计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(1):179−189 [doi: 10.15888/j.cnki.csa.009737] [CSTR: 32024.14.csa.009737]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

改进 YOLOv8 的道路损伤检测①

王瀚毅, 李春彪, 宋　衡

(南京信息工程大学 人工智能学院, 南京 210044)
通信作者: 李春彪, E-mail: goontry@126.com

摘　要: 针对道路损伤检测面临的多尺度目标、复杂的目标结构、样本分布不均及难易样本对边界框回归的影响等问题, 本研究提出了一种基于改进 YOLOv8 的道路损伤检测算法. 该方法通过引入动态蛇形卷积 (dynamic snake convolution, DSConv) 替代原有 C2f (faster implementation of CSP bottleneck with 2 convolutions) 模块中的部分 Conv, 以自适应聚焦于细小而曲折的局部特征, 增强对几何结构的感知. 在每个检测头前引入高效多尺度注意力 (efficient multi-scale attention, EMA) 模块, 实现跨维度交互, 捕获像素级别关系, 提升对复杂全局特征的泛化能力. 同时, 增设小目标检测层以提高小目标检测精度. 最后, 提出 Flex-PIoUv2 策略, 通过线性区间映射和尺寸适应性惩罚因子, 有效缓解样本分布不均和锚框膨胀问题. 实验结果表明, 该改进模型在 RDD2022 数据集上的 $F1$ 分数、平均精度均值 ($mAP50$、$mAP50$-95) 分别提高了 1.5 百分点、2.1 百分点和 1.2 百分点. 此外, 在 GRDDC2020 和 China road damage 数据集上的验证结果显示, 该算法具有良好的泛化性.

关键词: 目标检测; YOLOv8; 道路损失检测; 改进损失函数; 边界框回归

引用格式: 王瀚毅,李春彪,宋衡.改进 YOLOv8 的道路损伤检测.计算机系统应用,2025,34(1):179–189. http://www.c-s-a.org.cn/1003-3254/9737.html

Road Damage Detection with Improved YOLOv8

WANG Han-Yi, LI Chun-Biao, SONG Heng

(School of Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: This study proposes an algorithm for road damage detection based on an improved YOLOv8 to address challenges in road damage detection, including multi-scale targets, complex target structures, uneven sample distribution, and the impact of hard and easy samples on bounding box regression. The algorithm introduces dynamic snake convolution (DSConv) to replace some of the Conv modules in the original faster implementation of CSP bottleneck with 2 convolutions (C2f) module, aiming to adaptively focus on small and intricate local features, thereby enhancing the perception of geometric structures. By incorporating an efficient multi-scale attention (EMA) module before each detection head, the algorithm achieves cross-dimensional interaction and captures pixel-level relationships, improving its generalization capability for complex global features. Additionally, an extra small object detection layer is added to enhance the precision of small object detection. Finally, a strategy termed Flex-PIoUv2 is proposed, which alleviates sample distribution imbalance and anchor box inflation through linear interval mapping and size-adaptive penalty factors. Experimental results demonstrate that the improved model increases the $F1$ score, $mAP50$, and $mAP50$-95 on the RDD2022 dataset by 1.5%, 2.1%, and 1.2%, respectively. Additionally, results on the GRDDC2020 and China road damage datasets validate the strong generalization of the proposed algorithm.

Key words: object detection; YOLOv8; road damage detection; improved loss function; bounding box regression

① 基金项目: 国家自然科学基金 (62371242); 南京信息工程大学引入人才启动基金 (2023r061)
收稿时间: 2024-06-12; 修改时间: 2024-07-18; 采用时间: 2024-07-25; csa 在线出版时间: 2024-11-15
CNKI 网络首发时间: 2024-11-18

道路损伤对行人和交通安全构成了重大挑战. 传统的手动目视检查方式不仅耗时费力, 还可能干扰交通流动, 并增加检查人员的安全风险. 由于专家资源有限和财政资源紧缺, 许多地方政府难以及时开展必要的检查工作. 尽管一些市政机构尝试通过高性能传感器实现自动化检测, 但高昂的成本限制了其在大规模道路网络上的应用. 因此, 迫切需要一种成本低且精度高的道路损伤检测系统.

深度学习技术在图像分析中取得了突破, 为交通工程领域的应用提供了新的可能性. 深度学习模型 R-CNN (region-based convolutional neural network)[1]、Faster R-CNN[2]、YOLO (you only look once)[3]和 SSD (single shot detector)[4]能够处理大规模复杂数据集, 并从原始数据中直接提取有用特征和模式. 这些模型支持端到端训练过程, 显著提高了道路损伤检测与分类的效率和准确度, 对公共安全和经济发展产生了积极影响.

Zhang 等[5]提出的基于稀疏表示和多特征融合的路面裂缝检测方法, 尽管在检测精度上有所提升, 但无法满足实时处理的需求. 同年, Tarabalka 等[6]提出的基于流域的裂缝检测算法, 通过分割方法来检测裂缝, 并通过消除噪声提高了识别效果. 该方法有效减少了噪音干扰, 但分割算法在处理复杂背景时可能存在挑战.

Zhang 等[7]将深度学习应用于道路损伤检测, 利用廉价智能手机收集数据, 并使用深度卷积神经网络 (convolutional neural network, CNN) 对图像贴片进行分类. 该方法降低了硬件成本, 但在噪声环境下的检测精度依赖于网络的鲁棒性. Pereira 等[8]提出了基于 CNN 的道路坑洼检测算法. 与传统的机器学习算法相比, 该方法在检测精度上有显著提升, 但在光照变化的情况下, 坑洼检测精度仍然较低.

Feng 等[9]利用深度学习技术进行裂缝检测, 通过识别像素块并保留裂缝的形状、长度和方向等信息. 虽然有效保留了裂缝的几何特征, 但在处理较大面积时可能需要更多计算资源. Nie 等[10]采用 Faster R-CNN 进行道路损伤检测, 并使用迁移学习进行模型优化, 但 Faster R-CNN 计算开销较高. 2020 年, Du 等[11]利用车载摄像头和轻量化改进的 YOLOv5 模型进行损伤检测. 该方法显著减小了模型参数体积, 但在高分辨率图像处理上仍需优化.

Katsaliros 等[12]利用四元数神经网络 (quaternion neural network, QNN)[13]进行裂缝检测, 但四元数神经网络的计算复杂度较高, 可能影响实时性. 李松等[14]引入 bottleneck Transformer 和 Ghost 模块使得模型轻量化, 在保持高性能的同时, 具有更低的计算成本和更小的模型尺寸, 但对细小曲折的裂缝检测精度较低.

近两年, 魏陈浩等[15]引入可变形卷积 (deformable convolution, DConv) [16]并增加了小目标检测头, 从而增强复杂背景下的目标学习能力, 武兵等[17]提出 YOLOv8-RDD 算法, 结合了 DConv、自注意力机制和坐标注意力机制, 以提高检测精度. 胥铁峰等[18]通过融合 Dconv 和改进的分类损失函数, 进一步提升了检测精度. 然而, Dconv 缺乏明确的约束条件, 可能导致网络对感知区域无目标漫游. 王海群等[19]提出了重参数化 YOLOv8 算法, 通过融合多尺度特征和重参数化技术, 提升了模型的检测精度. 但这些模型仅在国内相似道路上进行训练, 缺乏在复杂条件和多种地貌下的鲁棒性和泛化能力. Zeng 等[20]通过引入大可分离核注意力机制 (large separable kernel attention, LSKA) 和轻量化卷积与检测头来提高检测精度, 但精度提升较小.

Ultralytics 团队于 2023 年推出的 YOLOv8 版本[20], 以其实时性、高检测精度和相对轻量的网络结构, 成为道路损伤检测研究的理想选择. 因此, 针对道路损伤检测存在的问题和相关工作的不足, 本文基于 YOLOv8 模型, 提出了一种改进 YOLOv8 的道路损伤检测算法, 主要贡献包括以下 4 个方面.

(1) 在用于语义信息传递的 C2f 模块中, 将部分 Conv 替换为 DSConv[21], 通过自适应聚焦于细小且弯曲的局部特征, 增强了模型对复杂几何结构的感知能力.

(2) 在每个检测头前部署 EMA 模块[22], 利用跨维度的交互, 捕捉像素级别的关系, 确保空间语义特征在每个特征组内均匀分布, 提升对复杂全局特征的泛化能力.

(3) 新增小目标检测层, 专门针对小尺度目标进行优化, 提高模型的识别精度.

(4) 提出 Flex-PIoUv2 策略来替代原模型中的边界框损失函数, 结合线性区间映射和尺寸适应性的惩罚因子, 有效应对样本分布不均和难易样本对边界框回归的影响以及锚框膨胀问题, 进一步提高模型的检测性能.

1 改进 YOLOv8 算法介绍

考虑到模型尺寸的限制, 本研究选用了体积小且精度较高的 YOLOv8n 网络. YOLOv8n 模型的构架分为 4 大部分: 输入层 (input) 、主干网络 (backbone) 、颈部网络 (neck) 以及头部网络 (head). 通过一系列针对性的改进, 模型的结构如图 1 所示.

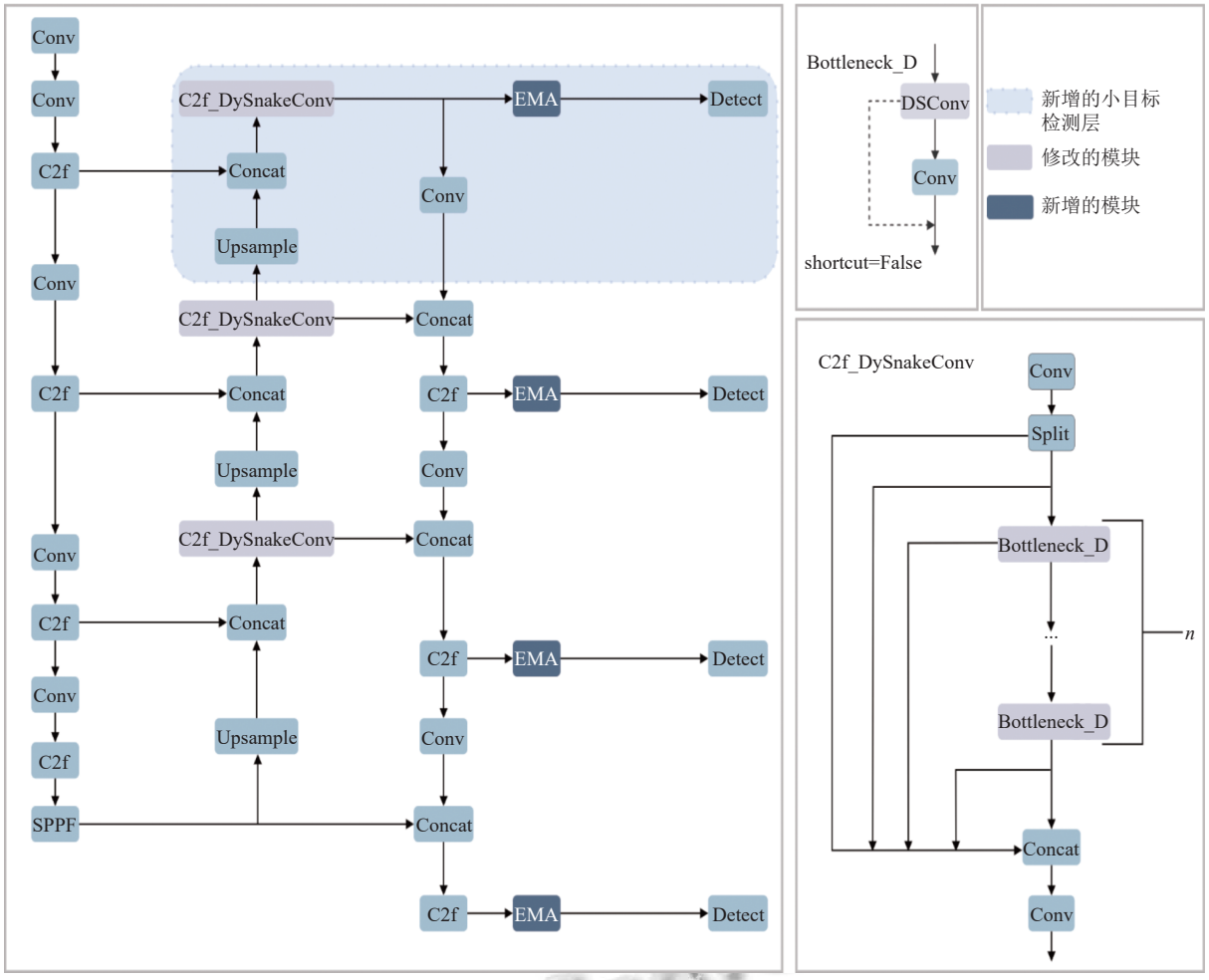

图 1 改进后的 YOLOv8 模型

为有效解决道路损伤检测中目标的细小及复杂的局部结构问题, 本研究引入了 DSConv. DSConv 通过其灵活的感受野, 并在自由学习过程中引入特定的约束, 确保对目标的连续聚焦, 避免因过大的变化偏移引起的感知范围过度扩散. DSConv 的灵活感受野、动态坐标计算和特定约束机制, 不仅提高了模型的几何结构感知能力, 还显著提升了整体检测性能.

此外, 为应对目标全局特征的复杂性和多样性, 本研究在各检测头前部署了 EMA 模块. EMA 模块通过将通道维度拆分为若干子特征组, 使每个组学习独特的语义信息, 从而强化对关注区域的特征表征. 同时, 双并行分支结构利用不同尺度的局部感受野收集空间信息, 实现空间语义特征在各特征组内的均匀分布, 显著提高了模型对复杂多变全局特征的泛化能力.

针对多尺度目标检测的挑战, 本研究增加了小尺度目标的检测层, 通过深层特征向浅层特征的有效传递及融合, 保留了更丰富的小目标轮廓及位置信息, 显著提高了模型在小目标检测精度上的表现, 有效弥补了原模型在此方面的局限.

最后, 为了解决样本分布不均和难易分布对边界框回归的影响以及锚框回归中的膨胀问题, 本研究提出了 Flex-PIoUv2 策略. 该策略通过线性区间映射方法增强了对不同质量锚框的区分能力, 使得损失函数能够更有效地聚焦于难易样本的处理. 此外, Flex-PIoUv2

通过调整参数来更好地处理样本不均衡问题, 确保模型在训练过程中对所有样本类型进行均衡学习. 同时, 引入具有尺寸适应性的惩罚因子, 精确引导锚框回归, 避免锚框不必要的膨胀, 提升了边界框回归的准确性.

2 算法改进

2.1 C2fDS 模块

传统卷积无法根据目标形状变化进行调整, 导致在细小结构检测方面表现不佳. DConv 通过其灵活的感受野和动态调整机制, 确保感受野集中在目标区域, 避免了传统方法中因固定感受野导致的信息丢失和误检问题. 然而, DConv 缺乏明确的约束条件, 可能导致网络对感知区域无目标漫游, 影响检测精度.

DSConv 在自由学习过程中引入了特定的约束条件, 这些约束有效引导卷积核在目标区域进行精确聚焦, 避免了对无关区域的漫游, 从而提升了模型的检测精度. 这些特定的约束条件不仅提高了细小结构的检测精度, 还能有效控制计算量, 确保模型在有限硬件资源上的高效运行.

DSConv 模块包含 4 个主要卷积操作: 普通卷积、自适应卷积、偏移卷积和 1×1 卷积. 假设输入特征图的大小为$H\times W$, 输入通道数为C_{in}, 输出通道数为C_{out}, 卷积核的大小为$K\times K$. 普通卷积和 1×1 卷积的计算复杂度分别为$\mathrm{O}(H\times W\times K^2\times C_{\text{in}}\times C_{\text{out}})$和$\mathrm{O}(H\times W\times 1^2\times C_{\text{out}}^2)$自适应卷积包含偏移卷积和批归一化, 计算复杂度为$\mathrm{O}(H\times W\times 3\times K\times C_{\text{in}}\times 2\times K)+\mathrm{O}(H\times W\times 2\times K)$, 以及动态蛇形卷积核和坐标映射与双线性插值的复杂度$\mathrm{O}(H\times W\times K\times C_{\text{in}}\times C_{\text{out}})+\mathrm{O}(H\times W\times K^2)$. 综合上述计算, DSConv 模块的总计算复杂度为$\mathrm{O}(H\times W\times(K^2\times C_{\text{in}}\times C_{\text{out}}+K\times C_{\text{in}}\times C_{\text{out}}+K^2+C_{\text{out}}^2))$. 相比之下, 原卷积模块的计算复杂度为$\mathrm{O}(H\times W\times(K^2\times C_{\text{in}}\times C_{\text{out}}+2\times C_{\text{out}}))$. DSConv 模块的计算量增加较为明显, 但精度提升显著.

如图 2 所示, DSConv 能够根据输入图像的特征动态调整卷积核的形状和大小, 从而适应不同目标的几何形状. 这种灵活的感受野使得 DSConv 在处理复杂局部结构的目标时, 比传统卷积具有更高的适应性. 此外, DSConv 通过引入特定的约束, 确保对目标的连续聚焦, 避免因过大的变化偏移引起的感知范围过度扩散问题. 具体坐标计算如式 (1) 和式 (2) 所示.

图 2 中x轴方向的变化为:

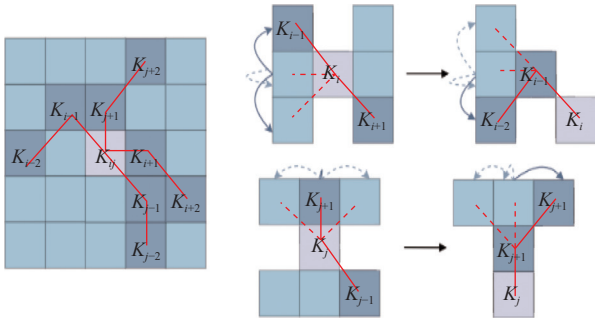

图 2 动态蛇形卷积核可选择的感受野和坐标计算

$$K_{i\pm c}=\begin{cases}(x_{i+c},\ y_{i+c})=\left(x_i+c,\ y_i+\sum_{i}^{i+c}\Delta y\right)\\(x_{i-c},\ y_{i-c})=\left(x_i-c,\ y_i+\sum_{i-c}^{i}\Delta y\right)\end{cases}\tag{1}$$

在y轴方向的变化为:

$$K_{j\pm c}=\begin{cases}(x_{j+c},\ y_{j+c})=\left(x_j+\sum_{j}^{j+c}\Delta x,\ y_j+c\right)\\(x_{j-c},\ y_{j-c})=\left(x_j+\sum_{j}^{j+c}\Delta x,\ y_i-c\right)\end{cases}\tag{2}$$

其中, c表示距离中心网格的水平距离. 卷积核K中每个网格位置$K_{i\pm c}$的选择是一个累积过程. 从中心位置K_i开始, 远离中心网格的位置取决于前一个网格的位置: K_{i+1}相对于K_i增加了偏移量$\Delta=\{\delta|\delta\in[-1,\ 1]\}$. 因此, 偏移量需要进行累加$\Sigma$, 以确保卷积核符合线性形态结构. 由于偏移量通常是分数, 然而坐标通常是整数形式, 因此采用双线性插值, 表示为:

$$V(x,y)=\sum_{i}^{n}\sum_{j}^{m}w_{ij}f(x_i,y_j)\tag{3}$$

其中, $V(x,y)$表示任意 (分数) 位置, (x_i,y_j)枚举特征图中的所有整数空间位置, 而w_{ij}是双线性插值核. w_{ij}是二维的, 它被分解为两个一维核, w_i和w_j为水平和垂直方向的权重, 如下所示:

$$w_{ij}=w_iw_j\tag{4}$$

本研究将 bottleneck 模块中的第 1 个 Conv 替换为 DSConv, 并将其命名为 Bottleneck_D. 然后, 将 C2f 模块中的所有 bottleneck 替换为 Bottleneck_D, 形成了 C2f_DySnakeConv (C2fDS) 模块, 最终替换用于语义信息传递的 C2f 模块 (如图 1 所示). 本模块更有效地利用了 DSConv 在捕捉局部特征方面的优势, 显著提升了目标检测任务的整体性能.

2.2 EMA 模块

本研究引入 EMA 模块, 来提高模型在有限计算资源下的检测精度和效率. 传统的注意力机制模块在处理复杂全局特征时, 往往面临计算成本高、通道信息丢失的问题, 难以有效捕捉细小而弯曲的裂缝. 例如, SE (squeeze-and-excitation network)[23]模块在处理包含细小横向裂缝的道路图像时, 可能由于全局池化操作而无法保留裂缝的精确位置信息, 导致检测结果不够准确. EMA 模块通过重塑和拆分通道维度, 实现特征的均匀分布和高效利用, 并通过双并行分支结构, 利用不同尺度的感受野, 增强了模型对复杂全局特征的捕捉能力. 这种设计不仅减少了通道信息的丢失, 降低了计算成本, 还显著提升了模型的检测性能. 具体结构如图 3 所示, EMA 模块被集成至 YOLOv8 每个检测头前. 该模块使得模型能够通过并行的卷积核更全面地利用中间特征图的上下文信息.

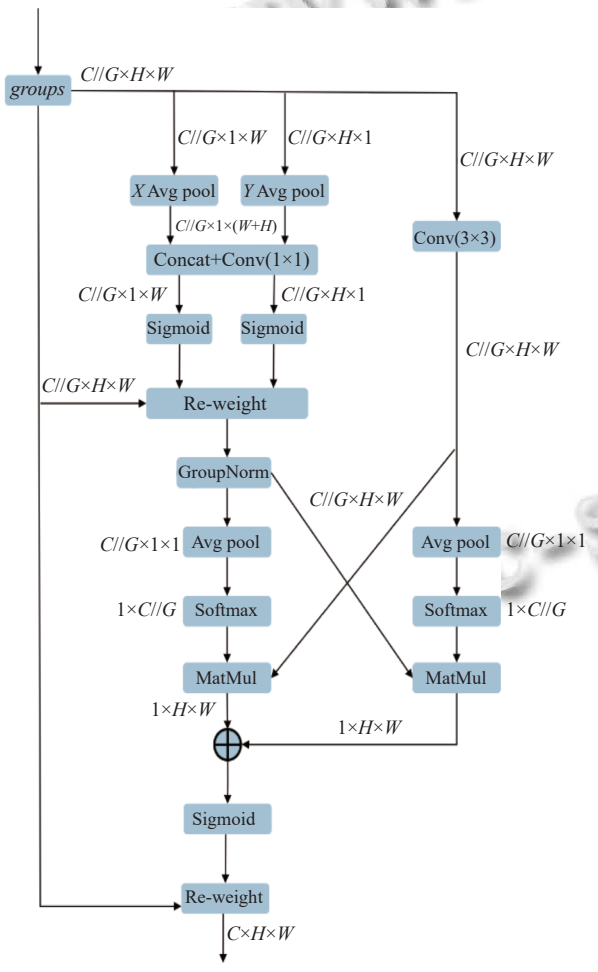

图 3 EMA 模块

EMA 模块的结构计算复杂度主要包括自适应平均池化、卷积、归一化和 Softmax 操作, 具体来说, 自适应平均池化层的计算复杂度为O($H\times W\times C/groups$), Concat 和 1×1 卷积计算复杂度都为O($(H+W)\times C/groups$), 3×3 卷积计算复杂度为O($H\times W\times 9\times C/groups$). Reweight 操作的计算复杂度为O($H\times W\times C/groups$). GroupNorm 计算复杂度为O($H\times W\times C/groups$). MatMul 计算复杂度为O($H\times W\times C/groups$). 此外, 矩阵乘法的计算复杂度为O($H\times W\times C/groups$). 最后的 Re-weight 计算复杂度为O($H\times W\times C$). 最终, 经过计算化简提取最大项, 其总计算复杂度为O($H\times W\times C\times(1+15/groups)$). 这表明 EMA 模块的计算复杂度依赖于输入特征图的高度和宽度、通道数以及分组因子, 且计算量很小, 适合部署提高性能.

通过在检测头前引入 EMA 模块, 模型能够在提炼道路损伤检测目标的复杂全局特征方面表现得更加出色, 显著提升了检测性能. 这种方法不仅提高了对各种类型道路损伤的检测精度, 同时也优化了计算效率, 使得在有限硬件资源上的部署成为可能.

2.3 小目标检测层

本研究通过在 YOLOv8 模型中增添小目标检测层, 提升了对小目标的检测精度, 并解决了下采样过程中局部细节特征及小目标信息丢失的问题. 此改进通过深层特征向浅层特征的有效传递及融合, 实现了更丰富的小目标轮廓及位置信息的保留, 显著提高了小目标的定位与识别准确性, 有效减少了小目标的漏检与误检率.

在道路损伤检测中, 小目标 (如细小裂缝和微小坑洼) 的检测是一大挑战. 这些小目标在不同尺度上表现出复杂的几何特征. P2 检测层能够更精准地捕捉这些细微特征, 提高了模型在复杂场景下的检测性能.

2.4 Flex-PIoUv2 函数

边界框回归在目标检测中占据着核心地位. 损失函数中的不合理惩罚因子会导致锚框在回归过程中进行不必要的膨胀. 例如, *IoU* (intersection over union)[24]在检测框和真实框不相交时, 无法反映两者间距且难以精确表征重合度. GIoU (generalized IoU)[25]通过引入最小外接框解决了在无重叠时损失为零的问题, 但在检测框与真实框出现包含现象时退化为 *IoU*, 并且在相交情况下收敛速度较慢. DIoU (distance IoU)[26]改进了 GIoU, 通过回归两个框中心点的欧氏距离来加速收敛, 但未考虑边界框的纵横比. CIoU (complete IoU) 在 DIoU 基础上增加了检测框尺度和长宽的损失, 但纵横

比描述的是相对值, 存在模糊性, 并且未解决难易样本的平衡问题.

针对这些挑战, 本研究提出了一种借鉴 Focaler-IoU[27]和 PIoUv2[28]方法的新型损失函数, 名为 Flex-PIoUv2, 旨在有效缓解由样本不均和难易样本对边界框回归的影响, 同时解决锚框在回归过程中的膨胀问题. Flex-PIoUv2 损失函数通过线性区间映射方法和适应性的惩罚因子降低边界框损失值, 显著提升了锚框的定位准确性和聚焦能力.

Focaler-IoU 函数使用了分段函数来重建 IoU 损失, 从而改善边界框回归. Focaler-IoU 的公式如下:

$$IoU^{\text{focaler}}=\begin{cases}0, & IoU<d\\ \dfrac{IoU-d}{u-d}, & d\leqslant IoU\leqslant u\\ 1, & IoU>u\end{cases} \tag{5}$$

其中, IoU 是原始 IoU 值, $[d,u]\in[0,1]$. 通过调整d和u的值, 我们可以使 IoU 专注于不同的回归样本. 其损失定义如下:

$$\mathcal{L}_{IoU^{\text{focaler}}}=1-IoU^{\text{focaler}} \tag{6}$$

在 Focaler-IoU 中, 设定一个阈值d, 使得 IoU 低于d的样本不被考虑. 这虽然可以减少低质量样本对损失的影响, 但也可能忽略了一些潜在有用的信息, 导致样本利用不充分. 道路损伤检测目标大部分都是中低质量样本, 显然不符合需求. 因此, 本研究提出了 Flex-IoU, 使用线性区间映射重新构建 IoU 损失, 其损失定义如下:

$$IoU^{\text{flex}}=\begin{cases}\dfrac{IoU}{u}, & IoU\leqslant u\\ 1, & IoU>u\end{cases} \tag{7}$$

$$\mathcal{L}_{IoU^{\text{flex}}}=1-IoU^{\text{flex}} \tag{8}$$

去掉阈值d后, Flex-IoU 函数所有样本都参与损失计算, 确保样本利用最大化. 这种方式更全面地反映了样本的分布情况, 有助于模型在不同质量的样本上均衡学习, 产生整体较高的回归效果. 通过线性区间映射, Flex-IoU 在低 IoU 范围内更为平滑, 避免了因 IoU 值较低而导致的损失曲线不连续问题, 从而促进模型更稳定地收敛.

为了解决锚框膨胀等问题, 引入了一种具有尺寸适应性的惩罚因子, 引导锚框直接有效地回归. 将这个惩罚因子与一个根据锚框质量调整梯度的函数结合起来, 称为 Powerful-IoU 损失. Powerful-IoU 损失直接最小化锚框的 4 个边与目标框的相应边之间的距离来优化边界框回归. 此损失定义如下:

$$p=\left(\frac{dw_1}{w_{\text{gt}}}+\frac{dw_2}{w_{\text{gt}}}+\frac{dh_1}{h_{\text{gt}}}+\frac{dh_2}{h_{\text{gt}}}\right)\Big/4 \tag{9}$$

$$\mathcal{L}_{\text{Powerful-IoU}}=\mathcal{L}_{IoU}+1-\mathrm{e}^{-p^2} \tag{10}$$

其中, p 综合考虑了锚框与目标框的定位误差和尺寸误差, 每个变量定义如图 4 所示.

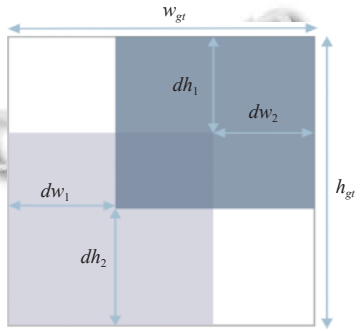

图 4 p 每个变量定义

在此基础上, 为了增强聚焦中等质量锚框的能力, 结合由单个超参数控制的非单调注意力层$m(x)$. PIoUv2 的损失定义如下:

$$q=\mathrm{e}^{-p},\ q\in(0,1] \tag{11}$$

$$m(x)=3x\cdot\mathrm{e}^{-x^2} \tag{12}$$

$$\mathcal{L}_{\text{PIoUv2}}=3\cdot m(\lambda q)\cdot\mathcal{L}_{\text{Powerful-IoU}} \tag{13}$$

其中, q用于衡量锚框的质量, 当$q=1$时, $p=0$, 表示锚框和目标框之间完全对齐. 随着p的增加, q逐渐减小, 表示锚框质量较低. λ是控制注意力函数行为的超参数, $m(\lambda q)$增强了 PIoU 聚焦于中等质量锚框的能力. PIoUv2 在每个锚框的回归过程中, 更加关注中等质量的锚框.

结合 Flex-IoU 和 PIoUv2 损失函数, 本研究提出了 Flex-PIoUv2, 其损失定义如下:

$$\mathcal{L}_{\text{Flex-PIoUv2}}=3\cdot m(\lambda q)\cdot\left(\mathcal{L}_{IoU^{\text{flex}}}+1-\mathrm{e}^{-p^2}\right) \tag{14}$$

本研究提出的 Flex-PIoUv2 策略通过线性区间映射方法, 在不同的 IoU 范围内进行调整, 从而增强对不同质量锚框的区分能力. 与 PIoUv2 相比, Flex-PIoUv2 在处理低质量锚框时表现出更好的自适应性. 通过调整参数值, 损失函数能够专注于特定范围内的样本, 从而更好地处理样本不均衡问题. 此外, Flex-PIoUv2 结合的非线性函数, 使损失函数在优化过程中具有平滑

的变化特性, 避免了由于不连续性引起的优化振荡问题, 促进模型更稳定地收敛. 这些改进使得 Flex-PIoUv2 在处理各种质量的锚框时表现出更优异的性能, 相较于 PIoUv2 具有明显的优势.

3 实验结果

3.1 数据集

实验数据集选自公共道路损伤数据集 RDD2022[29], 该数据集包含来自 6 个国家, 捕获了纵向裂缝、横向裂缝、鳄鱼裂缝和坑洼 4 种类型的道路损伤. 由于其中有部分图片没有符合要求的标注, 所以进行了数据清理. 样本数据集按照 7:3 比例随机划分. 最终有 23 767 张照片, 其中, 训练集 16 634 张图像, 验证集 7 133 张图像. 纵向裂缝 (D00) 13 548 个, 横向裂缝 (D10) 7 709 个, 鳄鱼裂缝 (D20) 8 412 个, 坑洼 (D40) 3 674 个.

GRDDC2020[30]数据集包含从印度、日本和捷克共和国收集的道路图像, China road damage 数据集仅包含国内收集的道路图像. 为与训练数据集标签保持一致, 本文进行了数据清理, 并按照 7:3 比例随机划分. 最终, GRDDC2020 验证集 2 381 张, China road damage 验证集 1 984 张. 这些数据集均已上传至飞桨 AI Studio 平台.

3.2 实验参数和评价指标

本实验采用的硬件环境为: RTX A5000 (24 GB) 软件环境为: CUDA 11.1, 深度学习框架为 PyTorch 1.9.0. 模型训练周期 (epoch) 为 300, 批量大小 (batchsize) 为 32, 优化器 (optimizer) 自动选择, 初始学习率为 0.01. 在 Flex-PIoUv2 函数中超参数$u = 0.95$, $\lambda = 1.5$. 为了客观评价实验结果, 本文选取了平均精度均值 (*mAP*) 和 *F*1 分数来衡量改进模型的优缺点. 计算公式如式 (16)–式 (19) 所示:

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = 2 \times \frac{TP \times FP}{TP + FP} \tag{17}$$

$$mAP = \frac{1}{c}\sum_{i=1}^{c}\int_{0}^{1} Precision(Recall)\mathrm{d}(Recall) \tag{18}$$

其中, *TP* 表示检测结果中正确目标的个数, *FP* 表示检测结果中错误目标的个数, *FN* 表示正确目标中缺失目标的个数.

3.3 DSConv 相关实验

本实验探讨了将 bottleneck 模块中不同感受野下的普通 Conv 替换为 DSConv, 以及将 neck 层中用于不同作用的 C2f 模块替换为 C2fDS 对模型性能的影响.

实验结果表明 (表 1), 当 C2fDS 模块替换用于语义信息传递的 C2f 模块 (YOLOv8n+C2fDS) 时, 模型性能显著提升, 主要是因为 DSConv 能有效捕捉细小且弯曲的局部特征. 然而, 将 bottleneck 模块中的第 2 个 Conv 替换为 DSConv, 组成新的 bottleneck 块, 并用这些 bottleneck 块替换 C2f 中的所有 bottleneck 块, 形成新的 C2f 模块用于语义信息传递 (YOLOv8n+C2fDS2), 效果并不理想. 这是因为第 2 个卷积层通常具有更大的感受野, 它需要处理来自第 1 层的综合和抽象特征. DSConv 在小感受野下 (第 1 层卷积) 更有效, 而在较大感受野下 (第 2 层卷积) 效果较差, 限制了其优势的发挥. 此外, 将 C2fDS 替换用于位置信息传递的 C2f 模块 (YOLOv8n+C2fDS (loca)) 也未能取得理想效果, 主要原因是 DSConv 的感受野设计不适合跨层级的位置信息传递.

表 1 DSConv 相关实验 (%)

Model	*F*1	*mAP*50					*mAP*50-95				
		D00	D10	D20	D40	All	D00	D10	D20	D40	All
YOLOv8n	56.8	57.9	54.9	66.6	45.3	56.2	31.9	27.7	35.2	19.7	28.6
YOLOv8n+C2fDS	57.2	57.9	56.0	66.7	46.0	56.6	32.2	28.0	35.3	20.1	28.9
YOLOv8n+C2fDS2	56.5	57.6	54.5	66.6	45.1	56.0	31.7	27.4	35.2	19.5	28.4
YOLOv8n+C2fDS (loca)	56.6	57.1	55.0	66.7	45.1	56.0	31.4	27.7	35.5	19.6	28.5

3.4 Flex-PIoUv2 相关实验

在本研究中, 通过综合分析训练与验证阶段的边界框回归损失、*mAP*50 以及 *mAP*50-95 等指标的动态变化, 探究了 Flex-PIoUv2 损失函数在目标检测任务中的性能影响. 模型包括基准 YOLOv8n、结合 C2fDS、P2 和 EMA 模块的 Improved 模型、进一步集成 PIoUv2 的 Improved 模型, 以及融入 Flex-PIoUv2 损失函数的最终模型.

实验结果表明, 采用 Flex-PIoUv2 损失函数的模型在边界框回归损失上相比于 CIoU 和 PIoUv2 损失函

数的模型表现出显著的减少, 如图 5 和图 6 所示. Flex-PIoUv2 在优化锚框回归过程中具有显著优势, 能够更有效地减少预测框与真实框之间的偏差. 图 7 和图 8 所示的 *mAP*50 和 *mAP*50-95 指标的提高, 进一步验证了 Flex-PIoUv2 损失函数在提高模型检测精度方面的有效性. 尤其是在较高 *IoU* 阈值的条件下, Flex-PIoUv2 仍维持着较高的检测性能, 这对于对定位精度要求高的目标检测任务极为重要.

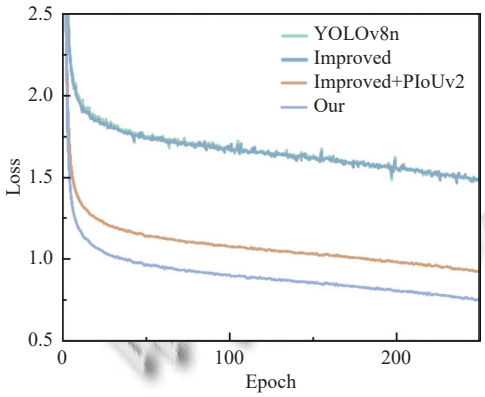

图 5 训练集边界框损失

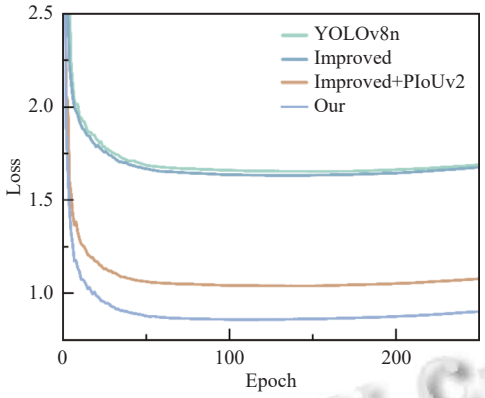

图 6 验证集边界框损失

综上所述, 通过引入 Flex-PIoUv2 损失函数, 可以显著提升目标检测模型的整体性能, 特别是在边界框回归和高 *IoU* 阈值检测精度方面表现出色.

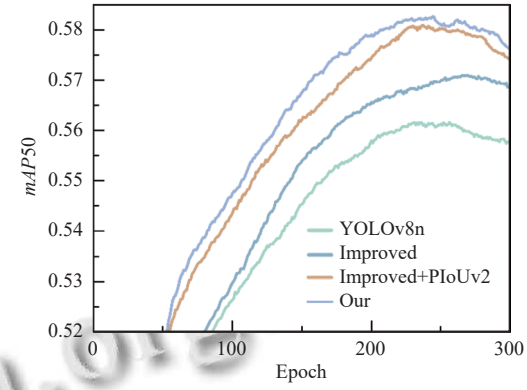

图 7 *mAP*50 值

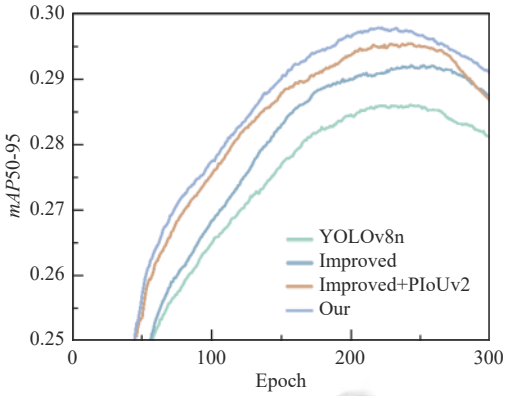

图 8 *mAP*50-95 值

3.5 消融实验

本研究通过一系列消融实验, 验证了改进模块 C2fDS、P2、EMA 以及 Flex-PIoUv2 损失函数在 YOLOv8n 模型中的有效性. 实验结果如表 2 所示, 综合应用这些模块的改进模型相比单独添加模块的变体在性能上表现更优, 尤其是在 *mAP*50 和 *mAP*50-95 指标上.

表 2 消融实验 (%)

Model	*F*1	*mAP*50					*mAP*50-95				
		D00	D10	D20	D40	All	D00	D10	D20	D40	All
YOLOv8n	56.8	57.9	54.9	66.6	45.3	56.2	31.9	27.7	35.2	19.7	28.6
YOLOv8n+C2fDS	57.2	57.9	56.0	66.7	46.0	56.6	32.2	28.0	35.3	20.1	28.9
YOLOv8n+P2	57.1	57.9	55.9	67.2	45.8	56.7	32.2	28.0	35.4	20.3	29.0
YOLOv8n+EMA	57.1	57.5	56.3	66.7	46.2	56.7	31.8	28.2	35.6	20.2	28.9
YOLOv8n+C2fDS+P2	57.1	58.4	55.1	66.5	45.9	56.4	32.6	27.4	35.6	20.1	28.9
YOLOv8n+C2fDS+EMA	57.0	57.6	55.6	66.6	46.3	56.5	32.1	27.6	35.6	20.3	28.9
YOLOv8n+P2+EMA	56.8	58.1	56.1	67.0	45.6	56.7	32.5	28.3	35.7	20.2	29.2
YOLOv8n+C2fDS+P2+EMA (Improved)	57.2	59.0	55.8	66.9	46.4	57.0	32.7	28.0	35.2	20.7	29.2
Improved+PIoUv2	58.1	59.3	56.1	67.5	49.2	58.0	32.8	28.3	35.4	21.6	29.5
Improved+Flex-PIoUv2 (Ours)	58.3 (+1.5)	59.8 (+1.9)	56.2 (+1.3)	67.8 (+1.1)	49.2 (+3.9)	58.3 (+2.1)	33.1 (+1.2)	28.0 (+0.3)	36.2 (+1.0)	21.8 (+2.1)	29.8 (+1.2)

具体而言, 替换为 Flex-PIoUv2 损失函数的 Improved 模型相比 YOLOv8n、Improved 以及 Improved+PIoUv2 展现出更显著的提升. 特别是, Flex-PIoUv2 损失函数对于难样本 D00 (纵向裂缝) 和数量较少的样本 D40 (坑洼) 的检测有明显的提升, 相较于基准模型 YOLOv8n, *mAP*50 分别提升了 1.9 和 3.9 百分点, *mAP*50-95 分别提升了 1.2 和 2.1 百分点. 损失函数方面, 同时使用 Improved 模型, Flex-PIoUv2 相较于 CIoU、PIoUv2, *mAP*50 分别高出 1.3 和 0.3 百分点, *mAP*50-95 分别高出 0.6 和 0.3 百分点. 这验证了 Flex-PIoUv2 损失函数在平衡样本分布和提升困难样本检测准确度方面的有效性.

综上所述, 通过引入 Flex-PIoUv2 损失函数和 C2fDS、P2、EMA 模块后的 YOLOv8n 模型在各种检测任务中均表现出更优的性能, 特别是在处理困难样本和少量样本时, 表现尤为出色.

3.6 与不同算法的实验

为了客观评估本研究所提出的道路损伤检测模型的性能, 采用了多个性能评估指标, 并将其与当前主流的目标检测算法进行了比较. 通过实验结果 (见表 3) , 与 YOLOv8n 相比, YOLOv8s 的参数量和计算量分别增加了 270% (3.0×10^6–11.1×10^6) 和 242% (8.3 GFLOPs–28.4 GFLOPs), 但 *F*1 分数仅提升了 3.0% (56.8%–58.5%), *mAP*50 提升了 4.8% (56.2%–58.9%), *mAP*50-95 提升了 5.6% (28.6%–30.2%). 而本文模型在参数量保持不变, 计算量仅增加了 54.2% (8.3 GFLOPs–12.8 GFLOPs) 的情况下, *F*1 分数提升了 2.6% (56.8%–58.3%), *mAP*50 提升了 3.7% (56.2%–58.3%), *mAP*50-95 提升了 4.2% (28.6%–29.8%). 这一结果表明, 本文模型在保持参数量不变的情况下, 以相对较小的计算量增加, 达到了接近 YOLOv8s 的性能提升, 从而验证了本文模型在效率和精度上的优越性, 并达到了良好的平衡.

表 3 与主流模型对比

Model	*F*1 (%)	*mAP*50 (%)	*mAP*50-95 (%)	Params (10^6)	GFLOPs
YOLOv5s	56.9	56.4	28.8	7.1	16.3
YOLOv5m	58.3	58.5	30.1	21.1	50.6
YOLOv7tiny	55.7	55.3	25.6	6.0	13.5
YOLOv8n	56.8	56.2	28.6	3.0	8.3
YOLOv8s	58.5	58.9	30.2	11.1	28.4
Ours	58.3	58.3	29.8	3.0	12.8

尽管本模型在性能上未能超越最先进的 YOLOv8s, 但其仍具有明确的意义和重要性. 具体而言, 本模型具备以下几个优势: 首先, 本文模型在保持较低计算资源增长的前提下, 实现了接近 YOLOv8s 的性能提升, 展示了更高的效率和性价比, 尤其适用于计算资源有限的应用场景. 其次, 本模型提供了一种根据道路损伤检测样本质量的具体情况来调整的 Flex-PIoUv2 函数, 为该领域的方法多样性做出了贡献. 最后, 本模型提供了对模型复杂性与性能之间权衡的洞察, 这对于实际部署至关重要.

3.7 通用性对比实验

为了验证本文提出的模型和损失函数的泛化能力和精度, 使用了 GRDDC2020 和 China road damage 在数据集进行实验对比. 实验结果如表 4 和表 5 所示.

表 4 GRDDC2020 数据集对比 (%)

Model	*F*1	*mAP*50	*mAP*50-95
YOLOv8n (CIoU)	65.0	68.7	36.6
Improved (CIoU)	66.9	71.1	37.9
Improved+PIoUv2	67.1	71.3	38.7
Improved+Flex-PIoUv2	**69.6**	**73.6**	**39.6**

表 5 China road damage 数据集对比 (%)

Model	*F*1	*mAP*50	*mAP*50-95
YOLOv8n (CIoU)	86.4	91.4	61.4
Improved (CIoU)	86.9	91.6	**62.6**
Improved+PIoUv2	87.0	92.2	61.5
Improved+Flex-PIoUv2	**88.2**	**92.7**	62.5

在 GRDDC2020 数据集上, 本模型相较于 YOLOv8n 模型在 *F*1 分数上提升了 4.6 百分点, *mAP*50 上提升了 4.9 百分点, 在 *mAP*50-95 上提升了 3 百分点. 损失函数方面, 同时使用 Improved 模型, Flex-PIoUv2 在 *F*1 分数上相较于 CIoU 和 PIoUv2 分别高出 2.7、2.5 百分点, *mAP*50 上高出 2.5、2.3 百分点, *mAP*50-95 上高出 1.7、0.9 百分点.

在 China road damage 数据集上, 本文的模型和损失函数也表现优异. 验证了其优秀性能和泛化能力.

3.8 可视化结果分析

在实际应用中, 道路损伤检测模型需要具备在各种天气条件、光照条件及复杂背景下的鲁棒性. 为了验证模型在这些条件下的性能, 本节对比分析了 YOLOv8n 模型和本文模型在高光照多目标、极小目标、雪地、阴天等情况的检测效果. 可视化结果如图 9 所示, 图 9(a) 为 YOLOv8n 模型的检测结果, 图 9(b) 为本文模型的检测结果.

在高光照多目标条件下, YOLOv8n 模型会出现漏

检问题, 在极小目标条件下, YOLOv8n 模型会检测不到小目标, 而本文模型能够精确识别. 在雪地、阴天等恶劣条件下, YOLOv8n 模型均未识别到目标, 虽然本文模型检测置信度不高, 但均已检测出目标. 这表明其在复杂背景下的适应能力较强. 在复杂背景条件下, 包括高光照、积水、积雪等多种复杂因素共同作用, 容易出现误检和漏检现象. 虽然本文模型会出现扩大锚框且置信度较低的情况, 但能够较好地排除背景干扰, 保持检测稳定性, 减少误检和漏检情况. 这表明本文模型在处理复杂背景时具备更高的稳定性和鲁棒性.

图 9 可视化结果

4 结论与展望

本研究提出了一种改进 YOLOv8 的道路损伤检测算法. DSConv 针对局部结构进行优化学习, EMA 模块则增强全局特征的表达. 两者关系互补, 既能捕捉细节, 又能理解整体结构. 进一步增加小目标检测层, 提高小目标检测精度. 最终, 本文提出的 Flex-PIoUv2 有效解决样本分布不均和难易样本等问题.

虽然本文模型在检测精度上相比原模型有了较大提升, 但在实际应用的可行性方面仍存在改进空间. 未来的改进方向包括: 在不增加计算负担的情况下优化 neck 层结构, 以更有效地融合语义和位置信息; 通过模型剪枝等轻量化技术减少参数量和计算需求, 以提升在资源受限环境中的推理效率; 以及进一步验证和优化 Flex-PIoUv2 损失函数在不同目标检测模型中的性能和通用性.

参考文献

1 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2013. 580–587. [doi: 10.1109/CVPR.2014.81]

2 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

3 Redmon J, Divvala SK, Girshick RB, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2015. 779–788.

4 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2015. 21–37.

5 Zhang Z, Xu Y, Yang J, *et al*. A survey of sparse representation: Algorithms and applications. IEEE Access, 2015, 3: 490–530. [doi: 10.1109/ACCESS.2015.2430359]

6 Tarabalka Y, Chanussot J, Benediktsson JA. Segmentation and classification of hyperspectral images using watershed transformation. Pattern Recognition, 2010, 43(7): 2367–2379. [doi: 10.1016/j.patcog.2010.01.016]

7 Zhang L, Yang F, Zhang YD, *et al*. Road crack detection using deep convolutional neural network. Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP). Phoenix: IEEE, 2016. 3708–3712.

8 Pereira V, Tamura S, Hayamizu S, *et al*. A deep learning-based approach for road pothole detection in timor leste. Proceedings of the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). IEEE, 2018. 279–284.

9 Feng H, Xu GS, Guo YH. Multi-scale classification network for road crack detection. IET Intelligent Transport Systems, 2019, 13(2): 398–405. [doi: 10.1049/iet-its.2018.5280]
10 Nie MX, Wang K. Pavement distress detection based on transfer learning. Proceedings of the 5th International Conference on Systems and Informatics (ICSAI). Nanjing: IEEE, 2018. 435–439.
11 Du YC, Pan N, Xu ZH, *et al*. Pavement distress detection and classification based on YOLO network. International Journal of Pavement Engineering, 2021, 22(13): 1659–1672. [doi: 10.1080/10298436.2020.1714047]
12 Katsaliros A, Panagos II, Sfikas G, *et al*. Road crack detection using quaternion neural networks. Proceedings of the 14th IEEE Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). Nafplio: IEEE, 2022. 1–5.
13 Parcollet T, Morchid M, Linarès G. A survey of quaternion neural networks. Artificial Intelligence Review, 2020, 53(4): 2957–2982. [doi: 10.1007/s10462-019-09752-1]
14 李松, 史涛, 井方科. 改进 YOLOv8 的道路损伤检测算法. 计算机工程与应用, 2023, 59(23): 165–174.
15 魏陈浩, 杨睿, 刘振丙, 等. 具有双层路由注意力的 YOLOv8 道路场景目标检测方法. 图学学报, 2023, 44(6): 1104–1111.
16 Dai JF, Qi HZ, Xiong YW, *et al*. Deformable convolutional networks. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 764–773.
17 武兵, 田莹. 基于注意力机制的多尺度道路损伤检测算法研究. 图学学报, 2024, 45(4): 770–778.
18 胥铁峰, 黄河, 张红民, 等. 基于改进 YOLOv8 的轻量化道路病害检测方法. 计算机工程与应用, 2024, 60(14): 175–186.
19 王海群, 王炳楠, 葛超. 重参数化 YOLOv8 路面病害检测算法. 计算机工程与应用, 2024, 60(5): 191–199.
20 Zeng JY, Zhong H. YOLOv8-PD: An improved road damage detection algorithm based on YOLOv8n model. Scientific Reports, 2024, 14(1): 12052. [doi: 10.1038/s41598-024-62933-z]
21 Qi YL, He YT, Qi XM, *et al*. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. 6047–6056. [doi: 10.1109/ICCV51070.2023.00558]
22 Ouyang DL, He S, Zhang GZ, *et al*. Efficient multi-scale attention module with cross-spatial learning. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes: IEEE, 2023. 1–5. [doi: 10.1109/ICASSP49357.2023.10096516]
23 Hu J, Shen L, Sun G, *et al*. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2017. 7132–7141.
24 Yu J, Jiang YN, Wang ZY, *et al*. UnitBox: An advanced object detection network. Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam: Association for Computing Machinery, 2016. 516–520.
25 Rezatofighi SH, Tsoi N, Gwak J, *et al*. Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 658–666.
26 Zheng Z, Wang P, Liu W, *et al*. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2019. 12993–13000.
27 Zhang H, Zhang SJ. Focaler-IoU: More focused intersection over union loss. arXiv:2401.10525, 2024.
28 Liu C, Wang KG, Li Q, *et al*. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. Neural Networks, 2024, 170: 276–284. [doi: 10.1016/j.neunet.2023.11.041]
29 Arya D, Maeda H, Ghosh SK, *et al*. RDD2022: A multi-national image dataset for automatic road damage detection. arXiv:2209.08538, 2022.
30 Arya D, Maeda H, Ghosh SK, *et al*. RDD2020: An annotated image dataset for automatic road damage detection using deep learning. Data in Brief, 2021, 36: 107133. [doi: 10.1016/j.dib.2021.107133]

(校对责编: 孙君艳)