

基于YOLOv8的校园自行车流监测模型

摘要

本文针对清华大学校园自行车车流问题,实现了深度学习校园自行车流监测。该模型以YOLOv8为框架,从公开数据集中获取数据进行训练,以视频流的方式实现对自行车的目标检测与识别,并且设计算法在目标检测的同时检测车流速度。对于重叠和残缺自行车,本文利用数据集标签实现了更精确的识别,完善了项目的功能,为自行车计数提供了基础。项目作为校园交通模型的组成部分初步实现了监测任务,但是在应对密集复杂情形,以及在算法复杂度等方面有更多改进空间。

1. 前言

1.1. 项目基础

目标检测是计算机视觉是目前的研究与应用热点,其基于纯视觉的方式检测与处理特定类型的目标,完成确定类别、分布情况的任务。

随着标注良好的图片数据集发布、壮大,深度学习在人工智能目标检测方面体现出巨大的优势。

计算机视觉的目标检测技术用于交通领域时,因其基于纯视觉的工作模式,具有终端设备要求简单、通用性好的特点,同时获得的信息可扩展性强,拥有丰富的信息挖掘潜力。随着基于深度学习的目标检测算法的发展以及算力的提升,人工智能在这方面的应用也迎来一波热潮。

1.2. 项目背景

基于目标检测的基础,结合交通方向专业知识,本项目立足于校园通勤交通这一具体场景。针对校园在高峰期拥堵这一痛点,有必要建立校园交通模型,了解车流信息并针对性地做出优化。作为优化建设的重要一环,本项目希望高效地获得车流信息,为后续的分析与优化问题提供数据支撑。

1.3. 项目方案与难点

本项目通过现场录制视频模拟视频监控车流的情境,

通过深度学习算法识别密集的自行车目标并获取车流的有关信息如流量、车速等。

本项目的最终实际应用场景应是实时的车流监控与分析,根据实际需求特点,本项目具有处理速度快,计算效率高的基本要求。在目前主流的目标检测算法中,主要有One-stage算法与Two-stage算法两类,前者直接通过主干网络给出类别和位置信息,如YOLO系列、SSD算法等;后者先进行区域生成,再通过卷积神经网络进行样本分类,如R-CNN、Fast-RCNN、Faster-RCNN等。Two-stage算法在部分应用场景准确率更扎实,但在实时计算、计算效率等方面不如One-stage算法。故本项目基于One-stage算法,并选择其中的YOLO系列模型作为项目的基础。

同时,实际场景中背景驳杂,干扰因素较多,并需要应对车流密集、动线复杂、遮挡严重问题。传统的cv框架多采用针对视频流的固定背景-像素移动的方法识别动态物体,但本项目的复杂特点限制了此类传统追踪方法的发挥。而本项目先对单帧进行检测识别,提高了适用性,但也提出了更高的准确率要求。

2. 相关工作

本项目基于的YOLOv8模型是YOLO模型系列的一种,目前目标检测领域有广泛的应用。以下为与本文有较强相关性的工作。

2.1. YOLO系列模型

YOLOv1最初于2015年由华盛顿大学的博士研究生Joseph Redmon提出^[1],使用了类似于GoogleNet^[2]的主干网络,其结构如图1所示,包含24个卷积层和2个全连接层。其核心原理是将图像分割为一个尺寸为 $s \times s$ 的网格单元,在目标对象的中心落入网格单元之一的情况下,该特定网格单元将负责检测该对象。每个网格单元负责预测2个边界框,同时预测20类不同的物体。其中,每个边界框包含五个变量 (x, y, w, h, c) ,分别为框中心的x坐标、y坐标、框的长度、高度、置信度。通过监督学习的梯度下降与反向传播算法对损失函数进行优化。该工作的损失函数为

$$\begin{aligned}
& \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
& + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} [(\sqrt{\omega_i} - \sqrt{\hat{\omega}_i})^2 \\
& + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{obj} (C_i \\
& - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\
& + \sum_{i=0}^{S^2} \mathbb{I}_i^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (1)
\end{aligned}$$

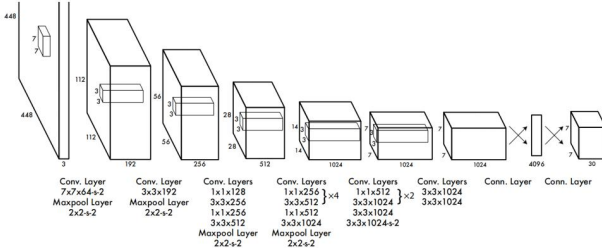


图 1 YOLOv1 模型结构

其中, λ_{coord} 和 λ_{noobj} 为参数, $(x_i, y_i, \omega_i, h_i, C_i)$ 为预测的边界框的参数, $(\hat{x}_i, \hat{y}_i, \hat{\omega}_i, \hat{h}_i, \hat{C}_i)$ 为标签; \mathbb{I}_{ij}^{obj} 为第i个网格的第j个边界框, 若负责预测目标则取 1, 否则为 0; \mathbb{I}_{ij}^{noobj} 为第i个网格的第j个边界框, 若不负责预测目标则取 1, 否则为 0; B为预测的边界框的数量; $p_i(c)$ 、 $\hat{p}_i(c)$ 分别表示第i个网格类别的预测值和标签值^[3]。

YOLOv1 在PASCAL VOC 2007 数据集上测试后, 具有 63.4%的平均精度均值(Mean Average Precision, mAP), 并且每秒处理帧数(Frames Per Second, FPS)为 45, 这说明YOLOv1 具有较好的检测能力与极快的检测速度。

继YOLOv1 之后, YOLOv2~YOLOv8 系列应运而生, 其主要变化是在YOLOv1 的基础上更改主干网络(Backbone), 增加新的颈部网络(Neck)于主干网络与预测的头部网络(Head)之间, 并进行了不同程度的调参与修改^[4,5]。本项目中所用到的YOLOv8 模型将在第三节进行介绍。

2.2. 基于YOLO系列模型的相关应用

Haiwei C, Guohui Zh, Huixin J基于改进的YOLOv8 模型对课堂视频中学生的行为实现检测^[6]。其通过结合Res2Net和YOLOv8 网络模型中的模块, 提出了一种新的C2f_Res2block模块, 该模块与MHSA和EMA一起集成到YOLOv8 模型中。改进后模型具有更好的检测性能,

平均精度提高了 4.2%。

周飞, 郭杜杜, 王洋等人针对目前复杂交通监控场景下车辆检测精度不足、检测速度慢的问题, 提出一种基于YOLOv8 模型的轻量级车辆检测算法^[7]。该工作将YOLOv8 的骨干网络替换为轻量级的FasterNet, 提高了模型的检测精度和推理速度; 在Backbone和Neck部分添加SimAM注意力模块, 提高了模型的特征融合能力; 针对密集车流下小尺寸车辆检测效果不佳的问题, 添加小目标检测头, 提高了密集车流下小目标车辆的检测精度; 使用可自适应调整权重系数的Wise-IoU作为损失函数, 提升了模型的收敛能力。在UA-DETRAC数据集的实验结果显示, mAP和FPS分别提高了 3.06%和 3.36%, 有效改善了复杂交通场景下小目标车辆检测效果不佳的问题。

付锦焱、张自嘉、孙伟等人针对航拍图像检测中易造成漏检和误检等问题, 提出了一种改进小目标检测算法CA-YOLOv8^[8]。文章设计了通道特征部分卷积模块CFPCConv (Channel Feature Partial Convolution), 减少了网络参数, 融合多尺度的小目标特征; 嵌入提升上下文聚合能力的模块CAM (Context Aggregated Module), 增强检测头对小目标信息的挖掘能力; 添加NWD损失函数, 将其与CIoU(Complete Intersection over Union)结合, 降低位置偏差的敏感性; 把原有检测头替换为DyHead(Dynamic head), 利用多注意头的优势。在VisDrone2019 数据集的测试中, 该工作较YOLOv8s原模型参数量降低了 33.3%, 检测精度mAP50 值和mAP50:95 分别提升了 8.7%和 5.7%。

贵向泉、刘世清、李立等人创建了TAPDataset行人检测数据集, 并基于YOLOv8 算法, 提出YOLOv8-L, 用于检测景区行人^[9]。该工作采用BiFormer注意力机制和上采样算子CARAFE, 提升了模型的回归合适收敛速度; 引入了深度可分离卷积模块, 降低了模型的参数量和计算量; 增加一层小目标检测层, 来提取更多的浅层特征, 改善模型对小目标的检测性能。与YOLOv8 相比, 模型的参数量减少了 18.06%, 在TAPDataset 上mAP@0.5 提高了 5.51%, mAP@0.5:0.95 提高了 6.03%。

YONGJUN L, SHASHA L, HAOHAO D等人基于YOLOv3 的高检测精度和速度, 通过添加注意力机制、完全交集损失函数(complete intersection over union, CIoU)、非最大值抑制(non-maximum suppression, Soft-NMS)和深度可分离卷积, 提出了一种新的和有效的检测器YOLO-ACN, 在检测小目标和遮挡物体时具有较高的准确性和速度^[9]。YOLO-ACN在MS COCO数据集上以 22ms的实时速度达到 53.8%的mAP, 并且在KAIST数据集上的单个类的mAP甚至达到了 80%以上。

3. 初步实验结果与分析

3.1. 实验模型

本项目中采用的模型主体为YOLOv8模型，其由Ultralytics团队与2023年发布，撰写本文时还未发布相关论文。YOLOv8最主要的特点在于采用了基于无锚点的检测方法，这意味着它直接预测目标的中心点和宽高比，而不是预测锚框的位置和大小。该方法可以减少锚框的数量，提高检测速度和准确性，其基本架构如图2所示^[6]。

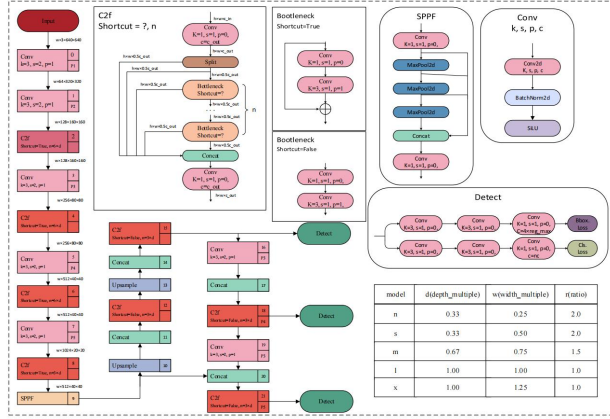


图2 YOLOv8 模型结构

具体来讲，YOLOv8 提供了一个全新的 SOTA 模型，其主要构成包含特征提取网络(Backbone)、特征强化网络(Neck)以及检测头(Head)三部分^[7]。Backbone部分在YOLOv5的基础上将第一层卷积改为 3×3 ，同时参考了YOLOv7^[12] ELAN的思想，采用C2f模块替换C3模块，并对不同尺度模型调整了不同的通道数，其主要优势在于既能获取更多梯度信息，又能保证模型参数轻量化^[9]，提升了模型的性能；Neck部分，该模型移除了一次上采样后的卷积层；Head部分，模型换成了目前主流的解耦头结构，从YOLOv5的Anchor-Based变成了Anchor-Free，将分类和回归独立，使用不同的损失函数；损失函数方面，模型在回归的损失函数中采用了分布式聚焦损失函数(Distribution Focal Loss, DFL)和CIOU损失函数。

Ultralytics 团队提供了多种服务于目标检测的YOLOv8模型，包括YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x等，本文采用YOLOv8n模型进行实验。

3.2. 数据集

本文采用Open Image Dataset V6数据集，其包含9M张图像。对于目标检测任务，数据集中含有近2M张图片，对象边界框包含600种类别。特别的，本项目基于Open Image Dataset V6数据集中18620张自行车图像，

将其划分为训练、验证和测试数据，比例为8:1:1。实际测试时使用实地拍摄的学堂路自行车流视频作为数据。

3.3. 评价指标

本文采用精确率(Precision, P)、召回率(Recall, R)、F1值(F1 Score)和mAP作为评价指标。其具体公式如下：

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \frac{P \times R}{P + R} \quad (4)$$

其中，TP表示正样本被判定为正的数量，FP表示负样本被判定为正的数量，FN表示正样本被判定为负的数量。

计算mAP需要引入平均精度(Average Precision, AP)，其含义是对不同召回率点上的准确率进行平均，在PR曲线图上表现为PR曲线下方的面积，公式为

$$AP = \int_0^1 P(R) dR \quad (5)$$

mAP即为对每一类的AP值进行平均，公式为

$$mAP = \frac{1}{S} \sum_{i=1}^S AP \quad (6)$$

特别地，mAP50和mAP50-95分别表示将IOU(Intersection over Union)阈值设为0.5、从0.5到0.95时AP的平均值。

3.4. 训练

训练时，导入yolov8n模型文件，在数据集上训练。具体的训练参数由模型文件决定，见下表。得到权重文件，训练结果和在测试集上的效果如下图。

表1 训练参数

optimizer	SGD
base learning rate	0.01
base weight decay	0.0005
optimizer momentum	0.937
batch size	128
training epochs	500
input size	640×640

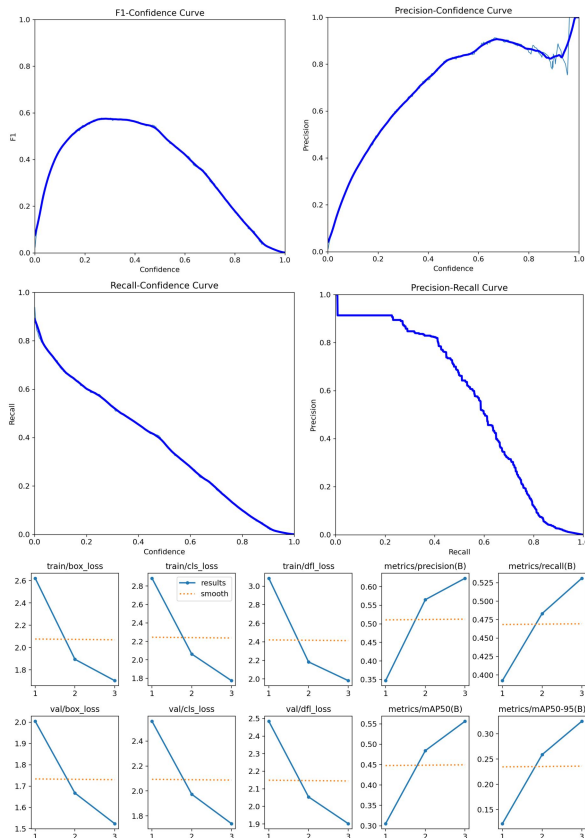


图 3 训练评价曲线

上图显示了不同分类阈值下的F1值、精确度、召回率等变化，以及损失的变化。随之置信度的增加，精确度越来越高，但在90%左右有一个明显的下凸；召回率和置信度近似成线性，没有明显的转折点；PR曲线不是很靠近右上角，说明了模型的性能不是十分理想。

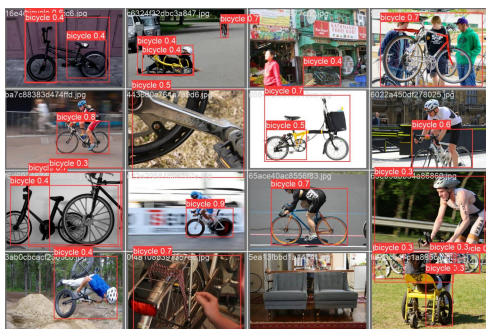


图 4 模型在验证集上的效果

3.5. 测试

使用训练好的参数识别学堂路上的自行车，效果如下图所示。可以看出，该网络能够比较准确的识别自行车，但是左侧自行车发生重叠的情况不能很好的识别。

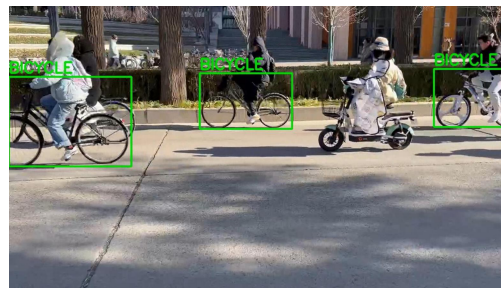


图 5 实地测试效果

4. 优化与完善

4.1. 识别重叠和残缺

数据集中除了类别、边界框坐标等参数外，还有一些参数，其中两个分别表示目标物体是否发生重叠(IsOccluded)和是否超出边界(IsTruncated)。基于此，本文将自行车分为两类，一类是完整的自行车，一类是残缺或重叠的自行车，进行训练。

由于数据集中完整的自行车数量较少，约8000个，为保证公平性，第二类自行车共选择10000个作为训练集。重复上述过程进行训练，结果如下图所示。

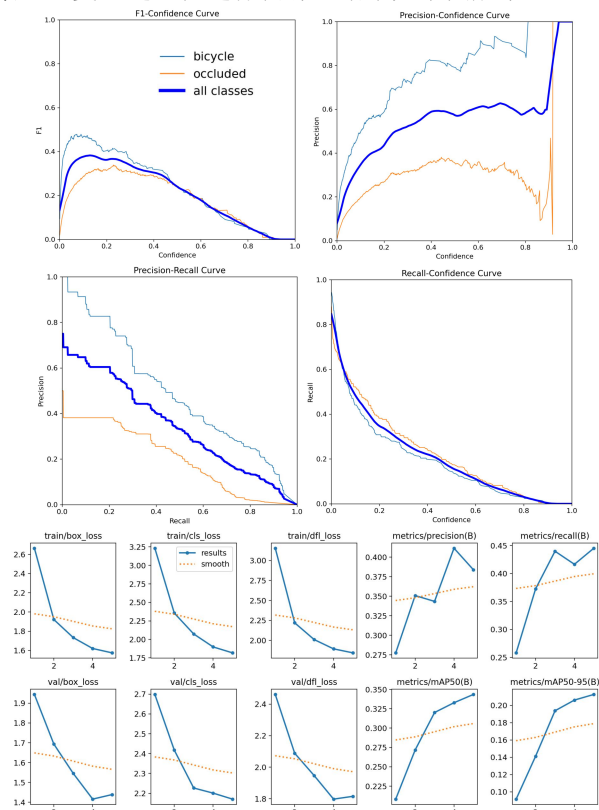


图 6 改进后的训练评价曲线



图 7 改进后实地测试效果

为了避免模型重复识别自行车的情况，尝试把完整自行车的置信度阈值调高。前两张图片展示了重叠自行车的识别情况，但是自行车分开后常常无法识别到，如第三张图片所示。基于此，可以采用两个模型分别识别重叠的自行车和完整的自行车，两个模型起到互补作用。

4.2. 车速测量

相机画面中部的均匀性很好，采用截取放大后的画面，可以利用识别边界框架的位置来获取车速。

在指定地点量取画面对应长度和一辆车通过的指定时间可以获得相应参数，对应边界框的移动速度（用像素表示），就可以在画面中表示出车的速度。

考虑到自行车在画面中主要是水平移动，为了应对遮挡问题，不妨规定边界框的左侧边作为测数标尺。

为了追踪到同一辆车，不妨设置条件找出中心点在两帧中相近的框就判断为同一辆。这一条件很容易实现，但是在两车超车这一遮挡过程中会短暂失灵或犯错。采用遍历算法也会直接抹去丢帧的数据。针对性解决办法是可以查看前几帧对应车辆的运行位置判断行驶方向。

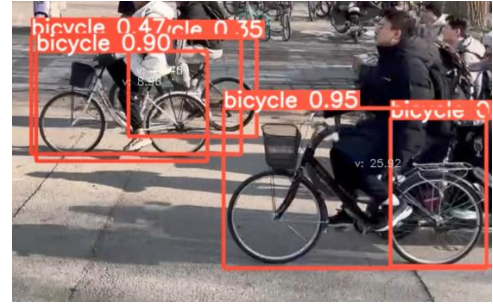


图 8 车速测量效果图

该方法的问题也要结合YOLO的运行特点。YOLO会先识别边界框中心位置，故更严格的方式是可以根据画面特点标定中心Y与单位像素表表征的速度的对应关系，例如：画面较远处（Y较大）处的物体更小，单位像素移动的距离更大。

标定对应关系的另一层要求是，拍摄位置、高度、俯角应当相对固定。这在交通应用较为容易实现。

然而在测试中可以发现，边界框位置本身并不稳定。这也是YOLO先确定中心后绘框的一大缺点。如果把一辆车以该方式的测得数据输出绘制成图则会发现其波动极大，但是再结合中心位置，其方向还是较为正确的。

考虑到实际运用是获取车流整体数据而非纠结于具体时刻与具体车辆，将各速度作平均值输出即可。

总的来说，本项目采用的算法逻辑比较简单，在特定情况会失灵。而本项目边界框定位的方法精度本身也不支持密集车流等复杂情况，只是一次有益的尝试。

4.3. 讨论与展望

当然，由于没有亲自拍摄数据集，网络上下载的数据不能完全满足训练实验的需求。例如，在查看数据集时，发生重叠的自行车并不都是自行车之间的重叠，有很大一部分图片是自行车被其他事物遮挡，这导致了许多的识别的错误。如果有更多的资源能够得到更好的数据集，模型的效果能够得到很大改善。

其次，视频的手持拍摄视角仍然不够高，这导致背景比较杂乱，增加了识别的难度，也无法使用常见的跟踪算法进行目标追踪。如果将摄像头装在高处，使用俯视角，可能会有更好的效果。

5. 总结

本项目针对自行车流监控这一场景，在了解多种基于深度学习的目标检测算法的基础上，选择YOLOv8模型展开工作。

本项目采用Open Image Dataset V6数据集中的自行车图像作为训练与验证集，并实地获取多角度拍摄视频作为测试集。模型实现了对测试集中完整自行车的识别跟踪，获得 95%以上置信度与准确率，并可以排除电动

车等干扰。在此基础上为模型制作了测速输出模块。对于平视视角重叠残缺自行车图形,项目利用训练集中的“完整性”参数对训练过程进行了优化,修改了置信度与结果分类方式,提高识别阈值。

项目作为校园交通模型的组成部分初步实现了监测任务,但是在应对密集复杂情形还有在算法复杂度等方面更多改进空间。

参考文献

- [1] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection. In Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on, pages 779-788. IEEE, 2016.
- [2] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, pages 1-9. IEEE, 2015.
- [3] 杨飞帆,李军.面向自动驾驶的YOLO目标检测算法研究综述.汽车工程师, 11: 1-11, 2023.
- [4] Hussain M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines, 11(7): 677, 2023.
- [5] 邵延华,张铎,楚红雨等.基于深度学习的YOLO目标检测综述.电子与信息学报, 44(10): 3697-3708, 2022.
- [6] Chen H, Zhou G, Jiang H. Student Behavior Detection in the Classroom Based on Improved YOLOv8. Sensors, 23(20): 8385, 2023.
- [7] 周飞,郭杜杜,王洋等.基于改进YOLOv8的交通监控车辆检测算法.计算机工程与应用, 1-13, 2024.
<https://link.cnki.net/urlid/11.2127.TP.20240105.0945.002.html>.
- [8] 付锦燚,张自嘉,孙伟等.改进YOLOv8的航拍图像小目标检测算法.计算机工程与应用 1-12,2024.
<http://kns.cnki.net/kcms/detail/11.2127.TP.20240105.0945.002.html>.
- [9] 贵向泉,刘世清,李立等.基于改进YOLOv8的景区行人检测算法.计算机工程,1-11, 2024.
<https://doi.org/10.19678/j.issn.1000-3428.0068125>.
- [10] Li Y, Li S, Du H, et al. YOLO-ACN: Focusing on Small Target and Occluded Object Detection. In IEEE Access, vol. 8, pages 227288-227303, 2020.
- [11] 孙方伟,李承阳,谢永强等.深度学习应用于遮挡目标检测算法综述.计算机科学与探索, 16(06): 1243-1259, 2022.
- [12] Wang C, Bochkovskiy A, Liao H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022.