

基于CAN-YOLOv11n的道路缺陷检测方法

丁富强, 秦喜喜, 董骏辉, 薛小维*

重庆文理学院数学与人工智能学院, 重庆

收稿日期: 2026年7月20日; 录用日期: 2026年9月7日; 发布日期: 2026年9月18日

摘要

针对道路缺陷检测中裂缝目标细长、尺度变化显著、背景纹理复杂以及小目标定位不稳定等问题, 本文在YOLOv11n的基础上提出CAN-YOLOv11n道路缺陷检测模型。首先, 在P4分支引入CAA模块, 增强缺陷区域的方向性上下文响应, 并抑制路面阴影、标线和纹理噪声干扰; 其次, 采用ASFF对多尺度特征进行自适应融合, 提升细小裂缝、坑洞和修补区域在不同尺度下的表达能力; 最后, 在边界框回归中引入NWD-CIoU, 使CIoU的几何约束与NWD对小目标位置偏移的鲁棒性形成互补。实验结果表明, 在RDD2022中国区域数据集上, CAN-YOLOv11n的R、mAP50和mAP50-95分别达到84.0%、88.6%和59.4%, 较YOLOv11n基线分别提升1.8、2.4和0.3个百分点。所提方法能够减少复杂路面场景中的缺陷漏检, 并提高道路缺陷的整体检测性能。

关键词

道路缺陷检测, YOLOv11n, CAA, ASFF, NWD-CIoU

Road Defect Detection Method Based on CAN-YOLOv11n

Fuqiang Ding, Xixi Qin, Junhui Dong, Xiaowei Xue*

School of Mathematics and Artificial Intelligence, Chongqing University of Arts and Sciences, Chongqing

Received: July 20, 2026; accepted: September 7, 2026; published: September 18, 2026

Abstract

To address the challenges of elongated cracks, large-scale variation, complex pavement textures, and unstable localization of small defects in road defect detection, this paper proposes CAN-YOLOv11n road defect detection model based on YOLOv11n. First, CAA is introduced into the P4 branch to strengthen directional contextual responses and suppress interference from pavement shadows,

*通讯作者。

文章引用: 丁富强, 秦喜喜, 董骏辉, 薛小维. 基于 CAN-YOLOv11n 的道路缺陷检测方法[J]. 人工智能与机器人研究, 2026, 15(5): 1231-1241. DOI: 10.12677/airr.2026.155112

lane markings, and texture noise. Second, ASFF adaptively fuses multi-scale features to improve the representation of small cracks, potholes, and repaired areas at different scales. Finally, NWD-CIoU is adopted for bounding-box regression, combining the geometric constraints of Ciou with the robustness of NWD to positional deviations of small targets. Experiments on the China subset of RDD2022 show that CAN-YOLOv11n achieves 84.0% recall, 88.6% mAP50, and 59.4% mAP₅₀₋₉₅, outperforming the YOLOv11n baseline by 1.8, 2.4, and 0.3 percentage points, respectively. The proposed method reduces missed detections in complex pavement scenes and improves overall road defect detection performance.

Keywords

Road Defect Detection, YOLOv11n, CAA, ASFF, NWD-CIoU

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

道路缺陷的及时识别是制定养护计划、评估道路服役状态和保障行车安全的重要基础。实际路面中的裂缝、坑洞与修补区域受材料老化、交通荷载和环境变化共同影响,常呈现尺度跨度大、形态不规则、边缘弱及背景纹理复杂等特征,使自动检测需要同时兼顾准确性、鲁棒性和实时性。现有道路损伤检测研究已形成从人工特征到深度学习的技术演进,但复杂场景下的小目标漏检和类别混淆仍是主要瓶颈[1]。针对沥青路面裂缝,光照变化、阴影、污渍和路面纹理会削弱裂缝与背景之间的对比度,细长裂缝还容易在下采样过程中丢失连续结构[2]。Maeda 等利用车载智能手机图像验证了深度神经网络开展道路损伤识别的可行性[3]。随后, RDD2022 汇集多国家、不同采集平台和道路环境的图像,为道路缺陷检测的跨场景研究提供了统一数据基础[4]。在检测框架方面, YOLO 系列凭借端到端预测和较高推理效率,已成为工程化道路巡检中常用的单阶段检测方法[5]。

围绕复杂背景、多尺度目标与轻量化部署,国内外研究对道路缺陷检测网络进行了多方面改进。Liu 等基于运动相机采集的街景图像开展路面病害检测,说明视角变化和真实道路背景会显著影响模型稳定性[6]。Xie 等提出 YOLO-SFT,通过特征扩散强化不同层级信息交互,在控制计算开销的同时提升道路损伤检测性能[7]。Wang 等构建 YOLO-RD,针对小尺度病害和复杂环境引入专门的特征增强策略,验证了任务导向结构设计的有效性[8]。对于裂缝的细长方向结构,Qu 等利用方向连通特征增强裂缝区域的连续表达[9]。Guo 等提出边界感知细化网络,通过强化边缘信息改善裂缝轮廓识别[10]。Yang 等设计渐进式自适应融合网络,以逐级融合不同层次信息提高复杂路面裂缝分割质量[11]。上述方法表明,增强有效区域响应、融合跨尺度信息和改善边界表达均能提升道路缺陷识别能力,但单一结构优化往往难以同时解决背景干扰、小缺陷漏检和定位不稳定问题。

注意力、多尺度融合与回归损失的协同优化为解决上述问题提供了可行路径。CAA 通过水平与垂直方向的条带卷积建立方向性长程上下文,可补偿细长裂缝在下采样过程中的连续性损失[12]; ASFF 通过空间位置相关的尺度权重抑制不同层级特征之间的冲突,使高分辨率细节与高层语义按目标区域自适应组合[13]; NWD 将小目标边界框表示为二维高斯分布,可降低少量像素偏移造成的相似度突变[14],而 Ciou 类约束则保留重叠程度、中心距离和长宽比等几何信息[15]。

CAA 增强后的 P4 特征同时进入中尺度检测分支和 ASFF 权重生成过程,可作为包含裂缝方向连续

性与背景抑制信息的隐式特征先验，引导 ASFF 在裂缝区域选择更合适的尺度来源；融合后的特征再由 NWD-CIoU 进行稳健定位，使“方向性表征 - 空间尺度选择 - 边界框回归”形成连续的优化链路。

基于上述分析，本文以 YOLOv11n 为基线提出 CAN-YOLOv11n 道路缺陷检测模型，围绕方向性上下文增强、空间自适应多尺度融合和小目标稳健回归构建协同优化方案。主要改进如下：

- 1) 针对路面阴影、标线和复杂纹理容易淹没缺陷响应的问题，引入 CAA 上下文锚点注意力模块，利用水平与垂直方向的长程上下文建模增强裂缝、坑洞和修补区域的有效特征，并抑制无关背景干扰。
- 2) 针对道路缺陷尺度差异显著以及小裂缝在特征融合中信息不足的问题，引入 ASFF 自适应空间特征融合模块，对多尺度特征进行位置相关的动态加权，使高分辨率细节与深层语义能够按目标尺度进行有效组合。
- 3) 针对细长和小尺度缺陷对像素级偏移敏感、边界框回归不稳定的问题，构建 NWD-CIoU 组合损失，以 Ciou 保持预测框的整体几何约束，并利用 NWD 提高小目标位置偏差下的回归鲁棒性。

2. YOLOv11 模型

YOLOv11 是 Ultralytics 在 2024 年推出的新一代实时目标检测模型，其核心改进主要集中在骨干网络和颈部网络的结构优化，设计目标是在控制参数数量与计算开销的同时，提高模型的特征提取能力和检测效率。具体而言，YOLOv11 采用 C3k2 作为关键特征提取单元，通过更灵活的卷积结构和跨阶段特征传递增强特征融合与梯度流动能力；同时在主干网络末端引入 C2PSA 模块，利用空间注意力机制强化关键区域响应并抑制无关背景信息。此外，模型通过 SPPF 模块扩展感受野，并利用颈部网络对不同尺度特征进行充分融合，最终由检测头完成目标类别预测和边界框回归。基于上述结构优化，YOLOv11 在检测精度、推理速度和模型复杂度之间具有较好的平衡。其中，YOLOv11n 作为该系列的轻量化版本，参数数量和计算量较低，适用于对实时性与计算资源较为敏感的道路缺陷检测场景。

3. CAN-YOLOv11n 模型

道路缺陷并非均表现为边界清晰、形态紧致的目标：纵向裂缝与横向裂缝通常呈稀疏、细长且方向性显著的线状结构，网状裂缝包含多方向分叉与交叉区域，坑洞和修补区域则具有尺度变化大、轮廓不规则及边界模糊等特点。在逐层下采样过程中，细长裂缝所占像素较少，连续的裂缝结构容易在下采样过程中退化为离散响应；不同尺度特征若采用固定方式融合，又可能将浅层纹理噪声与深层语义一并传递；对于小尺度或极端长宽比缺陷，预测框发生少量像素偏移还会引起 IoU 显著波动。针对上述相互关联的问题，本文在 YOLOv11n 基础上构建 CAN-YOLOv11n：在 P4/16 分支引入 CAA 形成 CAA-P4，在检测头前利用 ASFF 融合 P3、CAA-P4 与 P5 并生成 ASFF-P3，同时采用 NWD-CIoU 优化边界框回归。最终检测头的三尺度输入分别为 ASFF-P3、CAA-P4 和 P5，整体结构如图 1 所示。

在注意力机制选择与嵌入位置方面，SE 主要进行通道重标定，缺乏对裂缝空间走向的显式描述；CBAM 虽然结合通道与空间注意力，但对长距离线状结构的连续建模能力有限；Coordinate Attention 能够保留方向位置信息，但方向池化可能压缩局部细节。相比之下，CAA 利用水平与垂直条带深度卷积建立方向性长程依赖，更符合纵向裂缝、横向裂缝及网状裂缝的结构特征。考虑到 P3 层纹理噪声较多，而 P5 层空间分辨率较低，本文将 CAA 置于 P4 分支的 C3k2 之后，使其在空间细节与语义信息之间取得平衡，输出方向特征更突出、背景干扰更少的 CAA-P4。

在多尺度融合方面，直接相加难以区分不同尺度特征的重要程度，Concat 容易同时传递浅层噪声和冗余语义，而尺度级固定权重也难以适应道路缺陷在不同空间位置上的形态变化。因此，本文采用 ASFF 对 P3、CAA-P4 和 P5 进行自适应融合。其中，P3 主要提供细小裂缝的边缘与纹理细节，CAA-P4 提供方

向连续性和中层结构信息, P5 则补充坑洞、修补区域等目标所需的高层语义。ASFF 根据各空间位置的特征响应动态分配尺度权重, 并仅生成面向小目标检测的 ASFF-P3, 以增强小尺度缺陷表达, 同时保留 CAA-P4 和 P5 的独立检测路径。

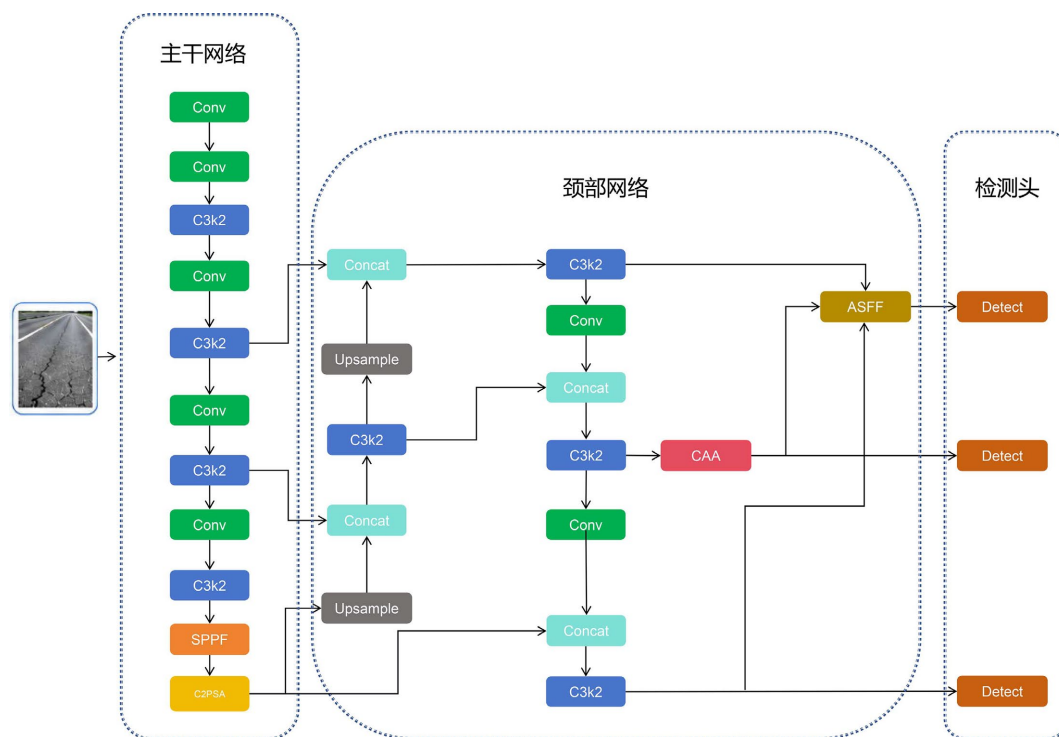


Figure 1. Structure diagram of CAN-YOLOv11n model

图 1. CAN-YOLOv11n 模型结构图

在边界框回归方面, CIoU 能够约束预测框的重叠程度、中心距离和长宽比, 但细长裂缝及小尺度缺陷对少量像素偏移较为敏感。NWD 通过高斯分布距离连续描述预测框与真实框的位置和尺度差异, 可在轻微错位或重叠不足时提供更加平滑的回归信号。因此, NWD-CIoU 既保留 CIoU 的整体几何约束, 又利用 NWD 提高细长、小尺度缺陷的定位稳定性, 并且仅在训练阶段参与损失计算, 不增加推理路径。

三项改进按照“方向特征提纯-空间尺度选择-边界框稳定回归”的顺序逐级协同。首先, CAA 在 P4 层对原始中尺度特征进行方向性上下文建模, 使沿裂缝延伸方向分布的离散响应得到联系, 同时削弱阴影、标线和粗糙纹理产生的无关响应, 形成 CAA-P4。该特征随后不仅用于中尺度检测, 还直接作为 ASFF 的一路输入参与权重生成。由于 ASFF 的尺度权重由对齐后的 P3、CAA-P4 和 P5 特征共同计算, CAA-P4 中更清晰的裂缝连续响应有助于使权重生成分支获得裂缝信息与尺度信息的判别依据。因此, CAA 所提取的方向性上下文实际上为 ASFF 的尺度感知提供了隐式先验: 在细长裂缝区域, ASFF 可结合 P3 的高分辨率边缘细节与 CAA-P4 的连续结构信息, 避免裂缝在融合过程中被背景纹理削弱; 在坑洞或修补区域, 则可更多引入 P5 的高层语义。融合得到的 ASFF-P3 进一步提升了小尺度缺陷区域的可分辨性, 而 NWD-CIoU 在训练阶段对预测框施加几何与分布双重约束, 并通过反向传播将定位误差传递至检测头、ASFF 和 CAA, 使前端学习到的方向与尺度特征更有利于边界框回归。因此, CAA 为 ASFF 提供经过方向增强的特征基础, ASFF 将该先验与不同尺度信息进行位置相关的选择和整合, NWD-CIoU 再将融合后的表征优势转化为更加稳定的定位结果, 形成完整的协同优化链路。

3.1. CAA 上下文锚点注意力模块

针对道路缺陷图像中裂缝细长、方向性明显且易受阴影、标线和路面纹理干扰的问题，本文在 YOLOv11n 的 P4/16 中尺度特征分支引入上下文锚点注意力(Context Anchor Attention, CAA)模块。CAA 位于颈部网络 P4 分支的 C3k2 模块之后、检测头之前；增强后的 CAA-P4 一方面作为中尺度检测分支输入，另一方面参与 ASFF-P3 的多尺度融合。P4 层兼具一定空间分辨率和较充分的语义信息，有利于在控制计算开销的同时增强中尺度裂缝、坑洞和修补区域的上下文响应。CAA 模块结构如图 2 所示。

CAA 模块结构图

Context Anchor Attention (used in CAN-YOLOv11n, P4/16, C=512)

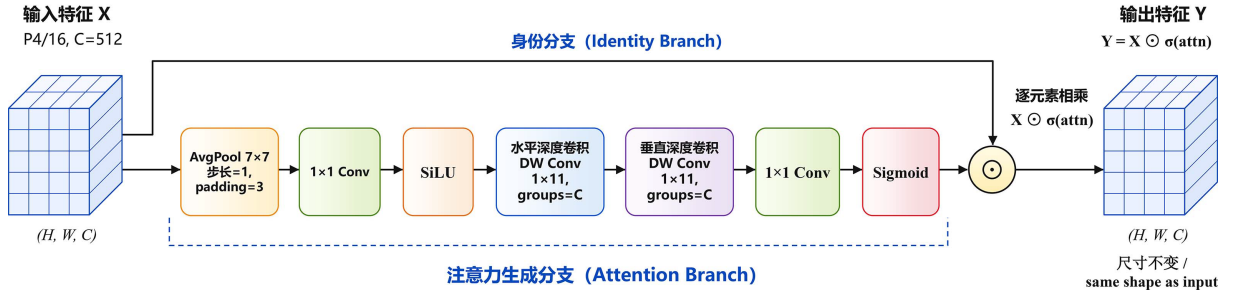


Figure 2. Structure diagram of the CAA module

图 2. CAA 模块结构图

设输入特征 X 的尺寸为 $C \times H \times W$ 。首先，CAA 采用核大小为 7×7 、步长为 1、填充为 3 的平均池化对局部邻域进行平滑聚合，在不改变特征图空间尺寸的前提下获得上下文锚点特征：

$$X_p = \text{AvgPool}_{7 \times 7}(X) \quad (1)$$

随后，使用 1×1 卷积对上下文锚点特征进行通道映射，并通过 SiLU 激活函数引入非线性表达，得到中间特征 X_c ：

$$X_c = \delta(W_1 * X_p) \quad (2)$$

为捕获道路裂缝沿水平和垂直方向的长程依赖，模块依次采用 1×11 的水平深度卷积和 11×1 的垂直深度卷积。两层卷积均设置 $\text{groups} = C$ ，使每个通道独立完成方向空间建模，从而在较低计算开销下扩大有效感受野：

$$X_h = \text{DWConv}_{1 \times 11}(X_c), \quad X_v = \text{DWConv}_{11 \times 1}(X_h) \quad (3)$$

完成水平与垂直方向建模后，模块再次利用 1×1 卷积整合通道信息，并通过 Sigmoid 函数将响应映射到 $[0, 1]$ 区间，形成注意力权重图 A ：

$$A = \sigma(W_2 * X_v) \quad (4)$$

最后，将注意力权重图 A 与原始输入特征 X 进行逐元素相乘，得到 CAA 模块的输出特征 Y ：

$$Y = X \odot A \quad (5)$$

式中，AvgPool 表示平均池化操作； W_1 和 W_2 分别表示前后两个 1×1 卷积的权重； $\delta(\cdot)$ 表示 SiLU 激活函数；DWConv 表示深度卷积； $\sigma(\cdot)$ 表示 Sigmoid 函数； $*$ 表示卷积运算； $\odot$ 表示逐元素相乘。由于平均

池化和两个方向深度卷积均采用步长 1 及对称填充, 输出特征 Y 与输入特征 X 的尺寸保持一致。CAA 通过水平与垂直方向的大核深度卷积建立长距离上下文联系, 能够增强细长裂缝及不规则缺陷区域的连续响应, 并抑制阴影、标线和冗余纹理干扰。同时, 本文仅在 P4/P6 分支使用该模块, 使特征增强效果与额外计算开销之间保持较好的平衡。

3.2. ASFF 自适应空间特征融合模块

针对道路缺陷在不同尺度特征层中存在语义信息与细节信息分布不一致的问题, 本文在 YOLOv11n 检测头前引入自适应空间特征融合(Adaptive Spatial Feature Fusion, ASFF)模块。与固定拼接或直接相加不同, ASFF 能够在每个空间位置自适应学习 P3、CAA-P4 和 P5 三种尺度特征的融合权重, 使网络根据目标尺寸及局部背景选择更有效的特征来源。根据本文模型配置, ASFF 的 level 设置为 0, 仅生成 P3/8 高分辨率融合特征; 融合后的 ASFF-P3、CAA-P4 与 P5 共同输入检测头。ASFF-P3 模块结构如图 3 所示。

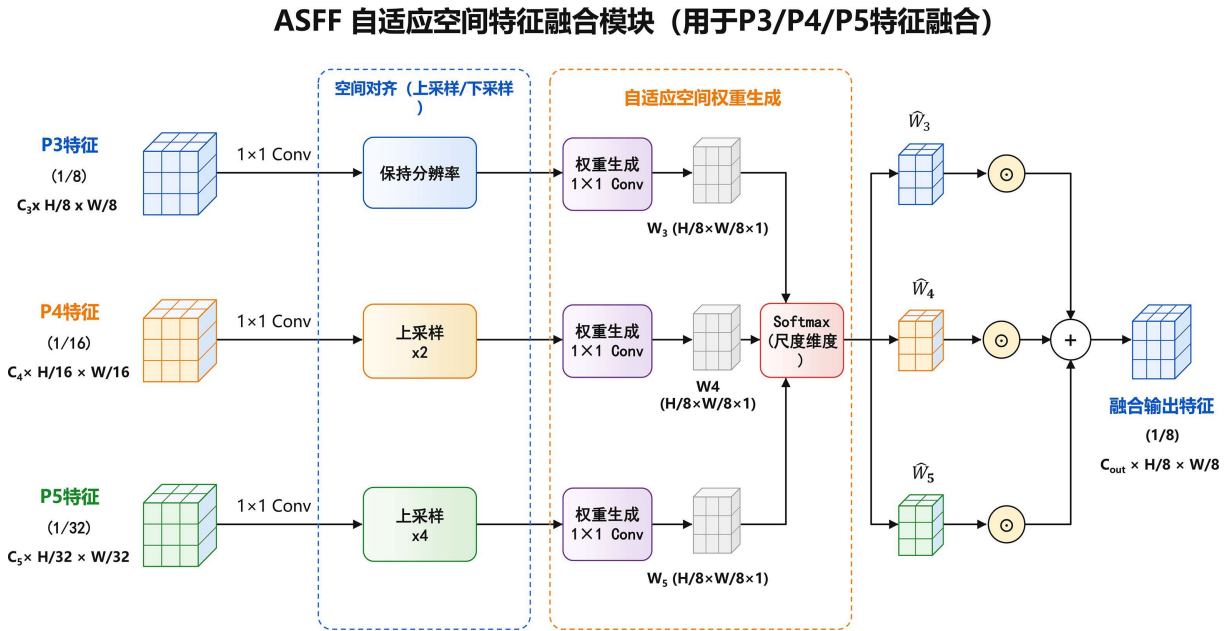


Figure 3. Structure diagram of the ASFF-P3 module

图 3. ASFF-P3 模块结构图

如图 3 所示, 设颈部网络输出的三尺度特征分别为 F_3 、 F_4 和 F_5 , 其中 F_4 为 CAA 增强后的 P4 特征, 其空间分辨率依次为 $H/8 \times W/8$ 、 $H/16 \times W/16$ 和 $H/32 \times W/32$ 。由于本文 ASFF 以 P3 为目标层级, 首先保持 P3 的空间尺寸不变, 并采用最近邻插值将 P4 和 P5 分别上采样 2 倍和 4 倍; 随后, 各分支通过 1×1 卷积完成通道映射, 使三路特征具有相同的空间尺寸和通道数。对齐后的特征可表示为:

$$\tilde{F}_3 = \phi_3(F_3), \quad \tilde{F}_4 = \phi_4(\mathcal{U}_2(F_4)), \quad \tilde{F}_5 = \phi_5(\mathcal{U}_4(F_5)) \quad (6)$$

其中, $\mathcal{U}_2(\cdot)$ 和 $\mathcal{U}_4(\cdot)$ 分别表示 2 倍与 4 倍最近邻上采样操作, $\phi_i(\cdot)$ 表示对应分支的 1×1 卷积通道映射。在完成尺度与通道对齐后, 模块分别利用 1×1 卷积将每一路特征压缩为单通道的未归一化空间权重图:

$$S_i = \psi_i(\tilde{F}_i), \quad i \in \{3, 4, 5\} \quad (7)$$

随后, 将三幅空间权重图沿尺度维度拼接, 并在尺度维度上执行 Softmax 归一化。对于空间位置 $(h,$

w), 第 i 个尺度对应的归一化权重为:

$$\alpha_i(h, w) = \frac{\exp(S_i(h, w))}{\sum_{j \in \{3, 4, 5\}} \exp(S_j(h, w))} \quad (8)$$

由此可保证同一空间位置处三个尺度权重之和为 1。归一化权重与对应的对齐特征进行逐元素相乘, 再对三路结果逐元素相加, 得到自适应融合特征:

$$F_{\text{fuse}} = \sum_{i \in \{3, 4, 5\}} \alpha_i \odot \tilde{F}_i \quad (9)$$

最后, 融合特征经过 3×3 卷积进一步整合局部信息, 形成用于小目标检测分支的 ASFF-P3 输出:

$$F_{\text{ASFF-P3}} = \varphi_{3 \times 3}(F_{\text{fuse}}) \quad (10)$$

式中, $\psi_i(\cdot)$ 表示第 i 个尺度的空间权重生成卷积, α_i 表示经过 Softmax 归一化后的空间权重, $\odot$ 表示逐元素相乘, $\varphi_{3 \times 3}(\cdot)$ 表示输出端的 3×3 卷积。该结构在保留 P3 高分辨率细节的同时, 自适应引入 P4 与 P5 的高层语义信息, 可缓解固定融合方式中不同层级特征相互干扰的问题, 从而增强模型对细小裂缝、坑洞边缘和小尺度修补区域的表达能力。由于本文仅使用 ASFF 生成 P3 检测特征, 中尺度分支继续采用 CAA-P4, 大尺度分支采用 P5, 因此能够在提升小目标检出能力的同时控制额外计算开销。

3.3. NWD-CIoU 边界框回归损失

道路裂缝与小尺度坑洞的标注框通常面积较小、长宽比较大, 预测框发生少量像素偏移时, IoU 值便可能出现明显下降, 从而导致回归损失波动。CIoU 虽然综合考虑了预测框与真实框的重叠面积、中心点距离和长宽比一致性, 但其本质仍依赖 IoU 度量, 对细长目标和小目标的位置偏差较为敏感。为提高边界框回归的稳定性, 本文在 CIoU 的基础上引入归一化 Wasserstein 距离(Normalized Wasserstein Distance, NWD), 构建 NWD-CIoU 组合边界框回归损失。

设预测框 B_p 和真实框 B_g 均由中心坐标、宽度和高度四个参数描述, 其中 c_x 、 c_y 表示边界框中心坐标, w 和 h 分别表示边界框宽度与高度。NWD 首先将边界框建模为二维高斯分布, 其均值向量反映目标中心位置, 协方差矩阵反映目标尺度信息:

$$\mathcal{N}_b = \mathcal{N}(\mu_b, \Sigma_b), \quad \mu_b = [c_x, c_y]^T, \quad \Sigma_b = \text{diag}\left(\frac{w^2}{4}, \frac{h^2}{4}\right) \quad (11)$$

由于道路缺陷检测采用水平边界框, 其协方差矩阵为对角矩阵, 因此预测框与真实框对应高斯分布之间的二阶 Wasserstein 距离可化简为中心坐标差与宽高差的联合度量:

$$W_2^2(\mathcal{N}_p, \mathcal{N}_g) = (c_{xp} - c_{xg})^2 + (c_{yp} - c_{yg})^2 + \left(\frac{w_p - w_g}{2}\right)^2 + \left(\frac{h_p - h_g}{2}\right)^2 \quad (12)$$

为消除目标绝对尺度对距离数值的影响, 将 Wasserstein 距离通过指数函数映射到 $(0, 1]$ 区间, 得到 NWD 相似度:

$$\text{NWD}(B_p, B_g) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_p, \mathcal{N}_g)}}{C}\right) \quad (13)$$

其中, C 为归一化常数。当预测框与真实框的位置和尺度越接近时, Wasserstein 距离越小, NWD 越接近 1。与直接使用 IoU 不同, 即使两个小目标框仅发生轻微偏移或暂时不存在重叠, NWD 仍能根据中心位置和尺度差异提供连续、稳定的优化信号。

为保留 IoU 类损失对边界框整体几何关系的约束, 本文同时采用 CIoU。设 $\rho(\cdot)$ 表示预测框与真实框中心点之间的欧氏距离, c 表示两框最小外接矩形的对角线长度, 则 CIoU 可表示为:

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(b_p, b_g)}{c^2} - \alpha v \quad (14)$$

其中, v 用于衡量预测框与真实框长宽比的一致性, α 为平衡系数, 其定义如下:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_g}{h_g} - \arctan \frac{w_p}{h_p} \right)^2, \quad \alpha = \frac{v}{1 - \text{IoU} + v} \quad (15)$$

在此基础上, 将 CIoU 损失与 NWD 损失按照 0.7 和 0.3 的权重进行组合, 得到本文采用的 NWD-CIoU 边界框回归损失:

$$L_{\text{box}} = 0.7(1 - \text{CIoU}) + 0.3(1 - \text{NWD}) \quad (16)$$

式中, $1 - \text{CIoU}$ 用于约束预测框的重叠程度、中心位置和长宽比, $1 - \text{NWD}$ 用于降低细长目标及小目标对像素级偏移的过度敏感性。二者结合后, 既保留了 CIoU 对整体几何一致性的约束能力, 又利用 NWD 提供更加平滑的尺度与位置度量, 有助于提高裂缝、小坑洞和小尺度修补区域的回归稳定性。

4. 实验与结果分析

4.1. 数据集与实验环境

实验采用 RDD2022 中国区域路面缺陷数据集集中的 Drone 与 MotorBike 两类采集场景数据。原始数据共 4878 张图像, 剔除无 XML 标注、空标注以及不包含目标类别的图像后, 最终保留 4373 张有效图像用于模型训练、验证与测试。数据集包含 5 类检测目标, 分别为裂缝(D00)、横向裂缝(D10)、网格裂缝(D20)、坑洞(D40)和道路维修(Repair)。为保证训练集、验证集和测试集中各类目标分布相对均衡, 采用多标签分层划分策略, 按照约 8:1:1 的比例划分数据集。实验环境为服务器 vGPU-32 GB, Ubuntu 22.04, Python 3.10, PyTorch 2.1.0, CUDA 12.1。主要训练参数如表 1 所示。

Table 1. Experimental parameter settings
表 1. 实验参数设置

参数	取值
epochs	200
batch	32
imgsz	640
workers	8
optimizer	SGD
lr0	0.001

4.2. 评价指标

为评价模型对道路缺陷的检出能力与定位精度, 本文采用召回率(Recall, R)、 mAP_{50} 和 mAP_{50-95} 作为主要指标。其中, TP 表示被正确识别的道路缺陷目标数, FN 表示未被模型检出的真实缺陷目标数, N_c 表示缺陷类别总数, AP_c^r 表示第 c 类目标在 IoU 阈值 τ 下的平均精度。召回率用于衡量模型对真实缺陷目标的覆盖程度, 其计算公式为:

$$R = \frac{TP}{TP + FN} \quad (17)$$

R 越高, 表示漏检的真实缺陷越少。对于道路巡检任务而言, 较高的召回率尤其有利于降低细小裂缝、低对比度裂缝以及远距离坑洞等目标被遗漏的风险。

mAP_{50} 表示当预测框与真实框的交并比阈值设为 0.50 时, 各缺陷类别平均精度的均值:

$$mAP_{50} = \frac{1}{N_c} \sum_{c=1}^{N_c} AP_c^{0.50} \quad (18)$$

该指标在匹配条件相对宽松的情况下综合反映分类与定位性能, 能够体现模型对纵向裂缝、横向裂缝、网状裂缝、坑洞及修补区域的整体检测能力。

mAP_{50-95} 则在 0.50 至 0.95 的十个 IoU 阈值上分别计算 mAP , 并对所有阈值和类别取平均:

$$mAP_{50-95} = \frac{1}{10N_c} \sum_{\tau \in \{0.50, 0.55, \dots, 0.95\}} \sum_{c=1}^{N_c} AP_c^{\tau} \quad (19)$$

与 mAP_{50} 相比, mAP_{50-95} 对预测框与真实框的重合程度要求更严格, 能够更充分地反映模型对裂缝边缘、坑洞轮廓和修补区域边界的精细定位能力。

4.3. 消融实验

为验证 CAA、ASFF 和 NWD-CIoU 三项改进的有效性, 本文以 YOLOv11n 为基线, 在相同数据集划分和训练配置下分别加入单个模块及其组合, 消融实验结果如表 2 所示。

Table 2. Ablation experiment results of CAN-YOLOv11n
表 2. CAN-YOLOv11n 消融实验结果

Baseline	CAA	ASFF	NWD-CIoU	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$
YOLOv11n	—	—	—	82.2%	86.2%	59.1%
	√	—	—	83.5%	87.5%	59.7%
	—	√	—	83.8%	87.1%	59.4%
	√	√	—	83.5%	87.6%	59.6%
	√	√	√	84.0%	88.6%	59.4%

由表 2 可知, CAA、ASFF 和 NWD-CIoU 三项改进均对模型检测性能产生了积极作用。单独引入 CAA 模块后, 模型的 R 由 82.2% 提升至 83.5%, mAP_{50} 由 86.2% 提升至 87.5%, mAP_{50-95} 由 59.1% 提升至 59.7%。CAA 通过水平和垂直方向的长程上下文建模, 强化了缺陷区域的连续特征响应, 有助于提升模型对纵向裂缝、横向裂缝和网状裂缝等方向性明显目标的识别能力, 同时增强复杂路面纹理背景下的缺陷定位效果。单独引入 ASFF 模块后, 模型的 R 达到 83.8%, 较基线模型提升 1.6 个百分点, mAP_{50} 和 mAP_{50-95} 分别达到 87.1% 和 59.4%。ASFF 能够根据不同空间位置自适应分配 P3、P4 和 P5 多尺度特征的融合权重, 在保留浅层高分辨率细节的同时引入深层语义信息, 使模型能够更加充分地表达不同尺度的道路缺陷。该结构尤其有利于细小裂缝、较小坑洞及修补区域边缘等目标的检测, 从而提高真实缺陷目标的检出率。当 CAA 与 ASFF 同时加入模型后, mAP_{50} 进一步提升至 87.6%, 说明上下文特征增强与多尺度自适应融合具有良好的互补作用。CAA 负责突出裂缝、坑洞和修补区域的有效响应, ASFF 则进一步整合不同层级的细节信息与语义信息, 使模型对复杂背景下多形态、多尺度道路缺陷具有更强的综合表征能力。

在 CAA 和 ASFF 的基础上进一步引入 NWD-CIoU 损失后, 最终构建的 CAN-YOLOv11n 模型在 R 和 mAP_{50} 指标上取得最优结果, 其 R 达到 84.0%, mAP_{50} 达到 88.6%。与基线模型相比, R 、 mAP_{50} 和 mAP_{50-95} 分别提升 1.8、2.4 和 0.3 个百分点。其中, R 和 mAP_{50} 的显著提升表明 NWD-CIoU 能够充分发挥 Ciou 几何约束与 NWD 小目标位置度量的互补优势, 增强细长裂缝和小尺度缺陷边界框的匹配稳定性。综合来看, CAA、ASFF 与 NWD-CIoU 分别从上下文特征增强、多尺度特征融合和边界框回归优化三个方面协同作用, 有效提升了模型对不同类型道路缺陷的整体检测能力。

4.4. 对比实验

为综合评估 CAN-YOLOv11n 的检测性能, 本文选取多个不同的 YOLO 模型进行对比, 并在相同数据集划分与评价指标下统计 R 、 mAP_{50} 和 mAP_{50-95} , 对比结果如表 3 所示。

Table 3. Comparison experiment results of different models
表 3. 不同模型对比实验结果

模型	$R/\%$	$mAP_{50}/\%$	$mAP_{50-95}/\%$
YOLOv5nu	83.7%	85.9%	57.6%
YOLOv6n	71.6%	81.0%	51.0%
YOLOv9t	80.8%	87.1%	59.5%
YOLOv11n	82.2%	86.2%	59.1%
YOLOv12n	80.0%	85.1%	57.5%
CAN-YOLOv11n	84.0%	88.6%	59.4%

CAN-YOLOv11n 在不同 YOLO 模型的对比中表现出较好的综合检测性能。其召回率(R)达到 84.0%, 为所有对比模型中的最高值, 分别较 YOLOv5nu、YOLOv6n、YOLOv9t、YOLOv11n 和 YOLOv12n 提升 0.3、12.4、3.2、1.8 和 4.0 个百分点。较高的召回率表明, CAN-YOLOv11n 能够检出更多真实道路缺陷目标, 有助于减少复杂路面环境下细小裂缝、坑洞和修补区域的漏检。在整体检测精度方面, CAN-YOLOv11n 的 mAP_{50} 达到 88.6%, 同样为表中最优结果。与 YOLOv5nu、YOLOv6n、YOLOv9t、YOLOv11n 和 YOLOv12n 相比, mAP_{50} 分别提升 2.7、7.6、1.5、2.4 和 3.5 个百分点, 说明所提模型在 IoU 阈值为 0.50 时具有更强的目标识别和定位能力。CAA 模块通过水平与垂直方向的上下文建模强化裂缝区域的连续响应, ASFF 模块通过自适应融合不同尺度特征提升小尺度缺陷与不规则目标的表达能力, NWD-CIoU 则进一步增强小目标边界框的位置匹配稳定性, 三项改进共同提升了模型对多类型道路缺陷的检测效果。在更严格的多阈值评价条件下, CAN-YOLOv11n 的 mAP_{50-95} 达到 59.4%, 高于 YOLOv5nu、YOLOv6n、YOLOv11n 和 YOLOv12n, 并保持了较高的定位性能。这表明模型不仅能够准确发现道路缺陷, 还能够不同 IoU 阈值下保持稳定的边界框回归能力, 对细长裂缝、小型坑洞以及边缘不规则的修补区域具有较好的适应性。本文提出的上下文特征增强、多尺度自适应融合和边界框回归优化策略能够形成有效协同, 提升复杂道路场景下不同尺度和不同形态缺陷的整体检测性能。

5. 总结

本文针对复杂路面背景下缺陷特征易受干扰、多尺度信息融合不足以及小目标定位不稳定等问题, 提出了 CAN-YOLOv11n 道路缺陷检测模型。该模型通过 CAA 增强方向性上下文特征, 利用 ASFF 融合不同尺度信息, 并结合 NWD-CIoU 提高小尺度和细长缺陷的边界框回归稳定性。实验结果表明, 模型的

R 和 mAP_{50} 分别达到 84.0%和 88.6%, 较 YOLOv11n 基线分别提升 1.8 和 2.4 个百分点, 说明所提方法能够有效减少缺陷漏检, 并提升复杂道路场景下的整体检测性能。目前模型在面对复杂道路环境的泛化能力还需进一步提升, 未来的研究将致力于拓展模型对更多复杂缺陷种类的检测, 尽量保证轻量化的同时提升模型的泛化性能。

基金项目

本研究得到了重庆市自然科学基金创新发展联合基金(项目编号: CSTB2025NSCQ-LZX0065)、重庆市教育委员会科学技术研究项目(项目编号: KJQN202201343、KJZD-K202401304)的资助。

参考文献

- [1] Abdelwahed, S.H., Sharobim, B.K., Wasfey, B. and Said, L.A. (2025) Advancements in Real-Time Road Damage Detection: A Comprehensive Survey of Methodologies and Datasets. *Journal of Real-Time Image Processing*, **22**, Article No. 137. <https://doi.org/10.1007/s11554-025-01683-1>
- [2] Huang, S., Chen, H., Yan, L., Zou, X., Li, B. and Bi, Y. (2025) A Review of the Progress in Machine Vision-Based Crack Detection and Identification Technology for Asphalt Pavements. *Digital Transportation and Safety*, **4**, 65-79. <https://doi.org/10.48130/dts-0025-0006>
- [3] Maeda, H., Sekimoto, Y., Seto, T., Kashiya, T. and Omata, H. (2018) Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. *Computer-Aided Civil and Infrastructure Engineering*, **33**, 1127-1141. <https://doi.org/10.1111/mice.12387>
- [4] Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D. and Sekimoto, Y. (2024) RDD2022: A Multi-National Image Dataset for Automatic Road Damage Detection. *Geoscience Data Journal*, **11**, 846-862. <https://doi.org/10.1002/gdj3.260>
- [5] Terven, J., Córdova-Esparza, D. and Romero-González, J. (2023) A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, **5**, 1680-1716. <https://doi.org/10.3390/make5040083>
- [6] Liu, Y., Liu, F., Liu, W. and Huang, Y. (2024) Pavement Distress Detection Using Street View Images Captured via Action Camera. *IEEE Transactions on Intelligent Transportation Systems*, **25**, 738-747. <https://doi.org/10.1109/tits.2023.3306578>
- [7] Xie, Y., Du, D., Wang, Z., Liu, Y. and Bi, M. (2025) YOLO-SFT: Road Damage Detection Algorithm Based on Feature Diffusion. *Journal of Transportation Engineering, Part B: Pavements*, **151**, Article 04025004. <https://doi.org/10.1061/jpeodx.pveng-1726>
- [8] Wang, W., Yu, X., Jing, B., Tang, Z., Zhang, W., Wang, S., et al. (2025) YOLO-RD: A Road Damage Detection Method for Effective Pavement Maintenance. *Sensors*, **25**, Article 1442. <https://doi.org/10.3390/s25051442>
- [9] Qu, Z., Wang, J.D. and Yin, X.H. (2025) A Directional Connectivity Feature Enhancement Network for Pavement Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*, **26**, 1039-1054. <https://doi.org/10.1109/tits.2024.3488656>
- [10] Guo, J.M., Markoni, H. and Lee, J.D. (2022) Barnet: Boundary Aware Refinement Network for Crack Detection. *IEEE Transactions on Intelligent Transportation Systems*, **23**, 7343-7358. <https://doi.org/10.1109/tits.2021.3069135>
- [11] Yang, L., Huang, H., Kong, S., Liu, Y. and Yu, H. (2023) PAF-Net: A Progressive and Adaptive Fusion Network for Pavement Crack Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, **24**, 12686-12700. <https://doi.org/10.1109/tits.2023.3287533>
- [12] Zhang, X., Wang, J., Jiang, D., Li, Y., Wang, X. and Zhang, H. (2025) Defects Detection in Screen-Printed Circuits Based on an Enhanced YOLOv8n Algorithm. *International Journal of Computational Intelligence Systems*, **18**, Article No. 101. <https://doi.org/10.1007/s44196-025-00815-6>
- [13] Qiao, Y., Guo, Y. and He, D. (2023) Cattle Body Detection Based on YOLOv5-Asff for Precision Livestock Farming. *Computers and Electronics in Agriculture*, **204**, Article 107579. <https://doi.org/10.1016/j.compag.2022.107579>
- [14] Xu, C., Wang, J., Yang, W., Yu, H., Yu, L. and Xia, G. (2022) Detecting Tiny Objects in Aerial Images: A Normalized Wasserstein Distance and a New Benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, **190**, 79-93. <https://doi.org/10.1016/j.isprsjprs.2022.06.002>
- [15] Tian, D., Han, Y., Wang, S., Chen, X. and Guan, T. (2022) Absolute Size IOU Loss for the Bounding Box Regression of the Object Detection. *Neurocomputing*, **500**, 1029-1040. <https://doi.org/10.1016/j.neucom.2022.06.018>