

DOI:10.19651/j.cnki.emt.2416269

多尺度特征融合增强的行人翻越护栏检测^{*}刘罡^{1,2} 侯恩翔² 黄孙港¹ 闫曙光² 黄应征²

(1. 无锡学院江苏省集成电路可靠性技术及检测系统工程研究中心 无锡 214105;

2. 南京信息工程大学电子与信息工程学院 南京 210044)

摘要: 针对行人翻越护栏检测在遮挡、多目标密集以及多人翻越的复杂场景下出现的漏检、误检和检测精度低的问题,提出了一种多尺度特征融合增强的行人翻越护栏检测算法。首先,设计了一种基于 Dual Vision Transformer 和 SCConv 构建的 SCDVT 模块,应用于主干网络,增强了对全局上下文信息和更细粒度信息的捕获,提升了网络的局部精细特征提取和特征融合能力;其次,提出多尺度特征融合增强模块 AM-SPPFCSP,弥补了最大池化带来的特征损失,提高了特征图的丰富性和完整性,增强了多尺度特征提取和特征融合能力;另外,对特征融合层再进行细化,使用 GSConv 替换普通卷积,并基于 GSConv 和 SCConv 设计了 VOV-GSCCSP 模块,有效的降低了计算成本和模型的复杂度,同时又保持较高的精度;最后在主干引入高效多尺度注意力 EMA,减少了复杂背景下无关目标的干扰,融合了多尺度信息,实现了更丰富的特征聚合。在自制行人翻越护栏数据集上的实验结果表明,本文所提算法在增加较少参数的情况下,其 mAP 达到了 93.6%,较原模型提高了 4.5%,并且检测速度为 108.5 FPS,改善了漏检、误检和检测精度低的问题,同时仍具有较高的实时性,更适用于行人翻越护栏的实时性检测。

关键词: 翻越护栏检测;Dual Vision Transformer;特征融合增强;EMA

中图分类号: TP391.4;TN914 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi scale feature fusion enhanced pedestrian crossing guardrail detection

Liu Gang^{1,2} Hou Enxiang² Huang Sungang¹ Yan Shuguang² Huang Yingzheng²

(1. Jiangsu Province Engineering Research Center of Integrated Circuit Reliability Technology and Testing System, Wuxi University, Wuxi 214105, China; 2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Aiming at the problems of omission, misdetection and low detection accuracy of pedestrian crossing guardrail detection in the complex scenarios of occlusion, dense multi-target situations as well as multiple people overtopping, a multi-scale feature fusion and enhancement algorithm for pedestrian overtopping guardrail detection is proposed. Firstly, an algorithm based on Dual Vision Transformer and SCConv, which is applied to the backbone network, enhances the capture of global context information and finer-grained information, and improves the local fine feature extraction and feature fusion capability of the network; second, a multi-scale feature fusion enhancement module AM-SPPFCSP is proposed, which compensates for the feature loss caused by maximal pooling, improves the feature map. The richness and completeness of the feature map is improved, and the multi-scale feature extraction and feature fusion capability is enhanced; in addition, the feature fusion layer is further refined by replacing the ordinary convolution with GSConv and designing the VOV-GSCCSP module based on GSConv and SCConv, which effectively reduces the computational cost and the complexity of the model, while maintaining a higher degree of accuracy; finally, a highly efficient multi-scale feature fusion module, AM-SPPFCSP, is introduced in the trunk to reduce the complex background and the complex background and the complex background and the complex background. Attention EMA, which reduces the interference of irrelevant targets in the complex background and fuses the multi-scale information to achieve richer feature aggregation. The experimental results on the homemade pedestrian over guardrail dataset show that the proposed algorithm in this paper achieves 93.6% mAP with the addition of fewer parameters, which is 4.5% higher than that of the original model, and has a detection speed of 108.5 FPS, which improves the problems of leakage, false detection and low detection accuracy, while still having a high real-time performance, and is more suitable for real-time detection of pedestrians crossing the guardrail.

Keywords: climb over guardrail detection;Dual Vision Transformer;feature fusion enhancement;EMA

0 引言

“行人过马路,要走斑马线”。这在日常出行中,已经是

大家都熟知的交通常识,但生活中还是有很多人为图方便,放着人行道不走,而是冒着危险选择翻越护栏抄近路的交通违法行为。道路交通护栏是为了实现人车分流和车道分隔

收稿日期:2024-06-19

* 基金项目:国家自然科学基金(62204172)项目资助

而设置的物理障碍物。当行人跨越护栏过街时,很容易与直车辆发生冲突,引发车辆碰撞和刮擦事故,导致车辆和交通设施受损,并引起道路交通混乱和拥堵,如果躲避不及时,还会造成行人受伤或死亡。因此,对行人在道路中间翻越护栏的行为进行检测,减少相关的一系列安全事故尤为必要。

有研究人员使用基于传感器的技术检测异常行为^[1]和使用专用分类器及混合高斯模型法来检测翻越行为^[2]。但这些传统的检测算法的实时性和检测精度不能够满足于对实时性要求较高的复杂道路场景中。随着深度学习的不端发展和进步,基于深度学习的检测算法已在交通违法检测中得到了广泛应用,检测性能得到了较大的提高。为了更好的保护行人、司机和避免出现交通混乱和拥堵的现象,研究基于深度学习的目标检测算法对行人翻越护栏的行为进行检测具有较大的社会意义。

基于深度学习的目标检测算法分为两阶段算法和单阶段算法。两阶段算法包括 R-CNN 系列^[3-5]和 R-FCN^[6]等算法,单阶段算法有 YOLO 系列^[7-10]、SSD^[11]、RetinaNet^[12]、EfficientDet^[13]等算法。两阶段算法通常拥有较高的检测精度,但由于它有两个独立的处理阶段,计算成本较大,难以满足实时性的检测。单阶段算法直接在图像上进行目标检测,检测速度相对更快。针对道路交通类和异常行为的检测,一般需要较高的实时性,所以许多国内外研究人员对单阶段算法进行各种改进。王林等^[14]结合改进的 YOLOv5 和 SlowFast^[15]的时空检测算法对翻越行为进行检测,将 Fast 和 Slow 分支的残差结构分别改为 AC 残差结构和 SE 残差结构,增强了模型对重要特征和细粒度特征的学习能力,检测性能得到了较大提升。王源鹏等^[16]在 YOLOv5s 骨干网络的最后一层引入了基于 Transformer 编码的 C3TR 模块,以更好地提取丰富的全局信息和融合不同尺度的特征;同时在特征融合网络中嵌入了通道注意力模块(squeeze-and-excitation, SE),以更好地关注主要信息,从而提高模型的精度。但上述方法均未考虑到实际场景中的小目标检测问题。针对这一问题,罗玉涛等^[17]在网络的特征提取部分,构建了基于通道注意力的 C3-TCA 模块,以提高模型对相似特征的辨别能力,在特征增强部分,通过双路径增强结构融合浅层和深层特征,并优化了预测分支的数量和下采样比例,从而提升了对微小交通标志的检测精度。井方科等^[18]提出了一种基于多尺度特征融合的小目标交通标志检测算法。设计了一种双向自适应特征金字塔网络,充分利用细节特征和跳跃连接,增强多尺度特征融合。然后又提出双头检测结构,聚焦于小目标特征信息,同时减少了模型的参数量。然而,对遮挡场景下的目标检测有一定的误检、漏检问题。薛珊等^[19]提出了一种改进 YOLOX 的遮挡检测算法,引入了坐标注意力机制,以增强无人机目标图像的显著度,并突出有用特征,抑制无用特征,接着,通过去除特征融合层中自下而上的路径增强结构,降低了网络的复杂度。同时,设计了自适应特征融合网

络结构,增强了有用特征的表达力,提升了遮挡场景下的检测精度。李富等^[20]针对遮挡场景,提出了一种改进 YOLOv7 的检测算法,首先将模型的 backbone 和 head 部分的 ELAN 模块替换为语义上下文增强模块(contextual transformer, CoT),以提高上下文语义信息的捕获能力。然后,将自注意力双向 Transformer 模块(Biformer)与 SPPCSPC 模块融合,提升了模型在复杂场景和遮挡目标下的处理能力。然而,该方法在密集场景下的检测效果仍存在一定的局限。为了提高密集场景下的目标检测不佳的问题,高昂等^[21]提出了一种改进 YOLOv8 的检测算法,通过引入可变形卷积改进骨干网络,以增强特征提取能力,并设计了遮挡感知注意力机制,以增强遮挡行人可见部分的特征。此外,为了解决密集行人场景中检测头定位不准的问题,设计了动态解耦头,以增强对多尺度行人特征的关注,从而提高检测头的表达能力。王泽宇等^[22]采用 MobileViT 作为主干网络,增强了模型在行人聚集区域的整体特征提取能力,使用排斥损失(repulsion loss^[23], RL)作为边界框损失函数,减少了在人群密集情况下小目标行人的漏检和误检。李忠科等^[24]提出了一种改进 YOLOv8n 的检测算法,将大目标检测层替换为小目标检测层,并调整网络结构,提高检测精度,并应用 GhostConv 与 DWConv 将原 C2f 模块设计为 C2f-GhostD 模块,减少计算成本,但对遮挡以及多目标密集情况下的检测问题未能改善,有一定的漏检情况;潘海鸿等^[25]基于 YOLOv8n 模型,构建了一种 MFBNet 轻量化骨干网络,引入了 MBConv 模块减少了模型计算量,同时加入 CBAM 注意力模块,提高检测精度,并基于 AKConv 提出了 AVCStem 模块,进一步减少了模型参数,但对遮挡和多目标干扰场景下的检测效果欠佳,存在误检和漏检问题。上述改进算法虽然具有一定的贡献,但检测性能仍然不佳,还存在一定的漏检、误检问题和检测精度低的问题,难以满足在遮挡、未翻越行人干扰、多目标密集以及多人翻越护栏等复杂道路场景下的实时性检测。

针对以上问题,本文基于 YOLOv8n 模型,提出了一种实时高效的行人翻越护栏检测算法。关于 SAGE-YOLO 模型将在第 2 章详细阐述,实验及结果分析将在第 3 章详细介绍。

1 YOLOv8n 模型介绍

YOLOv8n 算法是目前比较先进的一阶段目标检测算法之一,具有检测速度快、检测精度高的特点,在实时交通违法行为的检测中,具有重要的应用价值。所以选用 YOLOv8n 作为检测行人翻越道路护栏的基础模型,以更好满足实时性检测的要求。

YOLOv8n 的网络结构分为输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和检测头(Head)四部分。输入端采用 Mosaic 数据增强,在数据集中随机选择 4 张图像进行随机裁剪、缩放等变换,再按照一定的规则拼接成一个

新的图像,丰富了数据样本,增强了模型的泛化能力。Backbone 部分由 Conv、C2f、SPPF 模块组成。C2f 模块的主要作用是增强特征提取能力,它包含两个通过残差连接相连的卷积层,这种设计有效地缓解了深度神经网络中的梯度消失和表达能力受限的问题。颈部网络采用 PANet 结构,包含了特征融合层 FPN 和路径聚合层 PAN。Head 端,除了使用无锚框设计,还采用了解耦头(decoupled head, DH)的结构设计,将分类和回归进行解耦分离,这使得模型在处理各种任务时能够更加高效和准确。

2 SAVE-YOLO 模型结构

本文提出了一种基于 SAVE-YOLO 的行人翻越道路栏杆的检测算法,模型结构如图 1 所示。SAVE-YOLO 以 YOLOv8n 模型为基础,首先将主干网络末端的 C2f 模块

改进为 SCDVT 模块,增强了精细特征的提取能力和对全局上下文信息的捕获能力,SCDVT 模块由双路径视觉变换器(dual vision transformer^[26],DVT)和空间和通道重建卷积(spatial and channel reconstruction convolution^[27],SCConv)组成。其次,将原来的 SPPF 模块改进,提出了多尺度特征融合增强模块 AM-SPPFCSPC,减少最大池化后引起的特征信息丢失,使原始图像特征保留更完整的信息,提高模型的特征提取能力。再将特征融合层细化,颈部的 C2f 模块替换为由 SCConv 和 GSConv 构成的 VOV-GSCCSP 模块,Conv 改进为 GSConv,在减少较多参数量和计算量情况下,并保持较高的精度。最后在主干网络的前 3 个 C2f 中引入 EMA^[28] 注意力机制,以融合多尺度信息,实现更丰富的特征聚合,使模型聚焦于关键特征,减少无关特征干扰,提高行人翻越护栏检测的性能。

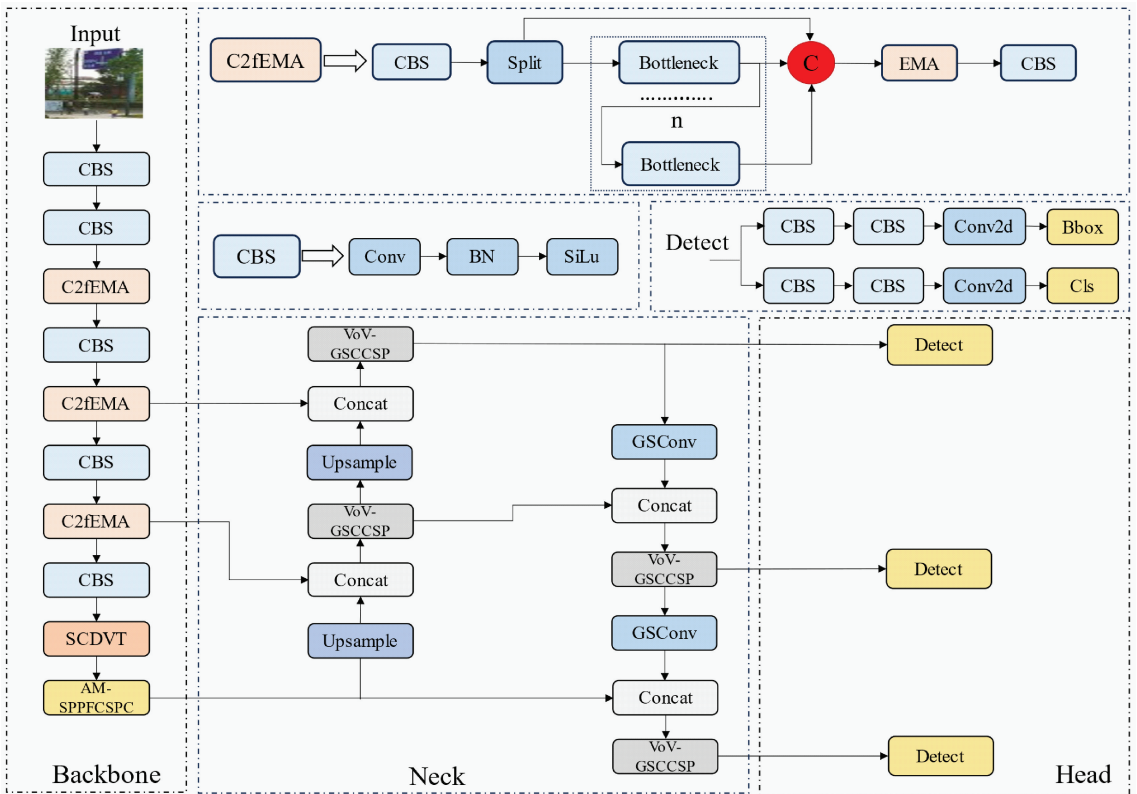


图 1 SAGE-YOLO 算法模型结构图

Fig. 1 Model structure of SAGE-YOLO algorithm

2.1 SCDVT 模块的设计

1) 融入双路径视觉 Transformer

在 YOLOv8n 的主干网络中,利用多层卷积和残差连接来提取图像特征信息,侧重对局部区域特征的提取,但对全局信息的获取和上下文信息的整合能力较差,这可能使得目标细节信息的丢失,尤其在密集的复杂道路背景下对行人翻越护栏有误检和检测准确度低的情况。

为解决上述问题,本文提出了一种融入 DVT 模块改进主干网络结构,DVT 通过两种交互路径来模拟自注意

力学习来增强主干的特征提取能力,有效地补偿了全局特征压缩的信息损失,同时还使得提取更精细的局部特征难度降低。

DVT 是一种多尺度 ViT 主干网络,与现有依赖局部自注意力或下采样操作的多尺度 ViT 不同,它将自注意力机制分解为像素路径和语义路径的建模。像素路径,是通过在像素级别细化输入特征来获取更细粒度的信息。语义路径在全局级别将特征抽象成高级语义标记。DVT 结构如图 2 所示。

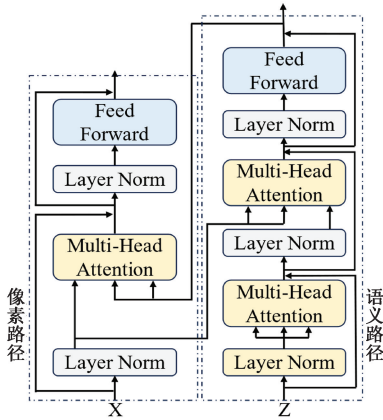


图 2 DVT 结构图

Fig. 2 Structure of DVT

形式上,对于第 l 个 DVT,使用额外的参数语义查询 $z_l \in \mathbb{R}_{m \times d}$ 来对其进行扩充。语义路径开始先使用自注意力来对语义查询进行上下文编码,接着利用交叉注意力来从精炼语义查询和输入特征 x_l 之间提取语义标记,然后通过前馈层进行处理。上述操作的具体过程如式(1)~(4)所示。

$$x_l^n = LN(x_l), z_l^n = LN(z_l), \quad (1)$$

$$z_l' = MHA(z_l^n, z_l^n, z_l^n) + z_l, \quad (2)$$

$$\tilde{z}_l = MHA(LN(z_l'), x_l^n, x_l^n) + z_l', \quad (3)$$

$$z_{l+1} = FFN(LN(\tilde{z}_l)) + \tilde{z}_l. \quad (4)$$

接下来,语义标记 z_{l+1} 被传递到像素路径中,作为高级语义的先验信息。同时,将这些语义标记视为增强的语义

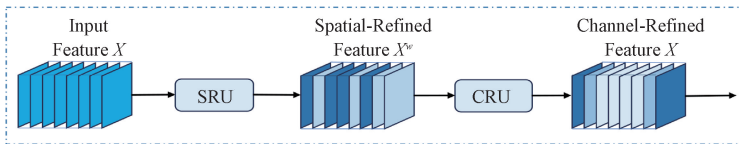


图 3 SCConv 结构图

Fig. 3 Structure of SCConv

用于空间冗余的空间重建单元 SRU,包括分离和重建操作。分离操作的主要目的是把包含丰富信息的特征图与空间内容相关的信息较少的特征图分离开来。首先对输入特征 X 进行标准化,方法是将其减去平均值 μ ,然后除以标准差 σ 。再在组归一化 GN 层中,使用可训练参数 $\gamma \in R^c$ 来度量各批次与通道的空间像素方差。能够表示不同特征图重要程度的归一化相关权重,由式(8)所示。

$$W_r = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^c \gamma_j}, i, j = 1, 2, \dots, C \quad (8)$$

然后,特征图由 W_r 重新加权的权重值通过 Sigmoid 函数映射到区间 $(0, 1)$,再通过门控进行阈值处理。最后,输入特征 X 与 W_1, W_2 分别相乘,获得了两个加权特征:信息量较为丰富的 X_1^o 和信息量比较少的 X_2^o 。

查询,并将其输入至下一个 Dual 块的语义路径之中。

对于像素路径,它与传统 Transformer 块的作用相似,但其不同之处在于还利用从语义路径派生的语义标记,通过交叉注意力机制来进一步精炼输入特征。详细来讲就是像素路径把语义标记 z_{l+1} 视作键/值,然后进行以下操作来执行交叉注意力,如式(5)~(7)所示。

$$x_l^n = LN(x_l), z_{l+1}^n = LN(z_{l+1}), \quad (5)$$

$$x_l' = MHA(x_l^n, z_{l+1}^n, z_{l+1}^n) + x_l, \quad (6)$$

$$x_{l+1} = FFN(LN(x_l')) + x_l'. \quad (7)$$

最后两条路径的输出会合并在一起,进一步输入多头自注意力。

因此,选择将 DVT 融入到 YOLOv8n 主干网络的 C2f 模块中,如图 2 所示,利用 DVT 的特性,使得增强的自注意力信息可以沿着两条路径并行传播,有效的促进了 YOLOv8n 模型对更精细特征信息的提取,以及对全局上下文信息捕获,能够达到准确性-延迟平衡。

2) 引入空间和通道重建卷积 SCConv

卷积神经网络(CNN)在各种目标检测任务中取得了显著的检测性能,但却耗费了巨大的计算资源,部分原因是网络中的卷积层提取了冗余特征。为了减少冗余计算,平衡 DVT 模块融入带来的额外参数量,并增强代表性特征学习,由此引入了 SCConv。

SCConv 由两个结构单元组成,空间重建单元(SRU)和通道重建单元(CRU)。具体来说,残差模块中的中间输入特征 X ,先经过 SRU 操作获得空间细化特征 X^o ,接着通过 CRU 操作获得通道细化特征 Y 。图 3 为 SCConv 的模块结构图。

在减少空间冗余的重构操作中,使用交叉重建操作,充分组合和加权两个不同的信息特征,以增强他们之间的信息传递。接着,将交叉重建后的特征 X_1^o 和 X_2^o 进行连接,生成空间细化的特征图 X^o 。整个重构过程用式(9)表示:

$$\begin{cases} X_1^o = W_1 \otimes X \\ X_2^o = W_2 \otimes X \\ X_{11}^o \oplus X_{22}^o = X^{\omega 1} \\ X_{21}^o \oplus X_{12}^o = X^{\omega 2} \\ X^{\omega 1} \cup X^{\omega 2} = X^o \end{cases} \quad (9)$$

式中: \otimes 逐元素乘法, \oplus 为逐元素加法, \cup 是 Concat 拼接操作。

SRU 操作应用到中间的输入特征 X 之后,有效地将

信息量丰富和信息量较少的特征都分离开来,并且将其重建来增强具有代表性的特征,并减少空间维度上冗余的特征。图 4 为 SRU 的模块结构图。

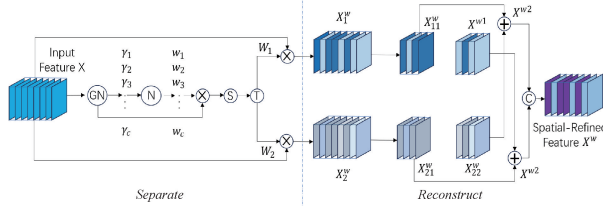


图 4 SRU 模块结构图

Fig. 4 SRU module structure

为了限制重构操作后特征 X^w 的通道冗余,进一步的减少网络参数量,在 SRU 后引入通道重建模块(CRU),它包括分割、转换、融合 3 种操作。CRU 的模块结构图如图 5 所示。

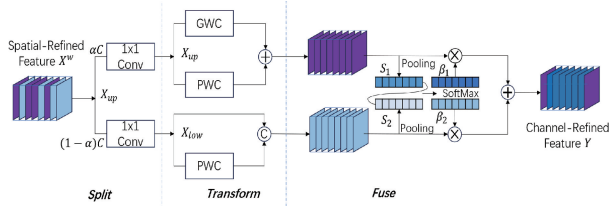


图 5 CRU 模块结构图

Fig. 5 CRU module structure

对于重构操作后的空间细化特征 X^w , 先将其通道分为 αC 通道和 $(1-\alpha C)$ 通道两部分,再通过一个 1×1 的卷积以压缩特征图的通道,并使用压缩比系数 r 来控制特征通道,平衡通道重建模块的计算成本。经上述操作后, X^w 被分成了上部 X_{up} 和下部 X_{low} 。转换阶段内, X_{up} 被传递到上层转换阶段,以高效的卷积运算(GWC 与 PWC)替换标准 $k \times k$ 卷积,来获取高级的代表性信息并且减少计算成本。接着,在同一个上部 X_{up} 去采取 $k \times k$ GWC,以及 1×1 PWC 的操作。最后,对输出进行逐元素加法合并为特征图 Y_1 。 X_{low} 被送到较低层转换阶段,应用 1×1 PWC 操作并重用 X_{low} 的特征,将其连接融合,得到特征图 Y_2 。在融合阶段,应用 SKNET 自适应地合并特征 Y_1 与 Y_2 , 再经全局平均池化聚合全局空间信息 $S_m \in \mathbb{R}^{c \times 1 \times 1}$, 将 S_1, S_2 进行堆叠后执行通道软注意力操作,得到特征重要性向量 $\beta_1, \beta_2 \in \mathbb{R}^c$, 再按照通道方向合并 Y_1 与 Y_2 得到通道细化特征 Y 。细化特征公式如下:

$$Y = \beta_1 Y_1 + \beta_2 Y_2 \quad (10)$$

在 SCCConv 卷积模块中,参数都集中于转换操作阶段,相较于传统卷积而言,SCCConv 在提高精度的同时又大幅减少了参数量,降低了运算复杂度。这一特性使得融入双路径视觉 Transformer 后参数量和计算量得到平衡,保证了引入 SCDVT 模块在提升精度的同时又不会产生较多参数量和运算量。

3) SCDVT 模块的组成

SCDVT 模块由 Dual-ViT 和 SCCConv 组成。根据原 C2f 模块中的结构,将 C2f 中的 2 个标准卷积替换为 SCCConv,然后用 n 个 Dual-ViT 模块替换主干网络的最后一个 C2f 模块,SCDVT 模块如图 6 所示。

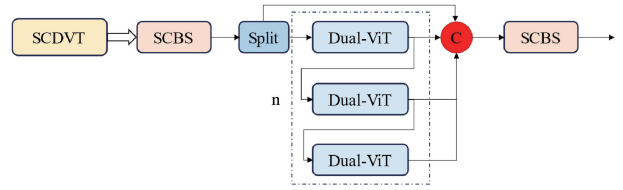


图 6 SCDVT 模块图

Fig. 6 SCDVT module diagram

综上所述,SCDVT 模块的融入,增强了更精细特征的提取,减少了全局特征压缩的信息损失,能够有效的捕获全局上下文信息,同时没有引入较多的参数量和计算量,提升了网络的整体性能以及在遮挡场景下对行人翻越栏杆的检测准确度。

2.2 多尺度特征融合增强模块

在 YOLOv8 中,使用了快速金字塔池化(SPPF)模块,处理不同尺度输入特征,更好的捕获不同层次的图像特征,从而提供更丰富的特征表达。虽然 YOLOv5、YOLOv8 中的 SPPF 模块和 YOLOv7 中的 SPPCSPC 模块能够较好的融合全局信息和局部信息,并保持空间信息,具有多尺度特征提取的能力,但只采取了最大池化提取特征的策略,容易导致局部细节信息的丢失,在解决行人翻越栏杆检测的遮挡问题上有着一定的局限性。为了应对上述问题,缓解遮挡和多人翻越等复杂背景下导致的漏检和检测精度低的情况,进一步提高多尺度特征提取和特征融合能力,增强特征图的丰富性,提出了多尺度特征融合增强模块 AM-SPPFCSPC,如图 7 所示。此模块从 SPPF 以及 SPPCSPC 的设计中获得灵感,新引入了一条新的平均池化支路,以应对多次最大池化后引起的部分特征信息的丢失,使原始图像特征保留更完整的信息。两种池化组合为混合池化 Hybrid Pooling 模块,如图 8 Hybrid Pooling 结构图所示,主要由 3 个串行的最大池化层和 3 个串行的平均池化层并行组成。Hybrid Pooling 丰富了特征的多样性,提升了模型的泛化能力以及特征的表达能力。此外,还将激活函数 SiLu 替换为 Mish 激活函数,它拥有良好的泛化能力和结果的有效优化能力,能够更好的传播梯度信息,提高模型的学习能力。

AM-SPPFCSPC 模块的工作流程,输入特征经由并联的两条路径进行处理,在其中一条路径中,输入特征依次通过卷积核大小为 $1 \times 1, 3 \times 3, 1 \times 1$ 的 3 个 CBM 卷积模块提取特征信息,然后进入混合池化模块中,进行多尺度的特征提取和融合。如图所示。随后由卷积核大小为 1×1 的卷积模块降低其通道数和减少参数量,再将处理后的

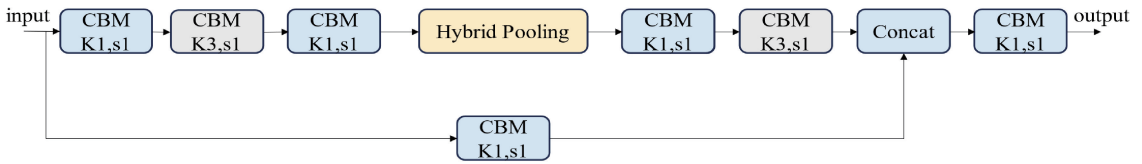


图 7 AM-SPPFCSPC 模块图

Fig. 7 AM-SPPFCSPC module map

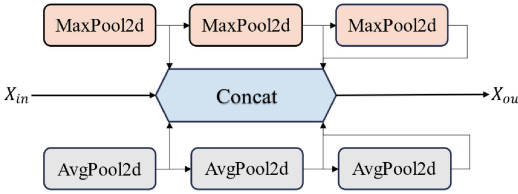


图 8 Hybrid Pooling 结构图

Fig. 8 Hybrid Pooling structure

输出利用 3×3 的卷积模块来恢复通道数得到一个输出。另外一只路径上,初始输入经过 1×1 的卷积模块后,与上一条支路中的输出连接,最后使用 1×1 的卷积模块以确保输出通道数与原始输入特征图相同。Hybrid Pooling 模块部分的计算步骤如式(11)所示。

$$\begin{aligned}
 S_1 &= C(MP(t), MP(MP(t)), MP(MP(MP(t)))) \\
 S_2 &= C(AP(t), AP(AP(t)), AP(AP(AP(t)))) \\
 S &= C(S_1, S_2, t)
 \end{aligned} \tag{11}$$

式中: S_1 和 S_2 分别为最大池化支路和平均池化支路的输出, S 表示两种池化后的输出与输入 t 拼接后的输出。

2.3 特征融合层的细化(RFFL)

传统标准卷积需要较多的参数来学习输入图像的特征,导致模型的参数增多,同时也增加整个模型训练和推理的计算量。一些轻量化网络模型 MobileNets、ShuffleNets 等,使用深度可分离卷积(depth-wise separable convolution, DSC),减少了计算量,显著提升了模型检测的速度和效率,但在通道卷积过程中会有一些信息损失,特征的表达能力也会减弱,导致模型检测精度降低。为解决上述问题,本文引入了轻量化卷积 GSConv,以减轻模型的大小,同时保持较高的精度。GSConv 结构如图 9 所示。

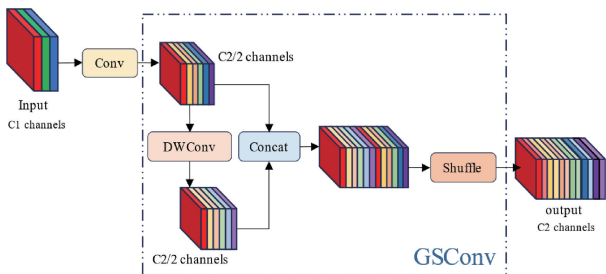


图 9 GSConv 结构图

Fig. 9 GSConv structure

GSConv 由标准卷积(Standard Convolution, SC)、DSC 和 shuffle 操作组成。通道数为 C_1 的输入特征首先经过一个标准卷积,通道数变为 $C_2/2$,接着通过 DW 卷积后通道数保持不变,再与经过标准卷积处理过的结果相结合,最后通过 shuffle 混洗操作,输出通道数变为 C_2 ,SC 生成的特征信息融入至 DSC 的输出结果中,增强信息的提取能力和通道的信息交互,提高特征的表达能力。

GSConv 最大程度的保留了特征图通道间的信息连接。但如果在模型的每个阶段都使用它,会使得网络层数更深,会增大对数据流的阻力,推理时间显著增加。当特征图进入特征融合 Neck 层时,已变得细长(通道维度达到最大,宽高维度达到最小)。所以本文在 Neck 层引入 GSConv,代替 SC,并在 GSconv 和 2.1 节中 SCConv 的基础上,设计了 GSCbottleneck 模块,如图 10(a)所示。并使用一次性聚合的方法设计了,以替换颈部的 C2f 模块,进一步降低计算成本和模型网络的复杂性,同时保持较高的精度。VOV-GSCCSP 模块如图 10(b)所示。

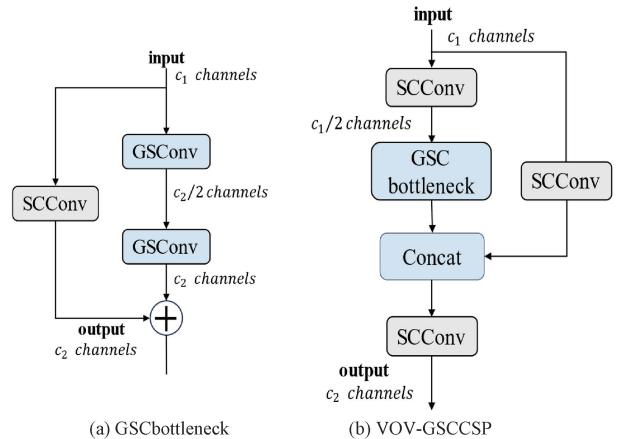


图 10 GSCbottleneck 模块和 VOV-GSCCSP 模块

Fig. 10 GSCbottleneck module and VOV-GSCCSP module

2.4 高效多尺度注意力

在存在遮挡的道路场景下,行人翻越护栏的检测效果容易受到影响出现误检、漏检的情况,为了减少无关特征的干扰,增强模型的特征提取能力,因此在模型中融入 EMA 注意力机制,它可以保留每个通道的信息并减少计算成本,使得模型更加关注感兴趣目标区域,以提升检测性能。EMA 注意力模块如图 11 所示。

在通道维度上,将输入特征 X 划分成 G 个子特征,

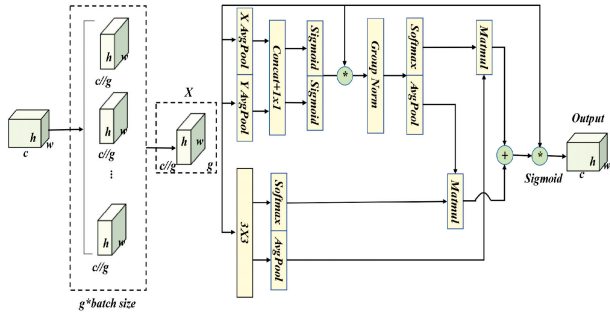


图 11 EMA 注意力机制图

Fig. 11 EMA attention mechanism diagram

$X = [X_0, X_i, \dots, X_{G-1}]$, $X_i \in \mathbb{R}^{C//G \times H \times W}$ ($G \ll C$), 然后用学习到的注意力权重来增强每个子特征中感兴趣区域的特征表示。为了捕获跨所有通道的依赖关系并减轻计算预算,对通道方向的跨通道信息交互进行建模,采用 1D 水平全局平均池化 X Avg Pool、1D 垂直全局池化 Y AvgPool、 3×3 卷积的 3 条并行路径进行处理,前两条路径的特征沿图像高度方向连接后共享 1×1 卷积,然后输出的两个特征向量经过 Sigmoid 函数激活,另一条路径使用 3×3 卷积操作捕获局部跨通道交互,以扩大特征空间。在跨空间学习部分,利用 2D 全局平均池化分别将 1×1 分支和 3×3 输出的全局空间信息编码,再通过非线性函数 Softmax,最小分支的输出转换为相应的维度形状,最后,生成的两个空间注意力的权重值通过跨空间交互模块进行聚合,建立远程依赖关系,捕获像素级成对关系并突出所有像素的全局上下文信息,融合了多尺度信息,实现更丰富的特征聚合。并通过 Sigmoid 函数输出特征映射,增强了模型对目标的重点关注度。

3 实验及结果分析

3.1 实验平台

本文对改进模型进行训练与测试的实验环境平台基于 Ubuntu(20.04)操作系统,具体环境配置如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

配置名称	版本参数
GPU	NVIDIA GeForce RTX 3090
CPU	Intel(R) Xeon(R) Platinum 8358P CPU@2.60 GHz
深度学习框架	PyTorch1.11.0
编程语言	Python3.9
CUDA	Cuda11.4

3.2 实验数据集介绍

目前未有行人翻越道路护栏的公共数据集,因此本文自建了 Cross 数据集,其中包括网络爬取和摄像机所拍摄

的行人翻越道路护栏图像,共有 3 426 张。图片的像素大小为 925×716 ,图片尺寸大小为 $10.5 \text{ cm} \times 6.1 \text{ cm}$,3 426 张图片中均至少含有一个行人翻越护栏的违法行为,其中有 279 张图片存在遮挡场景,存在行人且车辆密集的多人翻越场景 332 张,存在较多行人和车辆的复杂道路场景有 445 张,存在行人及遮挡的多人翻越场景有 217 张。然后用 Labeling 数据集标注工具标注,行人翻越护栏行为标注为 climb,完成后将图片输出为 YOLO 格式的 txt 文件。训练集、验证集和测试集,按照 7 : 2 : 1 的比例进行划分,其中训练集 2 398 张,验证集 685 张,测试集 343 张。

3.3 模型训练方法

本文在行人翻越护栏检测训练中,输入图像设置为 640×640 的分辨率,并采用 SGD 优化器。部分超参数如表 2 所示。

表 2 算法的部分超参数

Table 2 Some hyperparameters of the algorithm

超参数名称	数值
lr0	0.01
lrf	0.1
momentum	0.937
batch-size	16
epochs	300
workers	6

3.4 算法评估指标

为了准确评估模型检测行人翻越护栏的行为的准确性,采用准确率 (precision, P)、召回率 (recall, R)、参数量 (Params)、检测速度 (frames per second, FPS) 和平均精度均值 (mean average mAP, mAP) 作为算法评估指标。相关公式如式 (12)~(15) 所示。

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$AP = \int_0^1 P(r) d(r) \quad (14)$$

$$mAP = \frac{1}{n_c} \sum_{i=1}^{n_c} AP_i \quad (15)$$

TP 表示正确检测的正样本数,FP 表示错误检测的正样本数;FN 表示未检测到的正样本数; n_c 为总类别数。

3.5 实验结果与分析

1) 消融实验

为了验证 SAVE-YOLO 算法对行人翻越护栏检测的有效性,以 YOLOv8n 为基准模型,在自制数据集 Cross 上将各个改进方法以相同的实验环境配置进行消融实验,符号“ \checkmark ”表示往基线算法中添加相应的模块,实验如表 3 所示。

表 3 消融实验

Table 3 Ablation experiments

模型	SCDVT	AM-SPPFCSPC	RFFL	EMA	P/%	R/%	mAP/%	Param/M
YOLOv8n					89.4	81.6	89.1	3.21
Improve 1	✓				92.9	83.8	91.7	3.97
Improve 2		✓			92.0	82.3	90.6	3.31
Improve 3			✓		89.7	82.4	89.2	2.82
Improve 4				✓	91.8	81.8	90.2	3.24
Improve 5	✓		✓		92.6	86.5	92.8	3.52
Improve 6	✓	✓		✓	92.6	87.2	93.5	4.12
本文算法	✓	✓	✓	✓	92.8	87.4	93.6	3.72

由表 3 可以看出,引入 SCDVT 模块后,mAP 提高了 2.6%,表明 SCDVT 模块的添加,增强了更精细特征的提取能力,有效的捕获了全局上下文信息,在不产生较多参数的同时,较大的提升了复杂道路场景下对行人翻越栏杆的检测准确度;添加 AM-SPPFCSPC 模块后,其 mAP 比 YOLOv8n 算法提高了 1.5%,这表明 AM-SPPFCSPC 的添加,减少了局部特征细节信息的丢失,增强了多尺度特征提取特征融合的能力,从而提高模型的检测精度;对特征融合层的细化后,Params 减少了 0.39 M,mAP 提高了 0.1%,表明颈部网络在细化后,模型的参数大大减少,同时保持了较高的精度精度当 EMA 引入主干后,mAP 提升了 1.1%,这表明高效多尺度注意力 EMA 的引入后,模型能更加关注于行人翻越道路护栏的区域,提高了在多目标干扰的复杂道路场景中的检测能力;添加 SCDVT 模块并进行特征融合层的细化,Params 比只添加 SCDVT 模块时减少了 0.25 M,表明 GSConv 和 SCConv 组成的模块有效的平衡了 Transformer 引入后的参数量;在 SCDVT 模块、AM-SPPFCSPC 模块和 EMA 注意力模块同时引入时,mAP 提高了 4.4%,表明这 3 个模块的同时添加对模型的检测提升较大。对实验结果进行综合分析,4 种改进方法均有效的提升了 YOLOv8n 的检测精度。4 种改进方法的综合引入,在参数量增加不多的情况下,SAVE-YOLO 的 mAP 达到了 93.6%,比 YOLOv8n 算法高了 4.5%,准确率和召回率分别提高了 3.4%、5.7%,其检测效果最佳。

2) 注意力机制对比

为了进一步验证 EMA 注意力机制对行人翻越护栏检测的有效性,以 YOLOv8n 为基准模型,选取了一些主流注意力模块如 SE、CBAM、ECA、CA 与 EMA 在自制 Cross 数据集上以相同的环境配置进行对比实验。如表 4 所示。

由表 4 可知,上述 5 种注意力对行人翻越护栏检测的性能均有提升,但程度有所差别。与 Baseline 模型相比,SE、ECA 注意力机制添加到网络后,mAP 分别增加了 0.2%和 0.7%,GFLOPs 和 Params 均未增加;CBAM 加入后,mAP 增加了 0.5%,但 GFLOPs 和 Params 均比其他注意力机制增加的多;CA、ECA 加入后,模型的 mAP 分

表 4 不同注意力机制的对比实验

Table 4 Comparative experiments with different attention mechanisms

方法	mAP/%	GFLOPs	Param/M
Baseline	89.1	8.21	3.21
+SE	89.3	8.21	3.21
+CBAM	89.6	8.29	3.32
+ECA	89.8	8.21	3.21
+CA	90.1	8.27	3.26
+EMA	90.2	8.25	3.24

别增加了 1.0%和 1.1%,加入 ECA 后的 mAP 达到了最高的 90.2%,并且 GFLOPs 和 Params 相对增加的较少,性能优于其他注意力模块,验证了 EMA 注意力机制的有效性。

3) 特征融合模块性能对比

为了进一步验证 AM-SPPFCSPC 对行人翻越护栏检测的有效性,以 YOLOv8n 为基准模型,选取了一些特征融合模块如 SPPCSPC、SPPFCSPC、AM-SPPFCSPC 与在自制 Cross 数据集上以相同的环境配置进行对比实验。如表 5 所示。

表 5 不同特征融合模块性能对比

Table 5 Performance comparison of different feature fusion modules

方法	mAP/%	GFLOPs/G	Param/M
Baseline(SPPF)	89.1	8.21	3.21
SPPCSPC	89.9	11.24	3.28
SPPFCSPC	90.2	11.24	3.28
AM-SPPFCSPC	90.6	11.26	3.31

由表 5 可知,相比 Baseline 的 SPPF,改进成 SPPCSPC 和 SPPFCSPC 的 GFLOPs 和 Params 相比原来分别都增加了 3.03 G 和 0.07 M,mAP 分别增加了 0.8%和 1.1%,改进 SPPFCSPC 的模型要比改进为 SPPCSPC

的模型的检测性能要好。本文提出的 AM-SPPFCSPC 模块,虽然参数量和计算量略有增长,但 mAP 提升 1.5%,达到了最高的 90.6%,证明了此模块增强了多尺度特征提取和特征融合的能力,有效的提高了模型的检测性能。

4) 与其他算法的对比分析

为进一步验证 SAGE-YOLO 算法对行人翻越护栏检测的优势和有效性,在相同实验条件下,在自制电动车头盔数据集上将 SSD、PP-YOLOv2、YOLOX-s、YOLOv7、RT-DETR-r18 等 8 种单阶段检测算法和文献[33]中提出的基于 YOLOv8n 的改进算法与 SAGE-YOLO 算法进行实验对比分析,以 mAP、GFLOPs、Params、FPS 为评价指标。实验对比数据如表 6 所示。表 6 中的算法均为目前检测领域内性能表现较优的算法,文献[29-33]的算法与本文算法都是单阶段检测算法,实时性强且检测精度较高,适合于道路中交通违法行为的检测。

表 6 算法模型对比试验

Table 6 Algorithm model comparison test

模型	mAP/%	GFLOPs	Param/10 ⁶	FPS
SSD	75.8	124.3	26.45	76.3
CenterNet ^[29]	84.2	19.6	14.8	92.4
YOLOv5s	87.2	15.8	7.02	102.3
PP-YOLOv2 ^[30]	88.4	118.2	56.73	94.2
YOLOX-s ^[31]	87.7	25.6	8.95	98.7
YOLOv7	88.9	114.5	37.32	87.6
RT-DETR-r18 ^[32]	89.4	58.3	19.98	100.2
YOLOv8n	89.1	8.2	3.21	117.4
文献[33]	90.8	12.3	3.17	102.6
SAGE-YOLO(本文)	93.6	12.8	3.72	108.5

由表 6 可知,在 Cross 数据集上的对比实验结果中,SAGE-YOLO 算法的 mAP 明显优于其他主流算法,达到了 93.6%;与 SSD、CenterNet 算法的检测结果对比,SAGE-YOLO 的检测精度遥遥领先,推理速度较快,且计算量和参数量远远低于这两个模型;与 YOLOv5s、PP-YOLOv2、YOLOX-s、YOLOv7 模型相比,SAGE-YOLO 的检测精度明显高于它们,推理速度与 YOLOv5s、YOLOv7 相当,计算量和参数量仅为 PP-YOLOv2 的 10.8%和 6.6%,YOLOv7 的 11.2%和 10.0%;与推理速度相差较小的 RT-DETR-r18 模型相比,SAGE-YOLO 的 mAP 比其高了 4.2%,参数量和计算量都低于 RT-DETR-r18;与 YOLOv8n 模型相比较,SAGE-YOLO 在参数量和计算量略微上涨的情况下,其 mAP 提高了 4.5%,达到了 93.6%,FPS 仍然能达到 108,满足目标检测的实时性,检测性能得到了显著提升。与最新改进的都是行为检测算

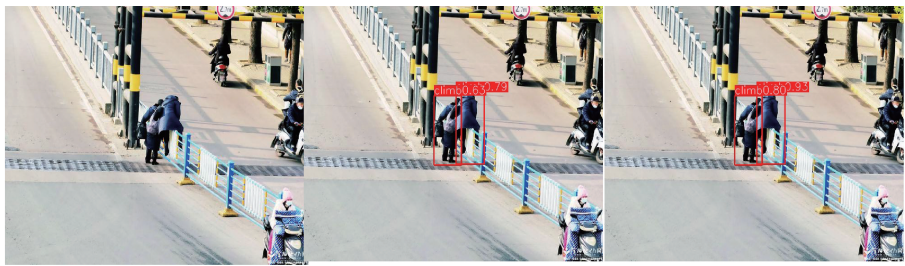
法的文献[32]相比,在运算量相当的情况下,本文算法在参数量上略高,但推理速度比其高 5.4%,mAP 比其高了 2.2%,在道路翻越护栏的行为检测中有较大的优势。综合上述模型的检测精度、复杂度和推理速度来看,本文算法在这些算法中拥有最为优异的表现,且能满足实时性检测,验证了 SAGE-YOLO 算法对行人翻越护栏检测的有效性和先进性。

5) 检测效果分析

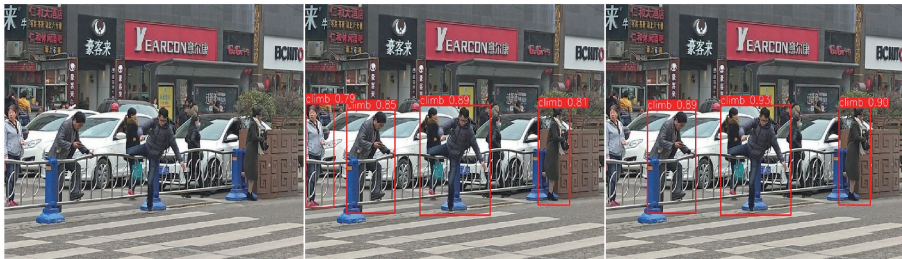
为了更好的检验 SAGE-YOLO 算法对行人翻越护栏的实际检测效果,取出部分测试集进行检测,并与 YOLOv8n 模型的检测结果对比,检测效果对比如图 12 所示。图 12 中最左侧图片均为测试集原图,中间图片为 YOLOv8n 模型的检测效果图,最右侧为 SAGE-YOLO 模型的检测效果图。

如图 12(a)为遮挡场景下的检测效果,可以看出 YOLOv8n 在此场景下的检测精度都低于 80%,而 SAGE-YOLO 模型的检测精度相对要高出很多,;如图 12(b)为存在未翻越行人且车辆密集的多人翻越场景下的检测结果,YOLOv8n 模型将图片最左侧的正常行人误检成翻越护栏的行为,而 SAGE-YOLO 正确检测出了翻越护栏行为,且无对其他目标的误检;图 12(c)为存在较多未翻越行人和车辆的场景的检测效果,由图可知,YOLOv8n 模型对斑马线旁未有明显翻越行为的黑衣男子误检成翻越护栏行为,而 SAGE-YOLO 正确检测出了图片中最左侧翻越护栏行为,且未对黑衣男子造成误检。如图 12(d)为存在未翻越行人及遮挡的多人翻越场景下的检测结果,YOLOv8n 模型仅将存在遮挡的蓝色衣服小伙和其右侧女子检测成一个翻越护栏行为,未将其两者的翻越行为全检测出来,并将最右侧正常行人误检为翻越护栏的行为,而 SAGE-YOLO 将翻越栏杆的行为全能正确检测出来,并未将最右侧正常行人误检为翻越护栏的行为。

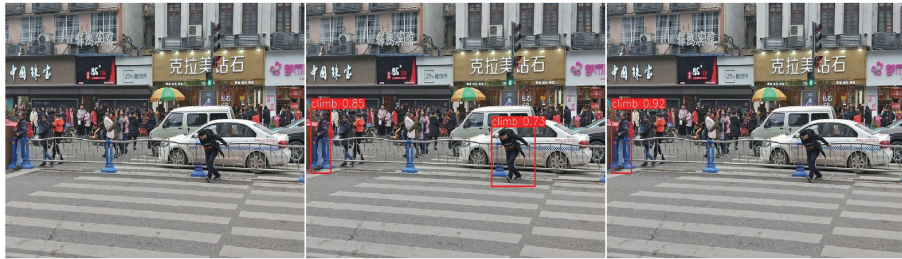
经上述检测结果分析,原因是添加 SCDVT 模块增强了局部精细特征的提取,减少了全局特征压缩的信息损失,能够有效的捕获全局上下文信息,提升了网络的整体性能以及在密集的复杂背景下对行人翻越栏杆的检测准确度和误检情况;多尺度特征融合增强模块 AM-SPPFCSPC,减少了局部细节信息的丢失,丰富了特征的多样性,能够更好的融合全局信息和局部信息,提升了模型的泛化能力、学习能力以及特征的表达能力,对遮挡和多人翻越场景的漏检和检测精度有良好的改善;引入 EMA 注意力机制,能够捕获像素级成对关系并突出所有像素的全局上下文信息,并融合多尺度信息,实现更丰富的特征聚合。增强了模型对行人翻越护栏行为的关注度,提升了模型对遮挡场景下的目标检测能力。因此 SAGE-YOLO 对行人翻越护栏检测有着更为优异的性能和鲁棒性。



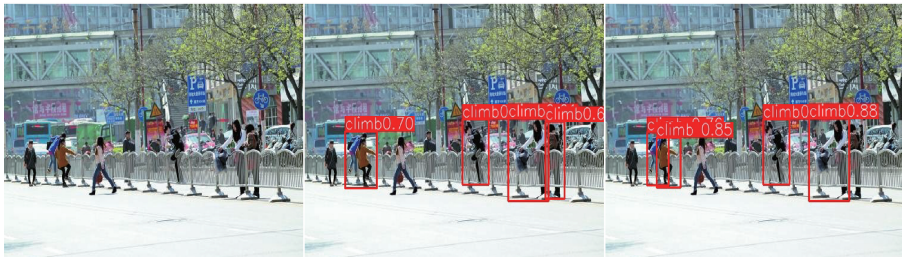
(a) 遮挡场景
(a) Shade scenes



(b) 存在未翻越行人且车辆密集的多人翻越场景
(b) Multiple rollover scenarios where there are no rollover pedestrians and dense vehicular traffic



(c) 存在密集未翻越行人和车辆的场景
(c) Scenarios where there is a dense concentration of uncrossed pedestrians and vehicles



(d) 存在未翻越行人及遮挡的多人翻越场景
(d) Presence of un-crossed pedestrians and sheltered multiple-crossing scenarios

图 12 原模型与改进模型在不同道路场景的实验结果对比图

Fig. 12 Comparison of experimental results between the original model and the improved model in different road scenarios

4 结 论

针对行人翻越护栏检测在有遮挡、车辆和行人密集以及多人翻越的复杂场景,出现的漏检、误检以及检测精度低的问题,提出了一种多尺度特征融合增强的行人翻越护栏检测算法。首先把提出的 SCDVT 模块应用于模型的主干网络,有效地补偿了全局特征压缩的信息损失,同时还提高了对更精细的局部特征的提取能力;其次,提出多尺度特征融合增强模块 AM-SPPFCSPC,弥补了特征信息的损失,丰富了特征的多样性,提升了模型的泛化能力以及特征的表达能力;另外,对特征融合层的细化,有效的降低

了计算成本和模型的复杂度,同时又保持较高的精度;最后引入 EMA 注意力,它可以保留每个通道的信息并减少计算成本,使得模型更加关注感兴趣目标区域,提高遮挡场景下的行人翻越护栏检测。在自制数据集 Cross 上进行检测分析,与原模型及其他主流模型相比,SAGE-YOLO 算法具有更好的检测性能。同时在有遮挡和密集车辆、行人的复杂道路场景下,漏检和误检情况更少,检测更为精准。但目前所建数据集仅在白天的场景下所获取,下一步将继续扩充数据集,在黑夜、雨、雪、雾等更为复杂的场景下,扩大实验范围与创新改进,提升行人翻越护栏检测的适应性和实用性。

参考文献

- [1] MATSUMURA S, HIGAKI H. Extension of rh2swl for collision-free data message transmissions by subsidiary channel in wide area wireless multihop networks[C]. 2010 IEEE Wireless Communication and Networking Conference. IEEE, 2010:1-6.
- [2] 张泰,张为,刘艳艳.周界视频监控中人员翻越行为检测算法[J].西安交通大学学报,2016,50(6):47-53.
ZHANG T, ZHANG W, LIU Y Y. Detection algorithm for person climbing over in perimeter video surveillance[J]. Journal of Xi'an Jiaotong University, 2016, 50(6): 47-53.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [5] HE K, GKIOXARI G, DOLLAR P, et al. MaskR-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2):386-397.
- [6] DAI J F, LI Y, HE K M, SUN J. R-FCN: Object detection via region-based fully convolutional networks [C]. Proceedings Systems, Barcelona Spain, 5-12 Dec, 2016. New York: Curran Associates Inc, 2016:379-387.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, realtime object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Patternrecognition, 2017:7263-7271.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv preprint arXiv: 1804. 02767, 2018.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv preprint arXiv: 2004. 10934, 2020.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016:21-37.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [13] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [14] 王林,赵甜.基于深度学习的翻越行为检测[J].计算机系统应用,2023,32(5):262-272.
WANG L, ZHAO T. Detection of climbing behavior based on deep learning [J]. Computer Systems and Applications, 2023, 32(5): 262-272.
- [15] FEICHTENHOFER C, FAN H Q, MALIK J, et al. Slow fast networks for video recognition[C]. 2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul: IEEE, 2019: 6201-6210.
- [16] 王源鹏,万海斌,黄凯,等.基于YOLOv5s的自动扶梯乘客异常行为实时检测算法[J].激光与光电子学进展,2024,61(8):211-218.
WANG Y P, WAN H B, HUANG K, et al. Real-time detection algorithm for abnormal behavior of escalator passengers based on YOLOv5s[J]. Laser & Optoelectronics Progress, 2024, 61(8): 211-218.
- [17] 罗玉涛,高强.基于通道注意力和特征增强的交通标志检测[J].华南理工大学学报(自然科学版),2023, 51(12):64-72.
LUO Y T, GAO Q. Traffic sign detection based on channel attention and feature enhancement[J]. Journal of South China University of Technology (Natural Science Edition), 2023, 51(12): 64-72.
- [18] 井方科,任红格,李松.基于多尺度特征融合的小目标交通标志检测[J].激光与光电子学进展,2024,21(9): 1-13.
JING F K, REN H G, LI S. Small traffic sign detection based on multiscale feature fusion[J]. Laser & Optoelectronics Progress, 2024, 21(9): 1-13.
- [19] 薛珊,王亚博,吕琼莹,等.基于YOLOX-drone的反无人机系统抗遮挡目标检测算法[J].工程科学学报, 2023, 45(9):1539-1549.
XUE SH, WANG Y B, LYU Q Y, et al. Anti-drone system occlusion-resistant target detection algorithm based on YOLOX-Drone[J]. Journal of Engineering Science, 2023, 45(9): 1539-1549.
- [20] 李富,徐凯,朱灵龙,等.基于上下文语义联合YOLOv7的分心驾驶检测算法[J].国外电子测量技术,2024,43(3):121-128.

- LI F, XU K, ZHU L L, et al. Distracted driving detection algorithm based on contextual semantics combined with YOLOv7 [J]. *Foreign Electronic Measurement Technology*, 2024, 43(3): 121-128.
- [21] 高昂, 梁兴柱, 夏晨星, 等. 一种改进 YOLOv8 的密集行人检测算法[J]. *图学学报*, 2023, 44(5): 890-898.
- GAO A, LIANG X ZH, XIA CH X, et al. An improved dense pedestrian detection algorithm based on YOLOv8[J]. *Journal of Graphics*, 2023, 44(5): 890-898.
- [22] 王泽宇, 徐慧英, 朱信忠, 等. 基于 YOLOv8 改进的密集行人检测算法: MER-YOLO[J]. *计算机工程与科学*, 2024, 22(5): 1-17.
- WANG Z Y, XU H Y, ZHU X ZH, et al. MER-YOLO: A dense pedestrian detection algorithm improved based on YOLOv8 [J]. *Computer Engineering and Science*, 2024, 22(5): 1-17.
- [23] WANG X, XIAO T, JIANG Y, et al. Repulsion loss: detecting pedestrians in a crowd[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7774-7783.
- [24] 李忠科, 刘小芳. 基于轻量级 YOLOv8n 网络的 PCB 缺陷检测算法[J]. *电子测量技术*, 2024, 47(4): 120-126.
- LI ZH K, LIU X F. PCB defect detection algorithm based on lightweight YOLOv8n network [J]. *Electronic Measurement Technology*, 2024, 47(4): 120-126.
- [25] 潘海鸿, 陈希良, 钱广坤, 等. 基于改进 YOLOv8n 的轻量化茶叶嫩芽检测方法[J]. *电子测量技术*, 2024, 47(7): 149-156.
- PAN H OU, CHEN X L, QIAN G K, et al. Lightweight tea shoot detection method based on improved YOLOv8n [J]. *Electronic Measurement Technology*, 2024, 47(7): 149-156.
- [26] YAO T, LI Y, PAN Y, et al. Dual vision transformer [C]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [27] LI J, WEN Y, HE L. Sconv: Spatial and channel reconstruction convolution for feature redundancy[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 6153-6162.
- [28] DALING O Y, SU H, GUO Z Z, et al. Efficient multi-scale attention module with cross-spatial learning [C]. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2023: 1-5.
- [29] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 6569-6578.
- [30] HUANG X, WANG X, LYU W, et al. PP-YOLOv2: A practical object detector [J]. *ArXiv preprint arXiv:2104.10419*, 2021.
- [31] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding yolo series in 2021 [J]. *ArXiv preprint arXiv: 2107.08430*, 2021.
- [32] ZHAO Y, LYU W, XU S, et al. Detsr beat yolos on real-time object detection[J]. *ArXiv preprint arXiv: 2304.08069*, 2023.
- [33] 白亮, 丁学文, 申明坤, 等. 基于改进 Yolov8 的摔倒行为检测算法[J]. *天津职业技术师范大学学报*, 2024, 34(1): 38-43.
- BAI L, DING X W, SHEN M K, et al. A fall behavior detection algorithm based on improved Yolov8 [J]. *Journal of Tianjin Vocational and Technical Normal University*, 2024, 34(1): 38-43.

作者简介

刘罡, 高级工程师, 硕士生导师, 主要研究方向为深度学习、机器视觉、移动通信、物联网等。

E-mail: liugang@cw Xu. edu. cn

侯恩翔, 硕士研究生, 主要研究方向为人工智能、机器视觉、目标检测。

E-mail: 2575072816@qq. com

黄孙港(通信作者), 高级工程师, 硕士生导师, 主要研究方向为计算机视觉、光电检测仪器与系统。

E-mail: hsg@cw Xu. edu. cn