

刘晶, 祁文哲, 吕德芳, 等. 改进 YOLOv5 的轨面伤损检测方法[J]. 智能计算机与应用, 2024, 14(6): 71-78. DOI:10.20169/j.issn.2095-2163.240610

改进 YOLOv5 的轨面伤损检测方法

刘晶¹, 祁文哲¹, 吕德芳^{1,2}, 李杉杉²

(1 兰州交通大学 机电工程学院, 兰州 730070; 2 兰州交通大学 机电技术研究所, 兰州 730070)

摘要: 针对目前钢轨表面缺陷检测中使用传统检测方法检测效率低, 定位精度不达标的问题, 提出了一种基于 YOLOv5 的轨面伤损检测算法。论文在 backbone 模块中加入 CA 模块提高检测精度, 在 Neck 部位添加 CA 模块赋予目标信息更大的权重。采用 *SIoU* 损失函数代替 *GIoU* 损失函数作为边界回归函数, 提高边界框的定位精度。通过改进算法与其他算法在相同实验环境下对轨面掉块、磨损、压溃、剥落、波磨这 5 种常见表面缺陷进行检测实验, 结果表明, 本文算法对 5 种缺陷的识别平均精度 (*mAP*) 达到 94.8%, 检测模型的 *FPS* 为 33 帧/s, 模型占用内存 37.8 MB, 相较于其他检测方法在检测精度、检测速度和内存占用方面都有显著的提高和改善, 为钢轨表面缺陷检测方法提供了新的检测思路。

关键词: 目标检测; *SIoU*; 注意力机制; YOLOv5

中图分类号: TH391

文献标志码: A

文章编号: 2095-2163(2024)06-0071-08

Improved track damage detection method for YOLOv5

LIU Jing¹, QI Wenzhe¹, LÜ Defang^{1,2}, LI Shanshan²

(1 School of Mechanical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;

2 Mechatronics T&R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Aiming at the problems of low detection efficiency and substandard positioning accuracy using traditional detection methods in rail surface defect detection, a rail surface damage detection algorithm based on YOLOv5 is proposed. In addition to adding CA module to the backbone module to improve the detection accuracy, the CA module is also added to the Neck part to allocate greater weight to the target information, which has a certain effect on the correction of the detection results. In the regression part of the bounding box, the *SIoU* loss function is used instead of the *GIoU* loss function as the boundary regression function to further improve the positioning accuracy of the bounding box. The results show that the average accuracy (*mAP*) of the algorithm for identifying the five defects reaches 94.8%, and the *FPS* of the detection model is 33 frames/s. The model occupies 37.8 MB in the memory, which is significantly improved in terms of detection accuracy, detection speed and memory occupation compared with other detection methods, which provides a new detection idea for rail surface defect detection methods.

Key words: object detection; *SIoU*; attention mechanism; YOLOv5

0 引言

随着国内铁路的快速发展, 交通枢纽城市的轨道交通运营里程和运行时间在不断增长, 伴随而来的安全运营压力和社会责任也越来越大。同时钢轨出现各类损伤的频率也会有所增加。而轨面一些常见的受损, 像掉块、磨损、压溃、剥落、波磨等, 这些损伤在普速、重载及高速铁路中都是普遍存在的^[1]。

如果轨面发生伤损就容易造成线路结构失稳、变形超限、断裂等事件^[2]。为了避免列车运行途中发生此类突发事件, 就需要对钢轨表面的状态进行实时监测。此前钢轨的检测主要依靠工人用眼睛检查, 时间成本和人工成本较高, 检测结果也受工人主观因素的影响。近年来随着轨检技术的发展, 针对钢轨伤损的检测方法有了一些无损检测方法, 比如超声探伤、涡流探伤、磁粉探伤、射线探伤、渗透探伤和

作者简介: 刘晶(1998-), 女, 硕士研究生, 主要研究方向: 深度学习, 机器视觉; 吕德芳(1995-), 男, 博士研究生, 主要研究方向: 深度学习, 计算机视觉, 人工智能等; 李杉杉(1997-), 女, 硕士研究生, 主要研究方向: 计算机视觉, 深度学习。

通讯作者: 祁文哲(1966-), 男, 教授, 主要研究方向: 机电设备控制与监测技术, 机械成套设备及机器人控制技术, 检测技术与自动化等。

Email: qiwez@mail.lzjtu.cn

收稿日期: 2023-07-03

机器视觉检测技术^[3]。目前国内应用的机器视觉无损检测设备以及相关技术明显优于传统的人工检测,更加具有全自动化、客观科学等突出特点。基于机器视觉的钢轨缺陷检测近些年来得到了国内外研究学者的广泛关注,相关研究成果在普速、高铁和重载等各铁路线路中推广使用^[4]。基于机器视觉的检测方法主要包括传统图像处理方法和基于深度学习方法^[5]。传统的图像采集处理的方法中特征数据的提取都主要还是依赖于人工设计的特征提取器,需要有一定的专业知识去处理复杂的调参数过程,而且每个数据提取处理方法也都是直接地针对具体应用场景去研发的,泛化学习的能力很低,造成算法鲁棒性都普遍较差。比如,岳彪等学者^[6]提出的一种基于 RetiNex 的钢轨表面的缺陷图像增强方法。并引入反射分量 $z - score$ 改进了 RetiNex 图像处理框架。实验验证这种该方法在提高缺陷与背景对比度的同时,还能有效克服阴影的干扰。闵永智等学者^[7]通过建立正负样本数据库以及样本特征数据库,设计缺陷分类器,对缺陷进行分类。通过实验验证得出该方法识别率波动明显较小,但识别时间较传统方法稍高,而且对道岔处的检测还有待研究。

胡文博等学者^[4]提出一种新的后处理方法实现伤损识别和分类的自动化,并从数据、方法和干扰因素三个方面对分类效果进行验证和评价。实验表明 ResNet-50 深度残差网络鲁棒性最强,但无法获得实时里程信息,有一定的局限性。刘孟轲等学者^[8]通过对算法进行数学推导并构建模型,使用深度学习框架对轨面缺陷进行检测及分类。但是,他无法区分是具体缺陷、还是污渍等,并且卷积神经网络模型还有待完善。李浪怡等学者^[9]提出了一种基于 YOLOv5 算法的轨面缺陷检测算法。对特征网络进行轻量化处理,并引入注意力机制对不同通道进行权衡。而且修改了检测输出层结构用于识别小目标。改进后的 YOLOv5 检测算法基本满足轨面缺陷精准检测和定位任务要求。

本文在分析前人创新工作基础上,通过对 YOLOv5m 检测算法对轨面缺陷进行检测,嵌入 CA 注意力模块,能够将横向和纵向的位置信息编码到 Channel Attention 中,使得移动网络能够关注大范围的位置信息又不会带来过多的计算量。并引用一种新的边界框回归损失函数 $SIoU Loss$, 解决了预测框在回归时随意匹配的问题,提高模型鲁棒性和泛化能力,加快网络的收敛速度。改进后的方法既提高

了特征提取及定位能力,同时也保证了较高检测速度。

1 YOLOv5m 目标检测算法

YOLOv5 是一种单阶段目标检测算法,具有均值权重文件更小、训练时间更短和推理速度更快的优势。YOLOv5 有 4 种不同模型。本文在研究比较后选择了准确率较高、运行速度较快的 YOLOv5m 模型对钢轨表面缺陷进行检测,并对原算法进行改进,提高算法的整体检测性能。

研究可知,输入端、Backbone、Neck、Prediction 四个部分共同组成了 YOLOv5m 模型的网络框架,如图 1 所示。

其中,输入端部分对数据集进行 Mosaic 数据增强与自适应图片缩放等操作。Backbone 部分由 Focus 与 CSP 等结构组成。Focus 结构先把图片切片再用卷积核对特征图进行卷积。模型不仅在 Backbone 部分添加 CSP1_X 结构,还创新性地 Neck 部分添加 CSP2_X 结构,极大地减少了模型的运算量。Neck 部分由 FPN 与 PAN 结构组成。FPN 从上向下对上层的强语义信息进行上采样,与不同层的输出进行特征聚合。PAN 自下而上的特征金字塔传递强定位信息。YOLOv5m 选择 $GIoU$ 作为损失函数, $GIoU$ 相对于 IoU 解决了模型生成的部分预测矩形框与图片标注的真实矩形框之间存在距离、不相交时损失为 0 的情况。同时 YOLOv5m 模型采用加权 NMS 对模型生成的所有预测矩形框进行筛选,选出预测效果最好的预测矩形框作为预测结果。

2 算法改进

2.1 融入注意力机制

随着 YOLOv5m 网络层数持续增加,输出端提取到的特征信息更加抽象,这使得模型对小目标的检测能力下降,因此引入了一种新型注意力机制、称为 Coordinate Attention(CA)。CA 注意力机制模块示意如图 2 所示。由图 2 可知,是在高度和宽度两个方向进行池化,获取到 2 个方向的特征矩阵,再通过卷积处理以及激活函数使得此时的特征矩阵包含了一部分图像位置信息^[10]。

2.1.1 Coordinate 信息嵌入

研究中通常使用的全局池化方法并不能有效地保存位置信息。为此,按照式(1)分解了全局池化:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (1)$$

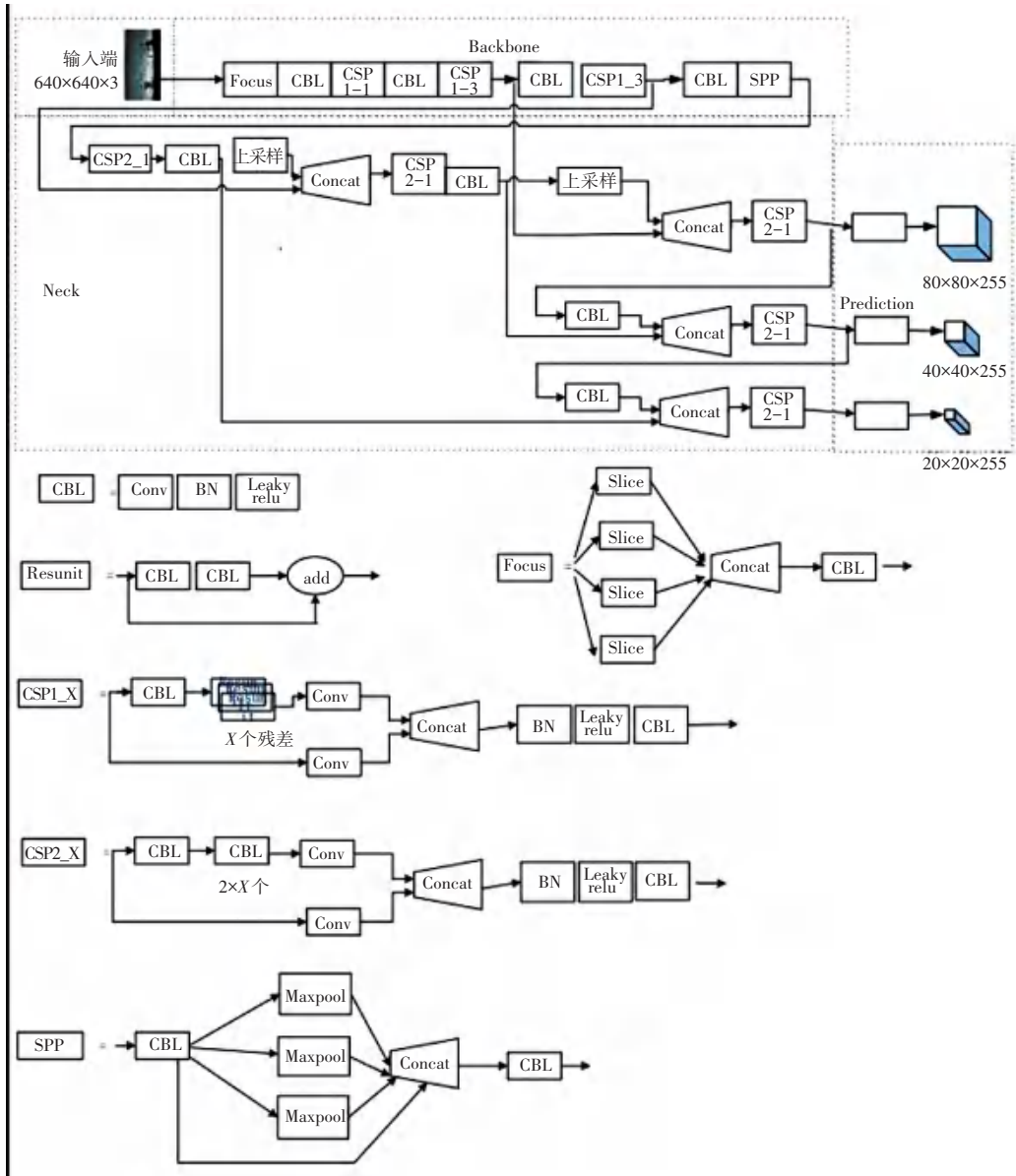


图 1 YOLOv5m 网络模型

Fig. 1 YOLOv5m network model

给定输入 X , 在水平方向使用尺寸 $(H, 1)$ 的 pooling kernel 进行水平方向编码, 在垂直方向使用尺寸 $(1, W)$ 的 pooling kernel 编码。因此, 高度为 h 的第 c 通道的输出可以表示为:

$$Z_c^h = \frac{1}{W} \sum_{0 \leq i < w} X_c(i, j) \quad (2)$$

同样, 宽度为 w 的第 c 通道的输出可以写成:

$$Z_c^w = \frac{1}{H} \sum_{0 \leq j < H} X_c(j, w) \quad (3)$$

2.1.2 Coordinate Attention 生成

Coordinate 信息嵌入后, 再进行 Concatenate 操作, 然后使用卷积变换函数对其进行变换操作:

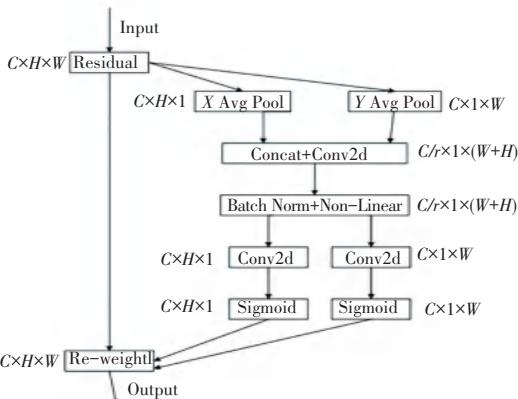


图 2 CA 注意力机制模块示意图

Fig. 2 Schematic diagram of CA attention mechanism module

$$f = \delta(F_1([Z^h, Z^w])) \quad (4)$$

其中, $[\cdot, \cdot]$ 表示沿空间维数的 Concatenate 操作; δ 表示非线性激活函数; f 表示中间特征映射。

然后沿着空间维数将 f 分解成 2 个单独的张量, 即:

$$f^h \in R^{C/\gamma \times H} \quad (5)$$

$$f^w \in R^{C/\gamma \times W} \quad (6)$$

其中, γ 表示缩减比。利用 2 个 1×1 卷积变换 F_h 和 F_w 将 f^h 和 f^w 变换为具有相同通道数的张量到输入 X :

$$g^h = \sigma(F_h(f^h)) \quad (7)$$

$$g^w = \sigma(F_w(f^w)) \quad (8)$$

其中, σ 表示 *sigmoid* 激活函数。

最后, 输出的 y 可以写成:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (9)$$

本文所需识别的轨面伤损情况相对于整张图片较小, 均属于中、小目标, 因此本文在模型的 Backbone 部位和 Neck 部位加入 CA 注意力模块, 引导模型更加关注中、小目标信息的特征, 减少模型对无关特征的关注, 以提升模型对中、小目标的检测性能。改进后的 YOLOv5m-CA 整体结构如图 3 所示。

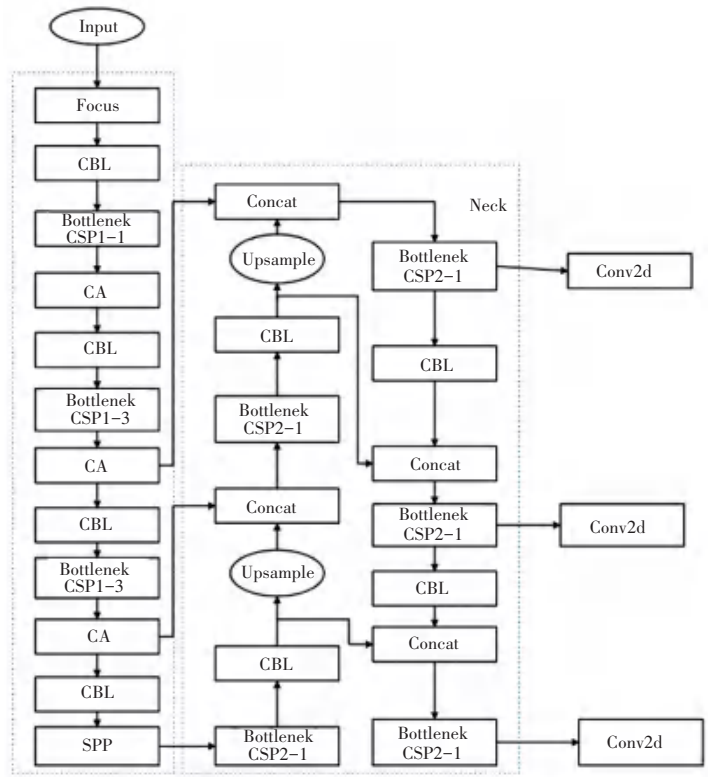


图 3 嵌入 CA 模块的 YOLOv5m 网络模型整体结构框架图

Fig. 3 The overall structure of the YOLOv5m network model embedded in the CA module

2.2 损失函数改进

神经网络中的损失 (*loss*) 反映预测值和真实值的差异, YOLOv5 中损失有 3 种: 回归损失 e_d , 分类损失 e_s 和置信度损失 e_k , 定义网络的总损失 l 为:

$$l = e_d + e_s + (a_1 + a_2 + a_3) e_k \quad (10)$$

$$e_d = 1 - IoU + \frac{s_2 - s_1}{s_2} \quad (11)$$

由此得到的 *GIoU* 示意如图 4 所示。

其中, 回归损失使用 *GIoU Loss*, 计算公式可写为:

$$GIoU = IoU - \frac{S_2 - S_1}{S_2} \quad (12)$$

$$L_{GIoU} = 1 - GIoU = e_d \quad (13)$$

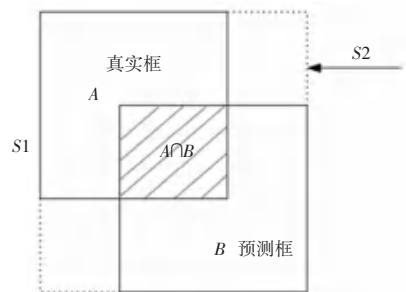


图 4 *GIoU* 示意图

Fig. 4 Schematic diagram of *GIoU*

该损失函数存在 2 个主要缺点。一是当真实框和预测框没有交集时,收敛速度较为缓慢;二是当真实框和预测框出现包含情况时,损失退化为 IoU 损失,收敛依然困难^[11]。

为此,引用一种新的边界框回归损失函数 $SIoU$ Loss, 解决了预测框在回归时随意匹配的问题,提高模型鲁棒性和泛化能力,加快网络的收敛速度^[12]。

损失函数 $SIoU$ 包含 4 部分:回归之间的向量角度、中心点距离、长宽比、重叠面积,引入向量角度解决了预测框与真实框随意匹配问题,最大限度减少与距离相关变量的数量。向量角度损失计算如图 5 所示,两者之间的计算公式如下所示:

$$\beta = \frac{\Pi}{2} - \alpha \quad (14)$$

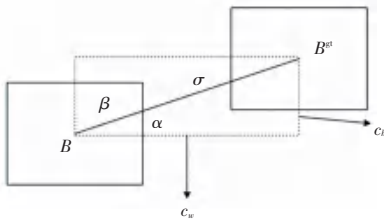


图 5 向量角度损失计算

Fig. 5 Vector angle loss calculation

同时,引入以下范式用来定义预测框与标注的真实边界框的向量角度损失函数,具体见式(15):

$$\Lambda = 1 - 2\sin(\sin^{-1}x - \frac{\Pi}{4}) \quad (15)$$

其中,

$$x = \frac{c_h}{\sigma} = \sin\alpha \quad (16)$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \quad (17)$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \quad (18)$$

距离损失定义的计算公式为:

$$\Delta = \sum t = x, y(1 - e^{-\gamma t}) \quad (19)$$

其中,

$$\rho_x = \frac{\frac{\partial b_{c_x}^{gt}}{\partial c_x} - b_{c_x} \ddot{\partial}^2}{e^{c_w} \ddot{\partial}} \quad (20)$$

$$\rho_y = \frac{\frac{\partial b_{c_y}^{gt}}{\partial c_y} - b_{c_y} \ddot{\partial}^2}{e^{c_h} \ddot{\partial}} \quad (21)$$

$$\gamma = 2 - \Lambda \quad (22)$$

在新定义的惩罚标准中可以看出随着角度的增大,会使得问题变复杂,所以又增加了一个随着角度增加优先于距离值的 γ 参数。同样,为了控制长宽比在总损失中的重要程度,也增加了一个超参数 θ ,通过

将遗传算法进行计算,实验上得到 θ 的值接近 4。

形状损失的惩罚标准可被描述为式(23)~(25):

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta \quad (23)$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \quad (24)$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (25)$$

IoU 损失也重新定义了惩罚标准,如式(26)所示:

$$L_{\text{box}} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (26)$$

其中, IoU 的计算如图 6 所示。推得的公式可写为:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (27)$$

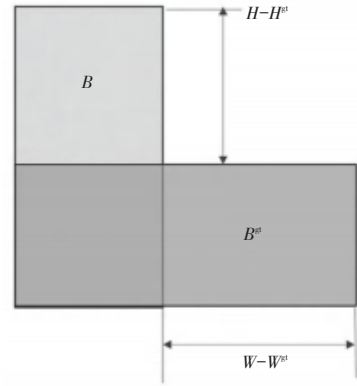


图 6 IoU 计算

Fig. 6 IoU calculation

最终的损失函数由 2 项组成,分别是:分类损失和框损失,可由式(28)表示为:

$$L = W_{\text{box}}L_{\text{box}} + W_{\text{cls}}L_{\text{cls}} \quad (28)$$

其中, L_{cls} 表示 Focal Loss, W_{box} , W_{cls} 分别表示框和分类损失权重。

3 实验对比与分析

3.1 数据集处理

3.1.1 数据集的采集

本课题主要针对铁路钢轨表面缺陷进行识别检测。数据集来自某铁路段,该数据集包括 3 种常见的钢轨表面缺陷,分别为:掉块 (spalling)、磨损 (fissure)、压溃 (crush)、剥落 (abrasion)、波磨 (corrugation)^[13]。训练集包括 1 000×3,测试集包括 100×3,共有 3 300 张图片。钢轨表面缺陷明显清

晰,对后续预处理有较大帮助。

3.1.2 数据集的标注

将采集到的钢轨表面缺陷数据集图片按照 YOLOv5m 算法模型的需求,使用 Labellmg 软件对

图片进行标注,缺陷类别分为 5 类,标签名称分别为 spalling(标签 0)、crush(标签 1)、abrasion(标签 2)、corrugation(标签 3)、fissure(标签 4)。图像标注情况如图 7 所示。



图 7 数据集的标注

Fig. 7 Annotation of the dataset

3.2 实验环境与评价指标

3.2.1 实验环境

本实验使用的操作系统为 Windows10, GPU 为 NVIDIA GeForce 2080Ti, 编译语言为 Python 3.9.12, 深度学习框架为 Pytorch 1.11.0、CUDA 版本为 11.6。batch_size 设置为 16, 训练次数为 400 次。训练结果显示最优权重文件以及最后一次权重文件, 采用测试集评价模型性能。

3.2.2 性能评价指标

本文实验中, 采用准确率(Precision), 召回率(Recall), 各类的平均精确 mAP 和每秒检测的图像帧数(FPS)作为评价指标来评估模型的实时性能。上述指标的计算方法公式为:

$$Precision = \frac{TP}{TP + FP} \quad (29)$$

$$Recall = \frac{TP}{TP + FN} \quad (30)$$

$$AP = \int_0^1 P(r) dr \quad (31)$$

$$mAP = \frac{\sum_i^n AP(i)}{N} \quad (32)$$

其中, AP 表示 PR 曲线所围的面积, N 表示类别个数。

3.3 实验结果

为验证本研究提出的对 YOLOv5m 的 2 种改进策略, 进行了消融实验, 以判断每个改进点是否有效, 依次在原有 YOLOv5m 模型基础上加上 CA 模块与 SIOU 损失函数, 数据集实验结果见表 1。

表 1 YOLOv5m 消融实验

Table 1 YOLOv5m ablation experiment

CA	SIOU	Precision/ %	Recall/ %	AP/ %
-	-	92.7	84.9	91.3
✓	-	93.8	86.6	92.4
-	✓	93.1	85.3	91.8
✓	✓	94.8	85.1	93.9

在网络中加入 CA 模块后, 召回率提高了 1.7%, 平均精度提高了 1.1%, 而在引入 SIOU 损失函数的提升表现比较弱。分析认为, CA 模块对位置信息进行编码, 不仅考虑了通道信息, 还考虑了方向相关的位置信息, 关注到许多易被淹没的语义信息^[14], 所以在检测精度上的提升较为明显, SIOU 提高了训练的速度和推理的准确性。

3.4 改进前后模型对比实验

不同模型训练结果对比如图 8 所示。从图 8 中可以看出, 本文所提出的改进算法的 mAP 比 YOLOv5m 提高了 2.6%, 准确率提高了 2.1%, 可以验证本文所提出的改进算法的有效性, 见表 1。

选择 YOLOv3、YOLOv4、YOLOv3-tiny 三种网络模型在相同的配置环境下利用相同的数据集, 进行对比实验。实验结果见表 2。

表 2 基于不同网络模型检测能力对比

Table 2 Comparison of detection capabilities based on different network models

模型	P	R	mAP@0.5	模型大小
Our	94.8	85.1	93.9	37.8
YOLOv3	86.3	77.8	84.4	234.0
YOLOv3-Tiny	91.9	82.6	87.9	33.1
YOLOv4	92.3	84.2	92.3	34.1

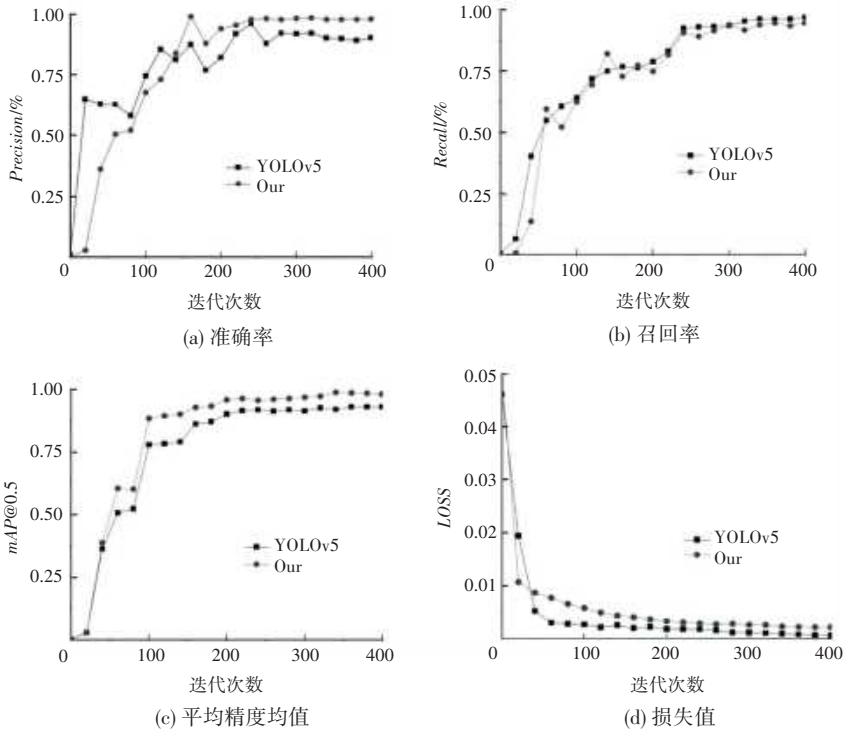


图 8 不同模型训练结果对比

Fig. 8 Comparison of training results of different models

3.5 改进前后模型精度对比实验

对原始 YOLOv5m 同时进行 CA 模块融合和 *SIoU* 损失函数改进, 实验结果见表 3, 可以看出同时引进 CA 和 *SIoU* 损失函数之后, 网络检测结果有了进一步提升, 均值平均精度 *mAP* 达到最高 0.939, 相较原始 YOLOv5m 提高了 2.6 个百分点。

在实际检测中, 与 YOLOv5m 算法进行检测对比, 效果对比如图 9 所示。对自建数据集的部分检测结果可视化, 可以看到原始网络对一些轨面缺陷

出现了漏检现象, 而本文提出的改进网络检测出的目标有明显改善。

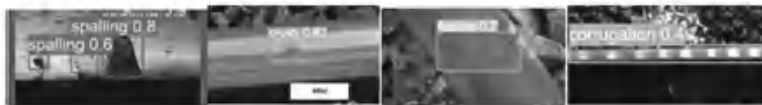
表 3 原始网络和改进网络结果对比

Table 3 Comparison of the results of the original network and the improved network

模型	Precision	Recall	MAP
YOLOv5m	92.7	84.9	91.3
Our	94.8	85.1	93.9



(a) YOLOv5m 模型识别效果



(b) 改进 YOLOv5m 模型识别效果

图 9 检测结果对比图

Fig. 9 Comparison of test results

4 结束语

本文使用改进的 YOLOv5m 网络检测钢轨表面

的缺陷目标, 将注意力机制 CA 模块与原始 YOLOv5m 主干网络融合, 提高模型网络对特征信息提取能力; 针对 *GIoU* 损失函数在有些情况下横向

与纵向收敛较为缓慢,以及两框出现包含时,退化为 IoU 损失,收敛依然困难的问题^[15],引入 $SIoU$ 损失函数,有效地提高了边界框定位精度,改善了漏检以及小目标缺陷识别效果差的问题。

对自建数据集进行训练、测试。试验结果表明,本文提出的改进方法比原始 YOLOv5m 在评价指标 mAP 上提高 2.6%,根据可视化结果,本文算法可以有效提高检测精度,并且降低漏检率。

参考文献

- [1] 刘佳朋,张银花,田常海,等. 我国高速铁路钢轨隐伤形成机理及维护策略[J]. 中国铁道科学, 2022,43(2): 40-47.
- [2] 田茂. 城市轨道交通设备系统建设一体化关键技术研究[D]. 北京:中国铁道科学研究院, 2019.
- [3] 张辉,宋雅男,王耀男,等. 钢轨缺陷无损检测与评估技术检测[J]. 仪器仪表学报, 2019,40(2): 11-25.
- [4] 胡文博,邱实,许馨月,等. 基于深度学习的钢轨伤损超声检测与分类[J]. 铁道学报, 2021,43(4): 108-116.
- [5] 张培培,王昭,王菲,等. 基于深度学习的图像目标检测算法研究[J]. 国外电子测量技术, 2020,39(8): 34-39.
- [6] 岳彪,闵永智,马宏锋,等. 钢轨表面缺陷检测系统中图像增强预处理方法研究[J]. 铁道科学与工程学报, 2018, 15(12): 3248-3256.
- [7] 闵永智,程天栋,马宏锋. 基于多特征融合与 AdaBoost 算法的轨面缺陷识别方法[J]. 铁道科学与工程, 2017, 14(12): 2554-2562.
- [8] 刘孟珂,吴洋,王逊. 基于卷积神经网络的轨道表面缺陷检测技术实现[J]. 现代计算机(专业版), 2017(29): 65-69, 77.
- [9] 李浪怡,刘强,邹一鸣,等. 基于改进 YOLOv5 算法的轨面缺陷检测[J]. 五邑大学学报(自然科学版), 2021, 35(3): 43-48, 54.
- [10] 张弘,刘保洋,高月. 基于改进自注意力神经网络的 X 光安检识别[J]. 激光杂志, 2023, 44(12): 47-55.
- [11] 孙建波,王丽杰,麻吉辉,等. 基于改进 YOLOv5s 算法的光伏组件故障检测[J]. 红外技术, 2023, 45(2): 202-208.
- [12] 李功,赵巍,刘鹏,等. 一种用于目标跟踪边界框回归的光滑 IOU 损失[J]. 自动化学报, 2023, 49(2): 288-306.
- [13] 中华人民共和国铁道部. 钢轨伤损分类: TB/T 1778-2020[S]. 北京:中国铁道出版社, 2010.
- [14] 白婷婷. 基于 MPSO-BP 神经网络的油层识别及储层参数预测研究[D]. 大庆:东北石油大学, 2022.
- [15] 赵璐璐,王学营,张翼,等. 基于 YOLOv5s 融合 SENet 的车辆目标检测技术研究[J]. 图学学报, 2022, 43(5): 776-782.