

基于多尺度特征融合的车辆及行人目标检测算法*

李相衡¹ 方虹苏² 杨娅琳² 杨炜²

(1. 河南理工大学, 焦作 454000; 2. 长安大学, 西安 710061)

【摘要】针对道路交通环境复杂多样, 车辆及行人检测易出现错检及漏检的问题, 提出一种基于多尺度特征融合的车辆及行人目标检测算法 YOLOv8-RC。首先, 在基础网络 YOLOv8 的结构中引入 RCS-OSA 模块代替原有模块, 对所提取的特征信息进行增强及融合, 并引入轻量级上采样算子内容感知特征重组 (CARAFE) 代替原上采样算子, 提高网络对全局多尺度信息的融合能力。其次, 通过公开数据集及网络收集的方式构建了由 6 000 张车辆及行人目标图片构成的检测数据集, 并采用准确率、召回率、平均精度均值 mAP50 及 mAP50-95 对算法检测效果进行定量评价, 相比于 YOLOv8-N, YOLOv8-RC 的精确率提升 1.7 个百分点, 召回率提升 1.2 个百分点, mAP50 提升 0.9 个百分点, mAP50-95 提升 0.5 个百分点, 证明了算法的有效性。

关键词: 深度学习 目标检测 YOLOv8 行人检测

中图分类号: U491 **文献标志码:** A **DOI:** 10.20104/j.cnki.1674-6546.20240224

Vehicle and Pedestrian Target Detection Algorithm Based on Multi-Scale Feature Fusion

Li Xiangheng¹, Fang Hongsu², Yang Yalin², Yang Wei²

(1. Henan Polytechnic University, Jiaozuo 454000; 2. Chang'an University, Xi'an 710064)

【Abstract】In response to the complex and diverse nature of the road traffic environment, where vehicle and pedestrian detection is prone to false and missed detections, this paper proposes a vehicle and pedestrian target detection algorithm YOLOv8-RC based on multi-scale feature fusion. Initially, the RCS-OSA module is introduced within the structure of the base network YOLOv8 to replace the original module, thereby enhancing and integrating the extracted feature information. Additionally, a lightweight Context-Aware Adaptive Feature Reorganization (CARAFE) is employed to replace the original upsampling operator, enhancing the network's capability for global multi-scale information fusion. Subsequently, a detection dataset consisting of 6 000 images of vehicle and pedestrian targets is constructed through public datasets and network collection. The algorithm's detection performance is quantitatively evaluated using accuracy, recall rate, mean Average Precision at a 50% intersection over union threshold (mAP50), and mAP50-95. Compared to YOLOv8-N, YOLOv8-RC demonstrates an improvement of 1.7 percentage in accuracy, 1.2 percentage in recall rate, 0.9 percentage in mAP50, and 0.5 percentage in mAP50-95, thus validating the algorithm's effectiveness.

Key words: Deep learning, Target detection, YOLOv8, Pedestrian detection

【引用格式】李相衡, 方虹苏, 杨娅琳, 等. 基于多尺度特征融合的车辆及行人目标检测算法[J]. 汽车工程师, 2024(8): 1-7.

LI X H, FANG H S, YANG Y L, et al. Vehicle and Pedestrian Target Detection Algorithm Based on Multi-Scale Feature Fusion[J]. Automotive Engineer, 2024(8): 1-7.

1 前言

车辆及行人检测作为当前无人驾驶领域的核心技术之一, 旨在自动识别图像或视频中的车辆及

行人区域。目前, 道路车辆及行人检测算法可分为传统检测算法和基于深度学习的检测算法。传统检测算法在无遮挡物、车辆及行人目标清晰的道路上检测效果较好, 但难以应用于无人驾驶场景。

*基金项目: 中央高校基金项目(300102229112)。

深度学习算法应用于图像处理领域,可以实现高精度的车辆及行人目标检测,为辅助驾驶系统提供安全基础。Liu 等^[1]提出一种用于检测被遮挡行人的关键点和可视部分融合网络,引入人体关键点和可见部分包围盒构建2个注意力模块,分别抑制行人的通道特征和空间特征中的遮挡部分。Liang 等^[2]提出一种基于增强YOLOv5s的道路目标检测算法,引入加权增强极化自注意力增强特征表达,提高了算法对复杂场景的适应性。Jain 等^[3]使用集成学习模型和深度卷积神经网络的鲁棒多模态行人检测模型实现了精确的行人检测。Kiran 等^[4]使用改进YOLOv4进行了不同光照条件下运动车辆的识别。Andika 等^[5]改进了轻量级YOLOv7模型中的特征提取网络,实现了低成本硬件上的实时车辆检测。胡倩等^[6]基于改进YOLOv5算法实现了密集场景下的行人检测。黄丹丹^[7]基于Transformer实现了双分支检测和重识别的多行人追踪。李林红^[8]等提出一种改进的YOLOv5s模型,用于对站口行人的高效检测。周力等^[9]将基于旋转边界框(Rotated Bounding Box, RDB)的YOLOv5模型应用于遥感图像车辆检测领域。沈正坤等^[10]提出基于YOLOv5的车辆目标检测算法,提升了模糊图像中小目标的检测精度。

上述文献利用不同检测算法实现了车辆及行人的检测,但在实际检测中,仍然会出现各种错检及漏检问题。基于此,本文提出一种基于改进YOLOv8的车辆及行人目标检测算法YOLOv8-RC,首先引入多尺度特征提取模块RCS-OSA(RepVGG/

RepConv ShuffleNet based One-Shot Aggregation)加强网络的特征融合能力,然后将YOLOv8的上采样算子替换为轻量级的内容感知特征重组(Content-Aware ReAssembly of FEatures, CARAFE)模块,从而在网络上采样的过程中再次提高其特征融合能力。最后,构建车辆及行人检测数据集,对改进算法进行综合评估。

2 基于多尺度特征融合的目标检测算法

2.1 YOLOv8-RC算法

YOLOv8目标检测算法^[11]总体架构包括3个部分:骨干(Backbone)模块对输入的图像进行下采样操作,可利用不同的网络深度得到不同层次的特征图信息;颈部(Neck)模块包含上采样及多个特征融合层等系列网络模块;头部(Head)模块用于输出检测结果,并对网络所提取的特征信息进行解码,同时负责预测每个候选框的类别及位置。在车辆及行人检测任务中,为对所提取的特征信息进行更好地融合及增强,本文在YOLOv8网络中引入一种多尺度特征融合模块RCS-OSA。此外,为更好地保留及提取特征,进一步加强网络的特征融合能力,将原网络的上采样算子改进为CARAFE模块,保持YOLOv8网络中快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)模块不变,形成改进算法YOLOv8-RC,其架构如图1所示。图1中所使用的卷积(Conv)模块中卷积核尺寸 k 为3、移动步长 s 为2、填充参数 p 设置为1。

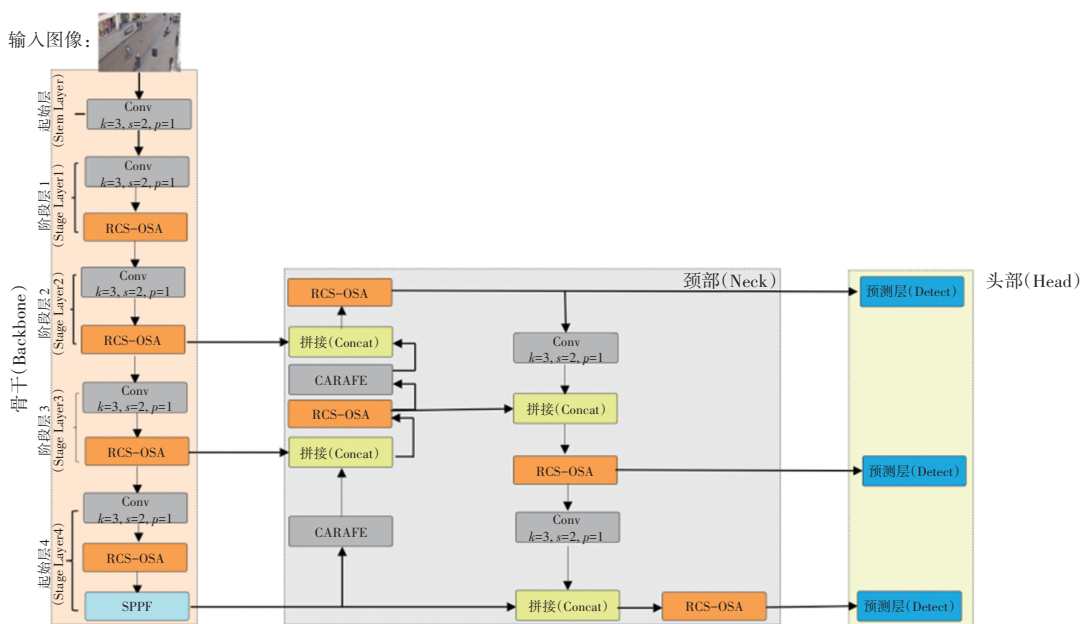


图1 YOLOv8-RC网络架构

2.2 多尺度特征融合模块

RCS(RepVGG/RepConv ShuffleNet)-YOLO 通过增强特征提取及融合能力,提高了目标检测的性能^[12]。故本文将该算法的 RCS-OSA 模块引入 YOLOv8 网络,其架构如图 2 所示,其中 C 为输入通道数量, n 为堆叠的 RCS 模块的数量。

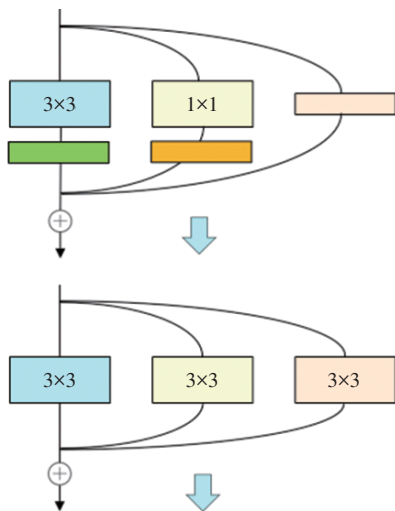
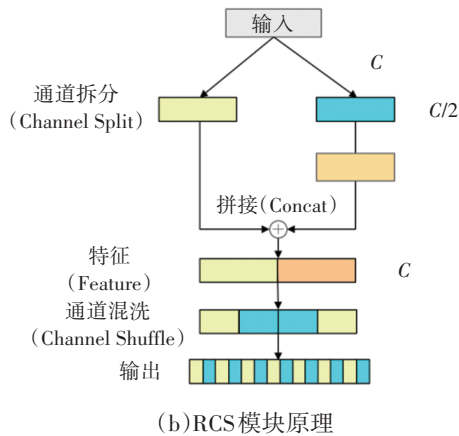
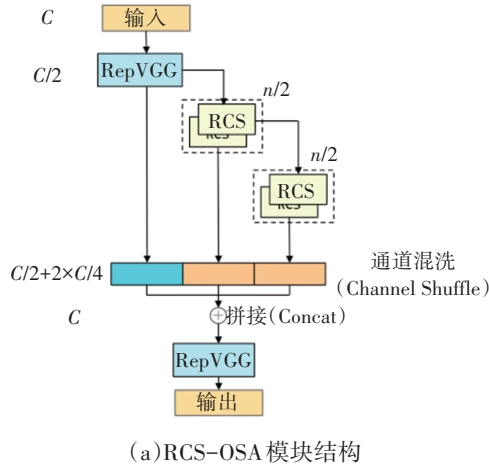


图 2 RCS-OSA 原理

RCS-OSA 模块首先对输入的特征图进行均匀分块,其中的一半特征图利用 RepVGG 网络的 3×3 卷积进行特征提取,另一半特征图信息用 RCS 模块进行特征提取。其中 RCS 模块将输入的特征再次进行分块,再采用通道拆分(Channel Split)及 RepVGG 操作将特征图相加得到新的特征图,最后利用通道混洗(Channel Shuffle)操作将特征图重新排列,从而使不同组之间的特征更好地交流及融合,以来增强特征的混合和传递效果。总体来说,RCS-OSA 模块结合了 3×3 卷积及丰富的特征融合模块,有助于更有效地捕捉图像特征中的重要信息。将其应用于 YOLOv8 网络中可对特征图信息进行更加充分的增强及融合,从而捕捉不同大小检测目标的细节信息。

2.3 上采样算子改进

YOLOv8 采用上采样(Upsample)算子执行上采样操作,目的在于提高特征图的空间分辨率,以提升网络的感知范围及定位精度。然而,该算子采用插值的方法填充新的像素,可能导致细节信息丢失,且不能准确还原原始图像细节。为解决这一问题,本文引入 CARAFE 轻量级上采样算子^[13]替换原上采样算子,其原理如图 3 所示。

CARAFE 上采样算子包含上采样核预测模块(Kernel Prediction Module)及特征重组模块(Content-aware Reassembly Module)。上采样核预测模块中的 H 、 W 及 C_m 分别为输入特征图的高度、宽度及通道数量, σ 为上采样率, σW 、 σH 分别为输出特征图的高度、宽度。内容编码器(Content Encoder)用于编码输入特征图的内容信息,内核归一化器(Kernel Normalizer)用于归一化处理预测的内核,通道压缩器(Channel Compressor)用于减少特征数量。特征重组模块中的 (x_i, k_{up}) 表示该位置的邻域特征, W_i 为用于重组操作的权重, X' 为通过内容感知重组模块后的输出特征图尺寸,其对应大小为 $(C, \sigma W, \sigma H)$ 。图 3 中, X 为输入特征图的尺寸。

该上采样算子的预测模块首先对输入的特征图进行通道压缩,然后利用卷积层进行内容编码及上采样核预测。其中上采样核的尺寸参数设定为:生成重组核的编码器卷积核大小 $K_{enc}=3$ 、重组核的大小 $K_{up}=5$,预测得到的上采样核通过归一化指数函数(Softmax 函数)进行归一化,以此确保权重之和为 1。此外,特征重组模块将输出特征图的每个位置映射回输入特征图,并提取以该位置为中心的区

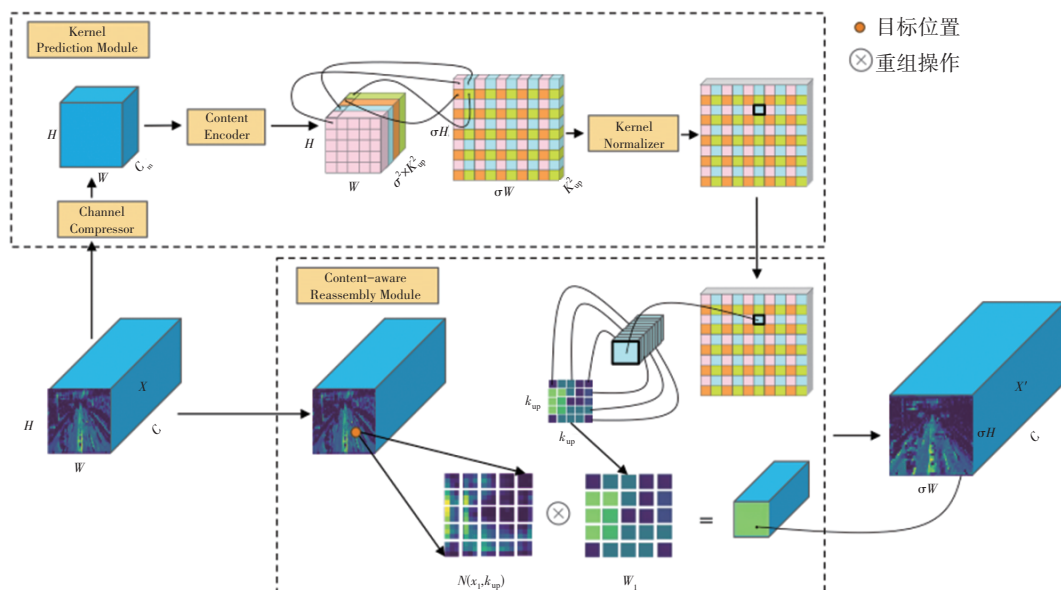


图3 轻量级上采样算子CARAFE原理

域。随后,该区域与相应位置的上采样核进行点积操作以获得输出值。最后,为保持特征的一致性及共享性,使不同位置的通道可共享相同的上采样核,在加入CARAFE算子后,可动态调整上采样核以适应不同的上采样需求,从而更好地保留和提取特征信息,在保持计算效率的同时,还可减少信息丢失。此外,以上过程通过融合不同尺度的特征图,还可获得更为丰富的上下文信息,以增强网络的特征表达能力,从而提高算法的感知范围及定位精度。

3 试验验证

3.1 试验环境及参数配置

为评估改进算法在车辆及行人检测任务中的表现,搭建了如表1所示的试验软硬件环境。此外,为控制变量,所有模型的训练均未采用预训练模型,同时均采用混合精度训练模式。

表1 试验软硬件环境

配置名称	版本参数
操作系统	Windows 11
图形处理器	NVIDIA GeForce RTX 4080
处理器	13th Gen Intel Core i9-13900H
深度学习框架	PyTorch 2.2.0
Python版本	Python 3.9.18
CUDA版本	CUDA 11.8

完成试验环境搭建后,本文所构建的所有网络模型在训练过程中也均采取相同的超参数设置,具

体配置如表2所示。

表2 各模型超参数

名称	数值
图片大小	480×480
初始学习率	0.01
批量大小(Batch Size)	16
优化器	随机梯度下降(SGD)
权重衰减系数	0.000 5
轮次/轮	200

3.2 数据库构建

为验证本文算法在道路交通环境下对车辆及行人目标的检测性能,选取多目标车辆检测公开数据集UA-DETRAC作为基础数据集,将该数据集中所有的车辆标注进行合并,并对包含行人的图像进行重新标注,得到仅含车辆及行人标注的目标检测数据集。为全面检验及分析所提出算法的效果并平衡数据集中的标签分布,通过网络收集的方式构建了更多包含行人目标的检测数据集,最后将两类数据集进行筛选整合,获得由6 000张检测目标较为均衡的车辆及行人图片构成的检测数据集,并将其划分为训练集、测试集及验证集,分别包含5 000张、500张、500张图片。

3.3 评价指标

本文采用采用精确率(Precision)P、召回率(Recall)R、平均精度(Average Precision, AP)和平均精度均值(mean Average Precision, mAP)对算法性

能进行定量评价:

$$P=N_{TP}/(N_{TP}+N_{FP}) \quad (1)$$

$$R=N_{TP}/(N_{TP}+N_{FN}) \quad (2)$$

$$p_{avg} = \int_0^1 P(R)dR \quad (3)$$

$$m_{p-avg} = \frac{1}{m} \sum_i p_{avg}(i) \quad (4)$$

式中: N_{TP} 、 N_{FP} 、 N_{FN} 分别为真正例、假正例、假负例的数量, m 为检测目标的类别数量, i 为类别序号, p_{avg} 为平均精度(AP), m_{p-avg} 为平均精度均值(mAP)。

AP作为单类别目标检测中常用的评价指标,其值为这一类目标的 P - R 曲线下的面积;mAP则代表了整个数据集中所有类别的AP均值,一般将交并比(Intersection over Union, IoU)阈值设置为0.5,即IoU大于0.5的预测框有效,用mAP50表示。在IoU的取值为0.50~0.95范围内,每间隔0.05计算一次mAP,并计算所有mAP的均值,记为mAP50-95。

3.4 对比试验

为验证改进模型对车辆及行人目标的检测效果,选取YOLO系列中的典型算法作为对比模型。此外,YOLO算法根据不同的网络尺度可分为N、S、M、L及X版本,考虑到模型在实际检测应用过程中的部署需求,将轻量级算法YOLOv8-N作为基础模型进行改进,同时选取其余对比模型的轻量级算法版本,其中YOLOv3则选取tiny版本。通过200轮的迭代训练,最后在测试集上得到各自模型的检测结果,如表3所示。

表3 对比模型试验结果 %

模型	P	R	mAP50	mAP50-95
YOLOv3-tiny	82.7	68.5	78.4	52.0
YOLOv5-N	87.0	81.4	90.2	66.2
YOLOv6-N	88.6	81.1	90.4	66.5
YOLOv8-N	86.2	82.2	90.3	67.7
YOLOv8-RC	87.9	83.4	91.2	68.2

由表3可知,改进算法YOLOv8-RC在所构建的车辆及行人检测数据集上的评价指标结果均高于基础算法YOLOv8-N。综合而言,YOLOv8-RC的各项评价指标在整体上得到了显著提升,较各对比模型具有更为优异的检测性能。

3.5 消融实验

为验证YOLOv8-RC中改进模块RCS-OSA及

CARAFE的提升效果,进行了如表4所示的消融实验。其中Model 1为加入RCS-OSA模块后的YOLOv8-N。

由表4可以看出:改进算法YOLOv8-RC在加入RCS-OSA模块后,除平均召回率外,其余指标均得到了显著提高;引入CARAFE后,可更好地融合特征信息,综合指标再次提升,更适用于车辆及行人检测任务。

表4 消融实验结果

模型	YOLOv8-N	Model 1	YOLOv8-RC
RCS-OSA	×	√	√
CARAFE	×	×	√
P /%	86.2	88.0	87.9
R /%	82.2	82.0	83.4
mAP50/%	90.3	90.8	91.2
mAP50-95/%	67.7	68.1	68.2

3.6 检测结果可视化分析

为验证改进算法YOLOv8-RC相比于其余对比模型的检测效果,从车辆及行人检测数据集的测试集中随机抽取2张典型的场景图片(分别为多车辆少行人场景及行人街道场景)来进行检测效果的可视化对比,其检测结果如图4所示。

由图4可知,在行人街道场景下,各算法均未出现漏检及错检情况,仅在预测置信度上存在部分差异,而改进算法YOLOv8-RC在多车辆少行人场景下对行人目标具备更加准确的检测性能,并且在车辆的预测上具备更高的置信度。



(a)YOLOv8-RC



(b)YOLOv8-N



(c)YOLOv6-N



(d)YOLOv5-N



(e)YOLOv3-tiny

图4 YOLOv8-RC 及对比模型检测结果可视化

4 结束语

本文提出了一种基于改进 YOLOv8 的车辆及行人检测算法 YOLOv8-RC,通过引入多尺度特征融合模块 RCS-OSA 加强全局特征信息的融合能力,并将基础网络 YOLOv8 的上采样算子改进为轻量级上采样算子 CARAFE,提高了网络的感知范围及定位精度。通过多目标车辆检测公开数据集 UA-DETRAC 及网络搜集的方式构建了由 6 000 张含有车辆及行人的图片构成的检测数据集,并选取 YOLO 系列的多个典型代表算法进行对比验证,结果表明,YOLOv8-RC 在车辆及行人检测任务上具备优异的检测性能。

虽然 YOLOv8-RC 算法相比于基础算法 YOLOv8-N 具备更为优异的检测效果,但在后续研究中,还可继续对数据集进行优化和扩充,以适应更多的应用场景。此外,车辆及行人被遮挡的情况仍是未来的研究重点,可进一步对网络算法进行优化调整,以实现综合检测效果持续提升。

参 考 文 献

- [1] LIU P Y, MA Y X. Key Points and Visible Part Fusion Attention Network for Occluded Pedestrian Detection in Traffic Environments[J]. Optoelectronics Letters, 2024, 20 (7): 430-436.
- [2] JAIN D K, ZHAO X D, GARCIA S, et al. Robust Multi-Modal Pedestrian Detection Using Deep Convolutional Neural Network with Ensemble Learning Model[J]. Expert Systems with Applications, 2024, 249.
- [3] LIANG W J. Research on a Vehicle and Pedestrian

- Detection Algorithm Based on Improved Attention and Feature Fusion[J]. *Mathematical Biosciences and Engineering*, 2024, 21(4): 5782-5802.
- [4] KIRAN K V, DASH S, PARIDA P. Vehicle Detection in Varied Weather Conditions using Enhanced Deep YOLO with Complex Wavelet[J]. *Engineering Research Express*, 2024, 6(2).
- [5] ANDIKA J L, KHAIRUDDIN A S M, RAMIAH H, et al. Improved Feature Extraction Network in Lightweight YOLOv7 Model for Real-Time Vehicle Detection on Low-Cost Hardware[J]. *Journal of Real-Time Image Processing*, 2024, 21(3).
- [6] 胡倩, 皮建勇, 胡伟超, 等. 基于改进 YOLOv5 的密集行人检测算法研究[J/OL]. *计算机工程*(2024-04-30)[2024-06-25]. <https://doi.org/10.19678/j.issn.1000-3428.0068753>.
- HU Q, PI J Y, HU W C, et al. Research on Dense Pedestrian Detection Algorithm Based on Improved YOLOv5[J/OL]. *Computer Engineering* (2024-04-30)[2024-06-25]. <https://doi.org/10.19678/j.issn.1000-3428.0068753>.
- [7] 黄丹丹, 张新茹, 刘智, 等. 基于 Transformer 的双分支检测和重识别的多行人追踪[J/OL]. *吉林大学学报(工学版)*(2024-03-28)[2024-06-25]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20231415>.
- HUANG D D, ZHANG X R, LIU Z, et al. Multi Pedestrian Tracking Based on Transformer's Dual Branch Detection and Re-Identification[J/OL]. *Journal of Jilin University (Engineering Edition)* (2024-03-28)[2024-06-25]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20231415>.
- [8] 李林红, 杨杰, 冯志成, 等. 面向站口行人检测的改进型 Yolov5s 算法[J]. *南京大学学报(自然科学)*, 2024, 60(1): 87-96.
- LI L H, YANG J, FENG Z C, et al. Improved Yolov5s Algorithm for Pedestrian Detection at Station Entrances[J]. *Journal of Nanjing University (Natural Science)*, 2024, 60(1): 87-96.
- [9] 周力, 惠飞, 张嘉洋, 等. 基于 RDB-YOLOv5 的遥感图像车辆检测[J]. *长安大学学报(自然科学版)*, 2024, 44(3): 149-160.
- ZHOU L, HUI F, ZHANG J Y, et al. Remote Sensing Image Vehicle Detection Based on RDB-YOLOv5[J]. *Journal of Chang'an University (Natural Science Edition)*, 2024, 44(3): 149-160.
- [10] 沈正坤, 王正超, 尹怀仙, 等. 基于 VOD-YOLOv5 的车辆检测算法研究[J]. *青岛大学学报(工程技术版)*, 2024, 39(1): 87-94.
- SHEN Z K, WANG Z C, YIN H X, et al. Research on Vehicle Detection Algorithm Based on VOD-YOLOv5[J]. *Journal of Qingdao University (Engineering Technology Edition)*, 2024, 39(1): 87-94.
- [11] DILLON R, JORDAN K, JACQUELINE H, et al. Real-Time Flying Object Detection with YOLOv8[EB/OL]. (2024-05-22) [2024-06-25]. <https://doi.org/10.48550/arXiv.2305.09972>.
- [12] KANG M, TING C M, TING F F, et al. RCS-YOLO: A Fast and High-Accuracy Object Detector for Brain Tumor Detection[EB/OL]. (2023-10-03) [2024-06-25]. <https://doi.org/10.48550/arXiv.2307.16412>.
- [13] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-Aware ReAssembly of FEatures[EB/OL]. (2019-10-29) [2024-06-25]. <https://doi.org/10.48550/arXiv.1905.02188>.

(责任编辑 斛 畔)

修改稿收到日期为2024年7月13日。