

文章编号:1007-2780(2022)09-1228-14

基于 YOLOv5s 的高速公路车辆实时检测模型

刘元峰¹, 姬海军², 刘立波^{1*}

(1. 宁夏大学 信息工程学院, 宁夏 银川 750021;
2. 宁夏路网监测与应急处置中心, 宁夏 银川 750021)

摘要:针对高速复杂环境下 YOLOv5s 算法细节特征学习能力弱、冗余信息过多、关键特征融合不足导致车辆目标检测精度低的问题,本文提出一种基于 YOLOv5s 的高速公路车辆实时检测模型(YOLOv5s-CRCP)。首先,在残差单元中嵌入卷积注意力模块,强化学习细节特征,抑制冗余信息干扰;然后,将卷积注意力融入金字塔网络中用以区分不同重要信息,加强关键特征融合。在构建的宁夏高速公路车辆数据集上进行实验,平均检测精度达到 91.2%,高出原算法 4.1%。实验结果表明,相较于 YOLOv5s 和主流实时车辆目标检测算法,本文方法具有更好的检测性能。

关键词:注意力机制;YOLOv5s;车辆目标检测;智能交通;神经网络

中图分类号:U495;TP391 **文献标识码:**A **doi:**10.37188/CJLCD.2022-0026

Real-time detection model of highway vehicle based on YOLOv5s

LIU Yuan-feng¹, JI Hai-jun², LIU Li-bo^{1*}

(1. School of Information Engineering, Ningxia University, Yinchuan 750021, China;
2. Ningxia Road Network Monitoring and Emergency Response Center, Yinchuan 750021, China)

Abstract: Aiming at the problems that the YOLOv5s algorithm has weak detailed feature learning ability, excessive redundant information, and insufficient key feature fusion in complex highway environments leads to low accuracy of vehicle target detection, a real-time detection model of highway vehicle is proposed based on YOLOv5s (YOLOv5s-CRCP). Firstly, convolutional attention module is embedded in the residual unit to strengthen the learning of detailed features and suppress the interference of redundant information. Secondly, convolutional attention is integrated into pyramid network to distinguish different important information and strengthen the fusion of key features. Experiments are conducted on the constructed Ningxia highway vehicle data set, and the average detection accuracy reaches to 91.2%, which is 4.1% higher than that of original algorithm. Experimental results show that the proposed method has better detection performance in comparison with YOLOv5s and the mainstream real-time vehicle target detection algorithms.

Key words: attention; YOLOv5s; vehicle target detection; intelligent transportation; neural networks

收稿日期:2022-01-24;修订日期:2022-03-16.

基金项目:宁夏重点研发计划(No. 2021BEG03024)

Supported by Key Research and Development Program of Ningxia(No. 2021BEG03024)

*通信联系人, E-mail:liulib@163.com

1 引言

精准、快速检测出车辆目标是高速交通控制和管理的重点、难点问题,具有重要的研究价值^[1]。为提升检测效率,研究人员提出了单阶段目标检测方法,主要代表有SSD^[2]、YOLO系列^[3-6]等。该类方法通过卷积网络提取特征并直接对目标定位识别,通过减少网络参数使检测速度得以提高,但却导致检测精度有所下降。因此,如何在车辆目标实时检测中既能保证检测速度又能提升检测精度,已成为研究的热点问题。

王滢暄等^[7]提出一种基于改进YOLOv4高速公路车辆目标实时检测方法,制作了各种复杂场景下的车辆数据集并对其进行数据增强,提升了检测精度。袁小平等^[8]提出一种改进YOLOv3的行人车辆目标检测算法,采用ResNeXt代替ResNet模块,并引入密集连接,提升了小目标的检测精度。于博等^[9]采用一种用于远红外图像的优化YOLO检测与分割网络模型,通过K-Means算法设定锚框尺寸以及使用自适应阈值分割方法,保证了实时检测要求,且能较有效地检测行人目标。Lai等^[10]提出了一种基于车辆检测和交通计数区域的二次检测框架,通过多角度、多路况数据采集,能够较精确地检测和统计出车辆信息。上述方法在车辆目标实时检测方面均有应用价值,但在检测过程中,模型对于细节特征学习能力较弱且提取了大量冗余信息,导致检测精度降低。

本文提出一种基于YOLOv5s的高速公路车辆实时检测算法,通过在残差单元中加入卷积注意力模块,强化模型学习细节特征,抑制冗余信息干扰;在金字塔网络中加入卷积注意力模型,加强关键特征融合;在自行构建的宁夏高速公路车辆数据集上进行实验,结果表明,本文方法能够准确、快速地检测出车辆目标,在高速公路车辆管理和控制方面具有一定的应用价值。

2 模型

YOLOv5系列包含s、m、l、x四种不同大小的网络模型^[11-13],其中YOLOv5s网络参数最少,检测速度更适合实时车辆目标检测场景。YOLOv5s主要由Backbone和Head两部分组成。Backbone部分用于提取图像特征,包含Focus、4个基础卷积(Convolution,Conv)、3个ResBCSPn模块和空间金字塔池化(Spatial Pyramid Pooling,SPP);Head部分由金字塔网络(Pyramid Attention Network,PANet)和分类回归预测模块Detect构成,分别用于多尺度特征信息融合和预测。整体网络结构如图1所示。

ResBCSPn作为Backbone的核心模块,主要由一个及以上残差单元(ResBottlebeck)和跨阶段局部模块(Cross Stage Partial,CSP)组成。其中,残差单元用于加深网络记忆,防止网络因层数加深而导致梯度消失;跨阶段局部模块用于组合特征、降低网络参数量。Head的核心模块是

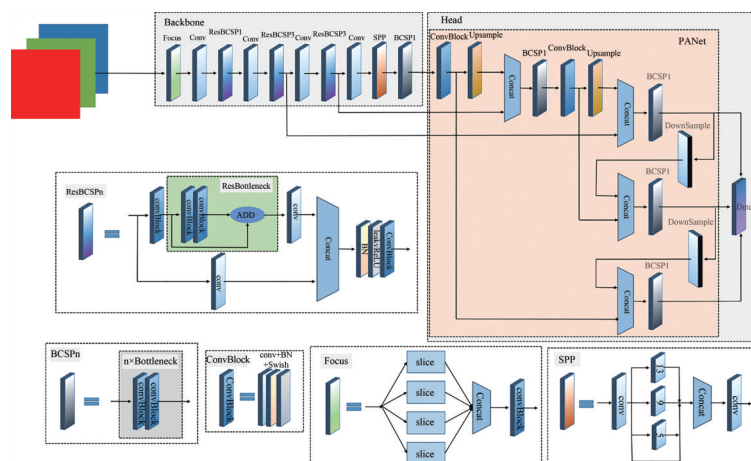


图1 YOLOv5模型架构

Fig. 1 Model architecture of YOLOv5s

金字塔网络,该部分采用自顶向下与自底向上联合的特征融合方式,更好地进行了多尺度融合,加强了不同尺度特征之间的语义关联性。

3 本文方法

3.1 注意力机制

注意力机制是计算机视觉领域常用的特征强化方法,源于人类大脑对视觉信号处理机制的模拟。人类在观察某一物体时,会重点关注目标有特点的细节部位,忽略全局和背景等冗余信息。这种选择性关注的方式与车辆目标实时检测任务中依靠车辆特点进行识别分类的理念相一致,近年来已在该领域得到了广泛应用。因此,为达到抑制图像中冗余信息、学习和融合关键细节特征进而提升模型检测精度的目的,本文引入一种多维注意力机制——卷积注意力模型(Convolutional Block Attention Module, CBAM)^[14-16],其包括通道域(Channel Attention Module, CAM)和空间域(Spatial Attention Module, SAM)两部分,分别处理图像的特征通道和特征空间。

由卷积网络生成的特征图中包含不同的特征空间和通道,每个空间、通道表示图像中的不同信息,其中可能包含与目标无关的背景冗余

特征。因此,让网络将注意力集中于提取和学习图像中关键细节特征,使其具备更高的权重系数,能有效提升模型的检测精度。本文在主干网络中使用CBAM提取特征的通道和空间生成注意力权重图,并将其与原特征图进行特征融合,将融合后的模块添加到残差单元之后,形成卷积注意力残差单元以获取并学习不同维度、更为丰富的关键细节特征。在金字塔网络中,使用相同的方法得到融合后的模块,将其放置在不同尺度特征图的上采样和下采样后,形成卷积注意力金字塔网络以加强不同尺度的关键细节特征融合。

3.2 卷积注意力残差单元

主干网络中由卷积函数生成特征图,由于卷积函数具有相同卷积核,导致无法辨析图像中关键细节特征。而卷积注意力模块作为一种特征强化方法,通过提高关键细节特征权重、降低冗余信息权重,使网络着重关注感兴趣区域,从而实现提升检测精度的目的。因此,本文将卷积注意力模块放置于残差单元后,形成卷积注意力残差单元(CBAMResBottleneck, CR),如图2所示。

卷积注意力残差模块主要由残差单元(ResBottleneck)、CAM和SAM三部分构成,分别如

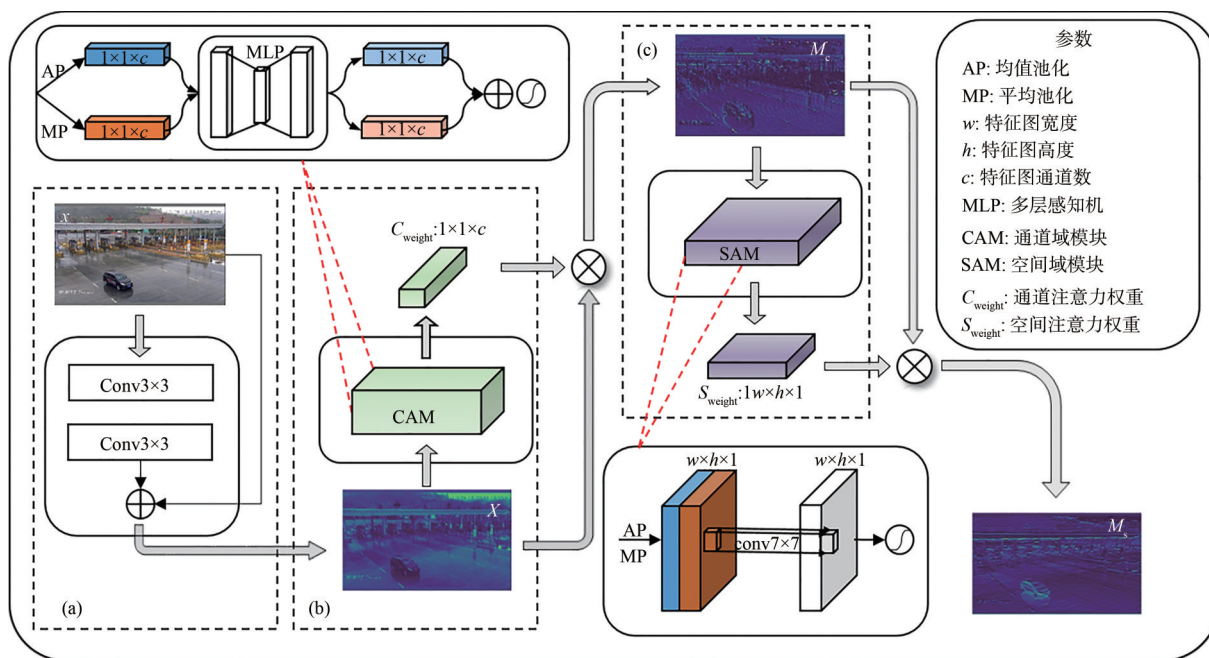


图2 卷积注意力残差模块。(a) 残差单元; (b) CAM; (c) SAM。

Fig. 2 Convolutional attention residual module. (a) ResBottleneck; (b) CAM; (c) SAM.

图2中(a)、(b)、(c)所示。输入特征 $x \in R^{w \times h \times c}$ 进入残差单元。残差单元通过两层 3×3 的卷积提取 x 得到具备高级语义信息的特征 $f(x)$ 后,使用跳跃连接将其与 x 相加得到 X ,这使流入到下一层网络的特征由 $f(x)$ 变为 $f(x)+x$,从而保留了浅层信息。该过程增强了特征信息。但并未改变特征维度,因此特征 $X \in R^{w \times h \times c}$ 。处理流程如图3(a)所示,具体公式如式(1)、(2)所示:

$$f(x) = \text{Conv}(\text{Conv}(x)), \quad (1)$$

$$X = f(x) + x, \quad (2)$$

式中: $w \times h$ 为特征维度大小, c 为特征通道数,Conv代表卷积核为 3×3 的卷积, X 是保留浅层信息的残差单元输出特征图。

经过残差单元叠加浅层信息的特征 X 由CAM单元处理,CAM重点关注特征的通道维度。不同卷积核生成的特征通道存在重要性差异,但网络公平捕获特征的方式忽略了这一问题。为此,CAM赋予各通道不同的权重系数来衡量其差异性,使网络根据权重系数聚焦关键通道特征。具体操作如下:首先,并行平均和最大池化将 X 的空间信息聚合压缩得到两个 $1 \times 1 \times c$ 大小的张量。然后,将其输送到共享网络MLP中获取通道注意力分布。MLP是包含一个隐藏层的感知机,利用其处理非线性数据分类的能力实现了通道注意力分布。接着对重新分配通道注意力权重的张量进行求和运算得到最终的通道权重 C_{weight} , $C_{\text{weight}} \in R^{1 \times 1 \times c}$ 。最后将其与 X 相乘融合得到通道注意力特征图 M_c , $M_c \in R^{w \times h \times c}$ 。处理流程如图3(b)所示,具体公式如式(3)所示:

$$M_c(X) =$$

$$\sigma(\text{MLP}(\text{AP}(X)) \oplus \text{MLP}(\text{MP}(X))) \otimes X, \quad (3)$$

式中:AP为平均池化,MP为最大池化, σ 为sigmoid激活函数,MLP为多层感知机, $\oplus$ 为矩阵求和运算, $\otimes$ 为矩阵乘积运算, M_c 是生成的通道注意力特征图。

经过CAM处理得到的特征图 M_c 进入SAM模块用以计算空间特征权重。SAM首先将不同通道在同一平面空间点上的值并行平均和最大池化分别获取两个维度为 $w \times h \times 1$ 的空间权重图,再通过卷积层和sigmoid函数对两个空间特征图进行非线性激活得到最终的空间权重图 S_{weight} , $S_{\text{weight}} \in R^{w \times h \times 1}$,最后将其与 M_c 相乘融合得

到空间注意力特征图。处理流程如图3(c)所示,具体操作公式如式(4)所示:

$$M_s(M_c) =$$

$$\sigma(\text{Conv7}(\text{AP}(M_c) \oplus \text{MP}(M_c))) \otimes M_c, \quad (4)$$

式中:Conv7是卷积核大小 7×7 的卷积, M_s 为生成的空间注意力特征图。

上述步骤产生的中间特征 X 、 M_c 和 M_s ,经过卷积等操作处理后,无法明确分析其中的差异性,为此给予中间过程特征的相应热力图^[17]进行直观描述,具体效果如图3所示。

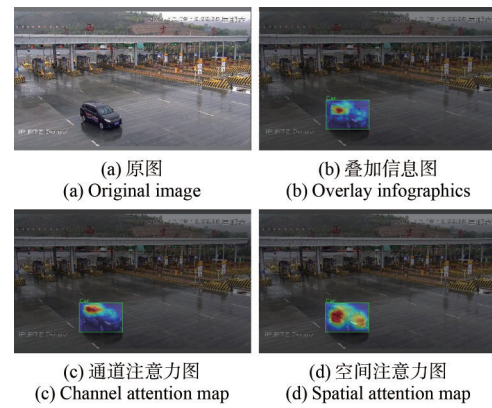


图3 中间过程效果图

Fig. 3 Intermediate process renderings

由图3可知,特征图 X 能够检测出车辆目标并将注意力集中在车侧上方,说明信息主要叠加在该区域;特征图 M_c 通过通道域处理,加强了各通道之间的依赖性,使注意力权重集中在信息丰富的区域,在图3(b)中已知车侧上方信息丰富,因此图3(c)中该位置高亮效果更为显著;特征图 M_s 经过空间域处理,突出了各通道上不同空间重要程度,在图3(c)基础上,额外获取了车头、车门等关键部位特征,增大了有效预测区域。

由于YOLOv5s将输入图像调整为 $640 \times 640 \times 3$ 的格式,因此在经过主干特征提取网络卷积后会得到维度分别为 $80 \times 80 \times 256$ 、 $40 \times 40 \times 512$ 、 $20 \times 20 \times 1024$ 的特征图 f_1 、 f_2 和 f_3 ,并将其输入到金字塔网络中进行多尺度特征融合。金字塔网络通过融合多尺度特征图加强了不同语义、分辨率特征图像的信息关联度,但破坏了特征的权重系数,网络无法着重融合关键位置信息。

3.3 卷积注意力金字塔网络

目标检测主要有两大任务:目标分类和定

位,两者分别依赖于深层网络中的高级语义信息和浅层网络中的车辆目标位置信息。在3.2节中,卷积注意力残差单元主要通过提取关键特征用以提升分类精确度,在本小节中,卷积注意力金字塔网络则侧重于解决目标定位问题。

金字塔网络主要通过采样操作融合不同层级之间的特征,以保证其兼顾高级语义和车辆位置信息,但如3.2节中提到的网络无法区分特征重要程度。为此,借鉴上述思想,通过在不同层次特征融合之前加入CBAM注意力机制,给定特征空间和通道方向上的一个权重,以保证金字塔网络

能够融合关键信息,进而提升模型准确度。具体方式是在金字塔网络上采样和下采样模块后加入卷积注意力机制,形成卷积注意力金字塔网络(CBAMPANet,CP)。卷积注意力金字塔网络结构如图4所示。其中,图4(a)为主干特征提取网络;图4(b)代表深层特征降通道维度拼接浅层位置信息的过程,中间特征为 $[P_3, P_2, P_1]$;图4(c)代表浅层升通道维度融合深层语义的过程,该过程会产生最终待分类回归特征 $[N_1, N_2, N_3]$ 。为清晰直观地描述中间过程特征,通过给予对应热力图进行分析,效果如图5所示。

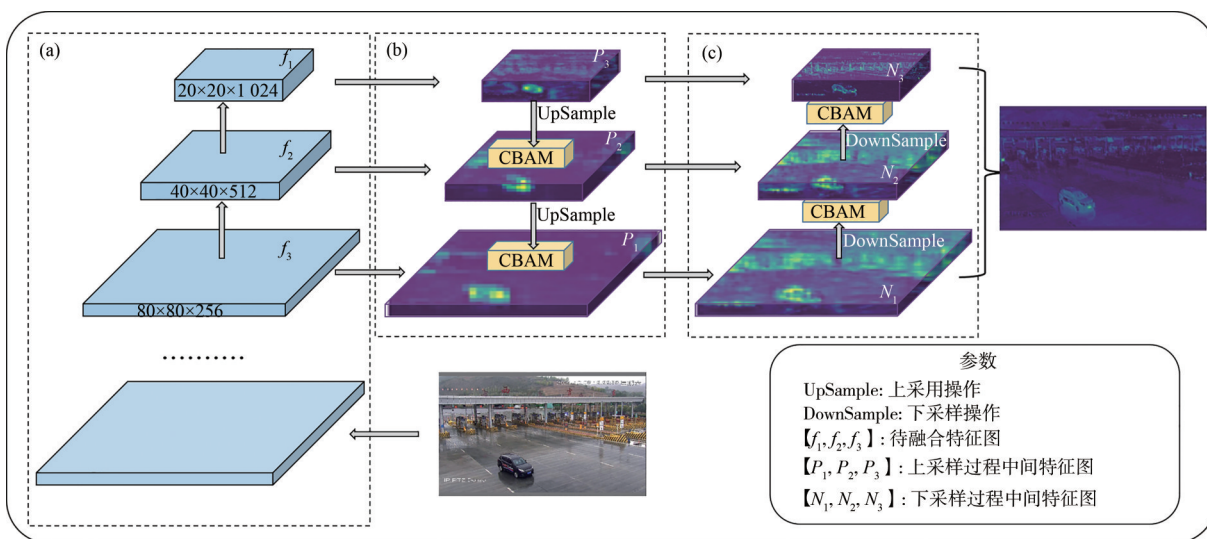


图4 卷积注意力金字塔网络。(a)主干特征提取网络;(b)自上向下的路径增强;(c)自下向上的路径增强。

Fig. 4 Convolutional attention pyramid network. (a) Backbone; (b) Up-bottom path augmentation; (c) Bottom-up path augmentation.

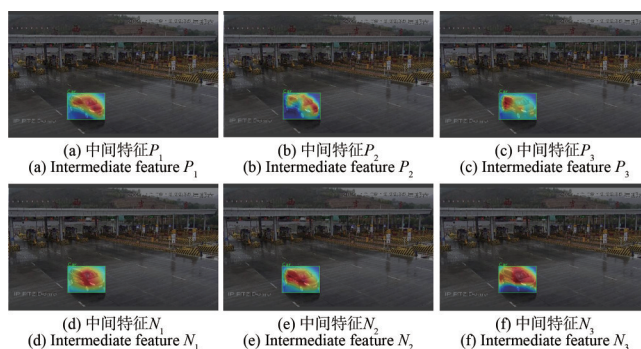


图5 中间过程热力图

Fig. 5 Intermediate process heat map

特征图 P_3 来自于深层特征 f_1 , 因此具备全网络中最为丰富的语义信息, 但由于深层网络目标位置信息欠缺, 导致最终预测区域并不理想, 仅

能检测到车侧上方部分, 这一效果与3.2节中最终效果类似, 说明图像中该区域特征在全局中的信息最为丰富; P_3 经上采样、CBAM处理后与特

征图 f_2 拼接形成 P_2 , f_2 相较于 f_1 ,兼顾了语义和位置信息,而 P_3 上采样过程中融入了CBAM模块,使用权重将注意力集中在关键区域,因此预测范围囊括了车身和车头等关键部位,基本覆盖整体车辆,具有较好的检测效果; P_1 来自于同级特征 f_3 和经CBAM处理的上级特征,由于 P_1 作为图4(b)中最后一步,包含的特征信息更为丰富,覆盖预测区域更为广泛且明显。 N_1 是基于 P_1 进行的卷积等操作,从而预测区域与 P_1 相似; N_2 检测效果较好的原因,一方面是来自同级特征 P_2 在上采样操作中包含的信息已较为丰富,另一方面下级特征 P_1 经CBAM处理后将注意力集中在特征关键部分; N_3 与 N_2 原理基本相同,作为图4(c)中最后一步,其包含了丰富的叠加信息,因此预测区域优于所有中间过程效果图。

在图4(b)中,浅层特征(如 f_3)车辆位置清晰,但由于卷积处理较少,特征语义信息欠缺,无法精确进行分类;深层特征(如 P_3)车辆信息模糊,但经过更多卷积提取,语义信息丰富;因此图4(a)过程主要利用深层特征的语义信息丰富浅层特征,使其能够精确定位和分类,融合后的特征如 P_1 所示,该过程可描述为:深层特征通过上采样降低特征维度、增强位置信息,然后经过CBAM进行赋权,最后与浅层特征进行拼接。而图4(b)过程的目的在于,通过利用浅层特征的位置信息丰富深层特征,使其具备明显的车辆信息,拼接后的特征如 N_3 所示,该过程可描述为:浅层特征通过下采样提升特征维度、丰富语义信息,经过CBAM进行赋权后与浅层特征进行融合。上采样(UpSample)用于扩展图像尺度并压缩通道,下采样(DownSample)用于压缩图像空间并扩张通道。具体公式如式(5)、(6)所示:

$$\text{UpSample} = \text{TransConv}(\text{Conv}(f_c)_{w \times h}), (5)$$

$$\text{DownSample} = \text{TransConv}(\text{Conv}(f_{w \times h})_c), (6)$$

式中:TransConv是卷积核 2×2 的反卷积,Conv是卷积核为 3×3 的卷积, f_c 代表当前特征的通道, $f_{w \times h}$ 为当前特征的宽高。

在残差单元之后加入CBAM,强化主干特征提取网络对目标细节的学习能力,抑制冗余信息干扰,获取关键的高级语义信息以保证提升目标分类准确度;在不同的层级特征融合之前加入CBAM,给予待融合特征重要性权重,丰富特征

的关键车辆位置信息来增强目标定位精准度。通过上述内容可知,卷积注意力残差单元和卷积金字塔网络分别用于完善目标分类和定位任务。为证明改进方法的有效性,在4.5节中将对各改进模块进行消融实验。

4 实验与结果分析

4.1 实验环境与数据集

实验过程在Pytorch框架下进行,训练及测试的计算机硬件配置GPU为NVIDIA Geforce RTX 3 090 Ti,操作系统为Ubuntu 16.04.12,python版本为3.6.5。使用宁夏地区2020年8月6日至2021年11月24日的高速公路视频数据,共计26.2G,视频显示格式为1 080 P,每秒传输帧数为25,囊括普通高速公路、分岔路口、收费站、盘山公路、隧道和高架桥等6类高速道路场景下的各类遮挡物与截断目标,可满足本文方法对于复杂场景数据的需求。

宁夏高速公路车辆数据集构建方式如下:首先通过人工清洗车辆缺失目标、模糊视频;然后使用Python脚本文件将视频截帧(截帧间隔2 s)并清洗,去除冗余、图像撕裂、马赛克等无效数据;接着使用labelimg工具标注图像车辆目标,并设置Car、Van、Bus和Truck四种车辆类别标签,标注完成1 200张图片后经数据增强得到3 000张;最后按照训练集:验证集:测试集=8:1:1的比例对数据集进行划分,得到训练集图像数量2 400张、验证集图像数量300张、测试集图像300张。各类车型数据如表1所示。

表1 数据集内容描述

Tab. 1 Datasets content description

种类	描述	数量
Car	私家轿车、4座及以内车型	3 576
Van	小货车、商务车、8座及以内车型	2 398
Bus	客车	3 074
Truck	大型卡车、大型货车	3 685

4.2 评价指标

本文使用每秒传输帧数(Frames Per Second, FPS)^[18]评估算法检测速度,采用均值平均精度(Mean Average Precision, mAP)^[19-24]对算法

检测精度进行定量评估。

FPS 即视频中的画面数,每秒帧数越多,所显示的动作就越流畅。通常情况下,FPS 高于 50 f/s 时,就可满足实时目标检测的要求。计算公式如式(7)所示:

$$FPS = \frac{\text{frameCount}}{\text{elapsedTime}}, \tag{7}$$

式中: frameCount 代表视频总帧数, elapsedTime 代表视频总时间。

mAP 由精确率(Precision)和召回率(Recall)求出。精确率又被称为查准率,用于衡量算法的准确度;召回率又称查全率,用于衡量算法的漏检率,计算公式如式(8)、(9)所示:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{9}$$

其中: TP 表示正确分类样本数量, FP 表示误分样本数量, FN 表示样本误检数量。

平均精度(Average Precision, AP)用于计算单类别的检测精度;其表示精确率与召回率所围成曲线的面积,均值平均精度用于计算多个类别的平均精度,计算公式如式(10)所示:

$$mAP = \frac{\sum_{j=1}^M AP}{M}, \tag{10}$$

其中, M 代表数据集的类别总数。

4.3 网络训练

本实验采用多任务损失函数优化网络参数,

利用 k-means 算法^[25]对宁夏高速公路车辆数据集进行聚类获得多尺度锚框,宽高分别为(15,12)、(23,18)、(41,18)、(36,28)、(65,30)、(78,59)、(122,51)、(154,90)、(189,171)。设置相同模型参数:输入图像尺寸为 640×640,初始学习率设置为 0.001,网络模型训练过程中选择余弦退火方式^[26]来降低学习率,训练批次大小(Batch size)设置为 32,迭代次数为 15 000 次。训练过程损失下降曲线如图 6 所示。由图 6 可知,网络训练初期下降较快,在 2 000 次左右到达拐点,损失下降梯度减缓,14 500 次左右损失下降趋于平稳,最终损失收敛在 0.025 左右。

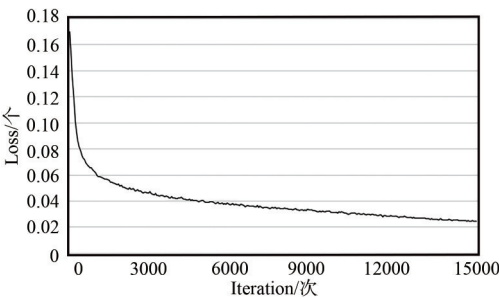


图 6 损失下降曲线

Fig. 6 Loss decline curve

4.4 对比实验

为验证和评估本文改进方法的性能,实验将本文方法(YOLOv5s-CRCP)与主流的车辆目标实时检测算法 SSD、YOLOv3、YOLOv4、YOLOv5s 在宁夏高速公路车辆数据集上进行对比。检测精度和检测速度的实验结果如表 2 和表 3 所示。

表 2 不同算法在宁夏高速公路车辆数据集的检测精度对比

Tab. 2 Comparison of detection accuracy of different algorithms on Ningxia expressway vehicle datasets

Detector	Car	Van	Bus	Truck	mAP
SSD	0.836	0.635	0.792	0.865	0.782
YOLOv3	0.906	0.658	0.889	0.927	0.845
YOLOv4	0.951	0.714	0.914	0.964	0.886
YOLOv5s	0.937	0.681	0.911	0.955	0.871
YOLOv5s-CRCP(ours)	0.986	0.744	0.923	0.995	0.912

由表 2 可知,本文方法通过在残差单元和金字塔网络中融合卷积注意力模块,提高了模型的检测精度,各类车辆的检测精度分别为 99.5%、

74.4%、92.3% 和 98.6%,均值检测精度高达 91.2%,优于表中任一主流车辆目标实时检测模型。SSD 模型均值检测精度仅为 78.2%,主要原

表3 不同算法在宁夏高速公路车辆数据集的检测速度对比

Tab. 3 Comparison of test results of different algorithms on Ningxia expressway vehicle datasets

Detector	Base Network	FPS
SSD	VGG16	26
YOLOv3	DarkNet53	36
YOLOv4	CSPDarkNet53	38
YOLOv5s	—	77
YOLOv5s-CRCP(Ours)	—	75

因是主干网络的低级特征卷积层数少,特征提取不充分,并且未对不同尺度特征进行融合。YOLOv3模型通过使用DarkNet53作为骨干网络以加深网络层数,充分提取特征信息,并使用特征金字塔网络充分融合不同尺度特征,将均值检测精度提升到84.5%。YOLOv4使用数据增强,扩充训练数据将模型均值检测精度提高到88.6%。YOLOv5s模型由于网络参数量大幅减少导致检测精度相较于YOLOv4下降1.5%,为87.1%,能达到实时目标检测的要求。

由表3可知,在同一环境下,SSD模型主要

采用VGG16作为特征提取网络,因参数量过多导致网络检测速度较慢,其检测速度仅为26 f/s。YOLOv3模型以DarkNet53作为骨干网络模型,网络中参数量大幅降低,检测速度为36 f/s,相较于SSD模型具有明显的提升。YOLOv4使用CSPDarkNet53模型分组处理特征,降低了网络参数量,使检测速度略微提升,达到38 f/s。YOLOv5s模型对主干网络模型进行深度和宽度压缩并使用focus结构分治处理输入特征,极大降低了网络参数量,检测速度达到77 f/s。本文方法在YOLOv5s的基础上添加了CBAM机制,由于CBAM中包含MLP结构,导致参数量有所增加,使检测速度略微降低,为75 f/s,但仍能达到实时检测的要求。

4.5 消融实验

本文以图3输入的特征图像为例,对上述各模块进行消融实验过程说明。整体处理流程为:(1)待检测图像经过包含卷积注意力残差单元的主干特征提取网络提取到多尺度特征;(2)输入到卷积注意力金字塔网络中进行多尺度特征融合。处理步骤(1)、(2)分别如图7所示。

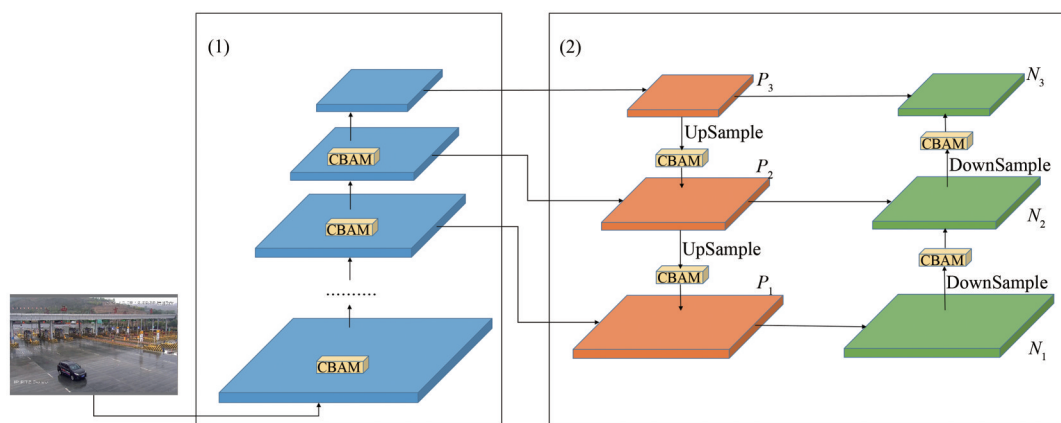


图7 整体实验流程图

Fig. 7 Overall process flowchart

4.5.1 卷积注意力残差单元实验分析

为验证卷积注意力残差单元融入主干特征提取网络后的性能,本文基于宁夏高速车辆数据集进行实验:实验1. 基准模型(YOLOv5s);实验2. 仅使用CAM模块;实验3. 仅引入SAM单元;实验4. 同时添加CAM和SAM模块(YOLOv5s-CR)。实验结果如表4所示。

由表4可知,加入CAM模块后,检测精度提

升了1.6%,这主要归功于MLP多层感知机使用全连接提取了更多的细节特征;引入SAM模块检测精度提升不明显,原因在于SAM仅通过串行池化的方式增大了感受野,但依旧采用卷积捕获特征,因此精度提升不大;同时引入SAM和CAM时,模型精度相较于原模型提升了2.2%,有着较好的检测效果,故最终本文采用实验4的模型架构。

表 4 卷积注意力残差单元的消融实验

Tab. 4 Ablation experiment of CR

Index	实验 1	实验 2	实验 3	实验 4
mAP	0.871	0.887	0.876	0.893
FPS	77	76	77	76

为进一步说明 CAM 和 SAM 的作用,对其引

入到主干特征提取网络前后的提取结果通过热力图进行对比分析,如图 8 所示。在不引入任何注意力机制时,具体实验结果如图 8(b)所示;引入 SAM 时,主干特征提取网络提取结果如图 8(c)所示;引入 CAM 时,提取特征实验结果如图 8(d)所示;同时引入 CAM 和 SAM 的提取结果如图 8(e)所示。

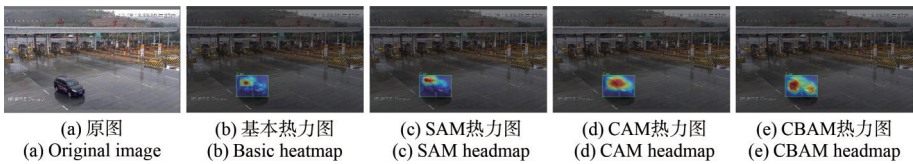


图 8 主干网络提取特征结果

Fig. 8 Backbone network extraction feature results

由图 8 可知,引入注意力机制后,网络能够更加关注目标关键特征,进而突出车辆轮廓,忽略背景冗余信息,因此特征的提取效果均高于原模型;使用 CAM 捕获特征的效果要高于 SAM,主要表现 CAM 能够将注意力集中在整体车身等明显的车辆特征部位,而 SAM 仅能突出车顶部位,对于其他关键部位的处理效果较弱;同时引入 CAM 和 SAM 时,整体车身、车头等关键部位均能关注,效果显著。

4.5.2 卷积注意力金字塔网络实验分析

为验证 CBAM 加入特征融合网络后的性能,本文进行如下实验:实验 1. 基准模型(YOLOv5s);实验 2. 仅在上采样单元添加 CBAM 模块;实验 3. 仅在下采样单元添加 CBAM 模块;实验 4. 在上、下采样后同时添加 CBAM 模块(YOLOv5s-CP)。最终实验结果如表 5 所示。

由表 5 可知,仅在特征融合网络中上采样后加入 CBAM 模块精度提升了 2.3%;仅在下采样后添加 CBAM,精度提升了 1.6%,低于实验 3 的原因,下采样过程在上采样特征融合完成之后,其目标位置信息并不清晰,因而提升效果略

表 5 卷积注意力残差单元的消融实验

Tab. 5 Ablation experiment of CP

Index	实验 1	实验 2	实验 3	实验 4
mAP	0.871	0.894	0.887	0.901
FPS	77	76	76	75

低;同时添加 CBAM 模块的检测精度较于原模型提升了 3%,在实际应用中表现良好,因而本文使用实验 4 的模型架构。

为进一步说明在特征融合网络的上下采样操作后加入 CBAM 的作用,对提取结果通过热力图进行对比分析,如图 9 所示。在不添加 CBAM 时,PANet 具体融合实验结果如图 9(b)所示;下采样后引入 CBAM 的实验结果如图 9(c)所示;上采样后添加 CBAM 的特征融合效果如图 9(d)所示;上、下采样后同时引入 CBAM 的提取结果如图 9(e)所示。

由图 9 可知,未添加注意力机制的特征图注意力主要集中在车顶部位,并不能很好地将注意力集中在车辆整体模块;在下采样操作后添加 CBAM,能够将注意力提升到车顶和车身分位

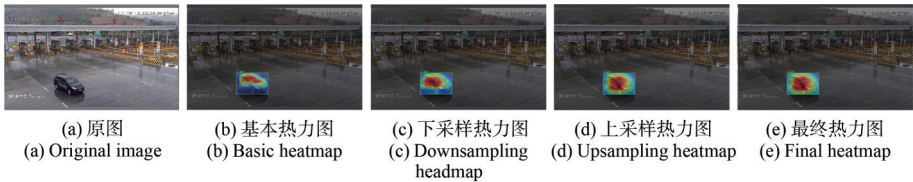


图 9 特征融合结果

Fig. 9 Feature fusion result

置,但效果依旧不理想;在上采样过程后添加CBAM的效果明显优于前者,网络将注意力扩散至车辆整体架构,对于车身能够整体覆盖;在上下采样后同时添加CBAM,突出重点位置基本与图9(d)相似,但是覆盖面积更为广泛,效果更好。

4.5.3 最终模型实验分析

为探究各改进模块对最终模型检测效果的影响,在宁夏高速公路车辆数据集上进行了以下实验:实验1. 基准模型(YOLOv5s);实验2. 加入卷积注意力残差单元;实验3. 加入卷积金字塔网络;实验4. 同时添加卷积注意力残差单元和卷积注意力金字塔网络。结果如表6所示。

在4.5.1和4.5.2节中详细分析了引入卷积

注意力残差单元和卷积注意力金字塔网络的效果,本实验主要关注同时融入两者的影响。从表6中可知,实验4通过同时在主干特征提取网络和特征融合网络中加入CBAM模块,分别用于提取关键特征和融合关键信息,最终检测精度达到了91.2%,相较于基准模型提升4.1%;在检测速度方面,由于添加了CBAM模块,导致网络参数略微增加,FPS降低为75,但与原模型差距不大,仍能保持实时检测。因此,本文改进模型能够在兼顾检测速度的同时,提升了检测精度,更能满足实际车辆目标检测场景。本文模型对宁夏高速公路数据集中的Car、Van、Truck和Bus四类车型进行检测,效果如图10所示。由图10可知,本文方法能够将注意力集中在车辆目标的关键部位,如车头、车身以及车顶等,其中Car和Truck效果最佳,基本可将车辆整体进行覆盖;Bus和Van可覆盖主要车身部分,但对于车头注意力较弱,主要由于数据量相对较少,特征提取不充分。

表6 最终算法消融实验

Tab. 6 Final algorithm ablation experiment

Index	实验1	实验2	实验3	实验4
mAP	0.871	0.893	0.901	0.912
FPS	77	76	75	75

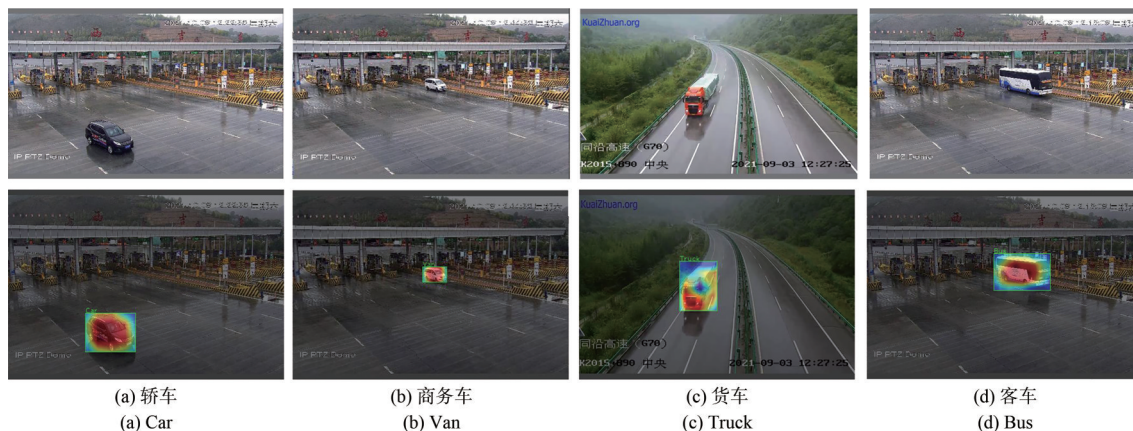


图10 特征融合结果

Fig10 Feature fusion results

4.6 车辆目标检测结果

为直观表述本文方法在不同道路环境下的有效性,选取宁夏高速公路测试集中普通道路、分岔路口、收费站、盘山公路、隧道和高架桥等不同道路场景进行定性分析测试,并将本文模型与主流实时目标检测算法进行对比。检测效果如图11所示,每组依次为原图、SSD、YOLOv3、YOLOv4、YOLOv5s和本文算法。

通过对比发现,在普通道路、分岔路口和

收费站口场景下,各算法均能正确检测出车辆类别,主要原因是道路情况较为简单,数据集中Truck和Car标签充足,但本文方法检测得分更高;在盘山公路场景下,SSD出现漏检车辆Car,其余算法均能正确检测出目标;在隧道中,SSD算法再次出现漏检现象,并且各算法均将Van误检为Truck,主要原因在于Van的特征与Truck相似、不易区分,而本文算法通过卷积注意力机制模块的添加则能很好学习车辆细节特征,最终正确检测出车辆类别;在高架桥



(a) 普通道路
(a) Ordinary roads



(b) 岔岔路口
(b) Fork in the road



(c) 收费站口
(c) Toll gate entrance



(d) 盘山公路
(d) Panshan highway

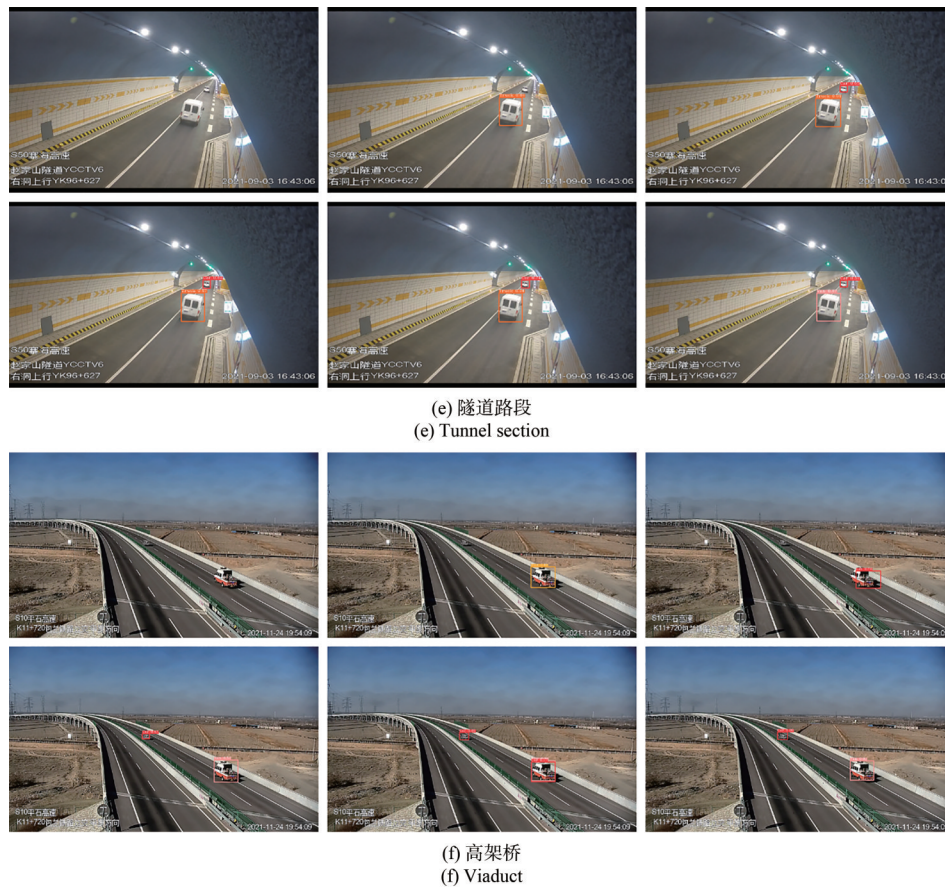


图11 宁夏高速公路车辆测试集上的效果图

Fig. 11 Renderings on Ningxia expressway vehicle test sets

中,SSD和YOLOv3将Van分别误检为Car和Bus并均存在漏检车辆,YOLOv4检测正确但检测精度较低,YOLOv5s则出现了重复标签现象,本文方法则取得了较高的检测精度。

5 结 论

针对复杂高速场景下YOLOv5s算法细节特征学习能力弱、提取冗余信息过多和关键特征融合不充分等问题,本文提出了基于YOLOv5s的

实时车辆目标检测模型,在YOLOv5s的残差单元和金字塔网络中加入卷积注意力机制模块,通过动态分配特征权重系数的方式,改善了上述问题。在自构建的宁夏高速公路车辆数据集上进行实验,通过评价指标mAP、FPS和测试效果图可知,相较于主流的实时目标检测算法,本文的改进模型具有更好的检测性能,但仍存在数据集不充分和极小目标车辆检测精度低等问题,因此在后续工作中将继续扩充数据集并提升极小目标车辆的检测精度。

参 考 文 献 :

- [1] ZHOU F Q, LI J, LI X S, *et al.* Freight car target detection in a complex background based on convolutional neural networks [J]. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2019, 233(3): 298-311.
- [2] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.
- [3] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016:

779-788.

- [4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 6517-6525.
- [5] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv*, 2018: 1804.02767.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. *arXiv preprint arXiv:2004.10934*, 2020.
- [7] 王滢暄, 宋焕生, 梁浩翔, 等. 基于改进的 YOLOv4 高速公路车辆目标检测研究[J]. *计算机工程与应用*, 2021, 57(13): 218-226.
WANG Y X, SONG H S, LIANG H X, *et al.* Highway vehicle object detection based on improved YOLOv4 method [J]. *Computer Engineering and Applications*, 2021, 57(13): 218-226.
- [8] 袁小平, 马绪起, 刘赛. 改进 YOLOv3 的行人车辆目标检测算法[J]. *科学技术与工程*, 2021, 21(8): 3192-3198.
YUAN X P, MA X Q, LIU S. An improved algorithm of pedestrian and vehicle detection based on YOLOv3 [J]. *Science Technology and Engineering*, 2021, 21(8): 3192-3198.
- [9] 于博, 马书浩, 李红艳, 等. 远红外车载图像实时行人检测与自适应实例分割[J]. *激光与光电子学进展*, 2020, 57(2): 021507.
YU B, MA S H, LI H Y, *et al.* Real-time pedestrian detection for far-infrared vehicle images and adaptive instance segmentation [J]. *Laser & Optoelectronics Progress*, 2020, 57(2): 021507.
- [10] LAI J H, LUO T T, WANG Y, *et al.* Traffic flow statistics method and practice based on roadside video with YOLOv3 [J]. *Journal of Highway and Transportation Research and Development (English Edition)*, 2021, 15(1): 89-97.
- [11] ZHANG X P, JI J H, WANG L, *et al.* People's fast moving detection method in buses based on YOLOv5 [J]. *International Journal of Sensors and Sensor Networks*, 2021, 9(1): 30-37.
- [12] WU T H, WANG T W, LIU Y Q. Real-time vehicle and distance detection based on improved YOLOv5 network [C]//2021 3rd *World Symposium on Artificial Intelligence (WSAI)*. Guangzhou: IEEE, 2021: 24-28.
- [13] SONG X Y, GU W. Multi-objective real-time vehicle detection method based on YOLOv5 [C]//2021 *International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)*. Xi'an: IEEE, 2021: 142-145.
- [14] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 3-19.
- [15] HONG Z H, YANG F, PAN H Y, *et al.* Highway crack segmentation from unmanned aerial vehicle images using deep learning [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 6503405.
- [16] DU S J, ZHANG B F, ZHANG P, *et al.* FA-YOLO: an improved YOLO model for infrared occlusion object detection under confusing background [J]. *Wireless Communications and Mobile Computing*, 2021, 2021: 1896029.
- [17] JETLEY S, LORD N A, LEE N, *et al.* Learn to pay attention [C]//*Proceedings of the 6th International Conference on Learning Representations*. Vancouver: ICLR, 2018.
- [18] MEHTA R, OZTURK C. Object detection at 200 frames per second [C]//*Proceedings of the European Conference on Computer Vision*. Munich: Springer, 2018: 659-675.
- [19] KIM K J, KIM P K, CHUNG Y S, *et al.* Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection [C]//2018 15th *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. Auckland: IEEE, 2018: 1-6.
- [20] PENG H L, GUO S, ZUO X Y. A vehicle detection method based on YOLOv4 model [C]//2021 2nd *International Conference on Artificial Intelligence and Information Systems*. Chongqing: ACM, 2021: 168.
- [21] WU W T, LIU H, LI L L, *et al.* Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image [J]. *PLoS One*, 2021, 16(10): e0259283.
- [22] SU H X, DONG Z H, YANG F, *et al.* Remote sensing vehicle detection based on embedded system [C]//*Proceedings of the SPIE 11719, Twelfth International Conference on Signal Processing Systems*. Shanghai: SPIE, 2021: 1171903.
- [23] 张炳力, 秦浩然, 江尚, 等. 基于 RetinaNet 及优化损失函数的夜间车辆检测方法[J]. *汽车工程*, 2021, 43(8):

1195-1202.

ZHANG B L, QIN H R, JIANG S, *et al.* A method of vehicle detection at night based on RetinaNet and optimized loss functions [J]. *Automotive Engineering*, 2021, 43(8): 1195-1202.

[24] ZHAN W, SUN C F, WANG M C, *et al.* An improved YOLOv5 real-time detection method for small objects captured by UAV [J]. *Soft Computing*, 2022, 26(1): 361-373.

[25] LIKAS A, VLASSIS N, VERBEEK J J. The global k -means clustering algorithm [J]. *Pattern Recognition*, 2003, 36(2): 451-461.

[26] 石健,成前,金立生,等. 面向视频结构化的细粒度车辆检测分类模型[J]. *汽车工程*, 2021, 43(10): 1427-1434.

SHI J, CHENG Q, JIN L S, *et al.* Fine-grained vehicle detection and classification model for video structuring description [J]. *Automotive Engineering*, 2021, 43(10): 1427-1434.

作者简介:



刘元峰(1997—),男,山东泰安人,硕士研究生,2020年于烟台南山学院获得学士学位,主要从事车辆目标检测理论与应用的研究。E-mail:1251666248@qq.com



刘立波(1974—),女,宁夏银川人,博士,教授,2010年于中国农业科学院农业信息研究所获得博士学位,主要从事计算机视觉、数字图像处理等方面的研究。E-mail:liulib@163.com