文章编号：1674-2974（2024）06-0029-11 DOI：10.16339/j.cnki.hdxbzkb.2024264

基于改进YOLOv5的鱼眼图像目标检测算法

韩彦峰†，任奇，肖科
（重庆大学 机械与运载工程学院，重庆 400044）

摘 要：针对自动驾驶场景下车载鱼眼相机采集到的图像存在畸变严重、场景复杂、尺度变化剧烈、小目标多以及传统的目标检测模型的检测精度不高的问题，提出了一种基于YOLOv5s改进的鱼眼图像检测模型YOLOv5s-R.首先，为解决小目标难识别的问题，提出随机裁剪多尺度训练的数据增强方法，该方法优于消融实验所得的最优数据增强方法.其次，为了提高模型的检测精度，在网络头部添加置换注意力机制与轻量化解耦头，增强模型对特征的提取能力与识别能力，并抑制噪声干扰.最后，模型额外增加角度预测项，实现旋转框目标检测.通过构建环形标签并用高斯函数对标签平滑，解决了旋转框角度的周期性问题；又对损失函数进行了优化，提出了RIOU，在CIOU的基础上增加角度惩罚项，提高了回归精度并加快了模型的收敛.实验结果表明，提出的YOLOv5s-R模型在WoodScape数据集上取得良好的检测效果，相比于原始的YOLOv5s模型，mAP@0.5、mAP@0.5∶0.95分别提升了6.8%、5.6%，达到82.6%、49.5%.

关键词：YOLOv5s；自动驾驶；鱼眼图像；旋转框目标检测
中图分类号：TP391 **文献标志码**：A

Object Detection Algorithm for Fish Eye Image Based on Improved YOLOv5

HAN Yanfeng†，REN Qi，XIAO Ke
（College of Mechanical and Vehicle Engineering，Chongqing University，Chongqing 400044，China）

Abstract：The images collected by fish eye cameras in autonomous driving scenarios have severe distortion，complex scenes，drastic scale changes，and many small targets，which lead to low detection accuracy of traditional object detection models. Therefore，YOLOv5s-R，an improved fish eye image detection model based on YOLOv5s，is proposed. Firstly，to solve the problem of difficult recognition of minor targets，the RCMS（Random Crop Muti Scale）data augmentation method is proposed，which performs better than the optimal data augmentation method obtained from ablation experiments. Secondly，to improve the detection accuracy of the model，SA（Shuffle Attention）and LDH（Light Decouple Head）modules are added to the network header to enhance the model's feature extraction and recognition capabilities，suppress noise interference. Finally，an additional angle prediction branch is added to realize the rotating box object detection，a circular label is constructed to solve the PoA（Periodicity of Angular）problem，and the label is smoothed with the Gaussian function. The RIOU is proposed to

* **收稿日期**：2023-07-23
基金项目：国家重点研发计划课题（2022YFB4702201），National Key R&D Program of China（2022YFB4702201）
作者简介：韩彦峰（1987—），男，山西忻州人，重庆大学副教授，博士
† 通信联系人，E-mail：fyh-0220@163.com

optimize the loss function by adding an angle penalty term on the basis of CIOU, which improves the regression accuracy and speeds up the convergence of the model. The experimental results show that the proposed YOLOv5s-R model achieves good detection performance on the Woodscape dataset. Compared to the original YOLOv5s model, mAP@0.5 mAP@0.5 is 0.95 increased by 6.8% and 5.6%, respectively, reaching 82.6% and 49.5%.

Key words: YOLOv5; autonomous driving; fish eye images; rotating box object detection

随着人工智能算法的不断创新发展,自动驾驶汽车受到越来越多的关注[1].作为未来汽车的发展方向,自动驾驶汽车拥有自主判断能力,能较大限度地减少人为失误.对实时目标的精准检测、识别并做出判断是保证自动驾驶汽车决策和判断的基础与核心[2].由于道路交通环境复杂,基于视觉的实时检测任务仍面临诸多挑战.

目前,基于深度学习的目标检测算法是主流[3],主要可以分为双阶段算法和单阶段算法两类.双阶段算法以RCNN系列[5-8]为代表,这类算法虽然精度高,但检测速度慢,不能满足实时性的要求.单阶段算法以YOLO系列[9-16]为代表,其优势是速度快,早期的单阶段算法精度不高,但随着后续的不断完善,在特定场景下的检测精度甚至优于双阶段算法.以YOLOv5为节点,后续的YOLO系列对特定模块做了优化,但其本质都是FPN+PAN的结构,对特定场景下的目标检测任务效果提升有限.如YOLOv6使用RepVGGBlock替换BottleNeck Block来作为主要的梯度流分支,YOLOv7使用ELAN Block替换BottleNeck Block,YOLOv8将C3模块优化为C2f模块,并取消了上采样前面的1×1卷积,百度提出的PP-YOLOE[17]则使用RepRes Block来替换BottleNeck Block作为主要的梯度流分支.

针对鱼眼图像的目标检测,主要有两种做法,一种是在图像畸变校正后进行检测,这类方法有一个先验知识,即假定鱼眼镜头的几何畸变是已知的.另一种是模型自适应畸变图像.由于基于图像畸变矫正的检测方法较烦琐,且泛化性较差,大多研究更关注模型自适应的检测方法.Li等[18]提出了一个自学习和基于轮廓检测器的鱼眼图像目标检测算法FisheyeDet,通过FM模块将不同分辨率大小的特征图融合,MFCM模块获取多尺度特征,但其骨干网络仍采用较老的VGG16;Kumar等[19]提出了一个多任务视觉感知网络OmniDEet,该研究表明联合训练的模型比各自单个任务版本表现得更好;Rashed等[20]提出针对鱼眼图像的网络FisheyeYOLO,该网络基于YOLOv3改进,可以输出不同形式的目标检测框,这给我们启发:旋转框目标检测能更好地适配鱼眼图像的畸变.有研究者对卷积核进行改进以适应畸变图像:Coors等[21]提出了SphereNet,通过将失真不变性编码到细胞神经网络的架构中,根据球面图像表示的几何形状调整卷积滤波器的采样网格位置,从而避免失真.Zhang等[22]提出Spherical Unet先采用等角矩形投影方式矫正图像后再使用球面卷积,实现了对全景视频进行显著性检测.Su等[23]提出了内核变换网络KTN,以有效地将透视图像中的卷积内核转换为360的等矩形投影;然而,球形模型不能为鱼眼透镜的复杂畸变提供精确的拟合,且复杂的卷积核会增加大量的运算量.也有的研究者根据鱼眼图像的不同位置的特征畸变特点,针对性地设计了不同的方法来处理:Dooley等[24]针对车辆在鱼眼图像边缘畸变严重,根据目标车辆的距离采取不同方法.Baek等[25]根据图像中心与边缘畸变程度不同,将完整的鱼眼图像分成左中右三个感兴趣区域,分别训练不同的模型来检测,但这类方法都较为繁琐.以上模型都存在检测精度低、计算复杂、不能满足自动驾驶场景下对实时性检测要求的问题.

YOLOv5目标检测算法发展至今,在通用目标检测数据集上的检测精度已经达到了较高的水平,但对于特定的任务,通常需要重新对网络结构进行优化.为提升YOLOv5在复杂海域中目标检测能力,王文亮等[26]提出多路径聚合网络结构MPANet,在自底向上特征传递过程中融合多层次特征信息以增强多尺度定位能力;秦晓辉等[27]提出了CDDPP下采样方法,解决了传统下采样丢失特征的问题,同时改进预测头,让每个分支专注自身任务,提高了对露天矿山

目标的检测精度;杨睿宁等[28]提出 EioU 损失,将宽高比的惩罚项拆分成预测框与真实框宽、高之差的惩罚项,提出上采样算子 CARAFE 减少上采样信息的损失,构建出适用于复杂交通场景下路侧目标检测任务的模型;刘春霞等[29]提出全局上下文特征提取模块 BoT3,提出 SIoU 损失函数避免了预测框的游荡现象,提高了模型对煤矿烟火的检测能力;杨晨等[30]针对小目标检测较困难的问题,提出了 Dense-DarkNet 和 EMFFN 有效融合不同尺度的特征图信息,并设计了一种新型的注意力机制 BFAM 减少噪声干扰,有效地提高了模型对遥感影像的目标检测效果.这些研究对 YOLOv5 的改进主要聚焦于:1)特征信息的流动方式.2)特征提取模块与注意力机制的构建.3)损失函数的优化.尽管这些改进算法在常规图像的检测取中得很好的效果,但并不适用畸变严重的鱼眼图像.

因此,本文提出一种适用于鱼眼图像的旋转框目标检测算法 YOLOv5s-R,在保证检测速度的同时提高了检测精度.本文的贡献如下:

1)提出随机裁剪多尺度数据增强方法,提高了模型对小目标和不同尺度目标的检测能力.

2)在 YOLOv5s 头部添加置换注意力机制与轻量化解耦头,增强模型对特征提取能力与识别能力,并抑制噪声干扰,提高了检测精度.

3)检测框增加角度预测,构建环形标签解决了角的周期性问题,提出 RIOU,在 CIOU 的基础上增加角度惩罚项,提高了回归精度并加快模型收敛.

1　自动驾驶数据集构建

1.1　数据集构建

研究使用公开的 WoodScape 数据集[31],该数据集通过车载环绕的四个鱼眼摄像头采集,得到前视(Front View, FV)、后视(Rear View,RV)、左视(Mirror Lelf View,MVL)和右视(Mirror Right View,MVR)图像,分辨率为 1 280×966,如图 1 所示.采集地点为美国、欧洲和中国,驾驶场景分为高速公路、城市驾驶和停车用例,包含五类目标:车辆(vehicle)、行人(person)、自行车(bicycle)、交通灯(traffic light)、交通符号(Traffic sign).统计各类别的标签数量,分别为 44 404、16 249、7 053、1 500、2 857,各类别物体示例如图 2 所示.由于标签存在漏标和误标,对部分图像重新标注.数据集共 8 234 张,按 95∶5 划分训练集和验证集,即训练集图像 7 822 张,验证集图像 412 张.

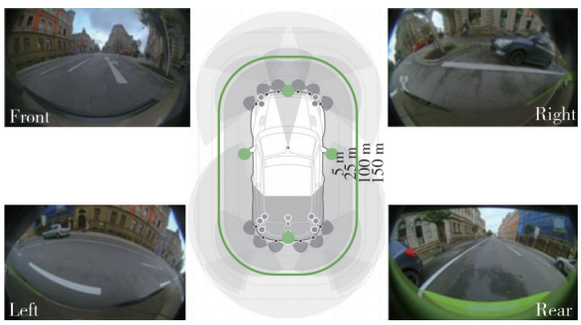


图 1　四个环视鱼眼相机构成的环视采集网络

Fig.1　Four fisheye camera images forming the surroundview camera network



图 2　不同类别目标示例

Fig.2　Examples of different categories of goals

对 WoodScape 数据集进一步分析,发现存在以下特点:以小目标为主;样本不均衡,车辆和行人居多;大部分目标分布在镜头周围;边缘目标存在角度倾斜等.在后续的研究中,主要根据这些特点来改进网络模型.

1.2　数据增强方法

为了提升模型的泛化能力,通常会使用数据增强方法来丰富数据集.常用的数据增强方法有 HSV 变换、旋转、平移、缩放、错切、透视、翻转、拼接、混合、复制粘贴.而要找到适合数据集的最优增强策略,需要做消融实验,流程如图 3 所示.

WoodScape 数据集以小目标为主,大多的数据增强方法都不能很好地提升模型的检测能力,因此提出一种针对小目标检测的随机裁剪多尺度(Random Crop Multi Scale, RCMS)数据增强方法,如图 4 所示.将原图以不同尺度(图中不同颜色框)随机裁

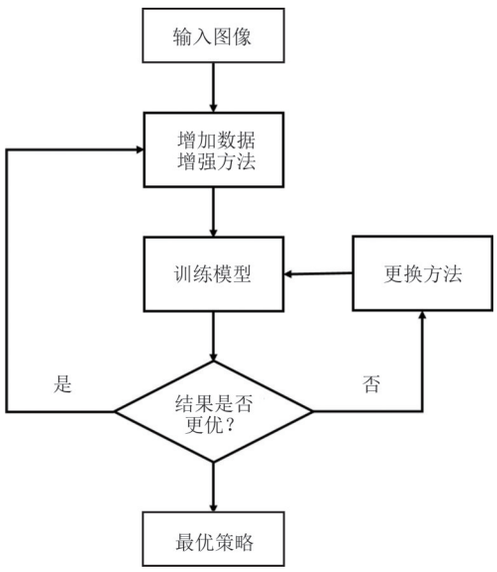


图3 消融实验流程

Fig.3 Ablation experimental process

剪图像,再缩放到统一的尺寸用于训练.该方法起着“放大镜”的作用,能使模型增强对小物体的特征的提取能力,多尺度则可以增强模型辨别不同大小物体的能力.

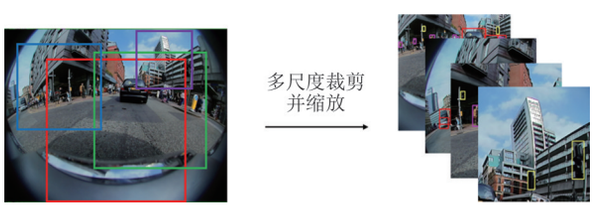


图4 随机裁剪多尺度方法实现

Fig.4 Implementation of random crop multi scale method

2 旋转框目标检测网络YOLOv5-R构建

2.1 网络结构改进

2.1.1 改进后的网络结构

YOLOv5根据模型的大小分为YOLOv5x、YOLOv5l、YOLOv5m和YOLOv5s,其中YOLOv5s参数量和计算量最小,推理速度最快,考虑到训练数据集大,训练资源有限,因此本文选用YOLOv5s作为改进模型,理论上改进方法同样适用于其他模型.模型主要由三部分构成:骨干、颈部、头部.骨干负责特征提取,占主要运算量;颈部负责特征融合,以提高模型的预测能力;头部输出检测结果,分为大、中、小三个尺度以检测不同大小的物体.

针对YOLOv5s存在小目标难识别、定位框精度低、误检与漏检的问题,对其改进,在原始YOLOv5s的基础上添加置换注意力机制(Shuffle Attention, SA)并构建轻量化解耦头(Light Decouple Head, LDH)以提升模型提取特征与识别特征的能力,改进后的网络结构如图5所示.

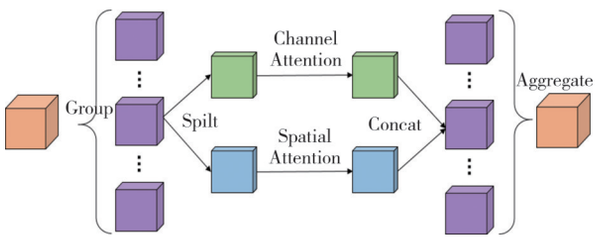


图5 SA模块

Fig.5 SA module

2.1.2 SA注意力机制

自动驾驶场景复杂,待检测目标小导致特征不明显,存在大量干扰,从而降低了模型识别精度.为了解决这一问题,在检测头位置加入SA[32]注意力机制,以抑制干扰因素,提高模型的特征提取能力,其结构如图6所示.

SA注意力机制首先将输入的特征图 $X \in R^{C\times H\times W}$ 进行分组操作,分为 G 组,即 $X=[X_1,\cdots,X_G]$, $X_k \in R^{C/G\times H\times W}$.对于每一组再划分(Spilt)为两个分支 X_{k1} 与 X_{k2}, $X_{k1},X_{k2}\in R^{C/2G\times H\times W}$.$X_{k1}$ 分支使用通道注意力机制,关注通道特征间的相关性,具体做法为:首先使用全局平局池化生成通道的权重信息 s,如式(1)所示,再对 s 做简单的非线性变换后与原通道特征相乘,如式(2)所示.X_{k2} 分支使用空间注意力机制,关注特征图的空间区域的相关性,具体做法为:对特征图做组归一化(Group Normalization, GN)获得空间特征,然后做非线性变化后与原特征图相乘,如式(3)所示.最后将两个分支拼接融合后得到每组的特征图,再将所有组的特征图都聚合得到输出的特征图.

$$s=\frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W}X_{k1}(i,j) \tag{1}$$

$$X'_{k1}=\sigma(W_1 s+b_1)\cdot X_{k1} \tag{2}$$

$$X'_{k2}=\sigma(W_2\cdot \mathrm{GN}(X_{k2})+b_2)\cdot X_{k2} \tag{3}$$

式中:H、W 分别为特征图的宽度和高度;i,j 为具体的元素;W_1,W_2 为线性变换的权重;b_1,b_2 为偏置;σ 为sigmoid激活函数.

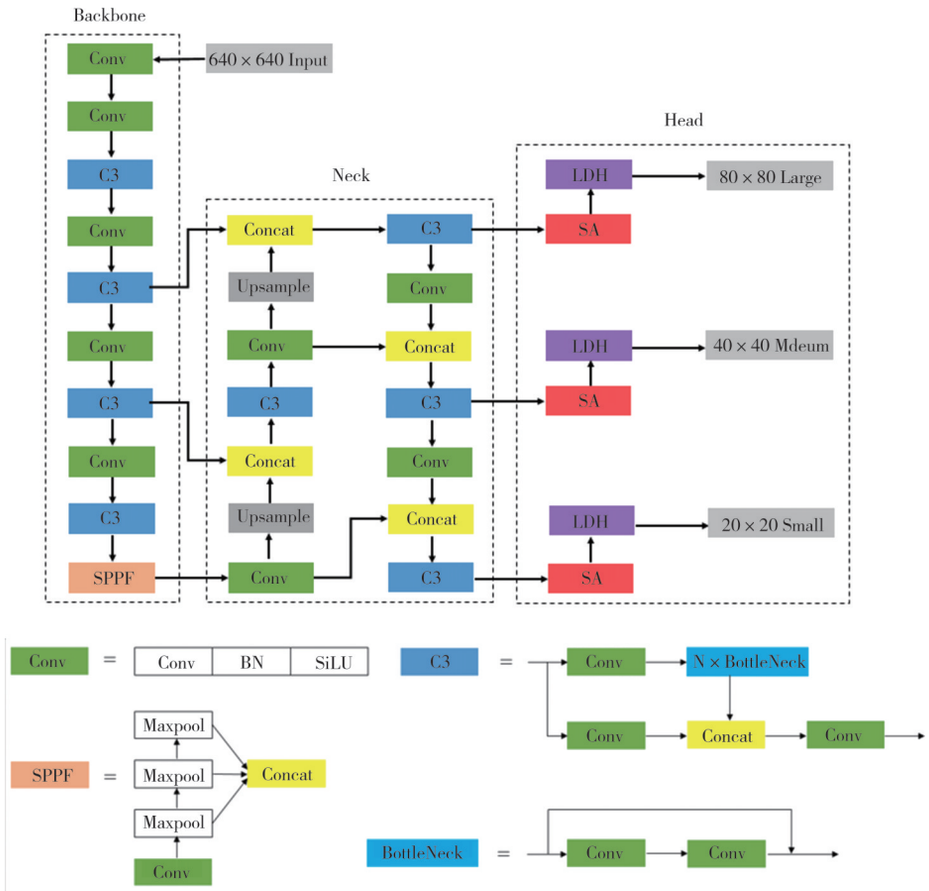


图 6　改进后的 YOLOv5s 网络结构

Fig.6　Improved YOLOv5s network structure

2.1.3　轻量化解耦头

相关研究表明[33],分类任务和定位任务关注的重点不同,分类任务更加关注所提取的特征与已有类别哪一类最为相近,而定位任务更加关注与真实预测框的位置坐标从而进行边界框参数修正.因此,可以在模型的头部添加解耦头来实现对各类任务更精准的判别,但增加解耦头的同时也增加了运算量,导致推理速度变慢.为了解决这一问题,参照 YOLOX[34]的解耦头,设计了一种轻量化的解耦头,用一个卷积模块替换原来的多个卷积模块以减少运算量,同时将通道数从 256 减少为 64.若增加角度预测项则额外添加角度模块,如图 7 所示.

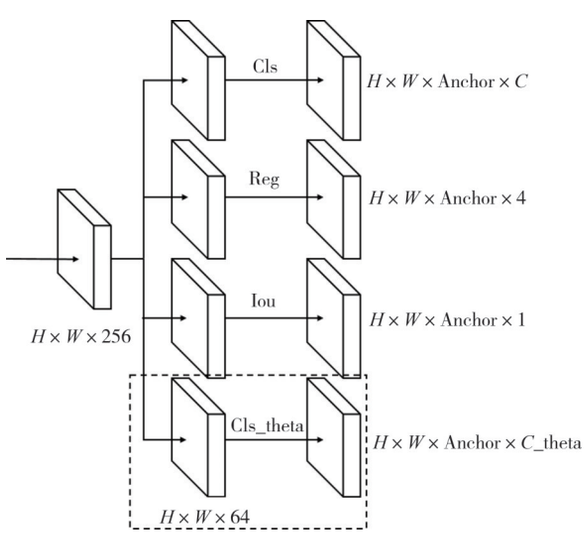


图 7　LDH 模块

Fig.7　LDH module

2.2　环形标签构建

使用长短边法构建旋转框,可以避免边的可交换问题(Exchangeability of Edges, EoE),预测的旋转框表示为(x, y, l, s, θ),其中(x, y)为旋转框的中心点坐标,l表示长边,s表示短边,θ表示长边与x轴的夹角,$\theta\in[0, 180)$.但长短边法存在角的周期性(Periodicity of Angular, PoA)问题,在边界处的角度损失会急剧震荡,网络难以训练,如图 8 所示.

为了解决 PoA 问题,引入环形标签(Circular Smooth Label, CSL)[35],将回归任务转为分类任务.角度分类数越多,预测角度越细致,但同时也会带来预测头厚重的问题;角度分类数越少,预测角度越粗

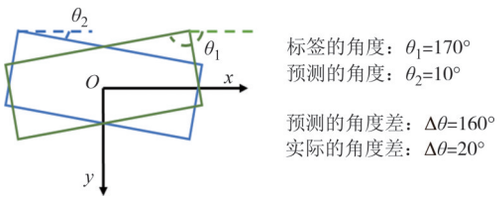


图 8 PoA 问题

Fig.8 PoA problem

糙.若按1°分为一类,则可以分为180类,环形标签首尾相连,消除了边界处的角度损失震荡,如图9所示.

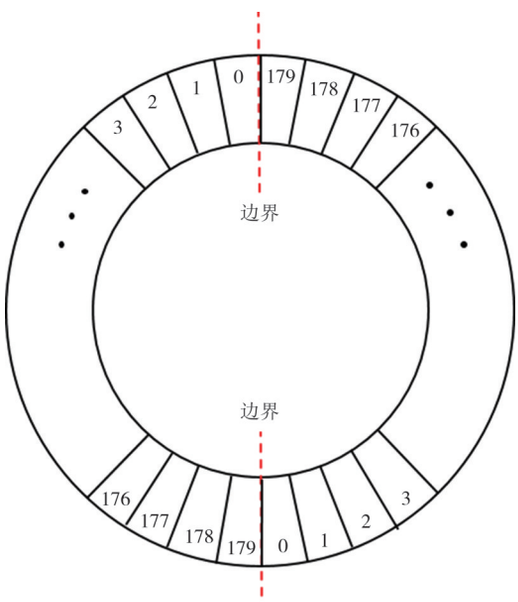


图 9 环形标签

Fig.9 Circular smooth label

使用高斯函数对标签平滑处理,以提高模型对预测角度的预测的泛化能力,高斯函数表达式为:

$$y = \mathrm{e}^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{4}$$

式中:μ为函数峰值出现的位置;σ为标准差,控制函数的峰形宽度.

当使用高斯函数对角度编码时,μ为实际的角度类别,x为其他任意的角度类别,根据角度的分类数设置合理的σ,代入公式可求得所有位置的平滑值.不同角度分类数c、不同标准差σ、不同角度标签值θ得到的高斯平滑标签如图10所示.

2.3 损失函数优化

YOLOv5s-R的损失函数由四项构成:IoU(Intersection over Union)损失、置信度损失、分类损失与角度损失.

YOLOv5s的原始IoU损失计算采用的是CIOU,它同时考虑了预测框与真实框的重叠面积、中心点

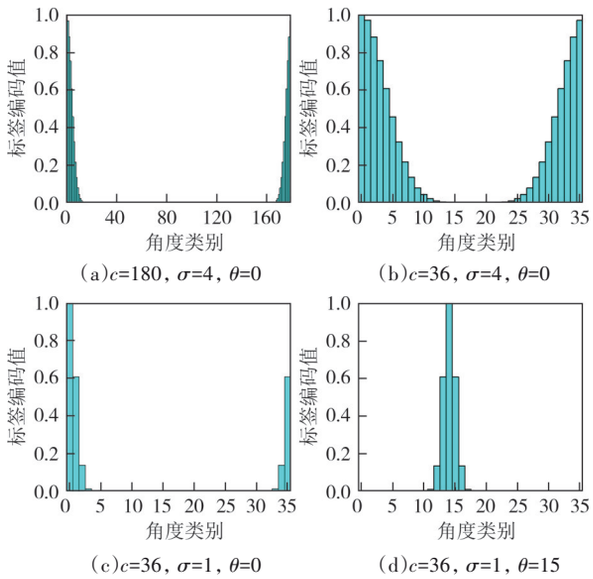


图 10 高斯标签平滑

Fig.10 Gaussian label smoothing

距离和长宽比.CIOU损失的公式为:

$$L_{\mathrm{CIOU}} = 1 - \mathrm{IoU} + \frac{\rho^2(p, p^{\mathrm{gt}})}{c^2} + \alpha v \tag{5}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}} - \arctan\frac{w}{h}\right)^2 \tag{6}$$

$$\alpha = \frac{v}{1 - \mathrm{IoU} + v} \tag{7}$$

式中:$\rho^2(p, p^{\mathrm{gt}})$表示预测框与实际框的中心点的距离;$c$表示包含预测框与实际框的最小外接矩形的对角线段距离;$\rho^2(p, p^{\mathrm{gt}})/c^2$即为距离惩罚项,它的作用是当预测框与真实框相互包含时,仍能提供梯度更新方向;v为真实边框与预测边框的宽高比损失;α为宽高比损失系数;w^{gt}和h^{gt}表示真实框的宽高;w和h表示预测框的宽高.

CIOU没有考虑到预测框与真实框的角度差,为了进一步优化旋转框的检测效果,提出RIOU,引入角度惩罚项:

$$L_{\mathrm{RIOU}} = 1 - \mathrm{IoU} + \frac{\rho^2(p, p^{\mathrm{gt}})}{c^2} + \alpha v + k\theta \tag{8}$$

式中:k为平衡系数,约束角度惩罚项,防止RIOU损失过大导致对其他损失的不敏感,实验中设置为0.1;θ为预测框的角度与真实框的角度差值,考虑到边界问题,θ的表达式为

$$\theta = \begin{cases} \dfrac{|\theta_{\mathrm{pred}} - \theta_{\mathrm{target}}|}{c}, |\theta_{\mathrm{pred}} - \theta_{\mathrm{target}}| \leqslant \dfrac{c}{2} \\ 1 - \dfrac{|\theta_{\mathrm{pred}} - \theta_{\mathrm{target}}|}{c}, |\theta_{\mathrm{pred}} - \theta_{\mathrm{target}}| > \dfrac{c}{2} \end{cases} \tag{9}$$

式中:θ_{pred}为预测的角度类别;θ_{target}为实际的角度类别;c为角度的分类数.对于环形标签,预测角度与目

标角度的差值的绝对值应小于分类数的一半,否则应从环形的另一边计算角度差.最后将角度差除以分类数c归一化到0到1之间,以保证角度分类数不同时的损失相同.

相比于CIOU,RIOU考虑了角度信息,如图11所示,能够更准确地衡量预测框(图中红色框)与真实框(图中绿色框)的差距.

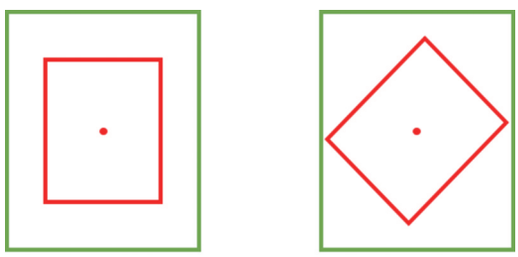


图 11　CIOU 损失与 RIOU 损失对比

Fig.11　Loss comparison of CIOU and RIOU

置信度损失为二值交叉熵损失,分类损失和角度损失均采用多分类交叉熵损失:

$$L_{obj} = -[\hat{C}\log(C) + (1 - \hat{C})\log(1 - C)] \tag{10}$$

$$\hat{C} = (1 - g) + g*\text{RIOU} \tag{11}$$

$$L_{cls/\theta} = \sum -p_i * \log(\hat{p}_i) \tag{12}$$

式中:$\hat{C}$为预测的置信度;C为真实的置信度标签;$\hat{C}$是1和CIOU的加权平均;g为训练难度系数;p_i表示属于第i类的标签值;$\hat{p}_i$是预测的概率.

综合以上四项损失,得到YOLOv5-R模型的总损失:

$$L_{all} = L_{RIOU} + L_{obj} + L_{cls} + L_{\theta} \tag{13}$$

3　实验结果与分析

3.1　实验环境

实验环境为Ubuntu20.04系统,CPU为Intel Xeon E5-2620v3 2.4 GHz,GPU为NVIDIA GeForce GTX 3090,PyTorch版本为1.10.1,CUDA版本为11.3,Python版本为3.8.

使用YOLOv5s模型,进入网络训练的图片大小都被设置为640×640,初始学习率设置为0.01,使用余弦退火动态调整学习率,学习率动量为0.937,权重衰减系数为0.000 5,批量大小为16,使用SGD优化器,预训练1轮后训练50轮,为使实验严谨,训练过程中固定了随机种子数.

3.2　评价指标

平均精度(Average Precision,AP)能同时反映定位能力与分类能力,相比于准确率(Precison,P)与召回率(Recall,R),更能反映模型的检测能力.采用mAP@0.5与mAP@0.5∶0.95作为模型检测精度的评价指标.AP@0.5指的是预测框与真实框IoU阈值大于0.5且分类正确的平均精度,通过计算PR曲线的面积得到.mAP@0.5是所有类别的AP@0.5的平均,mAP@0.5∶0.95是以5为间隔,mAP@0.5到mAP@0.5∶0.95的平均值.具体公式如下:

$$\text{AP} = \int_0^1 P(R)\mathrm{d}R \tag{14}$$

$$\text{mAP@0.5} = \frac{1}{n}\sum_{i=1}^{n}\text{AP@0.5}(i) \tag{15}$$

$$\text{mAP@0.5}:0.95 = \frac{1}{10}(\text{mAP@0.5} + \cdots + \text{mAP@0.95}) \tag{16}$$

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

式中:n为类别数;TP表示实际是正样本预测也为正样本的数量;FP表示实际是正样本预测为负样本的数量;FN表示实际是负样本预测为正样本的数量.

3.3　数据增强方法的消融实验与对比实验

3.3.1　消融实验

为了寻找最优的数据增强方法,使用常用的数据增强方法做消融实验.一般来讲,使用HSV的数据增强方法对提升模型的检测能力总是有益的,因此将其作为基准,实验结果如表1所示.

表 1　消融实验结果

Tab.1　Results of ablation experiment

数据增强方法	mAP@0.5	mAP@0.5∶0.95
基准(HSV)	0.687	0.434
旋转	0.712	0.414
旋转+平移	0.729	0.427
旋转+平移+缩放	0.719	0.405
旋转+平移+翻转	0.758	0.439
旋转+平移+错切	0.756	0.433
旋转+平移+错切+拼接	0.749	0.430
旋转+平移+错切+混合	0.746	0.431
旋转+平移+错切+复制粘贴	0.750	0.433

由表1可以看出,采用旋转+平移+翻转的组合增强方式是最优策略.分析原因:WordScape数据集的目标大多都带有一定角度,旋转增强了对旋转物体的检测能力,而平移和翻转一定程度上改变了目标的分布,丰富了数据集.

3.3.2 引入RCMS方法对比实验

为了验证RCMS方法的有效性,与消融实验所得的最优策略(Optimal Method, OM)做对比,实验结果如表2所示.

表2 OM方法和RCMS方法对比实验

Tab.2 Comparative experiment of OM and RCMS

数据增强方法	mAP@0.5	mAP@0.5:0.95
基准(HSV)	0.687	0.434
OM	0.758	0.439
RCMS	0.761	0.477
OM+RCMS	0.750	0.441

由表2可以看出,使用RCMS方法在mAP@0.5可以达到与OM方法差不多的得分,而mAP@0.5:0.95提升了3.8%,这意味着检测框的定位更准确,证明了该方法的有效性.此外,OM方法与RCMS方法存在冲突,效果反而下降.

3.4 模型改进的实验

3.4.1 引入注意力机制的实验

为了探究注意力机制位置对模型精度的影响,以使用RCMS方法训练YOLOv5s原始模型的结果为基准,分别在C3层、SPPF层、Head层添加SA注意力机制,实验结果如表3所示.

表3 不同位置SA注意力机制对比

Tab.3 Comparison of SA module in different positions

位置	mAP@0.5	mAP@0.5:0.95
无	0.761	0.477
C3	0.724	0.440
SPPF	0.738	0.447
Head	0.791	0.483

由表3可以看出,注意力机制添加在头部(Head)效果最好.分析原因:在训练时使用了官方权重文件,注意力机制添加在C3层或SPPF层会破坏网络结构,导致大部分预训练权重无法加载,而添加到Head层能加载大部分预训练权重,充分利用先验知识.

为了探究不同注意力机制的效果,使用常用的SENet、CBAM、CA、SA四种注意机制添加到头部做对比实验,实验结果如表4所示.

由表4可以看出,使用注意力机制大都能提升模型的检测能力,其中SA注意力机制的表现最好,且引入的参数量与运算量最少.使用SA注意力机制后,mAP@0.5、mAP@0.5:0.95分别提升了3.0%、0.6%.

表4 不同注意力机制对比

Tab.4 Comparison of different attention mechanisms

方法	mAP@0.5	mAP@0.5:0.95	参数/M	运算量/G
无	0.761	0.477	7.03	15.88
SENet	0.773	0.483	7.08	15.89
CBAM	0.769	0.474	7.08	15.89
CA	0.755	0.469	7.07	15.89
SA	0.791	0.483	7.03	15.88

3.4.2 引入轻量化解耦头的实验

为了验证解耦头的有效性,以使用SA注意力机制的结果为基准,使用原始的解耦头(Decouple Head, DH)和优化后的轻量化解耦头(LDH)做对比实验,实验结果如表5所示.

表5 DH与LDH对比

Tab.5 Comparison of DH and LDH

方法	mAP@0.5	mAP@0.5:0.95	参数/M	运算量/G
SA	0.791	0.483	7.03	15.88
+DH	0.788	0.488	7.77	56.42
+LDH	0.794	0.487	7.19	19.99

由表5可以看出,添加LDH后,mAP@0.5、mAP@0.5:0.95分别提升了0.3%、0.4%,这证明了解耦头的有效性.相比于DH,LDH的参数量和运算量大幅减少且表现更优.

3.5 引入旋转框的实验

3.5.1 标签平滑策略探究的实验

为了构建最优的标签编码格式,探究角度分类数和高斯函数的标准差对检测结果的影响,在改进模型的基础上增加角度预测项,实验结果如表6所示.

表6 不同高斯平滑标签对比

Tab.6 Comparison of different Gaussian smoothing labels

角度类别数	标准差	mAP@0.5	mAP@0.5:0.95
180	4	0.804	0.482
180	2	0.795	0.476
90	2	0.800	0.478
36	2	0.798	0.476
36	1	0.810	0.481
18	1	0.801	0.481

由表6可以看出,角度分为36类、标准差为1时,模型效果最好,mAP@0.5、mAP@0.5:0.95分别为0.810、0.481.分析原因:角度类别过多,会使预测头厚重,模型对角度参数不敏感;角度类别数过少则预测的角度粗糙,预测框偏差大;标准差的大小会影响

标签平滑范围,同样也影响模型对角度的判断;因此需要同时设置合理的角度类别和标准差.额外增加角度预测对参数量与运算量的影响微乎其微,因此不讨论.

3.5.2　引入 RIOU 的对比实验

为了验证 RIOU 的有效性,在采用最优标签编码格式下,分别使用 CIOU 和 RIOU 做对比实验,实验结果如表 7 所示.

由表 7 可以看出,使用 RIOU 作为 IoU 损失后,相比 CIOU 的 mAP@0.5、mAP@0.5∶0.95 分别提升了 1.6%、1.4%,这证明了 RIOU 损失的优越性.

表 7　CIOU 与 RIOU 对比

Tab.7　Comparison of CIOU and RIOU

IoU	mAP@0.5	mAP@0.5∶0.95
CIOU	0.810	0.481
RIOU	0.826	0.495

3.6　各模型的检测效果对比

为了更直观地展示改进后模型的性能,以车载鱼眼相机的前视、后视、左视、右视四个视角的图像为测试图例,对原始模型 YOLOv5s、改进后的水平框检测模型 YOLOv5s-H、改进后的旋转框检测模型 YOLOv5s-R 的检测结果进行对比,分别如图 12~图 15 所示.



(a)原始 YOLOv5s 检测效果　　(b)YOLOv5s-H 检测效果　　(c)YOLOv5s-R 检测效果

图 12　各模型对前视图的检测效果

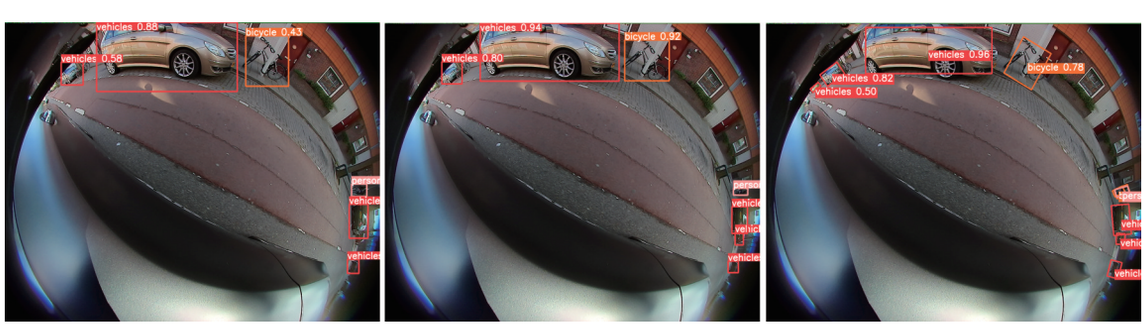
Fig.12　The detection results of different models on FV images



(a)原始 YOLOv5s 检测效果　　(b)YOLOv5s-H 检测效果　　(c)YOLOv5s-R 检测效果

图 13　各模型对后视图的检测效果

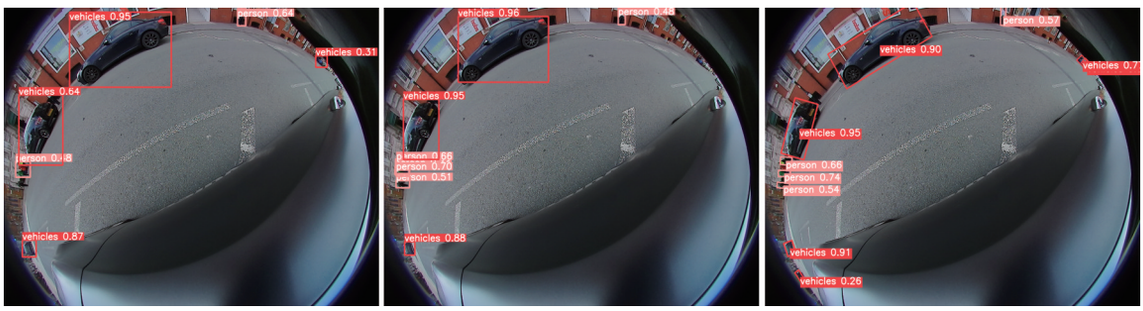
Fig.13　The detection results of different models on RV images



(a)原始 YOLOv5s 检测效果　　(b)YOLOv5s-H 检测效果　　(c)YOLOv5s-R 检测效果

图 14　各模型对左视图的检测效果

Fig.14　The detection results of different models on MVL images



(a)原始 YOLOv5s 检测效果 (b)YOLOv5s-H 检测效果 (c)YOLOv5s-R 检测效果

图 15 各模型对右视图的检测效果

Fig.15 The detection results of different models on MVR images

可以看到,原始的 YOLOv5s 模型定位精度和分类精度较低,存在漏检和误检的情况;水平框检测器 YOLOv5s-H 有效地优化了这一问题,检测框更紧凑,识别物体的能力更强,但对旋转物体的检测效果仍不佳,尤其是对左视图和右视图的检测存在框选出大量背景的情况;旋转框检测器 YOLOv5s-R 则更适配鱼眼图像的畸变特性,检测框的角度倾斜可以很好地适配畸变目标,与水平框相比,检测框与实际目标的 IOU 更小,检测效果更好.

以 mAP@0.5 为评价指标,各模型对不同类别检测结果如表 8 所示.可以看到,提出的改进模型 YOLOv5s-R 对各个类别检测精度均有较大的提升.此外,各模型对车辆和行人的检测精度较高,而对自行车、交通灯、交通符号的检测精度较低,这可能是因为这些类别的目标样本少且小目标多.

表 8 各模型在 mAP@0.5 的检测结果对比

Tab.8 Comparison of detection mAP@0.5 results of different models

模型	车辆	行人	自行车	交通灯	交通符号
YOLOv5s	0.906	0.881	0.757	0.591	0.658
YOLOv5s-H	0.919	0.876	0.771	0.696	0.710
YOLOv5s-R	0.925	0.895	0.784	0.718	0.727

4 结 论

首先,针对 WoodScape 数据集以小目标为主的特点,提出 RCMS 数据增强方法,为了验证该方法的有效性,与常用的数据增强方法消融实验的最优方法对比,结果优于最优方法.接着,为了进一步提升模型的检测效果,基于 YOLOv5s 模型进行改进,使用了 SA 注意力机制与 LDH 解耦头以提升模型对特征提取与识别的能力.最后,为适配鱼眼相机的畸变特性,构建了旋转框目标检测器 YOLOv5s-R,使用高斯函数平滑的环形标签解决 PoA 问题,并优化了损失函数,提出 RIOU 作为 IoU 损失,实现了检测框对角度的回归.本文提出的 YOLOv5s-R 模型对自动驾驶场景下的鱼眼图像的目标有很好的检测效果,与原始的 YOLOv5s 模型相比,mAP@0.5、mAP@0.5:0.95 分别提升了 6.8%、5.6%.本文研究的代码开源,供读者参考:https://github.com/renqi1/yolov5_woodscape.下一步的工作将考虑两个方面:一方面,进一步优化目标检测的精度,例如可以尝试更大的模型、对模型的优化、对数据的预处理.另一方面,可以构建多任务网络,同时实现目标检测、车道线检测、可行驶区域检测等任务,实现对行车环境的多任务实时检测.

参考文献

[1] 余世全. 无人驾驶汽车的发展现状和相关建议[J]. 专用汽车,2023(1):11-14.
YU S Q. Development status of driverless vehicles and relevant suggestions[J]. Special Purpose Vehicle, 2023(1):11-14.(in Chinese)

[2] 王艺帆. 自动驾驶汽车感知系统关键技术综述[J]. 汽车电器,2016(12):12-16.
WANG Y F. Overview on key technology of perceptual system on self-driving vehicles[J]. Auto Electric Parts, 2016(12):12-16.(in Chinese)

[3] 茅智慧,朱佳利,吴鑫,等. 基于 YOLO 的自动驾驶目标检测研究综述[J]. 计算机工程与应用,2022,58(15):68-77.
MAO Z H, ZHU J L, WU X, et al. Review of YOLO based target detection for autonomous driving[J]. Computer Engineering and Applications, 2022, 58(15):68-77.(in Chinese)

[4] 段续庭,周宇康,田大新,等. 深度学习在自动驾驶领域应用综述[J]. 无人系统技术,2021,4(6):1-27.
DUAN X T, ZHOU Y K, TIAN D X, et al. A review of deep learning applications for autonomous driving [J]. Unmanned Systems Technology, 2021, 4(6):1-27.(in Chinese)

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the 2014 IEEE Conference on

Computer Vision and Pattern Recognition, ACM,2014:580-587.
[6] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago,Chile: IEEE, 2015:1440-1448.
[7] 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.
[8] HE K M,GKIOXARI G,DOLLÁR P,et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Venice,Italy: IEEE,2017:2980-2988.
[9] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas,NV,USA: IEEE,2016:779-788.
[10] REDMON J,FARHADI A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu,HI,USA. IEEE,2017:6517-6525.
[11] REDOMON J, FARHADI A . YOLOv3: an incremental Improvement[J]. 2018.
[12] BOCHKOVSKIY A, WANG C Y, LIAO H. Yolov4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.
[13] JOVHER G. YOLOv5 [EB/OL]. https://github. com/ultralytics/yolov5, 2021.
[14] LI C Y,LI L,JIANG H L,et al. YOLOv6:a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.
[15] WANG C Y , BOCHKOVSKIY A , LIAO H Y M .YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// arXiv preprint arXiv:2207.02696, 2022.
[16] JOCHER G, CHAURASIA A, Q J. YOLO by Ultralytics (Version 8.0.0) [EB/OL]. 2023.https://github.com/ultralytics/ultralytics.
[17] XU S , WANG X , LV W ,et al.PP-YOLOE: An evolved version of YOLO[J]. 2022.
[18] LI T W,TONG G J,TANG H Y,et al. FisheyeDet:a self-study and contour-based object detector in fisheye images[J]. IEEE Access,2020,8:71739-71751.
[19] KUMAR V R, YOGAMANI S, RASHED H, et al. OmniDet: surround view cameras based multi-task visual perception network for autonomous driving [J]. IEEE Robotics and Automation Letters,2021,6(2):2830-2837.
[20] RASHED H,MOHAMED E,SISTU G,et al. Generalized object detection on fisheye cameras for autonomous driving: dataset, representations and baseline[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE,2021:2271-2279.
[21] COORS B, CONDURACHE A P, GEIGER A. SphereNet: learning spherical representations for detection and classification in omnidirectional images[C]//Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8 – 14, 2018,Proceedings,Part IX: ACM,2018:525 – 541.
[22] ZHANG Z H,XU Y Y,YU J Y,et al. Saliency detection in 360 $$^\circ$$ videos[C]//European Conference on Computer Vision. Cham:Springer,2018:504-520.
[23] SU Y C,GRAUMAN K. Kernel transformer networks for compact spherical convolution [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA,USA: IEEE,2019:9434-9443.
[24] DOOLEY D, MCGINLEY B, HUGHES C, et al. A blind-zone detection method using a rear-mounted fisheye camera with combination of vehicle detection methods[J]. IEEE Transactions on Intelligent Transportation Systems,2016,17(1):264-278.
[25] BAEK I, DAVIES A, YAN G, et al. Real-time detection, tracking, and classification of moving and stationary objects using multiple fisheye images [C]//2018 IEEE Intelligent Vehicles Symposium (IV). Changshu,China: IEEE,2018:447-452.
[26] 王文亮,李延祥,张一帆,等. MPANet-YOLOv5:多路径聚合网络复杂海域目标检测[J]. 湖南大学学报(自然科学版),2022, 49(10):69-76.
WANG W L, LI Y X, ZHANG Y F, et al. MPANet-YOLOv5: multi-path aggregation network for complex sea object detection [J]. Journal of Hunan University (Natural Sciences), 2022, 49(10):69-76.(in Chinese)
[27] 秦晓辉,黄启东,常灯祥,等. 基于改进YOLOv5的露天矿山目标检测方法[J]. 湖南大学学报(自然科学版),2023,50(2): 23-30.
QIN X H, HUANG Q D, CHANG D X, et al. Object detection method in open-pit mine based on improved YOLOv5 [J]. Journal of Hunan University (Natural Sciences), 2023, 50(2): 23-30.(in Chinese)
[28] 杨睿宁,惠飞,金鑫,等. 改进YOLOv5s的复杂交通场景路侧目标检测算法[J]. 计算机工程与应用,2023,59(16): 159-169.
YANG R N, HUI F, JIN X, et al. Roadside target detection algorithm for complex traffic scene based on improved YOLOv5s [J]. Computer Engineering and Applications, 2023, 59(16): 159-169.(in Chinese)
[29] 刘春霞,李超,潘理虎,等. 改进YOLOv5s的煤矿烟火检测算法[J]. 计算机工程与应用,2023,59(17):286-294.
LIU C X,LI C,PAN L H,et al. Improved coal mine smoke and fire detection algorithm of YOLOv5s[J]. Computer Engineering and Applications,2023,59(17):286-294.(in Chinese)
[30] 杨晨,余璐,杨璐,等. 改进YOLOv5的遥感影像目标检测算法[J]. 计算机工程与应用,2023,59(15):76-86.
YANG C, SHE L, YANG L, et al. Improved YOLOv5 object detection algorithm for remote sensing images [J]. Computer Engineering and Applications,2023,59(15):76-86.(in Chinese)
[31] YOGAMANI S, HUGHES C, HORGAN J, et al. WoodScape: a multi-task, multi-camera fisheye dataset for autonomous driving [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 9307-9317.
[32] ZHANG Q L, YANG Y B. SA-net: shuffle attention for deep convolutional neural networks [C]//ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada. IEEE, 2021: 2235-2239.
[33] WU Y,CHEN Y P,YUAN L,et al. Rethinking classification and localization for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA,USA: IEEE,2020:10183-10192.
[34] GE Z, LIU S, WANG F, et al. Yolox: exceeding yolo series in 2021[J].2021.
[35] YANG X, YAN J C. Arbitrary-oriented object detection with circular smooth label [C]//European Conference on Computer Vision. Cham:Springer,2020:677-694.