

DOI: 10.13382/j.jemi.B2508416

GMS2-YOLO:融合改进 MANet 和检测头的金属表面缺陷检测方法*

严 赫¹ 史晓亮¹ 吴超华¹ 曾 志¹ 罗 威¹ 孙存甫²

(1. 武汉理工大学机电工程学院 武汉 430070; 2. 武汉小出钢管有限公司 武汉 430090)

摘 要:针对工业现场生产中对金属表面缺陷检测存在精度低、误检和漏检的问题以及金属表面缺陷检测面临小目标难以分辨、复杂背景干扰强、噪声影响显著等挑战,尤其在小目标检测与多尺度特征提取方面仍存在不足,针对上述挑战提出 GMS2-YOLO 模型,以多尺度特征提取、特征融合增强与分离批归一化检测头相结合,提高目标缺陷检测精度。首先,将 GhostConv 与 HGBlock 模块结合,将 GHGnet 网络结构作为主干网络,提升模型对目标缺陷细节特征的提取能力;然后,在颈部结构创新型融合 MANet 与 StarBlock,使用 MSNet 替换 C3k2 模块,实现更加多样化和丰富的梯度流信息,增强模型的提取融合能力;其次,重新设计 YOLOv11n 检测头,采用分离批归一化检测头实现信息在不同层级间的交互与融合,准确识别缺陷目标;最后,使用新的损失函数 WIoU,通过提高对中等质量图片的关注度,增强了缺陷检测的精度。实验结果表明,此方法在 Self-Dataset 数据集上平均精度均值(mAP)mAP@0.5 指标达到了 80.8%,与 YOLOv11n 相比提高了 3.6%;mAP@0.5:0.95 指标达到了 53.5%,与 YOLOv11n 相比提高了 8.5%。此外,在 NEU-DET 数据集中,针对该数据集中的微小缺陷 CR,其平均精度(AP)值达到了 68.3%,与 YOLOv11n 相比提高了 15.8%。所提模型对缺陷检测精度有较大提升,在解决误检、漏检问题方面有明显优势。此外,与其他主流模型对比,改进模型在没有损失较多推理速度的前提下,提高了检测精度,具有良好的工程应用前景。

关键词: 缺陷检测;YOLOv11;特征提取;MSNet;SBN 检测头;损失函数

中图分类号: TP391.4;TN98 **文献标识码:** A **国家标准学科分类代码:** 520.2060

GMS2-YOLO: A metal surface defect detection method integrating improved MANet and detection head

Yan He¹ Shi Xiaoliang¹ Wu Chaohua¹ Zeng Zhi¹ Luo Wei¹ Sun Cunfu²

(1. School of Mechanical and Electrical Engineering, Wuhan University of Technology, Wuhan 430070, China; 2. Wuhan Koide kokan Co., Ltd., Wuhan 430090, China)

Abstract: Addressing the issues of low accuracy, false positives, and missed detections in metal surface defect detection in industrial field production, as well as the challenges faced by metal surface defect detection, such as difficulty in distinguishing small targets, strong interference from complex backgrounds, and significant noise impact, especially in the areas of small target detection and multi-scale feature extraction, this study proposes the GMS2-YOLO model to improve the accuracy of target defect detection by combining multi-scale feature extraction, feature fusion enhancement, and a separated batch normalization detection head. Firstly, the GhostConv is combined with the HGBlock module, and the GHGnet network structure is used as the backbone network to improve the model's ability to extract the detailed features of the target defects. Then, in the innovative fusion of MANet and StarBlock in the neck structure, MSNet is used to replace the C3k2 module to achieve more diversified and rich gradient flow information and enhance the extraction and fusion ability of the model. Secondly, the YOLOv11n detection head is redesigned, and the separated batch normalization detection head is used to realize the interaction and fusion of information at different levels, so as to accurately identify the defect target. Finally, using the new loss function WIoU, the accuracy of defect detection is improved by increasing the attention to medium-quality images. The experimental results show that the mean average precision mAP@0.5 index of this method on the Self-Dataset dataset reaches 80.8%, which is 3.6% higher than that of YOLOv11n. The mAP@0.5:0.95 index reached 53.5%, which was 8.5% higher than that of

收稿日期: 2025-05-27 Received Date: 2025-05-27

* 基金项目: 国家自然科学基金(52375201)、武汉理工大学产学研科技合作项目(20251hx0411)资助

YOLOv11n. In addition, in the NEU-DET dataset, the AP value of the small defect CR in the dataset reached 68.3%, which was 15.8% higher than that of YOLOv11n. The proposed model greatly improves the accuracy of defect detection, and has obvious advantages in solving the problems of false detection and missed detection. In addition, compared with other mainstream models, the improved model improves the detection accuracy without losing more inference speed, and has good prospects for engineering applications.

Keywords: defect detection; YOLOv11; feature extraction; MSNet; SBN Detect; loss function

0 引言

在工业智能化与新型工业化协同推进的背景下,工业产品的生产和加工取得了巨大进步,对其质量也提出了更高的要求。而在产品的加工生产过程中,受到原材料,生产制造设备和生产环境等因素的影响,产品往往会有生锈、裂纹以及伤痕等缺陷。这些缺陷会导致产品的性能下降,影响产品生命周期,同时也会影响后续的生产计划。因此,为了确保产品的安全性能和预防后续不必要的变化,对金属管套件表面缺陷检测也是生产过程中必要的一环,对缺陷的识别和分类也是当前的研究热点。传统的金属表面缺陷检测主要依靠人工检测,人工检测耗费时间,人力等资源,并且重复的工作导致视觉误判,造成漏检或误检情况,有着不稳定性因素。利用机器视觉的技术进行检测效率高、成本低,是未来发展的主要方向^[1]。

对于金属表面缺陷检测而言,重点在于准确识别出缺陷的类别。目前,基于深度学习的目标检测算法有一阶段目标检测算法和二阶段目标检测算法。一阶段目标检测算法直接在网络中同时预测类别和边界框的位置,无需单独生成候选区域,代表模型有 YOLO(you only look once)^[2]、单次多框检测器(single shot multibox detector, SSD)^[3];二阶段目标检测算法第1步利用区域提议网络生成一系列候选区域,第2步对这些区域进行分类和边界框回归,代表模型有 Faster R-CNN^[4]、Mask R-CNN^[5]。此外还有基于 Transformer 架构的目标检测算法。

到目前为止,诸多学者通过深度学习对目标检测展开了大量研究,提出了许多检测方法。诸多学者对二阶段检测算法在表面缺陷检测中的应用进行了研究,向宽等^[6]基于改进 Faster R-CNN 对铝材表面缺陷进行检测,在主干网络中加入特征金字塔网络(feature pyramid network, FPN)结构用来加强对小缺陷的特征提取能力,随后用感兴趣区域校准算法获得更准确的缺陷定位信息。马志程等^[7]针对光学元件中划痕检测的难点,使用 CSPRepResNet 替换 ResNet 并添加注意力机制,优化损失函数等方法改善 Mask R-CNN。同样也对 Transformer 架构进行了研究,赵云亮等^[8]提出 DPG-Transformer 模型,该模型将卷积神经网络的局部特征提取能力与 Transformer 的全局建模能力相结合,提出了一种基于深

度可分离卷积融合池化网格窗口注意力机制的金属表面缺陷识别网络架构,该模型在金属表面缺陷检测中准确率(accuracy, ACC)达到 99.3%。

一阶段目标检测算法相比于二阶段目标检测算法,拥有更快的检测速度,非常适合于金属表面缺陷检测任务。高嘉等^[9]针对起重机电引钢丝绳缺陷,提出 TERD-Net 模型,为加强模型对于缺陷语义信息的提取、对缺陷位置信息的敏感度,设计了跨层级加权特征金字塔网络结构,并采用 β -CIoU 改进 CIoU,提升检测精度。马燕婷等^[10]通过改进 YOLOv5s 模型,提出 MT-YOLOv5 进行带钢表面缺陷检测,MT-YOLOv5 引入自注意力机制,使主干网络更聚焦于图像全局特征信息的提取,将 Transformer 层与双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)结构相结合,增强特征提取与融合,最后增加预测层,检测不同尺度的目标。赵佰亭等^[11]为检测钢材表面缺陷,提出基于 YOLOv7 的 ECC-YOLO,该模型引入特征增强模块 ConvNeXt 和 C2fB 模块加强模型的特征提取能力,通过高效加性注意力(efficient additive attention, EAA)设计 MPCE 模块,削弱复杂背景对表面缺陷检测的干扰。赵林熔等^[12]提出 CBE-YOLOv5,该算法利用 BiFPN 作为特征提取网络,充分融合不同尺度的特征,增强小目标的缺陷检测,并采用 EIou 损失函数来计算模型损失,使模型能更精确的回归。胡波等^[13]针对铝镁合金表面缺陷检测,对 YOLOv5 模型进行改进,引入 Res2Net 替换 CSPDarkNet53 实现多尺度缺陷的有效检测,其次引入 CBAM(convolutional block attention module)卷积注意力模块。改进方法对铝镁合金表面缺陷检测的精度达到 75.8%。王素珍等^[14]针对钢铁表面缺陷任务,提出 YOLO-DWCSP_CA 算法,该算法采用 DWCSP 模块结合 CA(coordinate attention)注意力机制替换原 YOLO 主干网络,其次设计了 POLY-FLoss 损失函数优化二元交叉熵损失提升检测精度。赵小虎等^[15]针对金属表面缺陷检测,将坐标注意力机制优化为三通道坐标注意力机制,并优化损失函数为 MPD-IoU,改进模型对金属表面缺陷检测精度有一定的提升。

在工业生产中金属表面明显缺陷可以被肉眼察觉,但是微小缺陷往往会被忽视。在工业相机中,金属表面微小缺陷表现特征为像素点较少,且经过特征提取容易丢失。尽管现有方法在小目标识别与复杂背景分析层面

取得阶段性突破,但仍然面临诸多挑战。金属表面缺陷涵盖裂纹、凹坑等多种形态,其与金属基底的纹理相似度极高,形成强干扰性背景。其次,缺陷分布复杂且多样,小目标缺陷在强干扰性背景下的特征表达能力不足。其次工业现场采集图像中的随机高斯噪声与光照不均造成的亮度噪声,这两类噪声会导致缺陷边界模糊、灰度对比降低,从而影响检测精度,以上问题导致现有方法难以精确定位和识别这些细节,导致漏检率较高。

针对上述问题提出了 GMS2-YOLO 模型。针对 YOLOv11n 主干网络在细粒度特征提取不足、金属纹理与缺陷边界区分困难的问题,引入 GHGNet 网络结构替换原主干(backbone)。GHGNet 将 GhostConv 与 HGBlock 有机结合,利用主/次通道分离机制抑制特征冗余,通过层次化多尺度提取实现从低级边缘到高级语义的渐进建模,从而强化模型对裂纹、凹坑等微小缺陷的表征能力,提高复杂背景下的目标可分性。针对小目标特征在下采样过程中易丢失、梯度传播不充分的问题,在颈部结构(neck)中融合改进的 MSNet 模块替换原 C3k2 模块。MSNet 综合了 1×1 旁路卷积、深度可分离卷积(DSconv)与 StarBlock 的星型操作,通过并行分支机制扩大梯度流通道并增强非线性特征交互,实现多尺度语义信息的高效融合。该改进显著提升了特征图的表征能力和模型对噪声扰动的鲁棒性,使不同尺度缺陷的特征信息得到充分传递。针对原 YOLOv11n 检测头单尺度预测、跨层信息交互不足的问题,本文设计了全新的 SBN(separated batch normalization)检测头。通过共享卷积与分层独立批归一化机制,在多尺度特征层之间构建统一的特征空间,校正不同层级的统计偏差,实现跨层信息的高效融合。该结构能够自适应不同尺度缺陷的分布特征,有效降低漏检率与重复检测现象,提高整体检测精度。针对传统 CIoU 对低质量样本过拟合、对中等质量样本学习不足的问题,采用 WIoU(wise-IoU)替换原损失函数。WIoU 引入动态非单调聚焦机制与离群度调节因子,在训练过程中弱化高质量样本的梯度优势,强化中等质量样本的优化贡献,使模型在存在随机高斯噪声与光照噪声的环境下仍能保持稳定收敛。该设计显著提升了模型的总体鲁棒性与检测一致性。

1 YOLOv11 模型

YOLOv11 是目前最新的目标检测模型,该模型继承了一阶段目标检测算法的核心特点。本文所采用的模型为 YOLOv11,主要原因在于 YOLOv11 是最新的 YOLO 系列检测模型;在关于精度指标方面 YOLOv11 的效果是最优的;YOLOv11 在保证较高精度的同时也具有较高的检测速度。YOLOv11 相比于 YOLOv5、YOLOv8 等前作,其

特点主要体现在以下几个方面。首先,在主干网络中,YOLOv11 引入 C3k2 模块,该模块继承自类 C2F,C3k2 是在 C2F 的基础上进行了改进。当 C3k 参数为 False 时,C3k2 为 C2F 模块;当 C3k 参数为 True 时,其瓶颈结构会被替换为 C3 模块。C3 模块是一种具有特定结构的模块,能够实现高效的特征提取。其次在颈部网络新增 C2PSA 模块,该模块是对 C2F 的另外一种拓展。该模块结合点式空间注意力(pointwise spatial attention,PSA),在标准 C2F 模块中引入 PSA,实现更强大的注意力机制,提高模型捕捉重要特征的能力。此外 PSA 块利用多头注意力机制和前馈神经网络,能选择性地关注输入中的重要特征信息,增加模型的多尺度识别能力。

尽管 YOLOv11 已经做出了诸多优化,但是模型还存在有一些需要改进的地方。主干网络特征提取能力有限,网络无法保留目标的细节特征,导致特征提取不完整;小目标与复杂背景噪声相邻,对于小目标检测的准确率有限以及主干网络生产的特征图分辨率低,难以捕捉小目标缺陷,使检测精度下降。YOLOv11 包含有 n、s、m、l、x 5 种不同大小的模型,本文选择 YOLOv11n 作为改进的基线模型。

2 改进的 YOLOv11n 算法

首先介绍改进后的网络整体结构,然后依次介绍了 GHGNet 网络结构、MSNet 模块、SBN 检测头以及改进损失函数在 YOLOv11n 模型上的应用。

2.1 GMS2-YOLO 模型

本文提出 GMS2-YOLO 模型,该模型结构如图 1 所示。此外为避免改进方法的经验性与主观性,从特征提取理论、梯度传播理论等角度对所提出模型进行理论分析。GHGNet 主干网络依据多尺度特征学习与冗余特征抑制理论可在信息压缩与特征表达之间取得最优平衡,从理论上增强模型对小目标和复杂纹理的可分性。MSNet 的设计基于多分支梯度流动与特征非线性融合理论,它是对深层梯度流特性与非线性表达能力的系统优化。SBN 检测头设计采用的是归一化思想,用来确保跨层特征的统计一致性。基于动态非单调聚焦机制采取使用损失函数采用 WIoU 提高模型在噪声样本下的鲁棒性和泛化性能。本文提出的各改进方法均建立理论基础之上,形成从“特征提取—特征融合—检测头—损失函数”4 层递进的一种系统优化框架。

2.2 GHGNet 网络结构

2023 年 4 月,百度首次提出实时 DETR 模型(RT-DETR)^[16]。其中使用的 HGnetv2 网络结构具有强大的特征提取能力,能够提取更加丰富和深层次的特征,从而

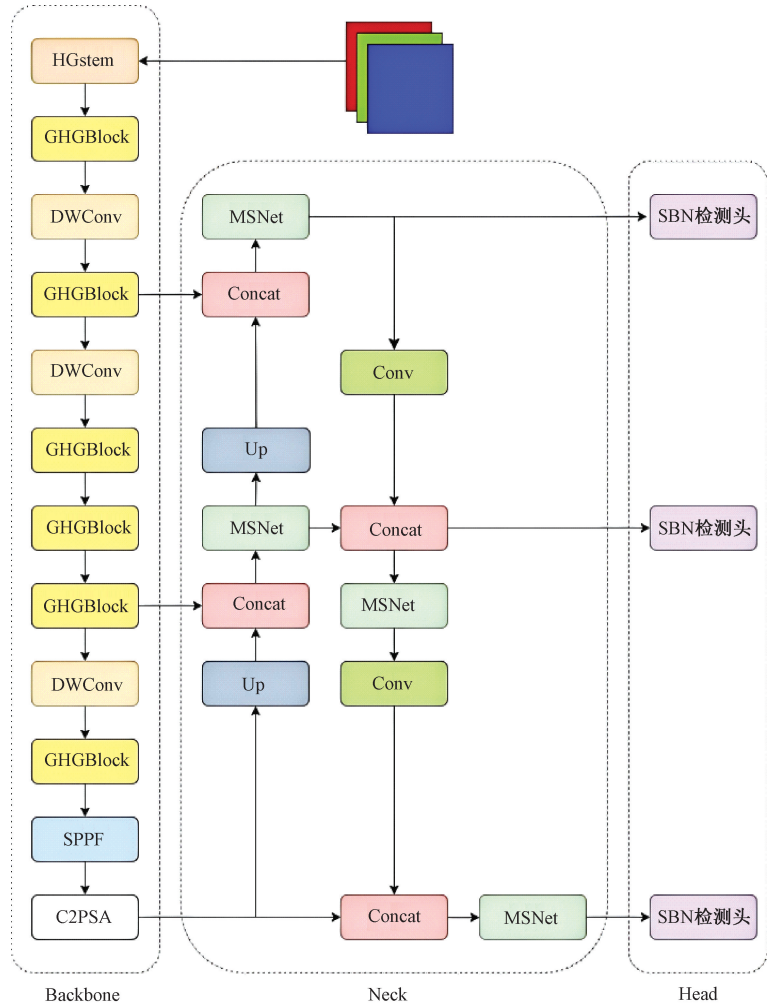


图 1 改进 YOLOv11n 模型结构
Fig. 1 Improved YOLOv11n model structure

更好地刻画目标缺陷的细节和语义信息更有利于后续模型从特征中进行学习。此外另一个网络 GHostNet^[17] 同样在特征提取领域取得了领先的效果,该网络重点使用主通道与次通道,主通道通常捕获图像的主要信息和特征,而次通道具有较少的卷积层,以细节和低级特征为目标。此外次通道充当正则化机制,有助于防止过拟合。受到两种网络结构的启发,本研究提出 GHGNet。GHGNet 是通过合并 GhostConv 和 HGBlock 模块而形成的。GhostConv 首先将输入进行卷积操作得到少部分本体卷积,之后通过线性变换生成幻影特征,保留原始特征图的关键信息,GhostConv 如图 2 所示。

使用 GhostConv 对原 HGnetv2 中的 HGBlock 进行优化,得到 GHGBlock 模块。GHGBlock 作为 GHGNet 网络的核心模块,采用分层处理机制,实现对输入数据的多尺度特征提取与融合。通过分层处理机制能够捕获不同层次的特征表示,逐步提取图像的低级边缘纹理到高级语义特征,同时降低特征图分辨率以保留关键信息。每个

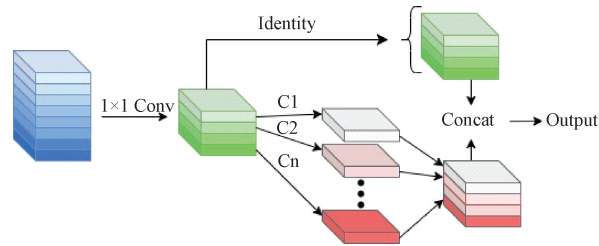


图 2 GhostConv 结构
Fig. 2 GhostConv structure

GHGBlock 包含多个特征处理层次,每个层次通过特定范围的感受野提取相应尺度的特征信息,从而实现对低级细节特征和高级语义特征的有效学习。其内部的 GhostConv 操作在提取基础特征的同时,还通过线性变换操作扩展基础特征得到更多复杂的特征,有效保留原始特征图关键信息。GHGBlock 公式如式(1)、(2)所示。

$$\begin{aligned}
Y_1 &= GHostconv_0(X_{in}) \\
Y_2 &= GHostconv_2(Y_1) \\
&\dots \\
Y_{n+1} &= GHostconv_{n+1}(Y_n) \\
Y &= Conv_1(Conv_0(X \parallel Y_1 \parallel \dots \parallel Y_{n+1} \parallel)) \\
Y_{out} &= X + Y
\end{aligned}
\tag{1}$$

$$\tag{2}$$

式中: X_{in} 表示输入; $Y_1, Y_2, \dots, Y$ 表示经过卷积之后的输出; Y_{out} 表示最终输出。GHGBlock 如图 3 所示。

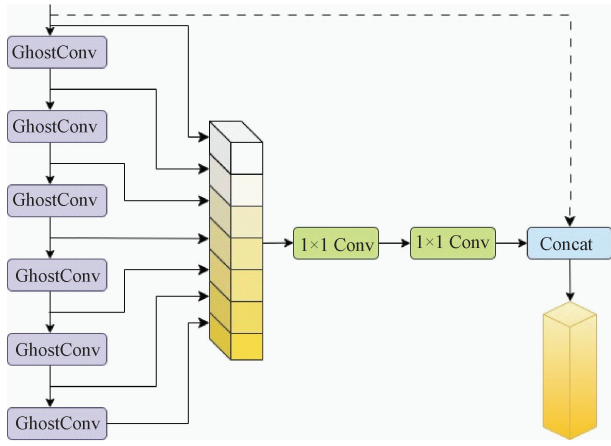


图 3 GHGBlock 结构

Fig. 3 GHGBlock structure

GHGNet 将 GhostConv 与 HGBlock 结合,通过“主通道 + 幻影特征”的机制减少冗余计算并保留关键特征信息;层次化多尺度特征提取机制使模型能从低层纹理到高层语义逐级学习。这种结合加强模型对金属表面小目标缺陷的感知与层次特征表达能力。

2.3 MSNet 模块

MANet^[18] 是一种用于图像处理的网络结构。StarNet^[19] 是一种基于星操作的神经网络模型。受到两种网络结构的启发,本文设计了 MSNet 模块,传统 C3k2 模块特征融合单一,梯度流不足,深层信息难以上传,导致小缺陷响应弱。MSNet 将 1×1 旁路卷积、深度可分离卷积与 Star 模块并行融合; 1×1 卷积完成通道校准,提升特征表达稀疏性;深度可分离卷积实现空间与通道解耦,提高计算效率;StarBlock 通过并行梯度分支构建无损梯度通路,缓解梯度消失。该结构的本质是特征融合多样化 + 梯度流富集,增强网络对不同尺度特征的敏感度。MSNet 在该模型中的优势如下。

1) 采用 1×1 旁路卷积可以融合不同通道的信息,实现通道间的非线性组合,增强特征表征能力;在深层网络中,不同通道可能学习到相似或者重复的特征,对通道式特征重新校准可以抑制低贡献通道同时保留高贡献通道,使特征分布更加稀疏且高效。

2) 深度可分离卷积由逐深度卷积(DWconv)和逐点卷积(PWconv)组成,逐深度卷积用于提取空间特征,逐点卷积用于提取通道特征。深度可分离卷积在特征维度上分组卷积,对各通道进行独立的 DWconv,并在输出前使用一个 PWconv 将所有通道进行聚合。相比于标准卷积耦合了空间滤波与通道交互,DSConv 通过分步操作实现高效解耦,既降低了冗余计算,又通过模块化设计增强了对复杂场景的适应性。

3) Star 模块采用特征分割与并行梯度处理机制,显著增强了神经网络中的梯度传播效率,提高了梯度流的丰富性和模型的检测精度。Star 模块通过 split 操作将特征解耦为两个独立分支,一个分支保持原始特征且不进行卷积操作直接向后传递。该操作在反向传播时能够保证完整的梯度流动,形成无损梯度通路,有效避免了因网络深度增加而引发的梯度消失问题。另一分支则使用 StarBlock 替换原有的 C2f 模块。在使用原有的 C2f 模块进行特征融合的时候,会导致非线性特征缺失。改进后的第 2 支路采用 StarBlock 模块替代传统 C2f 结构,针对特征融合过程中的非线性表达能力不足问题进行了优化创新。该模块独创的“星型运算”机制通过跨层特征的逐元素乘积操作,构建了独特的特征交互方法——在不扩展网络通道维度的情况下,将输入特征映射到高维非线性特征空间。该方法保证了在低维空间计算的同时考虑极高维的特征,为网络模型带来强大的特征融合能力。

同时该支路通过卷积操作进行特征变换,虽然造成了梯度衰减,但是得益于并行架构设计,这种衰减仅影响局部梯度。通过梯度流的双分支协同机制,Star 模块实现了深层网络中的动态梯度平衡,使梯度分布更加均匀,其独特的并行结构不仅优化了梯度分布均匀性,更重要的是确保了网络各层级都能获得充分的梯度反馈,从而在缺陷检测任务中展现出更优性能。MSNet 公式如式(3)、(4)所示。

$$\begin{aligned}
X_{mid} &= Conv_1(X_{in}) \\
Y_1 &= Conv_2(X_{mid}) \\
Y_2 &= DSConv(Conv_3(X_{mid})) \\
Y_3, Y_4 &= Split(X_{mid}) \\
Y_5 &= StarBlock_1(Y_4) + Y_4 \\
Y_6 &= StarBlock_2(Y_5) + Y_5 \\
&\dots
\end{aligned}
\tag{3}$$

$$\begin{aligned}
Y_{4+n} &= StarBlock_n(Y_{3+n}) + Y_{3+n} \\
Y_{out} &= Conv_0(Y_1 \parallel Y_2 \parallel \dots \parallel Y_{4+n})
\end{aligned}
\tag{4}$$

式中: X_{in} 表示输入; X_{mid} 表示第 1 次卷积输出; $Y_1, Y_2, \dots, Y_{4+n}$ 表示经过 3 条支路后的输出; Y_{out} 表示最终输出。MSNet 结构图如图 4 所示。

StarBlock 公式如式(5)、(6)所示。

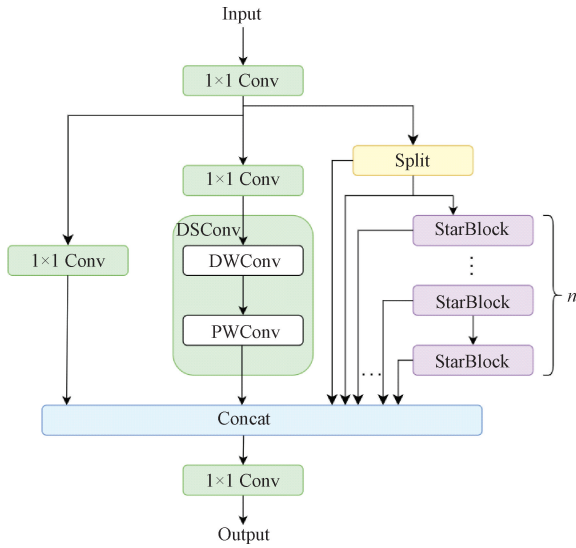


图4 MSNet 结构

Fig. 4 MSNet structure

$$Y_1 = \text{Relu6}(\text{Conv}_1(\text{DWConv}_1(X_{in})))$$

$$Y_2 = \text{Conv}_1(\text{DWConv}_1(X_{in})) \quad (5)$$

$$Y_3 = \text{DWConv}_2(\text{Conv}_2(Y_1 * Y_2))$$

$$Y_{out} = X_{in} + \text{DropPath}(Y_3) \quad (6)$$

式中: X_{in} 表示输入; Y_1 、 Y_2 、 Y_3 表示中间操作之后的输出; * 表示星操作; Y_{out} 表示最终输出。StarBlock 结构图如图5所示。

2.4 分离批归一化检测头

YOLOv11n 采用解耦头,将检测头分为两个分支,第1个分支会进行两个 $3 \times 3 \text{Conv}$ 操作和1个 $1 \times 1 \text{Conv}$ 操作

作,第2个分支采用两个深度可分离卷积操作和1个 $1 \times 1 \text{Conv}$ 操作,为了应对工业产品生产时的复杂环境,设计一个更为高效的检测头尤为重要。

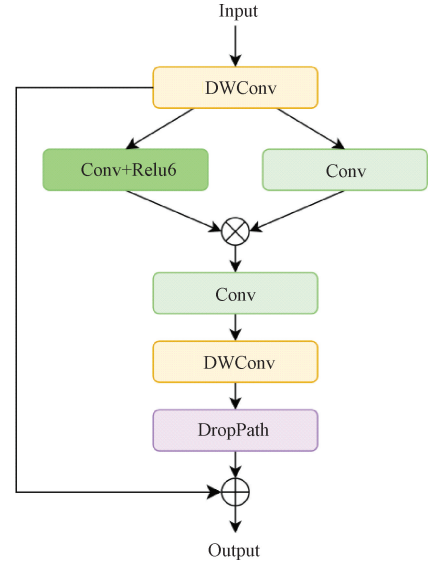


图5 StarBlock 结构

Fig. 5 StarBlock structure

YOLOv11n 原模型检测头采用传统单尺度预测结构,仅从一个尺度进行预测,从而忽略了其他尺度特征对检测的贡献,导致不同特征图之间的信息缺乏交互与相融,引起部分信息的丢失。为解决这个问题,本文设计了一个全新的检测头——SBN 检测头解决上述问题,其结构如图6所示。

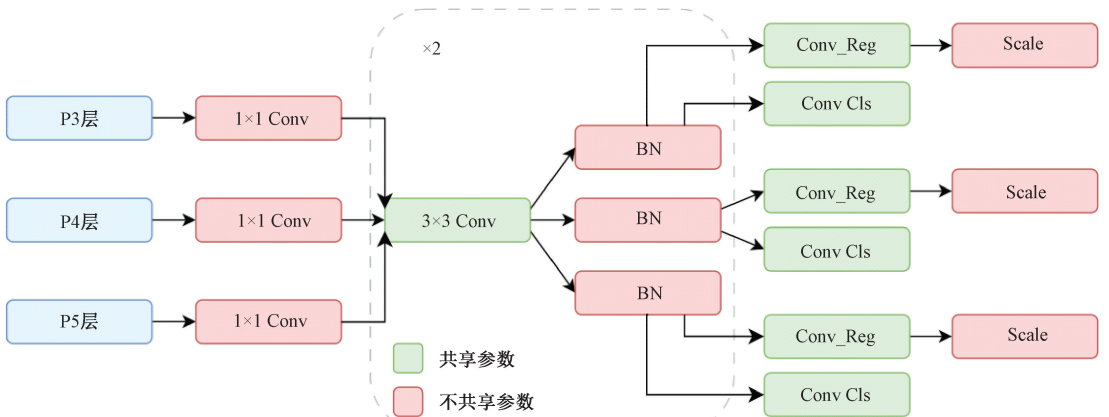


图6 SBN 检测头

Fig. 6 SBN detect

SBN 检测头将采用共享卷积模块替代传统检测头中的常规卷积模块。在共享卷积之后,不同层级之间特征的统计量会存在差异,为解决这个问题,在进行共享卷积

之后对其进行归一化处理。本文提出采用分层独立 BN (batch normalization) 机制解决该问题。直接采用 BN 会导致跨层统计偏差,本研究为每个检测层级构建独立批

归一化层,在保持检测精度的同时有效缓解多尺度特征分布偏移问题,同时减少不同特征层在统计量上的相互干扰,增强定位与分类任务的准确性。

SBN 检测头通过共享卷积 + 分层独立 BN 机制,在保证跨层特征交互的同时,为每个尺度构建独立统计量,从而消除分布偏移问题。BN 的独立化使网络在多尺度输入下保持稳定的均值方差估计,避免“跨层干扰”;共享卷积结构则强化了语义一致性。

2.5 损失函数改进

原 YOLOv11n 模型采用的损失函数是 CIoU,通过综合考虑预测框与真实框的长宽比、中心点距离以及重叠面积,可以在一定程度上提升对图像中目标的定位精度。在数据集中,样本质量的高低往往会影响到检测的精度,原模型 CIoU 损失函数在面对低质量样本时,如果一味地强化边界框对低质量示例的回归,会危害模型检测性能的提升。为解决该问题,本文将 YOLOv11n 的损失函数替换为 WIoU^[20],WIoU 是在 EIou 注意力机制基础上进一步发展的,通过动态方法来计算类别预测中的 IoU 损失。该损失函数通过使用动态非单调聚焦机制,提出利群度来替换传统 IoU 的评估指标。该策略通过抑制高质量样本的竞争优势,同时缓解低质量样本产生的负面影响,使 WIoU 能够将优化重心集中在中等质量样本上,从而有效提升检测模型的综合性能。WIoU 计算公式如式(7)所示。

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{S_u}$$

$$L_{WIoU} = r R_{WIoU} L_{IoU}$$

$$R_{WIoU} = \exp \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}$$

$$\beta = \frac{L_{IoU}}{L_{IoU}}$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}}$$

式中: W_i, H_i 表示预测框和真实框相交区域的宽度和高度; S_u 表示交叉区域的面积; IoU 表示真实框与预测框的重叠程度; L_{IoU} 表示 IoU 损失函数的值; x, y, x_{gt}, y_{gt} 表示预测框与真实框的中心坐标; W_g, H_g 表示预测框和真实框的最小外接框的宽度与高度; R_{WIoU} 表示权重因子; L_{WIoU} 表示加权 IoU 损失函数; β 表示离群度; r 表示梯度增益; α, δ 为超参数。

3 实验以及结果分析

3.1 数据集

本文通过自建数据集 Self-Dataset 上进行实验。该数据集源自某企业现场生产的工业产品,根据工厂实际

调查,选择实际生产中出现频率最高的 4 类金属表面缺陷,主要包括有锈迹、伤痕、凹坑、裂纹。为采集到金属表面缺陷数据集,搭建缺陷图像采集实验平台,该平台包括有工业相机,镜头,光源等设备,图像采集平台如图 7 所示,采集到的数据集共有 1 203 张缺陷图片,由于现场生产工业产品良品率高,为解决缺陷图片不足的问题,通过数据增强方式对数据集进行增强。增强后的数据集包含有 2 406 张缺陷图像。按照 7:1:2 的比例划分训练集,验证集,测试集。

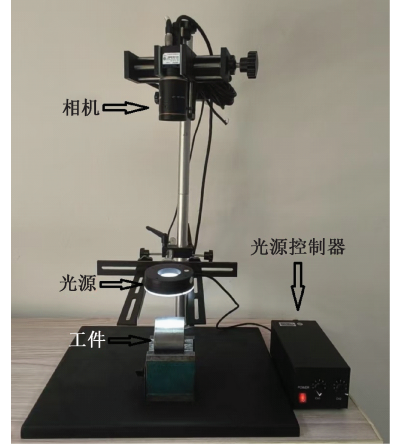


图 7 图像采集平台

Fig. 7 Image acquisition platform

3.2 实验环境

实验环境为 linux 系统,Python 版本为 3.10.15,PyTorch 版本为 2.3.0,CUDA 版本为 12.4,CPU 为 AMD EPYC7352,GPU 为 NVIDIA-GeForce-RTX-4090(24 G)。

模型训练参数设置如下:图片尺寸为 640×640,训练批次为 400epochs,批量大小 batchsize 为 64,优化器为 SGD,动量设置为 0.937,权重衰减为 0.000 5。

3.3 评价指标

为了评价 GMS2-YOLO 模型在 Self-Dataset 上的缺陷检测效果,选用的评价指标包括检测精度(precision, P)、召回率(recall, R)、平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)等。其计算公式如式(8)~(11)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$AP = \int_0^R P(R) dR \quad (10)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (11)$$

式中: TP 代表正确检测金属表面缺陷目标的数量; FP 代

表错误检测缺陷目标的数量; FN 代表漏检目标的数量; n 为缺陷目标的数量。

3.4 实验结果及分析

1) 损失函数对比实验

为了验证 WIoU 损失函数的优越性,对 YOLOv11n 网络使用一些主流的损失函数进行对比实验,如表 1 所示。实验结果表明,WIoU 损失函数更加适合金属表面缺陷检测,其总体检测效果最优,证明了 WIoU 损失函数的有效性。

为验证 WIoU 损失函数对边界框回归效率的影响,以 YOLOv11n 为基础模型,在只改变损失函数的前提下开展对比实验。将 WIoU 与 CIoU、Inner-SIoU、Shape-IoU、MpdIoU、Focal-IoU 损失函数进行对比,经 400 轮模型训练,绘制出不同损失函数对应的边界框回归损失曲线,如图 8 所示。WIoU 在 150 个 Epoch 内基本稳定,并且其损

表 1 损失函数对比

损失函数	mAP@ 0. 5/%
CIoU	77. 2
Inner-SIoU	76. 2
Shape-IoU	77. 0
MpdIoU	77. 2
Focal-IoU	77. 4
WIoU	78. 1

失值约为 0. 5。该损失函数通过动态非单调聚焦机制抑制高质量样本的过度优化,并提升中等样本的梯度权重,从而在存在噪声与低质量样本的数据集中依旧保持较高的收敛稳定性与检测精度。与其他损失函数比较,WIoU 在对复杂背景和模糊边界样本的可分性上更优,具有更强的鲁棒性。

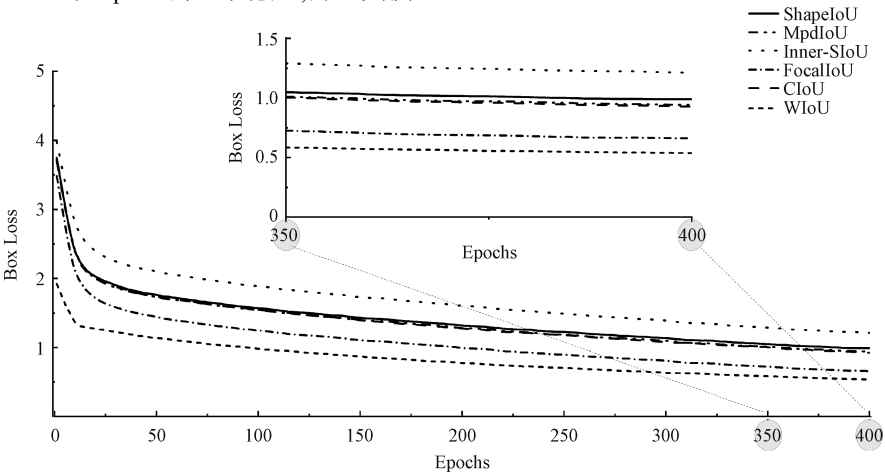


图 8 损失函数对比
Fig. 8 Comparison chart of loss functions

2) 消融实验

为了分析模型各个改进点对金属表面缺陷检测的有效性,在 YOLOv11n 的基础上设计了消融实验,即将改进的模块分别加入到 YOLOv11n 中,对每个模型得到的实验结果进行分析,如表 2 所示。表中“√”表示模型中使用了对应的模块,“×”表示模型中没有使用对应的模块。根据消融实验结果,采用改进后的网络模型,mAP@ 0. 5 从 77. 2% 提升至 80. 8%。在替换了主干网络后,模型的 mAP@ 0. 5 值由 77. 2% 提升至 79. 0%,这表明替换主干网络后,提升了模型的特征提取能力;在颈部网络融合 MSNet 模块之后,凭借其强大的特征融合能力以及丰富的梯度流信息,模型的 mAP@ 0. 5 值达到了 79. 5%;对检测头重设计之后,模型的 mAP@ 0. 5 值达到了 80. 2%,这表明新的检测头对检测任务有着更高的适配度;改进损失函数之后,模型的 mAP@ 0. 5 值达到了 80. 8%,说明

WIoU 损失函数对缺陷目标检测的优越性。改进模型相比于原模型,准确度提升 5. 5%,召回率提升 1. 2%,mAP@ 0. 5 提升 3. 6%,虽然其检测速度降低至 55 fps,但仍能满足工业检测需求,更好地应用于金属表面缺陷检测。

3) 对比实验

为了验证 GMS2-YOLO 模型的优越性和有效性,在相同的实验条件下,将目前主流的目标检测模型和改进模型在 Self-Dataset 上进行实验,使用检测精度、召回率、mAP@ 0. 5 和 mAP@ 0. 5: 0. 95,检测速度等多个指标进行比较,对比实验结果如表 3 所示。其次绘制精度与推理时间图直观感受各种模型的区别,精度与推理时间如图 9 所示。

由图 9 和表 3 可知,GMS2-YOLO 在精确度,召回率,mAP@ 0. 5,mAP@ 0. 5: 0. 95 4 个方面表现优异,精度

表 2 消融实验结果
Table 2 Results of ablation experiment

GHCnet	MSNet	SBN 检测头	WIoU	P/%	R/%	mAP@ 0. 5/%	mAP@ 0. 5;0. 95/%	检测速度/fps
×	×	×	×	78. 1	71. 3	77. 2	45. 0	67
✓	×	×	×	83. 3	73. 2	79. 0	50. 1	62
×	✓	×	×	83. 3	71. 1	77. 7	46. 4	60
×	×	✓	×	78. 0	69. 6	75. 4	45. 8	70
✓	✓	×	×	81. 8	75. 1	79. 5	50. 7	57
✓	✓	✓	×	80. 4	75. 1	80. 2	52. 4	55
✓	✓	✓	✓	83. 6	72. 5	80. 8	53. 5	55

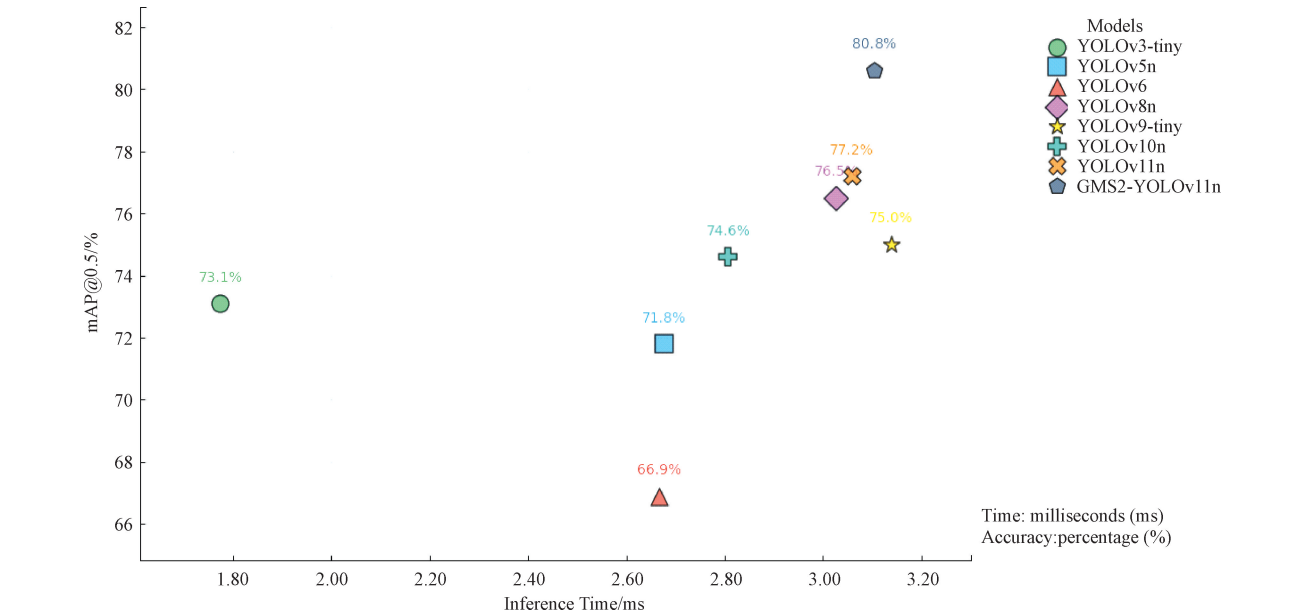


图 9 推理时间与模型精度
Fig. 9 Inference time vs model precision

表 3 对比实验结果
Table 3 Comparative experimental results

模型	P/%	R/%	mAP@ 0. 5/%	mAP@ 0. 5;0. 95/%	检测速度/fps	训练时间/h
YOLOv3-tiny	77. 8	67. 6	73. 1	43. 4	140	0. 927
YOLOv5n	69. 0	69. 8	71. 8	40. 7	74	0. 951
YOLOv6	68. 4	60. 7	66. 9	36. 4	78	0. 969
YOLOv8n	75. 8	70. 9	76. 5	44. 5	68	0. 984
YOLOv9-tiny	79. 3	70. 8	75. 0	45. 1	48	1. 099
YOLOv10	79. 2	67. 8	74. 6	45. 5	72	1. 004
Faster-RCNN	57. 8	42. 3	55. 2	25. 1	27	—
RT-DETR	69. 1	63. 1	66. 6	39. 2	123	—
YOLOv11n(Baseline)	78. 1	71. 3	77. 2	45. 0	67	0. 975
GMS2-YOLO	83. 6	72. 5	80. 8	53. 5	55	1. 027

达到 83.6%，召回率达到 72.5%，mAP @ 0.5 达到 80.8%，mAP@ 0.5;0.95 达到 53.5%。在检测速度方面，模型大小相比于原 YOLOv11n 仅减少 12 fps；在推理时间方面，大部分模型推理时间都在 2.6~3.2 ms 之间，改进模型在没有增加过多推理时间的情况下，mAP@ 0.5 有

着较高水平。综上所述 GMS2-YOLO 模型综合效果更优，在金属表面缺陷检测方面效果较为显著。

4) 热力图可视化分析

为了深入理解模型在目标检测中的工作原理，采用 EigenCAM^[21] 热力图可视化方法，对 YOLOv10n、

YOLOv11n 和 GMS2-YOLO 模型的检测结果进行了详细对比分析,探究模型在金属表面缺陷检测任务中的注意力权重分布特征。EigenCAM 是一种基于特征向量分析的类激活映射方法,该方法通过分析网络中间层的激活

图,生成热力图,直观展示输入图像中对模型预测影响最大的区域。生成的热力图以红色或深红色区域突出显示,表示模型对缺陷区域的关注程度,热力图结果如图 10 所示。

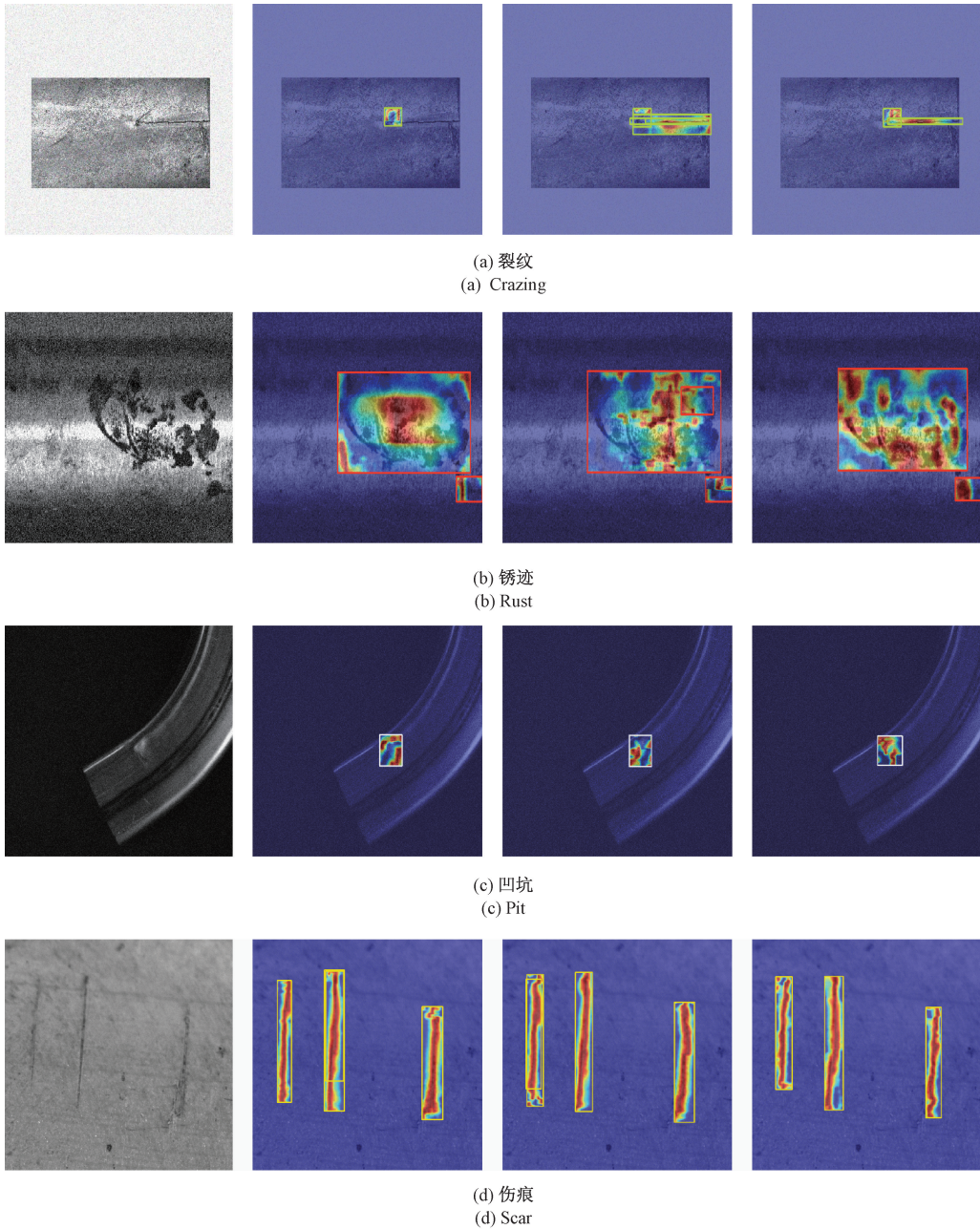


图 10 热力图

Fig. 10 Heatmap

根据热力图结果, YOLOv10n 的注意力存在有分散性过强的问题,难以精确定位关键区域,其次在面对复杂缺陷图像时,其定位能力和检测精度存在不足; YOLOv11n 在注意力分布方面取得了一定的改善,能够较为准确地聚焦于目标区域,但是其过于集中的注意力

导致了重复检测问题严重,由于其过于集中的注意力也导致注意力覆盖不完全,检测稳定性较差,部分细节特征未能被充分捕捉。由图 10(a)、(d) 可以看出, GMS2-YOLO 检测模型对于裂纹和伤痕缺陷,相比于 YOLOv10n 的漏检, YOLOv11n 的重复检测,改进模型能够识别缺陷

的同时,其深红色区域的范围更广,强度也更高;由图 10 (b)、(c)可以看出,锈迹属于复杂缺陷图像,锈迹分布广,分布不均匀,YOLOv10n 模型只能识别中心区域,YOLOv11n 模型相比于 YOLOv10n 有所改进,但是还是不能准确识别其分布,并且还有重复检测现象,改进模型能够准确识别锈迹缺陷。同时,其热力图分布也更加符合原图缺陷的分布情况。这表明改进后模型能够精确地定位并检测出缺陷,在复杂缺陷图像下表现尤其突出。

5) NEU-DET 数据集验证

为了验证改进模型中对微小缺陷目标检测的优势,在东北大学公开的数据集 NEU-DET^[22]进行了实验,NEU-DET 总共有 6 类缺陷,分别是裂纹(CR)、夹杂(IN)、斑块(PA)、麻点(PI)、氧化铁皮(RS)和划痕(SC),实验结果如表 4 所示。其中裂纹属于微小缺陷,改进模型相比于原模型,其 AP 值达到 68.4%,提高了 15.8%。

表 4 NEU-DET 实验结果							
Table 4 NEU-DET experimental results (%)							
模型	AP						mAP@0.5
	CR	IN	PA	PI	RS	SC	
YOLOv11n	52.6	83.1	93.1	77.1	69.7	92.2	78.1
GMS2-YOLO	68.4	81.6	90.0	78.3	64.9	94.3	79.6

6) GC-DET 数据集验证

为了证明改进模型的鲁棒性,在公开数据集 GC-DET 进行了实验,GC-DET 总共有 10 类缺陷,实验结果如表 5 所示,其 mAP@0.5 达到 65.8%,相比于原模型提高 1.6%;在检测速度方面,为 81 fps。实验结果表明模型具有较好的鲁棒性。

表 5 GC-DET 实验结果		
Table 5 GC-DET experimental results		
模型	mAP@0.5/%	检测速度/fps
YOLOv11n	64.2	97
GMS2-YOLO	65.8	81

4 结 论

本文围绕金属表面缺陷在实际工业场景中呈现出“小目标难检测、背景纹理干扰强、噪声显著影响”等长期未被充分解决的核心问题,构建了一个从主干网络、特征融合、检测头到损失函数的系统化优化框架,并据此提出了 GMS2-YOLO 模型。通过 GHGNet 实现细粒度表征能力增强,通过 MSNet 建立多尺度充分融合的特征表达,通过 SBN 检测头完成跨层级统计一致性与语义交互的

统一,通过 WIoU 提升模型在噪声样本与中等质量样本上的学习效率,使得整个检测体系在精度、稳定性与工程可部署性上均达到新的平衡。该方法的提出不仅验证了多尺度结构优化、统计归一化机制与动态损失设计在缺陷检测任务中的协同价值,也为小目标检测在复杂工业背景下的鲁棒性提升提供了一种可迁移的思路。

尽管模型在多项指标上取得了显著提升,但仍存在若干限制。例如,多分支与多模块结构在提升检测精度的同时带来了额外计算开销,在某些资源受限的边缘设备上仍需进一步压缩;此外,当前训练数据仍受场景类型、材料种类与采集条件的限制,对更极端噪声场景与更丰富缺陷类型的泛化能力仍需进一步验证。未来研究可围绕模型轻量化、跨场景迁移学习、少样本与零样本缺陷检测、缺陷细粒度刻画、结合物理先验的可解释结构设计等方向展开,并探索在工业产线中的实时闭环质量控制与自动化部署方案,从而推动金属表面缺陷检测走向更高的智能化与产业化水平。

参考文献

[1] 赵朗月,吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. 仪器仪表学报, 2022, 43(1): 198-219.

ZHAO L Y, WU Y Q. Research progress of surface defect detection methods based on machine vision[J]. Chinese Journal of Scientific Instrument, 2022, 43(1): 198-219.

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

[3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision-ECCV 2016. Berlin: Springer, 2016: 21-37.

[4] REN S, HE K, 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, 2016, 39(6): 1137-1149.

[5] HE K, GKIOXA RIG, DOLLA R P, et al. Mask R-CNN[C]. IEEE International Conference on Computer Vision (ICCV), 2017: 2961-2969.

[6] 向宽,李松松,栾明慧,等. 基于改进 Faster RCNN 的铝材表面缺陷检测方法[J]. 仪器仪表学报, 2021, 42(1): 191-198.

XIANG K, LI S S, LUAN M H, et al. Aluminum product surface defect detection method based on improved Faster R-CNN [J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 191-198.

- [7] 马志程,李丹,张宝龙. 基于改进 Mask R-CNN 的光学元件划痕缺陷检测研究[J]. 电子测量与仪器学报, 2023, 37(4): 231-239.
MA ZH CH, LI D, ZHANG B L. Research on scratch defect detection of optical elements based on improved Mask R-CNN [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(4): 231-239.
- [8] 赵云亮,唐东林,何媛媛,等. 基于 CNN 融合 PGW-Attention 的金属表面缺陷识别方法[J]. 电子测量与仪器学报, 2024, 38(8): 46-55.
ZHAO Y L, TANG D L, HE Y Y, et al. Metal surface defect recognition method based on CNN with PGW-Attention [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 46-55.
- [9] 高嘉,刘涛,王显峰,等. TWRD-Net: 一种用于曳引钢丝绳缺陷的实时检测网络算法[J]. 仪器仪表学报, 2023, 44(6): 223-235.
GAO J, LIU T, WANG X F, et al. TWRD-Net: A real-time detection network algorithm for traction wire rope defects [J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 223-235.
- [10] 马燕婷,赵红东,阎超,等. 改进 YOLOv5 网络的带钢表面缺陷检测方法[J]. 电子测量与仪器学报, 2022, 36(8): 150-157.
MA Y T, ZHAO H D, YAN CH, et al. Strip steel surface defect detection method by improved YOLOv5 network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(8): 150-157.
- [11] 赵佰亭,张晨,贾晓芬. ECC-YOLO: 一种改进的钢材表面缺陷检测方法[J]. 电子测量与仪器学报, 2024, 38(4): 108-116.
ZHAO B T, ZHANG CH, JIA X F. ECC-YOLO: An improved method for detecting surface defects in steel [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 108-106.
- [12] 赵林熔,甄国涌,储成群,等. 基于 CBE-YOLOv5 的钢材表面缺陷检测方法[J]. 电子测量技术, 2023, 46(15): 73-80.
ZHAO L R, ZHEN G Y, CHU CH Q, et al. Detection method of steel surface defects based on CBE-YOLOv5 [J]. Electronic Measurement Technology, 2023, 46(15): 73-80.
- [13] 胡波,王佳欣,杨青,等. 基于改进 YOLOv5 的铝锭合金表面缺陷检测技术研究[J]. 电子测量技术, 2024, 47(14): 121-126.
HU B, WANG J X, YANG Q, et al. Improved YOLOv5-based surface defect detection technology for aluminum ingot alloys [J]. Electronic Measurement Technology, 2024, 47(14): 121-126.
- [14] 王素珍,吕基岳,邓成禹,等. 基于改进 YOLOv5 的钢铁表面缺陷检测算法[J]. 国外电子测量技术, 2023, 42(10): 43-50.
WANG S ZH, LV J Y, DENG CH Y, et al. Improved YOLOv5-based surface defect detection algorithm for steel [J]. Foreign Electronic Measurement Technology, 2023, 42(10): 43-50.
- [15] 赵小虎,谢礼逊,慕灯聪,等. 基于 TCM-YOLO 网络的金属表面缺陷检测方法[J]. 计算机工程, 2025, 51(6): 338-348.
ZHAO X H, XIE L X, MU D C, et al. Metal surface defect detection based on TCM-YOLO network methods [J]. Computer Engineering, 2025, 51(6): 338-348.
- [16] ZHAO Y A, LYU W Y, XU S L, et al. Detsr beat yolos on real-time object detection [J]. ArXiv preprint arXiv: 2304. 08069, 2023.
- [17] HAN K, WANG Y, XU C, et al. GhostNets on heterogeneous devices via cheap operations [J]. International Journal of Computer Vision, 2022, 130(4): 1050-1069.
- [18] FENG Y, HUANG J, DU S, et al. Hyper-YOLO: When visual object detection meets hypergraph computation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2388-2401.
- [19] MA X, DAI X, BAI Y, et al. Rewrite the Stars [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694-5703.
- [20] TONG Z J CHEN Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv preprint arXiv: 2301. 10051, 2023.
- [21] MUHAMMAD M B, YEASIN M. EIGEN-CAM: Class activation map using principal components [C]. 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-7.
- [22] HE Y, SONG K C, MENG Q G, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69 (4): 1493-1504.

作者简介



严赫, 2023 年于三峡大学获得学士学位, 现为武汉理工大学硕士研究生, 主要研究方向为缺陷检测、机器视觉。

E-mail: 346602@whut.edu.cn

Yan He received his B. Sc. degree from China Gorges Three University in 2023. Now

he is a M. Sc. candidate of Wuhan University of Technology. His main research interests include defect detection and machine vision.



史晓亮, 2003 年于中国地质大学(武汉)获得博士学位, 现为武汉理工大学教授、博导, 主要研究方向为机器视觉识别、智能装备系统设计与优化、先进摩擦材料设计。

E-mail: sxl@whut.edu.cn

Shi Xiaoliang received his Ph. D. degree from China University of Geosciences, Wuhan, China, in 2003. Now he is a professor and Ph. D. supervisor at Wuhan University of Technology. His main research interests include machine vision recognition, intelligent equipment system design and

optimization, advanced friction material design.



吴超华(通信作者), 2020 年于武汉理工大学获得博士学位, 现为武汉理工大学副教授、博导, 主要研究方向为机器视觉、智能制造系统设计与工艺优化、材料界面优化设计与性能调控。

E-mail: wuchaohua@whut.edu.cn

Wu Chaohua (Corresponding author) received his Ph. D. degree from Wuhan University of Technology in 2020. Now he is an associate professor and Ph. D. supervisor at Wuhan University of Technology. His main research interests include machine vision, intelligent manufacturing system design and process optimization, material interface optimization design and performance control.