像组成。本文在此数据集中与其他通用目标检测网络进行实验。实验结果如表10所示。

表 6　MCA 模块消融实验

Table 6　Ablation study of the MCA module

模块	权重	mAP@0.5	mAP@0.5:0.95	F1-Score	FLOPs/G
CA	—	0.871	0.611	0.835	23.2
MaSA	—	0.869	0.622	0.840	22.4
MGC	—	0.866	0.605	0.838	**22.2**
MTC	—	0.867	0.623	0.828	22.6
CTM	—	0.858	0.614	0.830	22.6
CWM	0.4	0.866	0.616	0.839	22.6
	0.5	0.869	0.616	0.837	22.6
	0.55	0.867	0.613	0.835	22.6
	0.6	**0.877**	**0.626**	**0.853**	22.6
	0.7	0.859	0.610	0.825	22.6

注：加粗字体表示各列最优结果。“-”表示无权重数据。

表 7　CFC 模块消融实验结果

Table 7　Ablation study results of the CFC module

12层	15层	18层	mAP@0.5	mAP@0.5:0.95	F1-Score	FLOPs/G
√			0.867	0.604	0.833	24.9
	√		0.865	0.611	0.820	25.5
		√	0.851	0.598	0.836	23.8
√	√		0.865	**0.615**	0.825	**22.4**
	√	√	0.862	0.606	0.838	28.0
√		√	**0.879**	0.612	**0.841**	27.1
√	√	√	0.859	0.611	0.829	32.1

注：加粗字体表示各列最优结果。“√”表示使用了该模块。

表 8　RDD2022 全球数据集实验结果

Table 8　Experimental results of the RDD2022 global dataset

模型	mAP@0.5	mAP@0.5:0.95	F1-Score
YOLOv5s	0.799	0.571	0.772
YOLOv8s	0.801	0.572	0.791
YOLOv11s	0.812	0.582	0.806
MSANet(本文)	0.835	0.601	0.819
相对提升率/%	**+2.83**	**+3.26**	**+1.61**

注：加粗字体表示本文模型较于基线 YOLOv11s 的相对提升百分比。

表 9　CNRDD 数据集实验结果

Table 9　Experimental results of the CNRDD dataset

模型	mAP@0.5	mAP@0.5:0.95	F1-Score
YOLOv5s	0.280	0.124	0.536
YOLOv8s	0.281	0.126	0.325
YOLOv11s	0.287	0.125	0.350
MSANet(本文)	0.298	0.133	0.363
相对提升率/%	**+3.83**	**+6.40**	**+3.71**

注：加粗字体表示本文模型较于基线 YOLOv11s 的相对提升百分比。

UAPD（UAV-based asphalt pavement distress）(Zhu 等，2022)数据集由无人机采集，包含 6 类道路病害：横向裂缝、纵向裂缝、网状裂缝、斜裂缝、修复区域和坑洼，由 3 100 幅 512 × 512 像素的图像组成。本文在此数据集中与其他通用目标检测网络进行实

表10　UAV-PDD2023数据集实验结果

Table 10　Experimental results of the UAV-PDD2023 dataset

模型	mAP@0.5	mAP@0.5:0.95	F1-Score
YOLOv5s	0.918	0.657	0.886
YOLOv8s	0.924	0.671	0.899
YOLOv11s	0.940	0.746	0.920
MSANet(本文)	0.965	0.807	0.953
相对提升率/%	**+2.66**	**+8.18**	**+3.59**

注：加粗字体表示本文模型较于基线YOLOv11s的相对提升百分比。

表11　UAPD数据集实验结果

Table 11　Experimental results of the UAPD dataset

模型	mAP@0.5	mAP@0.5:0.95	F1-Score
YOLOv5s	0.247	0.179	0.290
YOLOv8s	0.251	0.182	0.268
YOLOv11s	0.258	0.176	0.260
MSANet(本文)	0.265	0.194	0.267
相对提升率/%	**+2.71**	**+10.23**	**+2.69**

注：加粗字体表示本文模型较于基线YOLOv11s的相对提升百分比。

验。实验结果如表11所示。尽管数据集之间存在显著的数据分布差异，本文方法在跨数据集实验中仍表现出稳定且具有竞争力的性能优势，并且mAP@0.5:0.95指标较mAP@0.5获得了更高的提升。

此外，本文模型在裂缝分割任务中依然表现显著优势，在Crack-seg裂缝分割数据集中，Box mAP@0.5:0.95提升2.16%，Mask mAP@0.5:0.95提升1.51%，详见表12，进一步说明了本文提出的模型在对裂缝等道路损伤目标的特征提取上具有显著优势。

表12　在Crack-seg数据集上泛化实验

Table 12　Generalization experiments on the Crack-seg dataset

模型	Box			Mask		
	mAP@0.5	mAP@0.5:0.95	F1-Score	mAP@0.5	mAP@0.5:0.95	F1-Score
YOLOv11s	0.832	0.649	0.805	0.650	0.199	0.717
本文方法	0.833	0.663	0.809	0.659	0.202	0.720
相对提升率/%	**+0.12**	**+2.16**	**+0.50**	**+1.38**	**+1.51**	**+0.42**

注：加粗字体表示本文模型较于基线YOLOv11s的相对提升百分比。

同时，为了验证模型在工程实践场景下的检测效果，对MSANet模型进行路面损伤检测实验。使用大恒工业相机MER-132-43U3C，采集图像大小为1 292×964像素。在2 m高度进行俯拍和1.5 m高度与地面呈45°—60°拍摄，以模拟车辆采集和无人机采集效果，采集了各类样本以评估模型的性能。同时，MSANet模型在实际道路检测中表现出良好的泛化能力和检测鲁棒性。本文使用在多源道路损伤数据集上预训练模型权重，对不同光照、纹理复杂背景下自收集数据集进行推理检测，仍有效检出目标，且本文模型在对目标的定位与置信度得分中表现出显著优势。但对于低对比度目标，依然存在漏检与误检的实例。具体检测结果如图11所示。

4　结　论

高效精准的道路损伤检测技术对智慧交通基础设施运维具有重要意义。本文提出MSANet——基于YOLOv11的多级优化道路损伤检测框架，通过协同空间感知强化与细节特征增强机制突破现有瓶颈。在所提出的方法中，MCA模块实现方向感知与局部细节增强的双重优化；跨尺度特征融合模块CFC构建非对称特征互补架构，解决细长目标表征损失问题；C3K2_CAC实现末端空间信息强化与融合。以上模块在不同阶段共同构建目标空间感知与细节强化协同的MSANet网络架构，在RDD2022_China数据集上取得87.7%的mAP@0.5的检测精度，mAP@0.5:0.95相较于基线提升了4.8%，同时

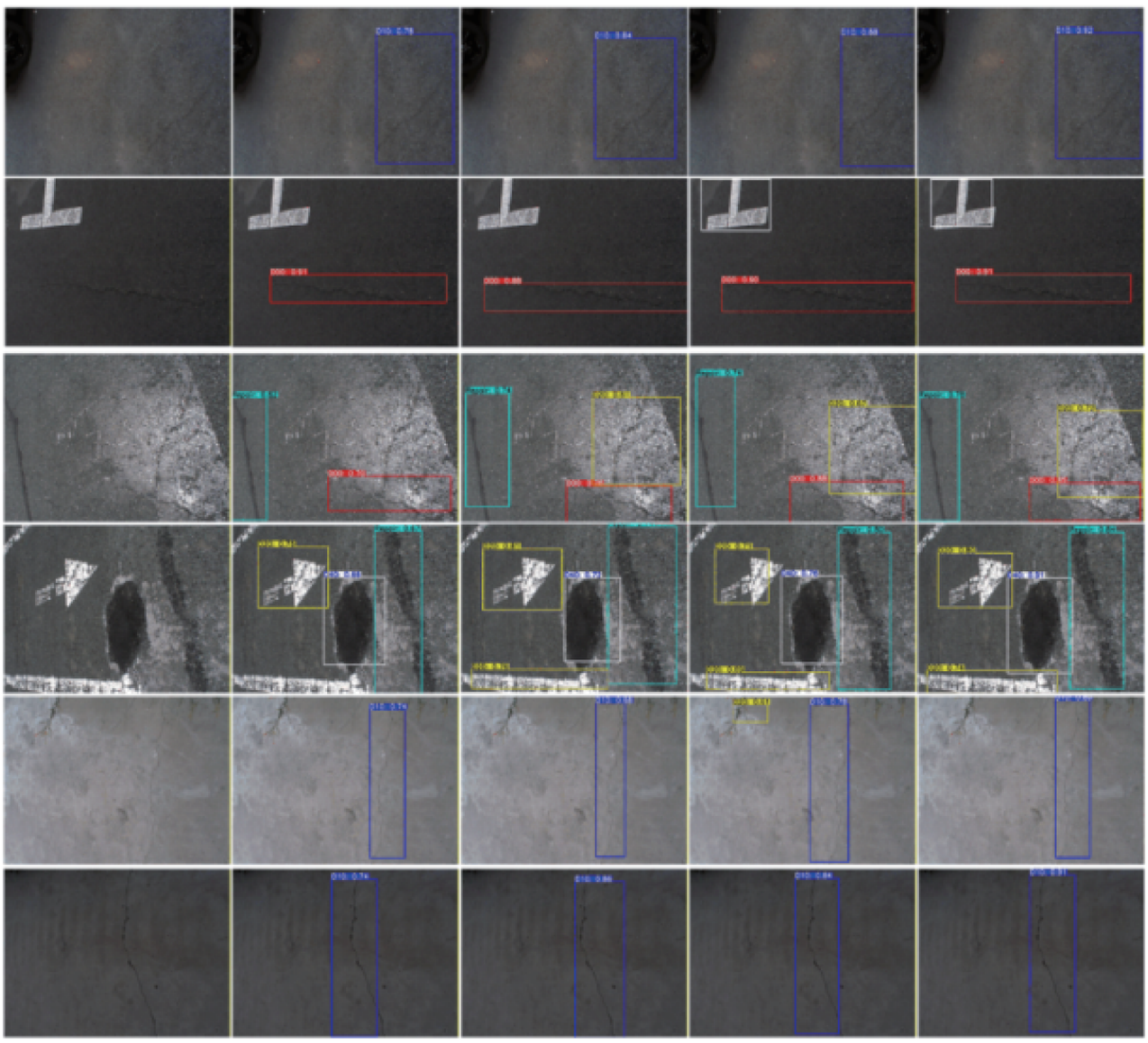


图 11　自收集数据检测结果

Fig. 11　Detection results on the self-collected dataset

保持 142 帧/s 实时性能，并在细粒度实验中取得优秀成绩。未来，在无人机低空巡检与自动驾驶车辆采集场景，该技术将为道路设施智能运维提供高效解决方案。后续研究将探索动态感受野机制与轻量化压缩，进一步提升工程实践中的鲁棒性和实时性能。

代码可用性声明

本论文代码已在中国科学院科学数据银行 ScienceDB 平台公开发布，访问地址为 https://www.scidb.cn/s/RZBnIz。

参考文献（References）

Arya D, Maeda H, Ghosh S K, Toshniwal D, Omata H, Kashiyama T, et al. 2020. Global road damage detection: state-of-the-art solutions//Proceedings of 2020 IEEE International Conference on Big Data (Big Data). Atlanta, USA: IEEE: 5533-5539 [DOI: 10.1109/BigData50022.2020.9377790].

Arya D, Maeda H, Ghosh S K, Toshniwal D, Omata H, Kashiyama T, et al. 2022. Crowdsensing-based road damage detection challenge (CRDDC'2022)//Proceedings of 2022 IEEE International Conference on Big Data (Big Data). Osaka, Japan: IEEE: 6378-6386 [DOI: 10.1109/BigData55660.2022.10021040]

Chu H H, Wang W and Deng L. 2022. Tiny-Crack-Net: a multiscale feature fusion network with attention mechanisms for segmentation of tiny cracks. Computer-Aided Civil and Infrastructure Engineering, 37(14): 1914-1931 [DOI: 10.1111/mice.12881]

de Abreu V H S, Santos A S and Monteiro T G M. 2022. Climate change impacts on the road transport infrastructure: a systematic review on Adaptation Measures. Sustainability, 14(14): #8864 [DOI: 10.3390/su14148864]

Fan Q H, Huang H B, Chen M R, Liu H M and He R. 2024. RMT: retentive networks meet vision transformers//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 5641-5651 [DOI: 10.1109/CVPR52733.2024.00539]

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 (CVPR). Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]

Han K, Wang Y H, Tian Q, Guo J Y, Xu C J and Xu C. 2020. GhostNet: more features from cheap operations//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 1577-1586 [DOI: 10.1109/CVPR42600.2020.00165]

Hou Q R, Zhou D Q and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

Jiang Y T, Yan H T, Zhang Y R, Wu K Q, Liu R Y and Lin C Y. 2023. RDD-YOLOv5: road defect detection algorithm with self-attention based on unmanned aerial vehicle inspection. Sensors, 23(19): #8241 [DOI: 10.3390/s23198241]

Khanam R and Hussain M. 2024a. What is YOLOv5: a deep look into the internal features of the popular object detector [EB/OL]. [2025-09-22]. https://arxiv.org/pdf/2407.20892.pdf

Khanam R and Hussain M. 2024b. YOLOv11: an overview of the key architectural enhancements [EB/OL]. [2025-09-22]. https://arxiv.org/pdf/2410.17725.pdf

Lau K W, Po L M and Rehman Y A U. 2024. Large separable kernel attention: rethinking the large kernel attention design in CNN. Expert Systems with Applications, 236: #121352 [DOI: 10.1016/j.eswa.2023.121352]

Lei M Q, Li S Q, Wu Y H, Hu H, Zhou Y, Zheng X H, et al. 2025. YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception [EB/OL]. [2025-09-22]. https://arxiv.org/pdf/2506.17733.pdf

Li C Y, Li L L, Jiang H L, Weng K H, Geng Y F, Li L, et al. 2022. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. [2025-09-22]. https://arxiv.org/pdf/2209.02976.pdf

Li T W and Li G Q. 2024. Road defect identification and location method based on an improved ML-YOLO algorithm. Sensors, 24(21): #6783 [DOI: 10.3390/s24216783]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]

Ma D, Fang H Y, Wang N N, Zhang C, Dong J X and Hu H B. 2022. Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF. IEEE Transactions on Intelligent Transportation Systems, 23(11): 22166-22178 [DOI: 10.1109/TITS.2022.3161960]

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]

Peng Y S, Li H B, Wu P X, Zhang Y Y, Sun X Y and Wu F. 2025. D-FINE: redefine regression task of DETRs as fine-grained distribution refinement//Proceedings of the 13th International Conference on Learning Representations. Singapore, Singapore: OpenReview.net

Ragnoli A, De Blasiis M R and Di Benedetto A. 2018. Pavement distress detection methods: a review. Infrastructures, 3(4): #58 [DOI: 10.3390/infrastructures3040058]

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]

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]

Swathi Y and Challa M. 2024. YOLOv8: advancements and innovations in object detection//Senjyu T, So-In C and Joshi A, eds. Smart Trends in Computing and Communications. Singapore, Singapore: Springer: 1-13 [DOI: 10.1007/978-981-97-1323-3_1]

Tian Y J, Ye Q X and Doermann D. 2025. YOLOv12: attention-centric real-time object detectors//Proceedings of the 39th Annual Conference on Neural Information Processing Systems. San Diego, USA: OpenReview.net

Wang A, Chen H, Liu L H, Chen K, Lin Z J, Han J G, et al. 2024a. YOLOv10: real-time end-to-end object detection//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #3429

［DOI：10.52202/079017-3429］

Wang C Y，Yeh I H and Liao H Y M. 2024b. YOLOv9：learning what you want to learn using programmable gradient information//Proceedings of the 18th European Conference on Computer Vision（ECCV）. Milan，Italy：Springer：1-21［DOI：10.1007/978-3-031-72751-1_1］

Woo S，Park J，Lee J Y and Kweon I S. 2018. CBAM：convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich，Germany：Springer：3-19［DOI：10.1007/978-3-030-01234-2_1］

Xu Z S，Lei X D and Guan H Y. 2023. Multi-scale local feature enhanced transformer network for pavement crack detection. Journal of Image and Graphics，28(4)：1019-1028（许正森，雷相达，管海燕. 2023. 多尺度局部特征增强 Transformer 道路裂缝检测模型. 中国图象图形学报，28(4)：1019-1028）［DOI：10.11834/jig.211129］

Zhang L L，Liao Y，Wang G X，Chen J and Wang H B. 2022. A multi-scale contextual information enhancement network for crack segmentation. Applied Sciences，12(21)：#11135［DOI：10.3390/app122111135］

Zhang X，Liu C，Song T T，Yang D G，Ye Y C，Li K，et al. 2023. RFAConv：innovating spatial attention and standard convolutional operation［EB/OL］.［2025-09-22］. https://arxiv.org/pdf/2304.03198v6.pdf

Zhang Z H，Wen Y N，Mu H W and Du X P. 2022. Dual attention mechanism based pavement crack detection. Journal of Image and Graphics，27(7)：2240-2250（张志华，温亚楠，慕号伟，杜小平. 2022. 结合双注意力机制的道路裂缝检测. 中国图象图形学报，27(7)：2240-2250）［DOI：10.11834/jig.200758］

Zhao Y A，Lyu W Y，Xu S L，Wei J M，Wang G Z，Dang Q Q，et al. 2024. DETRs beat YOLOs on real-time object detection//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Seattle，USA：IEEE：16965-16974［DOI：10.1109/CVPR52733.2024.01605］

Zhu J Q，Zhong J T，Ma T，Huang X M，Zhang W G and Zhou Y. 2022. Pavement distress detection using convolutional neural networks with images captured via UAV. Automation in Construction，133：#103991［DOI：10.1016/j.autcon.2021.103991］

作者简介

汪裕盛，男，硕士研究生，主要研究方向为机器视觉和目标检测。E-mail：2023101074@mails.cust.edu.cn

方明，通信作者，男，教授，主要研究方向为机器视觉和图像处理。E-mail：fangming@cust.edu.cn

冯俊龙，男，讲师，主要研究方向为计算机视觉。

E-mail：fengjunlong@cust.edu.cn