中图法分类号： **文献标识码：** A **文章编号：** 1006-8961(XXXX)XX-0001-17

论文引用格式： Wang Yusheng, Feng Junlong, Fang Ming. Multi-Stage spatial-aware network for road damage detection[J/OL]. Journal of Image and Graphics, XXXX:1-17. DOI: 10.11834/jig.250422. (汪裕盛, 冯俊龙, 方明. 多阶段空间感知道路损伤检测网络[J/OL]. 中国图象图形学报, XXXX:1-17. DOI: 10.11834/jig.250422.)[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,吉林省长春市 130022; 2. 长春理工大学人工智能学院,吉林省长春市 130022; 3. 长春理工大学中山研究院,广东省中山市 528437

摘 要：目的 道路损伤检测作为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,本文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142FPS实时推理速度的同时,mAP@0.5、mAP@0.5:0.95和F1-Score较现有最优方法分别提高了1.9%、4.9%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。论文代码将开源于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;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 City, Jilin Province 130022, China; 2. School of Artificial Intelligence, Changchun University of Science and Technology, Changchun City, Jilin Province 130022, China; 3. Zhongshan Institute of Changchun University of Science and Technology, Zhongshan City, Guangdong Province, 528437, China

Abstract: Objective Road damage detection constitutes a fundamental component of intelligent maintenance and management of transportation infrastructure. With the continuous modernization of global transportation systems and the rapid expansion of road networks, pavements inevitably undergo deterioration due to a variety of external and internal factors. Traditional inspection methods, such as manual surveys and existing automated approaches, still face considerable challenges in terms of scalability, efficiency, and accuracy, thereby limiting their applicability to large-scale deployments. Consequently, there is an urgent need to develop a high-performance, accurate, and robust road damage detection and maintenance framework. In practical detection scenarios, road surface damages are often characterized by elongated geometrical structures with high aspect ratios, low texture, low contrast, and significant heterogeneity. These attributes pose

收稿日期: 2025-09-04;**修回日期:** 2026-03-06

基金项目: 中山市科技局引进科研创新团队项目(项目编号:CXTD2023005)

Supported by: Zhongshan Science and Technology Bureau introduces research and innovation team project

fundamental obstacles for existing algorithms, which largely focus on local texture or semantic information extraction while overlooking the intrinsic physical properties of road damages. Specifically, such models fail to capture the directional feature coupling inherent in the continuous spatial distribution of cracks and the attenuation of edge information in blurred potholes. The lack of direction-aware features and ambiguous positional information weakens the ability to model long-range dependencies and reduces responsiveness to blurred-edge targets. Furthermore, the hierarchical convolution and downsampling operations inherent in convolutional neural networks (CNNs) cause progressive degradation of cross-scale features, leading to the loss of critical detail information in high-resolution shallow feature maps. **Method** To address these limitations, this study proposes a novel Multi-stage Spatial-Aware detection network (MSANet), which is designed on top of the YOLOv11 framework. The methodological design of MSANet follows a progressive principle of feature enhancement-fusion-optimization, tailored specifically to the physical characteristics of elongated cracks and blurred-edge potholes in road surfaces. In the direction-sensitive feature enhancement stage, a hybrid Manhattan self-attention and Coordinate Attention (MCA) mechanism is introduced. This mechanism incorporates Manhattan self-attention to model the spatial continuity of crack axes, while simultaneously leveraging coordinate attention to enhance the localization accuracy of blurred-edge defects. Together, these complementary components establish a dual-driven feature enhancement paradigm, reinforcing both directionality and detail representation, and enabling the model to capture long-range spatial dependencies with improved precision. In order to mitigate feature degradation caused by multiple downsampling operations, a cross-scale feature concat (CFC) module is integrated into the network's neck. By cascading feature maps across different resolutions, the CFC module effectively balances semantic completeness at larger scales with detail richness at smaller scales, thereby compensating for the loss of fine-grained information. Finally, within the detection head, a coordinate-aware convolution kernel (C3K2_CAC) is embedded in the C3K2 block. This design jointly optimizes spatial position and channel response through direction-aware spatial attention and dynamic channel reweighting, resulting in anisotropic representation enhancement in high-dimensional feature spaces. While maintaining computational efficiency, this module significantly improves edge clarity, detection robustness, and global feature discrimination capability. During training, a standardized experimental platform was established using Ubuntu 20.04 with PyTorch as the deep learning framework. The YOLOv11s network served as the baseline model for comparative and ablation studies. Across all experiments, consistent hyperparameters were applied: a batch size of 32, an initial learning rate decayed from 0.01 to 0.0001, image resolution fixed at 640×640, stochastic gradient descent (SGD) as the optimizer, and a training schedule of 300 epochs. Other parameters were maintained at their default network configurations to ensure fairness and reproducibility. **Result** Experiments were conducted on the publicly available RDD2022 Chinese subset, which includes multiple road damage categories: longitudinal cracks (D10), transverse cracks (D10), alligator cracks (D20), potholes (D40), and repaired regions. Model performance was objectively 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. Results demonstrate that the proposed MSANet model 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 is improved by 4.9% over the baseline, confirming the robustness of the design. In fine-grained damage detection tasks, MSANet further demonstrates adaptability, achieving a 4.15% improvement in mAP@0.5 (to 0.878), a 3.18% increase in mAP@0.5:0.95 (to 0.649), and a 5.36% gain in F1-score (to 0.629). Importantly, the model sustains a high inference speed of 142 FPS, with only a 6% reduction compared to the baseline, significantly outperforming real-time detectors such as RT-DETR. This balance of accuracy and efficiency ensures practical applicability in real-world engineering contexts, where both rapid decision-making and reliable predictions are essential. Ablation studies further validate the contributions of the proposed modules. The MCA module enhances model performance without incurring significant computational overhead. The CFC module effectively fuses low-level detailed features and high-level semantic features across network stages, introducing cross-scale contextual information that yields stable improvements of 0.5% in mAP@0.5 and 2.3% in mAP@0.5:0.95. The C3K2_CAC module, while introducing a modest increase in computational cost, enhances the spatial awareness of features output from the neck, thereby supplying the detection head with more discriminative representations. Visualizations of detection heatmaps further confirm

that MSANet provides improved target region focus, stronger suppression of background noise and irrelevant activations, and more complete coverage of damage regions. To assess generalization and deployment potential, experiments were also conducted on additional public datasets and field-captured road images, where MSANet achieved consistently promising results, demonstrating strong robustness and scalability across diverse environments. **Conclusion** Efficient and accurate road damage detection technologies are of critical importance for the intelligent maintenance of transportation infrastructure. By synergistically combining spatial awareness enhancement with detail preservation, the proposed MSANet framework addresses long-standing challenges of direction modeling and fine-grained feature retention. The model achieves substantial improvements in detection accuracy while preserving real-time inference capability, offering both theoretical innovation and engineering practicality. MSANet thus represents a promising solution for advancing intelligent road infrastructure management, and it provides valuable insights into the design of next-generation detection frameworks that aim to balance efficiency, accuracy, and scalability. With its strong adaptability to both controlled datasets and real-world imagery, MSANet establishes a foundation for future integration into large-scale intelligent transportation systems and smart city applications. The code for the paper will be available athttps://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等，2022)。若不能及时地发现与修复，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。传统人工巡检可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、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、Faster R-CNN)(Ross等，2014；Ren等，2016)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，但伴随处理流程复杂、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、YOLO系列)(Wei等，2016；Joseph等，2015)则将特征提取与目标分类、定位统一在同一网络中，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，成为制约检测算法性能提升的关键瓶颈。现有研究多借助注意力机制(如CBAM、LSKA模块)(Woo等，2018；Lua等，2023)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，卷积神经网络层级卷积与下采样过程中，使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中，在维持142FPS的高实时推理速度的同时，mAP@0.5、mAP@0.5：0.95和F1-Score较于YOLOv11s基准模型分别提高了1.9%、4.9%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等人（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（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。作为YOLO系列的最新演化，YOLOv11（Khanam等，2024）在YOLOv8（Swathi等，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，在无人机平台的道路裂缝检测任务中验证了其较高的识别准确性。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，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。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往往与周围背景纹理相似，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（Multi-stage Spatial-Aware road damage detection Network，MSANet）。网络总体结构如图1所示。该网络通过三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模块中引入坐标感知卷积核（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，我们通过坐标注意力机制（Coordinate Attention，CA）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。具体来说，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$\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，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$\boldsymbol{f}_h$和$\boldsymbol{f}_w$：

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

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

将$\boldsymbol{f}_h$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$\boldsymbol{M}_h\in\mathbb{R}^{C\times H\times 1}$和$\boldsymbol{M}_w\in\mathbb{R}^{C\times 1\times W}$，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，并与原始特征逐原始相乘，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$\boldsymbol{X}_{CA}$：

$$\boldsymbol{X}_{CA}=\boldsymbol{X}_{CA}\odot\sigma(\boldsymbol{M}_h)\odot\sigma(\boldsymbol{M}_w) \tag{3}$$

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，我们引入了一种基于曼哈顿距离的空间自注意力机制（MaSA），通过在注意力计算中融入曼哈顿距离衰减注意力权重，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$\boldsymbol{D}$：

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

其中(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：当两个位置距离越远时，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，我们利用输入特征$\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，经过独立线性变换得到的查询、键、值的注意力矩阵$(\boldsymbol{Q}=$

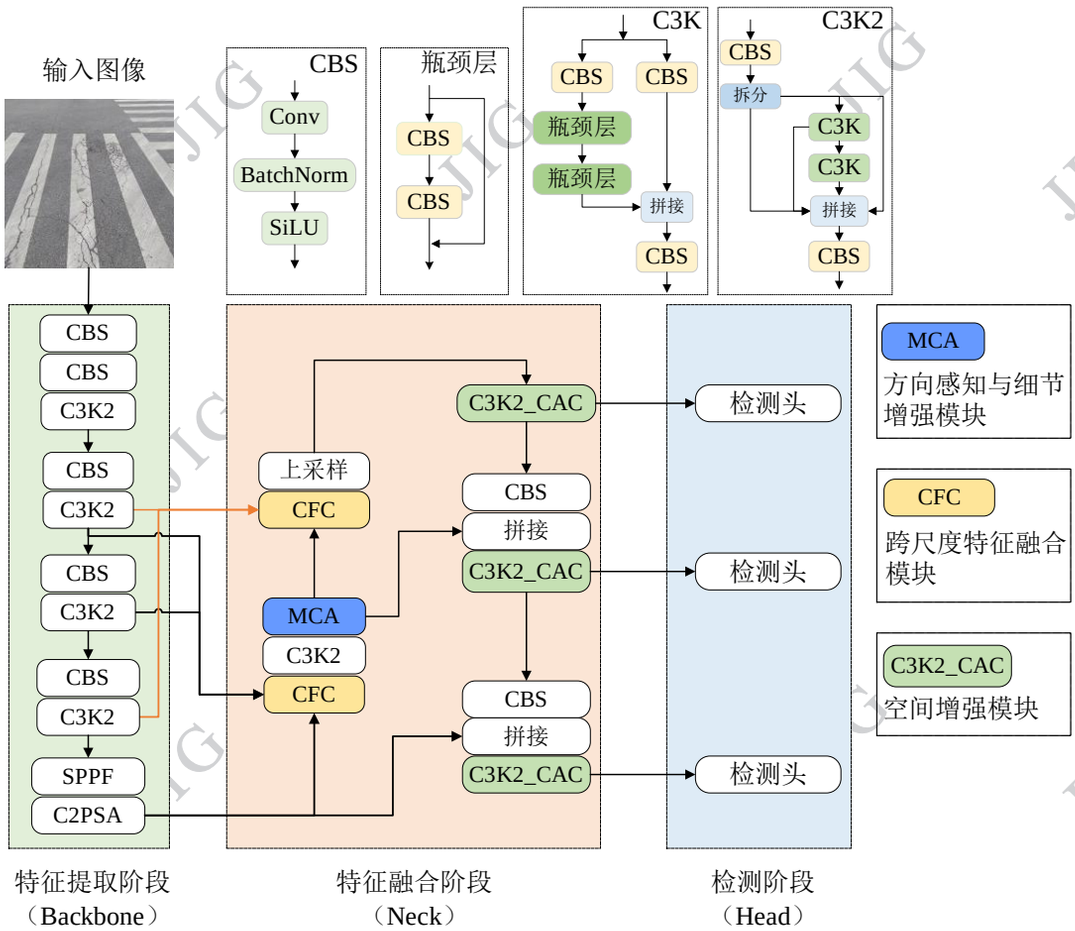


图1　MSANet模型总体结构图

Fig1　Overall architecture diagram of the MSANet model

$\boldsymbol{XW}_Q, \boldsymbol{K}=\boldsymbol{XW}_K, \boldsymbol{V}=\boldsymbol{XW}_V\big)$，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$Attn_W$）并作用于原始信息矩阵$\boldsymbol{V}$，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，又降低自注意力机制的计算复杂度。具体为：

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

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

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

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$Q, K$在水平方向的分解，其中$MaSA(X)$的输出为$\boldsymbol{X}_{MaSA}$。

最后，为了高效融合这两种互补的注意力信息，本文设计的混合注意力机制（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$\boldsymbol{F}_{out}\in\mathbb{R}^{C\times H\times W}$：

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

其中，$\boldsymbol{X}_{CA}$为来自CA注意力分支的输出特征；$\boldsymbol{X}_{MaSA}$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，和结合距离衰减策略突出邻域内的细节特征，使模型对全局方向和局部空间细节均具有敏锐的感知能力。

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/PANet）虽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

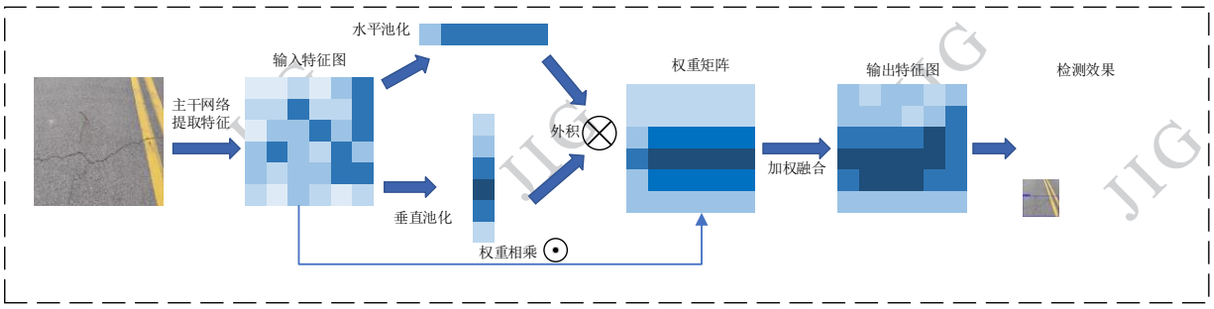


图2　CA坐标注意力机制原理

Figure 2　Principle of the CA coordinate attention mechanism

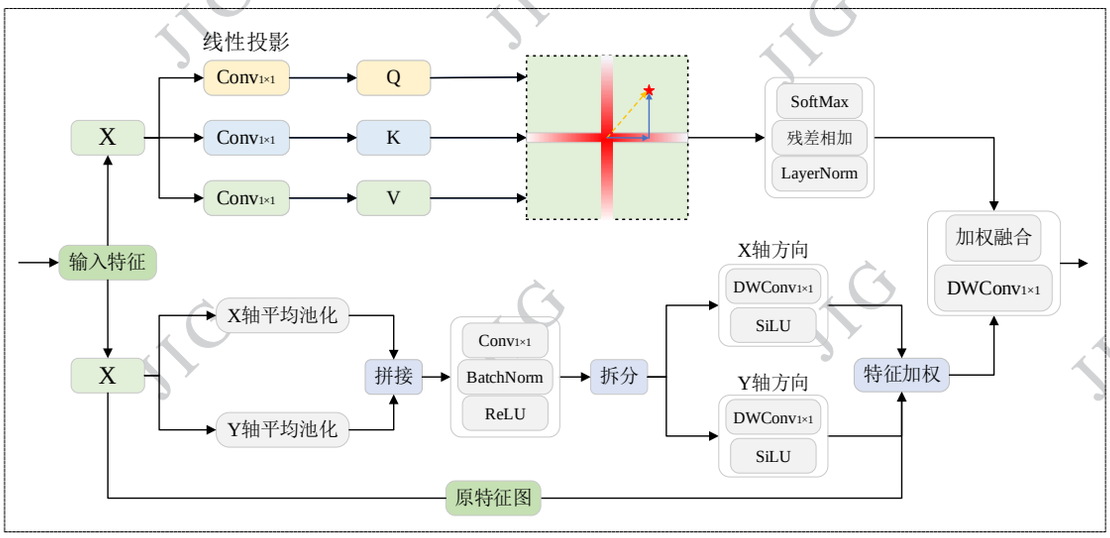


图3　方向感知与细节增强模块(MCA)结构图

Fig 3　Structure diagram of the Directional Awareness and Detail Enhancement Module (MCA)

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)、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,具体模块结构如图5所示),以替代传统的简单拼接操作。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,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$\boldsymbol{F}'_L$:

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

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。具体过程如下所示:

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

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

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

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。

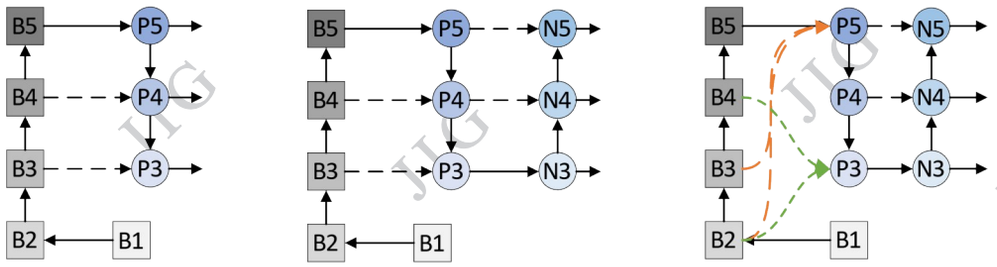


图4　不同特征融合网络结构对比

Figure 4　Comparison of different feature fusion network structures

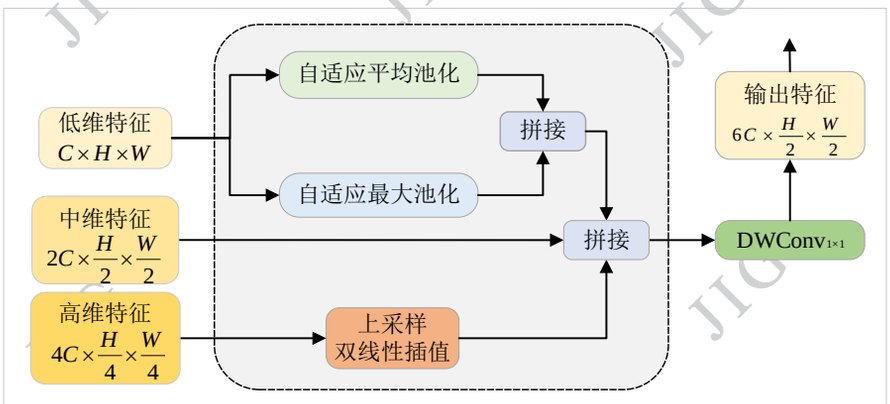


图5　跨尺度特征融合(CFC)模块结构图

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衰弱。因此,本文在C3K2模块中引入了坐标注意力卷积模块(Coordinate Attention Convolution,CAConv),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。

首先输入特征$\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$经过卷积-批归一化-激活的特征提取模块CBS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:

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

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

其中,$C3K$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,我们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:

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

其中⊙表示逐元素乘法,CA表示为坐标注意力编

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：

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

其中输出特征$\boldsymbol{X}_{out}\in\mathbb{R}^{C\times H\times W}$。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。

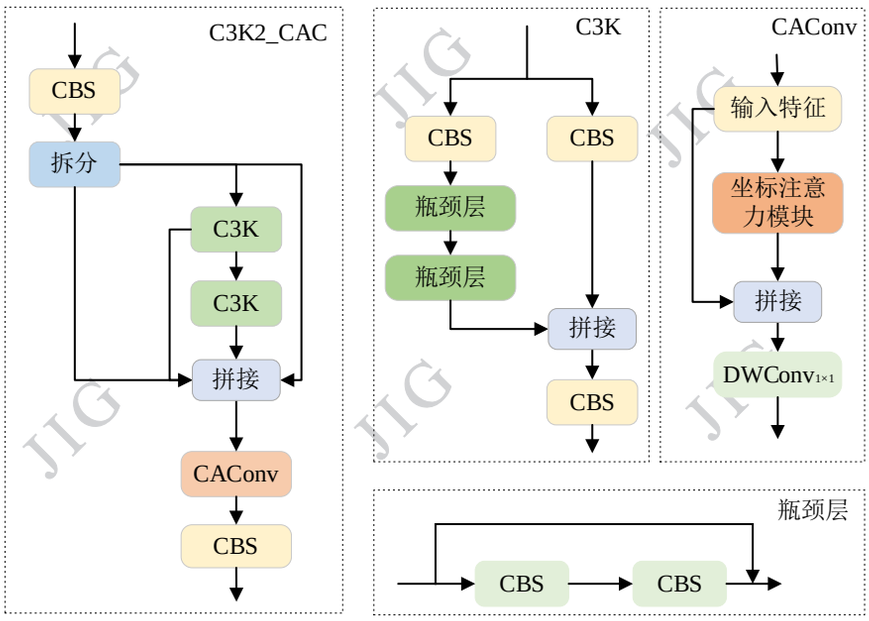


图6　空间增强模块结构图(C3K2_CAC)

Figure 6　Spatial Enhancement Module Structure Diagram (C3K2_CAC)

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，我们搭建了一个实验平台，其操作系统为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。

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。为了实验验证，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，它们具有多种形成原因和潜在风险，其中修补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。我们对数据进行清洗和筛选，最终保留五种类型的道路损伤(示例图片如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)、召回率(Recall)、mAP@0.5、mAP@0.5:0.95，F1-Score，FPS和GFLOPs这七类评价指标对模型的性能进行客观评估。

精确率刻画出模型所有预测中正负样本均被正确分类的比例，且易于计算和理解。召回率度量所有真实正样本中被正确检测出的比例，其公式表

表1　实验设置与超参数详情

Table 1　Details of Experimental Settings and Hyperparameters

操作系统	Ubuntu 20.04
深度学习框架	Pytorch2.2.2
编程语言	Python3.10
CPU	i7-7900X CPU
GPU	NVIDIA TITAN V
运行内存	64GB
学习率	0.01-0.0001
图像尺寸	640×640
优化器	SGD
批处理大小	32
随机种子	42
权重衰减	0.0005
训练轮次	300
其他超参数	默认

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	3686	416
横向裂缝	D10	2106	251
网状裂缝	D20	851	82
坑洞	D40	289	32
修补区域	/	945	100

注：修补区域无危险等级代码，标签中使用repair标记。

达为：

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

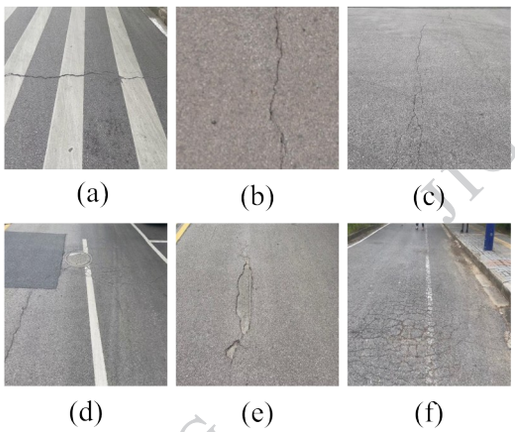
$$Recall = \frac{TP}{TP + FN} \tag{17}$$

式中TP、TN、FP、FN分别代表真阳性、真阴性、假阳性、假阴性。

平均精度mAP@0.5是在单IoU阈值0.5下，对每个类别计算P-R曲线下的面积（AP），再对所有类别取平均。其公式表达为：

$$AP_{0.5} = \int_0^1 P(r)\,dr \tag{18}$$

$$mAP_{0.5} = \frac{1}{C}\sum_{c=1}^{C} AP_{0.5}^{c} \tag{19}$$



(a)　(b)　(c)

(d)　(e)　(f)

(a)横向裂缝，(b)(c)纵向裂缝，(d)修补区域，(e)坑洞，(f)网状裂缝

(a) Transverse cracks, (b)(c) longitudinal cracks, (d) repair areas, (e) pits, (f) reticular cracks

图7　数据集不同缺陷图像示例

Figure 7　Examples of different defect images in the dataset

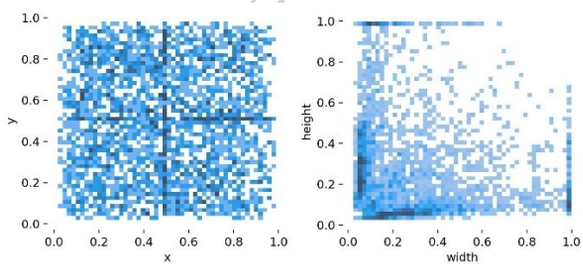


图8　数据集中目标标签位置分布与长宽比分布

Fig 8　Distribution of target label positions and aspect ratios

式中$P(r)$是预测的精度-召回率曲线，C是类别数。

因mAP@0.5的单阈值可能掩盖模型对高精度定位能力，引入mAP@0.5:0.95评测标准，以全面反映模型在不同定位精度要求下的性能。mAP@0.5:0.95分别在$IoU = 0.50, 0.55, \cdots 0.95$共十个阈值下计算AP，再取平均，其公式表达为：

$$mAP_{0.5-0.95} = \frac{1}{10C}\sum_{t\in\{0.50,0.55,\cdots,0.95\}}\sum_{c=1}^{C} AP_t^c \tag{20}$$

F1分数（F1-score）是精度和召回率的调和平均，对不平衡数据集更为准确或对召回率较高的模型具有惩罚作用，促进二者平衡。其公式表达为：

$$F1_{Score} = \frac{2 \times P \times R}{P + R} \tag{21}$$

帧率（FPS）衡量模型在给定硬件平台上每秒能处理图像的数量，反映系统的实时性能。计算量（GFLOPs）衡量模型在一次前向推理所需的浮点运算量，客观地评估模型复杂度。

研究中通过多维度指标全面衡量模型在检测精度、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。

表3　不同模型对比实验结果

Table 3　Comparative experimental results of different models

Method	P	R	mAP @0.5	mAP @0.5:0.95	F1_Score	GFLOPs(G)	FPS
Faster R-CNN(Ren等,2016)	0.701	0.783	0.732	0.453	0.740	940.7	11
YOLOv5s(Khanam等,2024)	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等,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,2024)	0.841	0.770	0.843	0.578	0.804	27.4	188
YOLOv10s(Wang等,2024)	0.872	0.752	0.845	0.602	0.807	24.8	158
YOLOv11s(Khanam等,2024)	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,2024)	0.834	0.782	0.844	0.599	0.807	72.6	71
MSANet(本文模型)	0.883	0.811	**0.877 (+1.9%)**	**0.630 (+4.9%)**	**0.845 (+1.8%)**	32.0	142

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模型的提升。

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4.9%，比mAP@0.5产生了更明显的增幅，反映模型对高质量预测的显著提升。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%的提升，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。在道路损伤各类损伤检测的细粒度任务中（表4），MSANet进一步体现场景目标适应性。mAP@0.5提升4.15%至0.878，mAP@0.5：0.95提升3.18%至0.649，F1-Score提升5.36%至0.629；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2.70%，F1-Score提升5.96%；坑洼检测（D40）：mAP@0.5提升7.13%，F1-Score提升8.38%。同时，模型在保持142FPS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8所示），本文模型展现出更高的目标区域聚焦度，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。实验结果证明，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。

表4　细粒任务对比实验结果

Table 4　Comparative experimental results on fine-grained tasks

模型	类别 指标	全部类别	裂缝类	坑洞	修补
基准模型	mAP@0.5	0.843	0.853	0.827	0.850
	mAP @0.5:0.95	0.629	0.578	0.521	0.595
	F1-Score	0.597	0.554	0.501	0.735
MSANet (本文模型)	mAP@0.5	0.878(**+4.15%**)	0.876(**+2.70%**)	0.886(**+7.13%**)	0.873(**+2.71%**)
	mAP @0.5:0.95	0.649(**+3.18%**)	0.601(**+4.01%**)	0.547(**+4.99%**)	0.610(**+2.52%**)
	F1-Score	0.629(**+5.36%**)	0.587(**+5.96%**)	0.543(**+8.38%**)	0.758(**+3.13%**)

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

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。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。虽然对mAP@0.5指标提升有限，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。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%。在改进的网络框架中，进一步对比了多种注意力机制。实验结果显示，所采用的MCA混合注意力模块相较于其他对比机制，取得了最优的性能表现。集成上述改进模块后，最终的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4.9%，F1-Score提升了1.8%（详情见表5所示），并保持142FPS的推理速度。

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	GFLOPs
			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：▮表示使用了该模块，加粗字体表示该指标下获得的最佳数据。

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）三种策略，具体网络结构如图9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。

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，两者不同程度上的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

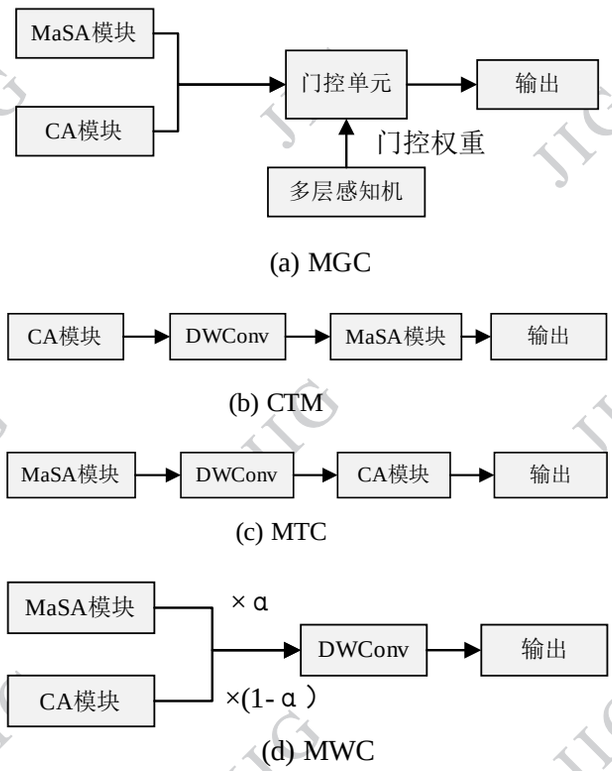


图9　不同融合策略网络结构图

Figure 9　Network structure diagrams of different fusion strategies

验结果如表6所示。其中CTM网络结构取得最差提升，我们分析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的全局感知信息中取得最佳融合。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　跨尺度特征融合模块(CFC)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，研究中系统性地对其部署位置及数量进行了消融实验(详见表7)。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，我们在多个公开道路损伤数据集进行跨数据集实验和在实地拍摄图像进行泛化性实验。

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。我们对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、坑洞四种缺陷类别的约三万张图片，并在本文的网络中进行训练和推理。我们在此数据集中与其他最佳目标检测网络进行实验。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，对数据集进行道路损伤检测更具挑战性。我们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数据集是由无人机在30米高度拍摄，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2592×1944分辨率图像组成。我们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。

UAPD数据集。由无人机采集，包含六类道路病害：横向裂缝，纵向裂缝，网状裂缝，斜裂缝，修复区域和坑洼，由3100张512×512的图像组成。我们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1.49%(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。

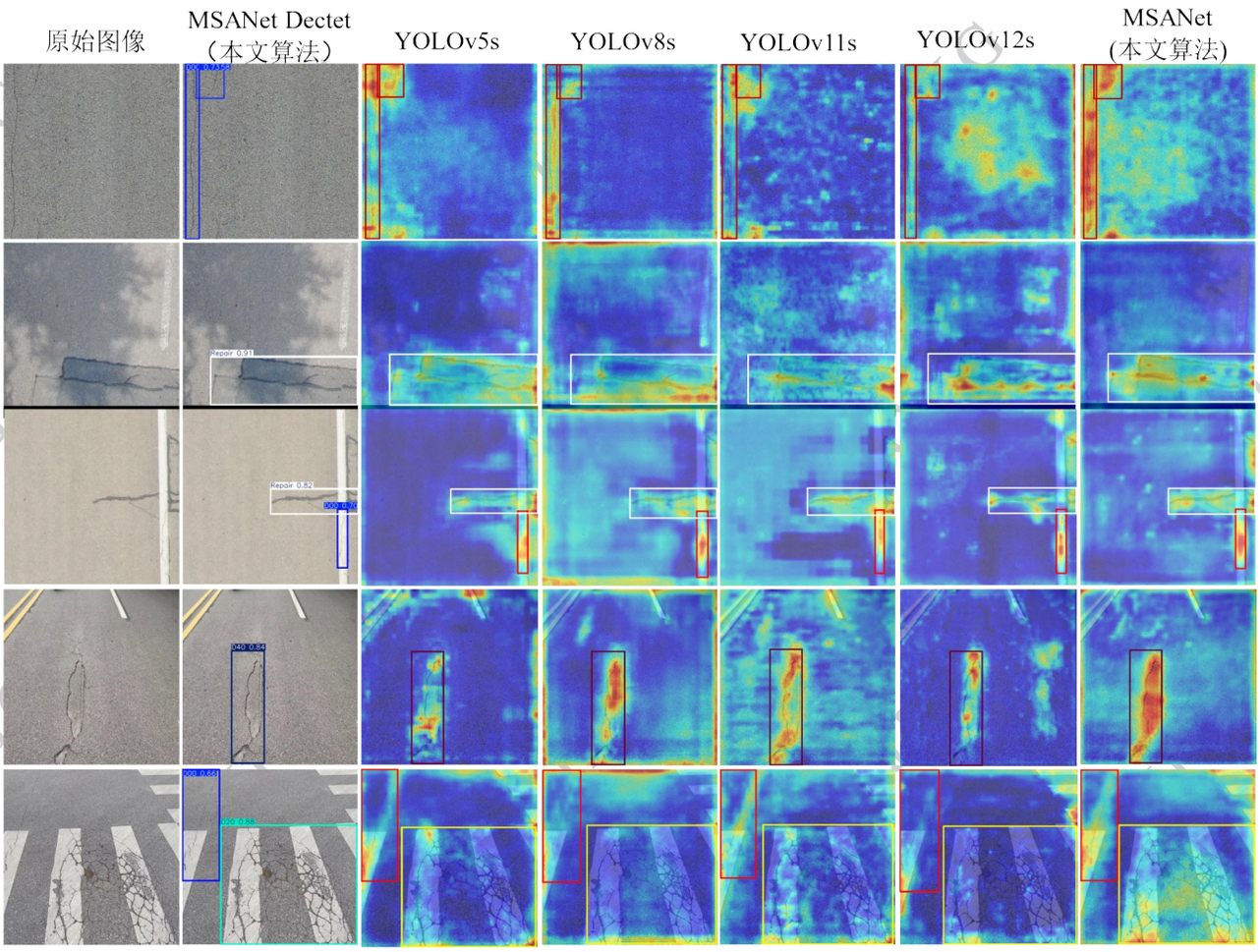


图8　不同模型热力图比较

Figure 8　Heatmap comparison of different model

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	GFLOPs
CA	—	0871	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	0.61	0.825	22.6

注：其他实验模块无权重数据，使用“—”替代，表中加粗文字为该指标下获得的最佳数据。

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	GFLOPs
▮			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：▮表示使用了该模块，加粗字体表示该指标下获得的最佳数据。

同时，为了验证模型在工程实践场景下的检测效果，我们对MSANet模型进行路面损伤检测实践。

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(Ours)	0.835 **(+2.83%)**	0.601 **(+3.26%)**	0.819 **(+1.61%)**

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

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	0.28	0.124	0.536
YOLOv8s	0.281	0.126	0.325
YOLOv11s	0.287	0.125	0.350
MSANet(Ours)	0.298 **(+3.83%)**	0.133 **(+6.40%)**	0.363 **(+3.71%)**

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

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	0.94	0.746	0.920
MSANet(Ours)	0.965 **(+2.66%)**	0.807 **(+8.18%)**	0.953 **(+3.59%)**

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

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(Ours)	0.265 **(+2.71%)**	0.194 **(+10.23%)**	0.267 **(+2.69%)**

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

使用大恒工业相机MER-132-43U3C，采集图像大小为1292×964。在2m高度进行俯拍和1.5m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。拍摄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。我们使用在多源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10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，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通路，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的精度，mAP@0.5:0.95相较于基线提升了4.9%，同时保持142FPS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。

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 mAP @0.5	Box mAP @0.5:0.95	Box F1-Score	Mask mAP @0.5	Mask mAP @0.5:0.95	Mask F1-Score
YOLOv11s	0.832	0.649	0.805	0.650	0.199	0.717
本文方法	0.833 **(+0.17%)**	0.663 **(+2.16%)**	0.809 **(+0.49%)**	0.659 **(+1.38%)**	0.202 **(+1.49%)**	0.720 **(+0.42%)**

注：加粗字体表示本文模型较于基线YOLOv11s的提升。

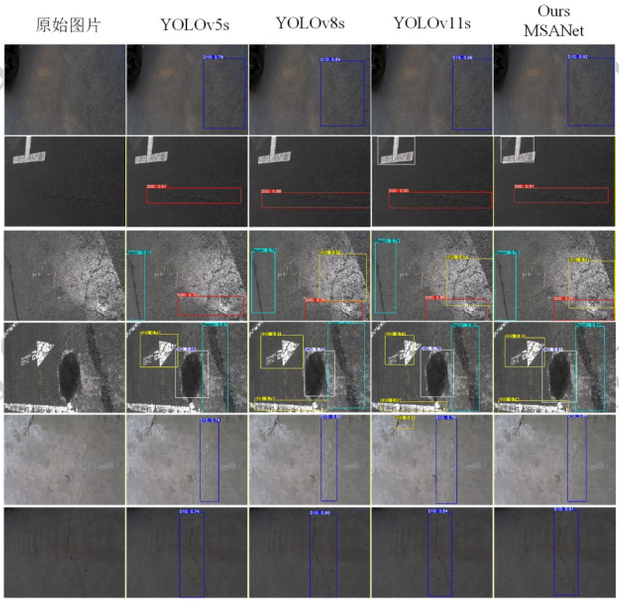


图10　自收集数据检测结果

Figure 10　Detection results on the self-collected dataset

参考文献(References)

Khanam R and Hussain M. 2024. YOLOv11: An Overview of the Key Architectural Enhancements.[EB/OL].[2024-10-23]. https://arxiv.org/pdf/2410.17725.pdf

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 Adaptation Measures.Sustainability,14(14): 8864 [DOI: 10.3390/su14148864]

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(04):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(04):1019-1028)[DOI:10.11834/jig.211129]

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]

Leng J X, Mo M J C, Zhou Y H, Ye Y M, Gao C Q and Gao X B. 2023. Recent advances in drone-view object detection. Journal of Image and Graphics, 28(09):2563-2586 (冷佳旭,莫梦竞成,周应华,叶永明,高陈强,高新波. 2023. 无人机视角下的目标检测研究进展. 中国图象图形学报,28(09):2563-2586) [DOI: 10.11834/jig.220836.]

Wu P, Wu J and Xie L Q. 2024. Pavement distress detection based on improved feature fusion network. Measurement, Volume 236: 115-119 [DOI: 10.1016/j.measurement.2024.115119]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: Convolutional Block Attention Module. [EB/OL].[2018-07-17]. https://arxiv.org/pdf/1807.06521.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]

Zhang, L L, Yang L, 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]

Fan Q, Huang H, Chen M, Liu H, and He R. 2024. Rmt: Retentive networks meet vision transformers//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle, WA, USA: IEEE: 5641-5651 [DOI: 10.1109/CVPR52733. 2024.00539]

Hou Q, Zhou D and Feng J. 2021. Coordinate attention for efficient mobile network design//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville, TN, USA, IEEE: 13713-13722 [DOI: 10.1109/CVPR46437. 2021.01350]

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]

Ma D, Fang H, Wang N, Zhang C, Dong J and Hu H. 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]

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］

Chu 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，38(5)：629－642［DOI：10.1111/mice.12881］

Arya D，Maeda H，Ghosh S K，Toshniwal D，Omata H，Kashiyama T and Sekimoto Y. 2020. Global Road Damage Detection：State-of-the-art Solutions. IEEE International Conference on Big Data（Big Data），Atlanta，GA，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）//2022 IEEE International Conference on Big Data（Big Data）. Osaka，Japan：IEEE：6378-6386［DOI：10.1109/BigData55660.2022.10021040］

Swathi Y and Challa M. 2024. YOLOv8：Advancements and innovations in object detection. Smart Trends in Computing and Communications，946：1-13［DOI：10.1007/978-981-97-1323-3_1］

Zhang X，Liu C，Yang D，Song T，Ye Y，Li K，et al. 2023. RFAConv：Innovating Spatial Attention and Standard Convolutional Operation.［EB/OL］.［2023-03-28］.
https://arxiv.org/pdf/2304.03198.pdf

Khanam R and Hussain M. 2024. What is YOLOv5：A deep look into the internal features of the popular object detector.［EB/OL］.［2024-07-30］.
https://arxiv.org/pdf/2407.20892.pdf

S Ren，K He，R Girshick and J Sun.2016.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］

Li C，Li L，Jiang H，Weng K，Geng Y，Li L，et al. 2022.YOLOv6：A single-stage object detection framework for industrial applications.［EB/OL］.［2022-09-07］.
https://arxiv.org/pdf/2209.02976.pdf

Han K，Wang Y，Tian Q，Guo J，Xu C and Xu C. 2020.Ghostnet：More features from cheap operations//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition（CVPR）. Seattle，WA，USA：IEEE：1580-1589［DOI：10.1109/CVPR42600. 2020.00165］

Zhao Y，Lv W，Xu S，Wei J，Wang G，Dang Q，et al. 2024. Detrs beat yolos on real-time object detection//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition（CVPR）. Seattle，WA，USA：IEEE：16965-16974［DOI：10.1109/CVPR52733.
2024.01605］

Peng Y，Li H，Wu P，Zhang Y，Sun X and Wu F. 2024. D-FINE：Redefine Regression Task in DETRs as Fine-grained Distribution Refinement.［EB/OL］.［2024-10-17］.
https://arxiv.org/pdf/2410.13842.pdf

Kang M，Ting C M，Ting F F，and Phan R C W. 2024. ASF-YOLO：A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. Image and Vision Computing，147：105057［DOI：10.1016/j.imavis.2024.105057］

Wang C Y，Yeh I H and Liao H Y M. 2024. YOLOv9：Learning What You Want to Learn Using Programmable Gradient Information // Proceedings of the European Conference on Computer Vision（ECCV）. Cham，Switzerland：Springer Nature Switzerland：1－21［DOI：10.1007/978-3-031-72751-1_1］

Wang A，Chen H，Liu L，Chen K，Lin Z and Han J. 2024. YOLOv10：Real-Time End-to-End Object Detection. Advances in Neural Information Processing Systems，37：107984－108011［DOI：10.52202/079017-3429］

Tian Y，Ye Q，and Doermann D. 2025. YOLOv12：Attention-Centric Real-Time Object Detectors.［EB/OL］.［2025-12-18］.
https://arxiv.org/pdf/2502.12524.pdf

Lei M，Li S，Wu Y，Hu H，Zhou Y，Zheng X，et al. 2025. YOLOv13：Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception.［EB/OL］.［2025-06-21］.
https://arxiv.org/pdf/2506.17733.pdf

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］

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

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］.

作者简介

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

方明，通讯作者，男，教授，主要研究方向为机器视觉，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