

文章编号: 1006-3080(2025)05-0657-14

DOI: 10.14135/j.cnki.1006-3080.20241210004

基于 GSL-YOLO 的钢材表面缺陷检测算法

王泽恺, 姜庆超, 颜学峰

(华东理工大学信息科学与工程学院, 上海 200237)

摘要: 钢材在现代工业生产中因加工和环境等因素易产生表面缺陷, 影响其质量和使用寿命。针对现有检测方法存在的精度不足与计算复杂度较高的问题, 本文提出了一种基于 YOLOv8 的改进算法 GSL-YOLO, 针对特征冗余和特征表达能力不足的问题, 在 C2f 模块中引入 Ghost Bottleneck 替代标准 Darknet Bottleneck, 减少冗余计算并提高特征提取效率, 采用 DynamicConv 替换标准 CBS 模块, 以增强对复杂场景的适应能力。为进一步提升模型对细小缺陷的检测精度, 在颈部网络中引入 SENetV1 注意力机制, 通过自适应调整通道权重, 加强对关键特征的关注。此外, 针对检测头参数量大、计算复杂度高的问题, 设计轻量级共享卷积检测头 LSDECD, 有效降低模型参数量和计算开销。实验结果表明, GSL-YOLO 在 NEU-DET 钢材表面缺陷数据集上的 mAP50 较 YOLOv8 提升 3%, 同时参数量减少 33%, 计算成本降低 37%。

关键词: 缺陷检测; YOLOv8; Ghost Bottleneck; DynamicConv; 注意力机制; 轻量级检测头

中图分类号: TP391

文献标志码: A

在现代工业生产中, 钢材作为重要的基础材料, 被广泛应用于建筑、交通、制造等多个领域。在生产和使用过程中, 钢材常因高温、机械加工和环境因素等多种影响, 导致表面产生裂纹、划痕、麻点等缺陷^[1]。这些缺陷不仅降低了钢材的外观质量, 还可能显著削弱其力学性能和使用寿命。因此, 及时、准确地识别和检测这些缺陷是保证钢材质量的重要前提。

传统的检测方法, 如人工视觉检查和电磁检测^[2], 由于受主观因素的影响, 往往难以实现高精度、高效率的检测。与此同时, 随着生产流程的加快和生产规模的扩大, 仅依赖人工检测的方法已无法满足现代工业的需求。因此, 迫切需要引入新型检测技术, 以提升检测的效率和准确性^[3]。

近年来, 计算机视觉和深度学习技术凭借非接触式检测和快速响应的优势, 已逐渐成为钢材表面缺陷检测的主流方法。其中, 卷积神经网络(CNN)在目标检测任务中展现出强大的性能, 广泛应用于钢材表面缺陷检测领域^[4]。目标检测算法主要分为

两阶段算法和单阶段算法两类, 两阶段检测算法通过区域建议网络生成候选区域, 并在候选区域上进行分类和边界框调整, 典型算法包括 FasterR-CNN^[5]等, 单阶段算法则基于特征图和锚框预测每个位置的类别和边界框坐标, 代表性算法有 SSD^[6]、RetinaNet^[7]、YOLO^[8] 系列等。尽管两阶段算法在检测精度方面优于单阶段算法, 但其训练和推理速度较慢, 效率相对较低, 难以满足工业场景中对实时检测的需求。而单阶段算法由于其高效性和端到端设计, 在钢材表面缺陷检测等对实时性和快速处理需求较高的工业场景中具有更大的应用潜力^[9]。

尽管单阶段检测算法在实时性和效率方面具有优势, 但以 YOLO 系列为代表的模型在计算成本和检测精度上仍存在不足, 其在复杂工业场景中对小目标和复杂背景的检测能力有待进一步提升。近年来, 针对 YOLO 模型的优化研究得到了广泛关注, 并提出了多种改进方案。Liang 等^[10] 通过引入 BotNet 自注意力机制, 结合 GSConv 轻量化卷积和 VoV-

收稿日期: 2024-12-10

基金项目: 装备预研教育部联合基金(8091B032230); 国家自然科学基金优秀青年基金(62322309)

作者简介: 王泽恺(2001—), 男, 河北石家庄人, 硕士生, 主要研究方向为计算机视觉。E-mail: christopher_wzk@163.com

通信联系人: 姜庆超, E-mail: qchjiang@ecust.edu.cn; 颜学峰, E-mail: xfyang@ecust.edu.cn

引用本文: 王泽恺, 姜庆超, 颜学峰. 基于 GSL-YOLO 的钢材表面缺陷检测算法[J]. 华东理工大学学报(自然科学版), 2025, 51(5): 657-670.

Citation: WANG Zekai, JIANG Qingchao, YAN Xuefeng. Steel Surface Defect Detection Algorithm Based on GSL-YOLO[J]. Journal of East China University of Science and Technology, 2025, 51(5): 657-670.

GSCSP 模块优化颈部结构,显著提升了 YOLOv5 模型在金属冲压件缺陷检测中的性能。Shuai 等^[11]提出了 SF-YOLO 模型,采用 CBAM-C3 注意力机制、BiFPN 特征融合策略和改进麻雀搜索算法优化超参数,提高了齿轮端面缺陷检测的精度和效率。Yang 等^[12]提出了 DCS-YOLOv8 模型,通过引入可变形卷积增强多尺度特征提取能力,并结合 CBAM 与 SCSE 模块优化通道与空间特征融合,显著提高了钢材表面缺陷检测的精度和效率。陶友凤等^[13]提出了 RCD-YOLO 模型,通过轻量化骨干网络 R_HGNet 降低模型复杂度,同时结合 CARAFE 上采样算子和 C2f-DWR 模块增强多尺度特征融合,并引入 Inner-CIOU 损失函数优化小目标定位精度,实现了钢材表面缺陷检测精度和效率的显著提升。赵佰亭等^[14]提出了 ECC-YOLO 模型,该模型通过引入 ConvNeXt 特征增强模块提升特征提取能力,设计 C2fFB 轻量化模块降低计算成本,并采用 ECA 注意力机制增强对复杂背景的适应性。该模型在保证检测精度的同时,实现了更优的模型效率和轻量化。吴亚尉等^[15]提出了 YOLO-GR 模型,在 YOLOv5s 基础上,引入 GhostNet 模块减少参数量,采用深度可分离卷积优化特征融合,并在检测头部分集成 RepLK 大卷积核以增强感受野。此外,引入 W-IoU 损失函数,通过平衡质量较好与较差样本的学习,增强模型对数据集的泛化能力。

尽管上述方法在提高钢材表面缺陷检测精度、增强特征表达能力以及优化特征融合等方面取得了一定进展,但这些改进往往伴随着计算复杂度的显著增加,难以满足对资源受限环境下轻量化部署的需求。因此,本文设计了一种钢材表面缺陷检测改进算法 GSL-YOLO。该算法通过设计 C2f-Ghost-Dynamic 模块替换主干中的 C2f 模块,有效提高了特征提取效率,在 Neck 网络结合通道注意力 SENetV1,增强模型对关键特征的专注能力。最后,采用轻量共享卷积检测头 LSDECD 替换基线模型的检测头,显著减少了模型参数量和计算开销。

1 钢材表面缺陷检测算法

1.1 基线模型 YOLOv8

本文对基线模型 YOLOv8 进行了改进,其网络结构如图 1 所示,主要包括 3 部分: Backbone、Neck 和 Head。

Backbone 由 CBS 模块、C2f 模块和 SPPF 模块组成,负责从输入图像中提取特征,其中, CBS 模块

结合卷积层、批归一化和 SiLU 激活函数, C2f 模块通过特征图的拆分和拼接提高特征提取效率, SPPF 模块利用快速空间金字塔池化捕捉多尺度信息。

Neck 采用 PAN-FPN 架构,融合来自 Backbone 不同阶段的特征图,以增强特征表示能力。PAN-FPN 通过路径聚合网络实现自底向上和自顶向下的特征融合,提升模型对不同尺度目标的检测能力。

Head 由大、中、小 3 个尺度的检测头组成,采用无锚框设计,并将每个检测头解耦为回归头和分类头,分别用于预测目标的边界框、置信度和类别信息。无锚框设计通过直接预测目标中心点,简化了训练过程并提高了检测效率。

1.2 改进模型 GSL-YOLO

鉴于钢材表面缺陷检测中存在背景复杂、小目标难以检测等挑战,且基线 YOLOv8 模型存在特征图冗余、参数量大、计算成本高、通道间特征依赖性不足等问题,本文提出以下改进方案:

(1)设计 C2f-Ghost-Dynamic 模块替换主干中的 C2f 模块,该模块结合 Ghost 模块的廉价操作生成冗余特征图,并根据不同输入动态调整卷积核权重,增强特征提取的灵活性和适应性,从而在降低计算量的同时,提高模型对钢材表面缺陷的特征表达能力。

(2)在颈部网络中引入通道级注意力机制 SENetV1,通过对通道特征的自适应校准动态调整各通道的重要性,使得关键通道特征得到增强,而无关或冗余特征被抑制。该机制通过显式建模通道间的依赖性,强化了多尺度特征的融合效果,显著提升了模型在复杂背景下对小目标和细节特征的检测能力。

(3)替换基线模型中的检测头为轻量共享卷积检测头(LSDECD),通过共享卷积层设计,减少模型的参数量和计算开销。共享卷积部分采用细节增强卷积(Detail-enhanced convolution),提升共享卷积层对特征细节的捕捉能力,在保证轻量化的同时,增强模型的检测精度和推理性能。

改进后的网络结构如图 2 所示。

1.2.1 C2f-Ghost-Dynamic 加强特征提取 C2f 模块在特征提取过程中存在计算冗余、特征表征能力不足的问题,为了降低计算开销并增强模型对复杂场景和细小目标的表征能力,本文基于 Ghost 模块^[16]和动态卷积^[17](Dynamic convolution)的设计思想对 C2f 模块进行了改进。具体而言,通过将动态卷积与 Ghost 模块相结合,设计了 C2f-Ghost-Dynamic 模块,并通过实验验证了其在特征提取能力和计算效率上的优越性。

动态卷积是一种自适应计算方法,能够根据输

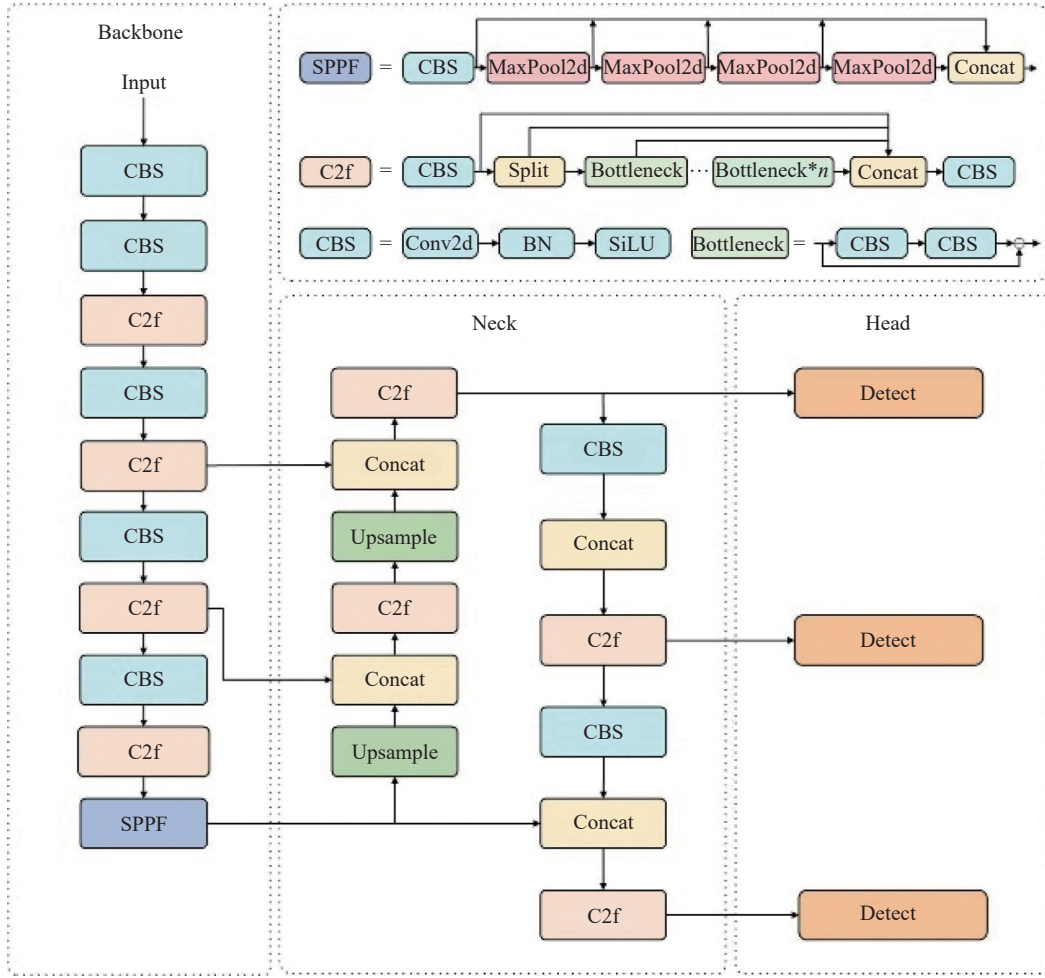


图1 基线模型 YOLOv8

Fig. 1 Baseline model YOLOv8

入特征的不同,自适应地选择多个不同的卷积核进行处理。具体来说,DynamicConv 模块首先对输入数据 $X \in \mathbf{R}^{C \times H \times W}$ 进行全局平均池化,将信息融合为一个特征向量,然后使用一个两层的全连接网络(MLP)生成权重 α ,这些权重将用于加权多个专家卷积核(Expert convolution),如式(1)所示:

$$\alpha = \text{softmax}(\text{MLP}(\text{Pool}(X))) \quad (1)$$

然后,使用生成的权重 $\alpha(i)$ 来动态组合不同的专家卷积核 W_i ,最终的输出可以表示为:

$$Y = \sum_{i=1}^M \alpha_i (X * W_i) \quad (2)$$

其中, M 是专家卷积核的数量, W_i 是第 i 个专家卷积核, $*$ 表示卷积操作,DynamicConv 模块结构图如图3所示。该机制使每个输入都能够动态选择最优的卷积核组合进行处理,从而提高了特征提取的灵活性与精度。而且动态卷积仅引入了轻微的计算开销,它在实际应用中模型效率的影响较小,但对模型性能的提升却十分显著。

传统的卷积层在目标检测任务中需要大量的参数和浮点运算,针对这个问题,Ghost 模块通过一种创新的低成本特征生成方法显著降低了计算开销,工作流程如图4所示。Ghost 模块的工作流程主要分为两个阶段:先是通过标准的操作生成较少的内在特征图:

$$Y' = X * W' \quad (3)$$

其中, W' 是卷积核, Y' 是生成的内在特征图。然后对内在特征图 Y' 进行简单的线性变换,生成额外的 Ghost 特征图:

$$y_{ij} = \Phi_{i,j}(y_i'), \forall i = 1, \dots, m; j = 1, \dots, s \quad (4)$$

最后,所有 Ghost 特征图与内在特征图进行拼接,形成最终的输出特征图:

$$Y = [y_{11}, y_{12}, \dots, y_{ms}] \quad (5)$$

该模块显著减少了冗余计算量,同时提高了特征表征能力,在此基础上,本文设计了由两个堆叠 Ghost 模块组成的 GhostBottleneck 模块,其结构如图5所示。第1个 Ghost 模块作为扩展层,通过增加

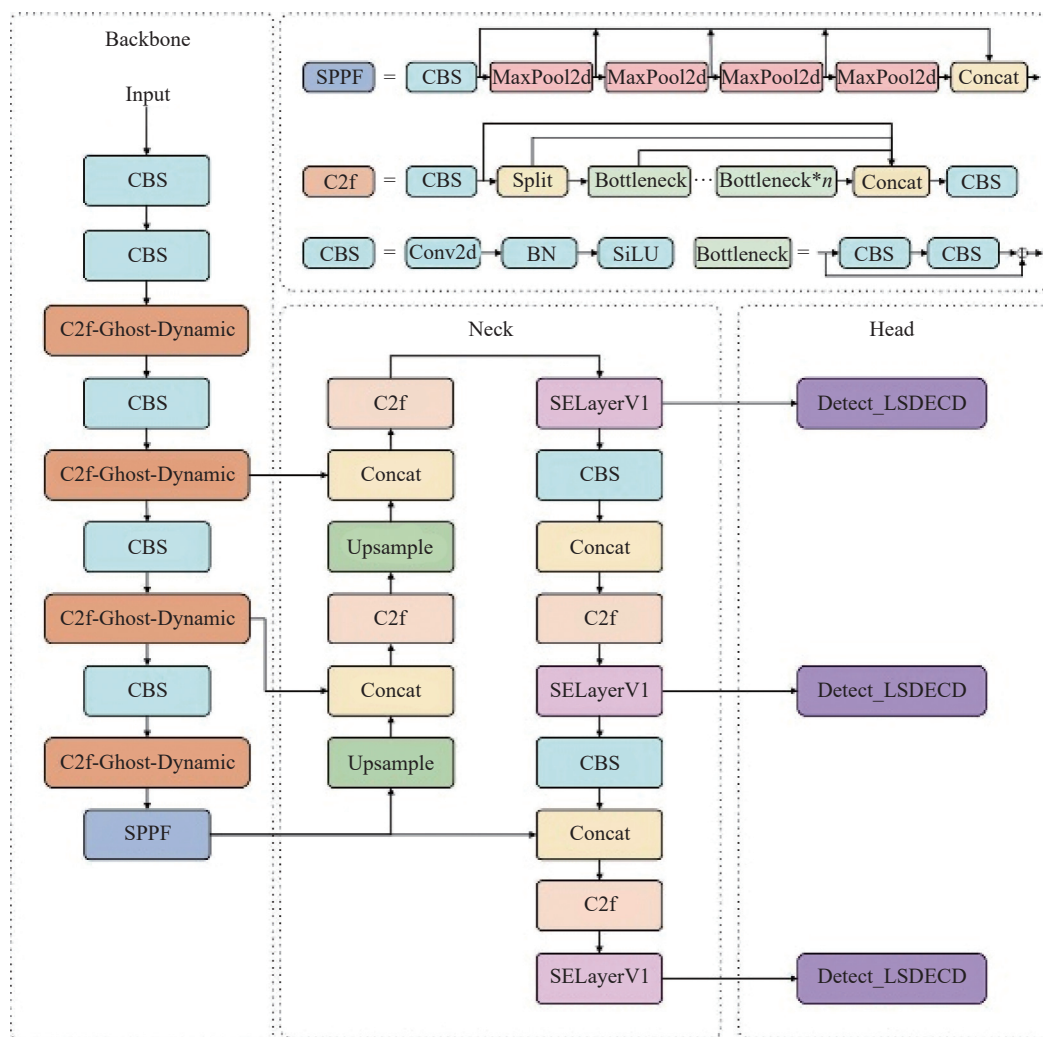


图 2 改进的 GSL-YOLO 结构

Fig. 2 Structure of improved GSL-YOLO

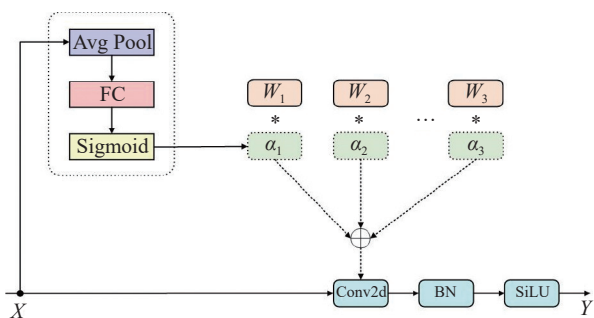


图 3 DynamicConv 模块结构

Fig. 3 Structure of DynamicConv module

通道数增强特征表征能力;第 2 个 Ghost 模块作为收缩层,通过减少通道数与快捷路径相匹配,从而控制模型复杂性并确保计算效率。两个 Ghost 模块之间加入快捷连接,用于促进信息传递和梯度流动,从而提升训练效果并优化模型性能。

为了增强 Ghost 模块的特征适应能力和计算效率,本文在其结构设计中引入了动态卷积和分组卷积的思想。Ghost 模块结构如图 6 所示,由两个动态

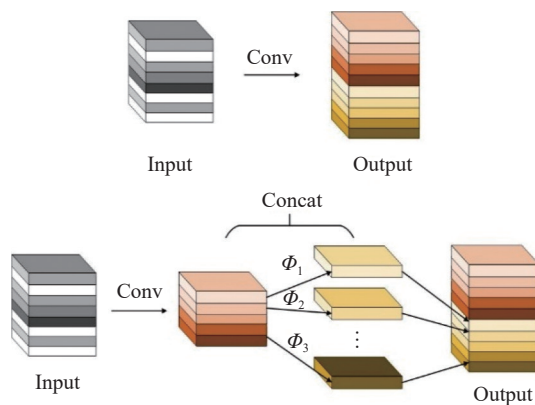


图 4 传统卷积层与 Ghost 模块的对比

Fig. 4 Comparison between traditional convolutional layer and Ghost module

卷积分支和特征拼接操作构成。主分支通过 1×1 动态卷积对输入特征进行通道变换和降维,生成内在特征(Intrinsic features);廉价分支则通过 3×3 分组动态卷积,对输入通道分组,每个分组的通道仅与对应的输出通道连接,两个分支均采用 BatchNorm 和 ReLU

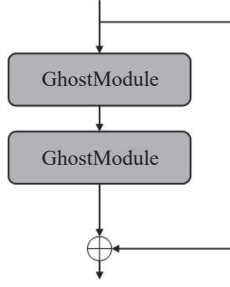


图 5 GhostBottleneck 模块结构

Fig. 5 Structure of GhostBottleneck module

进行特征归一化与非线性变换。最终,将主分支生成的内在特征与廉价分支生成的幽灵特征在通道维度上进行拼接(Concat),形成完整的特征表示。这种结构设计通过动态卷积提升了特征的自适应能力,同时利用分组卷积优化了计算资源的使用效率。

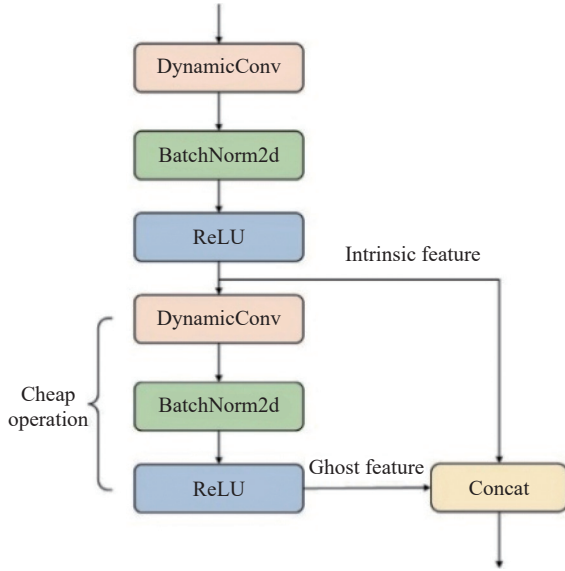


图 6 Ghost 模块结构

Fig. 6 Structure of Ghost module

基于上述改进,本文提出了 C2f-Ghost-Dynamic 模块,其核心在于结合动态卷积与 Ghost 模块的优势,通过设计改进的 GhostBottleneck 模块替代传统 C2f 模块中的 DarknetBottleneck 结构。与传统设计相比,C2f-Ghost-Dynamic 模块能够在显著优化计算资源使用的基础上,增强模型对复杂场景和多样化输入的适应能力。改进后的 C2f-Ghost-Dynamic 结构如图 7 所示。

1.2.2 SENetV1 自适应校准通道权重 基线模型 YOLOv8 颈部网络在特征融合过程中未能有效区分通道的重要性,导致对关键特征的表达能力有限。为解决这个问题,本文借鉴挤压-激励机制^[18](Squeeze-and-Excitation)构建 SENetV1 模块,通过动态调整通道权重,强化了模型对关键特征的关注能

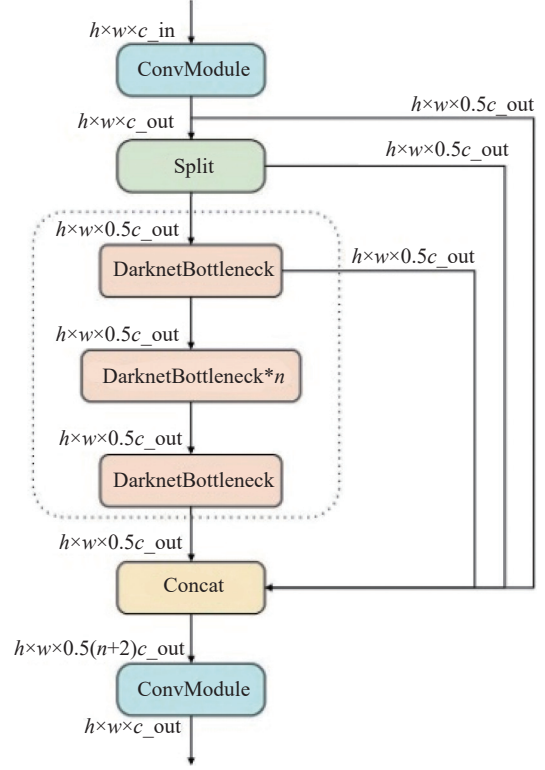


图 7 C2f-Ghost-Dynamic 模块结构

Fig. 7 Structure of C2f-Ghost-Dynamic module

力,并优化了多尺度特征融合的效果。

SE 模块通过建模通道间的依赖关系,实现对通道权重的自适应校准。具体而言,如图 8 所示,SE 模块分为挤压(Squeeze)和激励(Excitation)两部分。挤压操作通过全局平均池化将每个通道的信息压缩为全局描述符,从而提取特征图的全局分布信息。SE 模块对特征图 $U \in \mathbf{R}^{H \times W \times C}$ 进行全局池化,生成一个大小为 $1 \times 1 \times C$ 的描述符 z_c :

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

其中, z_c 为第 c 个通道的全局描述符。 $u_c(i, j)$ 表示输入特征图在空间位置 (i, j) 的值, H 和 W 分别表示特征图的高度和宽度。

在生成全局描述符 z_c 后,激励操作通过两个全连接层对通道权重进行动态调整。首先,降维操作通过缩减特征向量的维度以降低计算复杂度;然后,通过升维恢复到原始通道数,生成每个通道的动态权重 s :

$$s = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 z)) \quad (7)$$

其中, $\mathbf{W}_1$ 和 $\mathbf{W}_2$ 分别是降维和升维的全连接层权重矩阵, δ 为 ReLU 激活函数, σ 为 Sigmoid 激活函数。最终生成的通道权重 s 用于动态调整特征的重要性,根据权重 s 对原始特征图 u_c 进行逐通道加权,

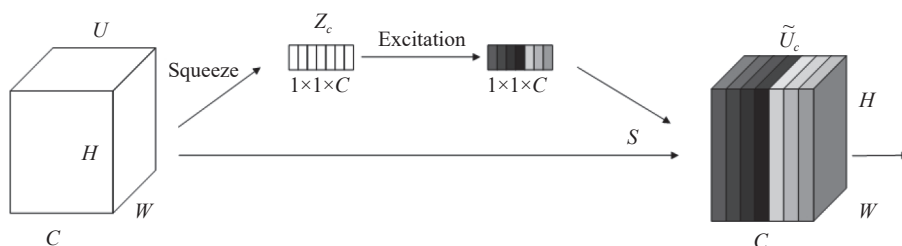


图 8 挤压-激励机制

Fig. 8 Squeeze-and-Excitation

形成校准后的特征图:

$$\tilde{u}_c = s_c \cdot u_c \quad (8)$$

其中, $\tilde{u}_c$ 表示校准后的特征图, s_c 为第 c 个通道的权重, u_c 为原始特征图。

本文在颈部网络的每个 C2f 模块后引入挤压-激励模块, 如图 9 所示, 构建了 C2f-SENetV1 模块。作

为一种通道级注意力机制, SENetV1 通过全局平均池化提取通道全局描述符, 并利用全连接网络动态生成权重, 对通道特征进行自适应校准。该机制显式建模通道间的依赖关系, 增强了网络对关键特征的关注意, 同时抑制了冗余信息的影响。在多尺度目标检测任务中, SENetV1 通过优化特征表达和融合, 显著提升了网络对不同尺度目标的检测能力与整体精度。

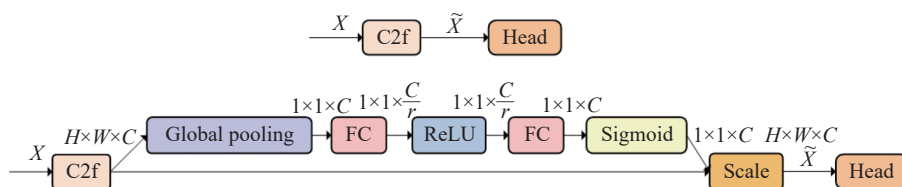


图 9 原始 C2f 模块和 C2f-SENetV1 模块

Fig. 9 Original C2f module and C2f-SENetV1 module

1.2.3 轻量级共享卷积检测头 在 YOLOv8 原始检测头中, 每个检测层独立执行卷积操作, 这种设计在保持较高检测精度的同时, 也导致了模型参数量庞大和计算复杂度较高的问题。这一限制显著影响了其在嵌入式设备等资源受限环境中的实际应用。为解决这个问题, 本文提出了一种轻量级共享卷积检测头 (Lightweight Shared Detail-Enhanced Convolutional Detection Head) LSDECD。通过共享卷积和细节增强卷积的结合, 该设计在显著降低参数量和计算复杂

度的同时, 有效提升了对小目标和复杂背景的检测能力, 其结构如图 10 所示。

LSDECD 的设计核心在于共享卷积模块与细节增强卷积^[19]模块的协同作用。共享卷积通过在检测头的多个分支中复用卷积参数, 显著减少冗余计算并降低了模型的参数量, 从而实现了计算资源的高效利用。这种方法不仅优化了模型的推理速度, 还增强了参数权重的泛化能力, 使模型能够在资源受限的嵌入式设备上满足实时运行的需求。此外, 为

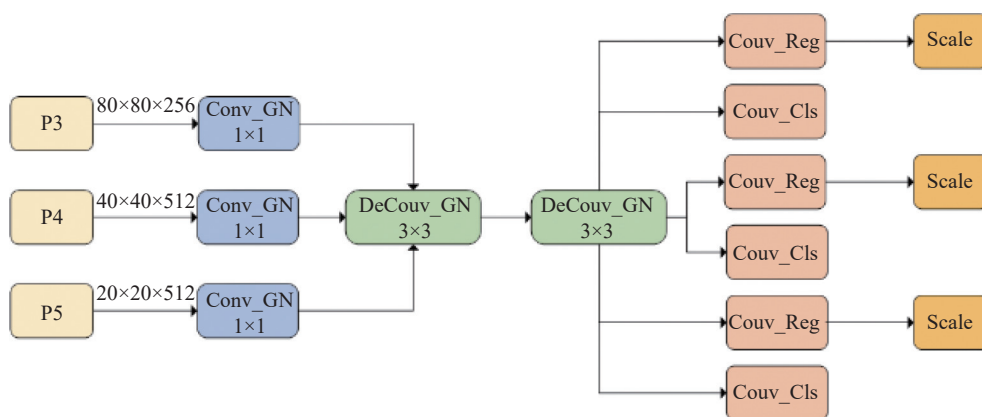


图 10 LSDECD 模块结构

Fig. 10 Structure of LSDECD module

了解决传统检测头在处理多尺度目标时存在的适配性不足问题, LSDECD 在每个检测层中引入了尺度层(Scale layer)。尺度层通过动态线性变换调整特征幅值和偏移, 确保共享卷积生成的特征能够适应不同尺度目标的检测要求。

轻量化设计可能导致特征提取能力的下降, 从而对模型的检测精度产生不利影响。为了弥补这个问题, LSDECD 引入了细节增强卷积模块(Detail-Enhanced Convolution, DeConv)。该模块结合标准卷积和多种差分卷积, 通过捕捉输入特征的局部梯度信息, 显式建模特征梯度变化, 从而提升了模型对细节特征的感知能力。DeConv 模块由 5 个并行路径组成, 包括标准卷积 Conv2d 和 4 种差分卷积(Difference convolution): 中心差分卷积 CDC、水平差分卷积 HDC、垂直差分卷积 VDC 和角度差分卷积 ADC, 其结构如图 11 所示。

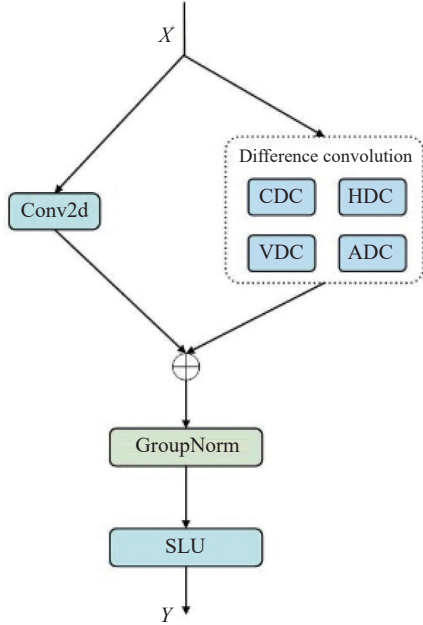


图 11 DEConv 模块结构

Fig. 11 Structure of DEConv module

细节增强卷积的输出特征可通过以下数学表达式表示:

$$F_{out} = \sum_{i=1}^5 F_{in} * K_i \quad (9)$$

其中, F_{in} 是输入特征图, K_i 表示第 i 路卷积核。

为进一步提升检测性能, LSDECD 在共享卷积和细节增强卷积模块中引入组归一化(Group Normalization, GN), 替代传统的批归一化(Batch Normalization, BN)。GN 通过分组归一化操作减少对批量大小的依赖, 在小批量训练条件下具有更高的稳定性, 降低了梯度爆炸和消失的风险^[20]。此外,

GN 优化了特征分布, 满足多尺度目标的特征表示需求, 并通过稳定梯度优化路径, 在复杂背景和小目标检测中提升了特征提取和目标定位的性能。

综上所述, LSDECD 通过共享卷积模块的高效特征复用、细节增强卷积对局部特征的精细化提取, 以及组归一化方法对分类和定位能力的优化, 有效提升了模型在小目标和复杂背景场景下的检测性能, 同时降低了模型的计算复杂度。这种设计为资源受限环境中的目标检测任务提供了一种高效且精确的解决方案。

2 结果与分析

2.1 数据集

本文基于东北大学公开发布的钢材表面缺陷数据集 NEU-DET 进行模型训练与测试。NEU-DET 数据集包含 6 种常见的钢材表面缺陷类别, 分别为裂纹(Crazing)、夹杂(Inclusion)、斑块(Patch)、点蚀面(Pitted surface)、轧制氧化皮(Rolled-in scale)和划痕(Scratch), 共 1800 张图像, 每种缺陷类别均包含 300 张样本图像, 图像分辨率统一为 640×640 像素。为了确保模型的训练和验证效果, 实验按照 7:2:1 的比例对数据集进行划分, 即训练集 1260 张图像, 验证集和测试集分别包含 360 张和 180 张图像。这样的划分方式保证了数据分布的一致性, 同时为模型的泛化能力评估提供了足够的测试样本。部分数据集的图像样例如图 12 所示。

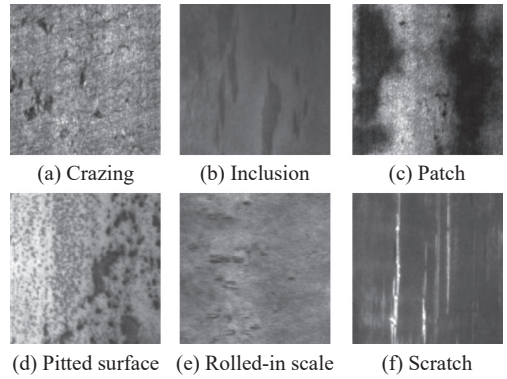


图 12 6 类缺陷示例

Fig. 12 Examples of six types of defects

2.2 实验环境与评价指标

本文网络程序基于 Pytorch1.13.1 深度学习框架实现, 对应的 CUDA 版本号为 11.6, 编程语言版本为 Python3.9。实验所用硬件设备为一台高性能服务器, 该服务器配备 NVIDIA GeForce RTX4090, 显存为 24 GB, 操作系统版本为 Ubuntu18.04。在模型训练过程中, 未采用任何预训练权重, 选择从零开始训

练 YOLO 模型。在训练中使用 Mosaic 数据增强技术,通过随机选择 4 张图像进行拼接,以增强训练样本的多样性并提升模型的鲁棒性。输入模型的图像在训练中被自动缩放至 640×640 的固定尺寸,以满足网络输入的尺寸要求并确保计算效率。实验参数设置如表 1 所示。

为了评估测试集上的模型性能,实验使用平均精度(AP)、平均精度均值(mAP)、模型参数量(Parameter)、计算开销(GFLOPs, GFLOPs 表示一次前向推理所需的浮点运算次数)、帧率(FPS)、精确率(Precision)、召回率(Recall)作为性能评估指标。部分指标计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$AP = \int_0^1 P(R)d(R) \quad (12)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (13)$$

其中, AP 表示精确率-召回率(Precision-Recall, P-R)曲线下的面积; mAP 表示在预测框与真实框交并比(IoU)超过设定阈值时,对数据集中所有类别的 AP 值取均值得到的指标; TP 表示正确检测的数量, FP 表示误报的数量, FN 表示漏报的数量, n 表示检测缺陷的类别数量。

2.3 实验结果及分析

2.3.1 消融实验 为验证改进模块对钢材表面缺陷检测的影响,本文设计了 4 组消融实验,每组实验采用相同的训练参数和环境条件,实验结果如表 2 所示。

表 1 实验参数

Table 1 Experimental parameters

Epoch	Batch size	Learning rate	Momentum	Optimizer	Input shape	Random seed
400	64	0.01	0.937	SGD	200×200	42

表 2 消融实验

Table 2 Ablation experiments

Method	mAP50/%	Parameter/M	GFLOPs/G	FPS
YOLOv8	72.5	3.00	8.1	62.2
YOLOv8+C2f-Ghost-Dynamic	73.7	2.58	6.7	60.4
YOLOv8+C2f-Ghost-Dynamic+SENetV1	74.6	2.60	6.7	60.2
YOLOv8+C2f-Ghost-Dynamic+SENetV1+LSDECD	75.5	2.00	5.1	59.9

基线模型 YOLOv8 的 mAP50 为 72.5%, 参数量为 3.00 M, GFLOPs 为 8.1 G, FPS 为 62.2。在引入 C2f-Ghost-Dynamic 模块后, mAP50 提升至 73.7%, 参数量和 GFLOPs 分别降低至 2.58 M 和 6.7 G, 但推理速度略微下降至 60.4, 表明该模块在显著减少计算复杂度的同时有效提升了特征提取能力。进一步引入 SENetV1 模块后, mAP50 提升至 74.6%, 参数量略微增加至 2.60 M, 表明注意力机制能够有效增强模型对关键特征的专注。在整合 C2f-Ghost-Dynamic、SENetV1 和 LSDECD 模块后, 模型的 mAP50 达到 75.5%, 相较基线模型提升了 3%, 各类缺陷检测的 mAP50 值几乎都有一定程度的提升, 如图 13 所示。参数量和 GFLOPs 分别减少至 2.00 M 和 5.1 G, 相较基线模型分别降低 33% 和 37%, 但推理速度下降至 59.9。实验结果表明, 各改进模块在降低模型复杂度

的同时有效提升了检测精度。尽管 FPS 出现了小幅下降, 但仍在可接受的实时检测范围内, 证明所提出的改进设计在检测性能与计算效率之间实现了良好的平衡。

图 14 对比了 YOLOv8 与 GSL-YOLO 在 NEU-DET 数据集上的 P-R 曲线, 展示了改进模型的整体

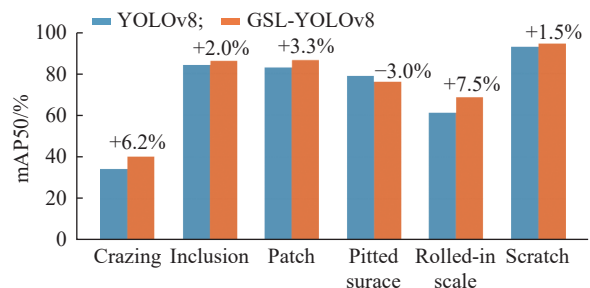


图 13 NEU-DET 数据集缺陷检测 mAP50 值对比

Fig. 13 mAP50 values of defects detected on NEU-DET dataset

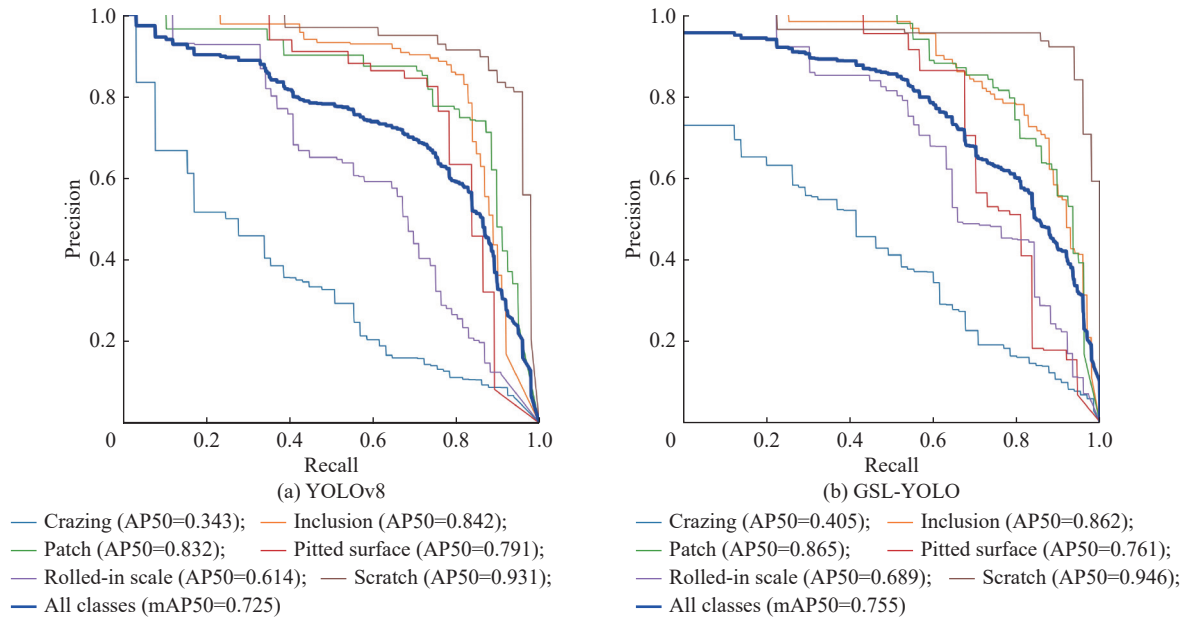


图 14 改进前后 P-R 曲线对比

Fig. 14 P-R curve comparison before and after improvement

性能优势。P-R 曲线下的面积反映了平均精度 AP 值的大小, 面积越大, 表明模型在精确率和召回率之间的平衡性能越优。相比基线模型 YOLOv8, GSL-YOLO 在复杂背景和小目标类别的检测中表现出更高的精确率和召回率, 优化了难检测类别的性能。此外, 改进模型在所有类别的检测结果中表现更均衡, 显著提高了整体检测性能的可靠性。

2.3.2 不同注意力机制对比实验 为评估不同注意力机制在目标检测任务中的性能表现, 本文对基线模型融入多种先进的注意力机制进行了对比实验, 主要包括 CA^[21]、LSKA^[22]、还有本文使用的 SENetV1, 结果如表 3 所示。

表 3 融合不同注意力模块对比实验

Table 3 Comparison of experimental results with different attention mechanisms

Method	mAP50/%	Parameter/M	GFLOPs/G	FPS
YOLOv8	72.5	3.00	8.1	62.2
YOLOv8+CA	72.4	3.05	8.2	60.4
YOLOv8+LSKA	74.0	3.17	8.4	57.7
YOLOv8+SENetV1	74.4	3.02	8.1	60.3

实验结果表明, 不同注意力机制对模型性能的提升效果存在一定差异。基线模型的 mAP50 为 72.5%, 参数量为 3.00 M, GFLOPs 为 8.1 G, 推理速度为 62.2。引入 CA 模块后, 各项指标均和基线模型保持接近, 表明 CA 模块对于模型的提升较为有限。相比之下, LSKA 模块显著提升了检测精度, 使 mAP50

达到 74.0%, 但略微提高了计算开销和参数量, 推理速度也有一定程度的下降。SENetV1 模块在多种注意力机制中表现出最优的综合性能, 检测精度提高至 74.4%, 在多尺度特征融合和通道注意力方面表现优异, 此外, SENetV1 模块在轻量化设计上表现优异, 其参数量更少, 仅为 3.02 M, 推理速度为 60.3。这表明 SENetV1 能够在提升检测精度的同时, 有效平衡模型计算复杂度与推理速度。

本文通过生成热力图, 将模型在输入数据不同区域的注意力分布进行可视化, 以评估其在钢材表面缺陷检测任务中的泛化能力, 并分析模型对关键区域的敏感性。如图 15 所示, 热力图结果表明 SENetV1 能够精准聚焦于复杂背景中的缺陷区域, 如裂纹、划痕的线性特征, 以及点蚀面和轧制氧化皮等小目标区域。相比之下, CA 模块的注意力分布较为分散, 未能充分突出关键区域的特征。LSKA 模块更倾向于捕捉大目标的整体轮廓, 对细节特征的响应能力相对较弱。进一步分析表明, SENetV1 通过显式建模通道间的依赖性, 增强了网络对关键特征的关注, 同时有效抑制了冗余信息的影响, 其精确的注意力机制提升了模型在复杂场景中的检测能力。这种特性不仅显著提高了模型对缺陷区域的关注强度和检测精度, 还增强了其对不同尺度目标的适应性, 为钢材表面缺陷检测任务提供了可靠且高效的解决方案。

2.3.3 不同数据集对比实验 为了验证本文提出的 GSL-YOLO 算法的有效性和优越性, 本文选用公开数据集 GC10-DET 进行对比实验, 实验在相同的硬

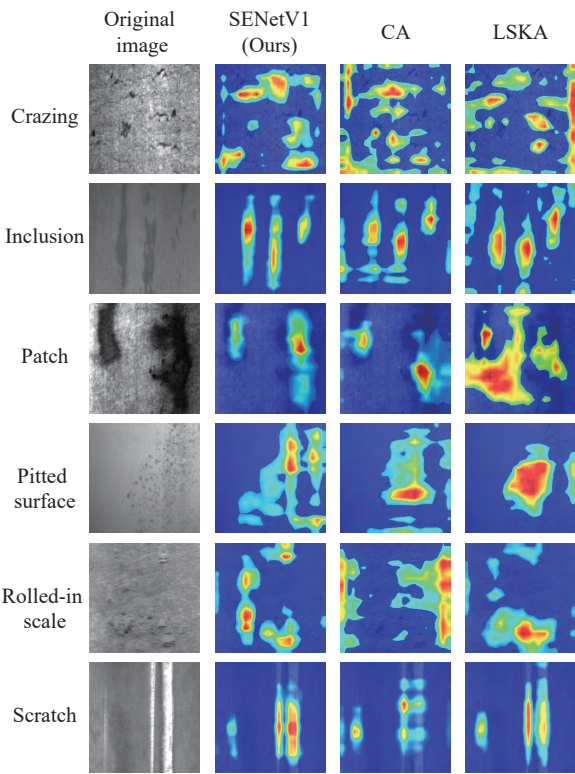


图 15 注意力机制热力图对比

Fig. 15 Heatmaps of attention mechanism comparison

件环境和超参数设置下, 比较了基线模型 YOLOv8 与改进后的 GSL-YOLO 模型在该数据集上的检测效果, 实验结果如表 4 所示。

从表 4 可以看出, 改进后的 GSL-YOLO 模型在

表 4 GC10-DET 数据集性能指标对比

Table 4 Comparison of GC10-DET dataset performance index

Model	mAP50/%	Parameter/M	GFLOPs/G	FPS
YOLOv8	57.5	3.00	8.2	71.0
GSL-YOLO	62.0	1.95	5.2	68.8

检测精度、模型轻量化和计算效率方面均优于基线模型 YOLOv8。具体而言, GSL-YOLO 的 mAP50 提升至 62.0%, 较基线模型提高了 4.5%。同时, GSL-YOLO 的参数量显著减少至 1.95 M, 较基线模型减少了 35%, GFLOPs 由 8.2 G 降至 5.2 G, 降低了 36.6%。尽管 FPS 从 71.0 略微下降至 68.8, 但仍满足工业实时检测的需求。以上结果表明, GSL-YOLO 在保证实时性的同时, 通过模块改进实现了检测精度与模型复杂度的良好平衡, 适用于资源受限的嵌入式设备和工业场景下的缺陷检测任务。

2.3.4 不同网络模型对比实验 为验证 GSL-YOLO 模型在钢材表面缺陷检测任务中的综合性能, 本文选择几种主流目标检测算法进行对比实验, 包括两阶段检测模型 Faster R-CNN、一阶段检测模型 RetinaNet、SSD300 以及多种 YOLO 系列模型。这些模型覆盖 Anchor-based 和 Anchor-free 两种预测模式, 从框架特性和实际性能两个维度全面评估其适用性, 实验结果如表 5 所示。

表 5 不同缺陷检测方法在 NEU-DET 数据集上的比较结果

Table 5 Comparison results of different defect detection methods on the NEU-DET dataset

Method	Backbone	AP/%						mAP50/%	Parameter/M	GFLOPs/G
		Crazing	Indusion	Patch	Pitted surface	Rolled-in scale	Scratch			
Faster R-CNN	ResNet50 ^[23]	34.1	84.2	87.6	68.1	55.7	95.4	70.8	41.4	134
RetinaNet	ResNet50 ^[23]	42.7	76.8	78.0	68.7	59.9	70.9	66.2	36.4	129
SSD300	VGG16 ^[24]	40.9	69.2	90.6	91.4	65.3	66.0	70.6	24.4	31
YOLOv5	CSPDarkNet(C3) ^[25]	36.7	83.5	84.3	73.7	59.1	96.1	72.2	2.5	7.1
YOLOv8	CSPDarkNet(C2f) ^[25]	34.1	84.2	83.2	79.1	61.4	93.1	72.5	3.0	8.1
YOLOv10	CSPDarkNet(C2f) ^[25]	38.6	76.7	88.3	71.8	63.7	93.5	72.1	2.3	6.5
YOLOv11	CSPDarkNet(C3k2) ^[25]	41.4	79.7	73.2	81.2	67.3	80.4	72.8	2.6	6.3
GSL-YOLO		40.4	86.2	86.5	76.1	68.9	94.6	75.5	2.0	5.1

实验表明, 传统两阶段检测模型 Faster R-CNN 在检测精度上表现良好, 其 mAP50 达到 70.8%, 但由于依赖区域提议网络 RPN 进行目标定位, 计算开销较高且模型复杂, 限制了其在轻量化场景中的应用。相比之下, RetinaNet 和 SSD300 等一阶段检测模型通过直接预测目标框提高了推理效率, GFLOPs 分

别降低至 129 G 和 31 G, 参数量也分别降至 36.4 M 和 24.4 M。然而, 这两种模型的整体检测性能较弱, 尤其在复杂场景下易受到背景干扰的影响。近年来广泛应用的 YOLO 系列模型在轻量化和实时性方面具有优势。实验结果显示, YOLOv5 和 YOLOv8 的 mAP50 分别达到 72.2% 和 72.5%, 显著优于传统一阶

段检测模型,同时保持了较低的计算复杂度。然而,这些模型在平衡检测精度与轻量化方面仍存在一定局限,尚难以完全满足工业场景对高效性能与低计算资源的需求。

相比之下, GSL-YOLO 在检测精度和模型复杂度方面均表现出显著优势。其 mAP50 达到 75.5%, 相比 YOLOv8 提升 3%, 同时参数量减少至 2.0 M, GFLOPs 降低至 5.1 G, 均为表中最低水平。这表明, GSL-YOLO 在提升检测精度的同时,进行了轻量化设计,能够在资源受限环境中兼顾高检测精度与低计算复杂度。为进一步验证其性能,本文在 NEU-DET 数据集上选取多种主流检测方法进行对比实验,检测结果如图 16 所示。实验表明, GSL-YOLO 在各类缺陷的检测任务中均表现出更高的检测精度,其预测结果与 GroundTruth 更加接近,尤其在复杂背景和小目标场景中表现优异。相比 RetinaNet, GSL-YOLO 有效改善了边界框重叠问题,并显著减少了漏检现象;与 Faster R-CNN 和 SSD300 相比,

GSL-YOLO 在裂纹、划痕等线性缺陷的检测中表现精准;在点蚀面和轧制氧化皮等小目标检测任务中, GSL-YOLO 的检测精度优于 YOLOv8 和 YOLOv11。这些结果进一步证明, GSL-YOLO 在复杂工业场景中具有较高的适用性与稳定性。

2.3.5 复杂场景适应性与检测效率分析 为了进一步检验改进后的 GSL-YOLO 模型在钢材表面缺陷检测任务中的适应性和鲁棒性,本文针对 NEU-DET 数据集进行了多种形式的增强^[26],包括噪声添加、随机遮挡、亮度调整、对比度变化及分辨率缩放,以模拟工业生产过程中可能出现的传感器噪声、光照变化和遮挡等复杂场景,因此本文分别引入高斯噪声与椒盐噪声,采用亮度和曝光度的调整来适配不同光照条件。针对钢材表面可能受到灰尘或污物遮挡的情况,采用了随机遮挡策略。此外,考虑到不同生产线所使用的摄像设备可能导致图像质量的差异,进一步引入了对比度调节和分辨率缩放,以全面评估模型在多样化输入条件下的泛化能力。

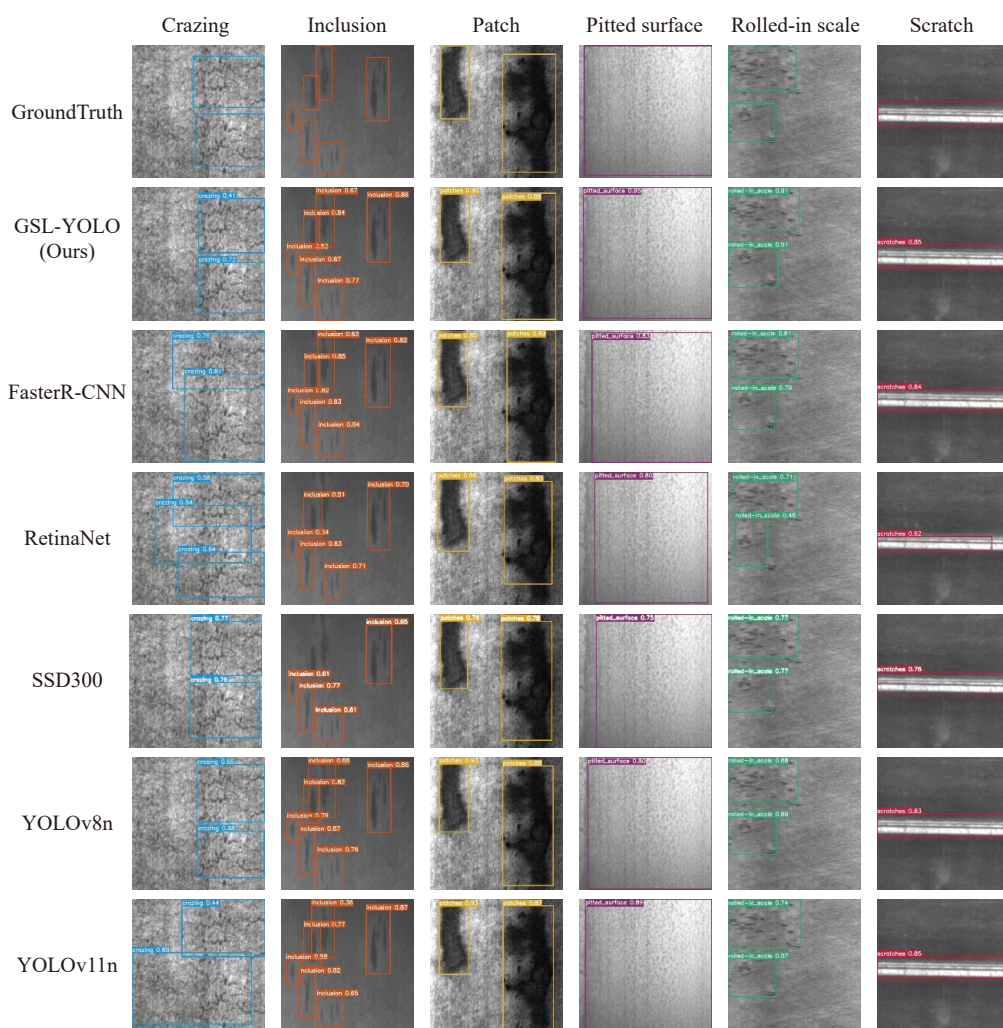


图 16 NEU-DET 数据集上多种方法检测结果可视化对比

Fig. 16 Visualization comparison of detection results using various methods on the NEU-DET dataset

为了验证 GSL-YOLO 在复杂场景下的泛化能力和适应性,本文在不同数据条件下进行了对比实验,实验采取 4 种数据配置:原始数据集、原始增强数据集、小样本数据集和小样本增强数据集。其中,数据增强指的是对训练集扩充 50%,扩充方式为在 5 种数据增强策略中随机组合 1~3 种,而验证集和测试集保持不变。小样本数据指的是从原始数据集中按比例分层随机抽取 30% 的样本。在上述基础上,分别评估 GSL-YOLO 与 YOLOv8 在 mAP50 指标上的检测性能,并对实验结果进行分析,实验采用统一的训练超参数设置,包括学习率、批量大小和训练轮次,以保证实验结果的可比性。实验结果如表 6 所示。

表 6 多种数据配置下模型适应性比较

Table 6 Comparison of model adaptability under various data configurations

Dataset type	Sample size	mAP50/%	
		YOLOv8	GSL-YOLO
Original	1 800	72.5	75.5
Original (Augment)	2 430	72.9	75.7
Small	540	64.0	65.5
Small (Augment)	729	64.5	66.7

实验结果表明,在原始数据集条件下,GSL-YOLO 的 mAP50 达到 75.5%,相比 YOLOv8 的 72.5% 提升 3.0%,体现了其在充足数据条件下对钢材表面缺陷的优越检测能力。随着数据集扩充至 2430 张,GSL-YOLO 的 mAP50 达到 75.7%,而 YOLOv8 则从 72.5% 上升至 72.9%,这表明数据增强在丰富样本多样性的同时,能够维持改进模型的稳定性能。在小样本条件下,GSL-YOLO 的 mAP50 为 65.5%,相较于 YOLOv8 的 64.0% 有一定提升,进一步采用数据增强后,GSL-YOLO 的 mAP50 提升至 66.7%,性能提升的幅度趋于平缓,表明在数据受限的情况下,模型的性能提升有限。整体而言,GSL-YOLO 在不同数据条件下均表现出良好的检测能力和较强的适应性,这表明改进模型在数据较少或检测环境较复杂的情况下,依然能够保持较好的检测效果。

除了数据增强,钢材表面缺陷检测系统的实际运行效率还受到预处理流程的影响。检测流程通常包括图像采集、预处理、模型推理和结果输出,其中预处理主要涉及去噪、归一化和格式转换等操作。虽然这些步骤有助于提高模型精度,但也增加了计算开销,进而影响系统的实时性。本文实验环境中预处理平均耗时约 15 ms,具体结果如表 7 所示。

表 7 检测时间对比

Table 7 Comparison of detection time

Model	Time/ms		
	Preprocessing	Inference	Total detection
YOLOv8	15	16.1	31.1
GSL-YOLO	15	16.7	31.7

实验数据显示,GSL-YOLO 的推理时间为 16.7 ms,略高于 YOLOv8 的 16.1 ms,总检测时间由 31.1 ms 增加至 31.7 ms。尽管推理时间有所增加,但 GSL-YOLO 的检测速度仍能满足实际生产流程中的运行速度。

3 结束语

本文提出了一种基于 GSL-YOLO 的钢材表面缺陷检测改进算法,针对传统目标检测算法在检测精度和计算复杂度之间平衡不足的问题,通过多种模块优化实现了显著改进。具体而言,设计了 C2f-Ghost-Dynamic 模块替代主干网络中的 C2f 模块,通过结合 Ghost 模块与动态卷积的特点,在降低计算开销的同时显著提升了特征提取能力;在颈部网络中引入通道注意力 SENetV1 模块,通过自适应调整通道权重增强了对关键特征的关注;并采用轻量级共享卷积检测头 LSDECD 替代传统检测头,有效减少了参数量和计算复杂度,同时提高了检测性能。

GSL-YOLO 在 NEU-DET 钢材表面缺陷数据集上的 mAP50 达到 75.5%,相比基线模型 YOLOv8 提升了 3%,同时参数量和计算复杂度分别减少了 33% 和 37%。尽管 GSL-YOLO 推理速度略有下降,并且还需考虑实际生产中对于图像的预处理时间,但整体仍能满足工业场景中的实时需求。与多种主流目标检测算法的对比结果表明,GSL-YOLO 在复杂背景场景和小目标检测任务中表现出较高的检测精度,进一步验证了其对小目标和复杂环境的适应能力。在 GC10-DET 数据集的测试中,GSL-YOLO 同样展现了良好的检测性能,证明了其在不同数据集间的泛化能力。在针对复杂场景的实验中,当应用数据增强策略时,模型展现出较强的检测能力和适应性。在小样本条件下,模型性能的提升趋于平缓,但在数据有限时依然能够保持较稳定的检测效果。

本文在检测精度、计算复杂度和实时性方面取得了良好进展,但在适应不同生产需求和工况方面,模型的适配性和鲁棒性仍需提升。工业生产对检测系统的性能要求各不相同,在检测速度、精度和计算

资源方面存在差异化需求。例如高速生产线对检测系统的实时性要求较高,需依托高性能 GPU 的并行计算能力,以实现高帧率和低时延的检测性能。而小批量精密检测则更加注重缺陷识别的高精度,可以通过增加模型的网络深度和宽度进一步提升对复杂缺陷的检测能力。在资源受限的嵌入式场景中,硬件功耗限制是主要挑战,当前研究主要依靠专用算子库优化计算效率,并通过硬件加速方案降低推理时的资源消耗。

针对这些挑战,未来的研究将围绕模型的参数优化以及适配不同环境需求的能力展开深入探索。通过剪枝、量化与知识蒸馏等技术,优化模型结构并减少计算负担。同时,结合针对性的预处理和数据增强策略,提升模型在复杂工业场景中的泛化能力。最终,通过设计通用且高效的跨硬件适配方案,使 GSL-YOLO 能够在不同硬件环境下保持稳定的检测性能,满足工业生产对效率和精度的综合要求。

参考文献:

- [1] 米春风,卢琨,汪文艳,等.基于机器视觉的带钢表面缺陷检测研究进展[J].安徽工业大学学报(自然科学版),2022,39(2):180-188.
- [2] 李少波,杨静,王铮,等.缺陷检测技术的发展与应用研究综述[J].自动化学报,2020,46(11):2319-2336.
- [3] 陶显,侯伟,徐德.基于深度学习的表面缺陷检测方法综述[J].自动化学报,2021,47(5):1017-1034.
- [4] 卢宏涛,张秦川.深度卷积神经网络在计算机视觉中的应用研究综述[J].数据采集与处理,2016,31(1):1-17.
- [5] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot multibox detector[C]//Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [7] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2980-2988.
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [9] 张慧,王坤峰,王飞跃.深度学习在目标视觉检测中的应用进展与展望[J].自动化学报,2017,43(8):1289-1305.
- [10] LIANG Y, FENG S, ZHANG Y, *et al.* A stable diffusion enhanced YOLOV5 model for metal stamped part defect detection based on improved network structure[J]. *Journal of Manufacturing Processes*, 2024, 111: 21-31.
- [11] SHUAI Y, LIN Z, CHEN W, *et al.* SF-YOLO: An evolutionary deep neural network for gear end surface defect detection[J]. *IEEE Sensors Journal*, 2024, 24(13): 21762-21775.
- [12] YANG S, ZHANG Z, WANG B, *et al.* DCS-YOLOv8: An improved steel surface defect detection algorithm based on YOLOv8[C]//Proceedings of the 2024 7th International Conference on Image and Graphics Processing. Beijing, China: ACM, 2024: 39-46.
- [13] 陶友凤,姜麟,大妹.基于RCD-YOLO的钢材表面缺陷检测算法[J/OL].控制工程,(2024-09-10)[2024-12-10]. 2024, <https://doi.org/10.14107/j.cnki.kzgc.20240426>.
- [14] 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.
- [15] 吴亚尉,明帮铭,何剑锋,等.基于YOLO-GR算法的轻量化钢材表面缺陷检测[J].组合机床与自动化加工技术,2023(11):107-111,115.
- [16] HAN K, WANG Y, TIAN Q, *et al.* GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1577-1586.
- [17] HAN K, WANG Y, GUO J, *et al.* ParameterNet: Parameters are all you need[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 15751-15761.
- [18] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [19] CHEN Z, HE Z, LU Z M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[EB/OL]. (2023-01-12)[2024-12-10]. <https://doi.org/10.48550/arXiv.2301.04805>.
- [20] TIAN Z, SHEN C, CHEN H, *et al.* FCOS: Fully convolutional one-stage object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, ROK: IEEE, 2019: 9626-9635.
- [21] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 13708-13717.
- [22] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN[J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [23] HE K, ZHANG X, REN S, *et al.* Deep residual learning for

- image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-9-4) [2015-5-7]. <https://doi.org/10.48550/arXiv.1911.11929>.
- [25] WANG C Y, LIAO H Y M, YEH I H, *et al.* CSPNet: A new backbone that can enhance learning capability of CNN [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: IEEE, 2020: 390-391.
- [26] 朱晓慧, 钱丽萍, 傅伟. 图像数据增强技术研究综述 [J]. 软件导刊, 2021, 20(5): 230-236.

Steel Surface Defect Detection Algorithm Based on GSL-YOLO

WANG Zekai, JIANG Qingchao, YAN Xuefeng

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Steel is prone to surface defects due to processing and environmental factors in modern industrial production, which adversely affect its overall quality and service life. To overcome the shortcomings of existing detection methods, such as, inadequate accuracy and high computational complexity, this paper proposes an improved YOLOv8-based algorithm, GSL-YOLO. To address the issues of feature redundancy and insufficient feature expression ability, GhostBottleneck is introduced in the C2f module to replace the standard DarknetBottleneck, thereby reducing redundant computation and enhancing feature extraction efficiency. DynamicConv is used to replace the standard CBS module to enhance adaptability to complex scenarios. To further improve the detection accuracy for minute defects, the SENetV1 attention mechanism is integrated into the neck network, which enhances the focus on key features by adaptively adjusting channel weights. In addition, a lightweight shared convolution detection head named LSDECD is designed to address the issues of large number of detection head parameters and high computational complexity, effectively reducing the number of model parameters and the cost of computation. The experimental results demonstrate that, on the NEU-DET steel surface defect dataset, GSL-YOLO achieves a 3% increase in mAP50 compared with YOLOv8, while the parameter quantity is reduced by 33% and the computational cost is reduced by 37% simultaneously. The proposed method achieves good lightweight design while improving the accuracy of model detection, meeting the application requirements of industrial real-time detection.

Key words: defect detection; YOLOv8; Ghost Bottleneck; DynamicConv; attention mechanism; lightweight detection head

(责任编辑: 李娟)