

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

无人机视角下的道路损伤检测算法MAS-YOLOv8n

王晓燕, 王禧钰, 李杰, 梁文辉, 牟建宏, 毕楚然

引用本文:

王晓燕, 王禧钰, 李杰, 等. 无人机视角下的道路损伤检测算法MAS-YOLOv8n[J]. 光电工程, 2024, 51(10): 240170.

Wang X Y, Wang X Y, Li J, et al. MAS-YOLOv8n road damage detection algorithm from the perspective of drones[J]. *Opto-Electron Eng*, 2024, 51(10): 240170.

<https://doi.org/10.12086/oe.2024.240170>

收稿日期: 2024-07-18; 修改日期: 2024-09-09; 录用日期: 2024-09-10

相关论文

自适应前景聚焦无人机航拍图像目标检测

肖振久, 吴正伟, 张杰浩, 曲海成

光电工程 2024, 51(9): 240149 doi: 10.12086/oe.2024.240149

改进YOLOv7的无人机视角下复杂环境目标检测算法

张润梅, 肖钰霏, 贾振楠, 陈中, 陈梓华, 袁彬, 曹炜威, 宋妮妮

光电工程 2024, 51(5): 240051 doi: 10.12086/oe.2024.240051

协同随机森林方法和无人机LiDAR空谱数据的盐沼植被“精灵圈”识别

韩江涛, 谭凯, 张卫国, 周若彤, 刘帅

光电工程 2024, 51(3): 230188 doi: 10.12086/oe.2024.230188

基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, 49(3): 210372 doi: 10.12086/oe.2022.210372

更多相关论文见光电期刊集群网站 

 | 光电工程
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE_Journal



Website

problem of the residual structure in the YOLOv8n model being prone to interference, resulting in information loss, a multi-branch mixed attention (MBMA) mechanism is introduced. This MBMA structure is integrated into the C2f structure, strengthening the feature representation capabilities. It not only captures richer feature information but also reduces the impact of noise on the detection results. Secondly, to address the issue of poor detection performance resulting from significant variations in road damage morphologies, the TaskAlignedAssigner label assignment algorithm used in the YOLOv8n model is improved by utilizing ShapIoU (shape-intersection over union), making it more suitable for targets with diverse shapes and further enhancing detection accuracy. Experimental evaluations of the MAS-YOLOv8n model on the China-Drone dataset of road damages captured by drones reveal that compared to the baseline YOLOv8n model, our model achieves a 3.1% increase in mean average precision (mAP) without incurring additional computational costs. To further validate the model's generalizability, tests on the RDD2022_Chinese and RDD2022_Japanese datasets also demonstrate improved accuracy. Compared to YOLOv5n, YOLOv8n, YOLOv10n, GOLD-YOLO, Faster-RCNN, TOOD, RTMDet-Tiny, and RT-DETR, our model exhibits superior detection accuracy and performance, showcasing its robust generalization capabilities.

Keywords: damage detection; YOLOv8n; attention mechanism; label allocation algorithm

1 引言

道路作为交通运输的基础设施,其安全性直接关系到车辆和行人的通行安全。道路损伤如裂缝、坑洞、路面破损等,不仅会影响行车的平稳性和舒适性,还可能引发交通事故。随着道路损伤检测设备的发展,无人机航拍图像采集的方法也逐渐被应用到道路损伤检测领域中,无人机在空中进行巡检,无需直接接触道路表面,避免了工作人员进入危险区域的风险,也大大提高了图像采集的效率^[1-2]。然而,由于无人机航拍视角下的道路损伤图像背景复杂、目标特征不够明显,并且道路损伤形态差异大、拓扑结构复杂,导致检测效果较差。针对这一问题,许多学者提出了多种道路损伤检测方法^[3-4],其中基于深度学习的道路损伤检测是目前主要的研究方向。

在基于深度学习的道路损伤检测领域,主要方法为单阶段目标检测算法和两阶段目标检测算法^[5-8]。单阶段目标检测算法如 YOLO 系列^[9-12]是基于回归的分类算法,避免了生成候选区域的步骤,相对两阶段目标检测算法,兼顾了准确率与实时性。为进一步提高检测效果,许多研究人员在 YOLOv8 模型的基础上进行改进。Jiang 等人^[13]将 Transformer 结构和 EVC 块(显性视觉中心块)进行整合,实现了特征增强,但对小目标检测效果不够好,且提高了模型复杂度。程期浩等人^[14]基于 YOLOv8 融合了非跨步卷积降低小目标特征的细粒度信息的丢失,设计了 C2f-LSK 模块,通过采用选择机制对空间特征进行有效加权,动

态地调整感受野,并加入 P2 小目标检测头以减少小目标特征丢失,但也由于 P2 检测头的加入,提升了计算复杂度。Li 等人^[15]利用 Bi-PAN-FPN 改进 YOLOv8s 中的颈部部分,并在骨干中采用 GhostblockV2 结构,取代了 C2f 模块的一部分,抑制了长距离特征传输过程中的信息丢失,显著减少了模型参数量,然而在多尺度与背景复杂情况下对小目标检测效果不够好。龙伍丹等人^[16]利用量化感知重参数化模块(QARepVGG)和加强注意力模块(AM-CBAM),加强浅层特征提取同时抑制了复杂背景干扰,提高表征能力,但没有考虑目标形态对检测效果的影响。孟鹏帅等人^[17]提出了基于可变形卷积的 DC-ELAN 与 DC-MP 模块,适应不同形状大小输入特征,提升网络对复杂背景下特征的解析能力,引入边框损失函数 EIou,解决 CIou 形状惩罚项失效问题,但也提高了模型复杂度,增加了计算代价。通过以上研究可以得出,目前无人机视角下的道路损伤检测研究仍存在复杂背景干扰、对形态变化不敏感、小目标特征丢失、计算代价提升等问题。

针对以上问题,本文提出了一种新的道路损伤检测模型 MAS-YOLOv8n。首先,针对复杂背景干扰和细节信息易丢失的问题,设计了一种多分支混合注意力机制,并将其引入 YOLOv8n 模型的 C2f 结构中,在捕捉到更为丰富的特征信息、加强特征的表达能力的同时,降低了噪声干扰,提高了模型的检测效果。此外,针对道路损伤尺度差异大、形态多变的特点,

利用 ShapeIoU 对 YOLOv8n 模型使用的 TaskAligned-Assigner 标签分配算法进行改进, 通过引入尺度因子和形状相关的权重, 适应不同形状和尺度的目标, 降低边界框本身的形状和尺度等固有属性对边界框回归的影响, 进一步提高检测精度。

2 模型改进

本文模型在 YOLOv8n 的基础上进行改进, YOLOv8n 模型由三个主要部分组成: Backbone、Neck 和 Head。Backbone 是用于提取图像特征的主干网络, Neck 部分是用于进一步整合和增强特征表示的中间层, Head 是目标检测层, 用于预测边界框和类别置信度, 结构如图 1 所示。

2.1 多分支混合注意力机制

注意力机制 (Attention mechanism) 在处理大量信息时, 能够聚焦于关键部分, 筛选重要特征信息。对于无人机视角下的复杂环境的目标检测, 在网络中融合注意力机制可以提高网络中被测物体的权重, 有选择性地增强与任务相关的特征, 同时抑制不相关的背景信息, 从而提高了模型的准确性和效率, 使模型的学习更具有目的性。目前深度学习领域常用的注意力机制有 SE^[18]、CBAM^[19]、CA^[20] 等, SE 结构简单有效, 但只关注到了通道维度, 无法捕捉空间上的注意力。CBAM 对输入的特征层, 分别进行通道注意力机制的处理和空间注意力机制的处理, 相比于 SENet 只关注通道的注意力机制可以取得更好的效果。前两种注意力机制在求取通道注意力的时候, 通道的处理一般是采用全局最大池化或平均池化, 容易损失物体的空间信息, 而 CA 注意力机制将位置信息嵌入到了通道注意力中, 能够较为准确地捕捉不同通道之间的关系。

为进一步增强特征提取和表达能力, 加强特征交互能力, 本文设计了一种多分支混合注意力机制, 命名为 MBMA 模块, 结构如图 2 所示。MBMA 模块是一种综合考虑了空间和通道之间关系的混合注意力机制, 通过多个分支对输入特征进行精细处理。其中一个分支首先对输入特征的高度 (H) 和宽度 (W) 分别进行自适应平均池化, 随后将池化后的特征进行拼接并通过卷积操作, 这一过程有效地关注并交换了 H 和 W 的空间信息, 使得特征在空间维度上得到更为充分的交互。通过多次的卷积和切分操作, 该分支能够使关注的特征更为精细、准确, 从而实现对特征的增强。另一个分支则利用了全局平均池化操作, 实现了对通道信息的加强, 这一操作有效地压缩了每个通道的空间信息, 强调了通道间的依赖关系。随后, 对两个分支所得的特征融合, 避免了特征信息的孤立, 并通过多头注意力模块进行进一步处理。

多头注意力机制允许模型在不同的表示子空间以及不同位置同时关注信息, 这通过并行地运行多个注意力函数来实现, 每个函数都被称为一个“头” (head), 利用并行处理机制能够捕获到输入序列中不同位置的不同表示方面的信息。对于输入查询 (Q)、键 (K) 和值 (V), 首先通过三个不同的线性变换, 将它们投影到 d_q, d_k, d_v 维空间, 对每个头使用缩放点积注意力机制计算输出, 计算公式如式 (1-2) 所示, 其中, X 输入序列的表示, W^Q, W^K 和 W^V 是可学习的权重矩阵。最后将所有头的输出被拼接起来, 输出结果如式 (3-4) 所示, 其中 Q_i, K_i, V_i 表示第 i 个头对应的查询、键和值, W^O 是输出的线性变换矩阵。多头注意力机制允许模块在不同的注意力头中关注输入数据的不同方面, 计算输入序列中任意两个位置之间的相关性, 从而更好地处理长距离依赖问题, 加强对特征上下文信息的提取, 捕获更丰富的特征表示。

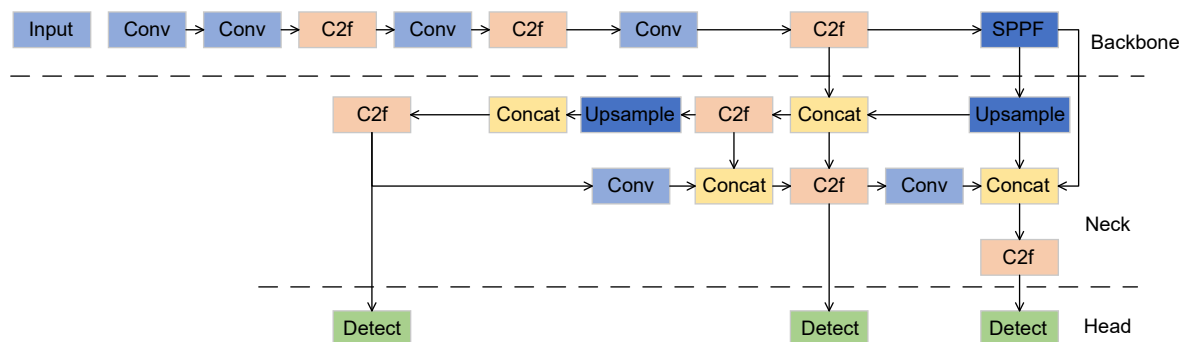


图 1 YOLOv8 模型结构
Fig. 1 YOLOv8 model structure

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \quad (1)$$

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V, \quad (2)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_n)W^O, \quad (3)$$

$$head_i = Attention(Q_i, K_i, V_i). \quad (4)$$

相比于其他注意力机制仅利用水平池化和垂直池化, 通过简单的通道拼接来获取空间注意力信息, MBMA 引入了一个额外的全局平均池化 (gap), 用于捕获全局特征, 并结合水平和垂直的池化信息进行更丰富的特征融合。还利用多个路径 (水平、垂直、全局) 来增强特征提取的多样性和鲁棒性, 并通过在池化后的特征上进行进一步的卷积操作和权重计算, 使得生成的权重更加细致, 能够更好地反映空间维度的复杂性。此外, 利用多头注意力机制允许模型通过多个独立的注意力头来关注输入数据的不同部分, 使得模型能够在不同的空间位置、通道或特征维度上捕获更加多样化的信息, 也能够减少单一注意力可能带来的偏差和噪声影响。相比于仅依赖于局部感受野的卷积操作, 多头注意力机制增强了特征的全局性, 也使模型能够更好地理解和利用输入特征之间的上下文关系。综上所述, MBMA 这种综合空间和通道注意力的设计, 增强了模块对特征的敏感性和表达能力,

提升了模块的鲁棒性, 使其能够更有效地处理复杂的输入数据, 提取关键特征。

2.2 C2f 结构改进

YOLOv8 的 C2f 结构如图 3(a) 所示, 该结构使用了残差结构。残差结构是一种在深度学习中广泛使用的结构, 它通过引入跳跃连接允许信息直接从一个层传递到另一个层, 从而缓解了梯度消失问题, 使得网络可以训练得更深。然而, 残差结构容易受到噪声干扰, 出现细节丢失等问题, 导致模型的检测效果下降。针对这一问题, 本文利用前文提出的 MBMA 注意力机制模块替换 Bottleneck 模块, 以到达降低复杂背景干扰、增强特征表达的目的, 改进后结构如图 3(b) 所示。

2.3 标签分配算法

2.3.1 原标签分配算法分析

YOLOv8n 模型目前使用的是 TaskAligned-Assigner 标签分配算法。TaskAlignedAssigner 的匹配策略简单总结为: 根据分类与回归的分数加权的分数选择正样本。TaskAlignedAssigner 的算法具体步骤如下:

1) 计算真实框和预测框的匹配程度, 计算方法如式 (5) 所示。其中, s 是预测类别分值, u 是预测框和真实框的 CIoU 值, α 和 β 为权重超参数, 两者相乘

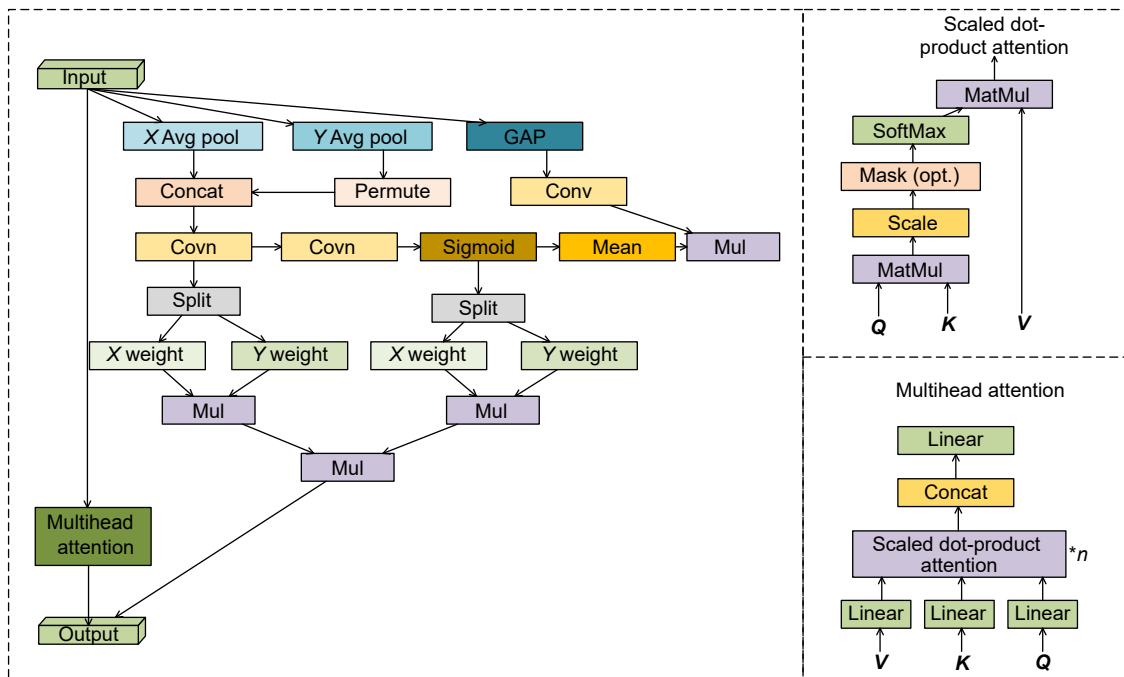


图 2 多分支混合注意力机制 (MBMA 模块) 结构

Fig. 2 Multi-branch hybrid attention mechanism (MBMA module) structure

就可以衡量匹配程度, 当分类的分值越高且 CIoU 值越高时, align_metric 的值就越接近于 1, 此时预测框就与真实框越匹配, 就越符合正样本的标准。CIoU^[21] 的计算方法如式 (6-9), 其中 w 、 h 、 (x, y) 分别表示预测框的宽、高尺寸和中心坐标; w_{gt} 、 h_{gt} 、 (x_{gt}, y_{gt}) 分别表示真实框的宽、高尺寸和中心坐标。IoU (intersection over union) 为真实框和预测框的交并比。 W_i 、 H_i 分别表示交集宽、高尺寸; W_g 、 H_g 分别表示最小边框宽、高尺寸; ε 为权重函数, 用于平衡参数, ν 为纵横比度量函数, 用于衡量高宽比一致性。

2) 对于每个真实框, 直接对 align_metric 匹配程度排序, 选取 topK 个预测框作为正样本。

3) 对一个预测框与多个真实框匹配情况进行处理, 保留 CIoU 值最大的真实框。

$$align_metric = s^\alpha * u^\beta, \quad (5)$$

$$CIoU = IoU - \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} + \varepsilon \nu \right), \quad (6)$$

$$IoU = \frac{W_i H_i}{wh + w_{gt} h_{gt} - W_i H_i}, \quad (7)$$

$$\varepsilon = \frac{\nu}{1 - IoU + \nu}, \quad (8)$$

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w}{h} - \arctan \frac{w_{gt}}{h_{gt}} \right)^2. \quad (9)$$

图 4 展示了边界框在回归过程中的一些示例。在样本 A 和 B 中的真实框尺度相同, 而在 C 和 D 中的真实框尺度也相同。A 和 D 的真实框形状相同, B 和 C 的真实框形状相同。C 和 D 中的边界框尺度大于 A 和 B 中的边界框尺度。在图 4(a) 中, 所有边界框的回归样本具有相同的偏差。图 4(b) 中所有边界框回归样本的形状偏差相同。IoU 表示真实框和预测框的交并比。在图 4(a) 中, A 与 B 的偏差相同, 但它

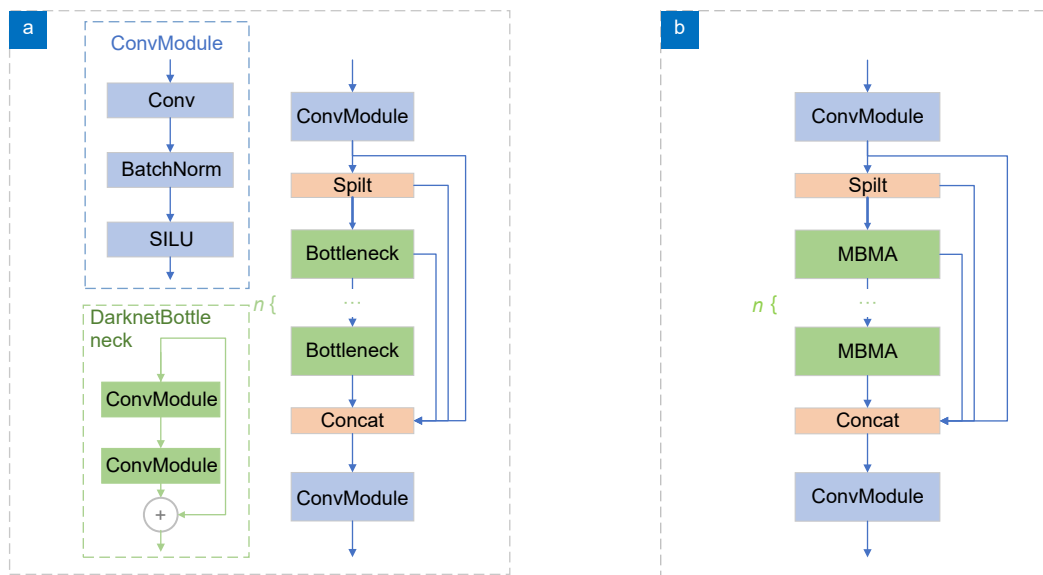


图 3 C2f 结构改进。(a) 原 C2f 结构; (b) 改进后 C2f 结构
Fig. 3 C2f structure improvement. (a) Original C2f structure; (b) Improved C2f structure

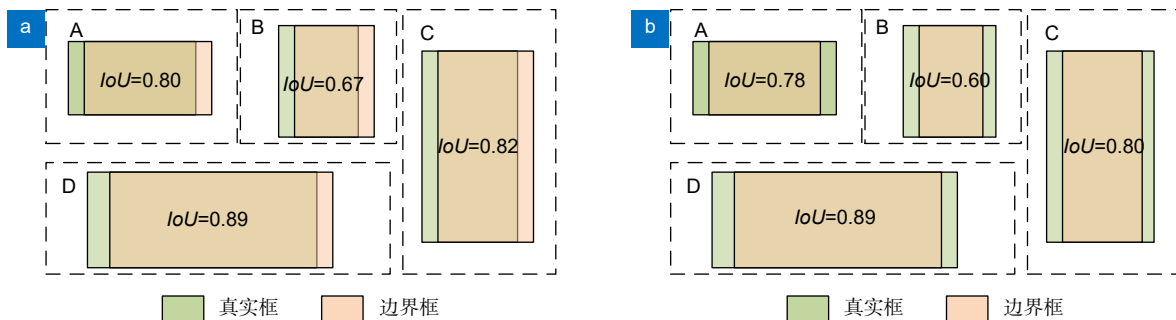


图 4 边界框回归示例
Fig. 4 Example of bounding box regression

们的 IoU 值有所不同, C 和 D 的偏差相同, 它们的 IoU 值也存在差异, 但与 A 和 B 相比, 差异并不显著。在图 4(b) 中, A 和 B 的形状偏差相同, 但 IoU 值不同, C 和 D 的形状偏差相同, IoU 值也不同, 并且与图 4(a) 中的 A 和 B 相比, IoU 值的差异同样不显著。A 和 B 的 IoU 值不同是因为它们的真实框形状不同, 偏差方向分别对应于它们的长边和短边方向。对于 A 来说, 沿着长边方向的偏差对 IoU 值的影响较小, 而对于 B 来说, 短边方向的偏差对 IoU 值的影响较大。小尺度边界框对 IoU 值的变化更为敏感, 真实框的形状对小尺度边界框的 IoU 值影响更显著。因为 A 和 B 的尺度小于 C 和 D, 所以在形状和偏差相同的情况下, 它们的 IoU 值变化更为显著。在图 4(b) 中, 分析形状偏差对边界框回归的影响可以发现, 真实框的形状在回归过程中会影响其 IoU 值。

通过以上分析发现: 当真实目标框不是正方形, 而具有不同的长边和短边时, 如果边界框回归样本的偏差和形状偏差相同且不是全部为零, 那么这些样本的边界框在形状和尺度上的差异会导致它们 IoU 值的显著不同。对于尺度相同的边界框回归样本, 如果在这些样本中偏差和形状偏差相同且不是全部为零, 边界框的形状将会显著影响它们的 IoU 值。特别是, 当偏差沿着边界框的短边方向时, 对应的 IoU 值变化会更加显著。此外, 对于形状相同的边界框回归样本, 如果它们的偏差和形状偏差也相同且不是全部为零, 与大规模的回归样本相比, 小规模边界框的 IoU 值会更显著地受到真实框形状的影响。这是因为小目标对于形状变化更为敏感, 即使是小的偏差也可能导致 IoU 值的较大变化。

2.3.2 标签分配算法改进

对于无人机视角下的道路损伤检测而言, 拍摄到的道路损伤形态不一, 小目标出现的概率相对较大, 出现以上几种情况是十分常见的, 对形状的变化也更为敏感。CIoU 考虑了真实框和预测框的距离、形状和角度对边界框回归的影响, 但忽略了边界框本身的形状和尺度等固有属性也会对边界框回归产生影响, 因此不能很好地适应形态多变的的目标。针对这一问题, 本文引入 ShapeIoU^[22] 代替 CIoU 对 YOLOv8n 中的标签分配算法进行改进。ShapeIoU 的计算公式如式 (10-14) 所示。

$$ShapeIoU = distance^{shape} - 0.5 \times \Omega^{shape}, \quad (10)$$

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}}, \quad (11)$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}}, \quad (12)$$

$$distance^{shape} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2, \quad (13)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta, \theta = 4, \quad (14)$$

$$\begin{cases} \omega_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases}$$

其中: IoU 为交并比, 是目标检测中衡量预测框与真实框匹配程度的常用指标; ww 和 hh 为水平和垂直方向的权重系数, 它们与真实框 (GT 框) 的宽度和高度的比例有关, 用于在计算距离形状项时对不同方向的差异进行加权, 这里的 w^{gt} 和 h^{gt} 分别是 GT 框的宽度和高度, $(w^{gt})^{scale}$ 和 $(h^{gt})^{scale}$ 是与数据集中目标尺度相关的尺度因子; $distance^{shape}$ 为距离形状项, 用于衡量预测框中心点与 GT 框中心点在水平和垂直方向上的距离, 并根据 GT 框的形状进行加权, x_c 和 y_c 是预测框中心点的坐标, x_c^{gt} 和 y_c^{gt} 是 GT 框中心点的坐标, c 是一个尺度因子; Ω^{shape} 为形状惩罚项, 用于惩罚预测框和 GT 框在宽度和高度上的差异, θ 是一个常数, ω_t 是根据预测框和 GT 框在 t 维度 (宽度或高度) 上的差异计算的权重; ω_w 和 ω_h 分别是宽度和高度方向上的惩罚权重, 根据预测框与 GT 框在宽度和高度上的差异进行调整, ShapeIoU 计算示意图如图 5 所示。

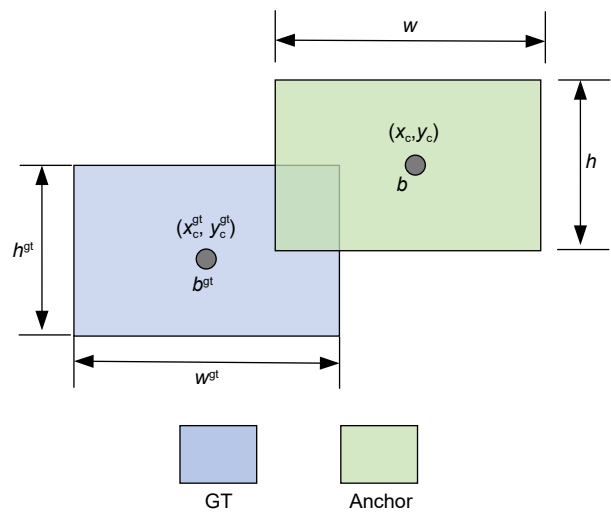


图 5 ShapeIoU 计算示意图

Fig. 5 Schematic diagram of ShapeIoU calculation

ShapeIoU 通过水平和垂直方向的权重系数 (w_w 和 h_h), 量化了真实框的形状特征, 使得损失函数能够根据目标的实际形状调整权重。距离形状项 ($distance^{shape}$) 根据真实框的形状对预测框中心点的偏差进行加权惩罚, 这样即使在 IoU 相同的情况下, 形状差异较大的预测框也会受到更大的损失。形状惩罚项 (Ω^{shape}) 对预测框和真实框在宽度和高度上的差异进行惩罚, 进一步推动预测框向真实框的形状靠拢。ShapeIoU 的计算综合了 IoU、距离形状项和形状惩罚项, 这不仅考虑了预测框和真实框的重叠程度, 还考虑了它们在形状上的一致性。ShapeIoU 方法通过引入尺度因子和形状相关的权重, 能够适应不同形状和尺度的目标。因此, 该方法更为关注检测框本身的形

状和规模, 利用 ShapeIoU 对标签分配算法进行改进也能进一步提高检测精度。

3 实验与分析

3.1 数据集介绍

本文采用的航拍道路损伤图像数据集为 China-Drone^[23], 该数据集在中国多个省份的多种道路场景中由无人机拍摄并采集, 共包含 2396 幅图像, 5 种道路损伤类型, 涵盖了较为全面的道路类型, 有助于提高模型的可迁移性。此外, 使用 RDD2022_Japanese 和 RDD2022_Chinese 数据集对模型的泛化能力进行验证。表 1 介绍了数据集的详细情况, 损伤类型样例如图 6 所示。后文中为表述方便, 称

表 1 数据集道路损伤详情
Table 1 Dataset road damage details

Damage type	Detail	Class name	Number of China-Drone	Number of dataset1	Number of dataset2
Crack	Longitudinal crack	D00	1426	3995	2678
	Lateral crack	D10	1263	3979	1096
	Alligator crack	D20	293	6199	641
Other corruption	Rutting, bump, pothole, separation	D40	86	2243	235
	Crosswalk blur	D43	—	736	—
	White line blur	D44	—	3995	—
Special signs	Manhole cover	D50	—	3553	—
	Repair	Repair	769	—	277

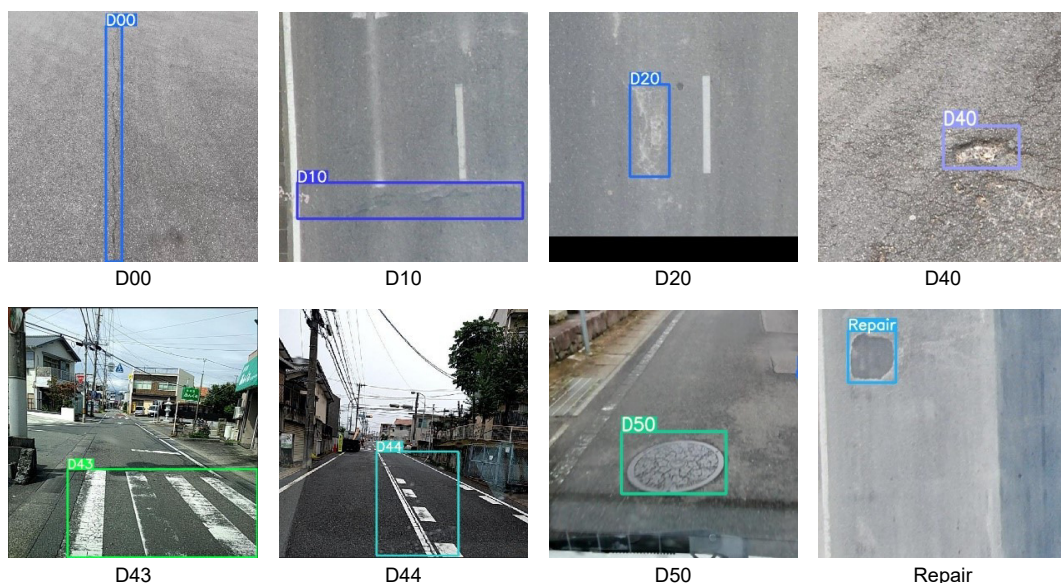


图 6 道路损伤类型样例
Fig. 6 Examples of road damage types

RDD2022_Japanese 为数据集 1, RDD2022_Chinese 为数据集 2。

3.2 实验环境

本实验使用 PyTorch 1.13.1 框架搭建网络模型, 在模型训练设置中, 迭代次数为 150 次, 每批次 (batch size) 包含 16 个样本, 按 7:1:2 比例随机划分数据集, 实验环境配置如表 2 所示。

表 2 实验环境配置
Table 2 Experimental environment configuration

Category	Environment condition
CPU	AMD Ryzen 7 5800X 8-Core Processor
GPU	NVIDIA GeForce RTX 3060
Graphics memory	12 G
Operating system	Ubuntu 22.04
CUDA version	CUDA 12.0
Scripting language	Python

3.3 模型评估指标

为了评估改进算法的性能, 本文选用的算法精度评价指标为平均精度均值 mAP (mean average precision), AP 指准确率 (precision, P) 和召回率 (recall, R) 曲线下的面积, mAP 由各类别 AP 的平均值得出, mAP 越大表示模型精度越高; 选取模型参数量、计算量 (GFLOPS) 和模型体积作为模型性能的评价指标, 模型参数量、计算量和模型体积越小, 则说明计算代价越低; FPS (每秒帧率, 即每秒内可以处理的图片数量) 作为速度评价指标, FPS 越大说明检测速度越快。

3.4 对比实验结果与分析

3.4.1 对比实验结果

为客观评估改进后模型的性能, 选取 YOLO 系

列中轻量级的 YOLOv5n, 基线模型 YOLOv8n, 经典的 Faster-RCNN 模型^[24], 最新的 YOLOv10n 模型、GOLD-YOLO 模型^[25]、TOOD 模型^[26]和 RTMDet-Tiny 模型^[27], 以及基于 Transformer 架构的目标检测模型 RT-DETR^[28], 与本文改进后的模型进行对比, 实验结果如表 3 所示。

从上述表格中可以看出, 本文模型在检测精度 (mAP@0.5) 和模型效率方面均展现出了卓越的性能。对比模型中, YOLOv8n、GOLD-YOLO、Faster-RCNN、TOOD、RT-DETR 检测精度相近, 而 YOLOv5n、YOLOv10n 和 RTMDet-Tiny 检测精度相对较低。Faster-RCNN 为两阶段检测方法, 检测精度较高但计算复杂度高且对小目标检测能力较差。TOOD 利用层注意力机制和可变形卷积, 对形态不一的道路损伤检测效果较好, 但模型复杂度较高、实时性差。YOLOv5n、YOLOv8n、YOLOv10n、GOLD-YOLO 为 YOLO 系列较新的版本, 考虑了实时性和检测精度, 有着广泛的适用性, 但对于道路损伤这类形态多变、细节多的目标针对性低, 且 YOLOv5n、YOLOv10 复杂度低但检测能力不够稳定, 在本文数据集检测效果较差。本文提出模型 MAS-YOLOv8n 加入了 MBMA 模块, 能够在不同的空间位置、通道或特征维度上捕获更加多样化的信息, 加强了特征提取能力和对小目标的关注度, 同时通过多个注意力头的组合, 减少了单一注意力可能带来的偏差和噪声影响。此外, 改进后的标签分配算法也使得模型对形态变化更敏感, 进一步提高检测准确率。在检测精度方面, 本文模型在 China-Drone 数据集上本文模型 mAP 达到 71.6%, 比对比模型中 mAP 最高的 TOOD 高出 2.6%, 比基线模型 YOLOv8 提高了 3.1%。在数据集 1 上本文模型 mAP 达到 67.3%, 与对比模型中

表 3 对比实验结果

Table 3 Comparative experimental results

Model	China-Drone mAP@0.5/%	Dataset1 mAP@0.5/%	Dataset2 mAP@0.5/%	Parameter/M	Model volume/MB
YOLOv5n	64.7	64.0	92.2	2.5	5.03
YOLOv8n	68.5	64.7	93.6	3.0	5.96
YOLOv10n	62.4	61.8	91.4	2.7	5.51
GOLD-YOLO	66.1	65.9	94.5	7.2	11.99
Faster-RCNN	67.8	66.4	94.7	34.6	310.24
TOOD	69.0	65.6	94.9	28.3	243.95
RTMDet-Tiny	65.6	64.1	93.0	4.4	77.76
RT-DETR	68.2	67.2	87.5	20.0	308
MAS-YOLOv8n	71.6	67.3	95.3	3.2	5.96

mAP 最高的 RT-DETR 相近, 比基线模型 YOLOv8 提高了 2.6%。在数据集 2 上本文模型 mAP 达到 95.3%, 对比模型中 mAP 最高的 TOOD 高出 0.4%, 比基线模型 YOLOv8 提高了 1.7%。在计算代价方面, 相比于基线模型 YOLOv8n 参数量增加了 0.2 M, 并没有大幅增加计算开销, 计算消耗仅次于 YOLOv5n 和 YOLOv10。综上, 本文模型 MAS-YOLOv8n 模型相对于对比模型在检测精度方面有所提高, 且没有大幅增加计算开销, 取得了较好的检测性能。

3.4.2 混淆矩阵

为了说明本文算法目标识别和分类的准确性, 对 YOLOv8n 和本文模型在 China-Drone 数据集检测后得到的混淆矩阵进行分析, 混淆矩阵结果如图 7、图 8 所示。本文的混淆矩阵是 5×5 的矩阵, 每一列代表一种真实类别, 每一行代表一种预测类别, 对角线上的数值表示模型正确分类的样本比例, 理想的混淆矩阵应为一个对角矩阵。图 7、图 8 的混淆矩阵中, 最后一行为图片中的道路损伤被误判为背景的样本比例, 这一行可展示道路损伤漏检情况, 其他非对角线上的数值则表示模型错误分类的样本比例, 可以体现道路损伤的错检情况。通过对比两个混淆矩阵, 能够发现本文模型混淆矩阵在对角线上的样本比例均高于 YOLOv8n 模型混淆矩阵, 分别高出 0.05、0.03、0.16、0.02 和 0.1, 且本文模型混淆矩阵在最后一行的样本比例均低于原 YOLOv8 模型, 分别降低了 0.07、0.02、0.07、0.01、0.2, 能够说明本文模型漏检、错检率更低, 检测效果较好。

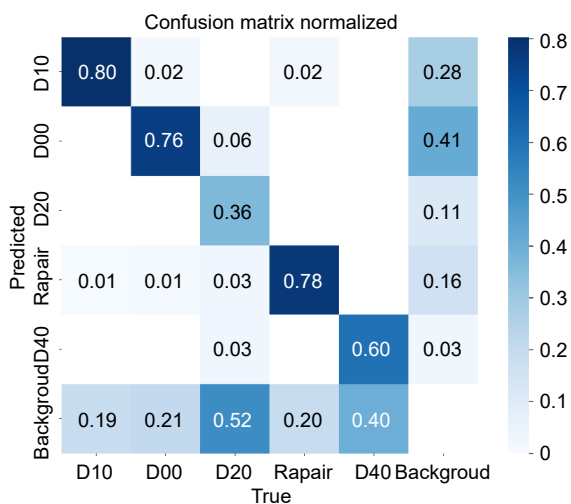


图 7 YOLOv8n 混淆矩阵
Fig. 7 YOLOv8n confusion matrix

3.4.3 实验结果分析

为了展示本文算法的检测效果, 选取了密集目标、识别困难、存在干扰等典型检测结果作为示例, 如图 9 所示。从图 9 第 1 列可以看出对于较为密集的目标, 对比模型中除 Faster-RCNN、YOLOv8n 和本文模型外都出现错检现象, 而 MAS-YOLOv8n 对上下文信息的理解能力更强, 检测效果较为准确, 对相对于其他模型错检情况减少; 从第 2 列可以看出, 对于难以辨识的道路损伤, 对比模型中除 GOLD-YOLO、YOLOv8n、RTMDet-Tiny 和本文模型外均出现重复检测情况, 而 GOLD-YOLO、YOLOv8 和 RTMDet-Tiny 模型检测出的范围不够完整, 相对而言, MAS-YOLOv8n 对目标的形态更为敏感, 检测的目标范围更为准确, 能够较为完整地框出检测目标; 从第 3 列可以看出当存在树枝等类似目标干扰的情况下, YOLOv10、GOLD-YOLO 和 RTMDet-Tiny 检测结果中出现将树枝和阴影错检为目标的情况, 而 MAS-YOLOv8n 对目标的识别能力更强, 能够更好地区分出需要检测的道路损伤; 第四列图片存在阴影、污渍、噪声等干扰, 且目标较小, 有一定识别难度, 除 YOLOv8、GOLD-YOLO、RT-DETR 和本文模型外均出现错检和漏检, 而 MAS-YOLOv8n 能够较好地排除噪声干扰, 并捕获到更为丰富的特征信息, 从而更为准确地检测出目标。结合表 3 对比实验结果和图 9 的检测样例能够得出, 本文模型在一定程度上改善了模型错检、漏检的情况, 并且在检测精度、计算效率和资源占用方面均展现出综合优势, 优于以上对比模型。

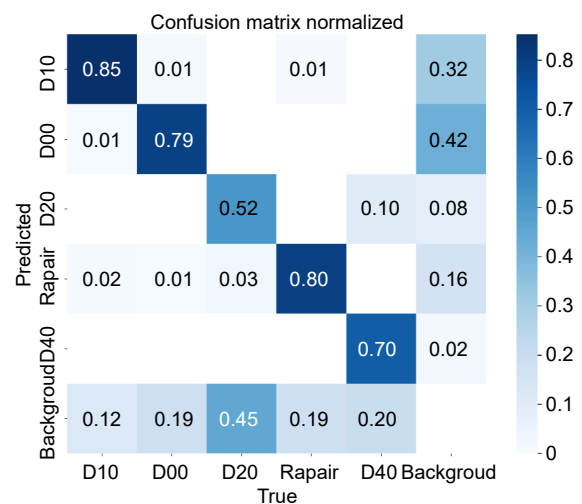


图 8 本文模型混淆矩阵
Fig. 8 Confusion matrix of the model in this article

3.5 消融实验结果与分析

3.5.1 注意力机制有效性验证

为验证本文提出的注意力模块在道路损伤检测模型上的有效性, 将 MBMA 模块与常见的几种注意力机制在两个数据集上进行比较, 实验结果如表 4 所示。MBMA 模块在三个数据集上的平均精度均值分别达到 70.7%、66.7% 和 94.8%, 与 SE、CMBA、CA 相比, 检测精度均有提升。图 10 为实验中使用不同注意力机制检测后的热力图可视化结果, 图中不同颜色的区域显示了模型关注的程度, 红色或黄色表示高度关注的区域, 蓝色或绿色则表示关注度较低的区域,

可以展示特征的提取效果。通过特征热力图的对比可以看出, 相比于其他注意力机制, 加入 MBMA 模块的模型更集中在道路损伤本身, 对重要特征的关注度更强, 能够关注到细小的特征, 从而提高了检测效果。

3.5.2 标签分配算法改进有效性验证

为验证标签分配算法改进有效性, 将改进前后的实验结果进行对比, 结果如图 11 所示, 图 11(a) 和图 11(b) 分别展示了标签分配算法改进前后在两个数据集上的损失值和 mAP 变化。能够看出, 利用 ShapeIoU 修改标签算法后, 初始损失值相对原损失值更低。这主要是由于道路损伤的尺寸、形态各异,

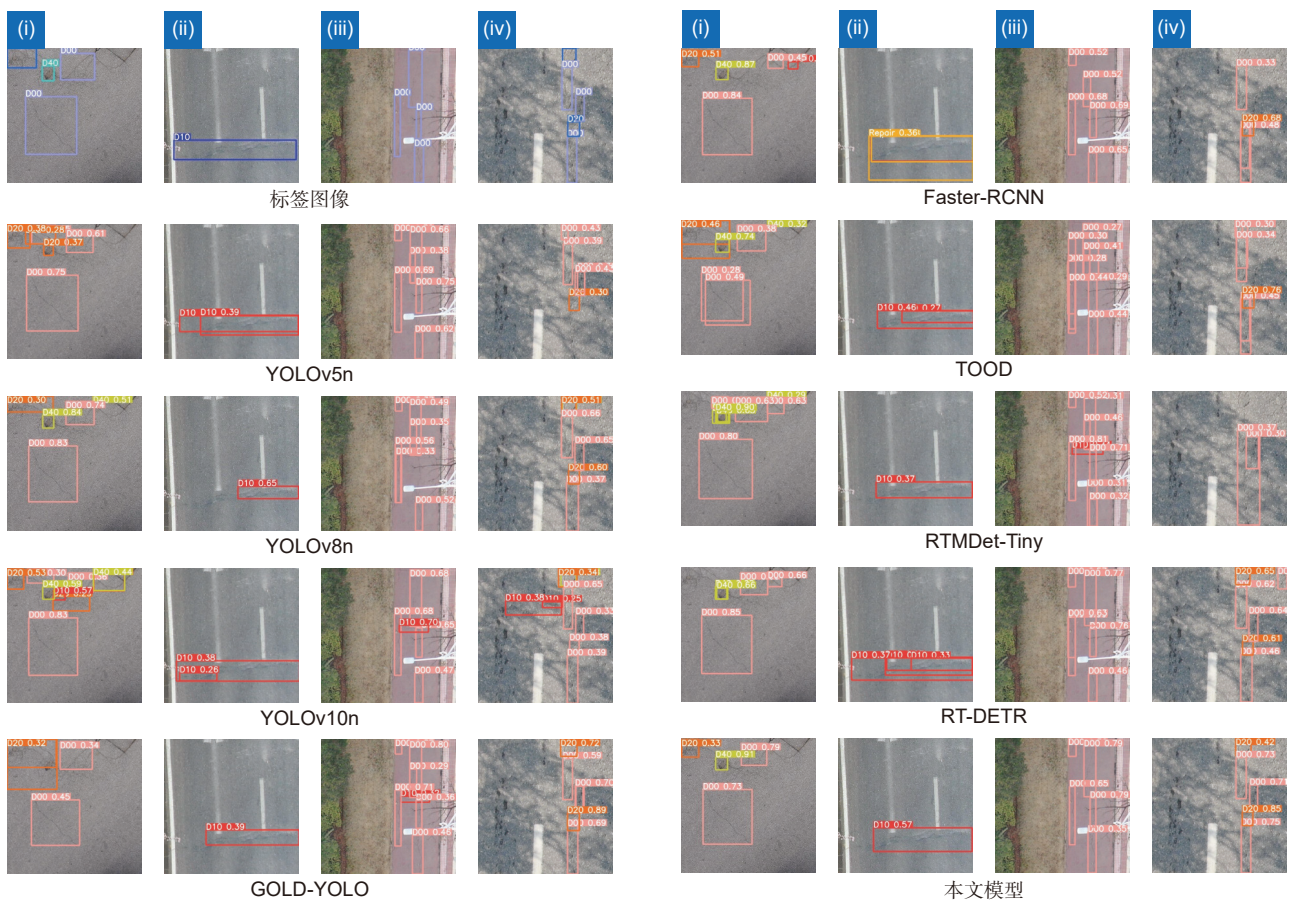


图 9 检测结果示例

Fig. 9 Example of detection results

表 4 注意力机制验证结果

Table 4 Verification results of attention mechanism

Attention mechanism	China-Drone mAP@0.5/%	Dataset1 mAP@0.5/%	Dataset2 mAP@0.5/%	Parameter/M
—	68.5	64.7	93.6	3.0
SE	69.1	64.1	93.5	3.1
CMBA	67.4	65.5	94.5	3.2
CA	68.8	65.7	94.5	3.2
MBMA	70.7	66.7	94.8	3.2

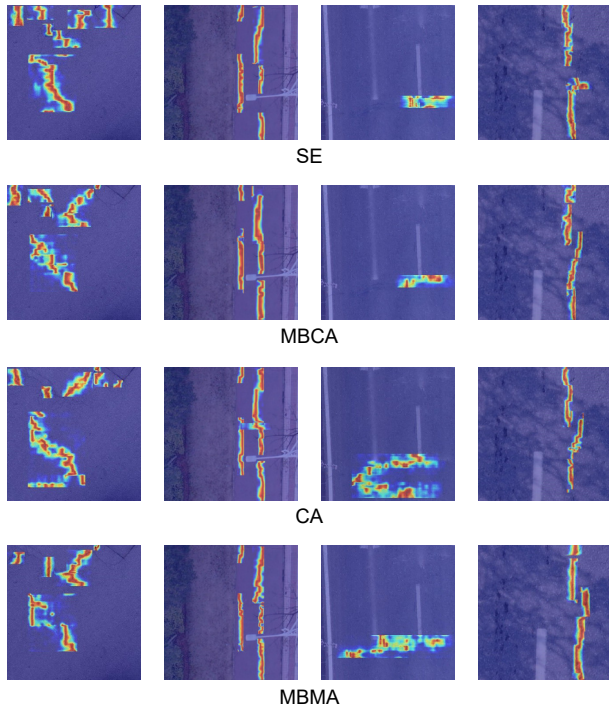


图 10 注意力机制可视化特征热力图

Fig. 10 Heat map of visual features of attention mechanism

而 ShapeIoU 对目标的形状和尺度敏感, 可以更精确地评估预测框与真实框之间的匹配程度, 通过更合理地选择或调整锚框, 更好地匹配真实目标的形状, 从而提高标签分配的准确性。训练过程中, 本文模型的收敛速度明显加快, 收敛后的损失值相对改进前更低, 并且在三个数据集上检测精度均有所提高, 这也能够说明用 ShapeIoU 改进后的标签分配算法更适合本文使用的数据集, 进一步证明了改进后算法的有效性。

3.5.3 消融实验结果

为验证改进模型各个改进部分的有效性, 本文通过消融实验对模型进行对比分析, 结果如表 5 所示。模型 1 为基线模型 YOLOv8n, 模型 2 为使用 MBMA 模块改进后的模型, 相比于基线模型, 使用 MBMA 模块后增强了特征的敏感性和表达能力, 捕获到了更为丰富的信息, 在三个不同的数据集上, mAP 分别提高了 2.2%、2% 和 1.2%。在计算代价方面, 模型 2 的参数量增加了 0.2 M, 计算量和模型体积均没有发生变化, 说明 MBMA 模块并没有大幅增加额外的计算开销。在计算速度方面, 模型 2 相比于

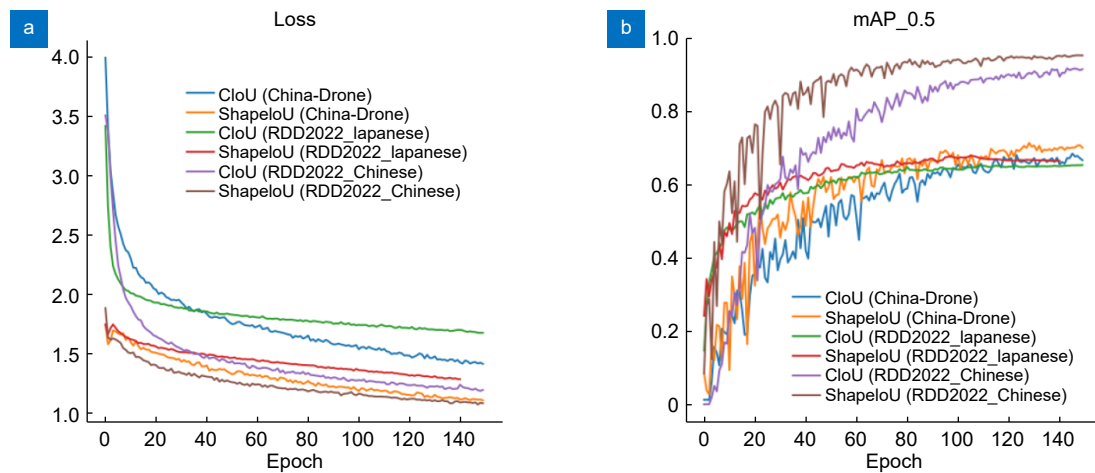


图 11 标签分配算法改进前后的实验结果。(a) Loss 值变化对比图; (b) mAP 变化对比图

Fig. 11 Experimental results before and after improvement of the label allocation algorithm. (a) Comparison of Loss value changes; (b) Comparison of mAP changes

表 5 消融实验结果

Table 5 Results of the ablation experiment

Model	China-Drone mAP@0.5/%	Dataset1 mAP@0.5/%	Dataset2 mAP@0.5/%	Parameter/M	GFLOPS	Model volume/MB	FPS
1 YOLOv8n	68.5	64.7	93.6	3.0	8.1	5.96	137
2 +MBMA	70.7	66.7	94.8	3.2	8.1	5.96	116
3 +ShapeloU	70.9	67.0	95.0	3.0	8.1	5.96	135
4 MAS-YOLOv8n	71.6	67.3	95.3	3.2	8.1	5.96	114

基线模型 YOLOv8n 检测 FPS 有所下降, 但仍能满足绝大多数应用场景的实时性要求, 证明了 MBMA 模块的适用性。模型 3 为使用 ShapeIoU 对标签分配算法改进后的模型, 相比于基线模型, 改进后的标签分配算法可以在分配过程中使用更精细的度量来确定最佳的锚框匹配, 从而提高训练效率和模型性能, 更适用于道路损伤这种目标尺度不一、形状多变的目标检测任务, 从实验结果中可以看出, 模型 3 相对于模型 1, 在三个数据集上 mAP 分别提高了 2.4%、2.3% 和 1.4%, 且不增加额外计算代价, 也没有降低检测速度。模型 4 为本文提出模型, 它综合了模型 2 和模型 3 的优点, 进一步提高了检测精度, 相对于基线模型, 在两个数据集上的 mAP 分别提高了 3.1%、2.6% 和 1.7%, 由于添加 MBMA 模块导致检测速度有所下降, 但仍能满足实时性需求, 且计算量和模型体积没有明显增加, 进一步验证了改进部分的有效性。

4 总结

无人机采集道路图像并对道路损伤进行检测极大地提升了道路损伤检测的效率, 对于保障道路交通安全、延长道路使用寿命等方面都具有重要意义。本文针对无人机采集的道路图像特点, 在 YOLOv8n 模型的基础上提出一种新的道路损伤检测模型 MAS-YOLOv8n。其中, 针对残差结构容易受到噪声干扰, 出现细节丢失等问题, 提出一种多分支混合注意力机制 (MBMA 模块), 使用该模块对 YOLOv8 中 C2f 结构进行修改。MBMA 模块通过多个分支处理输入特征, 每个分支关注不同的特征维度 (高度、宽度和通道), 从而能够更全面地捕捉特征信息, 同时利用多头注意力机制增强了模块对特征上下文信息的提取, 减少了单一注意力可能带来的偏差, 增强了模块的鲁棒性。此外, 针对道路损伤图像形态差异大、小目标出现概率高的特点, 对 TaskAlignedAssigner 标签分配算法进行改进, 引入 ShapeIoU, 考虑了边界框的形状和尺度, 提高了目标检测中预测框和真实框之间匹配的准确性, 使其更适应形态多变的目标, 进一步提升了模型的检测效果。将所提出的模型在 China-Drone、RDD2022_Japanese 和 RDD2022_Chinese 三个道路损伤数据集上进行验证, mAP 分别达到了 71.6%、67.3% 和 95.3%, 比基线模型 YOLOv8 提高了 3.1%、2.6%、1.7%。与 YOLOv5n、YOLOv8n、

YOLOv10n、GOLD-YOLO、Faster-RCNN、TOOD、RTMDet-Tiny、RT-DETR 相比, 本文模型表现最佳。该实验结果能够证明, 本文设计的道路损伤检测模型能够加强对数据特征的提取能力, 提高检测效果, 综合性能更优。

参考文献

- [1] Pan Y F, Zhang X F, Cervone G, et al. Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2018, 11(10): 3701–3712.
- [2] An X G, Dang J W, Wang Y P, et al. UAV image pavement disease detection based on improved YOLOv4[J]. *Radio Eng*, 2023, 53(6): 1285–1294.
安学刚, 党建武, 王阳萍, 等. 基于改进 YOLOv4 的无人机影像路面病害检测方法[J]. *无线电工程*, 2023, 53(6): 1285–1294.
- [3] Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, 51(5): 240051.
张润梅, 肖钰霏, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法[J]. *光电工程*, 2024, 51(5): 240051.
- [4] Kahmann S L, Rausch V, Plümer J, et al. The automatized fracture edge detection and generation of three-dimensional fracture probability heat maps[J]. *Med Eng Phys*, 2022, 110: 103913.
- [5] Di Benedetto A, Fiani M, Gujski L M. U-Net-based CNN architecture for road crack segmentation[J]. *Infrastructures*, 2023, 8(5): 90.
- [6] Arman M S, Hasan M M, Sadia F, et al. Detection and classification of road damage using R-CNN and faster R-CNN: a deep learning approach[C]//*Second EAI International Conference on Cyber Security and Computer Science*, 2020: 730–741. https://doi.org/10.1007/978-3-030-52856-0_58.
- [7] Girshick R. Fast R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
- [8] Lü B Y, Zhuo Z F, Han Y S, et al. Target detection based on faster region convolution neural network[J]. *Laser Optoelectron Prog*, 2021, 58(22): 2210017.
吕本远, 褚真福, 韩永赛, 等. 基于 Faster 区域卷积神经网络的目标检测[J]. *激光与光电子学进展*, 2021, 58(22): 2210017.
- [9] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, 49(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. *光电工程*, 2022, 49(3): 210372.
- [10] Ajmera S, Kumar C A, Yakaiah P, et al. Real-time pothole detection using YOLOv5[C]//*2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 2022: 1–5. <https://doi.org/10.1109/ASSIC55218.2022.10088290>.
- [11] Eriksson M. Road damage detection with YOLOv8 on Swedish roads[EB/OL]. [2024-07-10]. <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1803747>.
- [12] Pham V, Nguyen D, Donan C. Road damage detection and classification with YOLOv7[C]//*2022 IEEE International Conference on Big Data (Big Data)*, 2022: 6416–6423.

- <https://doi.org/10.1109/BigData55660.2022.10020856>.
- [13] Jiang Y T, Yan H T, Zhang Y R, et al. RDD-YOLOv5: road defect detection algorithm with self-attention based on unmanned aerial vehicle inspection[J]. *Sensors*, 2023, 23(19): 8241.
- [14] Cheng Q H, Chen D F, Wang X F. Small target detection in UAV images based on NDM-YOLOv8[J]. *Comput Technol Dev*, 2024.
程期浩, 陈东方, 王晓峰. 基于 NDM-YOLOv8 的无人机图像小目标检测[J]. *计算机技术与发展*, 2024.
- [15] Li Y T, Fan Q S, Huang H S, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. *Drones*, 2023, 7(5): 304.
- [16] Long W D, Peng B, Hu J, et al. Road damage detection algorithm based on enhanced feature extraction[J]. *J Comput Appl*, 2024, 44(7): 2264–2270.
龙伍丹, 彭博, 胡节, 等. 基于加强特征提取的道路病害检测算法[J]. *计算机应用*, 2024, 44(7): 2264–2270.
- [17] Meng P S, Wang F, Zhai W G, et al. UAV-to-ground target detection based on YOLO-DSBE[J]. *Aero Weapon*, 2024.
孟鹏帅, 王峰, 翟伟光, 等. 基于 YOLO-DSBE 的无人机对地目标检测[J]. *航空兵器*, 2024.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
<https://doi.org/10.1109/CVPR.2018.00745>.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]// *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, 2018: 3–19.
https://doi.org/10.1007/978-3-030-01234-2_1.
- [20] 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*, 2021: 13708–13717.
<https://doi.org/10.1109/CVPR46437.2021.01350>.
- [21] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]// *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020: 12993–13000. <https://doi.org/10.1609/aaai.v34i07.69993>.
- [22] Zhang H, Zhang S J. Shape-IoU: more accurate metric considering bounding box shape and scale[Z]. arXiv: 2312.17663, 2023. <https://doi.org/10.48550/arXiv.2312.17663>.
- [23] Arya D, Maeda H, Ghosh S K, et al. RDD2022: a multi-national image dataset for automatic Road Damage Detection[Z]. arXiv: 2209.08538, 2022. <https://doi.org/10.48550/arXiv.2209.08538>.
- [24] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137–1149.
- [25] Wang C C, He W, Nie Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[C]// *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023: 2224.
- [26] Feng C J, Zhong Y J, Gao Y, et al. TOOD: task-aligned one-stage object detection[C]// *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021: 3490–3499.
<https://doi.org/10.1109/ICCV48922.2021.00349>.
- [27] Xiang H F, Jing N F, Jiang J F, et al. RTMDet-R2: an improved real-time rotated object detector[C]// *6th Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 2024: 352–364. https://doi.org/10.1007/978-981-99-8555-5_28.
- [28] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[Z]. arXiv: 2304.08069, 2023.
<https://doi.org/10.48550/arXiv.2304.08069>.

作者简介



王晓燕(1980-), 女, 博士, 副教授, 研究方向为数字图像处理、数学在工程中的应用、大数据处理。

E-mail: wangxy252@163.com



【通信作者】李杰(1977-), 男, 博士, 教授, 研究方向为无人平台环境智能感知、车辆变速摩擦与制动技术、无人车辆智能控制技术、超高速激光熔覆与表面智能强化。

E-mail: lijie1@bucea.edu.cn



扫描二维码, 获取PDF全文

