

DOI:10.13382/j.jemi.B2508316

改进 YOLOv10 的密集人群检测算法研究*

康比比 董小明

(安庆师范大学电子工程与智能制造学院 安庆 246133)

摘要: 密集人群检测算法在公共安全监控、智能交通调度等领域意义重大。针对密集场景中目标遮挡、小目标检测精度低及漏检问题,以轻量化 YOLOv10 为基础,提出改进的 YOLOv10-SCD 算法。首先,融入卷积块注意力模块(convolutional block attention module, CBAM),通过通道-空间双维度加权,增强运动模糊处理能力并提升行人检测精度;其次,引入动态采样(dynamic sample, DySample)上采样器提升图像分辨率与处理效率,同时采用改进交并比(SIoU)优化损失函数,进一步提升定位精度与边界框回归速度;最后,在密集人群数据集上验证算法性能,并通过消融实验分析各模块作用。实验表明, YOLOv10-SCD 相较于原始 YOLOv10,核心指标显著提升,其中精确度提升 1.5%、召回率提升 2.9%,平均精度均值(mAP)mAP@0.5 提升 0.8%、mAP@0.5:0.95 提升 1.8%等。消融实验在两套数据集进行,自建数据集聚焦分析各模块对算法精度、效率的单独与协同影响;WiderPerson 公开数据集则验证算法泛化能力。因此, YOLOv10-SCD 能高效应对密集人群复杂场景,缓解目标遮挡、小目标难识别问题,显著提升目标检测鲁棒性与综合性能。

关键词: YOLOv10; Siou; 密集人群检测; CBAM; DySample

中图分类号: TP29; TN911.73

文献标识码: A

国家标准学科分类代码: 520.20

Pedestrian detection algorithm in dense scenes based on improved YOLOv10 algorithm

Kang Bibi Dong Xiaoming

(School of Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, Anqing 246133, China)

Abstract: The dense crowd detection algorithm is of great significance in fields such as public safety monitoring and intelligent traffic scheduling. Aiming at the problems of target occlusion, low detection accuracy of small targets, and missed detection in dense scenes, this paper, based on the lightweight YOLOv10, proposes an improved YOLOv10-SCD algorithm. First, the convolutional block attention module (CBAM) is integrated. Through channel-space two-dimensional weighting, the motion blur processing ability is enhanced and the pedestrian detection accuracy is improved. Second, the dynamic sample (DySample) up-sampler is introduced to improve the image resolution and processing efficiency. At the same time, the Siou optimization loss function is adopted to further improve the positioning accuracy and the bounding box regression speed. Finally, the performance of the algorithm is verified on the dense crowd dataset, and the role of each module is analyzed through ablation experiments. Experiments show that compared with the original YOLOv10, the core indicators of YOLOv10-SCD are significantly improved; the precision is increased by 1.5%, the recall rate is increased by 2.9%, mAP@0.5 is increased by 0.8%, and mAP@0.5:0.95 is increased by 1.8%. The ablation experiments are carried out on two sets of datasets, the self-built dataset focuses on analyzing the individual and synergistic effects of each module on the algorithm accuracy and efficiency; the WiderPerson public dataset verifies the generalization ability of the algorithm. Therefore, YOLOv10-SCD can efficiently cope with the complex scenes of dense crowds, alleviate the problems of target occlusion and difficult recognition of small targets, and significantly improve the robustness and comprehensive performance of target detection.

Keywords: YOLOv10; Siou; intensive crowd detection; CBAM; DySample

0 引言

在公共安全监控中,精准掌握人群聚集态势可及时预警踩踏风险;大型活动人流管控中,实时监测人流密度能科学疏导避免拥堵;智能交通调度里,准确检测路口行人流量是优化信号配时的关键。可见,密集人群的精准检测与分析是保障公共安全、提升管理效率的核心技术支撑。然而,高密度场景下的行人检测面临着严峻挑战:目标间的严重遮挡导致特征信息重叠模糊,小尺度行人(如远景目标)的特征细节微弱难以识别,运动模糊现象进一步加剧特征提取难度,同时还需兼顾检测精度与实时性的平衡。这些问题直接影响着公共安全保障的有效性和管理决策的科学性,因此,研究适用于密集人群场景的高效检测算法具有重要的现实意义和应用价值。

近年来,深度学习驱动的目标检测技术成为解决上述问题的主流方案,形成了双阶段检测框架与单阶段检测框架两大方向。双阶段框架(以 Faster R-CNN 算法^[1]、Mask R-CNN 算法^[2]等为代表),虽能通过候选区域生成提升检测精度,但复杂的两阶段流程导致速度较慢,难以满足密集场景的实时性需求。单阶段框架(以 YOLO(you only look once)算法^[3]、SSD 算法^[4]等为代表),通过卷积神经网络(convolutional neural network, CNN)直接输出目标的位置和类别信息,检测速度更快,更适合实时性要求较高的密集人群检测任务,因此成为该领域的研究热点。围绕 YOLO 系列算法,研究者们从不同角度展开改进以提升密集场景检测性能。在特征融合方面,高强等^[5]用双向特征金字塔网络(bi-directional feature pyramid network, BiFPN)替换路径聚合网络(path aggregation network, PAN),汤静雯等^[6]设计 BDR-FPN,王亚鹏等^[7]采用重复梯度流特征金字塔网络(repeated gradient flow feature pyramid network, RepGFPN)增强不同尺度特征融合效果,同时加入简单注意力模块(simple attention module, SimAM)聚焦关键特征。在注意力机制应用上,杨旭睿等^[8]引入坐标注意力(coordinate attention, CA)机制,并加深主干网络、增加检测头,力求聚焦关键特征并提升小目标识别能力;在网络结构调整上,胡伟超等^[9]引入可变形卷积网 v2(deformable convolutional network v2, DCNv2),并增加小目标检测头,期望提升小目标识别能力;在损失函数与辅助模块上,王洋等^[10]采用距离交并比(distance intersection over union, DIU)损失函数,宁爽等^[11]和石欣等^[12]结合超分辨率网络与引力模型,试图优化检测精度与抗干扰能力。但深入分析这些研究可发现,现有算法存在 3 大共性局限:1)对底层特征(如小目标边缘、运动模糊区域)的提取能力不足,难以有效应对密集场景中复杂的特征干扰,这是因

为多数改进更侧重高层特征融合,对底层细节特征的关注度不够;2)上采样过程中分辨率提升与处理效率的平衡问题尚未完全解决,现有上采样方法要么为追求分辨率牺牲速度,要么为保证效率降低特征质量,影响模型的实时性表现;3)损失函数对边界框回归的角度、形状等细节因素考量不够全面,像 DIU 损失主要关注距离因素,对边界框的角度差异和形状匹配度重视不足,导致定位精度和收敛速度有待进一步提升。选取 YOLOv10n 为基线模型,核心因其适配密集人群检测的轻量化与针对性优势,作为 YOLOv10 轻量版本,无需非极大抑制(non-maximum suppression, NMS),以一致双分配策略、创新标签分配与双头架构提效,结合大核卷积与部分自注意力(partial self-attention, PSA)机制降本,参数仅 2.7×10^6 、计算量 8.2 GFLOPs,兼顾精度与实时性;且底层细节特征保留更优,便于针对遮挡、小目标问题定制优化,聚焦行人检测细分任务。相较 R-CNN 系列双阶段框架, YOLOv10n 省去候选区域生成环节,推理速度提升数倍,避免因流程复杂导致的实时性不足,更适配人群密集场景的动态监测需求;相较 SSD 单阶段框架,其双头架构与特征融合设计对小目标(如远景行人)的检测精度更高,且参数与计算量更小,轻量化优势更显著。YOLOv11 和 YOLOv12 是较新的版本,与本文算法在改进重点、性能表现和适用场景上存在明显差异。在改进重点方面,本文算法聚焦密集人群检测难题,通过引入卷积块注意力模块(CBAM)、运用 DySample 上采样器及采用 SIOU 优化损失函数等进行针对性改进;而 YOLOv11 算法从多方面创新,包括改进骨干网络和颈部架构、在分类检测头增加深度可分离卷积等,注重整体性能提升和多任务支持;YOLOv12 以注意力机制为核心进行架构创新,采用区域注意力方法等,构建更简洁快速的架构。性能表现上,文中算法在密集人群检测场景优势显著,参数规模和计算量较小。针对密集人群检测中的上述共性问题,本文以 YOLOv10n 为基础提出针对性改进方案。首先,引入 CBAM 注意力模块以增强底层特征提取能力,通过动态加权关键特征缓解遮挡和运动模糊的影响;其次,采用 DySample 上采样器替代传统上采样方法,在提升小目标图像分辨率的同时保证处理效率;最后,将损失函数替换为 SIOU,通过综合考虑边界框的角度、距离和形状信息,优化定位精度并加速模型收敛。上述 3 方面改进协同作用,旨在提升算法对密集人群场景的适应性,实现高精度、高效率的行人检测。

1 密集人群检测算法

1.1 YOLOv10 目标检测算法

YOLOv10 系列^[13]包括 n、s、m、l 和 x 5 种模型,这些模型

的卷积层数量不同,并且模型的推理速度逐渐变慢,而推理精度则逐渐提高。YOLOv10 网络的整体结构由 3 个主要部

分组成,主干网络(backbone)、颈部网络(neck)和头部网络(head)。YOLOv10 的具体架构示意图如图 1 所示。

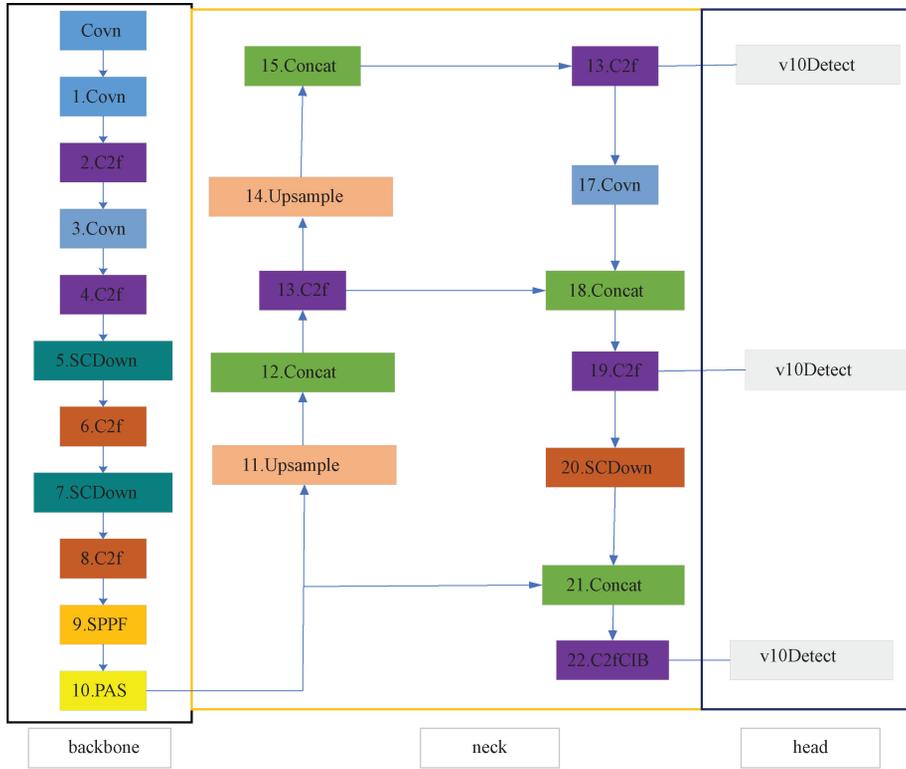


图 1 YOLOv10 网络结构图
Fig. 1 YOLOv10 network structure diagram

YOLOv10^[13]是最新实时端到端目标检测器,采用一致双分配训练策略和 NMS,通过创新的标签分配方法和双头架构提高推理效率。在匹配策略上,YOLOv10 结合一对一和一对多的优势,提供丰富的监督信号,并引入一致匹配度量以提升预测质量。为优化效率,该模型采用轻量化分类头、空间通道解耦下采样以降低延迟,并通过秩引导块设计,减少冗余。为了提升精度,YOLOv10 使用大核卷积扩展感受野,并引入 PSA 机制,降低计算复杂度和内存消耗,进一步增强性能与效率^[13]。

1.2 改进 YOLOV10 算法

1) 引入注意力机制 CBAM 模块

在 YOLOv10 中引入注意力机制 CBAM 模块^[14],该机制的核心思想在于通过赋予不同特征不同的权重来突出关键信息。具体而言,机制通过计算一组注意力权重(即注意力权重),这些权重反映了每个输入特征的相对重要性,并利用这些权重对特征进行加权求和,最终生成一个综合特征表示,如图 2 所示。

经过池化操作后,输入特征分别通过最大池化和平均池化,两种池化特征被送入共享的全连接层。再按元

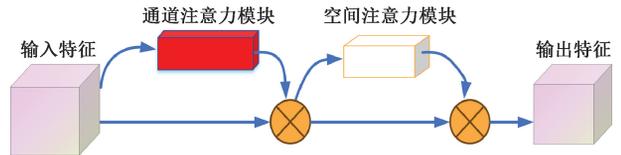


图 2 CBAM 模块结构
Fig. 2 CBAM module structure diagram

素求和,以整合不同池化提取的信息。经过 sigmoid 函数对求和结果进行非线性变换,最终得到通道注意力特征图,如图 3 所示,其公式如下:

$$M_c(\mathbf{F}) = \sigma(MLP(AvgPool(\mathbf{F})) + MLP(Maxpool(\mathbf{F}))) = \sigma(W_0(W_0(\mathbf{F}_{avg}^c)) + W_1(W_0(\mathbf{F}_{max}^c))) \quad (1)$$

式中: W_0 与 W_1 为对应的权重; r 为缩减率; σ 为 Logistid。

当输入特征分别通过最大池化与平均池化后,为了降低特征的维度,将上面的两个池化结果结合起来进行 3×3 卷积操作,最后通经 sigmoid 激活函数得到特征图 $M_s(\mathbf{F})$ 。表达式如下:

$$M_s(\mathbf{F}) = \sigma(f^{7 \times 7}([AvgPOOL(\mathbf{F})]) = \sigma(f^{7 \times 7}([\mathbf{F}_{avg}^s; \mathbf{F}_{max}^s])) \quad (2)$$

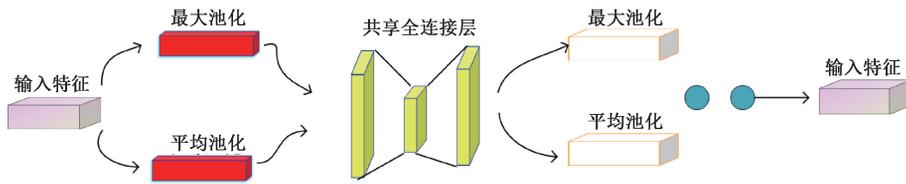


图 3 通道注意力模块

Fig. 3 Channel attention module

式中: $f^{7 \times 7}$ 为卷积核的大小。

注意力机制模块可以增强关键特征的提取,并压制非关键特征。在进行特征提取时,骨干网络 (Backbones) 是重要部分,它能够从输入数据中提取有效的特

征,并快速分析复杂场景中的信息。因此,在 Backbones 层引入 CBAM 模块。YOLOv10 引入 CBAM 模块位置如图 4 所示。

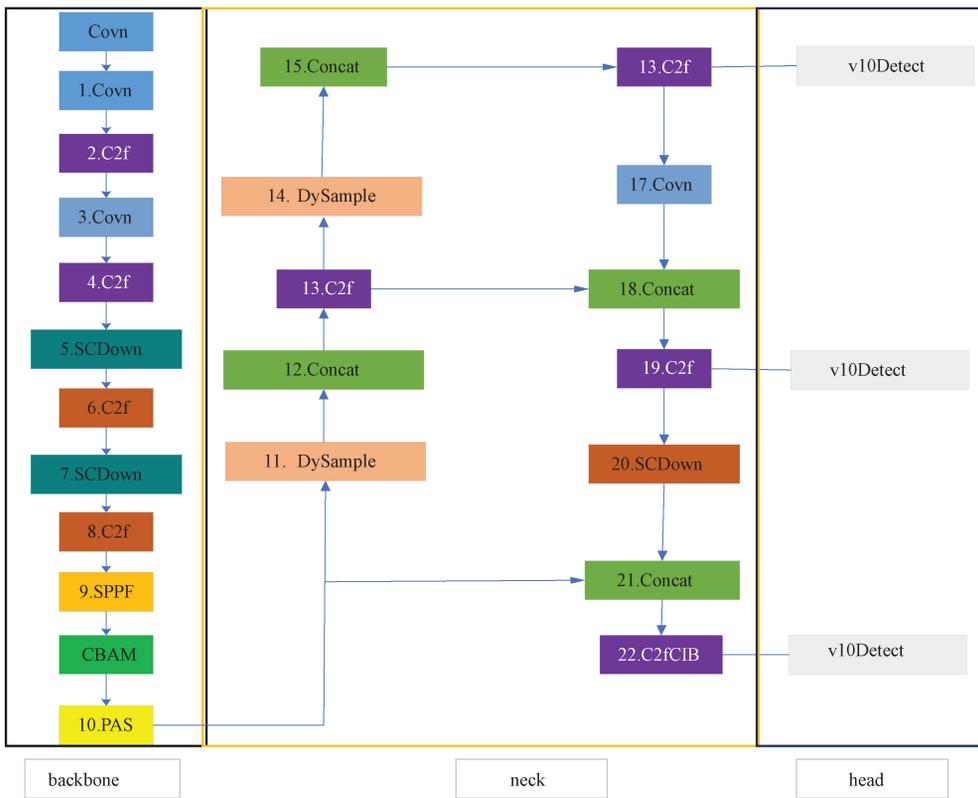


图 4 改进算法结构

Fig. 4 Improved algorithm structure diagram

2) 损失函数选择

YOLOv10 原使用 CIOU Loss 作为损失函数来计算真实框和预测框 (简称两框) 之间的距离。即使 CIOU 损失函数在设计上包含了两框相互重叠部分的面积、长度和宽度的比值等考量因素,但它并未真实反映长宽尺寸与置信度之间的差异,这在某些情况下可能会妨碍模型的改进。相对而言,SIOU^[15] 损失函数能够在回归过程中预测向量之间的夹角大小。并且 SIOU 进一步考虑了真实框和预测框之间的向量角度,重新定义相关损失函数,包

含 4 个部分,分别为角度损失 (angle cost)、距离损失 (distance cost)、形状损失 (shape cost) 和 IOU 损失 (IOU cost)。这不仅加快了回归过程的速率,而且提升了预测的准确性,因此,决定采用 SIOU。

角度函数表达式为:

$$A = 1 - 2\sin^2(\arctan(a) - \frac{\pi}{4}) \quad (3)$$

式中: a 为两框中心点的高度差和的比。距离函数表达式为:

$$\Delta = \sum_{t=a,b} (1 - e^{-\gamma q^t}) = 2 - e^{-\gamma q_a} - e^{-\gamma q_b} \quad (4)$$

式中: γ 值为 $2 - \Delta$; q_a 为外接框中最小宽度与两个框的中心位置之间的宽度差比值的平方数; q_b 为最小外接框的高度和两框中心点高度差比的平方数。

形状函数表达式为:

$$\Omega = \sum_{t=x,h} (1 - e^{-x^t})^\varphi = (1 - e^{-x^x})^\varphi + (1 - e^{-x^h})^\varphi \quad (5)$$

式中: x 下标 x 表示两框宽度差的绝对值与两框最大宽度之比; x 下标 h 表示两框高度差的绝对值与两框最大高度之比; φ 为常数。

由上所述, SIOU 损失函数公式为:

$$L_{SIOU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (5)$$

式中: IoU 为两框的交并比; Δ 为距离函数; Ω 为形状函数。

SIOU 在边界框回归中纳入了待回归向量间的角度信息, 通过引入角度损失因子实现对回归自由度的约束, 同时综合考虑距离损失、形状损失及 IOU 损失, 形成更全面的损失计算机制。这一机制不仅能够加快边界框回归速度以提升模型收敛速率, 还能通过优化惩罚策略提升参数学习效率与推理阶段的定位精度, 尤其在密集人群这类存在大量小目标及遮挡的检测场景中, 可有效增强边界框与行人目标的几何匹配度, 减少因目标重叠或尺度差异导致的定位偏差, 进一步提升检测性能。

3) 上采样设计

为了提高图像质量和数据精度, 本文引入 DySample 模块^[16] 作为上采样器替代原有的 Upsample 模块。DySample 是一种轻量级且有效的动态上采样器, 与传统基于内核的动态上采样器不同的是 DySample 采用了点采样的方法。由于密集人群里有大量且遮挡的小目标行人, 通常主干网络输出的为多尺度特征, 需要将低分辨率特征上采样到高分辨率。因此, 高效且轻量化的上采样器将有利于模型对目标的预测。不仅可以节约计算资源, 而且也不需要高分辨率的图像质量。DySample 的设计如图 5 所示, 采样点生成器如图 6 所示。

在 DySample^[16] 结构中, X 为输入特征, X' 为输出特征, s 为上采样比例因子, G 为原始表格, O 为生成的偏移量, σ 表示 Sigmoid 函数用于生成范围因子, δ 为采样函数。输入特征通过采样点生成器创建采样集, 然后利用网络样本函数对输入特征进行重新采样, 最后得到上采样特征。同一个采样点生成器生成的采样点具有相同的初始位置, 因此, DySample 使用双线性插值技术将输入特征转换为每个采样点的偏移量 O , 以便采样点在输入特征图上均匀分布, 再通过像素洗牌技术结合固定的范围因子来重新调整偏移量 O , 然后与原始网格位置 G 相加得到采样集。如果不做处理, 采样点偏移通常会重叠,

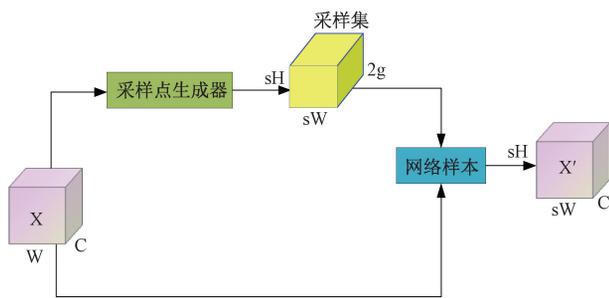


图 5 DySample 结构

Fig. 5 Structure diagram of DySample

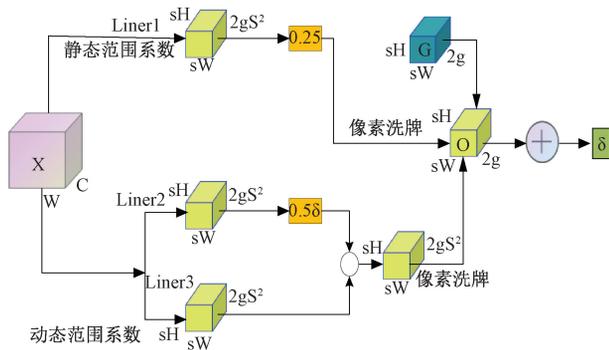


图 6 DySample 采样点生成器

Fig. 6 DySample point generator

从而导致预测伪影。除了线性层和像素洗牌, 还引入了动态范围因子, 用来调节偏移量 O 。偏移量 O 乘以范围因子 0.25 以减少重叠, 从而解决了采样点的初始位置及偏移范围的问题。为了适配密集人群中目标多、遮挡严重的检测场景, 对 DySample 模块核心参数进行针对性调节, 将上采样比例因子 s 设为动态模式, 对低分辨率特征图(远景小目标)采用 $s=4$ 强化细节, 对中分辨率特征图(中景行人)采用 $s=2$ 平衡效率; 将固定范围因子 0.25 改为自适应因子 α (0.15~0.3), 目标密集区取小值减少采样重叠, 稀疏区取大值保障灵活性; 调整采样点生成器网格步长为“密度敏感型”, 小目标区域步长缩小至 2~4 像素以提升采样密度, 大目标区域设为 6~8 像素避免冗余。这些调节使模块能动态适配人群分布特性, 在增强小目标特征提取的同时优化处理效率, 更适配密集人群检测需求。将 DySample 与颈部相结合, 提升了图像分辨率以及图像的处理效率。

2 实验结果与分析

2.1 实验数据集

行人数据集呈现出复杂多变性, 在现实场景中存在遮挡现象, 背景杂乱的现象。因此实验采用了密集场合

下人群的自建数据集和 WiderPerson 数据集^[17]。自建数据集利用网络抓取的方法和各类搜索引擎从网上获取公开图片以及视频截取帧图,收集密集场合下行人图片 2 000 张,利用 Labelimg 对图片进行标注,并按 YOLO 数据集格式制作了一个真实的行人数据集。该数据集包括商场、学校、交通路段等场合图像。同时,将数据集均按照 8 : 1 : 1 的比例划分为训练集、验证集和测试集,并且存在各种遮挡情况,包含远景行人和近景行人目标。

2.2 实验环境

为验证本文方法的有效性,搭建了相关实验平台。操作系统为 Windows 11,集成开发环境为 PyCharm 社区版 2024.2.4,编程语言为 Python 3.11.9,深度学习框架为 PyTorch 2.5.1,CPU 为 Intel(R) UHD Graphics 630,GPU 为 NVIDIA GeForce GTX 1650。学习率 0.01,学习动量为 0.937,使用 SGD 优化器进行参数优化,其权重衰减为 0.000 5。其余设置与原始的 YOLOv10n 数保持一致。

2.3 评价指标

本文选取精确度(Precision,P)、召回率(recall,R)和平均精度均值(mAP)mAP@0.5、mAP@0.5:0.95、模型计算量和每秒检测图片数量等评价指标来对本文改进的算法进行综合评价,其公式如下所示:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \sum_{k=0}^{k=n-1} [R_k - R_{k+1}] \times P_k \quad (9)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_k \quad (10)$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (11)$$

式中:TP(true positive)是真阳性;FP(false positive)是假

阳性;TP(true positive)是真阳性;FN(false negative)是假阴性。平均精度(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 实验设置

选择自建数据集和 WiderPerson 公开数据集进行的预训练获得预权重。总共训练 200 个 epoch,batch size 设置为 16,输入图片大小为 640 × 640。文中选择以 YOLOv10 中模型 YOLOv10n 为基础进行改进。

2.5 实验结果与分析

1) 与主流算法的对比试验

为了进一步验证改进 YOLOv10 算法在检测精度方面的性能,本研究将其与一些主流模型进行对比试验,这些模型包括 YOLOv5s、YOLOv8n^[18]、YOLOv10-BiFPN、YOLOv9-T、YOLOv10n、YOLOv10-Swin_Transformer。为了确保对比试验的公平性,YOLO 系列模型采用相同的超参数配置和训练策略,其他模型则采用默认参数配置,试验结果如表 1 所示,在自制数据集上的对比实验显示,文中改进的 YOLOv10-SCD 算法在密集人群检测任务中表现出显著综合优势:核心性能上,mAP@0.5 达 0.975,mAP@0.5:0.95 达 0.581,均优于 YOLOv8n、YOLOv10-BiFPN 等主流模型,尤其在高 IOU 阈值下对细节特征的捕捉能力突出;精确率(0.942)与召回率(0.935)实现高效平衡,召回率较 YOLOv5s、YOLOv9-T 分别提升 32.4% 和 6.1%,同时精确率优于 YOLOv10 系列改进模型;效率层面,参数规模(2.25 × 10⁶)和计算量(7.9 GFLOPs)为所有对比模型中最小,较 YOLOv8n 参数减少 25.2%、计算量降低 3.7%,较 YOLOv10-BiFPN 参数减少 13.5%、计算量降低 5.9%。综上,该算法通过多模块协同优化,实现了密集人群检测中高精度、高召回与高效率的三重突破,综合性能领先于当前主流 YOLO 系列及改进模型。

表 1 改进 YOLOv10 算法与当前主流算法在自制数据集上的对比试验结果

Table 1 Comparative test results of improving YOLOv10 algorithm and current mainstream algorithm on homemade data sets

模型	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Parameters/(×10 ⁶)	计算量/GFLOPs
YOLOv5s	0.951	0.706	0.775	0.465	7.01	15.8
YOLOv8n	0.947	0.920	0.970	0.562	3.01	8.20
YOLOv9-T	0.955	0.881	0.928	0.535	2.71	10.7
YOLOv10n	0.927	0.906	0.967	0.563	2.77	8.7
YOLOv10-Swin_Transformer	0.924	0.925	0.969	0.572	3.02	8.60
YOLOv10-BiFPN	0.914	0.942	0.972	0.568	2.60	8.40
本文	0.942	0.935	0.975	0.581	2.25	7.90

2) 不同损失函数对比试验

原始损失函数在训练过程中未纳入宽高比检测结

果的影响因素,导致模型收敛速率较慢。为此,在引入相关机制的基础上,本研究通过对比不同损失函数的性能

表现,分析其对模型检测效能的影响差异,试验结果如表 2 和图 7 所示。结果表明,同时考虑了预测框和真值框的角度、形状、和距离的 SIOU 损失函数在数据集上的表现优于其他损失函数,在 YOLOv10 模型引 CBAM 与 DySample 模块的基础上,采用 SIOU 损失函数时,模型性能优势显著。SIOU 召回率达 0.935 居对比损失函数之

首,精确率 0.942 仅略低于 DIOU;综合性能上,mAP@ 0.5 为 0.975 高于其他,mAP@ 0.5 : 0.95 达 0.581,与 GIOU、EIOU 持平且优于 CIOU、DIOU;检测帧率为 47.8 fps 是所有对比损失函数中最高,保证检测性能同时维持良好实时性,且 SIOU 与 CBAM、DySample 兼容性佳,实现检测性能与运行效率平衡。

表 2 不同损失函数在改进 YOLOv10 模型上的对比实验结果

Table 2 Comparative experimental results of different loss functions on improved YOLOv10 model

模型	Loss Function	Precision	Recall	mAP@ 0.5	mAP@ 0.5 : 0.95	检测帧率/fps
YOLOv10+CBAM+ DySample	CIOU	0.925	0.918	0.971	0.576	40.6
	DIOU	0.943	0.913	0.974	0.580	42.7
	GIOU	0.929	0.921	0.973	0.581	45.4
	EIOU	0.922	0.927	0.972	0.581	42.9
	SIOU	0.942	0.935	0.975	0.581	47.8

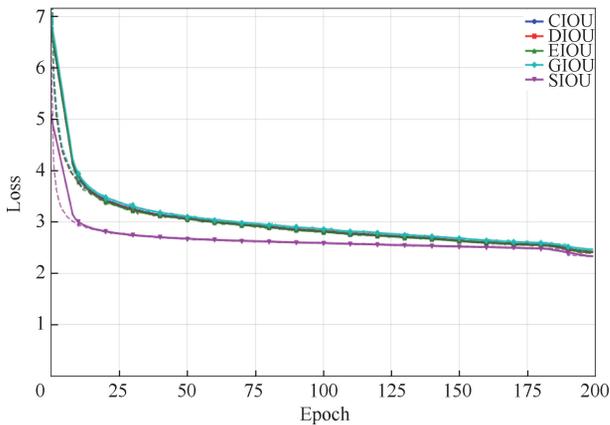


图 7 不同 IoU 损失函数试验结果对比

Fig. 7 Comparison of test results of different IoU loss functions

3) 不同注意力机制对比试验

在引入相关机制的基础上,为系统探究注意力机制对密集人群检测模型性能的影响规律,本文设计了多组对比实验,选取 CoordAtt、SA、CBAM、EMA、SimAM 等主流注意力机制,在统一实验框架下(相同数据集、训练超

参数及评估标准)分别嵌入模型架构。通过量化分析不同注意力机制作用下模型的 Precision、Recall、mAP 等核心性能指标,重点考察其在特征增强效能(对遮挡目标的分辨能力、小目标行人的识别精度,以及对运动模糊的鲁棒性)方面的表现差异,以此揭示各类注意力机制在密集人群检测场景中的适配特性,为优化模型架构、实现检测精度与运行效率的协同提升提供实验支撑,试验结果如表 3 所示。由表 3 可知,在密集人群检测任务中,当在主干网络引入 CBAM 后,模型在精度上,mAP@ 0.5 (0.975) 仅次于 EMA,mAP@ 0.5 : 0.95 (0.581) 为所有机制中最高,mAP@ 0.5 从 96.7% 提升至 97.5%对密集人群检测的复杂场景(如多尺度目标、遮挡环境)适配性更强,并且 CBAM 注意力机制的综合性能显著优于其他机制。这一结果表明,CBAM 通过协同关注通道维度的关键特征与空间维度的显著区域,能够精准增强密集人群中的有效目标特征,抑制冗余背景信息,从而有效提升对遮挡目标和小目标行人的检测能力。基于此,最终确定在 YOLOv10n 的主干网中加入 CBAM 注意力机制。

表 3 不同注意力机制试验结果

Table 3 Experimental results of different attention mechanisms

注意力机制	mAP@ 0.5	mAP@ 0.5 : 0.95	Parameters/(× 10 ⁶)	计算量/GFLOPs
+SA	0.974	0.580	2.26	8.0
+ CoordAtt	0.974	0.571	2.33	8.4
+EMA	0.977	0.579	2.90	8.4
+SimAM	0.972	0.575	2.24	7.9
+CBAM	0.975	0.581	2.25	7.9

4) 消融实验

为了验证提出的改进方法的有效性,进行了一系列的消融试验,结果如表 4 和图 8 所示,首先,以 YOLOv10n

为基线实验模型,其精度为 0.927,召回率为 0.906,mAP@ 0.5 为 0.967,mAP@ 0.5 : 0.95 为 0.563;实验 1 引入 CBAM 模块,精度为 0.907,召回率为 0.909,mAP@

0.5 升至 0.969, $mAP@0.5:0.95$ 提升至 0.575, 虽精确率略有下降, 但召回率和平均精度指标均有改善, 说明该模块能增强低层特征提取, 对缓解遮挡和运动模糊有一定作用; 实验 2 引入 SIOU 损失函数, 精度为 0.902, 召回率显著提升至 0.935, $mAP@0.5$ 升至 0.971, $mAP@0.5:0.95$ 为 0.568, 表明其在提升召回率和定位精度方面效果明显, 能加速边界框回归, 优化密集场景下的目标定位; 实验 3 引入 DySample 模块, 精度提升至 0.937, 召回率为 0.914, $mAP@0.5$ 升至 0.974, $mAP@0.5:0.95$ 提升至 0.577, 各项指标均有提升, 说明该模块能有效提升图像分辨率和处理效率, 对小目

标检测性能改善显著; 实验 4 结合 CBAM 和 DySample 模块, 精度为 0.925, 召回率为 0.918, $mAP@0.5$ 为 0.971, $mAP@0.5:0.95$ 为 0.576, 性能较单个模块呈现不同变化, 体现了特征增强与分辨率提升模块协同作用的特点; 实验 5 同时引入 CBAM、DySample 模块和 SIOU 损失函数(即文中算法), 精度提升至 0.942, 召回率为 0.935, $mAP@0.5$ 达 0.975, $mAP@0.5:0.95$ 提升至 0.581, 所有指标均为最高, 且参数和计算量降至最低(2.25×10^6 , 7.9 GFLOPs), 表明各模块合理组合能实现协同优化, 显著提升模型在密集人群场景下的检测性能, 证明整体改进方案有效。

表 4 在自建数据集消融实验结果

Table 4 Ablation experiment results on the self-built dataset

模型	Precision	Recall	$mAP@0.5$	$mAP@0.5:0.95$	Parameters/ $(\times 10^6)$	计算量/GFLOPs
YOLOv10n	0.927	0.906	0.967	0.563	2.77	8.7
YOLOv10n+CBAM	0.907	0.909	0.969	0.575	2.45	8.0
YOLOv10n+SIOU	0.902	0.935	0.971	0.568	2.71	8.4
YOLOv10n+DySample	0.937	0.914	0.974	0.577	2.70	8.0
YOLOv10n+CBAM+ DySample	0.925	0.918	0.971	0.576	2.26	8.0
本文	0.942	0.935	0.975	0.581	2.25	7.9

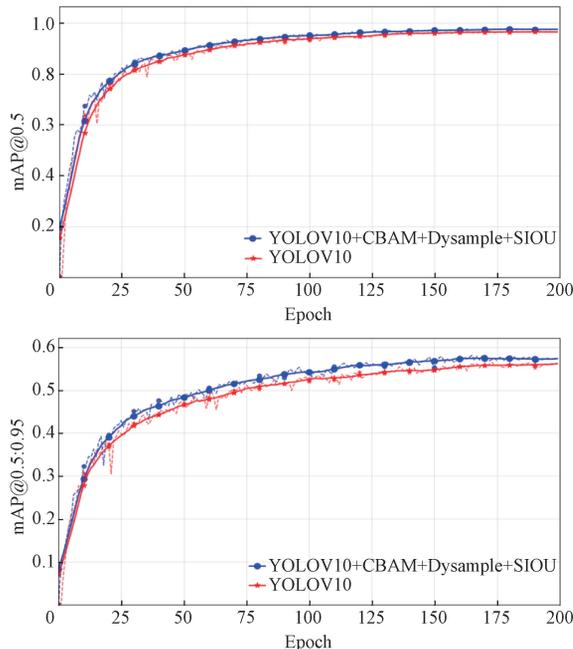


图 8 在自建数据集改进前后评价指标对比

Fig. 8 Comparison of evaluation indicators before and after improvement

5) 泛化实验

为验证改进模型的泛化能力与综合性能, 本文从数据集选择、实验结果分析及性能可视化 3 个维度开展验证。考虑到自制数据集规模有限易致模型过拟合, 选取密集行人检测领域广泛认可的 WiderPerson 公开数据集

进行泛化实验。该数据集含 13 382 张图像(划分为 8 000 张训练集、1 000 张验证集、4 382 张测试集), 标注约 40 万个目标实例, 其优势在于覆盖多样化背景环境、包含显著遮挡特征与复杂人群分布, 且标注类型涵盖行人、骑行者、部分遮挡目标等, 可有效规避自建数据集分布单一的局限, 既能检验算法在不同场景下的适应性, 也便于与主流算法横向对比, 增强结果说服力。实验结果如表 5 所示, 在 WiderPerson 数据集的泛化实验中, 本文算法相较于 YOLOv10 n 基线模型展现出更优的综合性能: 精确率(0.854 \rightarrow 0.858)、召回率(0.794 \rightarrow 0.796)均有小幅提升, 表明对密集目标的识别准确性与覆盖度有所增强; $mAP@0.5$ (0.884 \rightarrow 0.885)及 $mAP@0.5:0.95$ (0.626 \rightarrow 0.631)的提升, 印证了算法在不同 IOU 阈值下检测性能的稳定性, 尤其在高阈值场景中对目标细节的捕捉能力更优。同时, 模型参数规模($2.69 \times 10^6 \rightarrow 2.25 \times 10^6$)与计算量(8.4 GFLOPs \rightarrow 7.9 GFLOPs)均有所降低, 实现了性能提升与轻量化的协同优化, 充分验证了改进策略在提升密集人群检测泛化能力与运行效率方面的有效性。

6) 可视化结果对比分析

图 9 所示为原始 YOLOv10n 算法和文中所提改进算法 YOLOv10-SCD 的检测效果对比。从图 9 可以看出, 原始模型会出现检测精度低和小目标漏检的情况。而改进的 YOLOv10 算法的检测效果不仅精度高, 还能检测模糊或者小尺寸的行人, 证明了文中所提算法在密集场景行人检测的有效性。

表 5 在 WiderPerson 数据集泛化实验结果

Table 5 Experimental results of generalization on the WiderPerson dataset

模型	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Parameters/($\times 10^6$)	计算量/GFLOPs
YOLOv10n	0.854	0.794	0.884	0.626	2.69	8.4
本文	0.858	0.796	0.885	0.631	2.25	7.9



图 9 YOLOv10 算法与文中算法的检测结果对比

Fig. 9 Comparison of the detection results of YOLOv10 algorithm and this algorithm

3 结 论

本文针对 YOLOv10n 模型在密集人群场景下的检测优化问题,提出了基于 YOLOv10 改进方案。改进包括在 YOLOv10 网络中引入 CBAM 注意力模块以有效减少特征空间稀疏化带来的检测偏差并提升行人检测精度;同时,将 DySample 上采样模块应用于图像处理,通过动态点采样机制显著提升图像分辨率及图像处理效率,针对性解决密集场景中中小目标分辨率不足的问题;最后,将 CIOU 损失函数替换为 SIOU 边界框损失函数,不仅提升了行人定位的精准度,有效的改善了在密集场合下多目标漏检,错检等问题,并且进一步提升了低召回率场景下的检测性能,能够更有效地处理密集人群中的复杂场景,从而更好的提升整体目标检测的鲁棒性和性能。实验结果表明,改进方案在行人的检测精度方面表现出色,在密集人群复杂场景下实现低召回率场景下的检测能力提升,进一步完善了 YOLOv10 算法在密集人群中的鲁棒性和整体性能。尽管改进后的 YOLOv10 算法在密集人群检测方面取得了一定进展,但仍需拓展其面向不同密集人群数据集的训练策略以增强模型的鲁棒性。在接下来,将改进的 YOLOv10 算法与改进滤波的 DeepSort 算法相结合,最终实现更全面、更高效的目标检测跟踪系统。

参考文献

[1] 翟永杰,王璐瑶,郭聪彬. 基于 Faster R-CNN 的复杂背景下绝缘子目标检测[J]. 电子测量技术, 2023, 46(16): 187-194.
ZHAI Y J, WANG L Y, GUO C B. Insulator object detection in complex background based on Faster RCNN[J]. Electronic Measurement Technology, 2023, 46(16): 187-194.

[2] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2020, 42(2): 386-397.

[3] 邓亚平,李迎江. YOLO 算法及其在自动驾驶场景中目标检测综述 [J]. 计算机应用, 2024, 44(6): 1949-1958.
DENG Y P, LI Y J. Overview of YOLO algorithm and its target detection in automatic driving scene [J]. Computer Applications, 2024, 44(6): 1949-1958.

[4] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020, 48(6): 1230-1239.
LUO H L, CHEN H K. A Survey of object detection based on deep learning [J]. Acta Electronica Sinica, 2020, 48(6): 1230-1239.

[5] 高强,唐福兴,李栋,等. 基于改进 YOLOv5 的密集场景行人检测方法研究[J]. 国外电子测量技术, 2023, 42(4): 125-130.
GAO Q, TANG F X, LI D, et al. Pedestrian detection

- in dense scenes based on improved YOLOv5 [J]. Foreign Electronic Measurement Technology, 2023, 42(4):125-130.
- [6] 汤静雯, 赖惠成, 王同官. 远距离情形下的改进 YOLOv8 行人检测算法 [J]. 计算机工程, 2025, 51(4):303-313.
- TANG J W, LAI H CH, WANG T G. Improved YOLOv8 pedestrian detection algorithm for long distance cases [J]. Computer Engineering, 2025, 51(4):303-313.
- [7] 王亚鹏, 韩文花. 改进 YOLOv5 算法下的无人驾驶道路行人识别研究 [J]. 国外电子测量技术, 2024, 43(6):170-178.
- WANG Y P, HAN W H. Pedestrian recognition on driverless road based on improved YOLOv5 algorithm [J]. Foreign Electronic Measurement Technology, 2024, 43(6):170-178.
- [8] 杨旭睿, 冯宇平, 李悦, 等. 基于注意力机制改进 YOLO-V5 的多尺度行人目标检测 [J]. 青岛科技大学学报(自然科学版), 2024, 45(5):127-134.
- YANG X R, FENG Y P, LI Y, et al. Multi-scale pedestrian target detection based on attention mechanism improved YOLO-V5 [J]. Journal of Qingdao University of Science and Technology (Natural Science Edition), 2024, 45(5):127-134.
- [9] 胡伟超, 皮建勇, 胡倩, 等. 面向复杂场景密集行人检测的 YOLOv8 改进模型 [J]. 电子测量技术, 2024, 47(14):159-169.
- HU W CH, PI J Y, HU Q, et al. YOLOv8 improved pedestrian detection model for dense pedestrian in complex scenes [J]. Electronic Measurement Technology, 2024, 47(14):159-169.
- [10] 王洋. 无人驾驶车辆中基于深度学习的夜间横穿行人检测算法研究 [D]. 西安:西安工业大学, 2024.
- WANG Y. Research on night crossing pedestrian detection algorithm based on deep learning in autonomous vehicles [D]. Xi'an:Xi'an University of Technology, 2024.
- [11] 宁爽, 宋辉. 帧间方向梯度直方图特征关联的行人检测方法 [J]. 电子测量与仪器学报, 2024, 38(5):112-118.
- NING SH, SONG H. Pedestrian detection method based on inter-frame oriented gradient histogram feature association [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5):112-118.
- [12] 石欣, 卢灏, 秦鹏杰, 等. 一种远距离行人小目标检测方法 [J]. 仪器仪表学报, 2022, 43(5):136-146.
- SHI X, LU H, QIN P J, et al. A method for detecting small pedestrian targets in long distance [J]. Acta Instrumentation Sinica, 2022, 43(5):136-146.
- [13] WANG A, CHEN H, LIU L H, et al. YOLO-v10:Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2018:3-19.
- [15] GEVORGYAN Z. SIOU loss: More powerful learning for bounding box regression [J]. ArXiv preprint arXiv:2205.12740, 2022.
- [16] 张瑞芳, 刘占占, 程小辉, 等. 基于改进 YOLOv8 的无人机红外目标检测算法 [J]. 电子测量技术, 2025, 48(7):46-54.
- ZHANG R F, LIU ZH ZH, CHENG X H, et al. UAV IR target detection algorithm based on improved YOLOv8 [J]. Electronic Measurement Technology, 2025, 48(7):46-54.
- [17] ZHANG S F, XIE Y L, WAN J, et al. Wider-person: A diverse dataset for dense pedestrian detection in the wild [J]. IEEE Transactions on Multimedia, 2020, 22(2):380-393.
- [18] 方涛涛, 王池社, 王洁, 等. 基于 YOLOv8n 的探地雷达图像管线定位方法 [J]. 国外电子测量技术, 2023, 42(11):170-177.
- FANG T T, WANG CH SH, WANG J, et al. Ground penetrating radar image pipeline location based on YOLOv8n [J]. Foreign Electronic Measurement Technology, 2023, 42(11):170-177.

作者简介



康比比, 2024 年于河南工程学院获学士学位, 现为安庆师范大学硕士研究生, 主要研究方向为图像处理与机器视觉。

E-mail: 197599133@qq.com

Kang Bibi received his B. Sc. degree from Henan University of Engineering in 2024. He is now a M. Sc. candidate at Anqing Normal University. His main research interests include image processing and machine vision.



董小明 (通信作者), 1999 年于北京科技大学获学士学位, 2007 年于北京科技大学获硕士学位, 2010 年于中国科学院自动化研究所获博士学位, 现为安庆师范大学教授, 主要研究方向图像处理与机器视觉。

E-mail: 1006166992@qq.com

Dong Xiaoming (Corresponding author) received his B. Sc. degree from University of Science and Technology Beijing in 1999, M. Sc. degree from University of Science and Technology Beijing in 2007 and Ph. D. degree from Institute of Automation, Chinese Academy of Sciences in 2010. Now he is a professor in Anqing Normal University. His main research interests include image processing and machine vision.