

基于深度学习的高速公路小目标检测算法

徐慧智¹, 郝东升¹, 徐小婷², 蒋时森¹

(1. 东北林业大学 土木与交通学院, 哈尔滨 150040; 2. 浙江公路技师学院 试验检测中心, 杭州 310007)

摘要: 针对高速公路路侧摄像头拍摄的图像中, 远端的行人和车辆目标小、实时检测难问题, 提出一种改进的目标检测算法 YOLOv5s-3S-4PDH。首先, 采用 Shufflenetv2-Stem-SPPF 网络结构, 提高模型的运行速度; 其次, 引入加速归一化加权融合特征图和 160×160 小目标检测层, 优化小目标检测性能; 然后, 引入改进的解耦头机制, 提高小目标检测的定位和分类精度; 最后, 采用 Focal EIou 作为定位损失函数, 加快模型训练的收敛速度。在自建行人和车辆数据集上进行对比实验, 结果表明: 该算法与 YOLOv5s 基准网络算法相比, 计算量和参数量分别减少了 10.1% 和 24.6%, 检测速度和精度分别提高了 15.4% 和 2.1%; 在 VisDrone2019 数据集上进行的迁移学习实验表明, 该算法对所有目标类别的平均精度高于 YOLOv5s。YOLOv5s-3S-4PDH 算法在满足小目标检测实时性与精度的同时, 也具备泛化能力。

关键词: 交通运输规划与管理; 高速公路; 目标检测; 深度学习

中图分类号: U495 **文献标志码:** A **文章编号:** 1671-5497(2025)06-2003-12

DOI: 10.13229/j.cnki.jdxbgxb.20230939

Expressway small object detection algorithm based on deep learning

XU Hui-zhi¹, HAO Dong-sheng¹, XU Xiao-ting², JIANG Shi-sen¹

(1. School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China; 2. Test and Inspection Center, Zhejiang Highway Technician College, Hangzhou 310007, China)

Abstract: To address the challenging issue of real-time detection of small distant pedestrians and vehicles in images captured by roadside cameras on expressways, an improved object detection algorithm YOLOv5s-3S-4PDH was proposed. Firstly, the Shufflenetv2-Stem-SPPF network structure was used to improve the running speed of the algorithm. Secondly, the accelerated normalized weighted fusion feature map and the 160×160 small object detection layer were introduced to optimize the performance of small object detection; Then, the improved decoupling head mechanism was introduced to improve the localization and classification accuracy of small object detection. Finally, Focal EIou was used as the localization loss function of the algorithm to accelerate the training convergence speed of the algorithm. The results show that: compared with the YOLOv5s on the self-built pedestrian and vehicle dataset, the computation and parameter amount of the proposed algorithm are reduced by 10.1% and 24.6%, respectively, and the detection speed and accuracy are increased by 15.4% and 2.1%, respectively; Transfer learning experiment on the VisDrone2019 dataset shows that the proposed algorithm has better

收稿日期: 2023-09-05.

基金项目: 国家自然科学基金青年科学基金项目(62371170).

作者简介: 徐慧智(1977-), 男, 副教授, 博士. 研究方向: 交通环境感知理论与方法. E-mail: stedu@126.com

average precision for all categories. The proposed algorithm not only meets the real-time and accuracy requirements of small object detection, but also has generalization ability.

Key words: transportation planning and management; expressway; object detection; deep learning

0 引言

视频信息采集技术凭借其成本低廉、安装简便和提供信息丰富等优点,近年来在高速公路路侧得到了大范围部署,为基于视频技术的道路安全监控系统提供可靠、实时的数据支撑。然而现有的高速公路路侧监控设备拍摄范围大,导致远处车辆和行人目标在画面中像素占比小,以及互相遮挡现象严重,难以被精确检测。高速公路上的车辆运行速度快,对目标检测速度的要求更高。因此,如何在高速公路场景下寻求小目标检测的精度和速度平衡点,成为亟待解决的问题。

目前学术界针对小目标检测提出了多种改进方法,主要包括数据增强技术、锚框设计、多尺度检测和上下文信息利用等^[1]。数据增强技术是指对原始数据进行处理,以丰富小目标样本的多样性,提高算法对小目标的鲁棒性。VGG^[2]和 AlexNet^[3]分别采用了随机裁剪、随机缩放再裁剪、图片翻转等数据增强技术。王芋人等^[4]通过对图片进行高斯模糊、灰度处理、添加椒盐噪声等方法,进一步增强小目标特征信息。杨慧剑等^[5]通过对原始图像进行雾化处理,以提高在大雾天气下的小目标检测能力。

锚框设计最早应用于 Faster RCNN^[6]算法,通过设计不同尺寸的锚框来适配不同大小的目标,使锚框最大限度覆盖训练集中所有的目标。Singh 等^[7]提出一种 SNIP(Scale normalization for image pyramids)框架对锚框进行筛选,使大部分小目标能够在适当的尺度范围内进行训练,实现目标尺度和特征的归一化。Zhang 等^[8]提出一种稠密锚框策略,以确保不同尺度锚框的密度相同,提升小目标检测的召回率。

多尺度检测是指在不同尺度下检测目标,以便捕捉不同尺度的目标特征。王建中等^[9]针对目标的尺度和距离差异造成的多尺度问题,利用 YOLOv4 结合特征金字塔网络进行大、中、小目标的分离,实现不同尺度目标的检测。湛雨章等^[10]提出一种基于多速率空洞卷积的多尺度小目标检测方法,通过反卷积模块来重构图像中的细节,从而实现小目标分辨率的还原。李成豪

等^[11]通过多尺度感受野融合模块,对递归特征金字塔网络的输出进行独立处理,以提高算法对小目标的检测准确率。

上下文信息利用是指引入上下文信息来优化小目标检测。董亚盼等^[12]引入注意力机制模块来抑制背景杂波,增强小目标的特征。Qu 等^[13]通过将空洞卷积与特征融合相结合来强化浅层特征的语义信息,增强算法对小目标的检测效果。Li 等^[14]提出双通道特征融合网络,从两个不同的方向学习局部和上下文特征,以构建一种强有力的目标特征联合表达方法。

通过以上列举的方法可以提高小目标检测的性能,融合多种方法效果更好。在目标检测算法中,YOLO 系列^[15-18]是经典的单阶段算法,具有较高的检测精度和速度。针对特定的高速公路场景,以提高小目标检测精度和速度为目的,选择 YOLOv5s 为基准算法并进行改进。首先,从网络轻量化的角度出发,使用 Shufflenetv2-Stem-SP-PP 结构替代 YOLOv5s 主干网络,在不显著损失小目标检测精度的前提下,压缩网络的计算量和复杂度,提高运行速度;其次,增加一个 160×160 小目标层和引入权重参数加权特征图,加强算法对小目标语义特征的提取能力;再次,引入解耦头机制来加强算法的定位能力;最后,采用 Focal EIou 损失函数代替原预测框损失函数,加快模型训练速度。通过以上改进策略,能够提高算法对高速公路场景下小目标的检测能力。

1 YOLOv5s 算法

YOLO 系列主要目标检测算法已经更新迭代了 8 个版本,YOLOv5s 具有参数少、速度快、容易部署的特点,本文选取 YOLOv5s 作为高速公路监控设备检测基准算法。YOLOv5s 网络结构由输入端(Input)、主干网络(Backbone)、特征融合网络(Neck)和预测头(Head)4 部分组成。输入端主要负责读取图像和对应的标签文件,以及在输入网络之前对图像进行自适应缩放、数据增强和自适应锚框计算等预处理。主干网络的作用是提取目标的通用特征,由标准卷积层模块(Conv-

Batch normalization-siLU, CBS)、瓶颈层 C3 模块和快速金字塔池化(Spatial pyramid pooling-fast, SPPF)模块组成。特征融合网络将不同尺寸的特征进行融合,减少小目标在下采样过程中的信息缺失。YOLOv5s 在主干网络和预测头之间加入路径聚合网络(Path aggregation network, PANet)^[19],以进一步优化特征的融合。预测头模块对特征融合网络中 PANet 网络输出的大、中、小 3 个特征层进行回归分类预测,生成对应的预测框,然后使用 NMS 非极大抑制对预测框进行处理,选择最大 IoU 值对应的预测框作为目标在图片中的最终位置。

2 改进 YOLOv5s 的目标检测算法

针对高速公路上行人和车辆目标在图像中的位置分布随机,以及背景信息冗余等现象,从主干网络、特征融合、预测头和损失函数 4 个方面对 YOLOv5s 进行改进,提出 YOLOv5s-3S-4DPH 算法,以到达目标检测算法的高精度和高实时性平衡。改进后的算法网络结构如图 1 所示。

2.1 轻量化主干网络设计

YOLOv5s 主干网络使用大量的 CBS 模块和 C3 模块,导致网络的参数量和计算量庞大,运行速度慢,难以满足实时性需求。因此,提出 Shuf-

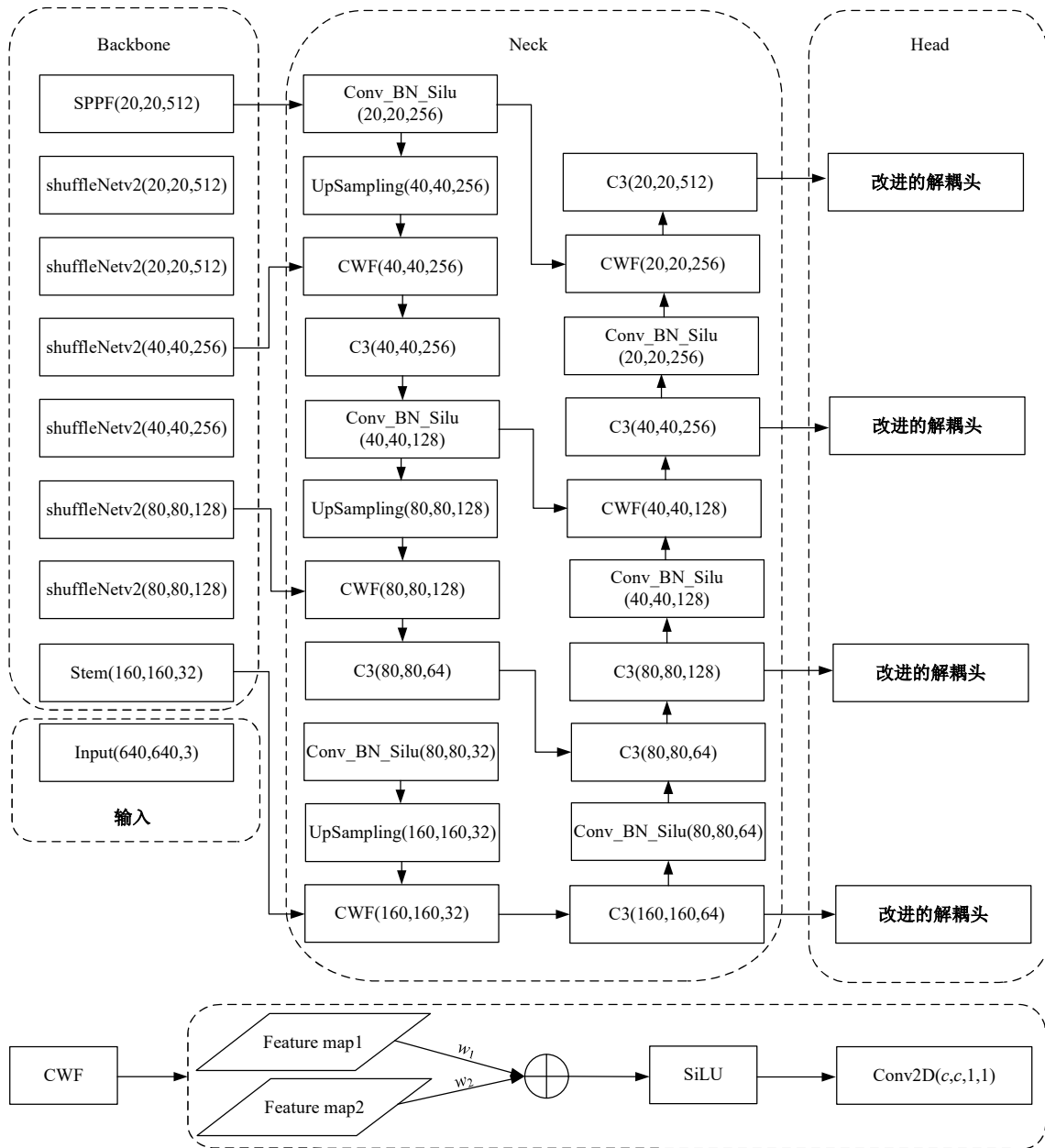


图 1 改进 YOLOv5s 网络结构图

Fig. 1 Network structure diagram of improved YOLOv5s

flNetv2-Stem-SPPF (3S)网络对主干网络进行改进,旨在保证精度损失较小的前提下降低算法的计算复杂度,满足嵌入式或移动设备轻量化部署需求。

ShuffleNetv2^[20]网络首先通过引入通道混洗(Channel shuffle)和通道拆分(Channel split)两大运算模块,在不增加网络参数数量的前提下,提升网络的有效容量和性能。其次,采用深度可分离卷积和逐点组卷积等技术,减少参数量,提高运行速度,使ShuffleNetv2网络成为一个高效和轻量级的深度学习网络算法。ShuffleNetv2网络由两个基本单元模块构成,每个单元包含两个分支,具体结构如图2所示。单元1首先通过通道拆分模块分成左右两个分支,其中左分支利用快捷连接执行恒等映射,加深模型的深度,提高模型的并行度。右分支经过两个1×1逐点组卷积层和一个3×3深度可分离卷积层,保证输入和输出通道数相同,减少网络参数量和降低运算复杂度。最后,将左右分支拼接并进行通道混洗操作,实现不同分支的特征信息交流,以提高检测精度。与单元1不同的是,单元2在分支之前直接引入输入,并同时采用步长为2的深度卷积层完成下采样,以实现特征图尺寸减半、通道数翻倍。

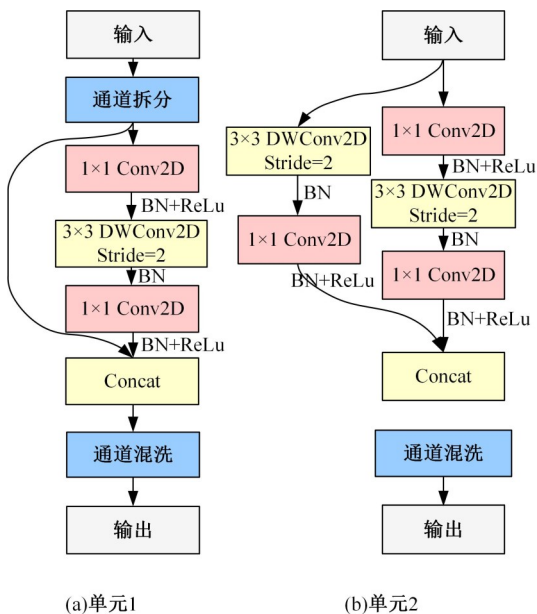


图2 ShuffleNetv2的单元模块
Fig. 2 Unit module of ShuffleNetv2

为了弥补网络轻量化所导致的精度损失,在原始ShuffleNetv2网络中使用Stem模块^[21]替代CBS模块,优化算法对目标的特征表达能力,并保留YOLOv5原有的SPPF模块,丰富特征图的

语义信息^[22]。Stem模块的网络结构如图3所示,其运行原理是使用两个分支提取图像的不同特征,增加网络的并行度和深度。在不明显增加计算量的前提下,通过下采样提取图像更深层次的语义特征,从而提高算法检测精度。

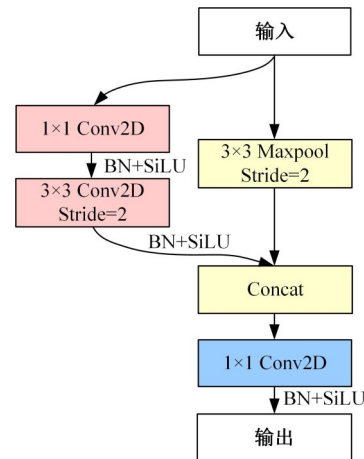


图3 Stem模块网络结构图
Fig. 3 Network structure diagram of Stem module

2.2 多尺度特征融合网络设计

在相机连续拍摄的图像中,同一目标在不同位置的尺寸不同,距离相机越远的目标尺寸越小,反之则越大。与常规目标相比,小目标在图像中存在可利用像素少、定位精度要求高、样本占比低等问题。随着模型网络层数的加深,小目标的特征和位置信息逐渐在深层网络中丢失,导致难以被检测。为了进一步提高算法对小目标的敏感性,减少漏检或误检情况的发生,提出以下两个优化方法:增加一个160×160小目标检测层;借鉴加权双向特征金字塔网络(Bidirectional feature pyramid network, BiFPN)^[23]结构思想,优化特征融合连接方式。

2.2.1 小目标检测层

在神经网络中,浅层特征图的感受野小,语义信息弱,上下文信息缺乏,但是可以获得更多空间和细节特征信息。基于此,提出多尺度目标检测算法,利用浅层特征图检测较小的目标,深层特征图检测较大的目标。在特征融合网络中将不同尺度的特征进行融合,使特征图兼顾浅层的表征特征和深层的语义特征,提升小目标检测性能^[24]。然后增加一个160×160的小目标检测层,来输出更小尺度特征,以进行小目标检测。使用四头输出端替代原来的三头输出端,具体修改后的网络结构如图4所示。

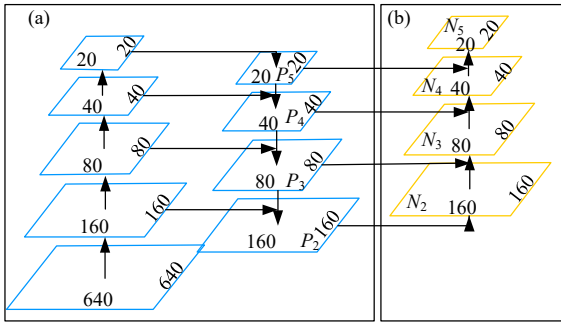


图 4 增加 160×160 层的多尺度特征融合网络

Fig. 4 Multi-scale feature fusion network for added 160×160 layer

2.2.2 加权融合

传统的特征融合方法如 FPN^[25]和 PANet 等, 仅将不同尺度的特征图进行叠加或相加操作, 忽视不同分辨率特征图在特征融合中的贡献差异。BiFPN 引入可学习的权重, 学习不同输入的特征图重要性, 并进行加权融合以提高算法对不同尺度特征融合的能力。BiFPN 将自顶向下和自底向上的融合路径模块化, 以实现可重复应用和更高层次的特征融合。本文引用 BiFPN 的加速归一化加权融合方法, 通过增设学习率 ϵ 来保证数值有效性, 提高 GPU 的运行效率。每层的输出特征计算公式如式(1)所示:

$$O_i = \sum_j \frac{W_j}{\epsilon + \sum_j W_j} \cdot I_j \quad (1)$$

式中: O_i 为第 i 层的输出特征图; I_i 为第 i 层的输入特征图; W_i 为第 i 层的输入特征图权重值。

2.3 解耦头引进

YOLOv5 存在目标定位和分类两个子任务, 且各自关注和感兴趣的区域存在空间错位问题, 其中分类任务更聚集目标特征, 定位任务更加关注真实框的坐标位置^[26]。YOLOv5 采用耦合结构的预测头, 将分类和边框回归任务在同一特征图上进行, 以共享预测头参数, 提升检测速度, 但会降低算法性能。

YOLOX 采用解耦头 (Decoupling head) 结构, 首先对特征融合网络输出的不同尺度特征图, 采用 1 卷积层将通道数减少至 256。然后使用两个并行分支的 3×3 卷积层, 来分别进行分类和定位任务。与使用一支全连接层的预测头分类相比, 参数量更少, 且加快了收敛速度^[27], 具体结构如图 5 所示。

YOLOv5 是基于锚框机制的目标检测算法,

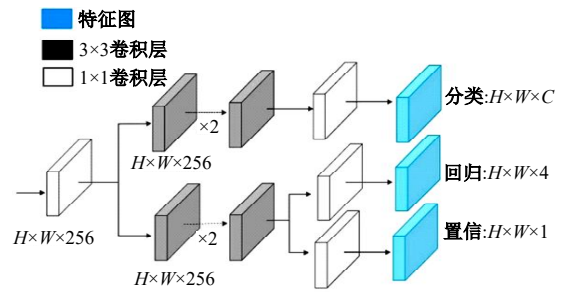


图 5 YOLOX 的解耦头结构示意图

Fig. 5 Schematic diagram of decoupling head of YOLOX

最后输出的特征图需要包括类别概率、预测框位置偏移量和置信度信息。这与基于 Anchor-free 的 YOLOX 不同, 因此, 需要在 YOLOX 解耦头的结构基础上将 3 种输出特征在通道方向拼接。同时, 在保证运行速度的前提下, 不增加过多的参数量和计算量, 删除两支并行的 3×3 卷积层。改进后的解耦头结构如图 6 所示。

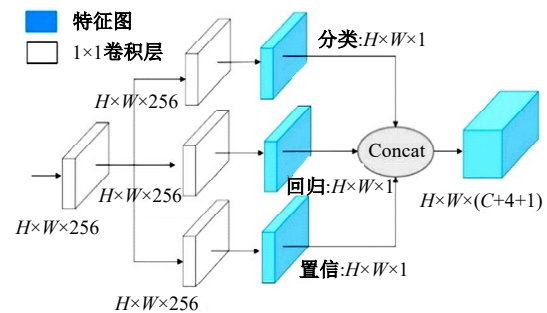


图 6 改进解耦头示意图

Fig. 6 Schematic diagram of improved decoupling head

2.4 Focal EIou 损失函数

YOLOv5-v7.0 的预测框回归采用 CIoU 损失函数, 通过引进真实框和预测框的重叠面积、中心点距离和宽高比, 来解决目标定位问题, 使预测框更接近真实框。CIoU 的计算公式如式(2)所示:

$$\begin{cases} L_{CIoU} = 1 - IoU + \frac{d^2(b, b^{gt})}{c^2} + \alpha v \\ IoU = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \\ v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \end{cases} \quad (2)$$

式中: d 为真实框与预测框中心点的欧氏距离; b 为预测框; b^{gt} 为真实框; c 为真实框与预测框最小外接矩形的对角线距离; α 为权重参数; v 为度量宽高比一致性; IoU 为真实框与预测框的面积交

并比; w^{gt} 为真实框的宽度; h^{gt} 为真实框的高度; w 为预测框的宽度; h 为预测框的高度。

CIoU 通过参数 v 反映真实框与预测框的宽高比差异性,但难以度量预测框与真实框的宽和高之间的真实差异,且有时会阻碍算法优化相似性。因此,引进 Focal EIoU 损失函数计算预测框回归损失。Focal EIoU 在 CIoU 的基础上提出两个优化方案:

(1) 将 CIoU 中预测框与真实框的宽高比损失项,拆分成预测框与真实框的宽度差值和高度差值,即用 $\frac{d^2(w, w^{gt})}{C_w^2} + \frac{d^2(h, h^{gt})}{C_h^2}$ 取代 αv 项,加快模型的收敛速度。

(2) 引入 Focal Loss 优化预测框回归任务中的样本不平衡问题,减小与真实框重叠较少的预测框对回归损失的影响,使回归损失倾向高质量预测框^[28]。Focal EIoU 损失函数计算公式如(3)和(4)所示:

$$L_{EIoU} = 1 - IoU + \frac{d^2(b, b^{gt})}{c^2} + \frac{d^2(w, w^{gt})}{C_w^2} + \frac{d^2(h, h^{gt})}{C_h^2} \quad (3)$$

$$L_{Focal-EIoU} = IoU^\lambda \times L_{EIoU} \quad (4)$$

式中: λ 为控制异常值抑制程度参数值; C_w 为预测框与真实框最小外接矩形的宽度; C_h 为预测框与真实框最小外接矩形的高度。

3 实验结果与分析

3.1 实验数据集

本文使用的行人和车辆数据集,是由多处高速公路卡口处电子交通摄像枪采集的车流视频,通过数据转换成图像后,再使用 LabelImg 软件进行标注自建而成。通过筛选重复、没有目标的图像,保留下的数据样本包含了白天、夜晚、晴天和阴天等高速公路场景下的 4 395 张车辆图像和 1 731 张行人图像。由于行人数据集包含的图像数量较少,采用数据增强技术单独扩充行人数据集,其中包括增加 1.5 倍的亮度、左右翻转和旋转 20° 等操作,使行人数据集扩充到 6 552 张图像。最终的行人和车辆数据集共计 10 947 张图像,按照 8:1:1 的比例随机划分为训练集、验证集和测试集。高速公路场景下的行人和车辆目标,在图像中的位置和标注框宽高标准化分布图如图 7 所示。

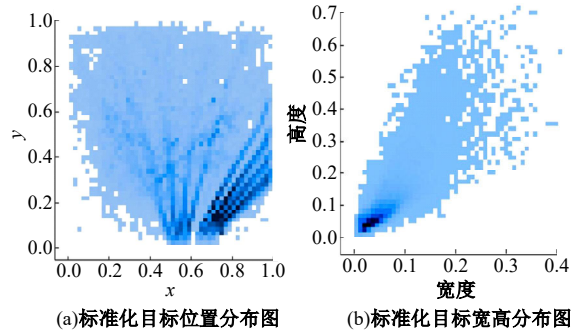


图 7 行人和车辆位置和宽高标准化分布

Fig. 7 Standardized distribution of pedestrian and vehicle positions and widths

由图 7(a)可知,行人和车辆目标大部分集中在 $[0.5 < x < 1.0, 0.0 < y < 0.6]$,即图像的右上方区域,符合实际情况。由图 7(b)可知,行人和车辆目标的宽高分布较为分散,主要集中在宽度小于 0.3、高度小于 0.5 的区域内,其中宽度和高度均小于 0.1 的目标占比最高,宽度大于 0.3 和高度大于 0.5 的目标占比小。这表明行人和车辆数据集中小目标样本数量较多,可以应用于本研究。自建行人和车辆数据集的部分数据如图 8 所示。



图 8 自建行人和车辆数据集

Fig. 8 Self-built pedestrian and vehicle datasets

3.2 实验环境

Pytorch 具有高模块化、易于修改等优点,因此,选用 Pytorch 作为深度学习框架,具体实验环境如表 1 所示。

3.3 模型超参数设置与评价指标

深度学习模型涉及大量的超参数,取值影响模型检测性能^[29]。为了验证本文提出的主干网络、特征融合、预测头和回归损失函数对模型性能的影响,对模型超参数取值采用了经验值。

在相关对比实验中,将训练时采用的超参数设置一致,其中训练总轮数为 300,训练批次大小为 64,图像输入图片尺寸统一为 640×640 ,使用

表 1 模型训练环境

Table 1 Model training environment

环境项	环境规格
CPU	Inter(R) Core(TM) i9-12900
内存	64 GB
显卡	NVIDIA RTX A4000
操作系统	Windows 11
编程语言	Python 3.9.10
深度学习框架	Pytorch 1.12.1
集成开发环境	Pycharm 社区版 2022.3.2
CUDA	12.0
CUDNN	8.3.2

SGD 优化函数和 warm-up 学习率,主要的超参数设置如表 2 所示。

表 2 实验超参数

Table 2 Experimental hyperparameters

超参数	系数值
初始学习率	0.01
循环学习率	0.01
动量	0.937
权重衰减系数	0.000 5
预热学习轮数	3.0
预热学习动量	0.8
预热初始偏置学习率	0.1
边界框回归损失系数	0.05
分类损失系数	0.5
置信损失系统	1.0
有无物体 BCE Loss 中正样本权重	1.0
分类 BCE Loss 中正样本权重	1.0
IoU 训练阈值	0.2
Anchor 的宽高比	4.0

选取平均精度 (AP) 和均值平均精度 (mAP) 评价模型检测精度,其中 AP@0.5 表示单个检测类别在 IoU = 0.5 时的平均精度, mAP@0.5 表示所有检测类别在 IoU = 0.5 时的平均精度, mAP@0.5:0.95 表示所有检测类别在不同 IoU 阈值(从 0.5 到 0.95,步长 0.05)上的平均精度;选取每秒内能处理的图像数量 (FPS) 评价模型运行速度;选取浮点数运算次数 (FLOPs) 评估模型的

时间复杂度;选取模型训练中所需要的训练参数总数 (Params) 评估模型的空间复杂度。相关公式如下:

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

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (6)$$

$$mAP@0.5 = \frac{1}{n} \sum_{i=1}^n AP_i (IoU = 0.5) \quad (7)$$

$$mAP@0.5:0.95 =$$

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{10} \sum_{t=0}^9 AP_i (IoU = 0.5 + 0.05t) \quad (8)$$

式中: n 为目标类别数; P 为精确率,表示正确预测为正样本的数量占全部预测为正样本的比例; R 为召回率,表示正确预测为正样本的数量占全部实际为正样本的比例。

3.4 实验结果与分析

3.4.1 YOLOv5 不同主干网络的对比实验

针对 YOLOv5 的 4 种不同主干网络进行对比实验,来验证采用 ShuffleNetv2-3S 对算法网络的改进效果。首先将主干网络 C3Net 替换为 ShuffleNetV2,得到 YOLOv5-1S 网络。然后,在主干网络添加 Stem 模块重构得到 YOLOv5-2S 网络。最后在主干网络添加 SPPF 模块,重构为 YOLOv5s-3S 网络。以下 4 个模型均采用 CIoU 做损失函数。在 YOLOv5 的 4 种不同主干网络上的评价指标如表 3 所示。

由表 3 可知,使用 Shufflenetv 2 作为主干网络时,模型的 Params 和 FLOPs 比 YOLOv5s 分别减少了 52.4% 和 53.8%,FPS 提高了 88.7%,但 mAP@0.5 降低了 3.8%,表明轻量化主干网络以牺牲精度为代价提升运行速度,对目标特征提取能力减弱。YOLOv5s-2S 在复杂度小幅度提升的基础上, mAP@0.5 较 YOLOv5s-1S 网络提高了 1.9%,行人目标 AP@0.5 提升幅度达 3.7%。YOLOv5s-3S 与 YOLOv5s 的 mAP@0.5 差距进一步缩小至 0.8%,但 FPS 提高了 74.8%。以上

表 3 不同主干网络对比实验结果

Table 3 Comparative experimental results of different backbone networks

模型	主干网络	AP@0.5/%		mAP@0.5/%	FPS	Params/ 10^6	FLOPs/ 10^9
		行人	车辆				
YOLOv5s	C3Net	89.3	95.3	92.3	26.6	7.02	15.8
YOLOv5s-1S	Shufflenetv2	83.0	93.9	88.5	50.2	3.34	7.3
YOLOv5s-2S	Shufflenetv2-Stem	86.7	92.3	90.4	48.7	3.36	9.0
YOLOv5s-3S	Shufflenetv2-Stem-SPPF	88.1	94.9	91.5	46.5	4.02	9.5

实验验证了 Shufflenetv2-Stem-SPPF 主干网络能有效降低网络的复杂度,缩短推理时间,弥补模型因轻量化导致的精度损失。

3.4.2 YOLOv5-3S 网络模型的消融实验

仅靠 Stem 与 SPP 模块无法完全弥补轻量化设计所造成的检测精度损失,因此,需要进一步提升网络的检测性能。为了探究本文提出的各改进

方法的有效性,在 YOLOv5-3S 的基础上,开展 5 组消融实验,以及 YOLOv8s 的对比实验。每组实验使用相同的超参数以及训练技巧,依次加入 Focal EIoU 损失函数、BiFPN 特征融合结构、解耦头、增加 160×160 检测层 4 种改进方法,其中“√”代表选择对应的改进方法进行训练,实验评价指标如表 4 所示。

表 4 消融和对比实验结果

Table 4 Ablation experiment results

序号	Focal EIoU	BiFPN	解耦头	增加 160×160 检测层	AP@0.5/%		mAP@0.5/%	FPS	Params/10 ⁶	FLOPs/10 ⁹
					行人	车辆				
1					88.1	94.9	91.5	46.5	4.02	9.5
2	√				88.3(+0.2)	95.2(+0.3)	91.8(+0.3)	45.4	4.02	9.5
3	√	√			88.6(+0.3)	95.3(+0.1)	91.9(+0.1)	40.1	4.17	10.2
4	√	√	√		90.1(+0.5)	95.4(+0.1)	92.7(+0.8)	38.6	4.40	10.9
5	√	√	√	√	93.1(+3.0)	95.5(+0.1)	94.4(+1.7)	30.7	5.31	14.2
YOLOv8s					91.8	95.9	93.8	29.6	11.1	28.4

由表 4 可知,采用 Focal EIoU 损失函数在不增加 FLOPs 的情况下,使 mAP@0.5 较第 1 组提高 0.3%,提升了模型对目标的定位准确度;采用 BiFPN 对特征融合网络进行改进,使 mAP@0.5 较第 2 组提高 0.1%;采用解耦头会降低 FPS,但模型 mAP@0.5 提升到了 92.7%,说明小目标的检测能力得到了改善,实现了不同尺度目标间的平衡;采用一个 160×160 检测层,mAP@0.5 和行人目标 AP@0.5 较第 4 组分别提升 1.7% 和 3.0%,FPS 减少至 30.7,证明小尺度特征图能够有效提高复杂背景下的行人小目标识别能力,但模型实时性降低。

从总体实验结果来看,改进算法可以有效提高小目标检测的精度和速度。YOLOv5-3S-4DPH 相对于 YOLOv5s,mAP@0.5 和 FPS 分别提高 2.1%、15.4%,Params 和 FLOPs 分别降低 24.4%、10.1%;YOLOv5-3S-4DPH 相对于 YOLOv8s,mAP@0.5 和 FPS 分别高 0.6%、3.7%,Params 和 FLOPs 分别低 52.2%、50.0%,如表 5 所示。

目前,主流的高速公路路侧摄像监控设备采

表 5 算法性能对比

Table 5 Algorithm performance comparison %

算法	mAP@0.5	FPS	Params	FLOPs
YOLOv5s	+2.1	+15.4	-24.6	-10.1
YOLOv8s	+0.6	+3.7	-52.2	-50.0

用 FPS 作为检测速度的评价指标(例如海康威视交通专用系列 iDS-2VS435-F832 采用 FPS=25 或 30),而 YOLOv5-3S-4DPH 在复杂高速公路场景下 FPS 达到 30.7(30.7>30),因此,检测速度满足视频交通目标实时检测的需求。

3.4.3 YOLOv5s-3S-4DFH 算法实验训练结果

如图 9 所示,该算法在验证集上的边界框、置信度和分类 3 类损失函数值,随着训练轮次逐渐趋向平稳收敛,且与训练集上的损失值差距不大,表明模型训练未出现过拟合现象,具有良好的稳定性。

如图 10(a)和(b)所示,该算法在验证集上的 mAP@0.5 和 mAP@0.5:0.95 值均随着训练轮次不断上升,且在 110 轮次之后趋于稳定。如图 10(c)和(d)所示,在训练 300 轮次时,改进算法的精确率为 90.7%,与 YOLOv5s 持平;召回率为 90.2%,比 YOLOv5s 微高。因此,YOLOv5s-3S-4DFH 算法在训练过程中比较稳定,且具有良好的检测性能。

3.4.4 泛化能力

为了验证 YOLOv5s-3S-4DPH 算法的泛化能力,将训练好的检测模型迁移到公开数据集上进行实验。选择 VisDrone2019 数据集作为迁移学习实验的数据集,该数据集由天津大学机器学习与数据挖掘实验室的 AISKEYEYE 团队构建,是一个大规模基准数据集,包含多种重要的计算

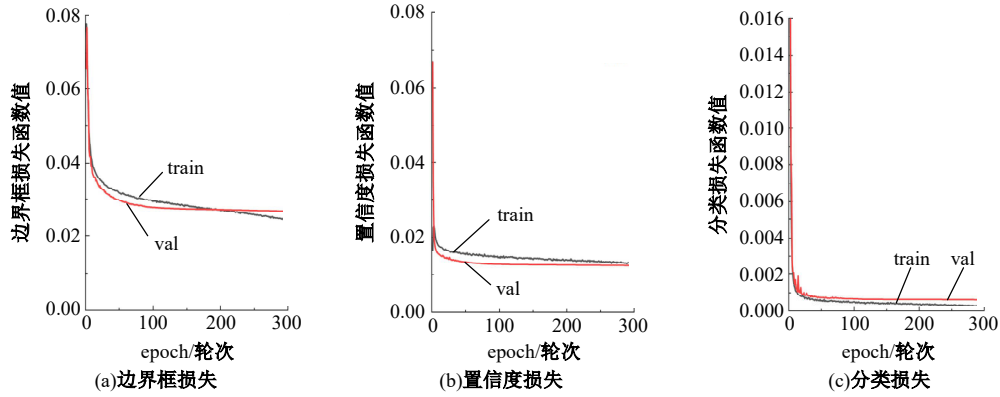


图 9 模型损失图

Fig. 9 Model loss diagram

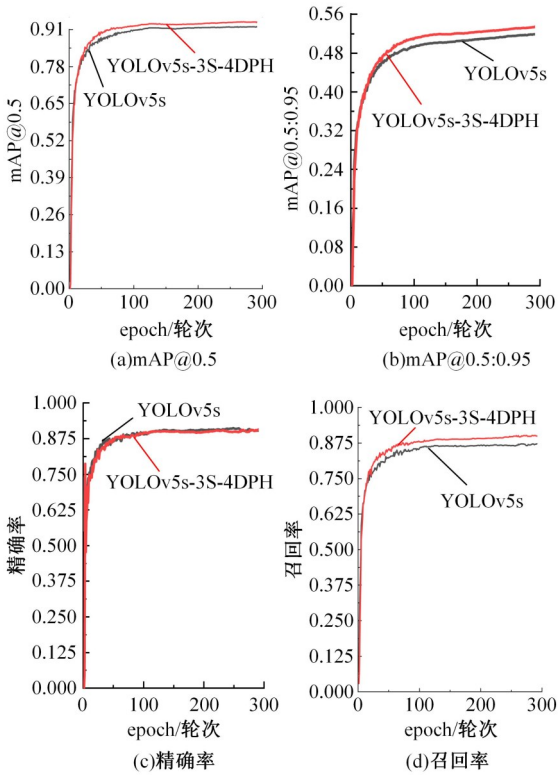


图 10 模型评价指标图

Fig. 10 Model evaluation indicator diagram

机视觉任务的精细标注。该数据集由各种装备摄像头的无人机捕获,手动标注了超过 260 万个常见目标的边界框,并提供了场景可见性、对象类别和遮挡等重要属性。冒国韬等^[30]采用 13 种算法在该数据集上进行了目标检测研究,用以评价模型性能。该数据集部分图像展示如图 11 所示。将 YOLOv5s-3S-4DPH 和 YOLOv5s 两种算法在 VisDrone2019 数据集上进行迁移训练,图片分辨率为 640×640 ,训练轮次 epoch 为 300,训练批次 batch size 为 64,实验结果如表 6 所示。



图 11 VisDrone 2019 数据集示例

Fig. 11 Example of VisDrone 2019 dataset

由表 6 可知,相较于 YOLOv5s, YOLOv5s-3S-4DPH 对 10 种目标在平均精度上均有提升, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提高了 4.0% 和 2.9%。综合分析上述结果,相比于 YOLOv5s, YOLOv5s-3S-4DPH 不仅提高了目标检测的性能,还具有以下几个优势:

(1) 适应性。YOLOv5s-3S-4DPH 可以适应不同的场景和任务,例如无人机视觉、行人检测、车辆检测等,而不需要针对每个任务进行特定的调整或优化。

(2) 可迁移性。YOLOv5s-3S-4DPH 可以在不同的数据集上进行有效的迁移学习,利用源数据集的知识来提高目标数据集的检测效果。

(3) 鲁棒性。YOLOv5s-3S-4DPH 可以抵抗特定的环境因素,例如光照变化、遮挡、背景干扰等,从而提高检测的准确性和鲁棒性。

3.4.5 检测效果

行人和车辆单张图片推理效果对比如图 12 所示。由图 12 可知, YOLOv5s-3S-4DPH 能有效检测远距离和被遮挡的车辆,检测效果优于 YOLOv5s。

表 6 迁移学习实验结果
Table 6 Results of the transfer learning experiment

种类	YOLOv5s			YOLOv5s-3S-4DPH		
	P/%	R/%	AP@0.5/%	P/%	R/%	AP@0.5/%
pedestrian	48.5	39.7	40.7	56.4	41.7	45.6
people	45.2	35.6	33.5	49.8	35.2	36.4
bicycle	29.1	16.9	13.8	29.0	16.7	14.5
car	64.0	73.5	74.4	67.2	80.9	81.1
van	47.5	36.9	36.8	50.7	41.8	42.2
truck	55.3	30.9	32.2	47.0	31.6	32.3
tricycle	40.7	23.1	19.9	44.2	25.1	24.6
awning-tricycle	24.0	11.6	10.4	26.7	15.4	13.3
bus	61.1	43.8	46.8	62.5	50.2	53.3
motor	48.0	43.2	39.1	53.7	45.3	44.6
mAP@0.5/%		34.8			38.8	
mAP@0.5:0.95/%		19.2			22.1	



图 12 不同算法检测部分效果对比图

Fig. 12 Comparison of detection effects of different algorithms

4 结束语

基于 YOLOv5s 提出一种改进的 YOLOv5s-3S-4DPH 目标检测算法,从主干网络、特征融合、预测头和回归损失函数 4 个方面进行改进。自建行人和车辆数据集,对比分析不同轻量化主干网络的算法检测性能,开展消融实验研究,验证了轻量化设计对模型压缩复杂度和提升模型检测速度的有效性;通过增加小目标检测层和引入权重参

数加权特征图,加强算法对小目标语义特征的提取能力,扩大小目标检测范围和提高检测精度;通过引入解耦头提高算法检测和识别精度,采用 Focal EIou 损失函数加快模型训练速度。在 Vis-Drone 2019 数据集上进行迁移学习实验,对比了 YOLOv5s 和改进算法在不同场景下的检测性能。结果表明:相较于 YOLOv5s 基准算法,本文提出的 YOLOv5s-3S-4DPH 算法在目标检测精度和速度上均有明显提升,对小目标以及被遮挡的目

标具有良好的检测效果,具有适应性、可迁移性以及鲁棒性。

参考文献:

- [1] 梁鸿,王庆玮,张千,等.小目标检测技术研究综述[J].计算机工程与应用,2021,57(1):17-28.
Liang Hong, Wang Qing-wei, Zhang Qian, et al. Small object detection technology: a review[J]. Computer Engineering and Applications, 2021, 57(1): 17-28.
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, 21(8): 91-103.
- [3] Krizhevsky A, Sutskever I, Hinton E G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] 王芋人,武德安.一种提高小目标检测准确率的数据增强方法[J].激光杂志,2021,42(11):41-45.
Wang Yu-ren, Wu De-an. Data augmentation method for improving the accuracy of small target detection[J]. Laser Journal, 2021, 42(11): 41-45.
- [5] 杨慧剑,孟亮.基于改进的YOLOv5的航拍图像中小目标检测算法[J].计算机工程与科学,2023,45(6):1063-1070.
Yang Hui-jian, Meng Liang. A small target detection algorithm based on improved YOLOv5 in aerial image[J]. Computer Engineering & Science, 2023, 45(6): 1063-1070.
- [6] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] Singh B, Davis L S. An analysis of scale invariance in object detection-SNIP[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018:3578-3587.
- [8] Zhang S, Zhu X, Lei Z, et al. Faceboxes: a CPU real-time face detector with high accuracy[C]//IEEE International Joint Conference on Biometrics, Denver, USA, 2017: 1-9.
- [9] 王建中,王加乐,于子博,等.士兵和装甲车目标多尺度检测方法[J].北京理工大学学报,2023,43(2): 203-212.
Wang Jian-zhong, Wang Jia-le, Yu Zi-bo, et al. Multi-scale detection method for soldier and armored vehicle objects[J]. Transactions of Beijing Institute of Technology, 2023, 43(2): 203-212.
- [10] 谌雨章,黄逸姿,张钧涵.基于多速率空洞卷积的多尺度水下小目标检测[J].计算机工程,2023,49(6): 257-264.
Chen Yu-zhang, Huang Yi-zi, Zhang Jun-han. Multi-scale underwater small object detection based on multi-rate dilated convolution[J]. Computer Engineering, 2023, 49(6): 257-264.
- [11] 李成豪,张静,胡莉,等.基于多尺度感受野融合的小目标检测算法[J].计算机工程与应用,2022,58(12): 177-182.
Li Cheng-hao, Zhang Jing, Hu Li, et al. Small object detection algorithm based on multiscale receptive field fusion[J]. Computer Engineering and Applications, 2022, 58(12): 177-182.
- [12] 董亚盼,高陈强,谌放,等.基于注意力机制的红外小目标检测方法[J].重庆邮电大学学报:自然科学版,2023,35(2): 219-226.
Dong Ya-pan, Gao Chen-qiang, Chen Fang, et al. Infrared small target detection method based on attention mechanism[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(2): 219-226.
- [13] Qu J S, Su C, Zhang Z W, et al. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images[J]. IEEE Access, 2020, 8: 82832-82843.
- [14] Li K, Cheng G, Bu S, et al. Rotation-insensitive and context-augmented object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(4): 2337-2348.
- [15] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779-788.
- [16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6517-6525.
- [17] Redmon J, Farhadi A. YOLOv3: an incremental improvement[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018:1-6.
- [18] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: optimal speed and accuracy of object detection[DB/OL]. [2023-06-05]. <https://arxiv.org/abs/2004.10934>
- [19] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//IEEE/CVF Confer-

- ence on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8759-8768.
- [20] Ma N, Zhang X, Zheng H T, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 116-131.
- [21] Yu C, Gao C, Wang J, et al. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129: 3051-3068.
- [22] 陈奎, 刘晓, 贾立娇, 等. 基于轻量化网络与增强多尺度特征融合的绝缘子缺陷检测[J]. 高压技术, 2024(3):1289-1300.
- Chen Kui, Liu Xiao, Jia Li-jiao, et al. Insulator defect detection based on lightweight network and enhanced multi-scale feature fusion[J]. 高压技术, 2024(3):1289-1300.
- [23] Tan M, Pang R, Le A V. EfficientDet: scalable and efficient object detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020:10778-10787.
- [24] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3):391-417.
- Gao Xin-bo, Momeng Jing-cheng, Wang Hai-tao, et al. Recent advances in small object detection[J]. Journal of Data Acquisition and Processing, 2021, 36(3):391-417.
- [25] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition Honolulu, USA, 2017: 2117-2125.
- [26] Song G, Liu Y, Wang X. Revisiting the sibling head in object detector[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11560-11569.
- [27] Ge Z, Liu S, Wang F, et al. YOLOX: exceeding YOLO series in 2021[DB/OL]. [2023-06-10]. <https://arxiv.org/abs/2107.08430>
- [28] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [29] 徐慧智, 宋爱秋, 武笑宇. 基于均匀设计的船舶目标检测深度学习模型训练方法[J]. 科学技术与工程, 2022, 22(25): 11241-11249.
- Xu Hui-zhi, Song Ai-qiu, Wu Xiao-yu. Training method of deep learning to ship target detection based on uniform design[J]. Science Technology and Engineering, 2022, 22(25): 11241-11249.
- [30] 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法[J]. 航空学报, 2023, 44(5): 273-283.
- Mao Guo-tao, Deng Tian-min, Yu Nan-jing. Object detection in UAV images based on multiscale split attention[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(5): 273-283.