

DOI: 10.13382/j.jemi.B2508657

基于 UAL-YOLO 的钢材表面缺陷检测算法*

张新君^{1,2} 赵志闯¹

(1. 辽宁工程技术大学电子与信息工程学院 葫芦岛 125105;

2. 辽宁省无线射频大数据智能应用重点实验室 葫芦岛 125105)

摘要:针对现有钢材表面缺陷检测模型参数量大,推理速度慢,且模型轻量化与检测性能难以兼顾的问题,提出一种基于 YOLOv11n 的 UAL-YOLO 模型。首先,在主干网络末端引入非参数化的统一三维注意力模块(unified 3D attention module, UAM),该模块通过解析特征分布的二阶统计特性动态推导出三维权重,利用神经元间的线性可分性精准捕捉特征空间中的缺陷信息,显著提升了关键语义特征的信噪比。其次,设计了全新的注意力引导融合(attention-guided fusion, AGF)颈部网络,利用深层特征生成引导权重,通过智能和选择性地聚合多尺度特征,增强了模型的特征融合效率与精度。最后,设计轻量级高效偏移上采样器(lightweight efficient offset-based upsampler, LEO)用来替代传统上采样算子,通过 LEO 模块来学习内容感知的采样点偏移,有效捕捉如细长裂纹边缘、微小麻点坑洼等微弱的纹理特征,提升了模型对低对比度细碎边缘的特征捕捉精度,同时降低了微小缺陷在复杂背景下的漏检率。实验表明,与基础的 YOLOv11 模型相比较,在 NEU-DET 数据集上 mAP@0.5 提升了 2.1%,召回率提升了 3.7%,同时模型的参数量、计算量和占用储存容量分别下降了 26.36%、15.87%和 23.64%。UAL-YOLO 模型在提升精度的同时实现了显著的轻量化,为高精度、低开销的钢材缺陷检测技术在工业场景的部署提供了新的解决方案。

关键词: YOLOv11n; 缺陷检测; 特征融合; 轻量化

中图分类号: TP183; TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Research on surface defect detection of steel strip based on UAL-YOLO

Zhang Xinjun^{1,2} Zhao Zhichuang¹

(1. School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China;

2. Liaoning Key Laboratory of Radio Frequency and Big Data for Intelligent Applications, Huludao 125105, China)

Abstract: Addressing the challenges of large parameter sizes, slow inference speeds, and the difficulty in balancing model lightweighting with detection performance in existing steel surface defect detection models, this paper proposes a UAL-YOLO model based on YOLOv11n. Firstly, a parameter-free unified 3D attention module (UAM) is integrated into the backbone. By deriving 3D weights from second-order statistics, it identifies outlier defect features via linear separability, significantly enhancing the signal-to-noise ratio of semantic information. Secondly, a novel attention-guided fusion (AGF) neck network is designed, which utilizes deep features to generate guiding weights, intelligently and selectively aggregating multi-scale features to improve the efficiency and accuracy of feature fusion. Finally, a lightweight efficient offset-based up sampler (LEO) is developed using content-aware offsets to recover fine textures of slender cracks and minute pits. This approach improves the extraction of low-contrast edges and effectively minimizes the missed detection rate of tiny defects in industrial scenarios. Experimental results demonstrate that compared to the baseline YOLOv11 model, the UAL-YOLO model achieves a 2.1% improvement in mAP@0.5 and a 3.7% increase in recall rate on the NEU-DET dataset, while reducing the number of parameters, computational load, and storage capacity by 26.36%, 15.87%, and 23.64%, respectively. The UAL-YOLO model significantly enhances precision while achieving notable lightweighting, providing a new solution for the deployment of

收稿日期: 2025-08-10 Received Date: 2025-08-10

* 基金项目: 辽宁省教育厅基本科研项目面上项目(LJKMZ20220678)、辽宁省教育厅重点项目(LJ212410147003)资助

high-precision, low-cost steel defect detection technologies in industrial scenarios.

Keywords: YOLOv11n; defect detection; feature fusion; lightweighting

0 引言

钢材是国家基础设施建设的重要的原材料,广泛应用于各个领域,发挥着重要的作用。目前的钢材产品大部分都是通过压力加工生产而成,在加工生产时,由于钢材本身制造成本高、设备工艺条件及现有科学技术水平的限制等,会造成钢材表面出现各类缺陷,这些缺陷的存在会影响钢材的抗压、抗拉、耐腐蚀性能,这就要求在交付使用前对钢材表面的缺陷进行检测,以保证产品质量。目前的钢材缺陷检测主要依靠深度学习来完成。深度学习模型会从数据中自动学习和提取特征,自动完成端到端的检测流程,既避免了手动设计特征,也具有不依赖环境的优势,简化了工业生产流程。目前主流的目标检测算法主要有单阶段方法与两阶段方法两大类。

两阶段目标检测算法以 R-CNN 系列为代表,如 R-CNN^[1]、Fast R-CNN^[2]、Faster R-CNN^[3] 及 Mask R-CNN^[4] 等。这类算法首先生成大量的候选区域 (region proposals),再对这些区域进行分类和精确定位,检测精度较高。然而候选区域的生成环节,检测速度慢,计算流程复杂,难以满足工业生产对于检测速度的要求。与之相对,单阶段目标检测算法有 SSD^[5] 和 YOLO^[6] 系列,将检测任务直接转化为端到端的回归问题,没有候选框生成环节,检测速度快,模型简单,更容易部署到工业流水线中。YOLO 算法自出现以后,不断有改进版本问世,兼顾了检测速度和精度,在工业流水线缺陷检测中被广泛使用,是目前工业流水线中主流的框架结构。朱成杰等^[7]引入聚焦调制模块替代 SPPF 模块,利用新的 C2f-MB 结构替换了原有的模块,在主干网络中加入了注意力模块,增强了模型的多尺度特征表达与抗背景干扰能力。崔晶楠等^[8]在 YOLOv5s 的主干网络中融合动态卷积 ODConv 以增强对重要区域的特征提取,并使用 WIoU 损失函数提高目标框的定位精度,以及通过添加更多的检测头解决对多尺度目标的漏检问题。曾勇杰等^[9]为提升风机叶片缺陷检测效果,采用高效多尺度卷积 (EMSCov) 降低特征图中尺度一致的冗余特征干扰,在检测头引入 Dynamic Head 注意力机制以增强目标感知,并设计新的 Inner-Wise-MPDIoU 损失函数以加速模型收敛和提升定位精度。高春艳等^[10]针对钢材表面缺陷背景复杂、目标尺度不一的问题,通过在 YOLOv7 主干网络中引入轻量化注意力机制以增强显著性特征提取,并利用改进的特征金字塔结构优化多尺度信息融合,提升了对复杂工业环境下微小缺陷的检测能力。李思思

等^[11]为提升 YOLOv7 的钢材缺陷检测效率,设计非对称解耦头以解决不同检测任务间的冲突,采用轻量化 FCSP 模块增强主干与颈部的特征提取及融合,并引入可学习参数促进动态特征融合以强化对小目标的检测能力。徐洪俊等^[12]为提升 YOLOv5s 对钢材表面异态缺陷的识别精度,引入协同注意力机制聚焦有效特征区域,并通过重构检测头的回归损失函数优化边界框定位,在保持模型轻量化的同时有效降低了缺陷的漏检率。

尽管现有研究在提升检测精度与模型轻量化方面取得了显著进展,但在实际工业场景中,如何平衡特征提取的深度与计算资源的有限性、多尺度特征间的语义对齐,以及上采样过程中的细节保持,仍是制约检测性能的核心瓶颈。上述缺陷检测方法难以在保持高精度的同时轻量化,导致计算能力有限的设备部署困难,为解决上述的问题,本文从特征精炼、智能融合与内容感知重构 3 个维度出发,提出了 UAL-YOLO 检测方法,主要工作包括:

1) 针对主干网络深层特征存在信息冗余与噪声干扰的问题,提出了一种统一三维注意力模块 (unified 3D attention module, UAM)。在主干网络的深层特征输出端嵌入该模块。该模块利用特征分布的二阶统计量方差动态推导三维权重的解析解,旨在不增加额外参数开销下实现对特征空间离群点的精准捕捉,从而显著提升模型对关键缺陷特征的辨识度与定位精度。

2) 针对传统颈部网络融合多尺度特征时效率低下的问题,设计了一种注意力引导融合颈部网络 (attention-guided fusion, AGF)。它通过深层特征的语义信息生成引导权重,通过智能地和有选择性地聚合来自不同层级的特征,解决了信息融合过程中的冲突与冗余,提升了模型对多尺度缺陷的综合表达能力。

3) 针对传统上采样算子易造成微小缺陷的关键细节丢失的问题,设计了轻量级高效偏移上采样器 (lightweight efficient offset-based upsampler, LEO)。它能根据特征内容动态调整上采样方式,最大程度地保留和恢复缺陷的边缘和纹理细节,提升了模型对微小、模糊缺陷的检测精度。

1 YOLOv11n 模型

本文选取先进的目标检测模型 YOLOv11n^[13] 作为基准模型进行对比实验。YOLOv11n 作为 YOLO 系列中的轻量化版本,在保持较高检测精度的同时,具有更快的推理速度,是工业界常用的高效检测器之一。该模型遵循了经典的 3 段式设计,主要由主干网络 (backbone)、颈部

网络 (neck) 和检测头 (head) 3 部分构成。其中主干网络用于提取输入图的多尺度特征信息。采用 C3K2 模块堆叠而成,并在网络的深层集成了 SPPF (spatial pyramid pooling fast) 模块与 C2PSA 模块捕获更丰富的上下文信息。颈部网络核心作用是融合主干网络在不同阶段输出的特征图。检测头负责对颈部网络输出的融合特征进行解析,并生成最终的预测边界框和类别。YOLOv11 的网络结构如图 1 所示。

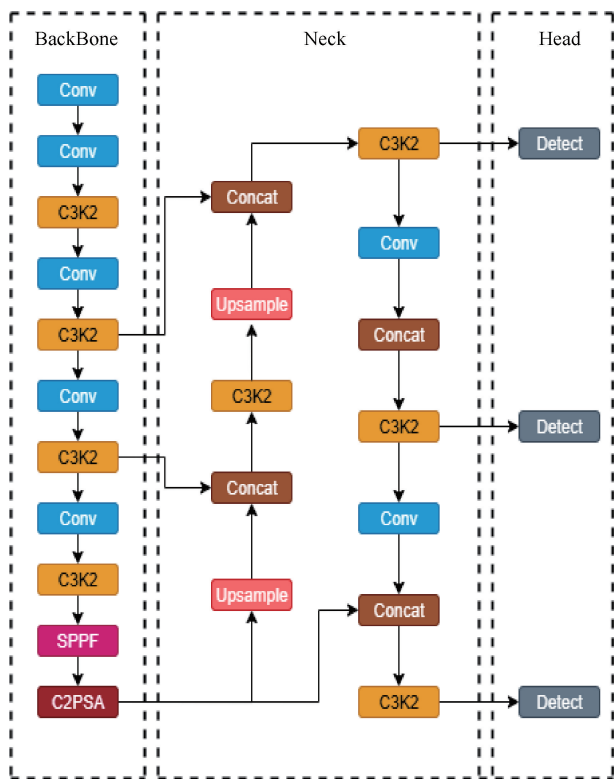


图 1 YOLOv11 网络结构

Fig. 1 Network Structure of YOLOv11

2 UAL-YOLO 钢表面缺陷模型

为了实现钢材表面缺陷检测模型在追求高精度的同时实现轻量化。基于 YOLOv11n 模型,提出了 UAL-YOLO 模型,其结构如图 2 所示。在主干网络末端 C2PSA 模块之前新增了 UAM 模块,来提升了对关键信息的捕获能力;对颈部网络的特征融合机制进行重构,利用设计的 AGF 模块替换了 YOLOv11n 原有的基于通道拼接的传统融合结构。实现了更高效、更精准的多尺度特征聚合,大幅削减了颈部网络的冗余参数与计算负载;替换了 YOLOv11n 颈部网络中传统的最近邻上采样算子,采用了 LEO 模块,加入到以动态点采样的全新范式替代传统上采样算子,在保证高效性的同时,最大程度地恢复了微小缺陷的细节纹理。

2.1 统一三维注意力模块

在高性能的目标检测器中,注意力机制已成为提升模型特征表达能力、聚焦关键信息的重要组件。但是现有的主流注意力机制在权重分配的维度上存在固有的局限性。为了更直观地理解本文所引入模块的创新性,本文对比几种注意力机制的权重维度,如图 3 所示。

如图 3(a) 所示,以 SE-Net 为代表的通道注意力 (channel-wise attention),主要作用是为输入的特征图 X (尺寸为 $H \times W \times C$) 生成一维的权重向量 (尺寸为 $1 \times 1 \times C$)。这说明同一通道内的所有像素共享完全相同的注意力权重。该机制虽然高效,但它只解决了看重哪些通道的问题,却忽略了不同空间位置的重要性差异。

如图 3(b) 所示,典型的空注意力 (spatial-wise attention) 其核心是为特征图 X 生成一个二维的权重图 (尺寸为 $H \times W \times 1$)。这说明了同一空间位置 (x, y) 的像素在所有通道上被赋予相同的权重。该机制解决了看重哪些位置的问题,但又无法区分不同通道在同一位置的贡献度。

为了克服上述机制将通道与空间维度割裂处理的局限性,本文提出了一种新颖的 UAM 模块。受生物视觉系统中的侧抑制 (lateral inhibition) 机制启发,视觉神经元在受到刺激时会对周围神经元产生抑制作用,从而突出对比度最高的区域。UAM 模块通过构造三维能量函数来模拟这种神经元竞争机制,旨在在不依赖额外卷积层的前提下,利用特征分布的二阶统计特性方差实现对缺陷显著性区域的自适应捕获与判别性增强。

如图 3(c) 所示,UAM 模块将输入的特征图 X 推导出一个尺寸完全相同的三维权重张量 (尺寸为 $H \times W \times C$)。特征图中的每一个独立的神经元在特定通道和特定位置的像素点都拥有一个专属的注意力权重。该模块的 3D 权重机制使得模型能够实现最精细化的特征调制,UAM 本质一种基于解析分布推导的非参数化注意力机制。通过数学解析解替代了传统注意力中冗余的参数学习过程。通过解析输入特征的二阶统计特性直接推导出权重张量,其生成过程不依赖任何卷积或全连接层。

尽管现有 3D 注意力机制也尝试统一空间与通道权重,但该模块数学机理上引入侧抑制理论。通过量化神经元与局部背景的线性可分性来实现特征提纯,而非简单的数值加权。使其在面对钢材生产中未见过的异态缺陷时,提高模型对缺陷的捕捉能力,有效避免了传统参数化注意力在小样本工业场景下易产生的过拟合瓶颈。

该模块的核心思想是通过定义一个能量函数,来量化每个神经元的独特性。该能量函数本质上是评估目标神经元与其所在通道内局部平均特征的线性可分性。特征图中的任意一个目标神经元 t ,通过以下能量函数来评估其重要性 e_t :

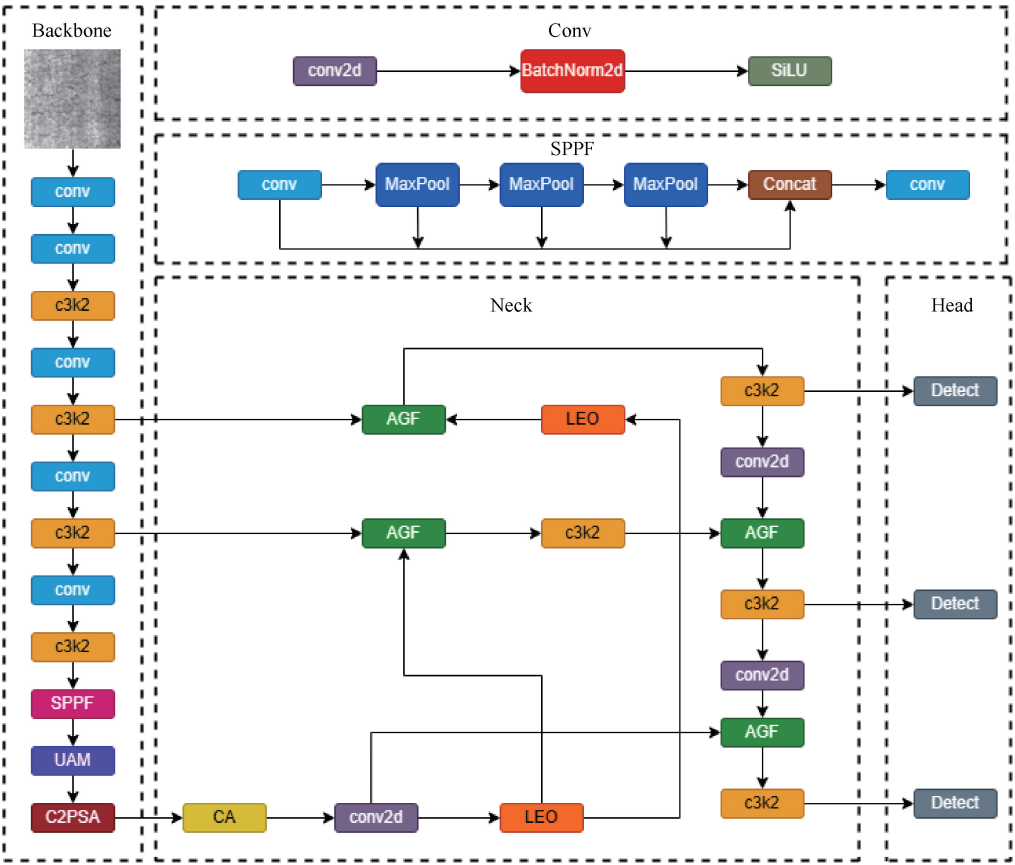


图 2 UAL-YOLO 模型结构
Fig. 2 Structure of the UAL-YOLO Model

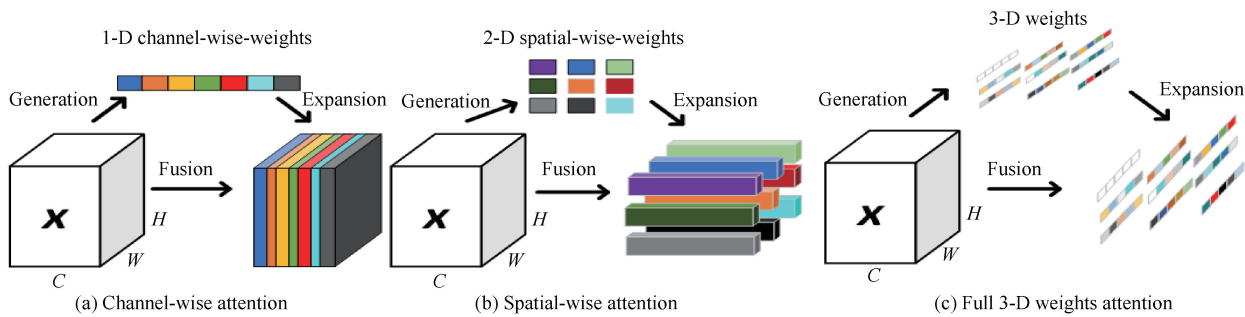


图 3 不同注意力机制权重维度
Fig. 3 Different dimensions of attention mechanism weights

$$e_t = \frac{4(\sigma^2 + \lambda)}{(\hat{i} - \hat{\mu})^2 + 2\sigma^2 + 2\lambda} \quad (1)$$

式中： $\hat{i}$ 为目标神经元的值； $\hat{\mu}$ 和 σ^2 分别是该通道内除 t 以外所有其他神经元的均值和方差，其中二阶统计量 σ^2 能够有效捕捉背景纹理中的波动信息，增强对缺陷边缘的敏感度。在钢材表面检测中，背景纹理通常表现为平稳分布，而缺陷边缘则引发局部方差的剧烈跳变。利用二阶统计量捕捉这种分布特征，使得 UAM 能够以极低的

计算开销实现对复杂背景下微小边缘的敏锐感知。 λ 为正则化项，防止分母为 0。从式(1)可以看出，当神经元 $\hat{i}$ 与均值 $\hat{\mu}$ 差异越大时，分母增大，从而导致能量值 e_t 降低。这表明该神经元偏离了背景分布，承载了更高的显著性信息，因此越重要。这一过程将缺陷识别建模为特征空间内的离群点(outliers)探测，通过最小化能量函数 e_t ，模块能够有效衡量目标神经元与局部背景的线性可分性(linear separability)。此外，式(1)中的二阶统计量

σ^2 承载了背景的波动信息,使得 UAM 能够通过捕捉方差的跳变,更灵敏地识别不规则缺陷的几何边缘。特征图中每一个神经元都计算出对应的能量值 e_i 后,将这些标量值在各自的空间和通道上进行组合,从而构建出与原始特征图 X 尺寸完全相同的张量 e 。最终的注意力权重被定义为能量的倒数。通过对原始特征图进行逐元素的乘法,UAM 便完成了对特征的精细化调制,其过程可总结为:

$$X_{refined} = X \otimes \text{Sigmoid}\left(\frac{1}{e}\right) \quad (2)$$

相比于基于卷积层学习的固定权重,UAM 能够根据输入特征的实时方差分布动态计算解析解,从而更敏锐地捕获钢材表面缺陷中不规则的几何边缘。本文将 UAM 创新性地应用于主干网络的末端,在不增加模型参数量和计算复杂度的情况下,对即将送入特征融合颈部的最高层语义特征进行了有效的 3D 注意力优化,提升了模型对关键缺陷特征的辨识度和定位能力,为后续注意力引导融合模块精准的缺陷检测奠定了坚实基础。

2.2 注意力引导融合模块

特征金字塔网络 (FPN) 与路径聚合网络 (PANet) 通过构建自顶向下与自底向上的双向路径,有效地融合了多尺度特征,提升了对不同尺寸目标的检测能力。然而,这些结构中普遍采用的特征融合方式本质上是一种无差别的信息聚合,如简单的逐元素相加 (Add) 或通道拼接 (Concat)。这种策略假设来自不同层级的特征具有同等的重要性,这在实际场景中往往并非最优解。不仅可能导致携带着丰富位置细节的浅层特征被深层特征图中复杂的背景纹理和噪声所污染,还可能在融合过程中漏掉对小目标检测至关重要的关键信息,从而限制了模型检测性能的进一步提升。

针对上述描述的传统融合方式的局限性,本文设计了一种注意力引导融合模块 (AGF) 并以此为核心组件构建了整个特征融合颈部。AGF 模块的结构如图 4 所示。图 4 中 S1~S5 表示主干网络特征图;N1~N5 表示自顶向下融合特征图;M1~M5 表示自底向上融合特征图;AGF 表示核心创新模块。AGF 模块的核心设计思想在于将特征融合过程从被动的数据混合转变为主动的信息引导。该模块利用深层特征图所具备的全局视野和高级语义理解能力,生成一个动态的引导信号,以此来智能地调控和筛选出来自浅层特征图的丰富细节信息,从而实现更高效、更有目的性的跨层级特征聚合。

AGF 的融合策略如图 5 所示 (AGF 内部图所示),给定一个作为引导流的输入高层特征 $f_{high} \in R^{C \times H \times W}$ 和一个作为内容流的输入低层特征 $f_{low} \in R^{C' \times H' \times W'}$,其详细融合流程如下。

1) 特征预处理与维度对齐。先将高层特征 f_{high} 的

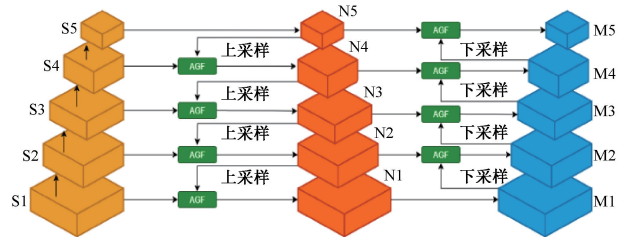


图 4 AGF 网络结构

Fig. 4 AGF network structure

空间分辨率与低层特征 f_{low} 对齐,采用可学习的转置卷积 (transposed Convolution) 对其进行上采样,得到与 f_{low} 相同尺寸的特征图 $f_{high_up} \in R^{C \times H' \times R'}$ 。同时对 f_{low} 和 f_{high_up} 应用 1×1 卷积以统一通道数,得到 f'_{low} 和 f'_{high_up} 。

2) 并行注意力处理:AGF 模块通过两条并行的注意力支路来分别处理内容与引导信息。首先引导权重生成:将对齐后的高层特征 f'_{high_up} 送入到通道注意力的模块 (CA) 中去。CA 模块通过分析全局的语义信息,经过 Sigmoid 激活函数,输出一个数值范围在 $[0, 1]$ 之间的注意力权重图 $W_{att} \in R^{C' \times H' \times W'}$ 中的每个值都代表着重要的分数,用于后续的融合。其次进行内容特征精炼:将对齐后的低层特征 f'_{low} 送入另一个结构相同的 CA 模块,增强其包含的有效缺陷细节特征,抑制其自身的背景噪声,得到精炼后的内容特征 $f_{content} \in R^{C' \times H' \times W'}$ 。

3) 引导融合实现:最终融合是通过逐元素相乘,将引导权重 W_{att} 应用于内容特征 $f_{content}$ 实现对内容的动态加权。之后将加权后的结果与高层特征 f_{high_up} 进行逐元素相加,加法操作不仅完成了信息的最终聚合,也构成了一个残差连接,有助于梯度的稳定传播。整个引导融合过程可用以下公式精炼地表示:

$$f'_{high_up} = \text{Upsample}(\text{CA}(\text{Conv}_{1 \times 1}(f_{high}))) \quad (3)$$

$$W_{att} = \sigma(\text{CA}(f'_{high_up})) \quad (4)$$

$$f_{out} = (\text{CA}(\text{Conv}_{1 \times 1}(f_{low}))) \otimes W_{att} + f'_{high_up} \quad (5)$$

式中:Upsample($\cdot$)代表上采样;CA($\cdot$)代表通道注意力模块;Conv $_{1 \times 1}$ ($\cdot$)代表 1×1 的卷积; σ 代表 Sigmoid 激活函数; $\otimes$ 代表逐元素相乘。

AGF 结构效率与轻量化分析。在后续的实验数据显示,引入 AGF 模块后,模型参数量由 2.58×10^6 降至 1.97×10^6 ,降幅达 23.6%。主要取决于以下 3 个方面:

首先是通道维度的约束,在传统的特征融合 (如 Concat) 往往导致通道数在融合后倍增,迫使后续卷积层 (C3k2) 必须在极高维度的空间中进行运算。然而 AGF 模块是在特征预处理阶段 (步骤 1)) 使用了 1×1 卷积进行维度对齐。并将其统一固定在较小的通道数 (256)。有效地控制了颈部整体的通道宽度,使得后验权重的参数规模随通道数的压减而大幅缩减。

其次是 AGF 模块核心采用基于注意力的逐元素相乘与残差加法替代了传统的深度卷积融合方案。仅通过动态生成的权重 W_{att} 信息进行加权筛选。这种机制在增强有效特征的同时抑制了冗余背景噪声,从而在保证检测精度的前提下,允许模型移除原头部结构中为了处理冗余信息而设置的深层卷积层。

最后是 CA 组件虽然引入极少量的参数增长,但其在结构上起到了杠杆作用。通过牺牲极小规模注意力权重参数,换取了后续大尺寸特征图卷积层通道数的显著下降。这种局部增参、整体降维的设计逻辑,最终使模型在复杂度和检测性能之间取得了更优的平衡。

为了进一步说明 AGF 模块对模型轻量化的贡献,从参数量进一步对比 AGF 模块与传统 Concat 融合结构。设 C_{in1} 与 C_{in2} 为输入特征通道数, C'_{out} 为 AGF 内部统一

的映射维度, K 为卷积核尺寸。

原生结构采用 Concat 导致通道数物理堆叠,其后卷积层的参数量 P_{base} 为:

$$P_{base} = K^2 \cdot (C_{in1} + C_{in2}) \cdot C_{out} \quad (6)$$

由于 K 通常为 3,参数量随输入通道的增加呈线性扩张。

AGF 模块将融合过程重构为特征对齐与动态加权,其总参数量 P_{AGF} 由维度映射层与注意力组件构成:

$$P_{AGF} = 1^2 \cdot (C_{in1} + C_{in2}) \cdot C'_{out} + P_{CA} \quad (7)$$

式中: P_{CA} 代表通道注意力模块的参数量。由 AGF 核心融合算子 ($\otimes, +$) 属于非参数化操作,且 1×1 卷积的理论参数开销仅为 3×3 卷积的 $1/9$,所以使得 AGF 在显著增强跨层级特征融合能力的同时,实现了模型整体参数基数的大幅下降。

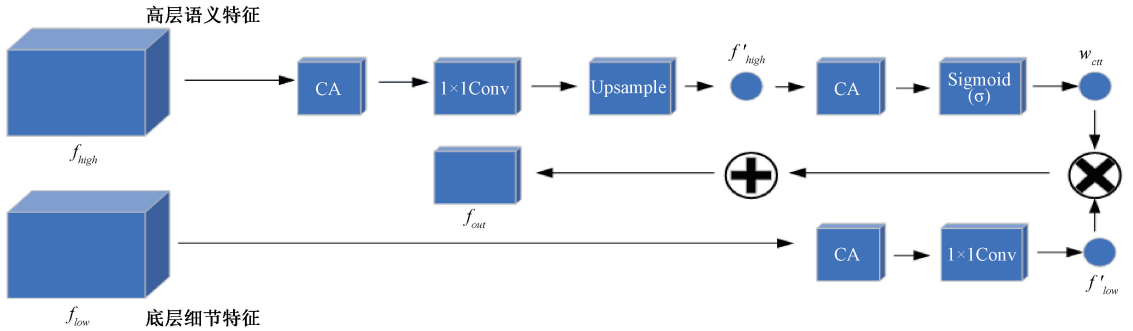


图5 AGF 内部图所示

Fig. 5 shown in the internal diagram of AGF

2.3 轻量级高效偏移上采样模块

近年来,为追求极致的性能,涌现了如 CARAFE、FADE 和 SAPA 等先进的动态上采样器。这些方法虽然有效提升了检测性能,但其实现路径却代价高,FADE 和 SAPA 依赖额外的高分辨率特征图进行指导,而 CARAFE 则需通过一个复杂的内部子网络来生成动态卷积核。这些设计不仅显著增加了模型的计算负担和内存开销,也限制了它们作为通用轻量化组件的适用场景。为解决这一问题,本文创新性地提出了一种轻量级高效偏移上采样器(LEO)。该模块根据输入的低分辨率特征自身来推导上采样方式。

LEO 模块的原理框架如图 6 所示。其中 $X \in R^{C \times H \times W}$ 代表输入的低分辨率特征图, s 为上采样尺度因子, g 为分组数。LEO 从动态点采样的角度来重新构建上采样过程,该过程解耦为两步:首先一个核心的采样点生成器(sampling point generator)负责根据输入特征 X 的局部内容,智能地生成一个采样点集合 $S \in R^{2g \times H \times s \times W}$;其次 grid sample 函数利用这个采样点集合 S 从原始特征 X 中进行重采样,从而得到高分辨率的输出特征 $X' \in R^{C \times sH \times sW}$ 。LEO 通过巧妙地规避耗时的动态卷积操

作,进一步实现了轻量化。

该模块的核心机制是采样点生成器,该机制详细结构如图 7 所示。目的是为规则的初始采样网格 G 学习一个与内容相关的、智能的偏移量图 O ,两者相加而成($S = G + O$)生成采样点 S 。输入的低分辨率特征 X 并行通过两个线性映射层(linear),将特征通道统一变换为 $2gs2$ 。这一特定的通道数是为了配合后续像素重组(pixel shuffle)操作而设计的,该操作在将特征图分辨率提升 s 倍的同时,会将 $2gs2$ 个输入通道高效地重塑为我们最终需要的 $2g$ 个输出通道。在通道变换后,其中一条支路用于生成动态范围因子 σ ,它与另一条主路提炼出的核心特征进行逐元素相乘,以实现偏移量的自适应范围约束。最后,将此约束后的特征送入像素重组操作,实现了从通道维度到空间维度的重塑,得到内容感知的偏移量图 O 。

整个过程如下:

$$O =$$

$$\text{PixelShuffle}(0.5 \times \text{Sigmoid}(\text{linear}_1(X)) \cdot \text{linear}_2(X)) \quad (8)$$

利用学习到的内容感知偏移量图 O ,使采样网格

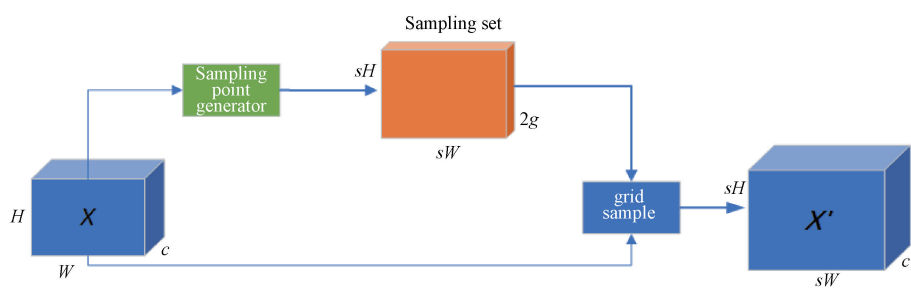


图 6 LEO 整体架构

Fig. 6 Overall architecture diagram of LEO

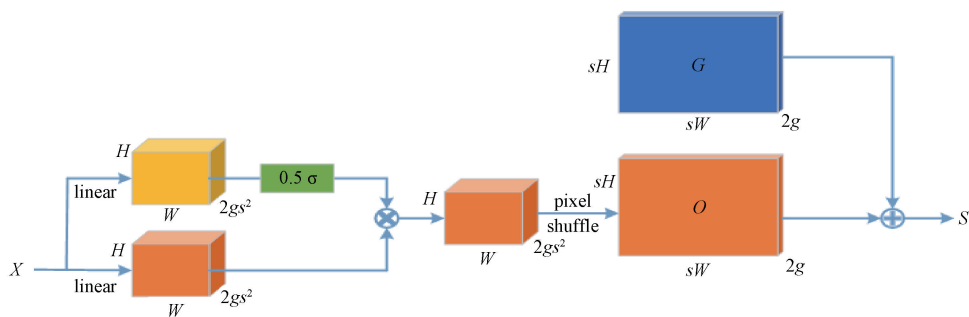


图 7 采样点生成器详细结构

Fig. 7 Architecture of the sampling point generator

G 突破几何位置限制,主动向缺陷边缘的高频纹理区域靠拢。保留了微小目标的边缘陡峭度,从而在下采样信息流失的情况下,依然能够维持缺陷特征在空间维度上的连通性,防止了因纹理模糊导致的检测框断裂或漏检。

在动态上采样领域,CARAFE 需构建复杂的编码网络来预测大尺寸重构核,参数量随核尺寸呈现平方级增长,FADE 则高度依赖高分辨率特征图的外部引导,显著增加了显存对齐与计算负荷。相比之下,LEO 仅通过两层轻量化线性映射推导 $2gs^2$ 并利用像素重组实现分辨率提升,使得 LEO 在参数量、计算量及推理延迟上均远低于上述算子。当 LEO 与 AGF 结合时,引导融合机制产生了协同效应。LEO 被用作 AGF 模块内部的上采样算子。该机制为 AGF 提供了细节更丰富的高质量上采样特征,使得 AGF 的注意力引导过程更为精准,检测能力得到了显著的提升。

3 实验结果及分析

3.1 实验环境

实验平台使用 NVIDIA A800 GPU, Intel (R) Xeon(R) Gold 6330N CPU @ 2.20GHz,操作系统为 Ubuntu。深度学习的框架为 PyTorch 2.2.2, Python

3.10.16,CUDA 12.1。实验环境参数如表 1 所示。

表 1 实验环境参数

Table 1 Experimental setup

实验参数	参数值
Epoch	300
Batch size	16
Optimize	SGD
Image size	640
Learn Rate	0.01
Momentum	0.937

3.2 实验数据集

本实验采用由东北大学发布的表面缺陷数据集 NEU-DET 和真实工业中收集的表面缺陷数据集 GC10-DET。GC10-DET 收集了 10 种不同类型的表面缺陷,如冲孔、焊缝、月牙状间隙、水斑、油斑等。NEU-DET 收集了热轧钢带的 6 种典型表面缺陷,即裂纹、斑块、夹杂、麻点、压入氧化皮和划痕,数据集一共 1 800 张数据集,按照 8 : 1 : 1 划分。包含 1 440 张训练集,180 张测试集和 180 张验证集。部分缺陷图像示例如图 8 所示。

3.3 评价指标

本实验采用精确率(P)、召回率(R)、平均精度均值(mAP)、参数量(parameters)、浮点计算量(floating

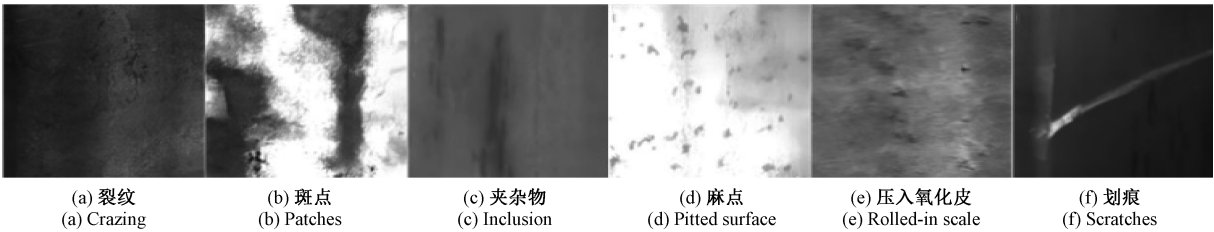


图 8 缺陷图像示例

Fig. 8 Examples of defect images

points of operations, FLOPS)、每秒帧数(FPS)。参数量: 评估模型的大小与复杂度;计算量:指模型在一次前向推理过程中执行的浮点运算次数。其评价指标公式如下:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P_{(R)} dR \tag{11}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{12}$$

式中: TP 为模型正确检测出的缺陷样本数量; FP 为模型将背景或无缺陷区域错误地识别为缺陷的样本数量; AP 为模型平均准确度; FN 为模型漏检的缺陷数量; N 为类别数; AP_i 为第 i 种缺陷类别的 AP 值, 其中 $mAP@0.5$ 为在 IoU 阈值为 0.5 时所有类别的平均检测精度。

3.4 消融实验及分析

表 2 中 YOLOv11n 是基础模型;YOLOv11n+UAM 为在主干网络的 SPPF 层之后引入了 UAM 模块,实验表明在引入 UAM 模块,能够在不增加任何模型成本的前提下,使模型的整体检测精度提升了 1.1%。YOLOv11n+AGF 为单独在基础模型上引入的 AGF 模块,这个改进带来了显著的轻量化效果,与基础模型 YOLOv11n 相比, YOLOv11n+AGF 的参数量从 2.58×10^6 下降到了 1.97×10^6 ,降幅达 23.6%,计算量从 6.3 GFLOPs 降至 5.8 GFLOPs;占用的储存容量也从 5.5 MB 减少至 4.3 MB。充分证明了 AGF 模块本身是一种计算开销和存储成本极低的、极为高效的特征融合模型。YOLOv11n+LEO 仅在基础模型的颈部用 LEO 替换标准上采样,召回率达到了 74.3%,提升了 1.4%,这是因为 LEO 能够更好地保留微小、模糊缺陷的边缘和纹理细节,从而减少了因信息丢失而导致的漏检的情况,使得模型发现潜在目标的能力得到了有效增强。

表 2 消融实验结果

Table 2 Ablation study results

模型	召回率/%	平均精度值/%	参数量/($\times 10^6$)	计算量/GFLOPs	推理速度/fps	占用储存容量/MB
YOLOv11n	72.9	77.4	2.58	6.3	90.9	5.5
YOLOv11n+UAM	68.0	78.5	2.58	6.3	90.1	5.5
YOLOv11n+AGF	68.3	75.3	1.97	5.8	86.9	4.3
YOLOv11n+LEO	74.3	76.0	2.59	6.3	71	5.5
YOLOv11n+UAM+AGF	72.2	75.9	1.97	5.8	84.7	4.3
YOLOv11n+UAM+LEO	73.8	77.3	2.59	6.3	73.4	5.5
YOLOv11n+UAM+AGF+LEO	76.6	79.5	1.90	5.3	84	4.2

在两两组合的实验中, YOLOv11n+UAM+AGF 在 AGF 模型的基础上加入了 UAM 模型,弥补了 AGF 轻量化所带来的性能损失, UAM 通过主干网络的提纯,为 AGF 提供了更高质量和噪声更少的输入,在不增加额外的成本的基础上,挽回了召回率和精度。YOLOv11n+UAM+LEO 表明 UAM 和 LEO 模型之前存在协同作用,单独使用 UAM 和 LEO 分别在精度和召回率上表现出偏科现象。当两者结合时,形成了优势互补;LEO 通过保留更多上采样细节提升了召回率,弥补了 UAM 因严格筛选

特征而导致的漏检问题;UAM 通过在主干末端提纯特征,为颈部网络提供了更高质量的输入,又反过来弥补了 LEO 可能引入的微小定位偏差,使得模型的整体精度得以恢复。最终 UAL-YOLO 模型平均精度值提升 2.1%;参数量和计算量分别减少了 26.36%和 15.87%,减少了设备性能占用;在推理速度上虽略低于基准模型,但仍高于工业实时检测常规标准的 2 倍多,说明了 UAL-YOLO 在牺牲极小的推理效率的情况下,换取了检测精度、模型体积与参数利用率深度优化,实现了性能与轻量化的

最优权衡,同时占用储存容量下降了 23.64%,极大地方便了模型在边缘计算平台或嵌入式设备上的移植与部署。此外,对 UAL-YOLO 训练的全过程进行了监测。在 300 个 Epoch 的训练周期内,模型的各项损失函数(Loss)收敛平稳,且未出现震荡或过拟合现象。mAP@0.5 曲线在训练中后期稳步攀升并最终达到平台期,最终精度优于基准模型。这也证实了引入的 UAM、AGF 与 LEO 模块在提升检测性能的同时,具备优异的数值收敛性与训练稳定性。综上,本文提出的 UAL-YOLO 模型相较于基础模型,3 个改进的策略在提升精度的同时实现高效地轻量化。

3.5 对比实验

为了验证该算法的优越性,在 NEU-DET 数据集以及相同的实验条件下,将 YOLOv5n、YOLOv8n、YOLOv11n、YOLOv12n、以及参考文献[14-17]算法进行对比,其中加粗表示最优值。

从表 3 可知,UAL-YOLO 算法在与 YOLOv5 和 YOLOv8 的比较中,参数量分别下降了 12.84% 和 29.10%,占用储存容量分别下降了 10.64% 和

26.32%,同时精度分别提升了 3.9% 和 1.3%,在工业检测关注的实时性方面,UAL-YOLO 的推理速度达到了 84 fps,远超工业流水线实时检测的阈值标准。虽然相较于基准模型 YOLOv11n 速度小幅下降,但是精度提升了 2.1%,参数量缩减了 26.18%。这说明模型在确保实时性的同时提供了更高的检测可靠性。相比与 YOLOv12n,UAL-YOLO 不仅推理速度更快,精度亦高出 3.6%。与 PIC2f-YOLO^[14] 相比,精度提升 1.5% 的同时,参数量和计算量分别下降了 50% 和 26.9%。在与 ESE-YOLO^[15] 相比,精度提升了 3.2%,参数量减少了 17.4%,虽然 ESE-YOLO^[15] 拥有低的计算量,但是 UAL-YOLO 以更少的模型参数实现了显著的精度优势。在与 YOLOv8n-CSG^[16] 的对比中,UAL-YOLO 的平均精度提升了 2.7%。与 YOLOX^[17] 相比,精度提升了 2.5%,而参数量仅为 26%,展现出更强的特征利用率。结果表明,本文提出的 UAL-YOLO 模型成功地将高精度检测能力与轻量化设计融为一体,其综合性能表现最佳,能够更出色地应对钢表面缺陷检测的实际需求。

表 3 不同检测网络对比实验
Table 3 Comparison with state-of-the-art detectors

模型	召回率/%	平均精度值/%	参数量/($\times 10^6$)	计算量/GFLOPs	推理速度/fps	占用储存容量/MB
YOLOv5n	73.0	75.6	2.18	5.8	104.1	4.7
YOLOv8n	72.4	78.2	2.68	6.8	111	5.7
YOLOv11n	72.9	77.4	2.58	6.3	91.7	5.5
YOLOv12n	70.2	75.9	2.55	6.3	70.9	5.6
PIC2f-YOLO ^[14]	—	78.0	2.6	10.6	82	—
ESE-YOLO ^[15]	—	76.3	2.3	3.2	—	—
YOLOv8n-CSG ^[16]	73.0	76.8	1.95	5.1	—	—
YOLOX ^[17]	—	77.0	7.23	20.7	100	—
UAL-YOLO	76.6	79.5	1.90	5.3	84	4.2

3.6 泛化实验

在 GC10-DET 钢材表面缺陷数据集上进行了进一步的实验,实验数据如表 4 所示,在 GC10-DET 的数据集检测结果中,本文的 UAL-YOLO 模型相较于基准 YOLOv11n 模型,展现出了全面的性能优势。其中在在针对水斑、油斑等边缘弥散的缺陷类别中,UAL-YOLO 对三维特征能量的动态聚焦,有效抑制了背景噪声的干扰,解决了原生模型在该类缺陷上频发的误检问题。通过深层语义引导,实现了对多尺度特征的选择性聚合,显著提升了检测的准确率,使得 UAL-YOLO 的平均精度值达到了 66.8%,相比基准模型的 64.9%,提升了 1.9%;UAL-YOLO 模型在处理 GC10-DET 数据集的月牙状间隙时,比原生最近邻上采样保留了更多的纹理细节使得召回率提升了 4.9%,在性能稳步提升的同时,参数量减少了

26.36%,计算量减少了 15.87%,模型占用储存容量减少了 25.45%。综上所述泛化实验的结果有力地证明了本文提出的 UAL-YOLO 模型并非只在特定数据集上有效,其通过改进所带来的高精度与轻量化的双重优势具备良好的泛化性,能够在不同的检测场景下稳定、高效地完成钢材表面缺陷检测任务。

表 4 泛化实验
Table 4 Generalization experiment

模型	召回率/%	平均精度值/%	参数量/($\times 10^6$)	计算量/GFLOPs	占用储存容量/MB
YOLOv11n	58.2	64.9	2.58	6.3	5.5
UAL-YOLO	63.1	66.8	1.90	5.3	4.1

3.7 检测结果可视化对比与改进机理分析

为了更准确的展示改进的 UAL-YOLO 模型对缺陷

检测的效果,从 NEU-DET 的数据集中随机抽取裂纹、夹杂、斑块、麻点、压入氧化皮和划痕,6 种不同缺陷的图片,分别使用 YOLOv11n 和改进的 UAL-YOLO 模型进行对比可视化^[18]。实验结果如图 9~11 所示。

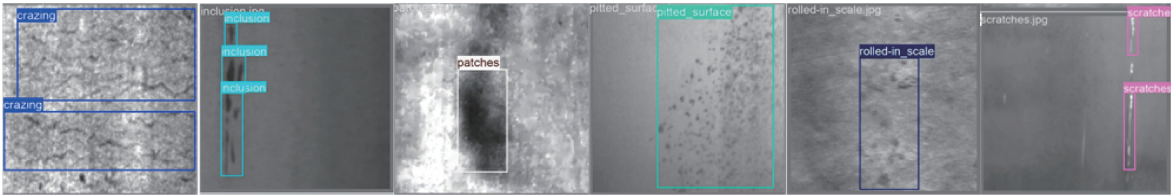


图9 缺陷图片
Fig. 9 Defect images

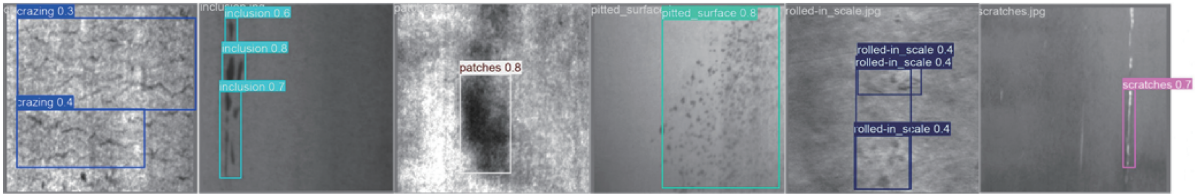


图10 YOLOv11n 的检测结果
Fig. 10 Detection results of YOLOv11n

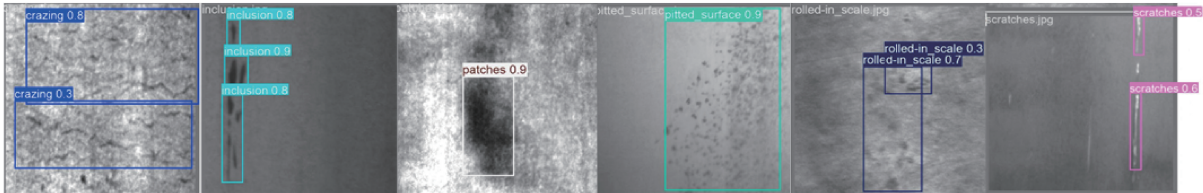


图11 UAL-YOLO 的检测结果
Fig. 11 Detection results of UAL-YOLO

通过对比实验结果图进行分析,UAL-YOLO 解决了微小缺陷的漏检问题,YOLOv11n 在检测划痕和裂纹等对比度极低、几何形态细长的缺陷时,容易产生漏检。而 UAL-YOLO 模型通过引入 LEO 模块,以内容感知的动态偏移采样替代了传统的最近邻插值,成功找回了这些原本无法检出的微小缺陷边缘。还解决了抑制背景噪声引发的误检与混淆,针对压入氧化皮类别中,原模型常因背景纹理复杂导致误检,或出现多个检测框重叠的情况。UAL-YOLO 凭借 UAM 模块对特征空间离群点的捕捉和 AGF 模块对浅层细节的选择性筛选,显著降低了信噪比,使模型能准确区分真实缺陷与背景噪声,消除了误检及框体重叠现象^[19]。通过 AGF 引导融合机制,实现了深层语义对浅层细节的主动调控,使得斑点与夹杂等形态多变的缺陷在融合过程中保留了更纯净的特征表达。直接解决了传统模型在复杂工业背景下精度与轻量化难以兼顾的瓶颈。

为了进一步可视化 UAL-YOLO 的优势,在 NEU-DET 数据集上与基础模型 YOLOv11n 进行对比,如图 12 所

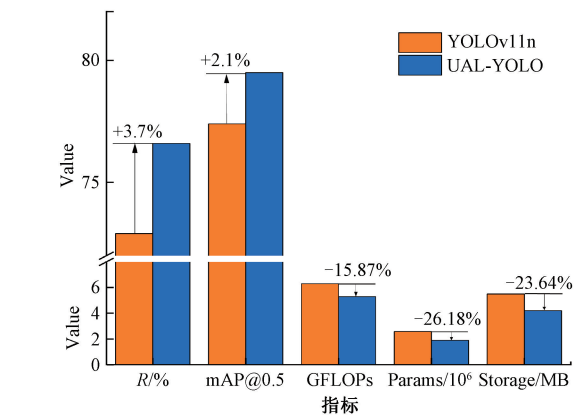


图12 模型检测性能对比
Fig. 12 Comparison of model detection performance metrics

示,UAL-YOLO 模型实现了精度和召回率的双重跨越,召回率提升了 3.7%,精度提升了 2.1%,性能的提升得益于 UAL-YOLO 模型对特征空间离群点的精准捕捉和对微小划痕边缘细节的有效恢复,在面对低对比度、细长裂纹缺

陷时易出现的漏检问题,使检测精度达到了更高维度的稳定。在精度提升的同时,其参数量降低 26.36%,计算量缩减了 15.87%,最终模型存储空间减少了 23.64%。这充分说明了 UAL-YOLO 在有限的算力下,通过重构特征融合机制,实现了比原始网络更高的参数利用效率。有效的解决了轻量化与检测性能难以兼顾的局限性^[20]。模型在大幅缩减权重体积的前提下,依然展现出更强的特征表达能力,能够更好地适配工业现场计算资源受限的实际场景。

4 结 论

针对现有钢材表面缺陷检测模型在追求高精度的同时造成模型结构复杂度过高、难以在资源弱化的工业场景中部署的问题。提出了 UAL-YOLO 模型,首先,在主干网络末端引入了 UAM,最高层语义特征进行 3D 注意力优化,提升了模型对关键信息的捕获能力。其次,设计了一种 AGF 颈部网络,通过深层特征生成引导权重来智能地调控浅层特征的融合过程,实现了更高效、更精准的多尺度特征聚合。最后,采用了 LEO 模块,以动态点采样的全新视角替代传统上采样算子,在保证高效性的前提下,最大程度地恢复了微小缺陷的细节纹理。实验结果表明,本文提出的 UAL-YOLO 模型在 NEU-DET 数据集上的平均精度值达到了 79.5%,提高了 2.1%;召回率提高了 3.7%。同时,模型的参数量、计算量和占用储存容量分别下降了 26.36%、15.87% 和 23.64%,验证了本文方法在提高检测性能的同时实现轻量化方面的双重优势。尽管本文模型获得了良好的效果,但仍存在进一步优化的空间。未来的研究将聚焦于扩充数据集中如麻点、细微划痕等识别难度大的缺陷样本,以提升模型的泛化能力和对困难样本的检测精度。

参考文献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] 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, 2015, 39(6): 1137-1149.
- [4] HE K, GKOIARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [5] TAN M, PANG R, LE Q. Efficientdet: Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [6] 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, 2016: 779-788.
- [7] 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.
- [8] ZHU CH J, LIU L L, ZHU H B. Surface defect detection algorithm for strip steel based on YOLOv8-NFMC [J]. Foreign Electronic Measurement Technology, 2024, 43(7): 97-104.
- [9] 崔晶楠,黄春艳,李艳玲. 融合动态卷积的 YOLOv5s 钢材表面缺陷检测 [J]. 制造业自动化, 2024, 46(11): 126-132.
- [10] CUI J N, HUANG CH Y, LI Y L. Fusion of YOLOv5s and dynamic convolution for steel surface defect detection [J]. Manufacturing Automation, 2024, 46(11): 126-132.
- [11] 曾勇杰,范必双,杨涯文,等. 改进 YOLOv8 算法在风机叶片缺陷检测上的应用 [J]. 电子测量与仪器学报, 2024, 38(8): 26-35.
- [12] ZENG Y J, FAN B SH, YANG Y W, et al. Application of improved YOLOv8 algorithm in fan blade defect detection [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 26-35.
- [13] 高春艳,秦桑,李满宏,等. 改进 YOLOv7 算法的钢材表面缺陷检测研究 [J]. 计算机工程与应用, 2024, 60(7): 282-291.
- [14] GAO CH Y, QIN SH, LI M H, et al. Research on steelsurface defect detection with improved YOLOv7 algorithm [J]. Computer Engineering and Application, 2024, 60(7): 282-291.
- [15] 李思思,葛华勇. 改进 YOLOv7 的钢材表面缺陷检测模型 [J]. 计算机技术与发展, 2024, 34(8): 78-85.
- [16] LI S S, GE H Y. Detection model of steel surface defect based on improved YOLOv7 [J]. Computer Technology and Development, 2024, 34(8): 78-85.
- [17] 徐洪俊,唐自强,张锦东,等. 钢材表面缺陷检测的 YOLOv5s 算法优化研究 [J]. 计算机工程与应用, 2024, 60(7): 306-314.
- [18] XU H J, TANG Z Q, ZHANG J D, et al. Research on optimisation of YOLOv5s algorithm for steel surface defect detection [J]. Computer Engineering and Application, 2024, 60(7): 306-314.

[13] KHANAM R, HUSSAIN M. YOLO11: An overview of the key architectural enhancements[J]. ArXiv preprint arXiv:2410.17725, 2024.

[14] 胡依伦, 杨俊, 许聪源, 等. PIC2f-YOLO: 金属表面缺陷检测轻量化方法[J]. 光电工程, 2025, 52(1): 89-103.
HU Y L, YANG J, XU C Y, et al. PIC2f-YOLO: Lightweight method for metal surface defect detection[J]. Opto-Electronic Engineering, 2025, 52(1): 89-103.

[15] 沈冰星, 黄洪琼. 基于 ESE-YOLO 的钢带表面缺陷检测研究[J]. 电子测量与仪器学报, 2025, 39(8): 126-135.
SHEN B X, HUANG H Q. Research on surface defect detection of steel strip based on ESE-YOLO[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 126-135.

[16] 赵佰亭, 张敏, 贾晓芬. YOLOv8n-CSG: 轻量化钢材表面缺陷检测算法[J]. 电子测量与仪器学报, 2025, 39(8): 115-125.
ZHAO B T, ZHANG M, JIA X F. YOLOv8n-CSG: Lightweight steel surface defect detection algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 115-125.

[17] LI C, XU A, ZHANG Q, et al. Steel surface defect detection method based on improved YOLOX [J]. IEEE Access, 2024, 12: 33748-33759.

[18] ZHAO C, SHU X, YAN X, et al. RDD-YOLO: A modified YOLO for detection of steel surface defects[J]. Measurement, 2023, 214: 112776.

[19] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[20] CUI Y, REN W, CAO X, et al. Focal network for image restoration [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 13001-13011.

作者简介



张新君, 2003 年和 2006 年于辽宁工程技术大学获得学士和硕士学位, 现为辽宁工程技术大学副教授, 主要研究方向为计算机视觉、人工智能安全。
E-mail: 64671356@qq.com

Zhang Xinjun received her B. Sc. and M. Sc. degrees from Liaoning Technical University in 2003 and 2006 respectively. She is currently an associate professor at Liaoning Technical University. Her current research interests include computer vision and artificial intelligent security.



赵志闯 (通信作者), 2024 年于文华学院获得学士学位, 现为辽宁工程技术大学硕士研究生, 主要研究方向为图像处理、缺陷检测。
E-mail: 2470433479@qq.com

Zhao Zhichuang (Corresponding author) received his B. Sc. degree from Wenhua College in 2024. He is now a M. Sc. candidate at Liaoning Technical University. His main research interests include image processing and defect detection.