

DOI:10.19651/j.cnki.emt.2313079

基于改进 YOLOv5s 的道路障碍物检测算法

冉险生 李锐 贺帅

(重庆交通大学机电与车辆工程学院 重庆 400074)

摘要: 道路障碍物检测是自动驾驶环境感知的重要内容。针对当前道路障碍物检测算法精度有待提升等问题,提出改进 YOLOv5s 的道路障碍物检测算法。首先引入改进坐标注意力模块,过滤多尺度特征图的无效信息,强化关注感兴趣区域。其次使用增强降采样模块缓解融合网络下采样过程的重要信息丢失,增强特征鲁棒性。最后优化算法回归损失,明智的梯度增益分配策略,提升了普通质量锚框损失贡献度。试验结果显示,改进模型在数据集上的平均精度均值较原 YOLOv5s 提高了 4.2%,达到了 78.6%,同时也优于 Faster R-CNN、YOLOX、YOLOv7 等算法。所提算法具有 42 帧/s 的检测速度,满足实时检测的要求。本研究提出的改进算法能够有效提高道路障碍物检测精度,具有实际应用潜力。

关键词: 道路障碍物检测;YOLOv5s;注意力机制;增强降采样;WIoU

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.20

Road obstacle detection algorithm based on improved YOLOv5s

Ran Xiansheng Li Rui He Shuai

(School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: Road obstacle detection is a crucial component of automatic driving environment perception. To enhance the precision of existing road obstacle detection algorithms, we propose an improved YOLOv5s road obstacle detection algorithm. The improved coordinate attention module filters invalid information from multi-scale feature maps and strengthens the focus on areas of interest. Additionally, the enhanced downsampling module alleviates the loss of essential information during sampling in the fusion network, thereby increasing feature robustness. The optimized algorithm's regression loss and wise gradient gain allocation strategy improve the contribution of common mass anchor frame loss. Experimental results demonstrate that the improved model's average accuracy on the dataset has increased by 4.2% to 78.6%, outperforming Fast R-CNN, YOLOX, YOLOv7, and other algorithms. With a detection speed of 42 frames per second, the algorithm meets real-time detection requirements. Therefore, the proposed improved algorithm in this study can effectively improve the accuracy of road obstacle detection and has practical application potential.

Keywords: road obstacle detection;YOLOv5s;attention mechanism;enhanced downsampling;WIoU

0 引言

环境感知作为自动驾驶车辆核心技术的基石,为车辆后续的路径规划、智能控制提供先决信息^[1]。随着自动驾驶技术的快速发展,环境感知在确保行车安全和提高驾驶体验方面发挥着至关重要的作用。优秀的环境感知系统不仅要准确及时的探测路面的车辆、行人等可移动障碍物,对于道路上出现的道路坑洞、减速带、三角锥桶、警示柱等常见静态障碍物做到实时精准识别也尤为重要。

近年来,由于深度学习理论的飞速发展,基于卷积神经

网络的道路图像目标检测技术层出不穷。现阶段目标检测算法分为两种,即基于回归的单阶段检测方法和基于区域建议的两阶段检测方法^[2]。两阶段方法,包括 Faster R-CNN^[3], Cascade R-CNN^[4] 和 D2Det^[5] 等,首先通过区域建议网络生成若干区域候选框,寻找前景,再针对生成的锚框进行回归;然后,针对上一步生成的候选框进一步分类,并对初步回归的结果进一步精确位置回归。国内外学者已经将两阶段检测算法应用于道路目标检测方面。Kortmann 等^[6]为实现道路伤损的自动化检测,在 RDD-2020 数据集上训练了 Faster R-CNN 模型,实现了对道路裂缝和坑洞

的检测,并在数据集上获得了 0.487 的 F1 得分,但模型每秒只能处理 7 幅图像。陈鹏等^[7]为解决极端光照条件下的坑洞检测难题,提出了 IFNet 模型。该模型融合了改进的 Inception v1 分类网络和 Faster R-CNN 网络,在包含强、弱光道路坑洞图像的混合数据集上获得了 0.717 的平均准确率。同时,模型对不同光照条件下道路坑洞的检测具有较强鲁棒性。然而,此类两阶段算法以牺牲检测速度的方式获得较高的检测精度,往往难以满足实时性检测需要。通常适用于对检测速度要求不高,对精度较高的场景。

SSD^[8]和 YOLO^[9]等主流单阶段检测方法,将检测任务处理成回归任务,使用端到端的特征提取网络来提取目标特征,并直接用于边界框分类和回归。由于单阶段方法端到端同时完成分类和定位任务,因而比两阶段方法效率更高。Suong 等^[10]针对直接将 YOLOv2 用于道路坑洞的检测存在检测精度不高,推理时间长等问题。优化网络结构,降低网络大小,减少计算量。使网络获得了 22% 的精度提升,同时网络参数下降 3×10^7 ,最终网络检测速度也仅为 21 fps。胡均平等^[11]考虑 YOLOv3 检测模型在有积水、阴影等复杂环境下路面坑洼的检测能力有限。改进特征提取网络,优化激活函数,并引入多种数据增强训练策略,得到的 YOLOv3-pt 模型获得了 13.99% 的精度提升,使得模型在应对复杂环境下的检测能力显著提升。高翔^[12]针对道路病害检测难点,在 YOLOv5 网络加入 ECA 注意力模块对病害特征加强学习。同时引入 DIOU-NMS 解决大量先验框的冗余问题。所提的优化模型明显提高了坑洞在内的不同病害类别的检测精度。许洁^[13]为实现路面质量的在线评估,将裂缝、坑洼和凹陷作为相关类别,轻量化优化和改进 YOLOX,最终模型在移动端实现了路面质量的准确高效评估。随着单阶段算法的迭代更新,其检测速度愈发满足实时性需要。近年来,精度也不亚于两阶段方法。

综上,过往研究关于道路图像的目标检测算法主要集中于对道路坑洞及路面病害的检测,但在实际驾驶场景中,实时、高精度地检测道路静态障碍物仍然面临诸多挑战。为了解决这些问题,本研究在道路坑洞检测的基础上丰富算法感知目标,提出一种改进 YOLOv5s 的检测算法,以实现实时、高精度的检测驾驶场景中遇到的道路静态障碍物。因此,本文首先针对研究目标制作规范的数据集用于后期模型的训练与评估。其次,根据数据集目标特点,提出一种基于 YOLOv5s 的改进检测算法。具体:

1) 引入改进的注意力机制模块,过滤复杂的背景信息,使信息丰富的目标特征获得更多关注。

2) 在特征融合网络中引入增强降采样模块,改善模型下采样过程的重要信息丢失问题,增强网络对目标形状特征的适应性。

3) 优化损失函数加速模型收敛,以训练更好的检测器,提高模型对道路障碍物目标的检测能力。

1 YOLOv5 算法结构介绍

2020 年,YOLOv5 问世,首次提出通道与层的控制因子,根据应用灵活配置模型,依次有 n、s、m、l 和 x 版本实现不同的检测精度和速度,版本越高,检测精度越高,速度相应越慢。

YOLOv5 模型架构如图 1 所示,它由 4 个部分组成:输入端、骨干网络(Backbone)、特征融合网络(Neck)和检测头(Head)。在输入阶段,采用数据增强策略进行图像预处理,丰富训练数据,增强训练网络的鲁棒性。Backbone 完成对输入图片的特征提取。使用 CSP 模块对底层特征映射进行划分,通过跨级的层次结构进行特征的融合,优化了因梯度信息重复导致推理过度的计算问题,在保证精度的同时降低计算量。

Neck 实现特征信息的融合,采用特征金字塔^[14](feature pyramid network, FPN)和路径聚合^[15](path aggregation network, PAN)结构。在 CNN 中,深层特征映射承载较强的语义特征和较弱的位置信息,而承载较强的位置信息和较弱的语义特征在较浅的特征层上。FPN 将深层的强语义特征自下而上的传达到浅层,从而在多个尺度上增强了语义表达。PAN 相反的传达强定位信息到深层,从而增强多个尺度上的定位能力,提升网络的检测性能。

检测头部实现目标类别和位置信息的预测。CIoU 作为预测阶段的损失函数,考虑了边界框高宽比信息,利用该信息估计预测框与真实框之间的真实差距,加速模型收敛。

2 道路障碍物检测算法

YOLOv5s^[16]是 YOLOv5 系列检测器里检测精度、速度各个方面较适中的检测器,被广泛应用于工业界实现缺陷、伤损以及各类目标的检测。本研究同样以 YOLOv5s 作为模型基础,对模型做出改进,以提高算法对道路障碍物的检测性能。改进后的模型结构如图 1 所示。

2.1 改进注意力机制

近年来,注意力机制被证实在提高卷积神经网络性能上具有很大的前景,其通过在通道或空间维度上对输入图像的特征信息进行矫正,从而强化有用特征。

在 SENet^[17]中提出了 SE 注意力模块对特征图进行通道特征加强,但是得到的特征图完全没有关注像素空间上的关系。后来 CBAM^[18]注意力模块增加关注了空间关系,但这也只是一种局部上的空间关系,缺乏长距离空间关系关注的的能力。

CA^[19](coordinate attention)则是一种将空间位置信息编码到特征通道的注意力机制,其将通道注意力分解为沿宽度和高度两个空间方向的一维特征编码再聚合的过程。以此方式,在沿一个空间方向编码像素间的长程依赖关系信息时,另一个向上像素间的准确位置信息得到保留。然

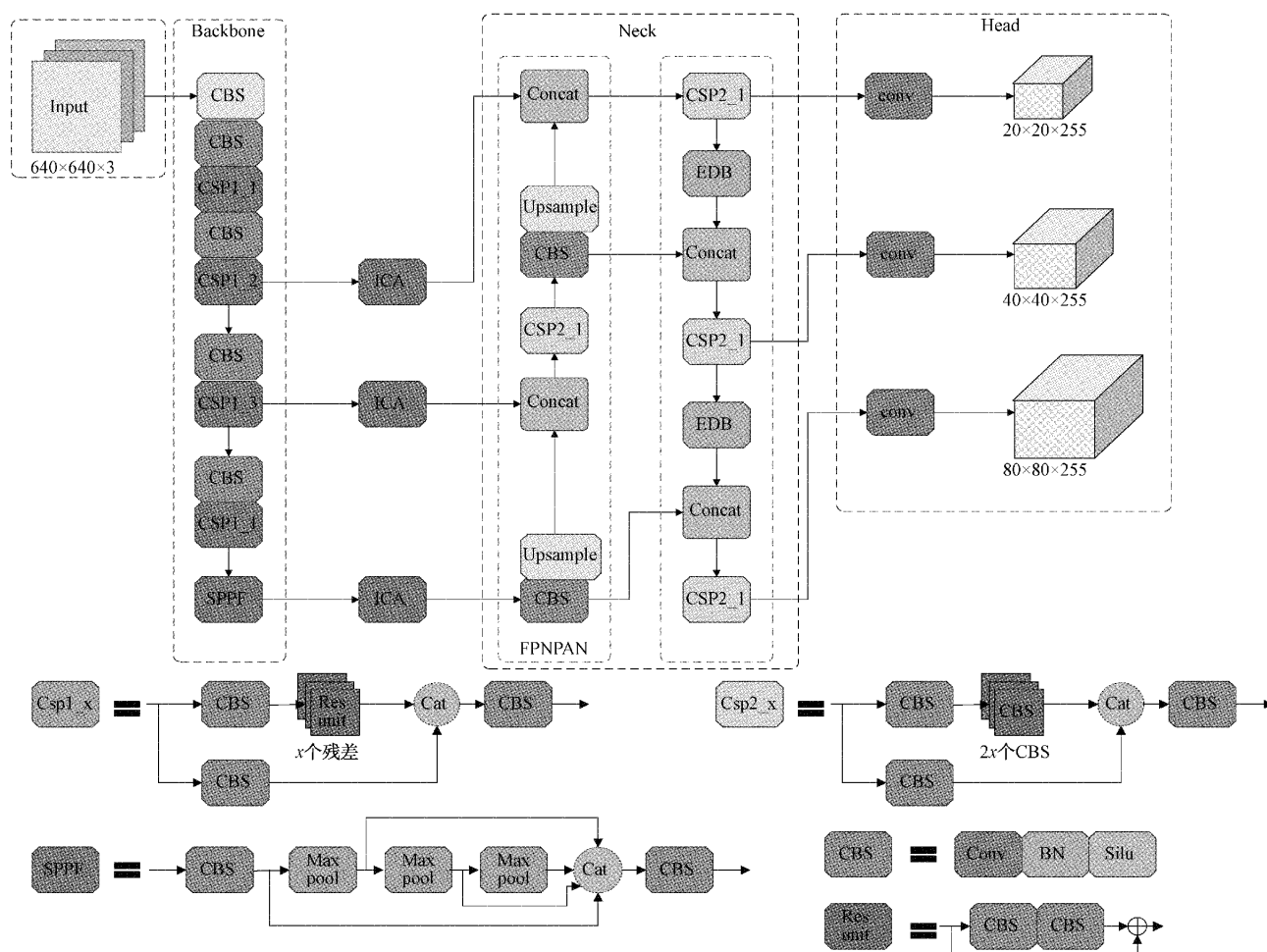


图1 改进后YOLOv5s网络结构

后所生成的对方向、位置敏感的注意力映射被互补地作用于输入特征图,以强化感兴趣目标的表示。

对于任意的特征张量可以表示为:

$$X = [x_1, x_2, x_3, \dots, x_c] \in R^{H \times W \times C} \quad (1)$$

输出同样尺寸的张量:

$$Y = [y_1, y_2, y_3, \dots, y_c] \in R^{H \times W \times C} \quad (2)$$

CA 为了能够捕获具有精准位置信息的远程空间交互,没有简单使用全局平均池化,而是对其进行分解得到 $z_c^h(h)$ 和 $z_c^w(w)$ 两个特征图,即:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (3)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (4)$$

将获得的宽度特征图 $z_c^h(h)$ 和高度特征图 $z_c^w(w)$ 进行拼接后送入 1×1 卷积核,使其通道降维为 C/r ,再 *Sigmoid* 激活函数激活得到特征图 f ,即:

$$f = \delta(F_1([z^h, z^w])) \quad (5)$$

然后沿空间维度,再将 f 进行 Split 操作划分成 f^h 和 f^w ,然后分别利用 1×1 卷积进行通道升维操作,再

Sigmoid 激活得到高度及宽度方向上的注意力向量 g^h 和 g^w ,即:

$$g^h = \sigma(F_h(f^h)) \quad (6)$$

$$g^w = \sigma(F_w(f^w)) \quad (7)$$

最后在原始特征图上分别乘上宽度和高度上的注意力向量,公式如下。

$$y_c(i, j) = x_c(i, j) \times g_c^w(i) \times g_c^h(j) \quad (8)$$

CA 注意力机制具有捕获位置信息以及不同通道之间关系的能力。为了进一步区分空间上不同像素之间的权重关系,强化感兴趣区域权重。基于此本文提出了一种新的改进坐标注意力(improve coordinate attention, ICA),在 CA 的基础上借鉴空间注意力模块^[18]思想,结构如图 2 所示。具体地,在通道维度上执行全局平均池化(global average pooling, GAP)以获得通道压缩的特征图 F_{avg}^c ,接着使用 7×7 卷积核细化空间局部特征,最后空间注意力权重 m_s 由 *Sigmoid* 激活得到,即:

$$m_s = \sigma(Con v^{7 \times 7}(F_{avg}^c)) \quad (9)$$

式中: σ 代表 *Sigmoid* 激活函数; $Con v^{7 \times 7}$ 代表 7×7 卷积。

由此,ICA 注意力机制的空间与坐标融合注意力权重 w_s 为:

$$w_s = x_i \times y_c \times m_s \quad (10)$$

最后,对于任意特征图 F ,经过 ICA 注意力模块后的输出 F_s 可表示为:

$$F_s = F \times w_s \quad (11)$$

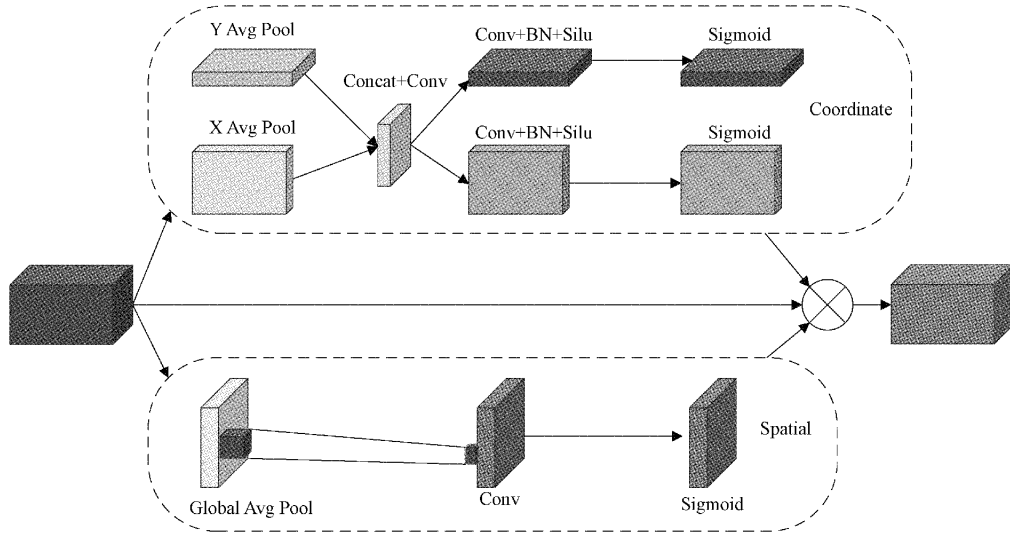


图 2 改进坐标注意力模块结构

2.2 增强降采样模块

YOLOv5s 算法颈部采用 FPN 和 PAN 网络结构,将来自主干的不同尺度特征层进行语义及位置信息的充分融合,进而提高特征表达力。PAN 自上往下传达强定位特征,需要通过两次降采样缩小特征图来匹配来自 FPN 的输入。常用的下采样模块是使用步长为 2 的 3×3 卷积或者步长为 2 的 2×2 的最大池化采样,旨在将特征图尺寸大小减半,通道数维持不变,便于与下一特征层堆叠融合。通常,下采样是隔点采样,这会导致一部分重要信息丢失。为了尽量获得下采样更多的重要信息,本文使用了一个新的增强降采样模块(enhanced down-sampling block, EDB)进行下采样,如图 3 所示,结合了两种常用的下采样方法。

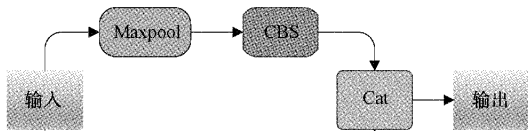


图 3 增强降采样模块结构

EDB 模块有两个分支,上分支使用了一个步长为 2×2 的最大池化和一个 1×1 卷积^[20],下分支分别使用了卷积核大小为 1×1 卷积和一个卷积核大小为 3×3 、步长为 2 的可变形卷积^[21](deformable convolution, Dconv2),完成之后将两个分支的结果堆叠输出。Dconv2 在感受野中引入可学习偏移量以及采样点加权,改善传统卷积存在感受野与目标外形轮廓不匹配以及规则点采样难以适应目标几何形变的问题,使得特征融合网络对目标几何形变的适应能力提升。

2.3 改进损失函数

本文算法训练的损失函数主要由三部分组成:置信度损失、分类损失和定位损失。YOLOv5s 则采用了 CIoU^[22] 作为定位损失函数对 Bbox 的中心点以及宽高进行回归。

目标检测作为计算机视觉的核心问题之一,其检测性能高度依赖损失函数的设计。定位损失函数的良好的定义将给目标检测模型带来显著的性能提升。文献[23]基于 IoU 提出了 Wise-IoU(WIoU),采用动态非单调的聚焦机制并提供了明智的梯度增益分配策略。本文将引入模型替换原有的定位损失,使模型在训练时降低高质量锚框的竞争力,同时也减小低质量示例产生的有害梯度,使得定位损失可以聚焦于普通质量的锚框,并提高检测器的整体性能。算法对 Bbox 的回归损失的定义过程如下:

1) 记预测框为 $\vec{B} = [x, y, w, h]$, 目标框为 $\vec{B}_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]$, 示意如图 4 所示。

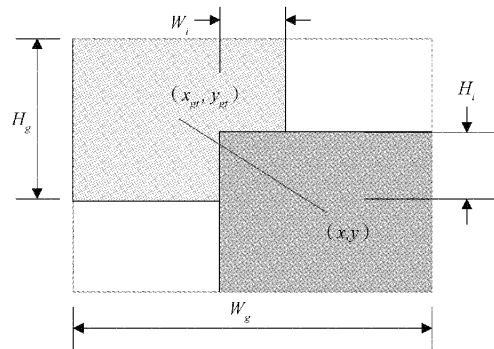


图 4 重叠度计算

IoU 用于度量目标检测任务中预测框与目标框之间的

重叠程度,定义为:

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{S_u} \quad (12)$$

$$S_u = w_h + w_{gt} h_{gt} - W_i H_i \quad (13)$$

式中: S_u 为预测框与目标框重叠面积。

2)为了在锚框与目标框重合较好时削弱几何度量的惩罚,不过多地干预训练,使模型有更好的泛化能力。因此,根据距离度量构建了距离注意力,得到了具有两层注意力机制的WIoU v1:

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (14)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (15)$$

式中:惩罚因子 $R_{WIoU} \in [1, e]$ 这将显著放大普通质量锚框的 L_{IoU} , $L_{IoU} \in [0, 1]$ 这将显著降低高质量锚框对梯度的贡献,并显著降低其对中心点距离的关注。

3)为描述预测框的质量定义了离群度 β ,定义为:

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \quad (16)$$

然后利用 β 构造非单调聚焦系数 r ,并将其应用于WIoU v1,得到最终的损失函数 L_{WIoU} ,即:

$$L_{WIoU} = r L_{WIoUv1} \quad (17)$$

$$r = \frac{\beta}{\delta \alpha^{\beta \delta}} \quad (18)$$

式中: $\overline{L_{IoU}}$ 为动量为 m 的滑动平均值, α, δ 是两个超参数,分别设置为1.9、3。 r 是动态非单调的聚焦系数,实现损失动态明智的梯度增益分配。

3 实验及结果分析

3.1 数据集与评估指标

现有关于道路目标检测的数据集主要集中在对车辆和行人目标的检测,针对本研究能直接使用的数据集几乎没有,因此采用自制数据集用于模型的训练与评估。RDD2020^[24]是全球道路缺陷检测挑战赛构建的官方数据集,数据样本包含大量道路坑洞及裂纹在内的道路图像。实验对其进行数据清理,挑选出具有道路坑洞的数据作为自制数据集的一部分。针对常见的道路减速带、三角锥桶、警示柱等障碍物目标,通过网络爬虫、线下拍摄等手段获得一定数量数据。最终构建了一包含2 512张图片的数据集,进行规范化标注后,自建数据集中共含有目标6 964个,各个类别实例数依次为2 863、975、1 386、1 740。以8:2的比例将数据集划分为训练集和测试集。

为验证本文提出的各项方法有效性,运用自建数据集进行训练和测试。实验采用mAP作为评价指标,mAP是所有类别的平均精度之和,然后除以类别数的均值。通过计算mAP的高低来衡量模型算法的有效性,越高的mAP值代表越好的检测性能,模型越好。计算公式为:

$$mAP = \frac{1}{k} \sum_{j=1}^k AP_j \quad (19)$$

式中: k 为类别数, AP_j 为第 j 类目标的平均精度。

AP由精准率P和召回率R共同决定的。精确率是指分类器在更可靠的情况下将样本预测为正样本,反映的是模型对负样本的分辨能力。召回率旨在找到实际为正例的样本中多少被预测为正例,反映模型对于正例样本的分辨能力,其值越高,则模型对正样本的区分能力越好。各自的计算公式为:

$$AP = \int_0^1 P(R) dR \quad (20)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (21)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (22)$$

式中: TP 、 FP 、 FN 分别代表正确识别出类别的数量,错误的识别成其他类别的数量以及未被识别出类别的数量。

3.2 实验环境与参数设置

具体实验环境如表1所示。

表1 实验环境

参数	实验环境
操作系统	Windows 11
GPU	GeForce RTX 3050Ti(4 G)
CPU	AMD R7 5800H @3.20 GHz
Python	3.9
深度学习框架	pytorch1.12.1、CUDA11.3

实验中采用SGD优化器训练300epoch, batchsize为4,初始学习率为0.01,采用自适应策略调整学习率,动量大小和权重衰减系数分别为0.937和0.0005。

3.3 实验结果对比分析

本文采用精准率P和召回率R以及IoU阈值为0.5时的mAP作为评估指标。在相同实验环境、参数设置、数据集下,使用改进的模块分别进行实验,YOLOv5s作为基础网络,最后同时添加所有使用的方法,实验结果如表2。使用ICA注意力作用在主干与颈部网络的连接部分,P、R、mAP提高了2.1%、2.9%、2.3%。从各个类别的mAP指标来看,改进对于坑洞的检测效果基本持平,在其他3种类别上均有不同程度的提高,说明改进的注意力对于道路图像的复杂背景信息起到了一定的过滤效果。通过引入增强降采样EDB模块,P、R、mAP分别提高了2.8%、5.6%、3.9%,各类别的mAP均有提高,尤其对于坑洞的检测首次有1.3%的涨点,证明EDB模块在颈部路径聚合网络中能够改善原有的采样方式对信息的丢失问题,增强网络对不规则形状特征的适应能力。使用WIoU损失替换原有的边界框回归损失,使得P降低了2%,但R和mAP这两项指标分别提高了5.2%、2.6%,说明改进加强了模型对于正样本的区分能力,总的检测性能得到

改善。最后,将 ICA、EDB、WIoU 融合改进原始模型,改进后的模型参数增加 0.95×10^6 , P、R 分别提高 2.2%、3.7%, mAP 实现了 4.2% 的提升,超越所有单点改进。

由此表明,加入不同的方法对道路障碍物检测都有一定的效果,同时加入所有方法最优,模型参数量仅略微增长。

表 2 消融实验结果对比

模型	ICA	EDB	WIoU	P/%	R/%	mAP/%	mAP/%				参数量/ ($\times 10^6$)
							c	p	s	w	
YOLOv5s	×	×	×	80.2	66.0	74.4	80.2	65.7	76.8	75.1	7.03
	√			82.3	68.9	76.7	81.9	65.5	79.4	80.2	7.18
		√		83.0	71.6	78.3	83.1	67.0	83.9	79.3	7.83
			√	78.2	71.2	77.0	83.2	65.1	79.9	79.9	7.03
YOLOv5s(本文)	√	√	√	82.4	69.7	78.6	84.4	66.2	84.0	79.9	7.98

注:加粗数据为最优值,c、p、s、w 分别代表目标类别三角锥桶、坑洞、减速带、警示柱。

为验证本文改进的 ICA 的优势,分别使用 SE、CBAM、CA 和本文的 ICA 进行对比,实验结果见表 3。结果表明,算法添加 ICA 比添加 SE、CBAM、CA 参数大致多

0.11×10^6 ,较基础 YOLOv5s 增加参数 0.15×10^6 。同时 P、R、mAP 分别达到了 82.3%、68.9%、76.7%,高于所有注意力机制,较基础模型分别提升 2.1%、2.9%、2.3%。

表 3 YOLOv5s 加入不同注意力对比

YOLOv5s	参数量/ ($\times 10^6$)	mAP/%				P/%	R/%	mAP/%
		c	p	s	w			
+SE	7.07	83.1	66.3	76.8	78.1	80.2	67.8	76.1
+CBAM	7.07	81.1	64.7	78.7	75.3	82.3	66.0	75.0
+CA	7.06	83.2	66.4	73.3	79.2	81.2	67.2	75.5
+ICA(本文)	7.18	81.9	65.5	79.4	80.2	82.3	68.9	76.7

注:加粗数据为最优值,c、p、s、w 分别代表目标类别三角锥桶、坑洞、减速带、警示柱。

为了验证改进算法性能优势,使用相同的数据集在不同的网络下训练,选用了经典的两阶段网络 Faster R-CNN,及 anchor free 网络 FCOS、YOLOX,单阶段网络 YOLO 系列的 YOLOv4、YOLOv5、YOLOv7 作对比实验,实验结果见表 4。实验结果表明,YOLOv5s 网络 mAP 优于 Faster R-CNN、YOLOv4、FCOS、YOLOX,低于最新模型 YOLOv7。改进后的算法 mAP 较 YOLOv7 和原模型分别有 2.9%、4.2% 的提升,改进后的 YOLOv5s 网络更具优势。同时改进的算法检测速度为 42 fps,满足实时性检测需要。

为了验证本文改进的 ICA 注意力的作用,使用 Grad-CAM^[25]方法,对 YOLOv5s 及不同注意力输出的热力图对比,结果如图 5 所示。通过可视化结果表明,YOLOv5s 网络在检测目标时对目标的关注不是特别明显。对于图像中同时存在多个目标时,在目标及附近出现成片高热,对道路图像中的树木及建筑也有关注,从而干扰检测,造成漏检和误检。在引入 SE、CBAM、CA 等注意力后,模型对于背景的关注有所减弱,对于路面目标的位置捕捉更好,热力图高热区域更多往目标附近聚集。但对于密集小目标依然存在定位不准的情况。可以发现改进的 ICA 注意力对目标位置定位更加准确,热力图覆盖更加明显,进一步区分了空间领域不同像素之间的权重关系,更多地关注感兴趣区域,减小次要区域的权重。本研究通过热力图可视化证明了 ICA 的有效性。

表 4 不同网络性能对比

模型	骨干网络	mAP/%	检测速度/fps
Faster R-CNN	ResNet50	65.1	8
YOLOv4	CSPDarknet53	68.9	43
FOCS	ResNet50	70.3	12
YOLOX	Darknet53	74.2	46
YOLOv5s	CSPDarknet53	74.4	69
YOLOv7	ELAN	75.7	71
YOLOv5s(本文)	CSPDarknet53	78.6	42

注:加粗数据为最优值。

为了更加直观地体现改进网络的涨点,本研究对检测结果做了可视化分析。如图 6 所示,对比 YOLOv5s 和改进后网络的结果,可以发现改进后的网络明显改善了漏检和误检情况的发生,对目标的标注也更加精确。

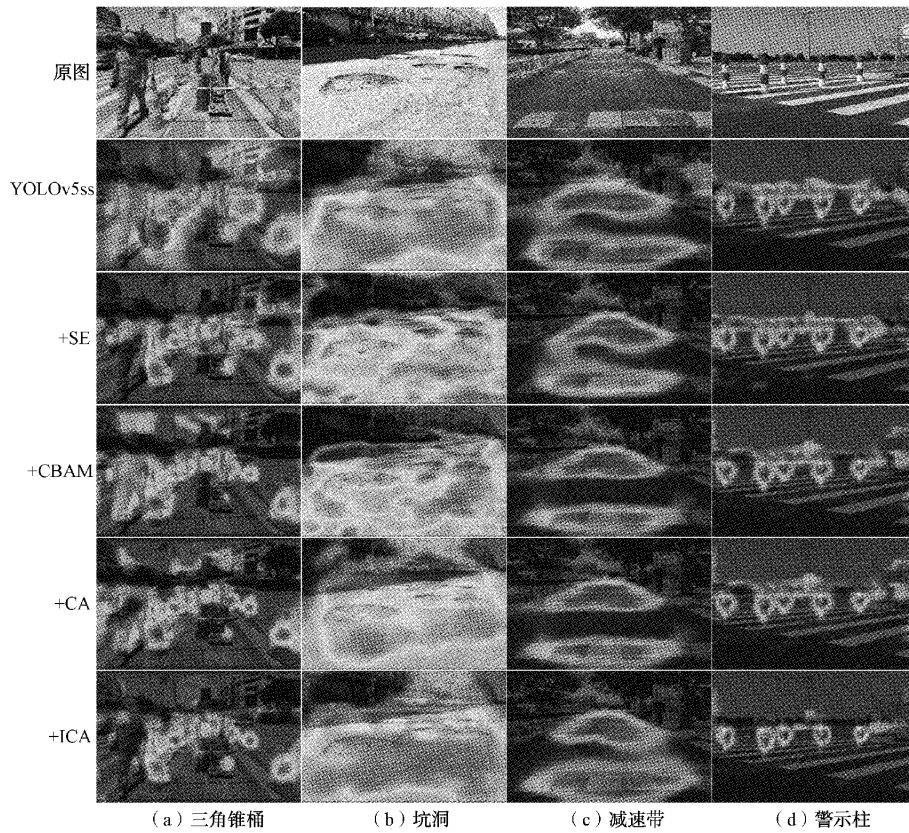


图5 不同注意力热力图结果对比

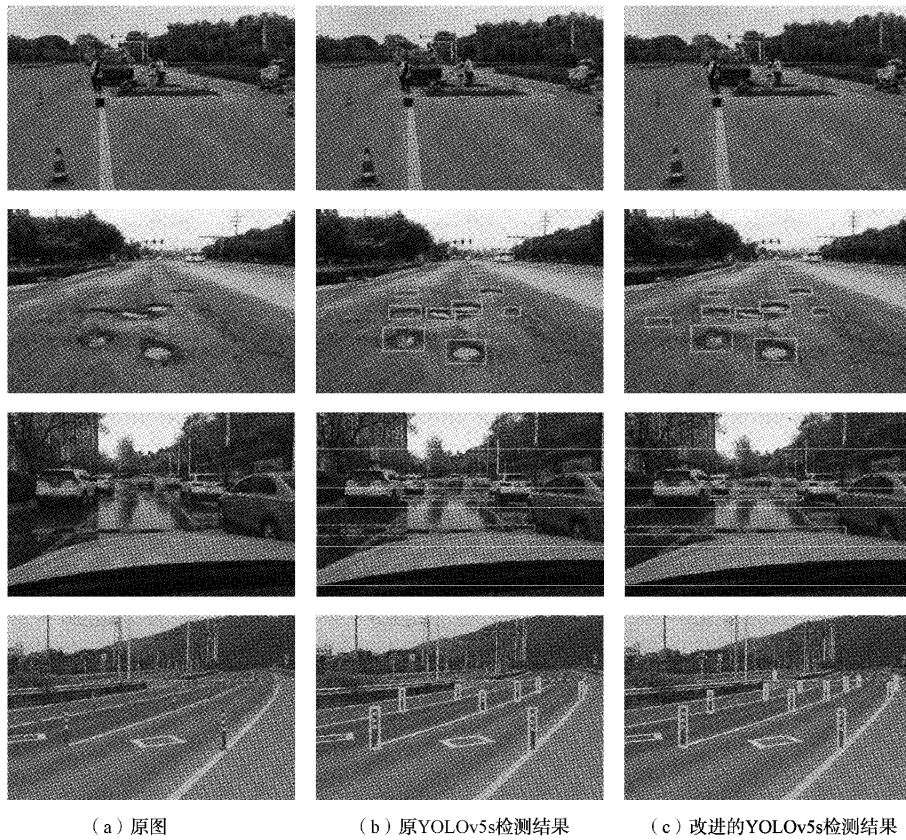


图6 算法改进前后检测结果对比

4 结 论

对于驾驶场景常遇到的道路障碍物目标,本文自制数据集用于模型的训练与测试。为了解决单纯利用 YOLOv5s 进行检测,漏检、误检率高,检测准确度低下等问题,提出了一种改进的 YOLOv5s 道路障碍物检测算法,提高了目标检测准确率,改善了漏检和误检问题。基于 YOLOv5s 模型,提出了 ICA 模块,将其融合到主干网络和颈部网络之间的 3 个前向通道中。网络更加关注路面上的目标位置,有效缓解了背景干扰问题。使用 EDB 模块对特征图下采样,获得了更多有效特征,提高了检测精度。通过 WIoU 实现边界框损失动态明智的梯度增益分配,加强普通质量的预测框对损失的贡献,最终提高检测器的整体性能。

本文的方法通过改进的注意力机制和增强降采样,强化有效特征,对道路障碍物精确检测, mAP 达到了 78.6%,相对于 YOLOv5s 提高了 4.2%。但在实验过程中发现,当目标存在重叠时,会出现漏检和误检,导致模型的检测效果不好。下一步将继续结合改进方法,提高模型在复杂背景下抗干扰的能力,进一步提高检测精度。

参考文献

- [1] 张新钰,高洪波,赵建辉,等. 基于深度学习的自动驾驶技术综述[J]. 清华大学学报(自然科学版), 2018, 58(4):438-444.
- [2] 吕辉,董帆. 基于 YOLOv4 的复杂交通状况下多目标检测算法[J]. 国外电子测量技术, 2022, 41(12): 41-47.
- [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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [4] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [5] CAO J, CHOLAKKAL H, RAO M A, et al. D2Det: Towards high quality object detection and instance segmentation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11482-11491.
- [6] KORTMANN F, TALITS K, FASSMEYER P, et al. Detecting various road damage types in global countries utilizing faster R-CNN [C]. IEEE International Conference on Big Data(Big Data), 2020, 5563-5571.
- [7] 陈鹏,应骏. 基于卷积神经网络的多场景道路坑洼图像检测[J]. 上海师范大学学报(自然科学版), 2020, 49(1):96-101.
- [8] BERG A C, FU C Y, SZEGEDY C, et al. SSD: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision. Amsterdam: Springer International Publishing, 2016: 21-37.
- [9] 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.
- [10] SUONG L K, JANGWOO K. Detection of potholes using a deep convolutional neural network[J]. Journal of Universal Computer Science, 2018, 24(9): 1244-1257.
- [11] 胡均平,黄强,张洪伟,等. 基于改进 YOLOv3 模型的车辆前方路面坑洼检测[J]. 传感器与微系统, 2022, 41(12):130-133,138.
- [12] 高翔. 基于深度学习的路面病害检测与分类识别研究[D]. 宁夏:宁夏大学, 2022.
- [13] 许洁. 基于深度学习的路面质量评估的研究和实现[D]. 成都:电子科技大学, 2022.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and pattern recognition, 2017: 2117-2125.
- [15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [16] 张宏群,班勇苗,郭玲玲,等. 基于 YOLOv5 的遥感图像舰船的检测方法[J]. 电子测量技术, 2021, 44(8): 87-92.
- [17] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(8):2011-2023.
- [18] WOO S, PARK J, LEE J Y et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [19] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 13708-13717.
- [20] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2261-2269.
- [21] ZHU X, HU H, LIN S, et al. Deformable ConvNets

- V2: More deformable, better results[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020: 9300-9308.
- [22] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021: 8574-8586.
- [23] TONG, Z J. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv Preprint, 2023, ArXiv:2301.10051.
- [24] ARYA D, MAEDA H, GHOSH S K, et al. RDD2020: An annotated image dataset for automatic road damage detection using deep learning[J]. Data in Brief, 2021, 36(1): 107133.
- [25] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. IEEE International Conference on Computer Vision, 2017: 618-626.

作者简介

冉险生, 副教授, 研究生导师, 主要从事摩托车动力学、智能汽车技术研究。

E-mail: cqrxs@qq.com

李锐, 硕士研究生, 主要从事机器视觉、图像处理技术研究。

E-mail: 2362713892@qq.com

贺帅, 硕士研究生, 主要从事深度学习、图像处理技术研究。

E-mail: 194576646@qq.com