

doi:10.12068/j.issn.1005-3026.2025.20230297

基于目标检测的复杂城市交通环境感知 技术及应用

艾散·西尔艾力¹, 车德福¹, 王 夺², 喻 甜³

(1. 东北大学 资源与土木工程学院, 辽宁 沈阳 110819; 2. 中冶沈勘工程技术有限公司, 辽宁 沈阳 110167;

3. 飞翼股份有限公司, 湖南 长沙 410600)

摘 要: 基于机器视觉的环境感知技术是智慧交通领域的关键任务之一。传统深度学习算法通常只能满足单一场景下的个别目标检测任务,难以应对复杂交通环境下的智能感知需求。为提高车辆在复杂环境下的智能感知能力,提出了一种改进的YOLOv8目标检测网络模型,结合注意力机制、优化器和可变形卷积层,实现了在复杂城市交通环境下的多目标检测。采用YOLOv4, YOLOv8及改进的YOLOv8算法对复杂交通环境样本图进行目标检测对比实验。结果表明,与YOLOv4, YOLOv8相比,改进的YOLOv8算法的平均精度分别提高了40.76%和16.92%。该算法的检测准确性与实时性满足实际应用需求,可通过多传感器信息融合,实现在复杂城市交通环境下的智能感知。

关键词: YOLOv8; 目标检测; 复杂城市交通; 环境感知; 智慧交通

中图分类号: P 23; TP 391.41 文献标志码: A 文章编号: 1005-3026(2025)05-0029-08

Perception Technology and Application of Complex Urban Traffic Environment Based on Target Detection

XIERAILI Aisan¹, CHE De-fu¹, WANG Duo², YU Tian³

(1. School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China; 2. Shen Kan Engineering & Technology Corporation, MCC, Shenyang 110167, China; 3. Feny Corporation Limited, Changsha 410600, China. Corresponding author: CHE De-fu, E-mail: chedefu@mail.neu.edu.cn)

Abstract: Machine vision-based environmental perception technology is one of the key tasks in the field of intelligent transportation. Traditional deep learning algorithms typically meet the detection needs of individual targets in simple scenarios. However, they are not capable of addressing the intelligent perception requirements in complex traffic environment. To improve the intelligent perception capability of vehicles in such environment, this paper proposes an improved YOLOv8 object detection network model, integrating attention mechanisms, optimizers, and deformable convolutional layers to achieve multi-target detection in complex urban traffic environment. To verify the effectiveness of the algorithm, comparative experiment were conducted using YOLOv4, YOLOv8, and the improved YOLOv8 algorithm on sample images from complex traffic environments. The results show that, compared to YOLOv4 and YOLOv8, the improved YOLOv8 algorithm increased the average accuracy by 40.76% and 16.92%, respectively. The detection accuracy and real-time performance of the improved YOLOv8 algorithm meet the practical application requirements, and through multi-sensor information fusion, it can realize intelligent perception in complex urban traffic environment.

Key words: YOLOv8; target detection; complex urban traffic; environment perception; intelligent transportation

收稿日期: 2023-10-27

基金项目: 国家自然科学基金资助项目(41871310); 中央高校基本科研业务费专项资金资助项目(N17241004, N2201007).

作者简介: 艾散·西尔艾力(1997—),男,新疆尉犁人,东北大学硕士研究生; 车德福(1970—),男,山东海阳人,东北大学教授,博士生导师; 王 夺(1980—),男,辽宁营口人,中冶沈勘工程技术有限公司教授级高级工程师。

近年来,随着信息融合与人工智能技术的飞速发展,基于机器视觉的环境感知技术在促进智能交通发展的基础上,也加快了智慧城市的建设速度^[1-2].由于城市交通环境下的路况千变万化,出现的目标类型多种多样、目标之间容易相互遮挡、目标运动轨迹错综复杂,车辆在行驶过程中对周围如此复杂环境信息的感知能力相当受限,这些因素给智慧交通的发展和交通安全带来了极大的挑战.此外,根据 Aufrère 等^[3]、Montemerlo 等^[4]、Leonard 等^[5]对智能汽车行驶环境感知的研究,车辆在驾驶过程中需要有效感知道路环境信息,识别周围动静目标,如其他车辆、行人、小动物和障碍物等.此外,还需要全面获取周围交通状况以及车辆自身定位等重要信息.为此,汽车需要配备多种传感器,以获取复杂交通环境下的多源信息.最终,通过信息融合技术感知当前行驶环境,确保车辆能够应对复杂的交通状况,从而为安全驾驶和自主导航提供坚实保障^[6].

这些研究有效解决了车辆智能感知任务中存在的一些问题,但具有一定的局限性,在复杂城市交通环境下对目标的多样性、复杂性考虑不足.为了提升车辆在复杂交通环境下的感知能

力,本文利用移动设备摄像头、固定摄像头等多元传感器采集复杂交通环境下的图像数据,运用计算机视觉、机器学习等先进技术,从视频图像处理方法入手,依赖改进的 YOLOv8 目标检测网络模型,针对城市复杂交通环境下多目标检测、分类、动态跟踪及定位等环境感知技术中的关键问题展开深入研究.

1 YOLO 模型概述

YOLO(you only look once)算法是一种基于深度学习回归方法的目标检测与分类算法.自 Redmon 等^[7]提出初代模型以来,YOLO 算法不断更新迭代并提升其性能,广泛应用于各类计算机视觉任务.YOLOv4 通过对损失函数^[8-9]、主干网络(backbone)和颈网络(neck)部分的优化,大幅提高了检测速度和精度^[10-11].YOLOv5 则通过自动化适配训练数据集,进一步提高了模型的训练速度和精度^[12].YOLO 算法现已迭代至第 8 代(YOLOv8),推出了更多优化和新功能,进一步提高了性能和灵活性,其结构如图 1 所示.

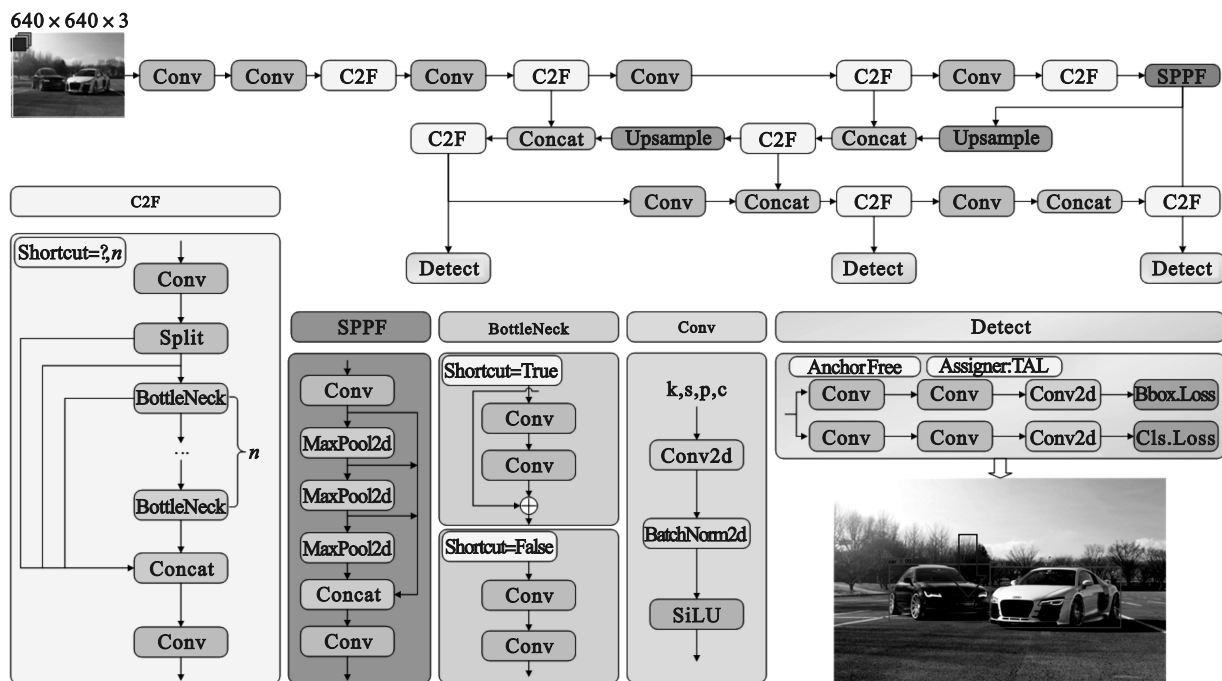


图 1 YOLOv8 网络结构
Fig. 1 YOLOv8 network structure

1.1 网络结构

YOLOv4 采用 CSPDarknet53 作为骨干网络,在 neck 部分引入了 SPP(spatial pyramid pooling)模块,并优化了检测过程中的损失函数与框筛选

策略^[13]; YOLOv8 的网络结构包括 C2F(cross stage partial with two fusion)特征融合模块与 SPPF(spatial pyramid pooling-fast)模块,进一步提升了特征提取和融合能力.Neck 部分同样采用

了 C2F 与 CBS(Conv+BN+SiLU)模块来优化来自不同层的特征图的融合,检测头(head)部分则通过解耦合头分别进行分类与回归操作^[14].

1.1.1 特征提取模块

本文改进的 YOLOv8 模型引入 C2F 模块(图 2),通过多分支特征融合与深度可分离卷积优化梯度流传递,结合上采样操作提升特征分辨率,在减少参数量的同时增强多尺度特征表达能力,从而提升检测速度与精度.

1.1.2 空间特征池化模块

相较于 SPP, SPPF 将简单的并行 Max pooling 改为串行加并行的方式,在提高检测速度的同时获取了更多的细节特征信息,其结构如图 3 所示.

1.1.3 检测模块

模型的检测模块主要由 neck 和 head 组成. backbone 和 neck 部分引入 C2F 和 SPPF 模块,并对不同尺度的模型调整了不同的通道数,在提

升模型性能的同时,平衡了模型的推理速度.而 head 部分则通过 Anchor-Free 方式进行类别预测与目标框回归,避免了 Anchor-based 方法的先验限制.

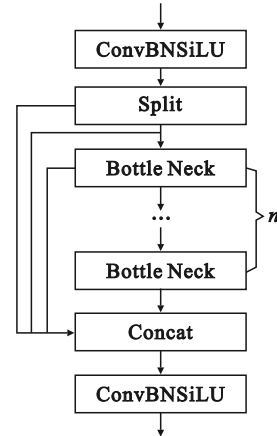


图 2 C2F 结构

Fig. 2 Structure of C2F

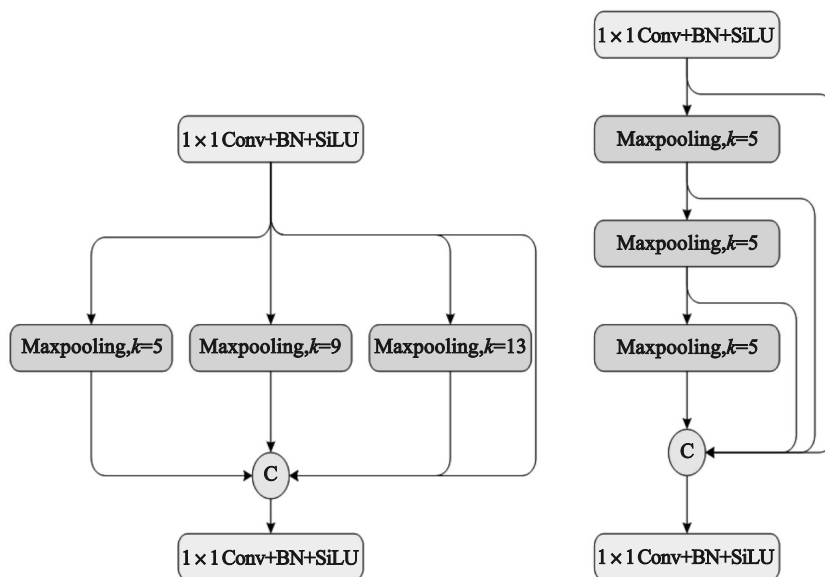


图 3 SPP 与 SPPF 结构

Fig. 3 Structure of SPP and SPPF

1.1.4 模型改进

本文提出了一种改进的 YOLOv8 模型,以更适应复杂交通环境下的智能感知技术要求.主要改进如下:

1) 引入全局多头自注意力机制.多头自注意力机制(multi-head self-attention, MHSA)是 Transformer 架构中的关键模块^[15].该机制由 n 个自注意力子模块构成,每个模块的线性变换矩阵 W_i^Q, W_i^K, W_i^V 分别与输入向量 X_i 相乘,以在多个空间维度上进行投影,从而提升模型的表达能力^[16],并生成相应的 Q (Query), K (Key), V (Value);接

着,所有模块的输出结果会被拼接在一起,与线性变换矩阵 W^o 相乘,最终得到自注意力的输出矩阵,如式(1),(2)所示.

$$\text{head}_i = \text{Attention}(X_i W_i^Q, X_i W_i^K, X_i W_i^V), \quad (1)$$

$$\text{MHSA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_i) W^o. \quad (2)$$

其中: $Q = XW_Q, K = XW_K, V = XW_V, Q, K$ 和 V 分别表示查询、键和值; X 是输入; W_Q, W_K 和 W_V 分别是 Q, K 和 V 的权重矩阵.

本文在 C2F 模块后面添加了 MHSA 机制,使得网络能够更加关注重要的特征区域. MHSA 允许模型共同关注来自不同表示子空间在不同位

置的信息.然而,单个注意力头(如缩放点积注意力头)在信息聚合时易将不同子空间信息平均化,限制了对丰富语义的捕捉.因此,MHSA可以更好地理解语义信息,可并行计算,降低计算复杂度^[17-18],更好学习远距离依赖.其结构如图4所示.

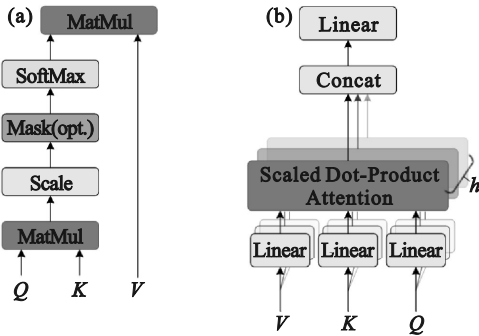


图4 缩放点积注意力与多头注意力机制结构

Fig. 4 Scaling point product attention and multi-head attention mechanism structure

(a)—缩放点积注意力;(b)—多头注意力.

2) 引入 Adam 优化器.优化器(optimizer)是引导神经网络更新参数的工具,深度学习在计算出损失函数之后,需要利用优化器来进行反向传播,以此完成网络参数的更新,找到最优的模型参数,使得损失函数最小化.卷积神经网络中常用的优化方法包括随机梯度下降法(stochastic gradient descent,SGD)^[19]、带有动量的随机梯度下降^[20-21]、AdaGrad (adaptive gradient)^[22]、RMSProp (root mean square propagation) 和 Adam (adaptive moment estimation)^[23] 优化器.本实验中,采用 Adam 优化算法替代迭代过程中存在一定随机性和震荡而导致准确度下降的 SGD 优化方法.

Adam是将 Momentum与 RMSProp融合于一身的算法,引入了 Momentum的一阶动量及 RMSProp的二阶动量,以累计梯度、加快收敛速度、缩小波动幅度.在此基础上增加了2个修正项,能够实现参数自动更新^[24].Adam 优化算法除了计算梯度平方 v_t 的指数衰减平均值,还计算梯度 m_t 的指数衰减平均值.计算公式如下:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (3)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \quad (4)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (5)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (6)$$

式中: β_1 和 β_2 分别是两个移动平均的衰减率,通常取值 $\beta_1=0.9$, $\beta_2=0.99$; g_t 为随机目标函数 t 时刻的梯度; \hat{m}_t 和 \hat{v}_t 分别是对 m_t 和 v_t 的校正,以此抵消偏差,尽量获取无偏估计.其梯度更新规则如下:

$$\theta_{t+1} = \theta_t - \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \cdot \eta. \quad (7)$$

其中: θ_t 为 t 时刻的模型参数; η 为学习率; ϵ 为常数项.

3) 采用可变形卷积.标准卷积的卷积核为固定的大小与形状,对于形状规则的物体可能会有更好的效果(图5a).可变形卷积在标准卷积规则的网状采样位置上引入可学习的偏移量,使得卷积核能够在特征图上自由变形,从而自适应地调整采样位置,更好地捕捉物体的非规则形状、尺度变化和旋转,显著提升了在处理复杂形状或变化较大的物体时的表现.从图5b~图5d中可以清晰地观察到它在尺度、纵横比和旋转变换等方面的能力.

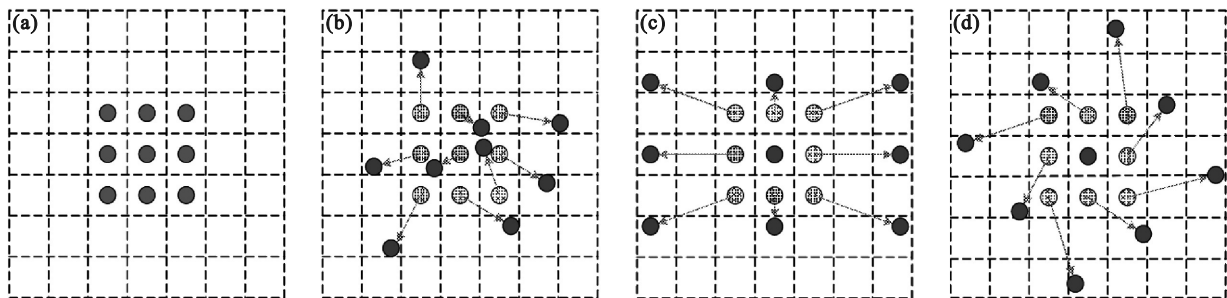


图5 标准卷积和可变形卷积的采样位置示意图

Fig. 5 Schematic diagram of sampling positions of standard convolution and deformable convolution

(a)—标准卷积规则采样网格;(b)—可变形卷积采样网格(带偏移量);
(c)—可变形卷积采样网格(尺度变换);(d)—可变形卷积(旋转变换).

从图6中可以看出,标准卷积中的感受野和采样位置都固定在顶部特征图上;在可变形卷积中根

据物体的尺度和形状进行自适应调整,使采样点更贴近物体的形状和尺寸,具有更强的鲁棒性^[25].

因此,本文在模型主干网络的后 3 个模块中增加了可变形卷积,使其可以根据实际情况调整自身的形状,更好地提取输入特征,提升对未知变化的适应性以及泛化能力.

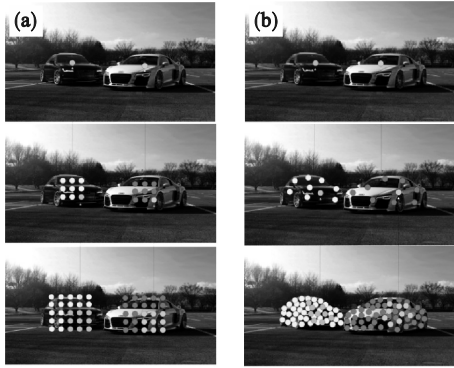


图 6 标准卷积与可变形卷积的卷积过程示例
Fig. 6 Examples of convolution process of standard convolution and deformable convolution
(a) —标准卷积; (b)—可变形卷积.

2 实 验

2.1 实验条件

本实验环境配置:操作系统为 Windows 10,处理器为 Intel Core i7-12700,内存为 32 GB,显卡为 NVIDIA RTX 3060 Ti,编程语言为 Python 3.7,计算机视觉库为 Python-OpenCV 3.4.8.

2.2 实验内容

2.2.1 制作数据集与预处理

本实验选用的城市交通环境图像由部分中国交通标志数据集 (Chinese traffic sign database, CTSDB) 以及个人采集的东北大学周边交通环境数据集构成,共 6 766 张交通环境图像.其中 72% (4 870 张) 为训练集,20% (1 354 张) 为测试集,8% (542 张) 为验证集.所有图像数据均采用 Labelimg 标注工具进行标注.本次实验收集到城市交通环境中的十几种目标类型,包括交通标志牌、交通信号灯、路面龟裂、路面破损、井盖、垃圾桶、路面垃圾、垃圾桶倾倒、雨水立算、雨水立算破损、井盖丢失、火焰、烟雾、违规摆摊.

2.2.2 数据增强

本文选用的是 Mosaic 数据增强方法,该方法是参考 CutMix 数据增强方法. CutMix 方法通过移除图像的部分区域,并用训练集中其他样本数据的像素值随机替代,而非填充 0 像素.分类结果则按照特定比例进行分配,如图 7 所示. Mosaic 方

法则是随机选择 4 张图,按一定比例取其部分,先分别对 4 张选取部分进行旋转、缩放、平移、错切、色域变化等数据增广操作,然后按照随机排列的方式进行拼接形成新图(见图 8). 图 9 中,4 种颜色代表 4 张样本原图,超出指定区域的部分将被舍弃. Mosaic 方法通过随机选取和组合,丰富了图像背景、增加了数据多样性;通过混合 4 张具有不同语义信息的图片,增强了模型鲁棒性;用 4 张图拼接生成 1 张新图,提升了小目标检测性能.



图 7 CutMix 方法示意图
Fig. 7 Schematic diagram of CutMix method

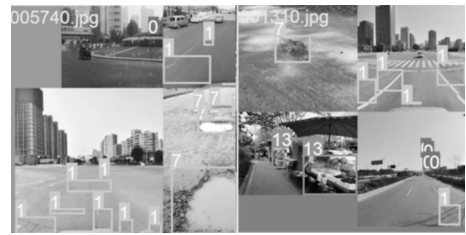


图 8 Mosaic 方法示意图
Fig. 8 Schematic diagram of Mosaic method

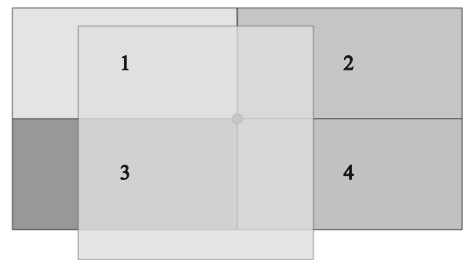


图 9 Mosaic 图像截取方式
Fig. 9 Image interception method of Mosaic

2.2.3 训练数据集

本实验训练了复杂城市交通环境下较为重要的十几种目标.训练参数设置主要包括:迭代次数(epoch)、输入图像大小(imgsize)、每组数据量(batch)、工作线程(workers)、优化器(optimizer)等.本实验对主要参数进行了设置和优化调整,参数设置见表 1.

表 1 训练参数设置
Table 1 Training parameter setting

训练参数	数值	训练参数	数值
epoch	100	imgsize	640
batch	16	workers	4
optimizer	Adam	patience	50
save_period	10	close_mosaic	10

2.3 评估指标指定

在复杂交通环境下基于目标检测的智能感知系统中,目标检测处于整个系统的前端,其精度直接影响后续操作的准确性,而检测速度则决定系统的实时响应能力.因此,本文将从多个维度,包括检测精度、模型规模及推理速度等方面,对改进的YOLOv8算法进行综合评估.

在目标检测中,交并比(intersection over union, IoU)通常用来衡量预测框与真实框(ground truth box, GTB)之间的匹配程度,其定义为目标的真实边界框与预测框交集与并集的比值^[26],取值范围在 $[0, 1]$ 之间.为了判断预测结果的正确性,目标检测任务中会设定一个交并比的阈值,只有高于阈值的预测结果被视为有效预测.其计算如式(8)所示.本文中IoU取值0.5,即当IoU值大于0.5时,则预测框视为有效,保留预测框;反之则视为无效,不保留预测框.

$$\text{IoU} = \frac{\text{Area}(\text{predict}) \cap \text{Area}(\text{label})}{\text{Area}(\text{predict}) \cup \text{Area}(\text{label})}. \quad (8)$$

引入平均精度均值 mAP (mean average precision)、每秒帧数 (frames per second, FPS) 和准确率-召回率 (precision-recall, $P-R$) 曲线作为

模型检测精度与实时性的评价指标,计算公式如下^[27].

$$P = \frac{\text{TP}}{(\text{TP} + \text{FP})} \times 100\%, \quad (9)$$

$$R = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100\%. \quad (10)$$

式中:TP表示真正例,是正确划分为正例的样本数量;FP表示假正例,是错误划分为正例的样本数量;FN表示假负例,是错误划分为负例的样本数量.

平均精度(average precision, AP)是对 $P-R$ 曲线上的precision值求均值,其值等于 $P-R$ 曲线与坐标轴围成的面积,代表各类别精度的平均值,最大值为1^[27].其计算公式如下:

$$\text{AP} = \int_0^1 p(r) dr. \quad (11)$$







由于精度与召回率始终介于0~1之间,因此,mAP也在0~1范围内,mAP数值越大说明检测效果越好^[26].

3 实验结果分析

3.1 实现成果

训练工作结束后,将已划分好的验证集输入到模型中,得到目标检测与目标类型预测的结果,如表2所示.可见,本文提出的算法能够准确检测出城市交通环境中的交通标志牌、路面状况、路边违规摆摊、火焰、烟雾等关键目标,为车辆精确感知周围环境提供有效信息.

表 2 实验结果
Table 2 Experimental results

测试图像	人工标记图像	本文模型检测结果
		
		

续表 2



3.2 实验评估

本文在相同实验环境配置条件下,进行了 YOLOv4、YOLOv8 及改进 YOLOv8 模型之间的对比实验.以上 3 种目标检测网络模型的性能对比结果如表 3 所示.

由表 3 可知,本文改进模型的检测性能提升效果明显,改进模型在保持较快检测速度的同时,对复杂交通环境中目标的检测效果更出色.

表 3 不同网络模型的性能对比
Table 3 Performance comparison of different network models

模型	主干网络	输入尺寸/像素	mAP/%	FPS	IoU
YOLOv4	CSPDarknet53	416×416	44.19	47	0.5
YOLOv8	CSPDarknet+C2F	640×640	53.20	87	0.5
改进 YOLOv8	CSPDarknet+C2F	640×640	62.20	78	0.5

YOLOv8 模型改进后,路面破损、井盖、井盖丢失、垃圾桶、垃圾桶倾倒、雨水立算、雨水立算破损、违规摆摊等目标的检测效果有明显提升.

实验中所涉及的 10 种目标检测精度均高于 YOLOv4 和 YOLOv8 模型.以上 3 个模型在同一数据集上的检测精度如表 4 所示.

表 4 实验结果对比
Table 4 Comparison of experimental results

目标类型	模型平均精度 mAP50			改进模型精度提升量	
	YOLOv4	YOLOv8	改进 YOLOv8	相对 YOLOv4	相对 YOLOv8
总平均	44.19	53.20	62.20	18.01	9.00
交通标志牌	55.59	68.40	68.30	12.71	-0.10
路面破损	57.60	73.60	81.10	23.50	7.50
井盖	50.95	47.50	51.10	0.15	3.60
井盖丢失	96.62	86.50	97.40	0.78	10.90
垃圾桶	53.95	71.80	75.80	21.85	4.00
垃圾桶倾倒	56.93	95.00	99.50	42.57	4.50
雨水立算	53.26	58.80	67.40	14.14	8.60
雨水立算破损	00.00	33.20	99.50	99.50	66.30
火焰	45.07	63.10	65.40	20.33	2.30
违规摆摊	44.19	47.00	54.70	10.51	7.70

4 结 语

本文提出了一种改进的 YOLOv8 目标检测模型,针对复杂交通环境的智能感知需求.该模型结合 MHSA 机制增强对重要特征的敏感度,采用 Adam 优化算法加速收敛并减少波动,同时通过可变卷积提高了模型的泛化能力和特征提取效果.实验结果表明,改进后的模型在多样化目标、远距离小目标、部分遮挡及复杂路况下优于对比模型,展示了良好的智能感知性能.未来研究将聚焦于提升检测精度,并探索激光雷达与摄像头数据的融合,进一步提升目标定位和细节识别能力,尤其在低精度目标的检测上取得突破.通过多传感器融合,模型可实现全息感知,为智能车辆在复杂交通环境下的自动驾驶提供有效辅助,增强其感知与决策能力.

参考文献:

- [1] 高德芝,段建民,郑榜贵,等.智能车辆环境感知传感器的应用现状[J].现代电子技术,2008(19):151-156.
(Gao De-zhi, Duan Jian-min, Zheng Bang-gui, et al. Application status of intelligent vehicle environmental sensing sensor [J]. *Modern Electronic Technology*, 2008 (19): 151-156.)
- [2] Wang K, Gou C, Zheng N, et al. Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives [J]. *Artificial Intelligence Review*, 2017, 48(3): 299-329.
- [3] Aufrère R, Gowdy J, Mertz C, et al. Perception for collision avoidance and autonomous driving [J]. *Mechatronics*, 2003, 13(10): 1149-1161.
- [4] Montemero M, Becker J, Bhat S, et al. Junior: the Stanford entry in the urban challenge [J]. *Journal of Field Robotics*, 2008, 25(9): 569-597.
- [5] Leonard J, How J, Teller S, et al. A perception-driven autonomous urban vehicle [J]. *Journal of Field Robotics*, 2008, 25(10): 727-774.
- [6] 谢志萍,雷莉萍.智能网联汽车环境感知技术的发展和研究现状[J].成都工业学院学报,2016,19(4):87-92.
(Xie Zhi-ping, Lei Li-ping. Development and research status of intelligent networked automotive environment awareness technology [J]. *Journal of Chengdu Technological University*, 2016, 19(4): 87-92.)
- [7] 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). Las Vegas, 2016: 779-788.
- [8] Asmaa B, Khalid Z. Optimizing CNN-BiGRU performance: mish activation and comparative analysis [J]. *International Journal of Computer Networks & Communications*, 2024, 16(3): 69-87.
- [9] Wang Y F, Hua C C, Ding W L, et al. Real-time detection of flame and smoke using an improved YOLOv4 network [J]. *Signal, Image and Video Processing*, 2022, 16(4): 1-8.
- [10] 张凯祥,朱明.基于YOLOv5的多任务自动驾驶环境感知算法[J].计算机系统应用,2022,31(9):226-232.
(Zhang Kai-xiang, Zhu Ming. Multi-task automatic driving environment perception algorithm based on YOLOv5 [J]. *Computer Systems & Applications*, 2022, 31(9): 226-232.)
- [11] Fei X, Li T H, Xiao Y G, et al. Research on YOLOv3 model compression strategy for UAV deployment [J]. *Cognitive Robotics*, 2024, 4: 8-18.
- [12] Guo K Y, Cheng B H, Min Y, et al. A pavement distresses identification method optimized for YOLOv5s [J]. *Scientific Reports*, 2022, 12(1): 1-15.
- [13] 郭振宇,高国飞.基于YOLO v4的复杂路口下人车混行检测算法研究[J].信息技术与信息化,2021(2):236-240.
(Guo Zhen-yu, Gao Guo-fei. Research on detection algorithm of mixed traffic between people and vehicles at complex intersections based on YOLO v4 [J]. *Information Technology and Informatization*, 2021(2): 236-240.)
- [14] Łysakowski M, Żywanowski K, Banaszczyk A, et al. Real-time onboard object detection for augmented reality: enhancing head-mounted display with YOLOv8 [C]//IEEE International Conference on Edge Computing and Communications. Chicago, 2023: 364-371.
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 6000-6010.
- [16] 李鸿,邹俊颖,谭茜成,等.面向医学图像分割的多注意力融合网络[J].计算机应用,2022,42(12):3891-3899.
(Li Hong, Zou Jun-ying, Tan Qian-cheng, et al. Multi-attention fusion network for medical image segmentation [J]. *Journal of Computer Applications*, 2022, 42(12): 3891-3899.)
- [17] Wang Z Y, Zhu H, Liu F. SMSTracker: a self-calibration multi-head self-attention transformer for visual object tracking [J]. *Computers, Materials & Continua*, 2024, 80(1): 605-623.
- [18] Vasanthi P, Mohan L. A reliable anchor regenerative-based transformer model for x-small and dense objects recognition [J]. *Neural Networks*, 2023, 165: 809-829.
- [19] Battiti R. First- and second-order methods for learning: between steepest descent and Newton's method [J]. *Neural Computation*, 2014, 4(2): 141-166.
- [20] Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning [C]// International Conference on Machine Learning (ICML). Atlanta, 2013: 1139-1147.
- [21] Qian N. On the momentum term in gradient descent learning algorithms [J]. *Neural Networks*, 1999, 12(1): 145-151.
- [22] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. *Journal of Machine Learning Research*, 2011, 12: 2121-2159.
- [23] Mohamed R, Amany M S. A modified Adam algorithm for deep neural network optimization [J]. *Neural Computing and Applications*, 2023, 35(23): 17095-17112.
- [24] Sarker I H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions [J]. *SN Computer Science*, 2021, 2(6): 420-420.
- [25] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks [C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 764-773.
- [26] 欧阳继红,王梓明,刘思光.改进多尺度特征的YOLO_v4目标检测方法[J].吉林大学学报(理学版),2022,60(6):1349-1355.
(Ouyang Ji-hong, Wang Zi-ming, Liu Si-guang. Improved multi-scale feature method for YOLO_v4 target detection [J]. *Journal of Jilin University (Science Edition)*, 2022, 60(6): 1349-1355.)
- [27] 魏东飞,熊峰,孔维畅.改进YOLOv4的轻量化目标检测方法[J].计量与测试技术,2022,49(11):18-22.
(Wei Dong-fei, Xiong Feng, Kong Wei-chang. Improved lightweight target detection method of YOLOv4 [J]. *Metrology & Measurement Technique*, 2022, 49(11): 18-22.)