

SLAW-YOLOv8:一种密集行人检测方法

陈汝熠, 闫河, 王艳, 沈其乐

(重庆理工大学两江人工智能学院, 重庆 401135)

E-mail:52232313181@stu.cqut.edu.cn

摘要:针对YOLOv8密集场景行人检测易出现漏检、误检现象,通过引入具有强特征提取能力的空间通道重建卷积模块(SCConv)和可分离大核注意力机制(LSKA)对其主干网络进行改进;分别在其颈部网络和检测头引入通道注意力模块(EMA)和自适应空间特征融合模块(ASFF),有效增强多通道上下文信息和目标尺度不变性能;提出一种新的Wise-MPDIoU边界损失函数,从而构建出一种新的密集行人检测方法:SLAW-YOLOv8.对比实验结果表明,在PASCAL VOC2012数据集上,mAP@0.5提高了4.4%,mAP@0.5:0.95提高了4.7%;在WiderPerson数据集上mAP@0.5提升了1.5%,mAP@0.5:0.95提高了1.9%,有效消除了密集场景行人检测的漏检、误检现象.

关键词:行人检测;YOLOv8;自适应空间特征融合;可分离大核注意力机制;边界损失函数

中图分类号:TP319

文献标识码:A

文章编号:1000-1220(2025)12-2941-08

SLAW-YOLOv8:a Method for Dense Pedestrian Detection

CHEN Ruyi, YAN He, WANG Yan, SHEN Qile

(School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401135, China)

Abstract:To address the problem of missed detections and false alarms in dense scene pedestrian detection with YOLOv8, we introduce a spatial channel reconstruction convolution module (SCConv) with strong feature extraction ability and a separable large kernel attention mechanism (LSKA) to improve the main body network. We also introduce channel attention modules (EMA) and adaptive spatial feature fusion modules (ASFF) in the neck network and detection head, respectively, to enhance multi-channel contextual information and target scale-invariant performance. We propose a new Wise-MPDIoU boundary loss function, thereby constructing a new dense pedestrian detection method: SLAW-YOLOv8. The experimental results show that on the PASCAL VOC2012 dataset, mAP@0.5 was improved by 4.4%, and mAP@0.5:0.95 was improved by 4.7%; on the WiderPerson dataset, mAP@0.5 was improved by 1.5%, and mAP@0.5:0.95 was improved by 1.9%, effectively eliminating the problem of missed detections and false alarms in dense scene pedestrian detection.

Keywords: pedestrian detection; YOLOv8; multi-scale feature fusion; attention mechanisms; loss function

0 引言

密集行人检测受相互遮挡及光照变化的影响,漏检、误检现象突出,制约了行人流量智能统计在智慧城市管理中的广泛应用.

SSD^[1]是最早出现的适应密集目标的检测器,但其检测准确率较低,随后陆续出现的YOLO^[2]系列检测器,均在提升检测准确率和减低模型参数量方面不断改进.其中,YOLOv3^[3]采用更深的卷积层和多尺度检测策略,在提升对小目标检测精度的同时,具有较高的推理速度,但其每秒浮点计算量大且检测精度有待提高;相比YOLOv3, YOLOv4^[4]引入CIoU损失函数、空间注意力(Spatial Attention Module, SAM)模块和PANet模块的改进方法,检测精度略有提高的同时,浮点数计算量明显降低;YOLOv5^[5]引入模型剪枝、数据增强、自动缩放多尺度检测等改进机制,进一步提高检测精

度和降低了浮点数计算量;为进一步提升检测精度,2022年在YOLOv5模型的基础上,YOLOv7^[6]增加了添加了空间金字塔池化(Spatial Pyramid Pooling, SPP)模块和SAM模块,但其浮点数计算量明显增加、帧率明显下降;2023年,YOLOv8^[7]引入梯度信息更丰富的C2f结构,采用解耦头检测结构,并优化模型通道分配策略,显著提升了推理速度和计算效率.

YOLO系列的不断优化提升了对密集行人检测的适应能力,但当前方法仍存在显著不足,模型对复杂场景中特征细节的捕捉能力不足,导致漏检和误检率偏高.为促使YOLO系列在密集行人检测任务中取得更出色的表现,国内外的学者已提出各种的改进方法.王^[8]等人在YOLOv4的基础上将像素块注意力机制模块(Convolutional-Pixel Block Attention Module, CBAM-PIX)嵌入主干网络CSPDarknet53,利用级联思想对特征融合网络进行改进,强化模型在空间和通道维度

收稿日期:2025-01-03 收修改稿日期:2025-02-27 基金项目:国家自然科学基金面上项目(61173184)资助;重庆市自然科学基金项目(cstc2018jcyjAX0694)资助. 作者简介:陈汝熠(通信作者),女,2000年生,硕士研究生,研究方向为深度学习、目标检测;闫河,男,1972年生,博士,教授,研究方向为计算机视觉、人工智能等;王艳,女,2000年生,硕士研究生,研究方向为目标检测;沈其乐,男,2003年生,研究方向为深度学习.

上提取像素信息的能力。姜^[9]等人在 YOLOv5 的基础上引入 SimAM 无参注意力机制,同时增设信息融合模块,以挖掘更丰富的特征信息;徐^[10]等人在 YOLOv7 的基础上通过引入转置卷积以及优化后的 Rep-ELAN-W 模块,显著提升了模型对中低维特征图中小目标特征信息的利用效率;王^[11]等提出了轻量级的密集行人检测算法 MER-YOLO,在 YOLOv8 架构中替换了主干网络,引入注意力机制并且对损失函数进行优化;而高^[12]等则基于 YOLOv8,采用可变形卷积对主干网络进行优化,同时设计了遮挡感知注意力机制和动态解耦头,使多尺度特征的检测效果得以提升。这些改进方法在针对遮挡与小目标检测方面虽有一定成效,但面对复杂密集场景,模型对目标细节特征的捕捉能力依然不足,漏检问题仍然明显。因此,如何在 YOLOv8 的基础上进一步增强对行人细节特征的提取能力,同时兼顾检测精度与计算效率,依然是密集行人检测领域需解决的关键问题。为解决上述问题,本文提出一种基于 YOLOv8 改进的密集行人检测算法 SLAW-YOLOv8,旨在提升复杂场景中的检测性能,同时保持较高计算效率。本文的改进点包括:

1) 在 C2f 模块中加入一种融合空间通道重建卷积 SCCo-
nv^[13]组成新的 C2f_SCCo-
nv 模块,通过提高目标细节特征提取能力显著提升了模型的检测性能;

2) 引入 LSKA 大核注意力机制^[14],增强目标细节多尺度特征融合能力;

3) 引入自适应空间特征融合方法 ASFF^[15]对特征融合网络进行改进,对特征图进行动态加权融合,全面捕捉多尺度信息,实现更准确的定位;

4) 引入 EMA^[16]注意力机制,提升模型对行人背景的识别能力,缓解遮挡对检测的影响;

5) 训练时采用高效的 Wise-MPDIoU 作为边界框损失函数,增加对低质量锚框的聚焦能力,加快模型收敛速度,进一步提高检测精度。

1 SLAW-YOLOv8 算法网络模型构建

1.1 YOLOv8 模型

YOLOv8 是基于 YOLOv5 由同一团队所提出的单阶段目标检测框架。它继承了 YOLO 系列模型在高效性和实时性方面的优势,在目标检测、图像分类以及语义分割等任务中得到广泛应用。与传统锚框机制不同,YOLOv8 通过自动学习边界框提升了模型的灵活性。同时,YOLOv8 进一步优化了参数量和计算量,并提供了 5 个版本分别为 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x。通过对模型速度与大小的综合考量,本文选用 YOLOv8n 来开展行人检测任务。

YOLOv8 的网络结构包括了 4 个部分,分别是输入模块 (Input)、主干网络 (Backbone)、颈部网络 (Neck) 以及检测头 (Head)。其中,Input 的作用是进行数据输入;Backbone 负责提取不同尺度的图像特征;Neck 主要承担融合多尺度特征的任务;而 Head 则用于生成最终的预测结果。

在输入端,YOLOv8 引入了 Mosaic 数据增强方法。这种方法通过将多张图片按照一定比例进行拼接来完成预处理,进而提升检测性能,增强模型的鲁棒性。YOLOv8 的主干网络

和颈部网络都经过了重新设计,主干网络引入了更为强大的 CSP (Cross Stage Partial Network, CSP) 结构,在减少计算量的同时,还提高了特征融合效果。用梯度流更为丰富的 C2f 结构对 YOLOv5 的 C3 结构进行替换,显著提升了特征提取能力,从而增强了模型对不同尺度目标的检测能力。

在 Neck 部分,YOLOv8 采用了 PAFPN 特征融合结构。该方法通过将深层特征的语义信息自上而下传递到浅层特征,同时把浅层特征的位置信息传递到深层特征,并且去除了 1×1 卷积的降采样层。在 Head 部分,YOLOv8 把 YOLOv5 的 Anchor-Based 机制替换为 Anchor-Free 机制,通过关键点或中心点结合边界信息的方式来表示物体,这种方式更适用于密集场景检测。同时,采用主流的解耦头结构 Decoupled-Head,将分类头与检测头分开,进一步提升了模型的精度与收敛速度。

1.2 Backbone 引入 SCCo- nv 卷积

在密集行人的检测中,行人之间由于互相遮挡及环境变化,目标的特征会出现大量干扰信息。卷积运算在各种视觉任务中表现优异,但容易对特征信息捕捉不准确,影响模型对目标特征的有效提取。C2f 模块是 YOLOv8 主干网络的基本构成单元,要负责特征提取与下采样,为了进一步提升对目标特征的捕捉能力,并减少因特征提取不足导致的检测错误,本文将 SCCo-
nv 卷积融入 C2f 模块构成一个新的卷积模块 C2f_SCCo-
nv,空间重建单元 (Spatial Reconstruction Unit, SRU) 和通道重建单元 (Channel Reconstruction Unit, CRU) 组合构建出 SCCo-
nv 模块,如图 1 所示。其中,SRU 单元会依据权重分配来分离和重构输入特征,达到既能抑制空间层面的冗余信息,还能增强特征的表征能力的目的。再通过 CRU 运算执行分裂、变换以及融合等操作,进一步降低通道维度的特征冗余,从整体上提升模型对目标的特征提取成效。

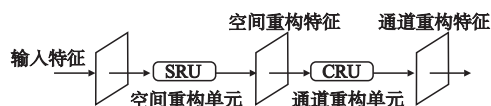


图 1 SCCo-
nv 模块

Fig. 1 SCCo-
nv module

如图 2 所示,输入特征首先经过一个 Conv 卷积块进行特征提取,然后沿通道维度分离为两条分支。其中一条分支通过 SCCo-
nv 模块提取特征后,与另一条分支的特征进行跳层连接,从而实现跨尺度的卷积融合。最终,这些特征通过一个 Conv 卷积块生成 C2f_SCCo-
nv 残差单元的输出特征。进一步用 C2f_SCCo-
nv 模块替换整个网络模型中的所有 C2f 模块。

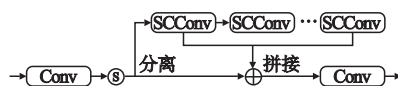


图 2 C2f_SCCo-
nv 结构图

Fig. 2 C2f_SCCo-
nv structure diagram

1.3 SPPF 引入可分离大卷积核注意力 LSKA

YOLOv8 中的 SPPF 结构是将 SPP (Spatial Pyramid Pooling, SPP) 和感知大内核卷积 (UniRepLK) 相结合,通过对输入特征图进行不同尺度的池化操作,实现各特征图之间的相互融合,进而提升模型的特征提取能力。由于密集行人检测背

景复杂,为了提升模型对复杂背景的应对能力,在 SPPF 层中引入 LSKA 可分离大卷积核注意力,更深层次捕捉目标的局部上下文信息,构成 SPPF-LSKA 模块以替代 YOLOv8 中 SPPF,如图 3 所示。

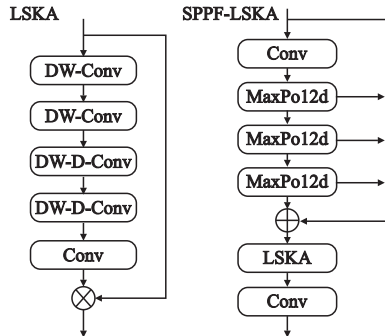


图 3 LSKA 及 SPPF_LSKA 模块结构图

Fig. 3 Structure diagram of LSKA and SPPF_LSKA modules

LSKA 模块由标准深度卷积(DW-Conv)、扩张深度卷积(DW-D-Conv)以及 1×1 通道卷积三部分组成,其卷积操作分为以下 3 步:1)把输入特征图和 $K \times K$ 卷积的核进行卷积拆分为两部分的二维卷积;2)将二维的深度与深度扩张卷积核继续拆分成一维的水平卷积核与垂直卷积核;3)将上述两步的 3 个卷积核级联起来。

1.4 Head 引入自适应融合 ASFF

YOLOv8 的颈部采用改进的 PAN-FPN(路径聚合网络-特征金字塔网络)结构,优化了上采样阶段的特征融合效率,但多尺度特征融合可能导致信息冲突和不一致性。复杂场景下的密集行人检测,待检测的行人会因为遮挡和光线等因素存在模糊轮廓,尺度差异和部分特征缺失的情况。为解决这一问题,在检测头引入自适应空间特征融合(Adaptively Spatial Feature Fusion, ASFF)策略,提升特征的尺度不变性和融合效果。ASFF 结构如图 4 所示。在此结构里,Level 1、Level 2 和 Level 3 代表特征金字塔中不同层级的特征,并且不同层级有不同的分辨率,经过特征融合得到 ASFF-1、ASFF-2 和 ASFF-3 这 3 个 ASFF 层。

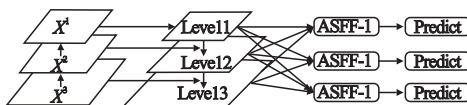


图 4 ASFF 模块结构图

Fig. 4 ASFF module structure diagram

ASFF 策略通过特征尺度归一化与自适应权重融合两个步骤实现多尺度特征整合,针对不同层级特征分别进行上采样和下采样操作。其中上采样使用 1×1 的卷积对特征通道进行压缩,使其与目标层级通道维度一致,再利用插值算法提升特征图空间分辨率,而比的下采样则使用步幅为 2 的 3×3 的卷积实现特征图尺寸减半与通道维度调整。该机制通过参数化方式统一多尺度特征的维度空间。比的下采样在比的下采样的基础上增加步幅为 2 的最大池化层,同样也使用 3×3 的卷积。

最终的特征融合结果经过特征缩放后的每个特征图与其对应的权重参数矩阵相乘,对应元素值相加的方式得到,计算

过程为:

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (1)$$

其中, $x_{ij}^{n \rightarrow l}$ 表示从 n 层映射到 l 层在 (i, j) 位置上的特征向量, $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1]$ 为可学习参数,分别为在 (i, j) 位置上 α, β, γ 映射在 l 层上的特征权重值,满足下列关系:

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha}^l a_{ij}}}{e^{\lambda_{\alpha}^l a_{ij}} + e^{\lambda_{\beta}^l b_{ij}} + e^{\lambda_{\gamma}^l c_{ij}}} \quad (2)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (3)$$

上式中 $x^{1 \rightarrow l}, x^{2 \rightarrow l}, x^{3 \rightarrow l}$ 进行特征缩放操作后的 3 个特征图层使用 1×1 的卷积计算得到 $\lambda_{\alpha}^l, \lambda_{\beta}^l, \lambda_{\gamma}^l$ 。经过上述操作,利用 ASFF 结构增强各个层级特征图上利于行人检测的语义信息,让不同尺度的特征达到更好的融合效果。

1.5 Neck 引入注意力机制模块 EMA

注意力机制可以对局部关键信息有效抓取,使模型能够把重点放在检测目标的核心特征上,在计算机视觉领域已被广泛应用。在颈部网络中引入了一种高效多尺度注意力(EMA)模块,此模块运用分组结构,无需进行降维就能完成特征处理。EMA 注意力机制采用通道维度重构策略,将部分特征通道映射至批处理维度,并通过多分支子特征划分,实现空间语义信息在跨组特征间的均衡化分布。借助特征分组和多尺度设计,EMA 模块成功建立起短期与长期的依赖关系。该模块极大地提升了检测器的性能,还保留各通道的信息的同时有效降低了计算成本,EMA 注意力模块如图 5 所示。

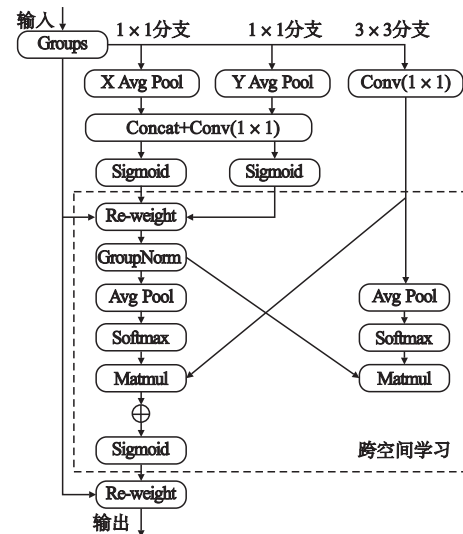


图 5 EMA 注意力模块

Fig. 5 EMA attention module

EMA 注意力采用并行结构,将输入特征图按通道维度分为 G 个子特征,包含两条 1×1 卷积和一条 3×3 卷积分支。 1×1 分支通过全局平均池化和 Sigmoid 函数提取二维特征分布, 3×3 分支捕获多尺度特征并生成新特征图。两分支特征通过平均池化编码后,利用 Softmax 归一化生成新特征,最后通过矩阵乘法与原特征融合,得到包含多尺度空间信息和精确位置信息的注意力图,实现跨分支信息交互并增强目标特征权重,输出与输入同维度的特征图。把 EMA 注意力模块放入 YOLOv8 网络的 neck 结构中,有利于增强网络通道维度上

的全局上下文信息,从而提升模型对行人背景特征的识别能力,对严重遮挡的行人检测有更好的性能提升。

1.6 损失函数的改进

YOLOv8 网络依赖于边界框回归 (boundary bounding, BBR) 模块确定目标位置。损失函数 IoU 是 BBR 的重要组成部分,是计算一个预测值与其真实值之间的损失的指标,以此衡量模型的准确性。通过合理设计损失函数,算法能够在训练过程中逐步优化模型性能从而提高检测精度,其中 YOLOv8 网络中采用了 CIoU^[17],计算公式如下:

$$IoU = \frac{|B^{pred} \cap B^{gt}|}{|B^{pred} \cup B^{gt}|} \quad (4)$$

$$CIoU = IoU - \left(\frac{\rho^2(B^{pred}, B^{gt})}{c^2} + \alpha v \right) \quad (5)$$

其中, $\alpha = \frac{v}{(1-IoU)+v}$, 作为平衡参数, $v = \frac{4 \times (\arctan \frac{w_g}{h_g} - \arctan \frac{w_p}{h_g})^2}{\pi^2}$, 作为修正因子, (w_g, h_g) 和 (w_p, h_p) 分别为目标框和预测框的宽高, $\rho(B^{pred}, B^{gt})$ 为预测框和真实框中心点的距离, c 为目标对象最小外接矩形的对角线长度。

然而,CIoU 中整合的长宽比约束项存在潜在局限性,它通过参数化建模预测框与真实框的宽高比差异,会让极端 IoU 区间样本的回归优化过程表现出显著偏差。MPDIoU^[18] 是一种改进算法,通过直接针对预测框与真实框的左上角、右

下角对应点间距离进行最小化处理,在边界框重叠还是分离的复杂场景下,都能展现出更好的适应性,有效加快模型的收敛速度,提升训练效率。

$$L_{MPDIoU} = 1 - MPDIoU \quad (6)$$

$$MPDIoU = IoU - \frac{\rho^2(P_1^{pred}, P_1^{gt})}{w^2 + h^2} - \frac{\rho^2(P_2^{pred}, P_2^{gt})}{w^2 + h^2} \quad (7)$$

其中, $\rho^2(P_1^{pred}, P_1^{gt})$ 是计算预测框从左下角到右下角这两点之间的距离, $\rho^2(P_2^{pred}, P_2^{gt})$ 是计算真实框从左下角到右下角这两点之间的距离。

本文引入动态非单调的聚焦机制 Wise-IoU^[19], 基于 WIoU 的思想设计 Wise-MPDIoU, 计算公式如下:

$$L_{Wise-MPDIoU} = R_{WIoU} L_{MPDIoU} \quad (8)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (9)$$

其中, x 和 y 为锚框的中心点坐标, x_{gt} 和 y_{gt} 表示目标框的中心点坐标, W_g 和 H_g 表示最小外接矩形的宽和高, $*$ 表示将 W_g 、 H_g 从计算图中分离,作用是为了防止 R_{WIoU} 产生阻碍收敛的梯度,该改进损失函数 Wise-MPDIoU 在关注高质量锚框的基础上增加了对低置信度目标的关注度,提升了模型处理困难样本的能力。

1.7 SLAW-YOLOv8

密集人群图像往往背景复杂、图像分辨率高且小目标人群多,原始 YOLOv8 算法提取到的目标行人信息少,模型易出现漏检的情况,上文对 YOLOv8 算法进行 5 个部分改进:

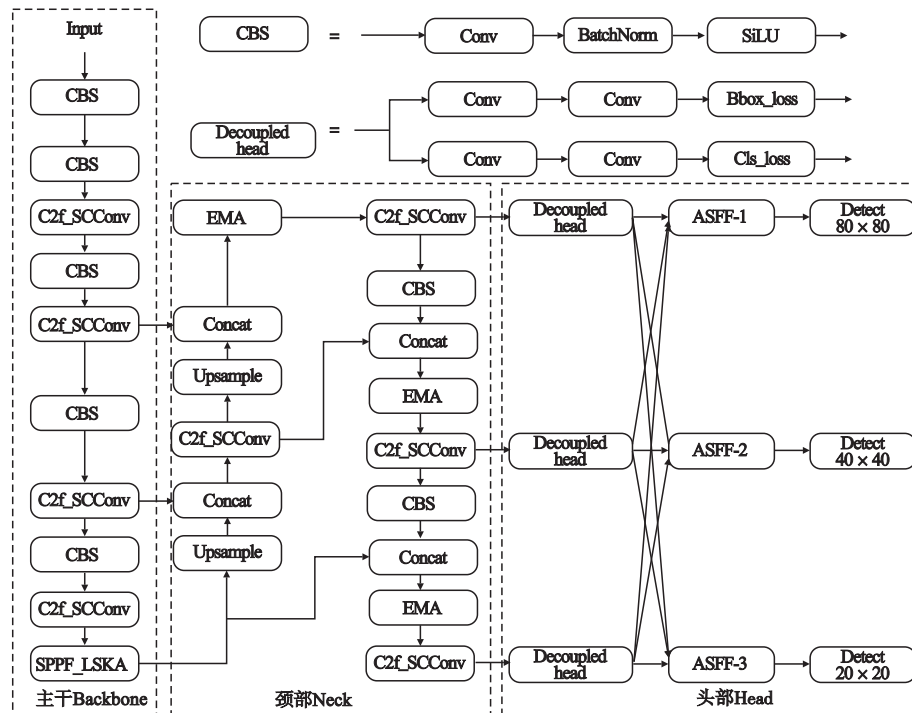


图6 SLAW-YOLOV8 网络结构图

Fig. 6 SLAW-YOLOV8 network structure diagram

设计 C2f_SCCConv 结构替换 C2f 结构、设计 SPPF_LSKA 结构替换 SPPF 结构、利用自适应空间特征融合方法改进检测头、

增加注意力机制和改进损失函数。

最终改进后的 SLAW-YOLOv8 如图 6 所示。

2 实验设计与结果分析

2.1 数据集

在消融实验中,本文选用 PASCAL VOC2012 数据集作为数据基础.该数据集含有 11K 张训练图像,共计 20 个类别,是计算机视觉领域广泛采用且被公认的标准数据集格式.为了全方位评估改进算法的有效性,本文还运用公开的 WiderPerson 密集场景行人数据集开展对比实验.这个数据集包含 13382 张图像,大约 40 万个存在不同遮挡程度的行人实例,其中训练集有 8000 张,验证集有 1000 张,测试集有 4382 张,涵盖行人、骑车人、人群等 5 个类别.而本文实验仅选取行人类别的数据用于测试.

在数据预处理环节,首先把图像随机裁剪成 640×640 像素的子图像,接着运用 Mosaic 方法进行增强,即随机挑选 9 张图片做随机排列与拼接,增加数据的多样性,提供更丰富的上下文信息. Mosaic 数据增强还能间接增大 Batchsize,进而提升模型的鲁棒性和泛化能力.

2.2 实验环境和参数

本文实验基于 Windows 11 操作系统,实验采用 NVIDIA GeForce RTX3090 24G 作为 GPU,实验环境为 Python3.8、CUDA11.7 及 Pytorch1.12.实验设置基本采用 YOLOv8 官方推荐的参数,设置如下:训练轮数为 300 epochs 以确保收敛;batchsize 设置为 48,输入图片统一尺寸为 640×640 ,为了避免消融实验误差和节约模型训练时间,将初始学习率设置为 0.01,将随机种子和数据增强均设置为 1,并且选用随机梯度下降策略(SGD)对网络参数加以优化.

2.3 评价指标

本文选取精确度 (precision, $P_{recision}$),召回率 (recall, R_{ecall}),和 mAP@0.5 (mean average precision, mAP), mAP@

0.5: 0.95、模型计算量 (Giga Floating-point Operations Persecond, GFLOPs) 和每秒检测图片数量 (frames persecond, FPS) 等评价指标来对本文改进的算法进行综合评价,其公式如公式(10)~公式(13)所示:

$$P_{recision} = \frac{TP}{TP + FP} \quad (10)$$

$$R_{ecall} = \frac{TP}{TP + FN} \quad (11)$$

$$AP = \int_0^1 P(R) dR \quad (12)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 P(R) dR \quad (13)$$

式中,TP 为预测正确的样本、FP 为预测错误的样本、FN 为漏检的样本;平均精度 (Average Precision, AP) 用以衡量模型在不同置信度阈值下的性能,其中 mAP@0.5,代表的是 IOU 为 0.5 时,各类别 mAP 的平均值.

2.4 消融实验

为了评估本文提出的改进方法的实际效果,分别对 YOLOv8n 的主干网络、Neck 端、检测头和损失函数进行了改进.原模型与改进后的模型在相同的实验平台上进行测试,并优化调整了相关参数.为了验证各个改进模块对检测效果的具体影响,开展了一系列对比实验.所有实验均基于 PASCAL VOC2012 数据集展开,以此全面分析不同改进方法在该数据集环境下呈现出的效果差异.以 YOLOv8n 作为基线模型进行一系列实验,分别对可变形卷积、特征金字塔、动态解耦头、多尺度注意力机制融合以及损失函数更换等方面进行了逐步改进,呈现了各模块对模型性能产生的影响,其中“√”表示在 YOLOv8n 网络模型上应用了相应的改进方法.实验结果如表 1 所示.

表 1 SLAW-YOLOV8 在 PASCAL VOC2012 上的消融实验
Table 1 Ablation experiment of SLAW-YOLOV8 on PASCAL VOC2012

模型	SCConv	LSKA	ASFF	EMA	Wise-MPDIoU	P	R	AP _{0.5}	AP _{0.5-0.95}	GFLOPs	FPS
①						67.1%	56.3%	61.5%	44.8%	8.2	55.3
②	√					68.5%	56.9%	62.6%	46.5%	9.0	42.5
③		√				69.7%	56.3%	63.1%	45.9%	8.3	104
④			√			73.3%	57.5%	63.5%	46.7%	10.3	41.5
⑤					√	70.5%	55.5%	62.1%	45.3%	8.2	108
⑥	√	√				71.3%	56.7%	63.4%	47.2%	9.6	41.2
⑦	√	√	√			70.4%	57.3%	64.4%	47.8%	11.7	43.7
⑧	√	√	√	√		72.9%	58.3%	64.9%	48.5%	11.8	41.9
SLAW-YOLOV8	√	√	√	√	√	73.9%	59.1%	65.9%	49.5%	11.8	41.2

由表 1 中的消融实验结果可知,将主干网络中的 C2f 模块改进为 C2f_SCConv 后,模型在 P、R 上分别提升了 0.6%,在 mAP@0.5、mAP@0.5:0.95 上分别提升了 1.1% 和 1.7%.这表明骨干网络的改进对模型的特征提取能力有显著提升,能够更有效地应对密集行人检测任务中的遮挡和干扰特征.由于改进后的 SCConv 模块引入了空间重建和通道重建操作,导致模型的深度增加,使计算量增至 9.0 GFLOPs,模型的推理速度有所下降.但 mAP 的显著提升证明了 SCConv 模块在密集行人检测任务中的实用性,能够在一定的计算代

价增加下实现性能的有效优化.

加入了 LSKA 模块后的改进模型相较于原始的 YOLOv8n,在 P、mAP@0.5、mAP@0.5:0.95 上分别提高了 2.6%、1.6% 和 1.1%,GFLOPs 保持不变,FPS 提高至 104.说明 LSKA 可分离大卷积核注意力模块在提升检测精度同时加快了检测速度,可以认为对特征金字塔结构的改进是有效的.

在颈部网络中加入了 ASFF 模块,将 3 个尺度大小的特征进行融合,增加对应权重,并进行学,可以看出改进过后的

算法在 P、R、mAP@0.5、mAP@0.5:0.95 都有不错的提升,分别提高了 6.2%、1.2%、2% 和 1.9%,说明 ASFF 自适应空间融合方式对于算法精度提升具有明显作用,但是 FPS 有所降低。

对于使用改进后的 Wise-MPDIoU 损失函数作为边界框回归函数,相较于基准算法,在不增加模型复杂度的情况下,各项指标都有所提高,提升了网络的检测精度,也大幅度提高了模型检测的速度,最后的结果可以看出对损失函数的优化对整体改进后的网络也有更好的提升,表明 Wise-MPDIoU 损失函数能够更好提升模型处理困难样本的能力。

此外,通过对比实验②和⑥可以发现,融合改进后的 C2f 模块和特征金字塔的模型在 P、mAP@0.5 和 mAP@0.5:0.95 上分别提高了 4.2%、1.9% 和 2.4%。对比实验⑤和⑥时,加入 EMA 注意力机制后,尽管模型略微牺牲了检测速度且增加了参数量,但召回率和 mAP 得到了提升。与实验①对比,实验⑦在牺牲一定检测速度的情况下,4 项评价指标均实现了全面提升。通过对比实验①和⑧,当采用所有 4 个改进模块时,相较于基础模型,尽管增加了参数量,所提改进模型在 P、R、mAP@0.5 和 mAP@0.5:0.95 上分别提升了 6.8%、2.4%、4.4% 和 4.7%。综合分析消融实验结果,证明本文所提出的各个改进模块有效,能够满足实时性和准确性的需求。

2.5 对比实验

为了验证本文提出的 SLAW-YOLOv8 密集行人检测算法在性能上的优越性,设计了一系列实验,将 SLAW-YOLOv8 与基准算法 YOLOv8n 以及当前目标检测领域代表性的算法进行对比分析。这些算法包括 SSD、YOLOv3、YOLOv4、YOLOv5s、YOLOv7 和 YOLOv8s,所有算法均在 WiderPerson 数据集上进行了测试,实验结果如表 2 所示。

表 2 在 WiderPerson 数据集中不同目标检测算法的对比实验
Table 2 Comparative experiments of different object detection algorithms in the WiderPerson dataset

模 型	AP _{0.5}	AP _{0.5:0.95}	GFLOPs	FPS
SSD	68.4%	-	87.7	59
YOLOv3	82.0%	47.6%	117.7	54.6
YOLOv4	84.9%	51.9%	17.7	65
YOLOv5s	87.2%	64.6%	13.7	64.6
YOLOv7	89.2%	64.3%	71.3	31.7
YOLOv8s	88.6%	67.8%	28.4	64.5
YOLOv8n (baseline)	87.8%	63%	8.1	55.3
SLAW-YOLOv8n	89.3%	64.9%	11.8	41.2

在相同条件下,SLAW-YOLOv8 相较于其他目标检测算法,如 SSD 和 YOLOv7,表现出显著的优势,拥有最高的 mAP@0.5。与 SSD、YOLOv3、YOLOv4、YOLOv5s 和 YOLOv8s 算法相比,本文算法的 mAP@0.5 分别提高了 20.9、7.3、4.4、2.9 和 0.7 个百分点。在计算量方面,SLAW-YOLOv8 相较于 SSD、YOLOv3、YOLOv4、YOLOv5s 和 YOLOv7 分别减少了 75.9、105.9、5.9、1.9 和 59.5 GFLOPs。虽然与 YOLOv8n 相比增加了 3.7 GFLOPs 的计算量,但其精度显著提高,进一步证明了该算法的优越性。

2.6 检测结果可视化

为了直观展现算法改进的效果,本文对 YOLOv8n 和 SLAW-YOLOv8 在训练 PASCAL VOC2012 数据集过程中的 P、R、mAP@0.5 和 mAP@0.5:0.95 等指标进行可视化对比,结果如图 7 所示。

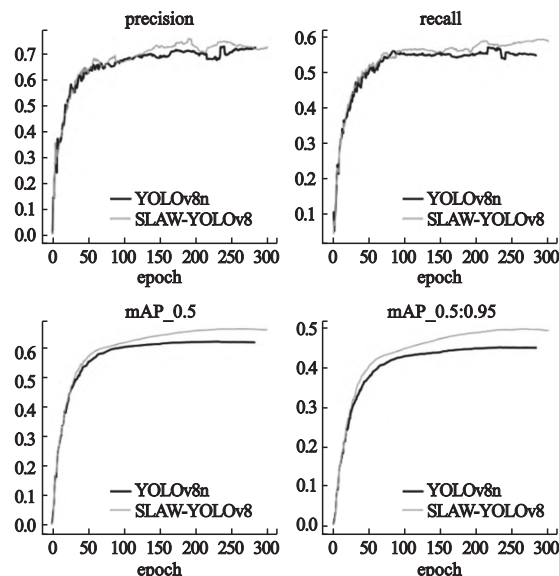


图 7 YOLOv8n 与 SLAW-YOLOv8 指标对比

Fig. 7 Comparison of YOLOv8n and SLAW-YOLOv8 indicators

从图中能够看出,随着迭代次数不断上升,这两种算法最后皆可达到收敛的状态。然而,改进后的 SLAW-YOLOv8 在 P、R、mAP@0.5 和 mAP@0.5:0.95 等指标上全面优于基准 YOLOv8n,这表明改进后的算法在检测精度上取得了显著提升。

为使本文改进算法的检测效果呈现得更为直观,分别运用 YOLOv8n 以及本文所提方法,针对多种密集人群场景展开检测工作。检测效果的对比情况,如图所示。图中左边呈现的是 YOLOv8 算法的检测效果图,而右边则是基于 YOLOv8 的本文改进算法的检测效果图。

在严重遮挡场景中(如图 8 所示),原始图像包含 102 个站立行人,YOLOv8n 的检测结果漏检现象明显,仅识别出前 3 排的站立行人,而后排因遮挡严重几乎全部漏检。而 SLAW-YOLOv8 显著提升了检测数量,检测框的置信度也较高。在人群密集场景中(如图 9 所示),由于人群拥挤,行人特征不够明显,YOLOv8n 仅检测出少量具有完整行人特征的目标。相比之下,SLAW-YOLOv8 检测到的 Person 类别数量增加了 16 个,同时保持较高的置信度。在背景昏暗场景中(如图 10 所示),YOLOv8n 对被水雾遮挡的行人完全无法检测,且置信度较低,而 SLAW-YOLOv8 检测出 23 个 Person 类别目标,比 YOLOv8n 多 9 个,同时对被水雾遮挡的目标和远处小目标的检测能力大幅提升。SLAW-YOLOv8 在处理严重遮挡、人群密集以及背景昏暗等复杂场景时展现出显著优势。

经分析可知,在小尺度密集遮挡行人检测任务中,YOLOv8n 出现漏检的情况较为频繁。与之相比,改进后的 SLAW-YOLOv8 漏检现象大幅降低,检测效果获得了显著提高。在行人密集的场景里,SLAW-YOLOv8 优势突出。综合来

看,相较于原算法,改进后的 SLAW-YOLOv8 在复杂场景下的检测性能更为出色。

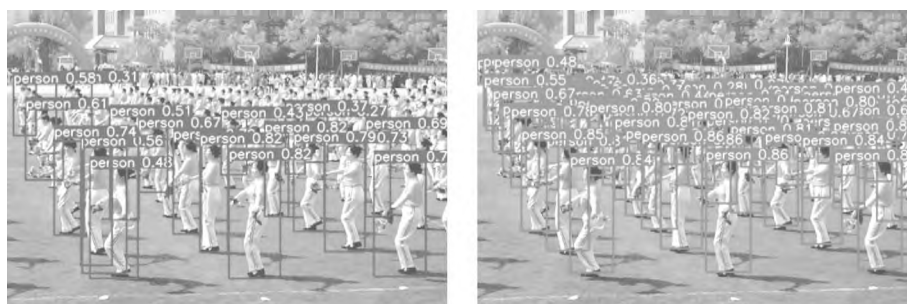


图 8 在严重遮挡时行人检测结果比较

Fig. 8 Comparison of pedestrian detection results in severe occlusion

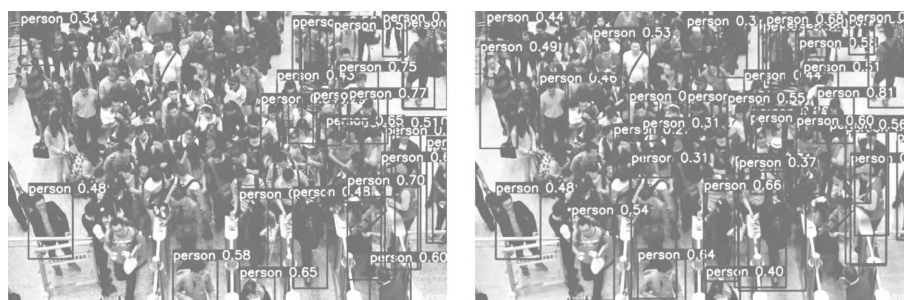


图 9 在人群密集时行人检测结果比较

Fig. 9 Comparison of pedestrian detection results in crowded times



图 10 在背景昏暗时行人检测结果比较

Fig. 10 Comparison of pedestrian detection results in a dim background

3 结 论

本文基于 YOLOv8n 网络进行改进,提出了一种适用于密集行人检测任务的新模型 SLAW-YOLOv8. 改进主要体现在 4 个方面:结合 SCAConv 模块增强特征提取能力,利用 LS-KA 模块实现特征融合,在检测头中引入 ASFF 模块,以及将 CIoU 损失函数替换为 Wise-MPDIoU 损失函数. 通过实验验证,本文所提出的模型能够高效、精准地完成密集行人目标的定位与识别任务,有效平衡了复杂场景下的目标定位准确率与检测实时性需求,本文在 WiderPerson 数据集上对多种目标检测算法进行了全面对比,尤其是针对小尺度密集遮挡行人检测场景,与多种对比模型相比,展现出了卓越的性能表现. 不过,需要注意的是,实际的密集行人场景中,目标分布情况更为复杂. 未来的研究方向将主要聚焦于实际场景的应用,进一步提升模型在小目标检测方面的性能表现与适应性. 就

目前而言,该算法仍然存在漏检的问题,并且检测速度也有待进一步提升. 接下来的工作重点将放在优化网络性能上,提高模型在密集场景下的检测速度,以便更好地适配嵌入式 GPU 平台的部署,推动该算法在更多实际密集行人场景中得到广泛应用.

References:

- [1] Liu W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector[C]//Computer Vision-ECCV: 14th European Conference, 2016: 21-37.
- [2] Redmon J. You only look once: unified, real-time object detection [C]//Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [3] Farhadi A, Redmon J. YOLOv3: an incremental improvement [C]//Computer Vision and Pattern Recognition, 2018: 1-6.
- [4] Gai R, Chen N, Yuan H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model[J]. Neural Computing

- and Applications, 2023, 35(19):13895-13906.
- [5] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:2778-2788.
- [6] 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]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:7464-7475.
- [7] Terven J, Córdova-Esparza D M, Romero-González J A. A comprehensive review of yolo architectures in computer vision: from YOLOv1 to YOLOv8 and yolo-nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4):1680-1716.
- [8] WANG C, WANG T J, ZHOU Z Y. A small target detection method for dense crowds based on YOLOv4 [J]. Applied Science and Technology, 2024, 51(2):82-89.
- [9] JIANG X, SU Z Y, XUE S. Dense pedestrian detection algorithm for intelligent driving [J]. Modern Computer, 2024, 30(10):1-10.
- [10] XU F X, FAN R, MA X L. Improved YOLOv7 algorithm for crowded pedestrian detection [J]. Computer Engineering, 2024, 50(3):250-258.
- [11] WANG Z Y, XU H Y, ZHU X Z, et al. Dense pedestrian detection algorithm based on YOLOv8; MER-YOLO [J]. Computer Engineering and Science, 2024, 46(6):1050-1062.
- [12] GAO A, LIANG X Z, XIA C X, et al. An intensive person detection algorithm for improved YOLOv8 [J]. Journal of Graphics, 2023, 44(5):890-898.
- [13] 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.
- [14] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: rethinking the large kernel attention design in cnn [J]. Expert Systems with Applications, 2024, 236:121352.
- [15] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection [EB/OL]. <https://arxiv.org/abs/1911.09516.pdf>, 2023-06-29.
- [16] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning [C]//ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023:1-5.
- [17] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52(8):8574-8586.
- [18] Wang H, Yang H, Wang J, et al. Multi-scale target detection algorithm of optical remote sensing image based on improved YOLOv8 [J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2024, 48(4):649-654.
- [19] Wang W, Liu W. Small Object detection with YOLOv8 algorithm enhanced by mobileViTv3 and wise-IoU [C]//Proceedings of the 12th International Conference on Computing and Pattern Recognition, 2023:174-180.

附中文参考文献:

- [8] 王 翀, 王同军, 周正一. 一种基于 YOLOv4 的密集人群小目标检测方法 [J]. 应用科技, 2024, 51(2):82-89.
- [9] 姜 鑫, 苏昭宇, 薛 松. 面向智能驾驶的密集行人检测算法 [J]. 现代计算机, 2024, 30(10):1-10.
- [10] 徐芳芯, 樊 嵘, 马小陆. 面向拥挤行人检测的改进 YOLOv7 算法 [J]. 计算机工程, 2024, 50(3):250-258.
- [11] 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, 46(6):1050-1062.
- [12] 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.