

文章编号: 1001-4632 (2026) 04-0105-14

引用格式: 秦勇, 彭根旺, 曹志威, 等. 基于 ADF-YOLO 模型的双模态轨道扣件缺陷检测方法[J]. 中国铁道科学, 2026, 47(4): 105-118.

Citation: Qin Yong, Peng Genwang, Cao Zhiwei, et al. Dual-modal rail fastener defect detection model based on ADF-YOLO[J]. China Railway Science, 2026, 47(4): 105-118.

基于 ADF-YOLO 模型的双模态轨道扣件缺陷检测方法

秦 勇¹, 彭根旺^{1,2}, 曹志威¹, 高 阳^{1,2}, 寇淋淋³, 王雅观⁴

(1. 北京交通大学 先进轨道交通自主运行全国重点实验室, 北京 100044;

2. 北京交通大学 交通运输学院, 北京 100044;

3. 北京地铁运营有限公司, 北京 100044;

4. 北京地铁运营有限公司技术创新研究院分公司, 北京 100044)

摘 要: 针对轨道扣件缺陷检测中单一可见光图像易受环境干扰、缺乏空间结构信息, 导致模型对小尺寸缺陷及扣件空间形变敏感性不足, 进而易产生误检和漏检的问题, 提出一种基于注意力机制与双模态融合 YOLO (Attention and Dual-Modal Fusion YOLO, ADF-YOLO) 模型的检测方法。首先, 构建 ADF-YOLO 模型, 通过双流骨干网络结构, 并行提取轨道扣件可见光图像的纹理特征和深度图像的空间结构特征; 其次, 为提高对小尺寸缺陷特征的捕获能力, 在 ADF-YOLO 模型骨干网络中引入注意力与多分支卷积协同模块, 扩大感受野, 强化细节特征表达; 最后, 为实现对异构特征的高效融合, 在 ADF-YOLO 模型颈部网络引入注意力引导的双模态特征融合模块, 充分挖掘双模态特征融合范式, 生成鲁棒性更高且对缺陷形态更敏感的融合特征。结果表明: 在扣件缺陷多模态数据集上, 基于 ADF-YOLO 模型的检测方法, 在交并比阈值为 0.5 和 0.5~0.95 时平均精度均值分别达 90.9% 和 70.5%; 与多种单模态和多模态目标检测模型相比, 基于 ADF-YOLO 模型的检测方法在检测精度上具有显著优势。提出的 ADF-YOLO 模型可为轨道扣件缺陷巡检任务提供可靠的技术支撑。

关键词: 铁路安全; 轨道扣件; 缺陷检测; 多模态融合; 注意力机制; YOLO

中图分类号: U216.3 **文献标识码:** A

doi: 10.3969/j.issn.1001-4632.2026.04.08

铁路是我国综合交通运输体系的骨干力量, 保障其运输安全是铁路发展的重中之重^[1]。扣件是轨道的关键部件, 起固定钢轨的作用, 其状态直接影响列车运行安全。在铁路系统中, 由于受列车荷载、材料老化等因素影响, 组成扣件的弹条、螺帽等零部件将出现松动、丢失等缺陷, 影响列车运行安全。因此, 定期检查和处理扣件缺陷, 对保障行车安全, 推动铁路发展具有重要意义。传统的扣件缺陷检测方法以人工巡检为主, 虽细致全面, 但存在天窗期工作强度大、效率低的缺陷。基于可见光

图像和深度学习的智能检测算法具有直观可视、部署灵活等优势, 已有众多国内外学者在该领域开展研究。

根据模型结构的不同, 当前研究方法可以分为两阶段方法和单阶段方法。相关学者通过对快速区域卷积神经网络 (Faster Region-Convolutional Neural Network, Faster R-CNN)^[2] 和掩码区域卷积神经网络 (Mask Region-Convolutional Neural Network, Mask R-CNN)^[3] 等两阶段模型进行针对性改进, 实现了对弹条形变、扣件缺失和松动等

收稿日期: 2025-08-14; 修订日期: 2026-07-28

基金项目: 先进轨道交通自主运行全国重点实验室自主课题 (RAO2025ZR001); 北京市自然科学基金资助项目 (L241078); 国家资助博士后研究人员计划项目 (GZC20251118)

第一作者: 秦 勇 (1971—), 男, 江苏徐州人, 教授, 博士研究生导师。E-mail: yqin@bjtu.edu.cn

通讯作者: 曹志威 (1993—), 男, 山东菏泽人, 博士。E-mail: zhiwei@bjtu.edu.cn

缺陷的检测^[4-6]。你只看一次 (You Only Look Once, YOLO) 视觉模型作为典型的单阶段模型, 被广泛应用在扣件缺陷检测上。针对不同扣件巡检场景和需求, 部分学者通过添加注意力机制^[7]、改进损失函数^[8]、轻量化模型框架^[9]等操作对 YOLO 系列算法进行了针对性改进。此外, 随着 Transformer 模型在视觉领域的广泛应用, 有学者将 Transformer 模型与卷积神经网络 (Convolutional Neural Networks, CNN) 相结合, 进行扣件缺陷检测。He 等^[10]提出了一种铁路扣件缺陷检测网络, 其通过构建 1 个 CNN 块和 1 个多尺度的 Vision Transformer 块交替提取扣件的局部细节特征和全局语义特征, 以增强缺陷识别效果。基于可见光图像的检测方法已经发展得较为成熟, 但可见光图像缺乏空间结构信息, 污渍遮挡、缺陷尺寸过小可能引起漏检和误检的问题。

针对可见光图像缺乏空间结构信息的固有局限, 可见光图像与深度图像融合检测提供了一条有效途径。深度图像能够提供轨道部件的三维结构信息, 通过将深度图像与可见光图像融合, 能够实现 2 种图像之间的信息互补, 从而提高缺陷的检测效果。按照融合阶段的不同, 融合方式可以分为像素级融合、决策级融合和特征级融合^[11]。Gao 等^[12]提出了一种基于像素级融合的扣件异常检测方法, 将灰度图像和相应的深度图像融合后输入网络进行检测。Wang 等^[13]提出了一种无砟轨道表面缺陷检测网络, 该网络利用融合后的可见光图像和深度图像检测复杂的缺陷对象。像素级融合虽能保留原始数据的完整信息, 但可能造成信息冗余。Ge 等^[14]提出了一种基于决策级融合的铁路轨道异常检测方法, 此方法由监督预检测、半监督再检测和决策级融合 3 部分组成, 将前 2 次检测结果在决策级进行融合。陈仕涛^[15]提出了一种基于多传感器信息融合的轨道异物检测方法, 该方法对摄像头和激光雷达采集的数据分别进行检测, 最后进行决策级融合。决策级融合方法的灵活性较强, 但依赖人工设定的决策规则, 存在较高的判定代价。赵宏伟等^[16]提出了一种基于特征级融合的轨道缺陷检测方法, 通过构建轨道表面缺陷的双模态深度学习检测网络, 分别融合强度图像和深度图像的多尺度特征, 进行钢轨顶面擦伤检测。孙铁强等^[17]提出了一种基于 YOLO 的多模态钢轨表面缺陷检测算法, 通过双模态特征融合模块对不同尺度的深度信息和可见光信息进行融合, 实现了对擦伤等 4 类钢轨缺

陷的检测。合理的特征级融合能够在保留关键语义信息的同时提升网络的特征表征能力, 但是现有的特征级融合方法以特征图相乘、通道拼接等简单的融合方法为主, 可能会造成不同模态特征之间的相互抑制, 导致检测效果下降。

本文构建基于注意力机制与双模态融合的 YOLO (Attention and Dual-Modal Fusion YOLO, ADF-YOLO) 模型, 并基于该模型提出一种轨道扣件缺陷检测方法。ADF-YOLO 模型以双流骨干网络并行提取可见光与深度图像的多尺度特征, 并通过注意力与多分支卷积协同模块 (Attention and Multi-Branch Convolution Synergy Module, AMCS 模块) 与注意力引导的双模态特征融合模块 (Attention-Guided Dual-Modal Feature Fusion Module, A-GDFF 模块) 实现模态间互补信息的充分整合, 以有效提升对小尺寸缺陷及扣件空间形变缺陷的检测精度。

1 轨道扣件缺陷检测方法

1.1 整体架构

注意力机制与双模态融合的 YOLO 模型整体架构如图 1 所示, 该模型基于 YOLOv8n 模型改造。YOLOv8n 模型由骨干网络、颈部网络和头部网络组成。其中, 骨干网络由卷积-批归一化-激活模块 (Conv+BatchNorm+SiLU, CBS 模块)、高效特征提取模块 (C2f 模块) 和快速空间金字塔池化模块 (Spatial Pyramid Pooling-Fast, SPPF) 组成, 负责特征提取。颈部网络由特征金字塔 (Feature Pyramid Network, FPN) 和路径聚合网络 (Path Aggregation Network, PAN) 组成, 负责特征融合。头部网络则负责处理融合后的特征, 得到目标检测结果。

在 YOLOv8n 的基础上, ADF-YOLO 模型首先对其骨干网络进行扩展, 构建双流骨干网络 (可见光骨干网络和深度骨干网络) 结构并行提取可见光图像特征和深度图像特征。随后, 在骨干网络浅层引入 AMCS 模块, 旨在利用注意力机制和多分支卷积结构捕获长距离上下文信息, 扩大感受野, 以提高模型对小目标的检测效果。同时, 在颈部网络引入 A-GDFF 模块, 通过多阶段双模态融合与增强策略, 增强和融合来自不同骨干网络同一层级的单模态特征, 减少特征差异, 获得融合特征

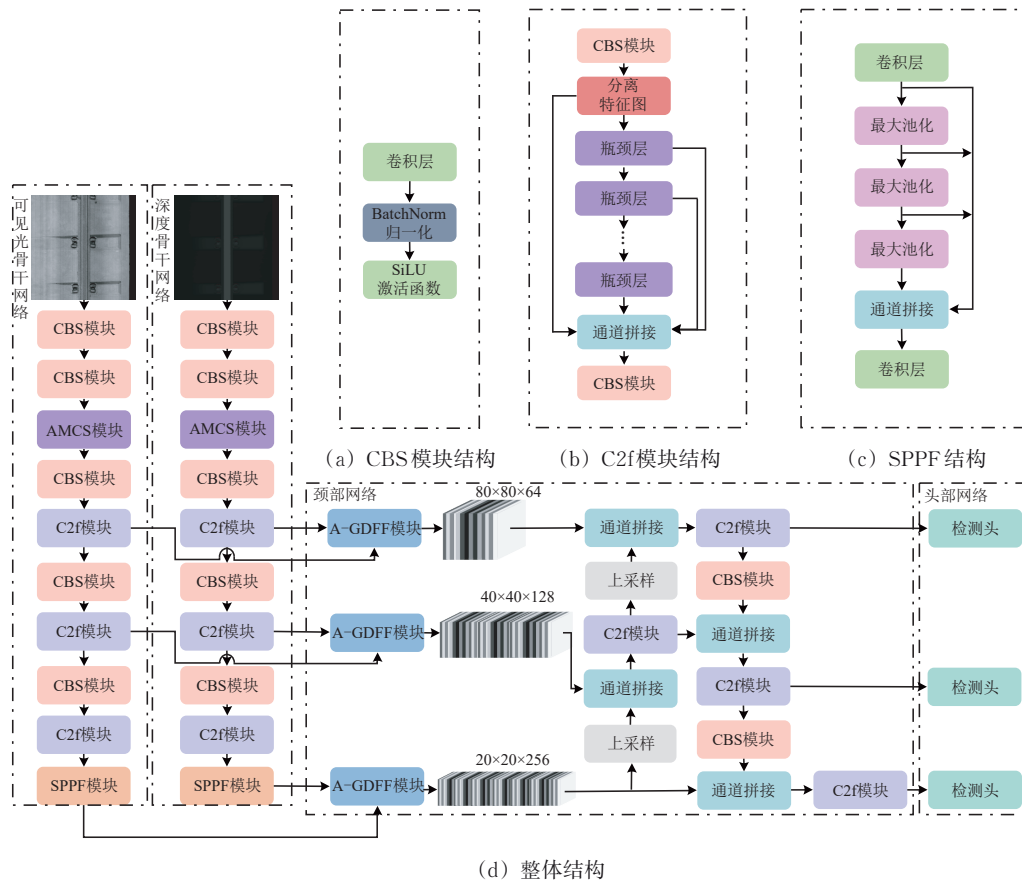


图1 ADF-YOLO模型架构

图。最后，将经过FPN和PAN结构处理的融合特征图输入到检测头中完成对扣件缺陷的检测。

1.2 AMCS模块

轨道巡检图像通常以俯视角度拍摄，轨枕、道床和钢轨等背景占据了图像中的大部分像素点，单个扣件在图像中的像素占比极小。在感受野较小的情况下，模型更聚焦于局部特征，导致全局建模能力较差，易发生误检。针对上述问题，受文献[18]和文献[19]的启发，在双流骨干网络的第3层引入AMCS模块，结合注意力机制和多分支卷积结构提取长距离上下文信息，扩大感受野，提高网络对小目标的检测效果。

AMCS模块的结构示意图如图2所示。首先，在原有C2f结构上以并行方式接入上下文锚点注意力机制(Context Anchor Attention, CAA)，与原输出融合。CAA通过平均池化提取图像的全局上下文信息，同时利用长条状深度卷积(Depthwise Convolution, DWConv)捕获扣件各部件与其周边更大范围的空间邻域之间的关联性，弥补网络上下文信息提取能力不足的问题。通过CAA生成的空间注意力图像对主分支的输出进行逐元素相乘加权

后再与原输出相加，引导网络聚焦于扣件及其周边区域，提高对小目标部件的识别效果。

其次，在AMCS模块的末端引入1层特征增强模块(Feature Enhancement Module, FEM)进行后处理。FEM层由4个分支组成，其中残差分支用于形成等效映射，保留扣件的整体特征；标准卷积分支用于提取扣件局部邻域内的基础纹理和边缘特征，从而捕捉螺帽、平垫圈等小尺寸部件的细微形变；2个空洞卷积分支则用于扩大感受野，使提取的特征图能够保留更多信息。通过多分支卷积结构能够提取多种具有判别性的语义信息，增加特征丰富性，提高网络对小目标的识别效果。

1.3 A-GDFF模块

不同模态的特征分布差异较大，现有的双模态特征融合方法容易忽略不同模态特征之间的关联，无法有效提取不同模态间的互补信息，容易造成信息冗余，影响模型性能。此外，与单模态特征相比，融合之后的双模态特征更为复杂，模型捕获小尺寸扣件缺陷特征的难度更大。针对上述问题，设计一种注意力引导的双模态特征融合模块A-GDFF模块。该模块采用多阶段双模态融合与增强策略，在

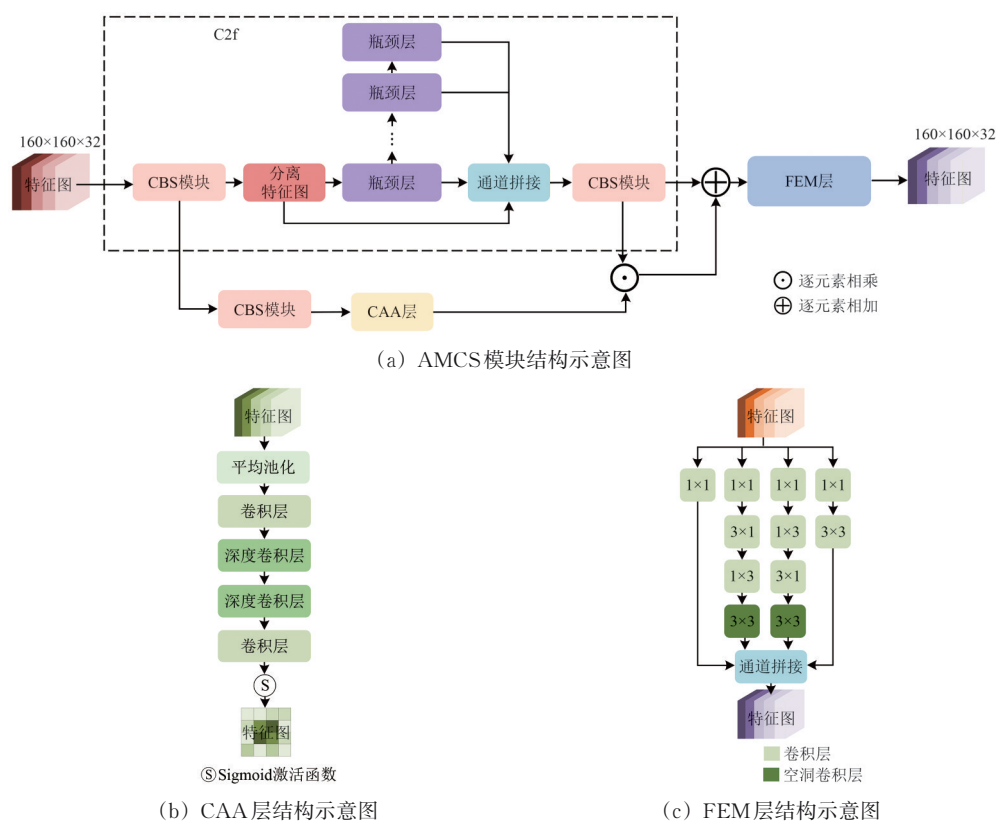


图 2 AMCS 模块及内部关键结构示意图

充分利用不同模态互补信息的同时减少特征差异，增强特征共享与对齐，以约束不同模态之间的预测一致性。A-GDFF 模块结构示意图如图 3 所示。

A-GDFF 模块基于金字塔结构提取融合特征图特征，生成混合特征权重，基于通道注意力机

制^[20]分别提取可见光图像和深度图像权重；随后将 2 种权重融合，分别对可见光特征和深度特征进行加权；最后将加权后的双模态特征相加得到融合特征图，具体融合步骤如下。

由于可见光图像提取的光谱纹理特征 F_{RGB} 与深度图像提取的三维几何特征 F_D 存在显著的语义差异，简单的特征加权往往会导致语义错位。因此，A-GDFF 模块首先将 F_{RGB} 和 F_D 在通道维度拼接，并通过 1 个 3×3 标准卷积层完成异构特征的初步语义融合与降维压缩，得到混合特征图 F_F ，如式 (1) 所示。这一步骤将原本处于 2 个独立特征空间的异构信息，强制映射到了 1 个统一的共享特征空间中。

$$F_F = C_{\text{onv } 3 \times 3}(C_{\text{oncat}}(F_{RGB}, F_D)) \quad (1)$$

式中： $C_{\text{onv } 3 \times 3}(\cdot)$ 为 3×3 标准卷积运算； $C_{\text{oncat}}(\cdot)$ 为通道拼接运算。

随后，设计深度卷积金字塔以提取包含双模态联合分布规律的混合特征权重，如式 (2) 所示。该金字塔由 3 个步长与卷积核大小相等，填充为 0 的深度可分离卷积 (Depthwise Separable Convolution, DSConv) 级联构成，尺度参数分别设置为 5×5 , 4×4 , 4×4 。在处理特征图时，初始的 5×5 较

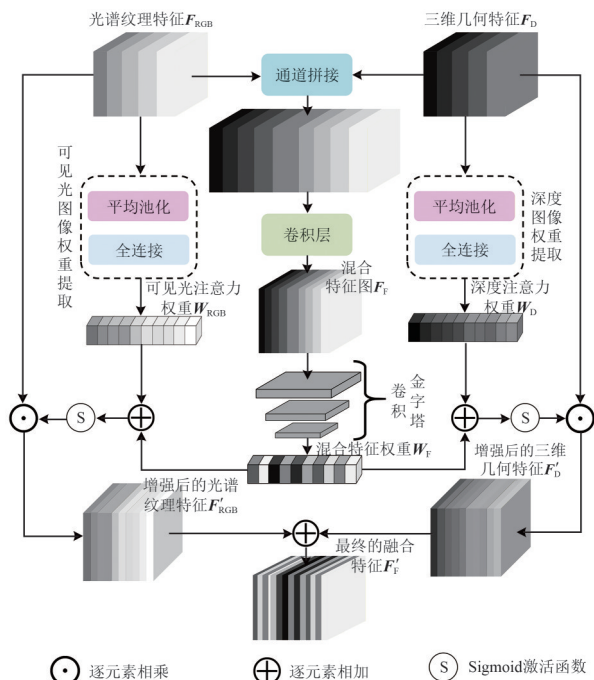


图 3 A-GDFF 模块结构示意图

大卷积核率先捕获扣件整体的宏观结构上下文，随后级联的2个 4×4 卷积则在逐层降维的过程中，精细聚合螺帽松动、平垫圈丢失等局部微小缺陷的微弱空间分布规律。该尺度组合在不干扰通道信息的前提下迅速降低特征图的空间维度，同时捕捉了可见光纹理信息和三维几何分布信息，完成对包含双模态联合分布规律的混合特征权重 W_F 的提取。 W_F 作为2种模态交互的桥梁，为后续的对齐提供全局一致性参考。

$$W_F = D_{SConv4 \times 4} \left(D_{SConv4 \times 4} \left(D_{SConv5 \times 5} (F_F) \right) \right) \quad (2)$$

式中： $D_{SConv_{n \times n}}(\cdot)$ 为卷积核大小为 $n \times n$ 的深度可分离卷积， n 为4或5。

同时，独立提取可见光和深度特征的自身注意力权重 W_{RGB} 和 W_D 。该环节旨在采用通道注意力机制显式地建模通道之间的相互依赖性，自适应地重新校准通道特征响应，从而提高特征图表示的质量。具体通过1个自适应平均池化层提取全局特征向量，之后将其输入1个2层的全连接网络中分别生成可见光权重 W_{RGB} 和深度权重 W_D ，分别为

$$W_{RGB} = F_C(A_{vgPool}(F_{RGB})) \quad (3)$$

$$W_D = F_C(A_{vgPool}(F_D)) \quad (4)$$

式中： $F_C(\cdot)$ 为全连接操作； $A_{vgPool}(\cdot)$ 为自适应平均池化。

之后，混合特征权重与单模态权重加权融合。为了保证可见光信息与三维几何信息在融合过程中的语义一致性，设计基于共享上下文信息的一致性约束机制。具体而言，在对单模态特征进行加权时，网络将该模态自身注意力权重与提取到的混合特征权重进行逐元素相加，并经激活函数激活后作为最终引导权重。在该过程中，网络无需设定固定的加权比例超参数，而是通过端到端的反向传播，自适应地调整各分支特征通道的响应幅值。利用最终的引导权重分别对可见光图像和深度图像进行逐元素相乘，得到 F'_{RGB} 和 F'_D ，完成对单模态特征的加权增强，即

$$F'_{RGB} = F_{RGB} S_{igmoid}(W_{RGB} + W_F) \quad (5)$$

$$F'_D = F_D S_{igmoid}(W_D + W_F) \quad (6)$$

式中： $S_{igmoid}(\cdot)$ 为Sigmoid激活函数。

这种全局联合语义空间的强制约束有效消除了单一模态在铁路巡检场景下产生的独立视觉偏差，促使异构特征在融合前向统一的物理共识靠拢，从而隐式且精准地实现了双模态特征的深度对齐。最后，将经过一致性对齐的 F'_{RGB} 和 F'_D 相加，输出高

质量的融合特征图 F'_F ，即

$$F'_F = F'_{RGB} + F'_D \quad (7)$$

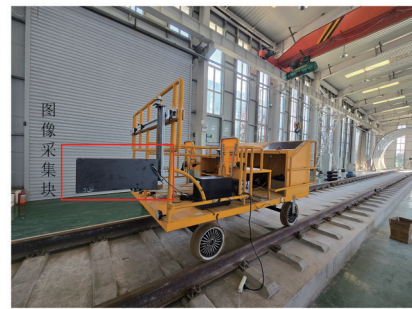
2 试验验证

2.1 扣件缺陷多模态数据集

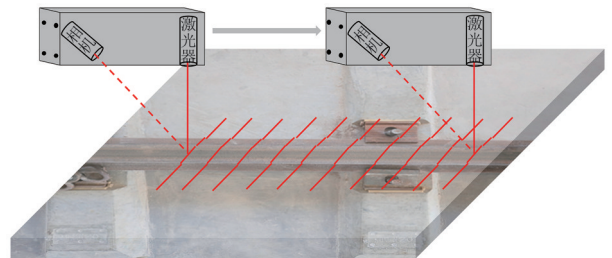
为全面评估ADF-YOLO模型在实际巡检场景中的检测性能，构建扣件缺陷多模态数据集，数据采集场景及原理如图4所示。数据集中的图像通过搭载了图像采集模块的轨检小车采集于1条全长60 m的真实比例高速铁路无砟轨道试验线上。试验线采用标准的 $60 \text{ kg} \cdot \text{m}^{-1}$ 钢轨，轨道扣件采用广泛应用于高速铁路的WJ-8B型扣件。数据采集时天气为晴天，在室内光照度为500 lx的正午强光条件下共采集2 200对图像，光照度为50 lx的傍晚暗光条件下共采集905对图像。



(a) 数据采集场景



(b) 搭载图像采集模块的轨检小车



(c) 图像采集模块工作原理

图4 图像数据采集现场及采集原理

数据采集时轨检小车上配备了由3D结构光相机、线性激光器等设备组成的图像采集模块。激光器发出的激光线垂直照射到轨道表面,通过3D结构光相机拍摄该激光线得到轨道部件的激光断面图像;随后,激光器和相机随车移动等间距采集的轨道断面图像,获取轨道的灰度信息和深度信息。最后对采集的轨道断面图像进行拼接,分别生成反映二维特征的灰度图和体现三维特征的深度图。

数据集共3 150对(2 048×2 048)像素的图像,每张图像通常包含4个扣件、2个轨枕、1条钢轨和1块轨道板,数据集中的图像示例如图5所示。按8:1:1的比例将数据集划分为训练集、验证集和测试集。

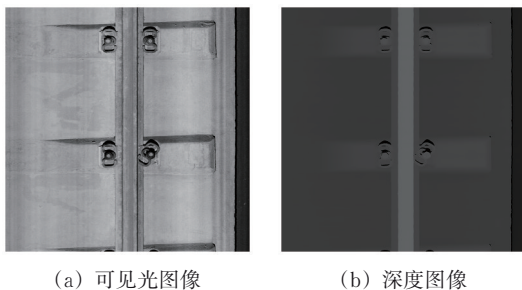


图5 数据集中的某组图像示例

数据集图像中各轨道部件的像素占比详见表1。由表1可知:单个扣件在图像中的像素占比仅为1.2%,发生在扣件局部的螺帽或平垫圈缺陷,其像素占比仅为0.13%。因此,将螺帽丢失、螺帽松动和平垫圈丢失这3个数据集中尺寸最小且表观变化最微弱的缺陷定义为小目标。

表1 数据集图像中各轨道部件的像素占比 %

轨道部件	像素占比
轨道板	72.4
钢轨	12.7
轨枕	10.1
扣件	4.8

数据集中各类样本的整体分布情况见表2。从表2可以看出:数据集中正常扣件的数量显著高于其余12类缺陷样本,缺陷样本的数量分布相对均衡。该类别不平衡现象源于轨道巡检的实际场景特征:正常扣件在轨道结构中占主导地位,缺陷发生频率相对较低,且单幅图像中所有扣件同时出现缺陷的概率极低。

2.2 试验条件

试验系统采用Ubuntu 20.04, GPU采用NVIDIA GeForce RTX 4090,试验环境的Python

表2 数据集中各类别样本的整体分布情况 个

样本类别	样本数量
扣件(fastener)	10 772
工具(tool)	896
异物(foreign matter)	839
零部件(component)	559
弹条丢失(elastic strip miss)	310
弹条松动(elastic strip loose)	525
螺帽丢失(nut miss)	201
螺帽松动(nut loose)	553
平垫圈丢失(plain washer miss)	310
垫板丢失(plate miss)	263
垫板松动(plate loose)	294
挡板丢失(baffle miss)	302
挡板松动(baffle loose)	342

版本为3.8.20, PyTorch版本为2.4.1, CUDA版本为11.3。训练时,输入图像统一缩放至(640×640)像素,采用SGD优化器,批处理大小设为32,初始学习率为0.001,共训练300轮。针对样本类别不平衡问题,训练过程中采用马赛克(Mosaic)数据增强策略,通过随机拼接扩充少数类缺陷的样本多样性;同时,采用YOLO框架的解耦头设计与动态样本分配机制,引导模型优先聚焦稀有难例,有效提升稀有类别的特征学习效率。

2.3 评价指标

采用精确率 P (Precision),召回率 R (Recall),平均精度均值 m_{AP} (Mean Average Precision),小目标类别的平均精度 A_{PS} (Average Precision to Small)评价模型性能。 P 为检测结果中正确检测的比例, R 为所有目标中被成功检测的比例, m_{AP} 为平均精度 A_P (Average Precision)的平均值,是评价模型综合性能的关键指标,其中 A_P 为模型识别某一类别目标的平均精度。一般采用 $m_{AP@0.5}$ 和 $m_{AP@0.5:0.95}$ 展示 m_{AP} 的大小,二者分别表示当预测框与真实框的交并比(Intersection over Union, IoU)为0.5时 m_{AP} 的大小,以及当IoU从0.5到0.95以0.05为步长变化时,各IoU值下的 m_{AP} 的平均值。 A_{PS} 指小目标类别的平均精度,计算方式为提取螺帽丢失、螺帽松动和平垫圈丢失3类小目标缺陷在IoU=0.5:0.95阈值下的单类平均精度,计算其算术平均值。各评价指标的计算式分别为

$$P = \frac{T_p}{T_p + F_p} \quad (8)$$

$$R = \frac{T_p}{T_p + F_n} \quad (9)$$

$$A_p=\int_0^1P(R)dR \tag{10}$$

$$m_{AP}=\frac{1}{k}\sum_{i=1}^kA_{P_i} \tag{11}$$

$$A_{PS}=\frac{1}{3}(A_{P_{螺帽丢失@0.5:0.95}}+A_{P_{螺帽松动@0.5:0.95}}+A_{P_{平垫圈丢失@0.5:0.95}}) \tag{12}$$

式中： T_p 表示真阳性； F_p 表示假阳性； T_N 表示真阴性； F_N 表示假阴性； k 表示所检测样本类别。

2.4 对比试验

2.4.1 定量对比

将 ADF-YOLO 模型分别与部分最新的 YOLO 模型^[21]、基于 YOLO 的双模态模型^[22]以及多模态目标检测领域代表性模型迭代式交叉注意力引导特征融合（Iterative Cross-Attention Guided Feature Fusion, ICAFusion）模型^[23]和超轻量级双模态自适应融合网络（Super-Lightweight Bimodal Adaptive Fusion Network, SLBAF-Net）^[24]进行对比。试验时，单模态模型的输入仅为可见光图像，双模态模型输入为可见光图像和深度图像（RGB and Depth, RGBD）。ADF-YOLO 模型与对比模型在验证集上的对比结果见表 3。由表 3 可知：与对比模型相比，本模型的 $m_{AP@0.5:0.95}$ 指标均达到最

表 3 ADF-YOLO 模型与对比模型在验证集上的对比结果

	%				
模型	P	R	$m_{AP@0.5}$	$m_{AP@0.5:0.95}$	A_{PS}
YOLOv8n(基准)	83.6	87.5	90.2	65.8	42.1
YOLOv10n	74.6	79.4	83.6	60.5	39.7
YOLOv11n	80.3	84.6	85.4	61.9	31.6
YOLOv12n	83.8	84.8	88.4	65.1	41.5
YOLOv8n-RGBD	85.6	83.8	84.1	64.1	30.4
YOLOv10n-RGBD	74.3	77.3	81.0	60.2	33.9
YOLOXn-RGBD	86.4	90.3	90.7	68.0	50.1
YOLOv12n-RGBD	87.8	83.4	88.9	67.0	44.2
ICAFusion	75.1	89.1	87.7	60.9	38.4
SLBAF-Net	71.3	76.9	78.8	57.6	23.9
ADF-YOLO	87.5	90.3	90.9	70.5	52.7

优，比基准模型 YOLOv8n 提升了 4.7%；ADF-YOLO 模型的 A_{PS} 相较于基准模型 YOLOv8n 实现了 10.6% 的提升，明显高于全类别的平均提升水平，这得益于 AMCS 模块有效扩展了感受野并成功捕获了高频细节，以及 A-GDFF 模块有效融合了三维几何特征，极大缓解了微小缺陷在二维高分辨率背景中易被淹没的语义退化问题，展现出卓越的微弱特征表征能力。

在验证集上对比 ADF-YOLO 模型与单模态 YOLO 模型在 13 类样本上的 $m_{AP@0.5:0.95}$ 指标，结果见表 4。

表 4 ADF-YOLO 模型与单模态 YOLO 模型在验证集上各样本的 $m_{AP@0.5:0.95}$ 指标对比

模型	不同缺陷样本的 $m_{AP@0.5:0.95}$												平均值
	FA	TO	FM	CO	EM	EL	NM	NL	PWM	PM	PL	BM	BL
YOLOv8n	95.6	84.7	78.4	74.6	69.3	87.3	38.5	55.1	32.7	54.5	67	66.9	50.1
YOLOv10n	95.5	83.4	76.9	72.4	65.9	80.6	40	42.6	36.6	43.9	49.2	62.2	36.6
YOLOv11n	95.2	82.7	79.1	72.4	69.6	86.4	36.3	39.6	18.9	50.8	62.9	62.7	48.3
YOLOv12n	95.5	84.9	78.5	73.6	69.1	86.6	45.8	49.2	29.6	51.4	68.5	67.1	46.8
ADF-YOLO	96.3	89	82.9	77.8	73.6	89.2	47.3	64.4	46.4	56	67.2	69.9	56

注：FA 表示扣件，TO 表示工具，FM 表示异物，CO 表示零件，EM 表示弹条丢失，EL 表示弹条松动，NM 表示螺帽丢失，NL 表示螺帽松动，PWM 表示平垫圈丢失，PM 表示垫板丢失，PL 表示垫板松动，BM 表示挡板丢失，BL 表示挡板松动，AV 表示平均值。

从表 4 可以看出以下结论。

(1) 与基准模型 YOLOv8n 相比，ADF-YOLO 模型在各个类别的检测精度上均有不同程度的提升，平均提升幅度达 4.7%。从缺陷的物理尺寸与形态特征来看，这种提升具有明显的类别差异性。首先，对于平垫圈丢失、螺帽松动这类小目标缺陷，由于其在二维平面上视觉特征较弱，单模态模型极易因缺乏三维信息而发生漏检。ADF-YOLO 模型引入深度图像，为这些几何状态变化提供了空间梯度信息。同时，AMCS 模块的引入有效扩大了感受野，增强了对微小目标空间特征的捕获能力，从而使本模型相较于次优模型分别获得了高达

9.8% 和 9.3% 的提升。

(2) 对于工具、异物、零件和弹条丢失等具有明显三维凸出属性的不规则缺陷，A-GDFF 模块的双模态融合机制能够同时捕获它们在可见光中的异常纹理和深度图像中的高度突变，其提升幅度也稳定在 3.2%~4.1% 之间。对于正常的扣件，由于其在 2D 图像中特征已足够明显，单模态检测已趋于饱和，因此提升相对平缓。整体而言，ADF-YOLO 模型对于突破单模态算法在复杂背景和小目标缺陷检测上的瓶颈作用显著。

在验证集上对比 ADF-YOLO 模型与其他双模态模型在 13 类样本上的 $m_{AP@0.5:0.95}$ 指标，结果见表 5。

表5 ADF-YOLO模型与双模态对比模型在验证集上各样本的 $m_{AP@0.5:0.95}$ 指标对比

%

模型	不同缺陷样本的 $m_{AP@0.5:0.95}$													
	FA	TO	FM	CO	EM	EL	NM	NL	PWM	PM	PL	BM	BL	平均值
YOLOv8n-RGBD	96.1	86.4	79.2	75.6	69.6	87.3	23.3	54.6	13.2	53.3	66.9	69.5	58.3	64.1
YOLOv10n-RGBD	96.1	84.1	78.2	72.5	67.4	85.7	30.3	38.4	32.9	34.9	54.0	66.1	42.6	60.2
YOLOXn-RGBD	95.6	81.4	76.0	69.4	71.9	87.3	45.2	61.8	43.4	52.3	63.5	65.2	53.7	66.7
YOLOv12n-RGBD	96.1	86.7	79.9	75.3	71.1	87.1	40.2	58.1	34.2	46.5	69.0	68.3	58.6	67.0
ICAFusion	95.1	81.0	77.2	69.3	64.5	81.0	30.3	49.3	35.5	39.0	61.6	67.7	40.0	60.9
SLBAF-Net	94.5	79.6	70.7	72.5	61.0	80.1	8.9	40.5	22.2	46.5	66.0	62.0	44.2	57.7
ADF-YOLO	96.3	89.0	82.9	77.8	73.6	89.2	47.3	64.4	46.4	56.0	67.2	69.9	56.0	70.5

注：FA表示扣件，TO表示工具，FM表示异物，CO表示零件，EM表示弹条丢失，EL表示弹条松动，NM表示螺帽丢失，NL表示螺帽松动，PWM表示平垫圈丢失，PM表示垫板丢失，PL表示垫板松动，BM表示挡板丢失，BL表示挡板松动。

由表5可以看出以下结论。

(1) 相较于次优模型，ADF-YOLO模型在宏观指标上获得了3.5%的提升，这表明本模型在进行双模态特征融合时具有更优的架构设计。

(2) 针对零件、弹条丢失、垫板丢失和挡板丢失等包含复杂上下文信息的空间突变型缺陷，对比算法的特征融合方法容易导致可见光纹理与深度几何边缘产生语义错位。本模型通过A-GDFF模块促使异构特征在融合前向统一的物理状态靠拢，有效实现了可见光的高频细节与深度图的三维特征的精确对齐，使检测效果取得了进一步提升。

(3) 针对螺帽丢失、螺帽松动和平垫圈丢失这3类小目标缺陷，ICAFusion与SLBAF-Net出现了严重的性能衰减，而本模型在上述3类缺陷上的检测精度分别达到47.3%，64.4%和46.4%。得益于AMCS和A-GDFF模块的引入，本模型在应对轨道场景中占比极低且易受背景干扰的小目标缺陷时，其检测性能优于ICAFusion和SLBAF-Net等多模态目标检测领域的代表性方法。

(4) 针对垫板松动和挡板松动这类扁平状缺陷，其松动往往仅表现为二维平面上的轻微平移或错位，深度模式的梯度变化并不明显。A-GDFF模块强耦合的特征引导机制可能使深度特征对可见光特征产生了一定的语义稀释，导致相较于弱耦合的双模态对比模型略有衰退。

总体而言，ADF-YOLO模型在绝大多数扣件缺陷类别上均实现了跨模态特征的高效互补与增强，显著提升了轨道扣件缺陷检测的综合性能。

2.4.2 可视化对比

为了更加直观地展示本模型的优越性，在测试集上将ADF-YOLO模型和部分其他模型的检测结果进行可视化对比，结果如图6—图9所示。

从图6—图9可以看出：仅ADF-YOLO模型在

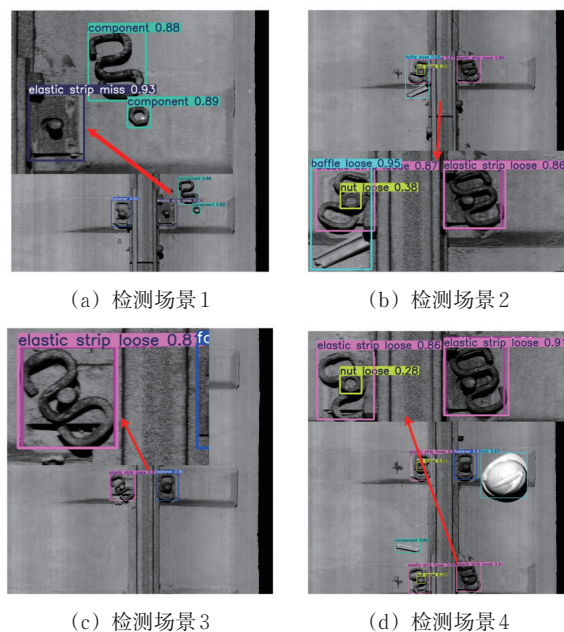


图6 YOLOv11n模型检测结果

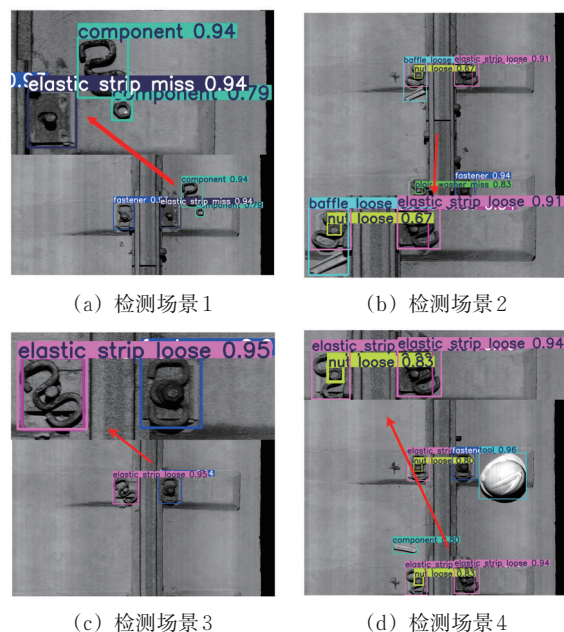


图7 YOLOv8-RGBD模型检测结果

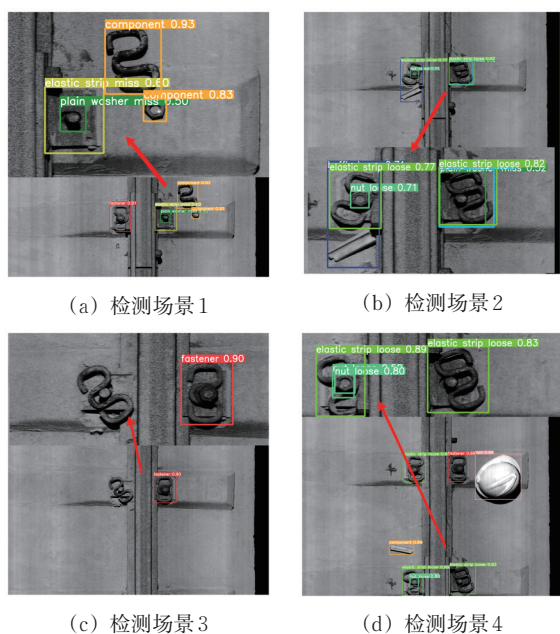


图8 SLBAF-Net模型检测结果

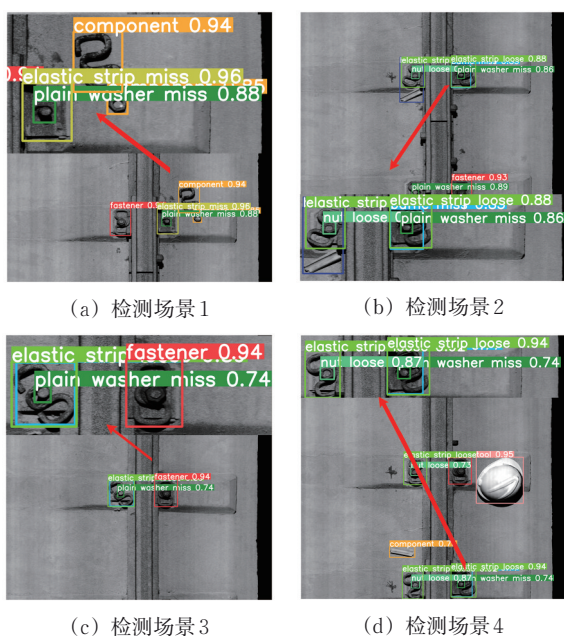


图9 ADF-YOLO模型检测结果

4组图像中实现了对螺帽松动、丢失和平垫圈丢失这3类小目标缺陷的全检出；相比之下，对比模型普遍存在漏检或误检现象，特别是在平垫圈丢失这一小目标检测任务中，对比模型的漏检问题尤为严重，说明本文提出的AMCS模块可以有效提升模型对小尺寸缺陷的检测效果；同时，ADF-YOLO模型以较高的置信度检测出了4组图像中的所有缺陷，证明了A-GDFF模块可以有效融合双模态特征，约束不同模态之间的预测一致性，提高模型检测精度。

为验证模型在实际巡检场景中的鲁棒性，基于测试集构建了涵盖大面积阴影、强光过曝及水雾遮挡3种典型工况的模拟数据集，可视化对比分析结果如图10—图13所示。

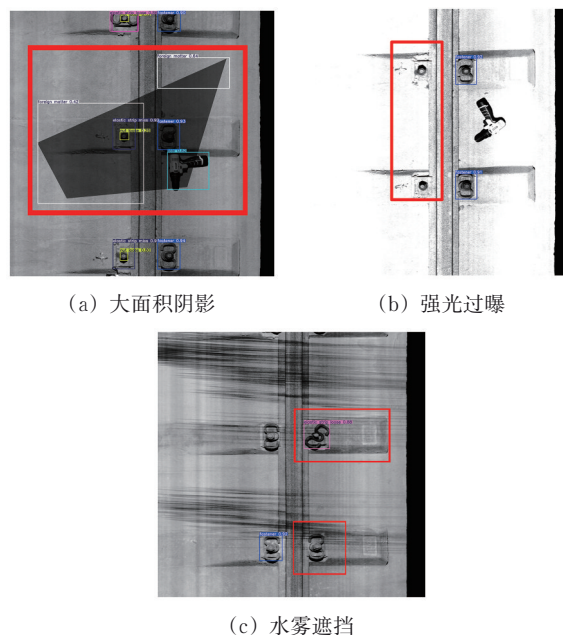


图10 YOLOv11n模型在3种典型工况下的检测结果

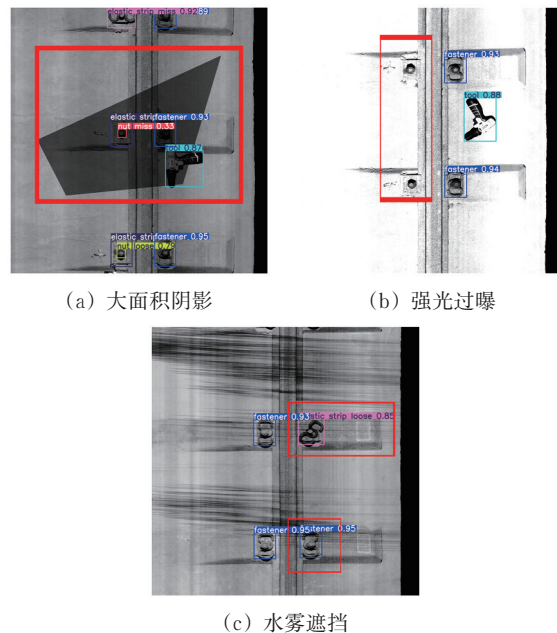


图11 YOLOXn-RGBD模型在3种典型工况下的检测结果

从图10—图13可以看出以下结论。

(1) 在大面积阴影干扰下，由于阴影边界强烈的光度梯度干扰，除YOLOXn-RGBD模型外，其余对比模型均将阴影边缘检测为异物或工具，而YOLOXn-RGBD模型则将阴影下的螺帽松动误检为丢失。

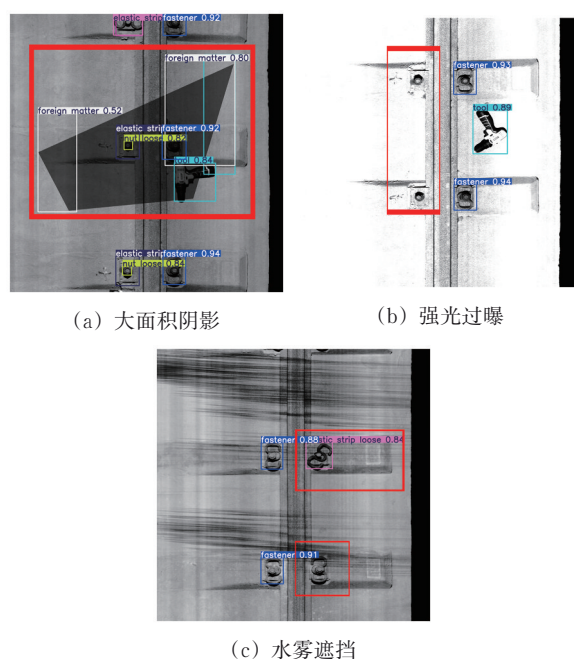


图12 YOLO12n-RGBD模型在3种典型工况下的检测结果

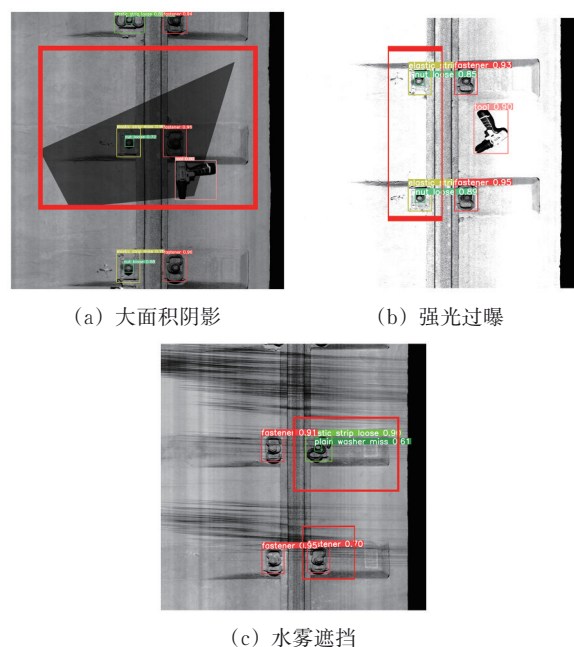


图13 ADF-YOLO模型在3种典型工况下的检测结果。

(2) 在强光过曝场景下,单模态模型因可见光纹理严重退化导致弹条丢失和螺帽松动2种缺陷均漏检;双模态对比模型虽引入了深度信息,但由于缺乏有效的跨模态对齐机制,易受强光区域边缘高频噪声影响,同样出现了漏检的情况。相比之下,ADF-YOLO模型的A-GDFF模块利用深度图不受光照干扰的三维几何先验补偿受损的可见光特征,使模型在极端光照干扰下依然保持了极高的定位精度与判别能力,准确检测到所有缺陷。

(3) 在水雾遮挡场景中,水雾形成了强烈的方向性条纹噪声与局部纹理畸变,掩盖了扣件真实的二维边缘,导致常规网络发生严重的特征提取混乱。所有对比模型均未检测到右上角扣件的平垫圈丢失缺陷,且部分模型同时出现了漏检扣件的情况。而ADF-YOLO模型通过AMCS模块扩大了感受野,跳出了局部条纹噪声的干扰;同时通过A-GDFF模块利用深度图几何先验作为强制约束,滤除了可见光分支的伪影噪声,成功检出所有缺陷。

尽管ADF-YOLO模型在绝大多数场景下展现出了优异的检测性能,但在面对极端的物理遮挡与环境伪装时,仍存在一定的局限性。为了客观评估本模型的适用边界,提取测试集中3个典型的失败案例进行分析,如图14所示。

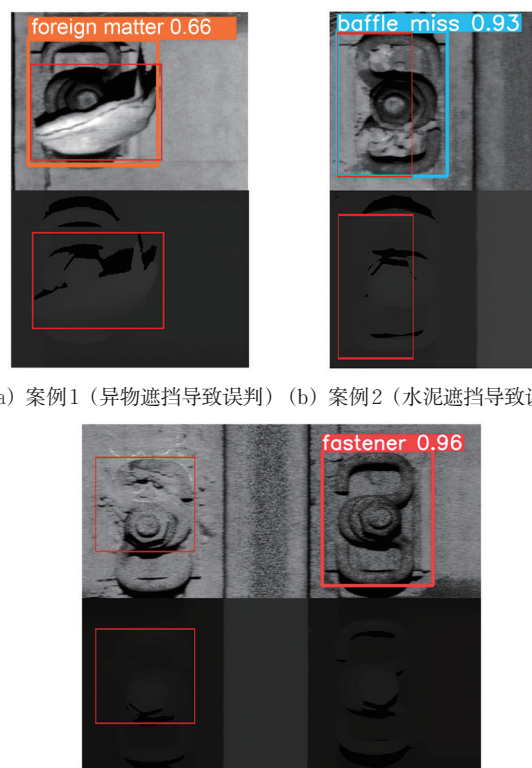


图14 典型失败案例示意图

从图14可以看出:第1个案例中,由于异物紧贴并覆盖扣件,模型将二者整体识别成了异物,在密集的空间重叠下,异物掩盖了扣件的纹理与三维几何特征,导致在特征提取过程中,两者的局部特征在同一感受野内发生了高度的空间耦合,网络倾向于利用视觉显著性更强的异物特征主导类别判定,从而导致复合目标被单一化归类的误判;第2个案例中,正常挡板因泥浆覆盖被误检为丢失;第

3 个案例中，左侧扣件因泥浆覆盖遮挡导致漏检，因为泥浆抹除了可见光图像的边缘与纹理，同时填平了扣件原本的三维空间结构，导致深度传感器无法捕捉其有效几何形态，在此极端工况下，可见光与深度信息同时发生语义欺骗与结构坍塌，最终引发误检与漏检。

2.5 消融试验

在进行消融试验以验证所提各模块的有效性时，基准融合方式的选择至关重要，常见的多模态特征融合方式包括特征图拼接、相加和相乘。以 YOLOv8n 为基准模型，对比上述 3 种融合方式的检测性能，结果见表 6。

表 6 常见多模态特征融合方式在验证集上的试验结果 %

融合方式	$m_{AP@0.5:0.95}$
特征图相乘	65.3
特征图相加	63.2
特征图拼接	64.3

表 7 验证集上的消融试验结果 %

模型	P	R	$m_{AP@0.5}$	$m_{AP@0.5:0.95}$
基准模型	82.6	82.6	87.7	65.3
基准模型+AMCS 模块	84.3	89.7	89.7	68.4
基准模型+A-GDFF 模块	86.1	88.7	89.6	69.2
基准模型+AMCS 模块+A-GDFF 模块(ADF-YOLO 模型)	87.5	90.3	90.9	70.5

从表 7 中可以得到如下结论。

(1) 在基准模型的双流骨干网络中引入了 AMCS 模块，对比基准模型 $m_{AP@0.5:0.95}$ 提升了 3.1%。

(2) 将基准模型的多模态融合方式由特征图相乘替换为 A-GDFF 模块，对比基准模型 $m_{AP@0.5:0.95}$ 提升了 3.9%。证明了与简单融合方法相比，A-GDFF 模块基于注意力引导和卷积金字塔的融合机制可以更有效地融合双模态特征，促进双模态信息交互，提高模型检测效果。

(3) 在基准模型中同时引入 AMCS 模块和 A-GDFF 模块，得到最终的 ADF-YOLO 模型。与基准模型相比，改进后模型的 $m_{AP@0.5:0.95}$ 提升了 5.2%，其他指标也有明显提升，证明本文提出的 2 个模块可以提高模型在扣件缺陷检测任务上的精度。

此外，为证明 AMCS 模块针对小目标的优化效果，对比了基准模型在引入 AMCS 模块前后，螺帽松动、螺帽丢失和平垫圈丢失这 3 类小目标缺陷的 $m_{AP@0.5:0.95}$ 指标，结果见表 8。由表 8 可知：引入 AMCS 模块后，模型针对 3 类缺陷的检测指标均有明显提升，证明了该模块可以有效提高模型对小目标的检测精度，进而提高模型整体的检测效果。

由表 6 可知：采用特征图相乘策略时，模型的 $m_{AP@0.5:0.95}$ 指标显著优于相加和拼接，这是因为轨道扣件属于刚体结构，其缺陷在可见光图像与深度图中具有严格的空间位置一致性；从特征图相乘算法机理上看，只有当可见光与深度特征在同一空间位置均产生高响应时，融合后的特征才会被激活；反之，若某一模态在背景区域产生了独有噪声，另一模态的低响应可有效抑制该干扰；相比之下，相加或拼接易将单模态噪声无条件传递至下游网络。因此，选用更为稳健的特征图相乘策略作为基准融合方式。

随后，为系统评估 AMCS 模块和 A-GDFF 模块对 ADF-YOLO 模型的具体贡献，在试验条件不变的前提下设计 4 组消融试验，分别验证各改进模块的有效性。试验时，模型的输入均为可见光图像和深度图像，基准模型采用多模态融合方式为特征图相乘的 YOLOv8n 模型，试验结果见表 7。

表 8 验证集上前 2 组试验针对 3 类小目标缺陷的 $m_{AP@0.5:0.95}$ 指标对比结果 %

模型	不同样本类别的 $m_{AP@0.5:0.95}$			平均值
	NM	NL	PWM	
基准模型	42.3	55.8	36.0	44.7
基准模型+AMCS 模块	47.5	59.0	38.2	48.2

注：NM 表示螺帽丢失，NL 表示螺帽松动，PWM 表示平垫圈丢失。

最后，为验证 CAA 模块和 FEM 模块对 AMCS 模块的贡献，以引入 A-GDFF 模块的 YOLOv8n 模型为基准模型，对 AMCS 模块进行消融试验，结果见表 9。

表 9 AMCS 模块在验证集上的消融试验结果 %

模型	P	R	$m_{AP@0.5}$	$m_{AP@0.5:0.95}$
基准模型	86.1	88.7	89.6	69.2
基准模型+FEM 层	86.9	89.6	90.4	70.0
基准模型+CAA 层	87.5	90.3	90.9	70.5

由表 9 可知：在添加 FEM 层后，模型在验证集上的 $m_{AP@0.5:0.95}$ 提高了 0.8%；在添加 FEM 层的基础上，以并行的方式添加 CAA 层，与 Baseline 相

比模型的 $m_{AP@0.5:0.95}$ 提高了1.3%。消融试验结果表明CAA层和FEM层的协同作用优化了模型的检测效果。

3 结论与展望

(1) 提出的基于ADF-YOLO模型的检测方法,通过双流骨干网络结构并行提取可见光特征和深度特征,创新性地引入注意力与多分支卷积协同模块以增强模型对小尺寸缺陷的检测效果,引入注意力引导的双模态特征融合模块以实现双模态特征的高效融合。

(2) 采用提出的检测方法对扣件缺陷尤其是对平垫圈丢失、螺帽松动等小尺寸缺陷的检测效果提升明显。在扣件缺陷多模态数据集上,ADF-YOLO模型在交并比阈值为0.5时平均精度均值达到了90.9%,在交并比阈值为0.5至0.95时平均

精度均值达到了70.5%,与当前主流的单模态和双模态模型相比均具有显著优势,为铁路系统的智能化巡检提供了一种可靠的技术方案,对提高铁路运维效率,保障行车安全具有重要意义。

(3) 然而,本研究数据集主要采集于实验室环境,实际线路场景中复杂的光线、天气和背景干扰以及多样化扣件类型可能对模型的性能带来挑战。针对上述局限,未来的研究将从数据域泛化、模型架构优化与在线持续学习3个维度持续推进。首先,构建涵盖典型气象场景、多种扣件型号的大规模、多模态真实线路数据集,通过丰富数据提升模型的鲁棒性;其次,引入多帧图像联合分析机制,弥补单帧图像在极端物理遮挡情况下的信息缺失;最后,引入增量学习机制,使模型能够在实际巡检任务中随现场巡检数据的积累进行持续自适应更新与优化。通过上述多维度的迭代优化,使模型更好地支撑全天候、全场景的轨道智能巡检工程应用。

参 考 文 献

- [1] Qin Yong, Cao Zhiwei, Sun Yongfu, et al. Research on active safety methodologies for intelligent railway systems[J]. Engineering, 2023, 27: 266-279.
- [2] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [3] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [4] 梁楠,张伟,刘洋龙,等.基于改进Faster R-CNN的高铁扣件弹条缺陷检测[J].太赫兹科学与电子信息学报,2024,22(11):1221-1227,1269.
Liang Nan, Zhang Wei, Liu Yanglong, et al. Fastener clips defect detection based on improved faster R-CNN in high-speed railway[J]. Journal of Terahertz Science and Electronic Information Technology, 2024, 22(11): 1221-1227, 1269.
- [5] 黄午祥.基于深度学习的铁轨表面缺陷检测技术研究[D].武汉:湖北工业大学,2023.
Huang Wuxiang. Research on rail surface defect detection technology based on deep learning[D]. Wuhan: Hubei University of Technology, 2023.
- [6] 许贵阳,李金洋,白堂博,等.基于改进Mask R-CNN的轨道扣件状态检测方法[J].中国铁道科学,2022,43(1):44-51.
Xu Guiyang, Li Jinyang, Bai Tangbo, et al. Detection method of track fastener state based on improved mask R-CNN[J]. China Railway Science, 2022, 43(1): 44-51.
- [7] 陈修忻.基于改进YOLOv5s算法的钢轨扣件状态检测方法[J].城市轨道交通研究,2024,27(增刊1):147-151.
Chen Xiuxin. Rail fastener status detection method based on improved YOLOv5s algorithm[J]. Urban Mass Transit, 2024, 27(S1): 147-151.
- [8] 吴送英,刘林芽,江家明,等.复杂背景下铁路扣件的改进YOLOv5s检测算法[J].中国铁道科学,2023,44(3):53-63.
Wu Songying, Liu Linya, Jiang Jiaming, et al. Detection algorithm of improved YOLOv5s of railway fastener under complex background[J]. China Railway Science, 2023, 44(3): 53-63.
- [9] 武福,蒋鹏民,李忠学,等.基于轻量化YOLOv8s的轨道扣件状态检测方法[J].北京交通大学学报,2024,48(5):59-68.
Wu Fu, Jiang Pengmin, Li Zhongxue, et al. Lightweight detection method for track fastener status based on improved YOLOv8s[J]. Journal of Beijing Jiaotong University, 2024, 48(5): 59-68.
- [10] He Jin, Wang Wei, Lv Fengmao, et al. Multi-scale CNN-transformer hybrid network for rail fastener defect detection[J].

- IEEE Transactions on Intelligent Transportation Systems, 2025, 26(6): 8894-8906.
- [11] Gómez-Chova L, Tuia D, Moser G, et al. Multimodal classification of remote sensing images: a review and future directions[J]. Proceedings of the IEEE, 2015, 103(9): 1560-1584.
- [12] Gao Yang, Cao Zhiwei, Qin Yong, et al. Railway fastener anomaly detection via multisensor fusion and self-driven loss reweighting[J]. IEEE Sensors Journal, 2024, 24(2): 1812-1825.
- [13] Wang Jingpeng, Song Kechen, Zhang Defu, et al. Collaborative learning attention network based on rgb image and depth image for surface defect inspection of no-service rail[J]. IEEE/ASME Transactions on Mechatronics, 2022, 27(6): 4874-4884.
- [14] Ge Xuanyu, Cao Zhiwei, Qin Yong, et al. An anomaly detection method for railway track using semisupervised learning and vision-lidar decision fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5023215.
- [15] 陈仕涛. 基于多传感器信息融合的轨道异物检测研究与实现[D]. 南京: 南京邮电大学, 2023.
Chen Shitao. Research and implementation of track abnormal object detection based on multi-sensor information Fusion[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.
- [16] 赵宏伟, 郑嘉俊, 赵鑫欣, 等. 基于双模态深度学习的钢轨表面缺陷检测方法[J]. 计算机工程与应用, 2023, 59(7): 285-293.
Zhao Hongwei, Zheng Jiajun, Zhao Xinxin, et al. Rail surface defect method based on bimodal-modal deep learning[J]. Computer Engineering and Applications, 2023, 59(7): 285-293.
- [17] 孙铁强, 魏光辉, 宋超, 等. 基于YOLO的多模态钢轨表面缺陷检测方法[J]. 电子测量技术, 2024, 47(21): 72-81.
Sun Tieqiang, Wei Guanghui, Song Chao, et al. Multi-modal rail surface defect detection method based on YOLO[J]. Electronic Measurement Technology, 2024, 47(21): 72-81.
- [18] Cai Xinhao, Lai Qiuxia, Wang Yuwei, et al. Poly Kernel Inception Network for Remote Sensing Detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-21, 2024, Seattle, WA, USA. New York: IEEE Press, 2024: 27706-27716.
- [19] Zhang Yin, Ye Mu, Zhu Guiyi, et al. FFCA-YOLO for small object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5611215.
- [20] Hu Jie, Shen Li, Sun Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [21] Jocher G. Ultralytics[CP/OL]//(2025-04-01)[2025-08-15]: <https://github.com/ultralytics/ultralytics>.
- [22] Wan Dahang, Lu Rongsheng, Fang Yang, et al. YOLOv11-RGBT: towards a comprehensive single-stage multispectral object detection framework[EB/OL]. (2025-06-18).<https://doi.org/10.48550/arXiv.2506.14696>.
- [23] Shen Jifeng, Chen Yifei, Liu Yue, et al. ICAFusion: iterative cross-attention guided feature fusion for multispectral object detection[J]. Pattern Recognition, 2024, 145: 109913.
- [24] Cheng Xiaolong, Geng Keke, Wang Ziwei, et al. SLBAF-Net: super-lightweight bimodal adaptive fusion network for UAV detection in low recognition environment[J]. Multimedia Tools and Applications, 2023, 82(30): 47773-47792.

Dual-modal rail fastener defect detection model based on ADF-YOLO

Qin Yong¹, Peng Genwang^{1,2}, Cao Zhiwei¹, Gao Yang^{1,2},
Kou Lingling³, Wang Yaguan⁴

(1. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China;

2. School of Transportation, Beijing Jiaotong University, Beijing 100044, China;

3. Beijing Subway Operation Co., Ltd., Beijing 100044, China;

4. Technology Innovation Research Institute Branch, Beijing Subway Operation Co., Ltd., Beijing 100044, China)

Abstract: In rail fastener defect detection, single-modal visible-light images are susceptible to environmental interference and lack spatial structural information. This limitation reduces the model's sensitivity to small-sized

defects and spatial deformations of fasteners. Consequently, it increases the rate of false and missed detections. To address these issues, this paper proposes a detection method based on the Attention and Dual-Modal Fusion YOLO (ADF-YOLO) model. First, the ADF-YOLO model is constructed, which employs a dual-stream backbone architecture to concurrently extract texture features from visible-light images and spatial structural features from depth images of railway fasteners. Second, an Attention and Multi-Branch Convolution Synergy Module is introduced into the backbone network to improve the capture of small-sized defect features. This module expands the receptive field and enhances the representation of detailed features. Finally, an Attention-Guided Dual-Modal Feature Fusion Module is integrated into the neck network for the efficient fusion of heterogeneous features. This module fully exploits the dual-modal feature fusion paradigm. Consequently, it generates more robust fused features that are highly sensitive to defect morphology. The results demonstrate that on a multimodal fastener defect dataset, the ADF-YOLO model detection method achieves a mean average precision (m_{AP}) of 90.9% and 70.5% at IoU thresholds of 0.5 and 0.5–0.95, respectively. The ADF-YOLO method significantly outperforms various single-modal and multi-modal object detection models in detection accuracy. The proposed ADF-YOLO model provides reliable technical support for railway fastener defect inspection tasks.

Key words: Railway safety; Rail fasteners; Defect detection; Multimodal fusion; Attention mechanism; YOLO

(责任编辑 张 琪)