

基于改进 YOLOv8 的车辆漆面缺陷检测

郝友胜, 文贞慧, 冯小溪, 邓泽华, 黄清宝

(广西大学电气工程学院, 广西 南宁 530000)

摘要: 针对车辆漆面缺陷检测精度低、检测算法参数量大、难易样本不均匀等问题, 提出一种基于改进 YOLOv8 的车辆漆面检测算法。首先, 为了提升划痕缺陷检测能力并降低模型规模, 将 DAT (Deformable Attention Transformer) 注意力机制引入主干网络来增强长距离特征依赖关系, 同时使用幻影卷积 (GhostConv) 替换网络中的卷积 (Conv) 模块。然后, 为了提升特征提取能力并进一步降低模型规模, 结合 FasterBlock 模块与高效多尺度注意力 (EMA) 机制提出 C2f-E (C2f Based on EMA) 模块。接着, 为了提高小目标检测性能, 基于双向特征金字塔网络 (BiFPN) 进行设计, 并增加小目标检测头与多尺度特征融合支路, 提出 BiFPN-D (BiFPN with Small Object Detection Head) 颈部金字塔结构。最后, 为了解决难易样本的平衡问题并提高针对小目标缺陷的检测性能, 使用 WIoUv3 (Wise-Intersection over Union version 3) 作为训练网络的损失函数。在自建的车辆漆面缺陷数据集上进行训练并开展对比实验。实验结果表明, 相较于 YOLOv8n, 改进模型的均值平均精度 (mAP@0.5) 提高了 5.5 个百分点, 规模减小了 1.4×10^6 。

关键词: YOLOv8 算法; 车辆漆面缺陷; 目标检测; 双向特征金字塔网络; 损失函数

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070032

Vehicle Paint Defect Detection Based on Improved YOLOv8

HAO Yousheng, WEN Zhenhui, FENG Xiaoxi, DENG Zehua, HUANG Qingbao

(School of Electrical Engineering, Guangxi University, Nanning 530000, Guangxi, China)

【Abstract】 To address the issues of low accuracy in vehicle paint defect detection, excessive parameters in detection algorithms, and the uneven distribution of easy and hard samples, a vehicle paint detection method based on an improved YOLOv8 is proposed. To enhance scratch defect detection capabilities and reduce model size, a Deformable Attention Transformer (DAT) mechanism is introduced into the backbone network, and Ghost Convolution (GhostConv) replaces the standard Convolution (Conv) modules. Subsequently, to improve feature extraction capabilities and further reduce model size, a C2f Based on Efficient Multiscale Attention (EMA) (C2f-E) module is proposed by combining the FasterBlock module and the EMA attention mechanism. Moreover, to enhance the detection performance for small objects, a network based on the Bidirectional Feature Pyramid Network (BiFPN) is designed. Additionally, by adding a small-object detection head and a multiscale feature fusion branch, a neck pyramid structure named BiFPN with Small Object Detection Head (BiFPN-D) is proposed. Finally, to address the balance issue between difficult and easy samples and improve the detection performance for small object defects, Wise-Intersection over Union version 3 (WIoUv3) is employed as the loss function for training the network. The improved network is trained on a self-built dataset of vehicle paint defect images and subjected to comparative experiments. The results show that, the improved model achieves an increase of 5.5 percentage points in terms of mean Average Precision (mAP@0.5) and a reduction of 1.4×10^6 in terms of parameter count, compared to YOLOv8n.

【Key words】 YOLOv8 algorithm; vehicle paint defect; object detection; Bidirectional Feature Pyramid Network (BiFPN); loss function

0 引言

随着经济的发展, 汽车数量不断增加, 人工汽车喷涂逐渐被机器人自动化喷涂取代。然而, 由于车辆漆面缺陷检测与定位精度不足、速度较慢, 智能化

补漆技术的应用仍面临一定挑战^[1]。同时, 在汽车质检领域, 传统缺陷检测方法存在精度低、易漏检等问题^[2]。因此, 如何高效且准确地进行车辆漆面缺陷检测, 使其能够适用于实际工业场景, 成为一个具有重要研究价值的课题。

基金项目: 国家自然科学基金 (62276072); 广西研究生教育创新计划项目 (JGY2023016); 广西大学研究生教育教学改革计划。

作者简介: 郝友胜 (CCF 学生会员), 男, 硕士研究生, 主研方向为计算机视觉、多模态智能; 文贞慧、冯小溪, 硕士研究生; 邓泽华, 硕士; 黄清宝 (通信作者, CCF 高级会员), 教授、博士。

收稿日期: 2024-06-24

修回日期: 2024-07-23

E-mail: qbhuang@gxu.edu.cn

随着人工智能的发展,智能化缺陷检测逐渐取代了人工缺陷检测。早期主要使用激光检测、阈值分割等方法。WU 等^[3]提出了一种利用差分激光三角测量法检测波纹板表面缺陷的方法。FENG 等^[4]提出一种基于改进阈值分割算法的表面缺陷识别方法,该方法有效提高了检测工件的准确率。然而,这些方法普遍具有价格昂贵、光线需求高、检测区域较小等问题。

近些年来,基于卷积神经网络(CNN)的目标检测算法得到了广泛应用,在一定程度上解决了这些问题,主要分为单阶段和双阶段两类。双阶段检测算法先生成候选区域再进行精细的分类与回归,代表网络有 R-CNN^[5]、SPP-Net^[6]、Faster R-CNN^[7]等。文献[8]提出了一种改进的 Faster R-CNN 算法,该算法融合了 CBAM(Convolutional Block Attention Module)注意力机制和特征金字塔结构,检测精度高,但模型规模大、推理时间长,工业场景下实用性低。单阶段目标检测算法在输入图像上直接完成目标检测,有效提高了检测速度,代表网络有 YOLO^[9]、SSD^[10]等。文献[11]提出了一种 YOLOv8-MDC 算法,该算法采用了特征融合网络与 CA(Coordinate Attention)注意力机制,同时提高了钢材表面缺陷的检测精度与检测速度,但对于小目标缺陷检测效果不佳。文献[12]提出了一种改进的 YOLOv7-tiny 检测算法,该算法提高了钢表面小目标缺陷的检测精度,但没有同时兼顾检测精度和模型规模。文献[13]提出了以 YOLOv8n 作为基准模型的 PGS-YOLO 算法,该算法通过增加小目标检测头以提高手机屏幕瑕疵的检测能力,但针对大尺度缺陷的检测性能不佳。

目前,在众多学者专家的努力下,缺陷检测算法已经取得了较高的准确率。然而,若将现有的缺陷检测算法应用到汽车漆面检测领域,存在两方面不足:一方面,汽车漆面点状缺陷多为小目标而划痕状缺陷多为大尺度目标,现有的缺陷检测网络没有同时针对这两类目标进行设计;另一方面,现有的检测网络在提升模型精度的同时难以降低模型的大小。因此,需要有针对性地网络进行设计,才能适配工业中目标检测任务的需求。

基于以上问题,本文选择将 YOLOv8 作为基础网络,以压缩模型规模、提升模型精确度为方向,针对车辆漆面缺陷检测的应用场景,对网络进行设计。考虑到 YOLOv8n 在参数量方面的优势,本文选择在 YOLOv8n 的基础上进行设计,提出 EDW-YOLOv8,主要贡献包含以下 4 个方面:

1)构建了一个包含 1 684 张图像的数据集,为车辆漆面缺陷检测提供了数据支持。

2)为了在增强划痕状缺陷检测能力的同时降低模型参数量,提出了 C2f-E 模块,用幻影卷积(GhostConv)替换网络中的卷积(Conv)模块,在主干网络中引入 DAT(Deformable Attention Transformer)注意力机制,以加强关注长距离的特征依赖关系。

3)为了增强小目标检测能力,提出双向特征金字塔网络(BiFPN)-D(BiFPN with Small Object Detection Head)颈部网络,首先在浅层网络中增加小目标检测头,然后将颈部网络替换为 BiFPN 的连接方式,并引入加权特征融合方法,最后增加了一路多尺度特征融合支路,以充分利用浅层的位置特征信息。

4)针对车辆漆面小目标缺陷检测难度较大的问题,本文将 WIoUv3(Wise-Intersection over Union version 3)损失函数纳入边界框回归损失中,有效降低了高质量样本和低质量样本的梯度增益。

1 YOLOv8 算法

Ultralytics 团队在 YOLOv5 网络的基础上进行了模型改进,将 YOLOv5 中的 C3 模块替换为 C2f,提出了 YOLOv8 网络,并根据网络深度划分出 n、s、m、l、x 5 种规模。如图 1(a)所示,该网络包含输入端、主干网络、颈部网络和检测头 4 个部分。

输入部分对输入图片进行 Mosaic^[14]数据增强和自适应图片缩放。Backbone 包含 Conv、C2f 和 SPPF(Spatial Pyramid Pooling Fast)等模块,负责从输入图像中提取高级语义信息。颈部网络能够接收来自主干网络不同尺度的特征图,并通过卷积操作或采样操作来确保模型能够同时处理不同分辨率的信息,从而提高目标检测的鲁棒性。检测头负责将经过主干网络和特征金字塔处理的特征图转化为目标检测的预测结果,生成目标的类别概率、边界框位置、置信度得分等信息。

2 EDW-YOLOv8 设计方法

EDW-YOLOv8 网络结构如图 1(b)所示。首先,为了降低模型参数,引入 GhostConv 和 C2f-E 模块。然后,为了增加对大尺度目标的检测能力,引入 DAT 注意力机制放置在主干网络最后一层。再者,为了增加小目标检测能力,新增小目标检测头,并替换 BiFPN 的连接方式。最后,为了提高特征融合能力,新增一条多尺度特征融合支路,从主干网络

的第一个 Conv 模块引出后,与 P2、P3 层的特征信息进行融合,作为小目标检测头的输入。在图 1(b)

中,BiFPN-Contact 2 和 BiFPN-Contact 3 模块分别对输入的 2 和 3 种特征图进行加权特征融合。

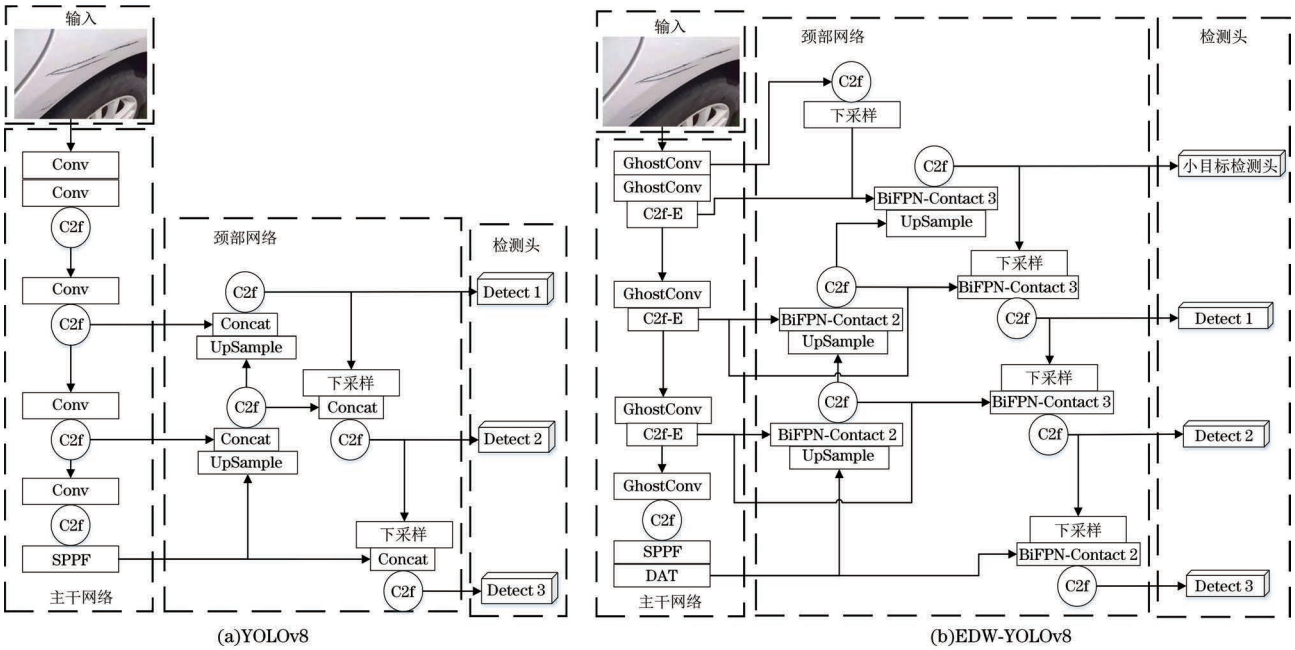


图 1 改进的 YOLOv8 网络结构

Fig.1 Structure of improved YOLOv8 network

2.1 C2f-E 模块

C2f 模块具备良好的特征提取功能,但其复杂度较高且在 YOLOv8 网络中使用频繁,导致网络的参数量增加。为了在提高模型检测性能的同时减小参数量,本文对 C2f 模块进行改进,提出 C2f-E。改进方法如图 2 所示。首先,在 C2f 模块的第一个 Conv 后添加了高效多尺度注意力(EMA)模块,以提升模块的特征提取能力;然后,将 Conv 替换为

GhostConv,以降低模型参数量;最后,将 C2f 中的 Bottleneck 层替换为 FasterBlock,在降低模块的参数量的同时提高特征提取能力。其中,EMA 是文献[15]提出的新型注意力机制,有助于模型理解和利用特征图,增强特征图的语义信息和空间信息。EMA 的结构如图 3 所示。首先,对于任意给定的特征图输入,利用 3 条并行路径来提取每组特征图的注意力权重描述,其中,2 条位于 1×1 分支,1 条

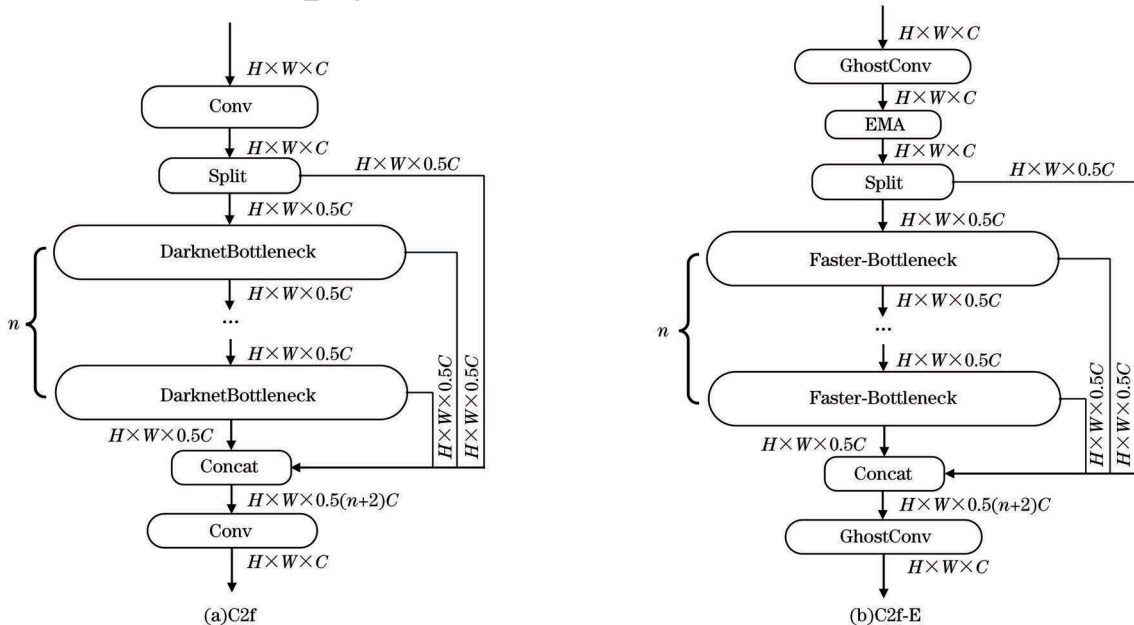


图 2 改进的 C2f 模块结构

Fig.2 Structure of improved C2f module

始特征直接保留下来,不进行额外处理;另一部分通过深度可分离卷积(DSConv)^[19]进行处理,在减少计算量的同时提取复杂特征。最后,将两部分特征结合起来,形成了最终的输出特征图。输出结果既包含了未经处理的原始信息又包含了经过深度处理的复杂信息。

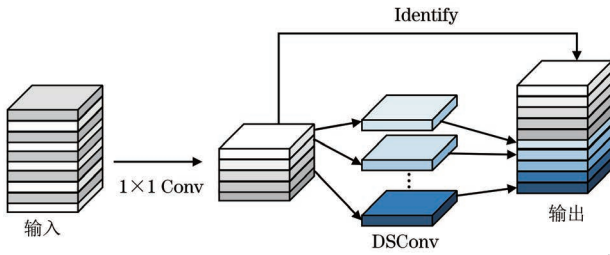


图 5 GhostConv 模块结构

Fig. 5 Structure of GhostConv module

2.2 DAT 注意力机制网络

划痕状缺陷为大尺度目标,其特征分布在图像中相对较远的位置,需要较大卷积核进行特征提取以获得足够的感受野。虽然 YOLOv8 网络通过多层卷积操作逐步扩大感受野,但由于每一层的卷

积核较小,在逐步提取过程中容易丢失特征依赖关系。

为了解决这一问题,本文在 YOLOv8 骨干网络中引入一种 DAT 注意力机制模块^[20],以数据依赖的形式选择注意力机制中键和值的相对位置,从而能够灵活地关注相关区域,捕获更多有用的特征信息。如图 6 所示,DAT 注意力机制包含了 4 个阶段,在输入图像后,经过一个步长为 4 的 4×4 非重叠卷积进行嵌入。在阶段 1 和阶段 2 中,采用窗口局部注意力机制来提取局部特征,并利用 Swin Transformer^[21] 中的移位窗口注意力机制将局部注意力的计算结果在窗口内进行移动与信息整合,以增强特征提取的效果。在阶段 3 和阶段 4 中,特征图像继续通过窗口局部注意力机制进行处理以集中局部信息,并通过可变形注意力(Deformable Attention)机制^[22]来模拟局部信息之间的全局关系,灵活调整注意力的范围和形状,实现局部与全局感受野的有效交替,从而使模型在关注长距离特征关系的同时增加计算效率。

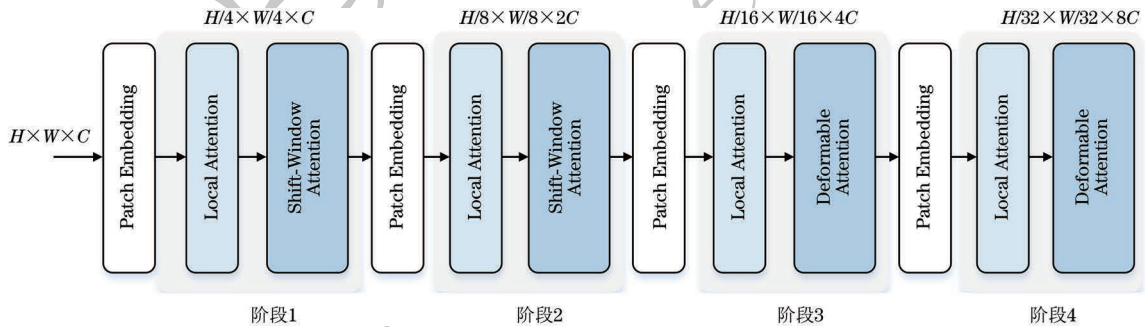


图 6 DAT 模块结构

Fig. 6 Structure of DAT module

2.3 BiFPN-D 颈部网络

YOLOv8 网络经过多次下采样操作压缩图像信息,使网络感受野变大,导致小目标信息被更大的背景信息所淹没。因此,小目标的特征难以被识别出来。为了弥补 YOLOv8 在小目标检测上的不足,本文设计了 BiFPN-D 颈部网络。首先,增加小目标检测头,处理高分辨率的特征图,保留更多的细节信息;然后,设置 BiFPN 带权特征融合网络,将浅层网络的高分辨率细节特征与深层网络的高级语义特征相结合,以提高对各种尺度目标的检测性能;最后,设置了一条多尺度特征融合分支,以获取更浅层的小目标特征信息。

2.3.1 小目标检测头

YOLOv8 的检测头分布在 P3、P4、P5 层,分别提供了 80×80 、 40×40 、 20×20 像素 3 种尺度,用以检测不同大小的目标。若输入图像大小为 $640 \times$

640 像素,使用 80×80 像素的检测头检测时感受野为 8×8 像素、使用 40×40 像素的检测头时感受野为 16×16 像素、使用 20×20 像素的检测头时感受野为 32×32 像素。在构建的汽车漆面缺陷数据集中,小目标点状缺陷较多,其中许多高和宽小于 8 像素,这些缺陷特征难以识别。因此,如图 7(a)所示,为了增加对小目标缺陷的检测能力,本文在 P2 层增加一个 160×160 像素的检测头,其感受野为 4×4 像素。相应地,在颈部网络 P2 层增加一条特征融合支路,检测头与颈部网络之间通过 C2f-E 进行连接。

2.3.2 BiFPN 带权特征融合网络

传统特征融合采用 FPN 网络通过自上而下的方式聚合多尺度特征,受到单向信息流的限制;YOLOv8 在 FPN 的基础上增加了一个自底向上的聚合网络,充分利用了多尺度要素,形成 PAN 网

络。然而,这也增加了更多的参数和计算量。为了在增强小目标检测能力的同时降低模型的参数量,本文采用了 BiFPN^[23] 网络,同时使用加权特征融合的方式,有效地融合来自不同尺度的特征图,显著提升不同大小缺陷的检测精度。

如图 7(b)所示,该结构删除了只有一条输入边的节点,简化了网络结构,并在原始输入和输出节点之间添加额外连接,在融合更多特征信息的同时不会显著增加计算成本。经过 BiFPN 金字塔自上而下传递深层特征的语义信息,自下而上传递浅层特征的位置信息进行双尺度特征融合。此外,在 BiFPN 金字塔网络结构的基础上,采用带权特征融合算法,用于学习不同特征的重要性,并根据特征权

重对输入有区分地进行融合。以 P6 层特征为例,特征融合的输出如下:

$$f_{P6_{th}} = C\left(\frac{\omega_6 f_{P6_{in}} + \omega_7 R_1(f_{P7_{in}})}{\omega_6 + \omega_7 + d}\right) \quad (1)$$

$$f_{P6_{out}} = C\left(\frac{\omega_6 f_{P6_{in}} + \omega_7 R_2(f_{P5_{out}}) + W_6 f_{P6_{th}}}{\omega_6 + W_6 + \omega_7 + d}\right) \quad (2)$$

式中: $f_{P6_{th}}$ 是自顶向下路径第 6 层的特征; $f_{P6_{out}}$ 为自底向上路径第 6 层的特征; ω_6 为第 6 层特征权重; W_6 为自顶向下路径第 6 层权重; C 为卷积操作; R_1 为上采样操作; R_2 为下采样操作; d 为任意常数。

所有的特征融合都以类似的方式构建。

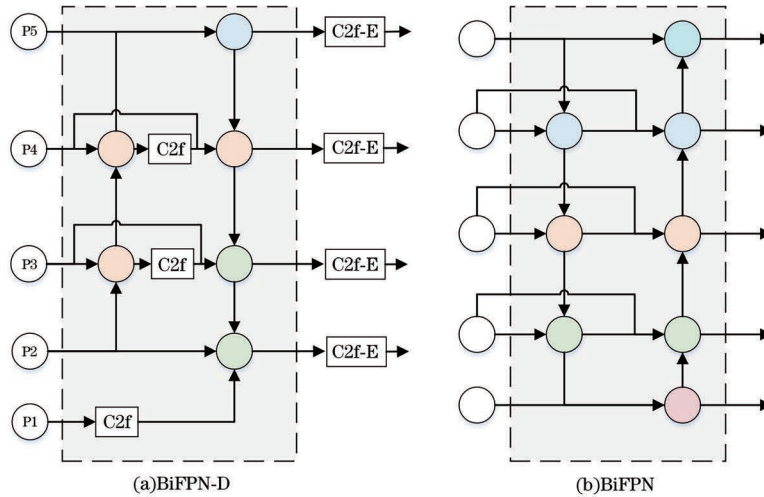


图 7 特征融合模块结构
Fig. 7 Structure of feature fusion module

2.3.3 多尺度特征融合分支

为了充分利用更浅层网络的特征信息,在不过多增加参数量的同时进一步加强对小目标检测的能力,如图 7(a)所示,在颈部网络的 P1 层增加了一条特征融合分支,经过下采样处理后与 P2 层、P3 层的特征信息进行特征融合。这样 BiFPN-D 网络充分利用了浅层网络的小目标特征信息,在不影响大尺度目标检测性能的前提下,显著提升了 YOLOv8 对小目标的检测性能。

2.4 改进的损失函数

在车辆漆面缺陷数据集中,小目标样本占比较高。这些小目标样本的像素数量有限,容易被背景噪声覆盖或与其他物体混淆,导致特征提取难度增加。目前,YOLOv8 使用 CIoU(Complete-IoU) 计算边界框的回归损失^[24],虽然引入了纵横比,但是没有解决难易样本的平衡问题,会出现容易识别的实例占有较大权重的损失值,与识别较为困难的实例竞争。因此,有必要选择更加合适的损失函数。

本文提出使用 WIoU^[25] 替换 YOLOv8 中的 CIoU,使网络充分学习小目标缺陷样本的特征,有效平衡难易样本。在 WIoUv1 中,作者提出 L_{IoU} ,专门用于平衡不同难易度样本所获得的梯度增益。WIoUv1 定义如下:

$$L_{IoU} = 1 - R_{IoU} \quad (3)$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (4)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (5)$$

式中: $R_{WIoU} \in [1, e]$ 显著地降低了高质量锚框对损失函数 L_{IoU} 的影响。

当锚框与真值框重合时, R_{WIoU} 将进一步降低高质量锚框的权重。WIoUv2 在 WIoUv1 的基础上引入了一个非单调聚焦系数,以防止网络过早得到收敛,同时使用了归一化因子 L_{IoU} 加快后期网络的收敛。WIoUv2 定义如下:

$$L_{WIoUv2} = L_{WIoUv1} L_{IoU}^\gamma, \gamma > 0 \quad (6)$$

WIoUv3 提供了一种更灵活的梯度更新策略, 将小梯度增益分配给低质量样本的锚框以减少有害梯度, 并聚焦于普通质量样本的锚框, 有效解决对弱小目标难以检测的问题, WIoUv3 定义如下:

$$\beta = \frac{L_{IoU}}{L_{WIoU}} \in [0, +\infty) \quad (7)$$

$$L_{WIoUv3} = rL_{WIoUv1}, \quad r = \frac{\beta - \alpha}{\beta + 1} \quad (8)$$

式中: β 为离群度; α 以及 r 是引入的非单调聚焦超参数。

3 实验结果与分析

3.1 车辆漆面缺陷数据集

由于 COCO、VOC、ImageNet 等公开数据集没有针对车辆漆面缺陷的数据, 本文控制环境因素, 拍摄了大量的图片, 制作出一套适用于漆面检测的数据集。数据集涵盖了不同大小、不同形状的车辆漆面缺陷来确保样本的丰富性。在删除模糊图片和相似图片后, 按照 7:3 划分训练集和验证集。原始数据共有 421 张图片, 为了进一步扩大训练模型的泛化能力, 提高模型性能, 通过对原始数据进行增强将数据集图片扩增至 1 648 张。图 8(a) 为原始图片示例, 通过翻转操作获得对称特征, 得到图 8(b); 通过裁切操作获得更多不同尺度的缺陷样本, 得到图 8(c); 通过引入高斯噪声来增加训练模型的鲁棒性, 得到图 8(d)。

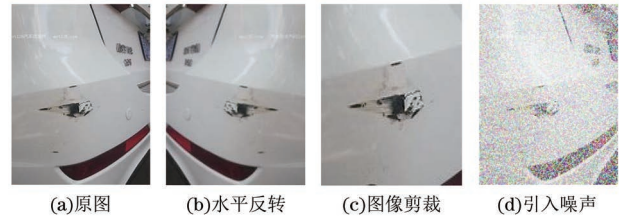


图 8 数据增强示意图

Fig. 8 Data enhancement schematic diagram

在采集好原始图像后, 使用开源的数据标注软件 LabelImg 进行人工标注, 标注示例如图 9 所示, 针对点状类缺陷标注为 Flaw, 针对划痕状缺陷标注为 Scratch。然后, 以矩形勾勒出车辆漆面缺陷实例的边缘, 矩形对应的合围区域即为车辆漆面缺陷区域。

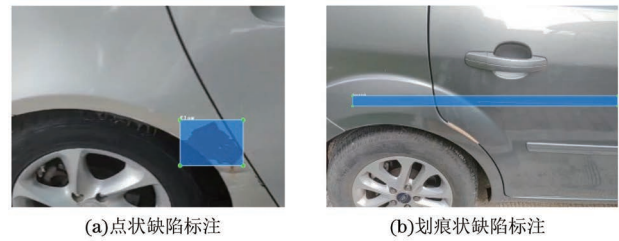
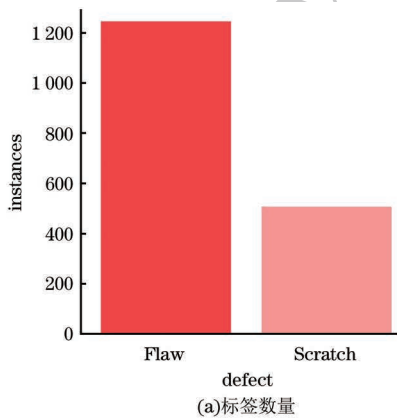


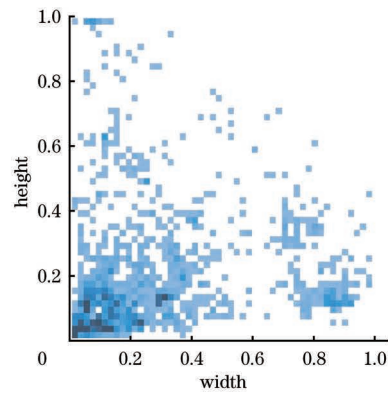
图 9 数据集标注示例

Fig. 9 Examples of dataset annotations

为了有针对性地设计网络结构, 本文统计了数据集标签数量与样本尺度分布情况。如图 10 所示, 统计结果表明不同种类缺陷尺度相差较大, 存在小尺度目标与大尺度目标缺陷。



(a) 标签数量



(b) 标签大小分布

图 10 标签数量与缺陷大小分布

Fig. 10 Number of tags and distribution of defect sizes

3.2 模型评价指标

本文使用精确率 (Precision, P) 和召回率 (Recall, R) 作为漆面检测的评价指标。其中, 精确率是在所有被分类为正类别的样本中实际为正类别样本的比例, 召回率是在所有实际为正类别的样本中, 被正确预测为正类别样本的比例, 计算公式分别如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (9)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (10)$$

式中: N_{TP} 表示实际目标为车辆漆面缺陷且被正确地预测为车辆漆面缺陷的数量; N_{FP} 表示实际目标为背景但被错误地预测为车辆漆面缺陷的数量;

N_{FN} 表示实际目标为车辆漆面缺陷但被错误地预测为背景的数量。

使用平均精度 (AP, P_{AP}) 衡量精确率-召回率曲线下的面积, 将每个类别的平均精度取均值即得到均值平均精度 (mAP, P_{mAP}), 公式如下:

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

$$P_{mAP} = \frac{\sum_{i=1}^N P_{AP}}{N} = N^{-1} \sum_{i=1}^N \int_0^1 P(R) dR \quad (12)$$

式中: N 为图片中的物体类别。

3.3 实验环境与参数配置

在实验环境方面, 使用的是 Windows 10 操作系统、ADM Ryzen 7 5800H with Radeon Graphics 3.20 GHz CPU、NVIDIA Geforce RTX 3060 显卡和 16 GB 内存。在参数配置方面, 采用 PyTorch 2.10 深度学习框架, 输入图像尺寸为 640×640 像素, 迭代次数 (epoch) 为 200, 批处理大小 (batch size) 为 16, 初始学习率 (lr) 为 0.01, 权重衰减因子 (weight_decay) 为 0.0005, 学习率动量 (momentum) 为 0.937。

3.4 相关模块性能对比

3.4.1 卷积性能比较

为了比较 Conv 模块的性能, 分别用 DWConv^[19]、GhostConv、AKConv^[26]、PConv^[17] 替换 YOLOv8 网络中的 Conv 模块, 并在车辆漆面缺

陷数据集上训练。训练结果如表 1 所示, 与其他轻量化卷积相比, GhostConv 综合性能表现最优。

表 1 Conv 模块性能对比

Table 1 Comparison of Conv module performance

Model	mAP@0.5/ %	Params/ 10^6	FLOPs/ 10^9	Model size/ 10^6
Conv	79.4	3.0	8.2	6.2
DWConv	78.8	2.4	7.3	4.7
AKConv	79.2	2.6	7.9	5.0
PConv	80.1	2.9	8.2	6.1
GhostConv	80.3	2.7	7.6	4.9

3.4.2 注意力机制性能比较

为了测试 DAT 注意力机制对划痕状缺陷的检测作用, 并验证 EMA 机制的优越性, 本文进行了注意力机制模块的对比实验。在 YOLOv8 主干网络中引入 SE (Squeeze-and-Excitation)^[27]、CA^[28]、CBAM^[29] 等注意力机制模块, 并在车辆漆面缺陷数据集上进行了对比实验。

实验结果如表 2 所示, 可以看出: 引入 EMA 机制后, 在参数量不变的前提下, mAP@0.5 提升了 1.9 个百分点; 引入 DAT 注意力机制后, 在不影响小目标检测性能和模型规模的前提下, 针对划痕状缺陷的 mAP@0.5 提升了 7.3 个百分点。

表 2 注意力机制性能对比

Table 2 Performance comparison of attention mechanism

Model	mAP@0.5/%	mAP@0.5(Flaw)/%	mAP@0.5(Scratch)/%	Params/ 10^6	FLOPs/ 10^9	Model size/ 10^6
Base	79.4	77.8	83.8	3.0	8.2	6.2
SE	78.4	76.1	84.0	3.0	8.3	6.3
CA	80.4	77.5	88.9	3.0	8.3	6.3
CBAM	79.8	78.0	86.5	3.1	8.3	6.3
EMA	81.3	79.1	87.8	3.0	8.2	6.3
DAT	82.2	78.9	91.1	3.2	8.4	6.6

3.5 模型对比

为了验证 EDW-YOLOv8 与 YOLOv8 的优越性, 本文分别将其与传统目标检测模型和缺陷检测模型进行对比, 对比结果如表 3 所示。未改进的 YOLOv8n(Base) 算法模型结合了 YOLO 系列以往算法模型的优势, 对比 YOLOv5n、YOLOv3-tiny、YOLOv3-spp 在精度上有明显的提升。文献[11]在 YOLOv8n 的基础上引入特征融合方法与 CA 注意力机制, 明显提升了对划痕状缺陷的检测效果, 但对小目标缺陷的检测效果不佳。文献[12]在 YOLOv7-

tiny 的基础上引入 Conv 模块, 对两种缺陷的检测效果都有所提升, 但相较于 YOLOv8n(Base) 模型规模增大 3.0×10^6 。文献[13]在 YOLO7-tiny 的基础上设计小目标检测头, 并引入 SE 注意力机制和 PConv 卷积, 与本文改进思路相似, 在简化模型规模的同时提升了网络对小目标的检测能力, 但针对划痕状缺陷的检测精度不佳。EDW-YOLOv8 在提高两种漆面缺陷检测效果的同时, 模型规模对比 YOLOv8n(Base) 减小 1.4×10^6 , mAP@0.5 提高 5.5 个百分点, 在所有模型中综合性能表现最优。

表 3 目标检测算法模型对比

Table 3 Comparison of object detection algorithm models

Model	mAP@0.5/%	mAP@0.5(Flaw)/%	mAP@0.5(Scratch)/%	Params/10 ⁶	FLOPs/10 ⁹	Model size/10 ⁶
YOLOv3-tiny	69.1	67.3	74.8	12.1	18.9	12.2
YOLOv3-spp	76.2	74.3	81.5	20.3	283.8	41.4
YOLOv5n	71.1	70.5	73.2	2.5	7.2	5.1
YOLOv9c	83.5	80.7	88.2	22.4	102.0	10.3
YOLOv8n ^[11]	80.1	76.1	88.3	3.1	9.3	5.3
YOLOv7-tiny ^[12]	82.1	81.6	83.7	6.1	12.4	9.2
YOLOv8n ^[13]	80.1	79.3	83.2	2.0	10.8	4.4
YOLOv8s	83.0	81.7	85.1	11.1	28.6	11.5
YOLOv8n(Base)	79.4	77.8	83.8	3.1	8.2	6.2
EDW-YOLOv8	84.9	83.6	89.1	2.2	12.9	4.8

VOC-2012 数据集样本分布情况如图 11 所示, 提供了 5 717 张训练集与 5 823 张验证集, 同时包含了 20 个类, 适用于模型的性能评估。为了测试 EDW-YOLOv8 的通用性, 选择不同网络在公共数据集 VOC-2012 上训练 200 个 epoch。YOLOv5n、

YOLOv8n、EDW-YOLOv8、YOLOv8s 的训练结果如表 4 所示。EDW-YOLOv8 的 mAP@0.5 与 YOLOv5n 相比提高 5.5 个百分点、与 YOLOv8n 相比提高 2.1 百分点。这证明 EDW-YOLOv8 在通用数据集上也表现出一定的优越性。

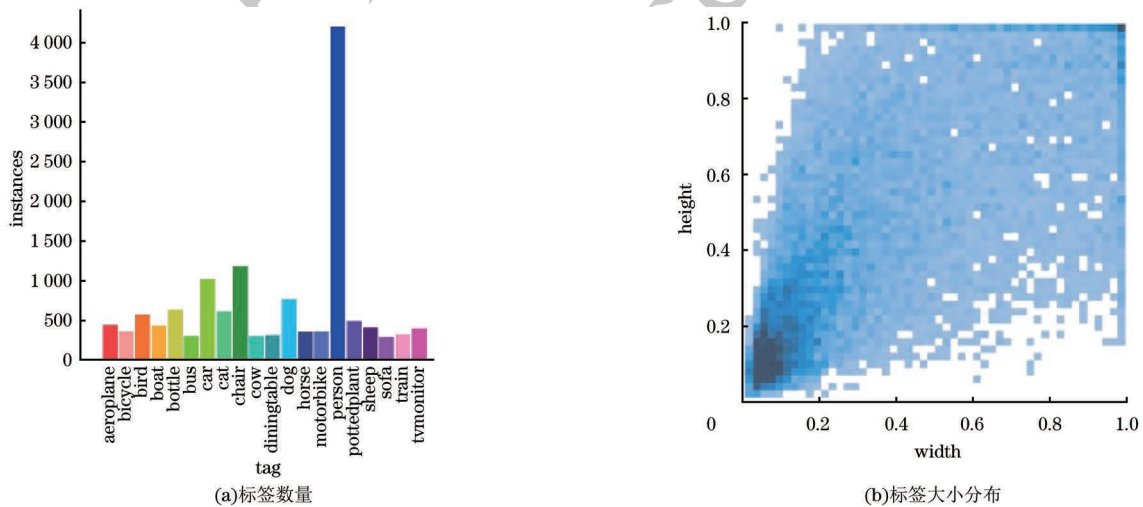


图 11 VOC-2012 数据集样本分布情况

Fig.11 Sample distribution of VOC-2012 dataset

表 4 通用性对比实验

Table 4 Comparative experiment of generality

Model	mAP@0.5/ %	Params/ 10 ⁶	FLOPs/ 10 ⁹	Model size/ 10 ⁶
YOLOv5n	58.7	2.5	7.1	5.3
YOLOv8n	62.1	3.0	8.2	6.3
YOLOv8s	65.6	11.1	28.6	11.5
EDW-YOLOv8	64.2	2.2	12.9	4.8

为了进一步验证 EDW-YOLOv8 网络的优越性, 在实际场景中进行了对比实验。实验结果如

图 12 所示, 针对小目标与大目标的检测效果均优于其他基础网络。

3.6 消融实验

3.6.1 BiFPN-D 网络消融实验

为了验证本文所提 BiFPN-D 的有效性, 选择 YOLOv8n 作为基础模型, 逐步引入改进部分。实验结果如表 5 所示, 其中 A 代表添加小目标检测头, B 代表引入 BiFPN 结构, C 代表新增特征融合分支, D 代表引入 GhostConv 模块和 C2f-E 模块。实验结果表明, BiFPN-D 有效地提升了小目标缺陷检测性能, 并减小了模型规模。



图 12 模型检测效果对比
 Fig. 12 Comparison of model detection effect

表 5 BiFPN-D 网络消融实验
 Table 5 Ablation experiments on BiFPN-D network

Model	mAP@0.5/%	mAP@0.5(Flaw)/%	mAP@0.5(Scratch)/%	Params/ 10^6	FLOPs/ 10^9	Model size/ 10^6
Base	79.4	77.8	83.8	3.0	8.2	6.2
+A	82.3	81.1	84.7	3.0	12.6	6.5
+A+B	82.7	80.3	86.2	3.1	12.5	6.5
+A+B+C	83.8	84.1	82.9	3.4	13.1	6.9
+A+B+C+D(Ours)	83.3	83.8	81.6	2.5	11.7	5.8

3.6.2 EDW-YOLOv8 网络消融实验

为了验证改进方法对 EDW-YOLOv8 网络的影响,本文设置了一系列消融实验。保持相同的实验环境,在车辆缺陷数据集上进行训练,根据模型评价标准对比模型在加入不同模块后的优化效果。

为了展现各个改进模块的有效性,分别在 YOLOv8n 中引入 DAT、GhostConv、C2f-E、BiFPN-D 与基础模型进行对比。实验结果如表 6 所示,其中,

“√”表示使用该模块,“×”表示不使用该模块。由表 6 可知:单独引入 DAT 注意力机制后,mAP@0.5 提升 2.8 百分点,模型规模增大 0.4×10^6 ;单独引入 GhostConv 模块后,mAP@0.5 提升 0.9 百分点,模型规模下降 1.3×10^6 ;单独将网络中 C2f 模块替换为 C2f-E 模块后,mAP@0.5 提升 2.0 百分点,模型规模下降 1.0×10^6 ;单独将颈部网络替换为 BiFPN-D 结构后,mAP@0.5 提升 3.9 百分点,模型规模下降 0.4×10^6 。

表 6 EDW-YOLOv8 消融实验结果

Table 6 Results of ablation experiments for EDW-YOLOv8

Model	DAT	GhostConv	C2f-E	BiFPN-D	Wiou	mAP@0.5/ %	Precision/ %	Recall/ %	Params/ 10^6	FLOPs/ 10^9	Model size/ 10^6
YOLOv8n(Base)	×	×	×	×	×	79.4	82.0	72.1	3.0	8.2	6.2
YOLOv8n+DAT	√	×	×	×	×	82.2	85.7	77.5	3.3	8.4	6.6
YOLOv8n+GhostConv	×	√	×	×	×	80.3	77.3	83.4	2.7	7.6	4.9
YOLOv8n+C2f-E	×	×	√	×	×	81.4	82.6	80.5	2.4	7.3	5.2
YOLOv8+BiFPN-D	×	×	×	√	×	83.3	87.3	75.2	2.5	11.7	5.8
YOLOv8+Wiou	×	×	×	×	√	80.3	76.9	84.6	3.0	8.2	6.2
YOLOv8+BiFPN+Wiou	×	×	×	√	√	84.5	79.2	86.2	2.5	11.7	5.8
EDW-YOLOv8	√	√	√	√	√	84.9	80.3	85.8	2.2	12.9	4.8

为了展现改进部分的相容性,本文在 YOLOv8 的基础上逐步加入改进方法。首先,将损失函数更换为 WIoU,并在此基础上将颈部网络替换为 BiFPN-D 后, mAP@0.5 提升 3.2 百分点、模型规模下降 0.4×10^6 ; 然后,引入骨干部分网络改进后, mAP@0.5 提升 0.4 百分点,模型规模下降 1.0×10^6 。实验结果证明了各个模块在网络中均有正向作用且相互兼容。

4 结束语

本文构建了包含 1 684 张图像的数据集,并在 YOLOv8n 模型的基础上进行了改进,针对车辆漆面缺陷检测存在的问题,分别针对小目标缺陷与划痕缺陷进行了设计,同时考虑了难易样本的平衡问题,提出一种基于改进 YOLOv8 的车辆漆面检测算法,为车辆漆面缺陷检测提供了一种有效的解决方案。本文通过对比实验验证了改进算法的优越性和通用性,通过消融实验验证了改进算法的有效性与相容性。为降低过拟合的风险,未来本团队将进一步扩充数据集图片数据,并添加不同种类的缺陷标签,以提高模型的鲁棒性。

参考文献

- [1] PERERA R, GUZZETTI D, AGRAWAL V. Optimized and autonomous machine learning framework for characterizing pores, particles, grains and grain boundaries in microstructural images[J]. *Computational Materials Science*, 2021, 196: 110524.
- [2] 陆玉凯, 袁帅科, 熊树生, 等. 汽车漆面缺陷高精度检测系统[J]. *吉林大学学报(工学版)*, 2024, 54(5): 1205-1213.
- [3] LU Y K, YUAN S K, XIONG S S, et al. High precision detection system for automotive paint defects[J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2024, 54(5): 1205-1213. (in Chinese)
- [4] WU C X, CHEN B J, YE C S. Detecting defects on corrugated plate surfaces using a differential laser triangulation method[J]. *Optics and Lasers in Engineering*, 2020, 129: 106064.
- [5] FENG X L, GAO X W, LUO L. A surface defect identification method based on improved threshold segmentation algorithm[J]. *Journal of Physics: Conference Series*, 2020, 1651(1): 012072.
- [6] 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*. Washington D. C., USA: IEEE Press, 2014: 580-587.
- [7] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [8] REN S 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*, 2017, 39(6): 1137-1149.
- [9] 圣文顺, 余熊峰, 林佳燕, 等. 融合注意力与特征金字塔的小尺度目标检测算法[J]. *计算机工程*, 2024, 50(1): 242-250.
- [10] SHENG W S, YU X F, LIN J Y, et al. Small-scale object detection algorithm integrating attention and feature Pyramids[J]. *Computer Engineering*, 2024, 50(1): 242-250. (in Chinese)
- [11] 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)*. Washington D. C., USA: IEEE Press, 2016: 779-788.
- [12] BERG A C, FU C Y, SZEGEDY C, et al. SSD: single shot multibox detector [C] // *Proceedings of European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 21-37.
- [13] 彭菊红, 张弛, 高谦, 等. 基于改进的 YOLOv8 算法的钢材缺陷检测[J]. *计算机工程*, 2025, 51(7): 152-160.
- [14] PENG J H, ZHANG C, GAO Q, et al. Steel defect detection based on improved YOLOv8 algorithm [J]. *Computer Engineering*, 2025, 51(7): 152-160. (in Chinese)
- [15] 阳丽莎, 李茂军, 胡建文, 等. 基于改进 YOLOv7-tiny 的带钢表面缺陷检测算法[J]. *计算机工程*, 2025, 51(1): 208-215.
- [16] YANG L S, LI M J, HU J W, et al. Strip steel surface defect detection algorithm based on improved YOLOv7-tiny [J]. *Computer Engineering*, 2025, 51(1): 208-215. (in Chinese)
- [17] 周思瑜, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 的手机屏幕瑕疵检测算法: PGS-YOLO [J]. *计算机工程*, 2025, 51(5): 326-339.
- [18] ZHOU S Y, XU H Y, ZHU X Z, et al. Mobile phone screen defect detection algorithm based on improved YOLOv8n: PGS-YOLO [J]. *Computer Engineering*, 2025, 51(5): 326-339. (in Chinese)
- [19] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS [J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.
- [20] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning [C] // *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Washington D. C., USA: IEEE Press, 2023: 1-5.
- [21] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington D. C., USA: IEEE Press, 2023: 12021-12031.
- [22] MA X L, GUO F M, NIU W, et al. PConv: the missing but desirable sparsity in DNN weight pruning for real-time execution on mobile devices [C] // *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, USA: AAAI Press, 2020: 5117-5124.
- [23] CAO J S, BAO W S, SHANG H X, et al. GCL-YOLO: a GhostConv-based lightweight YOLO network for UAV small object detection[J]. *Remote Sensing*, 2023, 15(20): 4932.
- [24] GUO Y H, LI Y D, WANG L Q, et al. Depthwise convolution is all you need for learning multiple visual domains [C] // *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, USA: AAAI Press, 2019: 8368-8375.
- [25] XIA Z F, PAN X R, SONG S J, et al. Vision Transformer with deformable attention [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington D. C., USA: IEEE Press, 2022:

- 4784-4793.
- [21] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2021: 9992-10002.
- [22] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable Transformers for end-to-end object detection [EB/OL]. [2024-05-08]. <https://arxiv.org/abs/2010.04159>.
- [23] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 10781-10790.
- [24] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 12993-13000.
- [25] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [EB/OL]. [2024-05-08]. <https://arxiv.org/pdf/2301.10051.pdf>
- [26] ZHANG X, SONG Y, SONG T, et al. AKConv: convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters [EB/OL]. [2024-05-08]. <https://arxiv.org/abs/2311.11587v1>.
- [27] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D.C., USA: IEEE Press, 2018: 7132-7141.
- [28] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2021: 13708-13717.
- [29] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the 2018 European Conference on Computer Vision. Berlin, Germany: Springer International Publishing, 2018: 3-19.

文字编辑 陆燕菲
栏目编辑 赖玉玲