

基于 YOLO 的多模态钢轨表面缺陷检测方法 *

孙铁强^{1,2} 魏光辉¹ 宋超^{1,2} 肖鹏程³

(1. 华北理工大学人工智能学院 唐山 063210; 2. 华北理工大学河北工业智能感知重点实验室 唐山 063210;

3. 华北理工大学河北工业智能感知重点实验室 唐山 063210)

摘要: 针对钢轨表面缺陷区域与背景区域特征相似特性导致的模型检测性能下降问题,本文研究高实时性轻量级目标检测网络 YOLOv8n,提出一种基于 YOLO 的多模态钢轨表面缺陷检测算法 RailBiModal-YOLO。改进 YOLOv8n 模型:构建双流主干网络结构并行提取多尺度深度信息和 RGB 信息;为降低低质量图像特征相互干扰并能充分利用双模态互补信息,设计了一种即插即用的双模态特征交互修正融合模块;在多尺度特征构建阶段引入 EVCBlock,增强 RGB-D 特征层的层内信息交互,提高小缺陷检测能力。以东北大学 NEU-RSDDS-AUG 作为实验数据集,将数据集自定义划分为 4 种典型缺陷类型,以平均精度均值 mAP、每秒检测帧数 FPS、参数量作为主要评价指标,实验结果表明:所提模型与原模型相比,在保证高检测速度的同时,mAP@50, mAP@50:95 分别提高 1.8% 和 3.2%,并具有更强鲁棒性。

关键词: YOLOv8n; 缺陷检测; 多模态; RGB-D; 特征融合; 深度学习

中图分类号: TP391.41; TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi-modal rail surface defect detection method based on YOLO

Sun Tieqiang^{1,2} Wei Guanghui¹ Song Chao^{1,2} Xiao Pengcheng³

(1. College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

2. Hebei Provincial Key Laboratory of Industrial Intelligent Perception, North China University of Science and Technology, Tangshan 063210, China; 3. College of Metallurgy and Energy, North China University of Science and Technology, Tangshan 063210, China)

Abstract: To tackle the performance decline of models detecting rail surface defects due to the similarity between the characteristics of defect areas and background areas, this paper explores the high real-time, lightweight object detection network YOLOv8n and proposes a multi-modal rail surface defect detection algorithm, named RailBiModal-YOLO. Improvements to the YOLOv8n model involve the construction of a dual-stream backbone network structure that allows for the parallel extraction of multi-scale depth and RGB information; a plug-and-play dual-modal feature interaction and revision fusion module is designed to minimize the interference of low-quality image features and to fully leverage the complementary information from both modalities; the EVCBlock is introduced during the multi-scale feature construction phase to enhance the intra-layer information interaction within the RGB-D feature layers, thereby improving the detection of small defects. The Northeastern University NEU-RSDDS-AUG dataset is utilized for experiments, which has been custom-divided into four typical defect types, with mean average precision (mAP), frames per second (FPS), and the number of parameters serving as the primary evaluation metrics. It is demonstrated by the experimental results that the proposed model, in comparison to the original model, not only maintains high detection speed but also achieves enhancements in mAP@50 and mAP@50:95 by 1.8% and 3.2%, respectively, along with exhibiting increased robustness.

Keywords: YOLOv8n; defect detection; multimodal; RGB-D; feature fusion; deep learning

0 引言

直接关系到列车运行的安全和稳定。随着我国铁路交通运输业快速发展,钢轨常处于高频率、高负荷运行压力下,导致其表面易产生多种缺陷。这些缺陷如果不及时被检测到

钢轨作为铁路交通系统的重要组成部分,其质量状况

并对其进行处理,将会对铁路交通系统的安全运行构成严重威胁。因此,提高钢轨缺陷的检测效率和检测精度对保障铁路交通运输安全有至关重要的作用。

早期主要通过人工巡检的方式进行钢轨缺陷检测,这种方法效率低下,且检测结果易受检测人员主观看法影响。随后,磁粉检测法,超声波探伤法,电涡流检测等无损探伤技术^[1]的出现使得钢轨缺陷检测效率和检测精度显著提升,但这些检测方法存在抗干扰能力差、设备维护成本高等缺点。除此之外,基于传统图像处理和机器学习的机器视觉技术也常被应用于钢轨缺陷检测任务^[2],虽然其具有无接触、检测结果可视化等优点,但其需要专业人员分析复杂的缺陷特征并设计对应的特征提取器和分类器,当检测环境发生变化需要重新设计算法,泛化能力弱。

近年来,基于深度学习^[3]的目标检测技术飞速发展,由于其具备强大的特征自动提取能力、泛化能力、鲁棒性等特点,已经被广泛应用于钢轨缺陷检测任务。这些目标检测算法具体可分为单阶段和双阶段检测算法两种,双阶段算法为基于候选区域提议的卷积神经网络(region-based convolutional neural network, RCNN),其将整个检测过程分为两个阶段,首先生成候选框,然后对候选框进行分类和定位,代表性的算法有 Faster R-CNN^[4] 和 Mask-R-CNN^[5] 等。Chen 等^[6]建立钢轨缺陷数据集并训练 Faster R-CNN 模型用来检测钢轨缺陷。罗晖^[7]等为降低天气、光照等环境因素对检测结果的影响,对图像应用降噪算法和图像关键特征信息增强算法,并改进 Faster R-CNN 网络用于钢轨缺陷检测。Wang 等^[8]主要针对小尺寸缺陷检测问题,改进 Mask R-CNN 设计了新的特征金字塔结构用于检测钢轨缺陷。单阶段检测算法则将整个检测过程放在一起进行,没有生成候选框阶段,因此其检测速度相比于双阶段网络要快很多,但精度略低,常用的算法有 YOLO(you only look once)^[9] 系列算法以及 SSD(single shot multibox detector)^[10] 算法。Song 等^[11]将 YOLOv3 模型应用于钢轨表面缺陷检测任务。Zhang 等^[12]提出了基于改进 SSD 的轨面缺陷检测方法。Hu 等^[13]将 Transformer 结构的编码器模块应用于 YOLOv5 模型用来检测钢轨缺陷。杨佳佳等^[14]对 YOLOX 进行轻量化设计,使模型保证精度的同时尽可能提升检测速度。杜少聪等^[15]改进 YOLOv5,引入多头自注意力机制,构建跨层加权级联结构,将浅层信息融入到深层网络中,以提升模型的抗干扰能力和检测精度。

虽然以上基于深度学习的钢轨缺陷检测技术取得了一定进展,但基本都是基于二维视觉信息,即 RGB 图像或灰度图像提供的颜色和纹理信息进行检测,无法避免缺陷与背景区域颜色和纹理特征相似造成的检测性能下降问题。第三维深度信息作为判断钢轨缺陷的一种关键信息,具有很大的研究价值,因此通过将两种不同模态图像的优势特征进行互补融合完成检测是一种有效的解决方案。Wang 等^[16]首次将深度信息引入并与视觉信息完成互补融合,提

出了一种协作学习注意力网络(collaborative learning attention network, CLANet)用于分割钢轨表面缺陷,有效提高了分割效果,但图像分割模型本身即具有结构复杂,检测效率低,鲁棒性较差等局限性,同时处理两种模态信息这些缺点会更加明显,且该方法无法定位缺陷的同时识别缺陷类型。赵宏伟等^[17]引入深度信息并提出了一种基于目标检测的模型检测钢轨缺陷,取得了较好的效果,但其仅简单的将两个骨干网络提取出的双模态特征图进行通道级连接融合,这种融合方式忽略了双模态特征差异性较大,以及采集到的 RGB 图像和深度图像中有一部分图像质量较差导致的特征彼此干扰问题。此外,融合之后的双模特征相较于单一模态特征更为复杂,小缺陷特征更难被准确捕获。上述方法均未针对此问题做出研究。

针对上述问题,为保证高检测效率的同时有效提高检测效果,本文将实时性单模态目标检测网络 YOLOv8 改进为双模态缺陷检测网络,其作为单阶段检测模型 YOLO 系列的最新版本,将实时性和准确率做到了很好的平衡,已被证明适用于要求高效性和高准确性的工业缺陷检测领域,如:李忠科等^[18]改进 YOLOv8 并进行轻量化设计用于检测 PCB 板缺陷,朱成杰等^[19]改进 YOLOv8 用于带钢表面缺陷检测任务。因此本文将对 YOLOv8 模型进行改进,具体改进如下:

- 1) 构建双流主干网络结构提取多尺度 RGB 图像和深度图像特征用于下一阶段融合。
- 2) 设计双模态特征交互修正融合模块(interaction revision and fusion module, IRFM),先修正原始低质量图像特征之后进行特征互补融合。
- 3) 引入显示视觉中心模块(explicit visual center block, EVCBlock),增强 RGB-D 特征层层内信息交互,用于后续预测,加强小缺陷检测效果。

1 YOLOv8 网络模型

YOLOv8 模型规模从小到大分为 n、s、m、l、x 五个版本。网络主体由骨干网络(Backbone)、颈部网络(Neck)、检测头(Head)三部分组成,具体结构如图 1 所示(针对不同规模的 YOLOv8 模型,网络宽度(width, w),缩放因子(ratio, r)不同,本文对 YOLOv8n 模型改进, $w = 0.25$, $r = 2$)。

YOLOv8 使用 CSPDarknet53 作为骨干网络,如图 1 所示,骨干结构延用 YOLOv5 的跨阶段局部融合(cross stage partial, CSP)结构设计思想,主体由 CBS、C2f 和 SPPF 三部分组成;颈部采用路径聚合网络(path aggregation network, PAN)结构设计理念,构建双流特征金字塔网络(feature pyramid networks, FPN)^[20],对主干网络在 3 个不同阶段提取到的特征图进行处理和融合,从而加强特征表示以便更好的捕获不同尺度的信息;YOLOv8 的头部模块采用解耦头结构,使用两个独立的分

支进行对象分类和边界框预测,对这两个分支的任务应用不同的损失函数,同时采用无锚框策略,不再需要生成大量

候选框,这种头部结构设计一定程度上提高了网络训练速度和检测精度。

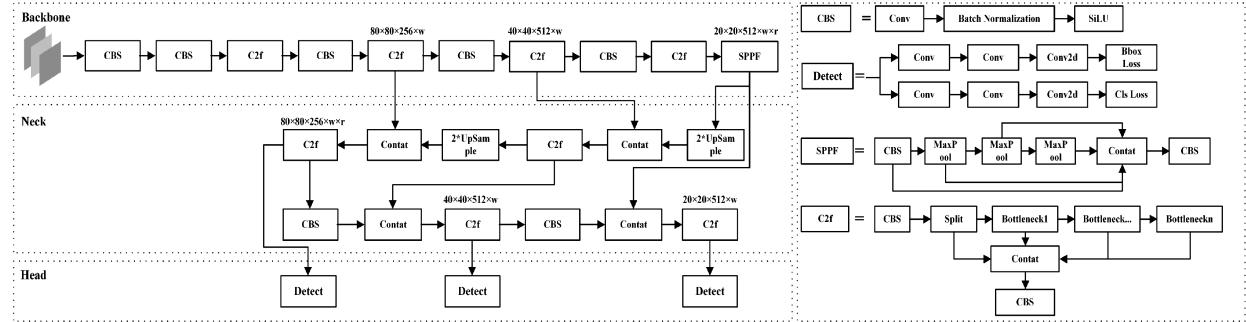


图 1 YOLOv8 网络结构图

Fig. 1 YOLOv8 network structure diagram

2 相关工作

2.1 特征提取

本文采用中期特征级融合策略,在特征提取之后对提取到的双模态图像特征进行融合。这种融合策略可以保证骨干网络能够保留各模态的独立关键特征,这些特征对于下一阶段的特征交互修正融合有重要作用。

使用原骨干网络 Darknet53 作为特征提取器,构建两

个并行的骨干网络实例用于提取双模态图像特征,特征提取结构如图 2 所示。颜色变化是判断钢轨缺陷的重要特征,应尽可能使缺陷区域与背景区域颜色特征具有更强的差异性,因此将包含丰富颜色特征的 RGB 图像作为视觉特征提取骨干 Visual Backbone 的输入。对于 RGB 图像,Visual Backbone 提取其颜色、纹理等视觉特征,对于深度图像,几何特征提取骨干 Geometric Backbone 提取图像中场景的三维结构和缺陷边界特征信息。

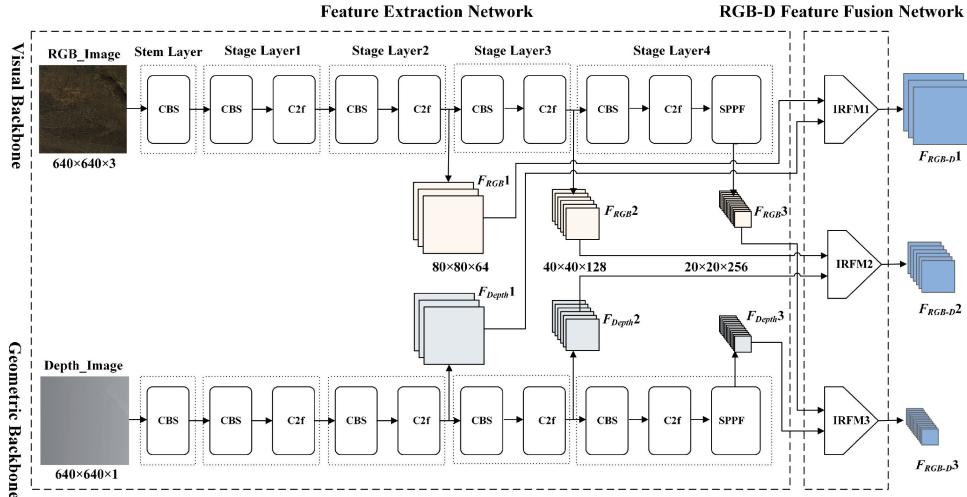


图 2 双流骨干网络结构

Fig. 2 Dual-stream backbone network structure

将双流骨干网络分别在 Stage Layer2, Stage Layer3, Stage Layer4 层后面输出的 3 个不同尺度特征图作为目标特征层用于下一阶段融合,分别记为 $\{F_{RGB}1, F_{RGB}2, F_{RGB}3\}, \{F_{Depth}1, F_{Depth}2, F_{Depth}3\}$ 。将每个尺度成对 RGB 特征和深度特征输入融合模块 IRFM 得到 $\{F_{RGB-D}1, F_{RGB-D}2, F_{RGB-D}3\}$ 3 个不同尺度的场景三维空间特征。

2.2 双模态特征融合模块 IRFM 设计

当前深度相机成像技术有限,易受被测区域表面光线反射率、异物材质、颜色深浅以及光滑程度等多种因素影

响,可能会产生深度值异常的深度图像,同样,由于钢轨缺陷自身特性,采集到的 RGB 图像往往具有缺陷与背景区域特征相似的特点。常用的特征融合方式(如相加、连接等)往往会忽略这些低质量图像的影响以及不同模态特征之间的差异,且不能充分利用双模态信息的互补特性,导致其中一种模态特征抑制另一种模态特征的表示,或是相互补充有误导性的特征信息,反而起到负面作用。

综合以上两点,为避免双模态信息相互干扰并充分利用双模态信息的互补特性,设计引入了通道空间注意力模

块(convolutional block attention module, CBAM)^[21]的IRFM 模块, 模块结构如图 3 所示。为了降低模块参数量

且避免模型过拟合, 由深度可分离卷积(depthwise separable convolution, DSConv)^[22]代替传统卷积。

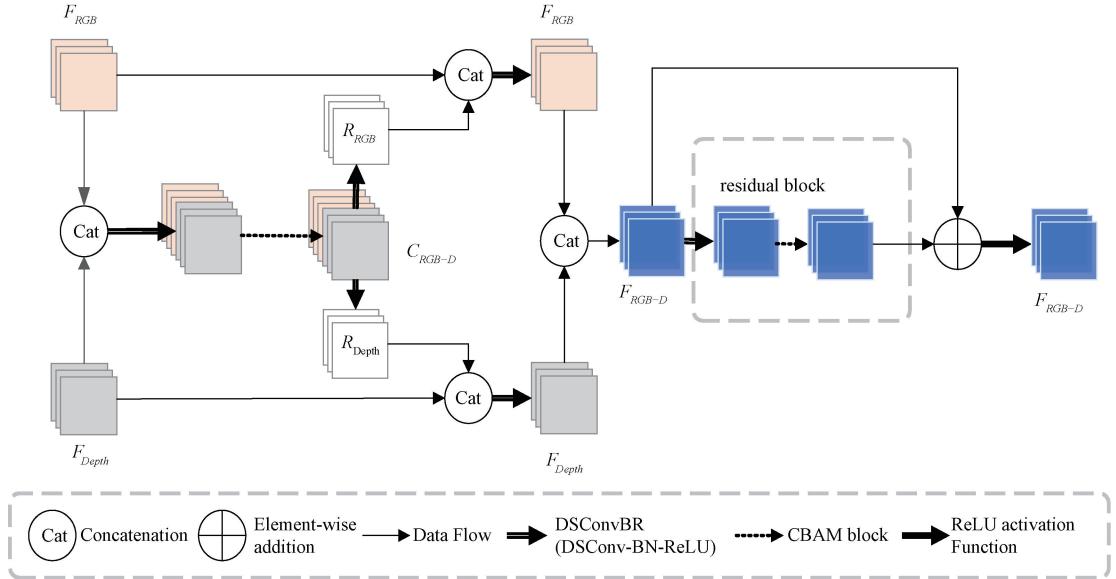


图 3 IRFM 模块结构图

Fig. 3 IRFM module structure diagram

特征融合分为两个阶段, 第 1 阶段通过特征层之间的交互修正原始 RGB 特征图和深度特征图, 第 2 阶段进行关键信息互补融合。

在第 1 阶段, 将双流主干网络提取到的 RGB 特征图和深度特征图 F_{RGB} 、 F_{Depth} 进行拼接操作, 得到双模态特征图 F_{RGB-D} , 如式(1)所示。

$$F_{RGB-D} = \text{Concat}(F_{RGB}; F_{Depth}) \quad (1)$$

接下来通过 DSConvBR_{3×3} 卷积模块和 CBAM 注意力模块串行连接的结构对 F_{RGB-D} 执行卷积操作, 整合 RGB 信息和深度信息, 捕获各自模态独有的关键特征并学习这些特征之间的联系, 构建相关性表示, 相关性矩阵记为 C_{RGB-D} , 通过两个独立的 CBR_{1×1} 模块(Conv_{1×1}-BN-ReLU, 通道数为 C_{RGB-D} 的 1/2)将相关性矩阵分解成两个分别包含高度区分度的 RGB 信息和深度信息修正单元 R_{RGB} 和 R_{Depth} , 如式(2)所示。

$$R_{RGB}, R_{Depth} = \text{Split}(C_{RGB-D}) \quad (2)$$

其中, R_{RGB} 主要包含关键深度特征, R_{Depth} 主要包含 RGB 图像的关键特征。将修正单元 R_{Depth} 与原始深度特征图 F_{Depth} 拼接并使用 DSConvBR_{3×3} (DSConv_{3×3}-BN-ReLU) 模块对原始深度特征图进行修正, 通过 RGB 图像丰富的纹理和颜色信息抑制深度特征图异常区域的影响, 并突出深度值正常区域特征表示, 得到新的深度特征图 \tilde{F}_{Depth} 。同样的, 通过 R_{RGB} 修正 RGB 特征图 F_{RGB} , 使其缺陷区域和伪缺陷区域具有更高区分度, 得到新的 RGB 特征图 \tilde{F}_{RGB} 。将修正之后的两个特征图进行拼接操作得到包含更少误导性信息的 RGB-D 特征图 F_{RGB-D} , 修正过程如

式(3)所示。

$$\begin{aligned} \tilde{F}_{RGB} &= \text{DSConvBR}_{1×1}(\text{Concat})(R_{RGB}, F_{RGB}) \\ \tilde{F}_{Depth} &= \text{DSConvBR}_{1×1}(\text{Concat})(R_{Depth}, F_{Depth}) \\ F_{RGB-D} &= \text{Concat}(\tilde{F}_{RGB}, \tilde{F}_{Depth}) \end{aligned} \quad (3)$$

在第 2 阶段, 构建残差连接结构并在残差块尾部引入 CBAM, 进一步过滤冗余信息, 突出关键信息, 构建出缺陷在三维空间中的特征表示, 如式(4)所示。

$$F_{RGB-D} = F_{RGB-D} + \text{Res}(F_{RGB-D}) \quad (4)$$

2.3 多尺度 RGB-D 特征增强

钢轨表面缺陷形态多变, 尺度差异性大, 且由于融合双模态信息之后的三维空间特征比单一模态信息更为复杂, 模型难以学到小目标的有效特征, 针对小缺陷检测效果仍然较差。

为了增强 RGB-D 预测特征层的多尺度特征表示, 在构建 FPN 阶段引入 EVCBlock^[23]。其由两个并行连接的模块组成: 轻量级多层感知机 (multilayer perceptron, MLP) 和可学习的视觉中心 (learnable visual center, LVC), 如图 4 所示。

其中 MLP 负责捕获图像的全局长距离依赖关系, 其由深度卷积和通道 MLP 两个残差模块组成。首先使用 stem 块对原始图像特征进行平滑处理得到特征 X_{in} , 之后使用深度卷积模块接收该特征, 通过深度卷积层进行处理之后与 X_{in} 进行残差连接。接着, 通道 MLP 模块进一步提取高级别的全局特征, 以获取更丰富的图像信息, 经通道 MLP 模块处理之后与原特征进行残差连接作为 MLP 模块的输出, 如式(5)、(6)所示。

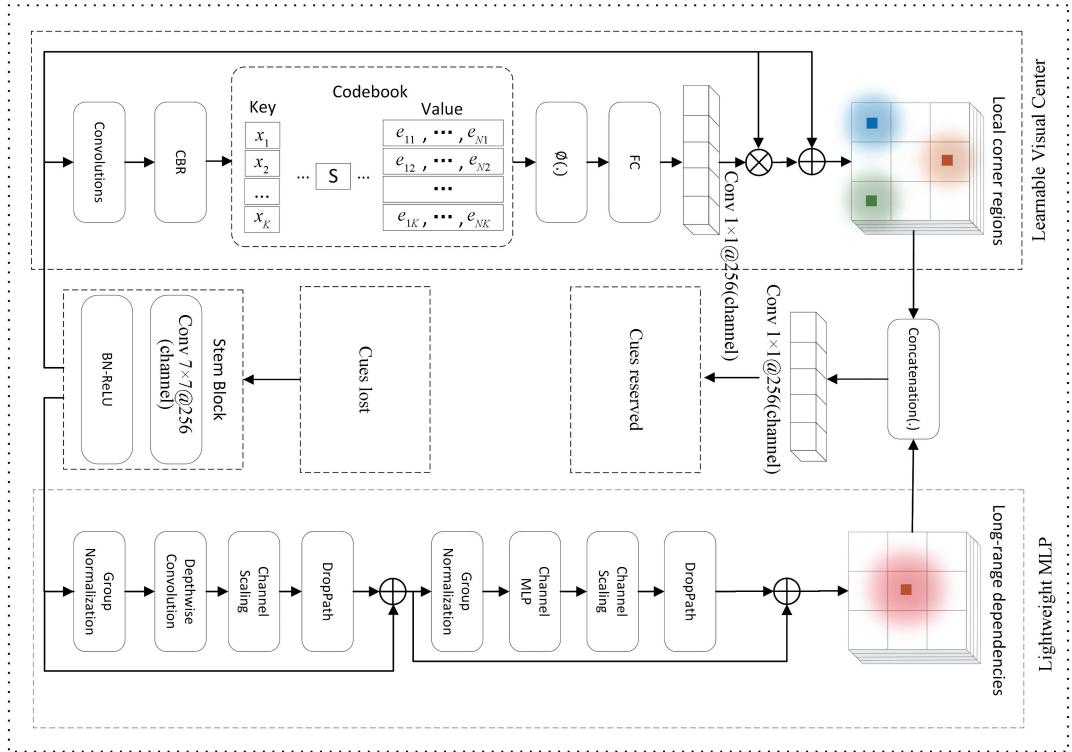


图 4 EVCBlock 结构图

Fig. 4 EVCBlock structure diagram

$$\tilde{X}_{in} = DConv(GN(X_{in})) + X_{in} \quad (5)$$

$$MLP(X_{in}) = CMLP(GN(\tilde{X}_{in})) + \tilde{X}_{in} \quad (6)$$

两个残差块最后进行了通道缩放和 DropPath 正则化操作, 以提高特征的泛化能力和鲁棒性。

LVC 是一个内置字典的编码器, 通过编码器和字典对局部信息进行编码和聚合, 尽可能地保留输入图像的局部角落区域信息。具体来说, 经 stem 块平滑之后的特征 X_{in} 首先通过一组卷积层进行编码, 再通过 CBR (Conv-BN-ReLU) 块处理得到 X_{in} 特征层, 最终将特征层分为 k 个特征嵌入到码书, 通过 k 个缩放因子对其映射相应的位置信息生成 k 个特征向量, 如式(7)所示。

$$e_k = \sum_{i=1}^N \frac{e^{-s_k} \|x_i - b_k\|^2}{\sum_{j=1}^k e^{-s_k} \|x_i - b_k\|^2} (x_i - b_k) \quad (7)$$

其中, $N = H \times W$ (H 为特征图的高度, W 为特征图的宽度), 通过融合算法将这些特征向量组合起来作为码本的输出, 如式(8)所示。

$$e = \sum_{k=1}^k \phi(e_k) \quad (8)$$

其中, ϕ 为 BN 层和激活层组合, 进一步把该输出输入到一个全连接层和一个 1×1 卷积层中。随后, 使用原特征 X_{in} 与缩放因子系数 $\delta(\cdot)$ 进行通道间的乘法运算。再对原特征 X_{in} 和局部角区特征 Z 之间执行通道加法运算,

如式(9)、(10)所示。

$$Z = X_{in} \otimes (\delta(\text{Conv}_{1 \times 1}(e))) \quad (9)$$

$$LVC(X_{in}) = X_{in} \otimes Z \quad (10)$$

最终, 将 MLP 和 LVC 的输出沿通道维度拼接在一起作为 EVCBlock 的输出, 得到包含全局远程依赖关系和局部角区信息的特征映射, 如式(11)所示。

$$X = \text{cat}(MLP(X_{in}); LVC(X_{in})) \quad (11)$$

通过该特征映射来调节其他特征层, 调整策略如图 5 所示。用最深层 IRFM 融合模块输出的包含丰富语义信息和全局信息的 F_{RGB-D3} 特征层构建 EVC 信息, 将该 EVC 信息通过两次独立的上采样操作和标准 1×1 卷积调整到与其他两个浅层预测特征层相同的尺寸和通道数, 共得到 3 个包含关键特征映射信息的特征向量 **EVC1**、**EVC2**、**EVC3**。分别将这 3 个特征向量与 3 个 IRFM 融合模块输出的 3 个原始预测特征层 $\{F_{RGB-D1}, F_{RGB-D2}, F_{RGB-D3}\}$ 执行通道级连接操作, 并通过 3 个独立的 $Conv_{1 \times 1}$ 卷积块对连接之后的特征进行调整, 为降低参数量并过滤无用特征进行通道缩放操作, 将卷积块通道数设为拼接之后特征通道数的 $1/2$, 以此为 3 个原始预测特征层注入全局远程依赖关系信息, 并同时增强其层内信息交互, 构建出更加全面、差异化的特征表示, 使模型能够有效区分不同尺度下目标的特征, 从而提高模型对密集小目标缺陷的检测能力。

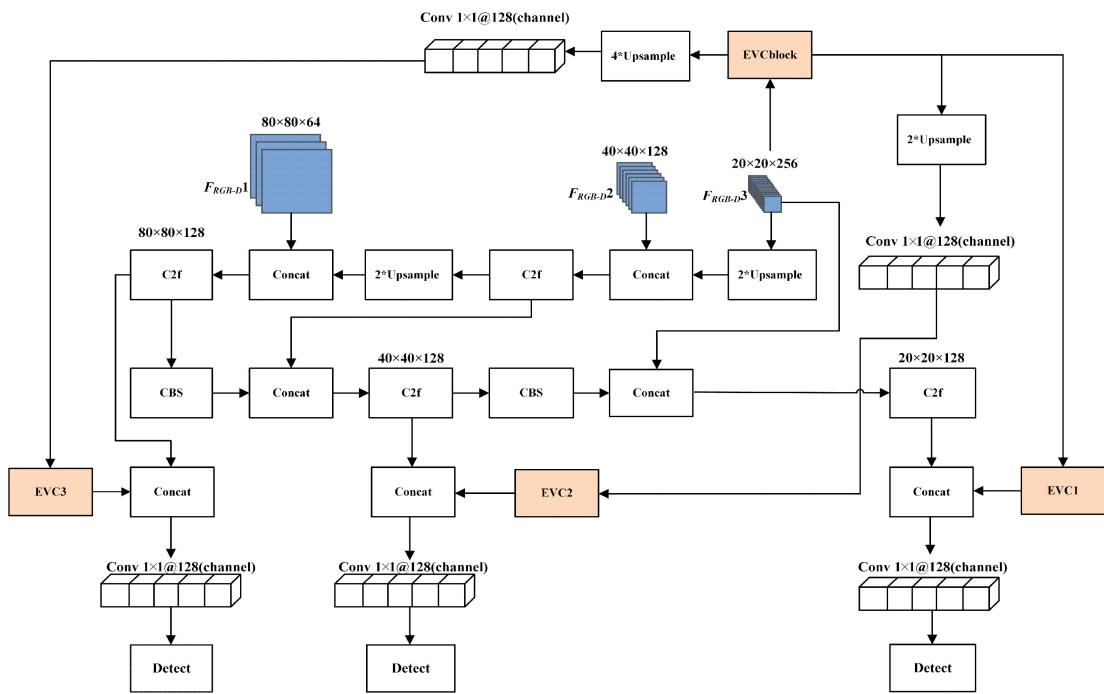


图 5 EVCBlock 调整预测特征层

Fig. 5 EVCBlock adjusted prediction feature layer

3 实验

3.1 实验环境

实验所使用的软件、硬件以及版本如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

软/硬件	版本
操作系统	Ubuntu 18.04
GPU	NVIDIA GeForce RTX 3080(24 GB)
CPU	Intel(R) Xeon(R) Platinum 8255C
编程语言	Python 3.8.10
深度学习框架	PyTorch 1.10.0
计算平台	CUDA 11.1
GPU 加速库	CUDNN 8005

3.2 实验数据集

本次实验使用工业 RGB-D 数据集 NEU-RSDDS-AUG^[16]作为实验数据集, 其中钢轨缺陷部位均是专业人员所认定并标注。该数据集是基于双目立体视觉系统, 从 15 个长度在 130~400 mm 之间, 质量为 60.64 kg/m 的常用轨道 P60 重轨上采集, 其中采集到的深度图像具有 14 微米的高精度深度信息, 能够很好的满足模型性能测试和实际应用需要, 共采集到 113 张成对的像素级对齐的 RGB 图像和深度图像。由于获取到的图像数量远不能满足模型训练的需要, 且原始图像分辨率非常高, 因此使用

MATLAB 通过随机网格对图像进行分割操作以完成数据增强, 最终把质量较差的图像过滤之后共获得 1 862 对图像, RGB-D 数据集如图 6 所示。这些图像分辨率随机且缺陷位置随机, 更符合缺陷检测模型实际应用环境。根据缺陷的公共特性将数据集自划分定义 hole、gutter、scar、protrusion 四种缺陷类型并使用 LabelImg 标注工具标注缺陷用来测试模型性能, 训练集和测试集数量按 8:2 的比例随机划分, 分别为 1 490 和 372 张图像, 各类型缺陷数量及占比如表 2 所示。

3.3 超参数设置

在图像输入网络之前将图像分辨率动态的调整到 640×640 像素大小, 训练过程中开启 Mosaic 和 Mixup 数据增强操作提高模型泛化性, 避免过拟合。使用 sgd 优化器更新模型参数。此外, 将初始学习率(learning rate, lr)、迭代次数(epoch)分别设置为 2×10^{-2} 、300, 其中前 50 次迭代为冻结训练, 仅训练除双流骨干网络之外的参数, 批量大小(batch size)为 64, 之后解冻主干网络进行训练, batch size 变为 32, 学习率的调整采用余弦退火算法(cosine annealing, Cos) 加自适应矩估计(adaptive moment estimation, Adam)优化算法相结合的策略, 以更好地调整学习率以加速模型的收敛, 其中动量(Momentum)设置为 0.937。

3.4 评估指标

针对钢轨表面缺陷检测任务, 本文以平均精度均值(mean average precision, mAP)、每秒检测图像数量(frames per second, FPS)、参数量作为主要评价指标,

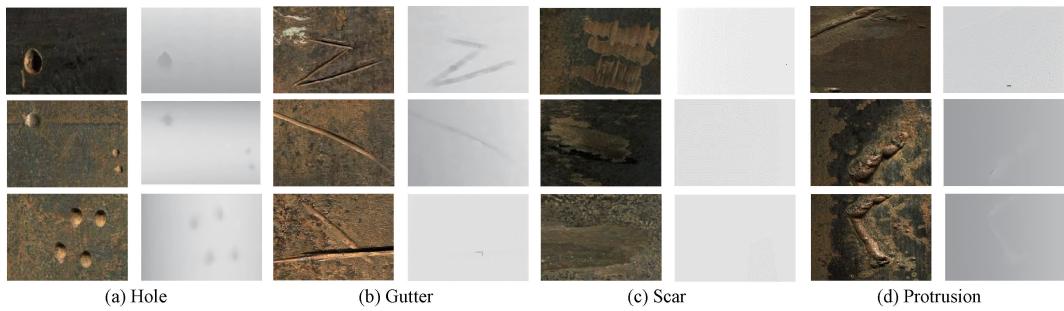


图 6 NEU-RSDDS-AUG 数据集

Fig. 6 NEU-RSDDS-AUG dataset

表 2 数据集划分及缺陷数量

Table 2 Dataset division and defect quantities

缺陷类型	hole	scar	gutter	protrusion	all
数量/张	536	978	1 017	621	3 152
占比/%	17	31	32	20	100

mAP 计算方法如式(12)~(15)所示。

$$P(\text{Precision}) = \frac{TP}{TP + FP} \quad (12)$$

$$R(\text{Recall}) = \frac{TP}{TP + FN} \quad (13)$$

$$AP = \int_0^1 P(R) d(R) \quad (14)$$

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (15)$$

其中, TP 、 TN 、 FP 、 FN 分别代表真阳性、真阴性、假阳性、假阴性, n 代表所检测样本缺陷类别数。

3.5 实验结果与分析

1) 消融实验

为了验证改进策略的有效性, 本文通过消融实验对每个改进点进行测试, 实验结果如表 3 所示(表中实验数据均为设置合适的超参数后且模型收敛之后的结果)。

表 3 消融实验

Table 3 Ablation experiment

方案	Add	Concat	IRFM	IRFM(+CBAM)	EVCBlock	mAP@50/%	mAP@50:95/%	GFLOs/G	Params/M
1	√					94.8	71.6	11.42	4.28
2	√				√	95.5	73.2	15.05	8.62
3		√				95.1	72.7	11.74	4.46
4		√			√	95.5	73.0	15.37	8.79
5			√			96.1	74.7	15.54	6.49
6			√		√	97.0	76.0	19.18	10.83
7				√		96.7	75.4	15.55	6.50
8				√	√	97.2	76.7	19.18	10.84

使用直接相加和拼接操作进行特征融合会使模型检测效果略有下降, 其中相加操作会使模型效果下降更为明显, 证明了两种模态信息差异较大, 简单的相加和拼接会造成特征层内存在信息干扰情况。方案 2、4 分别对相加、连接两种融合方式应用 EVCBlock 模块调整 RGB-D 预测特征层, 相比于直接在原始特征层上进行预测效果有所提升。方案 5 使用 IRFM 模块融合两种不同模态的信息, 与原模型相比 mAP@50 提高了 0.7%, mAP@50:95 提高了 1.2%, 在 IRFM 模块引入 CBAM 之后, 与未加入该注意力机制的 IRFM 模块相比, 在参数量和运算量基本保持不变的情况下 mAP@50 提高了 0.6%, mAP@50:95 提高了 0.7%, 显然 CBAM 可以有效的突出关键信息, 过滤冗余信息。应用 IRFM(+CBAM)融合双模态特征之后在网络颈

部对每个预测特征层应用 EVCBlock 调整其层内信息, 实验证明该模块可以有效的提高模型精度, 最终使模型的 mAP@50 达到了 97.2%, mAP@50:95 达到了 76.7%, 与未加入 EVCBlock 模块相比 mAP@50 提高 0.5%, mAP@50:95 提高 1.3%。

2) 对比实验

将该模型与其他基于单一模态的钢轨缺陷检测领域常用模型进行比较, 实验结果如表 4 所示。

与原模型 YOLOv8n 相比该模型 mAP@50、mAP@50:95 分别提高了 1.8% 和 3.2%, 相比于在钢轨缺陷检测任务中综合性能表现最好的 YOLOv8s, 本模型的 mAP@50、mAP@50:95 分别高 0.6%、0.8%, FPS 高 11, 证明了该模型可以在保证高检测效率的情况下有效提高检测效果。

表 4 对比实验

Table 4 Comparison experiment

网络模型	mAP@50/%	mAP@50:95/%	GFLOPs/G	FPS	Params/M
Faster-RCNN	95.8	73.2	—	26	41.37
YOLOv8n	95.4	73.5	8.23	117	3.01
YOLOv8s	96.6	75.9	28.71	82	11.14
YOLOv5s	96.0	75.2	17.27	77	7.05
本文	97.2	76.7	19.18	93	10.84

3) 鲁棒性测试

为评估模型在面对真实环境中可能遇到的干扰因素(如传感器噪声、传输过程干扰等)时的表现,对测试集加入噪声评估模型的鲁棒性。高斯噪声由于其数学特性,即噪声值服从正态分布,能够很好的模拟这些现实中的随机干扰,因此选择对深度图像和 RGB 图像添加高斯噪声作为模拟噪声。

模型表现结果如表 5 所示,加入高斯噪声之后,YOLOv8n, YOLOv8s 的 mAP@50 分别下降 56.7% 和 60.4%,本文提出的模型下降 24%。实验结果表明,当 RGB 图像信息和深度图像信息受到环境干扰时,IRFM 模块仍能使得双模态图像剩余的有价值信息彼此修复被干扰的像素区域,该模型具有更强的鲁棒性。

表 5 加入噪声前后模型检测性能对比结果

Table 5 Comparison of model detection performance before and after adding noise

网络模型	加噪声前		加噪声后	
	mAP@50/%	mAP@50:95/%	mAP@50/%	mAP@50:95/%
YOLOv8n	95.4	73.5	38.7	21.4
YOLOv8s	96.6	75.9	36.2	19.8
本文	97.2	76.7	73.2	48.4

4) 模型可视化检测效果对比

为了更直观地观察改进算法检测效果,将改进前后的模型检测结果进行可视化对比,如图 7 所示。从两组图像

检测效果对比可以看出,针对缺陷与背景区域颜色与纹理特征相似,原模型无法准确判断缺陷的图像,深度信息与IRFM 模块的引入可以使 RGB 特征和深度特征有效抑制

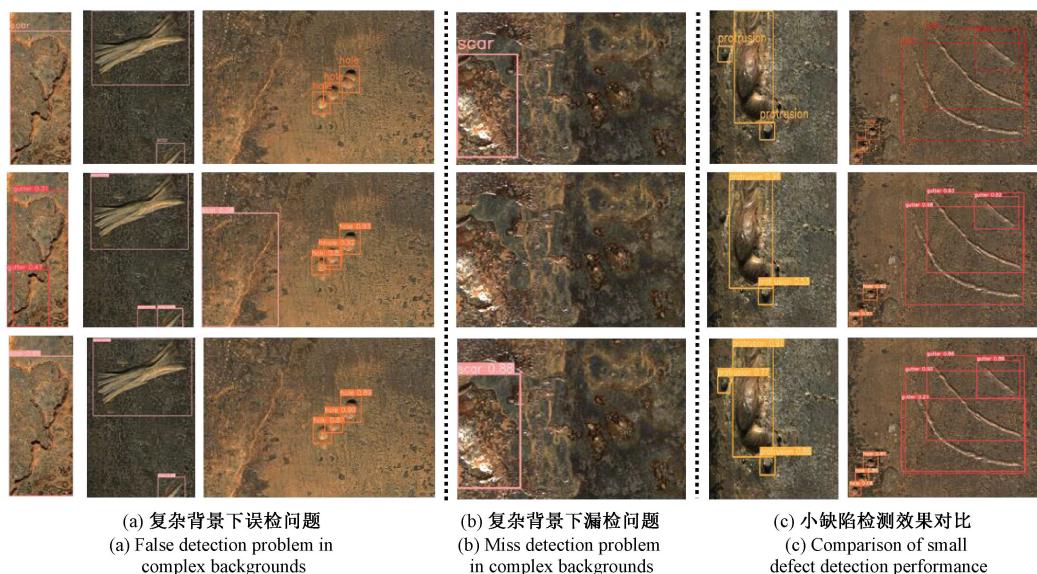


图 7 模型检测结果对比(上、中、下 3 组图像分别对应真实标签、YOLOv8n 模型检测结果、本文改进模型检测结果)

Fig. 7 Comparison of model detection results (the top, middle, and bottom images represent the ground truth labels, YOLOv8n model detection results, and the detection results of the proposed improved model, respectively)

彼此低质量信息的表达,同时增强有价值信息的表示,从而增强前景与背景特征的区分度,进一步降低模型的误检和漏检现象发生,如图 7(a)和(b)所示。此外,通过EVCBlock 对 RGB-D 预测特征层的调整,有效增强了小目标缺陷的检测效果,如图 7(c)所示。综合来看,本文模型相较于与原模型有更好的检测效果。

4 结 论

本文提出了一种基于 YOLO 的多模态钢轨表面缺陷检测算法 RailBiModal-YOLO。通过改进高实时性轻量级网络 YOLOv8n,构建双流主干网络结构并行提取双模态特征;设计 IRFM 特征融合模块,交互修正原始特征之后进行特征融合,从而降低低质量图像对检测结果的影响,充分融合不同模态的关键互补信息;引入 EVCBlock 进一步调整预测特征层,增强特征层层内信息交互,提高其多尺度特征表征能力,保证模型对密集缺陷和小缺陷有更好的检测效果。

实验证明了本文改进的有效性,模型在保证高检测效率情况下显著提升了检测效果,并且通过鲁棒性测试证明该模型具有更强的环境适应性。

参考文献

[1] 张辉,宋雅男,王耀南,等.钢轨缺陷无损检测与评估技术综述[J].仪器仪表学报,2019,40(2):11-25.
ZHANG H, SONG Y N, WANG Y N, et al. Overview of non-destructive detection and evaluation techniques for rail defects [J]. Chinese Journal of Scientific Instrument, 2019, 40(2): 11-25.

[2] 闵永智,岳彪,马宏锋,等.基于图像灰度梯度特征的钢轨表面缺陷检测[J].仪器仪表学报,2018,39(4):220-229.
MIN Y ZH, YUE B, MA H F, et al. Detection of rail surface defects based on image gray-scale gradient features[J]. Chinese Journal of Scientific Instrument, 2018, 39(4): 220-229.

[3] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[4] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137-1149.

[5] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision(ICCV), 2017: 2980-2988.

[6] CHEN X, ZHANG H. Rail surface defects detection based on Faster R-CNN [C]. 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 2020: 819-822.

[7] 罗晖,徐广隆.基于图像增强与深度学习的钢轨表面缺陷检测[J].铁道科学与工程学报,2021,18(3):623-629.

[8] LUO H, XU G L. Detection of rail surface defects based on image enhancement and deep learning[J]. Journal of Railway Science and Engineering, 2021, 18(3): 623-629.

[9] WANG H, LI M J, WAN ZH B. Rail surface defect detection based on improved Mask R-CNN [J]. Computers and Electrical Engineering, 2022, 102: 108269.

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Cham: Springer, 2016: 31-37.

[12] SONG Y N, ZHANG H, LI L, et al. Rail surface defect detection method based on YOLOv3 deep learning networks [C]. 2018 Chinese Automation Congress(CAC), 2018: 1563-1568.

[13] ZHANG B, FANG SH Q, LI ZH X. Research on surface defect detection of rare-earth magnetic materials based on improved SSD[J]. Complexity, 2021, DOI:10.1155/2021/4795396.

[14] HU Q, TANG B, JIANG L, et al. Rail surface defects detection based on YOLOv5 integrated with transformer [C]. 2022 IEEE 5th International Conference on Electronics Technology(ICET), 2022: 1131-1135.

[15] 杨佳佳,许贵阳,白堂博.基于改进 YOLOX 的轻量级钢轨表面缺陷检测算法[J].铁道建筑,2023,63(7):34-39.
YANG J J, XU G Y, BAI T B. Lightweight rail surface defect detection algorithm based on improved YOLOX[J]. Railway Engineering, 2023, 63 (7): 34-39.

[16] 杜少聪,张红钢,王小敏.基于改进 YOLOv5 的钢轨表面缺陷检测[J].北京交通大学学报,2023,47(2):129-136.
DU SH C, ZHANG H G, WANG X M. Detection of rail surface defects based on improved YOLOv5[J]. Journal of Beijing Jiaotong University, 2023, 47(2):

129-136.

[16] WANG J, SONG K CH, ZHANG D, 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.

[17] 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 H W, ZHENG J J, ZHAO X X, et al. Rail surface defect detection method based on bi-modal deep learning [J]. Computer Engineering and Applications, 2023, 59(7): 285-293.

[18] 李忠科,刘小芳.基于轻量级 YOLOv8n 网络的 PCB 缺陷检测算法[J].电子测量技术,2024,47(4):120-126.
LI ZH K, LIU X F. PCB defect detection algorithm based on lightweight YOLOv8n network [J]. Electronic Measurement Technology, 2024, 47 (4): 120-126.

[19] 朱成杰,刘乐乐,朱洪波.基于 YOLOv8-NFMC 的带钢表面缺陷检测算法[J].国外电子测量技术,2024,43(7):97-104.
ZHU CH J, LIU L L, ZHU H B. Steel strip surface defect detection algorithm based on YOLOv8-NFMC[J]. Foreign Electronic Measurement Technology, 2024, 43 (7): 97-104.

97-104.

[20] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[21] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision(ECCV), 2018: 3-19.

[22] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.

[23] QUAN Y, ZHANG D, ZHANG L Y, et al. Centralized feature pyramid for object detection[J]. IEEE Transactions on Image Processing, 2023, 32: 4341-4354.

作者简介

孙铁强,博士,教授,主要研究方向为机器视觉、图像处理。

E-mail: hblgdxs@163.com

宋超(通信作者),博士,讲师,主要研究方向为机器视觉。
E-mail: 15633616170@163.com

魏光辉,硕士研究生,主要研究方向为计算机视觉、深度学习。

E-mail: weiglory@163.com