

MSS-YOLO: 通关场景下遮挡目标检测算法

陈泊睿¹, 张 梅², 李昕蕊¹

(1. 中国人民警察大学研究生院, 廊坊 065000; 2. 中国人民警察大学移民管理学院, 廊坊 065000)

摘 要: 目标检测是智慧口岸实现对通关人员行为识别的前置任务, 为了解决快捷通道处排队拥挤、闸机设备等带来的遮挡导致现有模型检测精度下降, 以及兼顾边缘设备要求等问题, 提出了一种改进YOLOv12n的遮挡目标检测算法MSS-YOLO。首先, 在主干网络中设计了自适应多尺度特征提取(Adaptive multi-scale feature extraction, AMFE)模块替换下采样模块扩大模型的感受野, 并自适应加权关注关键人体特征区域, 增强多尺度特征的提取能力缓解特征信息丢失和遮挡干扰的问题; 其次, 在颈部网络引入改进的特征增强模块A2C2f_SEAM, 学习遮挡与未遮挡区域的关系, 动态调节关注点; 最后, 将CIoU替换为Alpha-Shape IoU, 通过引入幂次变换和几何约束, 进一步优化边界框回归性能。本文分别在CrowdHuman实验子集和自采口岸快捷通道翻越行为数据集CBCD上进行实验验证。结果表明, 相较于基线模型, 本文提出的MSS-YOLO模型在两个数据集上的精确率分别提升了1.4和6.1个百分点, 召回率分别提升了7.1和4.3个百分点, mAP@0.5分别提升了5.4和5.0个百分点, mAP@0.5:0.95分别提升了4.5和12.7个百分点, 同时模型参数量减少了 0.23×10^6 。MSS-YOLO在提升口岸通关场景下遮挡人员检测精度的同时保持了较小的参数规模, 为后续端侧部署与优化提供了基础。

关键词: 智慧口岸; 遮挡目标检测; YOLOv12n; 多尺度特征提取

中图分类号: TP391

文献标志码: A

引用格式: 陈泊睿, 张梅, 李昕蕊. MSS-YOLO: 通关场景下遮挡目标检测算法[J]. 数据采集与处理, 2026, 41(5): 1349-1365. CHEN Borui, ZHANG Mei, LI Xinrui. MSS-YOLO: Occluded object detection in customs scenarios[J]. Journal of Data Acquisition and Processing, 2026, 41(5): 1349-1365.

引 言

为提升口岸智能化, 通过计算机视觉技术实现对通关人员的自动检测、识别与行为分析, 以提升通关效率与风险防控能力^[1]。在这一过程中, 目标检测是异常行为识别的前置环节, 其准确性与鲁棒性对整体系统性能具有决定性影响^[2-4]。然而, 与开放道路或普通公共场所不同, 口岸快捷通道具有高度结构化且拥挤的空间特征。在实际运行中, 行人往往在闸机区域呈现出密集排队状态, 人体关键部位容易被闸机设备、行李箱以及前后通行人员遮挡。这类复杂遮挡与尺度不一致现象使得行人目标呈现出局部可见、边界模糊和形态不完整等特点, 增加了目标检测的难度, 出现检测精度下降、漏检等问题^[5]。

近年来, 遮挡问题作为行人检测领域的核心难点, 严重干扰了检测器对目标的精准定位与分类, 相关研究已成为目标检测领域的重要方向。Zhang等^[6]提出FC-Net, 通过自激活与特征校准模块, 动态强化行人可见部分的语义表征, 从而增强遮挡场景下对行人可见部分的表征能力, 提高遮挡行人检测性能。刘毅等^[7]提出了一种新的基于注意力机制的UAST R-CNN模型, 利用改进的Faster-RCNN框架,

结合 Swin-Transformer 主干网络,提升全局的感受野,进一步加强了多尺度特征提取能力,并采用渐进式焦点损失,增强了对遮挡行人的检测精度。王家辉等^[8]针对视频目标检测中目标遮挡进而使跨帧特征聚合失效的问题,引入轻量级跟踪器与特征选择模块融合检测信息和运动先验,并利用加权框融合 (Weighted boxes fusion, WBF) 合并检测框与跟踪预测框以生成更可靠的候选区域,随后在轨迹上进行时序特征聚合,从而提升遮挡场景下的检测鲁棒性与精度。尽管在不同方法上缓解了遮挡带来的困难,但部分方法多依赖复杂主干网络或 Transformer 结构,模型计算开销较大,难以满足口岸快捷通道对实时性与部署成本的实际要求,对遮挡条件下边界框回归精度的针对性建模不足,难以准确刻画通关人员在部分遮挡、尺度变化条件下的真实空间形态^[9-10]。

因此,本研究提出基于改进 YOLOv12n^[11]的对存在遮挡情况的通关人员检测算法 MSS-YOLO。该算法以 YOLOv12^[11]为基准模型,融合新设计的自适应多尺度特征提取 (Adaptive multi-scale feature extraction, AMFE) 模块、空间增强注意力模块 (Spatially enhanced attention module, SEAM) 机制^[12],以及改进的 Alpha-Shape IoU 模块。MSS-YOLO 会自适应地调整空间维度上的边缘特征权重和通道维度上高对比度区域的权重,从而突出人员目标的显著特征,提高遮挡情况下的检测率。模型在提高遮挡目标检测精度的同时保持了较小的参数规模,为其在口岸端侧设备上的进一步应用提供了基础。具体贡献如下:

(1) 设计了 AMFE 模块并替换原有骨干网络下采样卷积层。该模块通过并行的多尺度卷积路径与轻量化注意力机制,在下采样阶段扩大感受野同时构建局部细节特征与大范围上下文信息,有效缓解了遮挡条件下行人特征丢失的问题,可为后续特征融合提供更具判别力的多尺度表示。

(2) 由于通关人员特征部分丢失,导致检测器无法准确区分人员与背景,因此将 SEAM 机制引入 A2C2f 模块,通过双路径并行注意力架构,模型能学习遮挡区域与未遮挡区域之间的关系。SEAM 机制帮助模型在遮挡情况下动态调整关注点,从而更好地处理被遮挡的行人信息,提升遮挡场景下的检测精度。

(3) 现有损失函数 CIoU 对目标形状与尺度差异的细粒度刻画仍有限,导致边框回归精度不高,引入 Shape-IoU^[13] 损失函数后,模型能够专注于预测框与真实框之间的形状和尺度差异。同时引入 α 损失函数^[14]的思想构建 Alpha-Shape IoU,旨在改善梯度调节和边界框回归,通过聚焦边框自身的几何特征和尺度优化边框回归,提升模型在复杂遮挡场景中的定位精度。

1 MSS-YOLO 模型

YOLOv12 是一种单阶段目标检测算法,通过优化网络结构实现了实时性与检测精度的平衡。该模型由主干网络 (Backbone)、特征融合网络 (Neck) 和检测头 (Head) 3 部分组成,其中 Backbone 负责提取多层次特征信息,Neck 用于多尺度特征融合,Head 完成目标分类与边界框回归任务。此外,其引入区域注意力机制 (AreaAttention)^[15]以增强特征建模能力,并采用残差高效层聚合网络 (R-ELAN) 结构以提升网络训练稳定性和检测效率^[16]。YOLOv12 作为新一代单阶段目标检测算法,通过深度优化网络架构,实现了实时性与检测精度的协同提升,结构设计如图 1 所示。

本文针对 YOLOv12 基线模型感受野单一,在遮挡场景下局部特征提取能力弱,以及边界框回归精度受限的缺点提出改进模型 MSS-YOLO 模型,模型结构如图 2 所示。MSS-YOLO 模型架构与基线模型改进点如下:引入设计的 AMFE 模块替换主干网络中部分卷积层扩大感受野以提升下采样过程中的多尺度特征提取能力;将颈部网络中部分 A2C2f 模块替换为融合 SEAM 机制的 A2C2f_SEAM 模块,以增强中、高层特征对遮挡目标的建模能力;采用 Alpha-Shape IoU 损失函数替换原来的 CIoU^[17],以提升边界框对目标形状与尺度差异的建模精度。

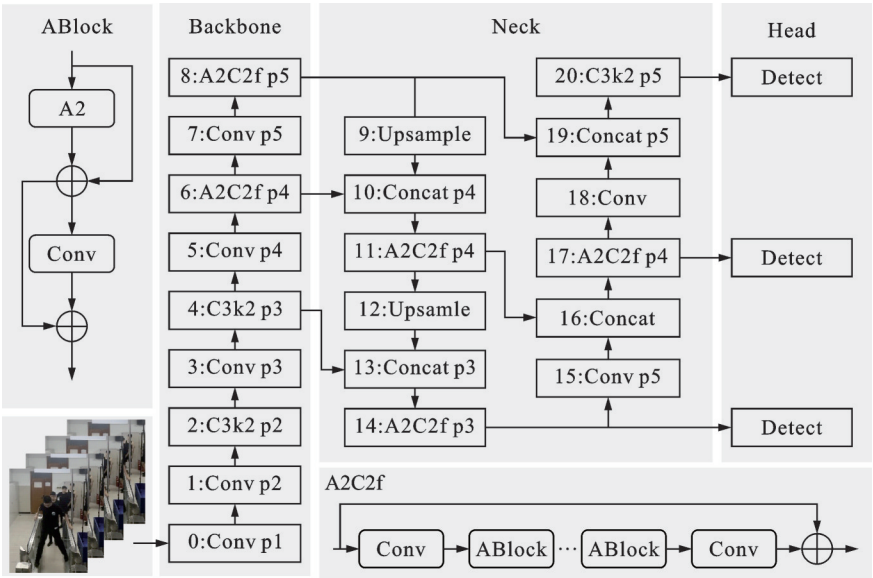


图1 YOLOv12模型结构
Fig.1 YOLOv12 model structure

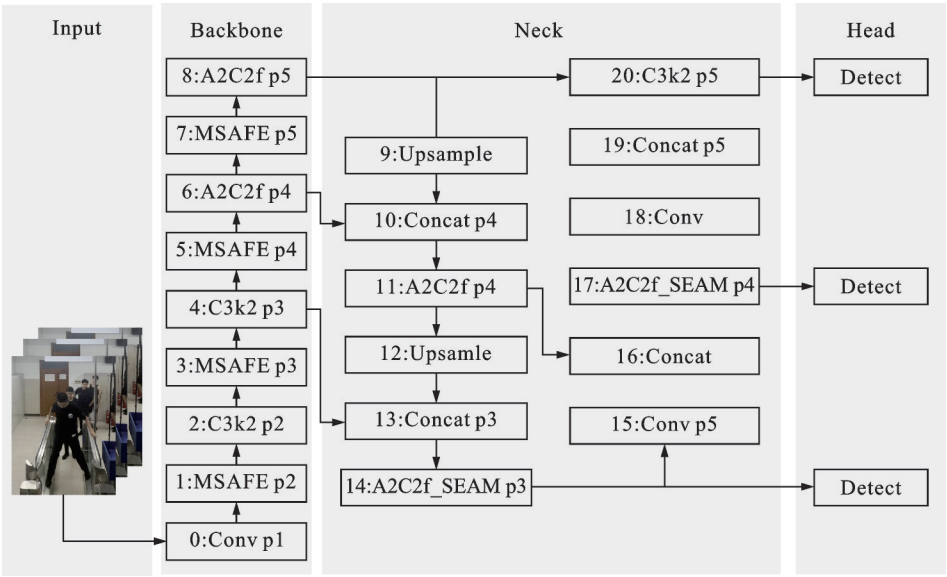


图2 MSS-YOLO模型结构图
Fig.2 MSS-YOLO model structure

1.1 AMFE 模块

YOLOv12的特征提取能力在遮挡场景下存在不足,难以适应遮挡目标检测对多尺度信息的复杂需求。基础卷积单元仍采用Conv 3×3-BN-SiLU串联结构,该结构在通道变换和局部空间特征提取上具有较高的计算效率,但感受野由固定的3×3卷积核决定,由于在通关场景下,通关人员会出现遮挡,往往只暴露头肩等局部区域,单一的感受野难以同时兼顾局部细节和较大范围上下文,导致遮挡的行人学习不敏感,出现漏检、检测置信度偏低的问题。为此,本文在YOLOv12的Backbone结构中引入了

AMFE 模块, 以实现多尺度特征建模和遮挡区域特征增强, 缓解特征信息丢失和遮挡干扰的问题, 如图 3(a) 所示。

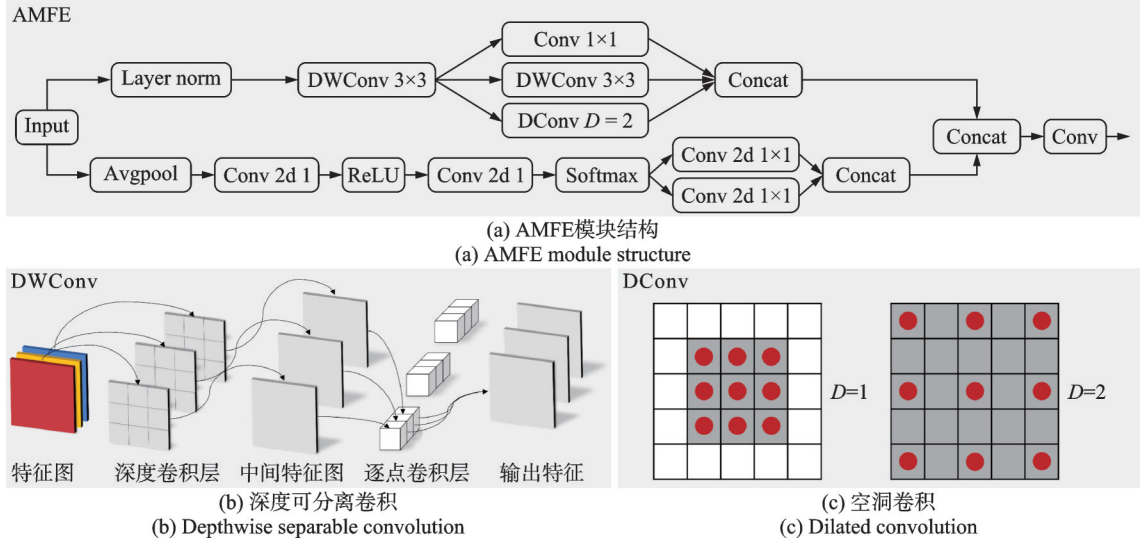


图3 AMFE 模块

Fig.3 AMFE module

AMFE 模块主要由两部分组成: (1) 多尺度卷积分支, 通过引入不同感受野的并行卷积路径, 在轻量化前提下捕获从局部纹理到全局上下文的多尺度人体特征; (2) 轻量级通道-空间注意力分支, 通过自适应加权突出关键人体区域, 抑制遮挡背景噪声。同时, 在模块输入端引入 LayerNorm 对特征进行归一化, 有助于稳定深层网络的训练过程, 减少梯度波动和特征分布的不均衡, 从而提升网络对遮挡行人的表征能力。

设输入特征为 $X \in \mathbb{R}^{C \times H \times W}$, 输入通道数为 C , 输入尺寸为 $H \times W$ 。首先, 为稳定训练对其进行 LN (LayerNorm) 归一化处理, 表达式为

$$\text{LN}(X) = \frac{X - E[X]}{\sqrt{\text{Var}[X] + \epsilon}} \odot \gamma + \beta \quad (1)$$

式中: $E[X]$ 和 $\text{Var}[X]$ 分别为特征在通道维度上的均值和方差, ϵ 表示极小的正常数, γ 和 β 为可学习的仿射参数, “ $\odot$ ” 为 Hadamard 积, 表示逐元素乘法操作。随后, 通过一个 3×3 的深度可分离卷积 (Depthwise separable convolution, DWConv)^[18] 进行初步空间编码, 输出经过 BN 批归一化得到基础特征 F_0 , 即

$$F_0 = \text{BN}(\text{DWConv}_{3 \times 3}(\text{LN}(X))) \quad (2)$$

F_0 被送入 1×1 点卷积、深度可分离卷积、空洞卷积并行的卷积分支, 每一分支对应不同的感受野和建模侧重点, 从而实现多尺度特征的充分提取。

1×1 点卷积分支主要负责通道维度上的线性变换与重映射。遮挡场景下, 目标的有效信息往往集中于少量关键区域, 直接进行空间卷积容易引入冗余响应。通过 1×1 卷积对通道特征进行压缩与筛选, 可以强化与行人相关的判别性语义, 削弱无关背景干扰, 为后续多分支特征融合提供更加紧凑的语义表示, 即

$$F_1 = \text{BN}(\text{Conv}_{1 \times 1}(F_0)) \quad (3)$$

为了更好地保留边缘、轮廓和局部结构特征,该分支引入DWConv提取局部空间信息,结构如图3(b)所示。DWConv主要用于提取局部空间细节信息。对于遮挡行人,头肩轮廓和躯干边缘等局部结构往往是完成检测的重要依据,因此需要增强模型对细粒度纹理与边界特征的感知能力。相较于标准卷积,深度可分离卷积能够在较低参数开销下保留局部空间建模能力,更适合在轻量化条件下强化局部可见区域的特征表达。该分支输出可表示为

$$F_2 = \text{BN}(\text{DWConv}_{3 \times 3}(F_0)) \quad (4)$$

感受野的大小直接影响神经元在特定位置的感知能力,增大感受野能够提升模型对特征的提取能力。为了进一步扩大模型的有效感受野并增强上下文信息的建模能力,本分支引入空洞卷积(Dilated convolution, DConv)^[19]。空洞卷积通过在卷积核内部按照膨胀率 d 生成的间隔,使卷积能够在不增加参数量和计算量的前提下感知更大范围的特征结构。

图3(c)直观展示了空洞卷积的特性,本文采用空洞率为 $d=2$ 、卷积核为 3×3 的空洞卷积,相比等效 5×5 标准卷积感受野覆盖的范围,能降低64%的计算成本。不同空洞率对应不同感受野尺寸,通过调节空洞率参数,可灵活控制等效感受野的尺寸,实现感知范围扩大及全面捕获上下文信息。本分支的最终输出形式可以表示为

$$F_3 = \text{BN}(\text{DConv}_{3 \times 3, d=2}(F_0)) \quad (5)$$

3类分支在功能上互补,可共同提升模型对局部可见目标和尺度变化目标的表征能力。多尺度特征融合过程表示为

$$F_{\text{ms}} = \text{Concat}(F_1, F_2, F_3) \quad (6)$$

式中:特征 F_{ms} 同时蕴含全局语义信息、局部纹理和细节特征以及扩展感受野的上下文信息。与原YOLOv12中采用 3×3 Conv-BN-SiLU构成的单一路径卷积模块相比,多尺度卷积分支通过并行卷积路径提取不同感受野下的特征,在一定程度上增强了对尺度变化的适应能力。

尽管多尺度融合能够增强特征表达,但在复杂背景和严重遮挡条件下,目标区域与干扰区域仍可能产生相近响应。为进一步增强目标可分性,AMFE模块在多尺度卷积分支的基础上并联引入轻量化注意力机制,对融合特征进行自适应加权,增强人体区域响应,抑制遮挡和背景干扰,从而提升检测的稳定性与准确性。以输入特征 X 为基础,首先通过全局平均池化操作提取通道方向的全局统计量,有

$$z = \text{AvgPool}(X) \quad z \in \mathbf{R}^{C \times 1 \times 1} \quad (7)$$

随后,将 z 输入由两个串联的 1×1 卷积和ReLU激活函数构成的非线性映射中,得到通道注意力向量 u ,有

$$u = \sigma_1\left(\text{Conv}_{1 \times 1}^2\left(\text{ReLU}\left(\text{Conv}_{1 \times 1}^1(z)\right)\right)\right) \quad (8)$$

式中 σ_1 为Softmax函数,用于在通道维度上生成归一化的注意力权重,确保不同通道对最终响应的贡献具有可解释性和可区分性。

为了进一步在空间维度上引入显著性建模,本文对 u 进行空间扩展,分别采用并行的 1×1 卷积和 3×3 卷积获得两种感受野下的空间注意力图 $A_{1 \times 1}$ 和 $A_{3 \times 3}$,并进行拼接融合,表达式为

$$A = \text{Concat}(A_{1 \times 1}, A_{3 \times 3}) \quad (9)$$

式中: 1×1 卷积更关注局部像素级响应,而 3×3 卷积则在保留空间结构一致性的基础上捕获小范围邻域信息,两者结合既保证了细节捕获能力,又兼顾了空间上下文的平滑性。

最终,注意力图 A 作为权重作用于多尺度融合特征 F_{ms} ,得到经过通道-空间联合重标定的输出特征 F_{out} 为

$$F_{\text{out}} = \text{Conv}_{1 \times 1}(\text{Concat}(F_{\text{ms}}, F_{\text{ms}} \odot A)) \quad (10)$$

通过这种方式,网络在保持模块轻量化的前提下,实现了对关键人体区域的显式关注,能够在复杂遮挡背景下抑制无关区域的响应,提高行人目标的可分性和检测置信度。

1.2 A2C2f_SEAM 模块

在复杂口岸通行场景中,受闸机结构与人群密集度影响,行人常出现严重遮挡,仅身体局部可见,易引发特征对齐偏移、局部信息缺失及同类目标特征混叠问题,使模型难以区分真实行人与背景或遮挡物。YOLOv12中的A2C2f是Backbone与Neck间的重要特征融合单元,承担着中高层特征建模与信息传递的作用。然而,原始A2C2f主要依赖单一路径的注意力建模,在复杂遮挡场景下对关键人体区域的响应不够充分,容易受到背景和遮挡物干扰。为缓解遮挡导致的特征退化问题,本文提出一种改进的特征增强模块A2C2f_SEAM,结构如图4所示。将A2C2f内的ABlock替换为集成SEAM的ABlock_SEAM,并通过多层堆叠与残差连接增强遮挡行人的特征表达。

与原始ABlock相比,ABlock_SEAM的主要优势在于构建了一种双路径并行的注意力架构,实现了互补的特征捕获机制。其中,Attn路径通过区域划分建模空间区域级的全局注意力,有助于获取整体结构信息;SEAM路径则通过学习遮挡区域与未遮挡区域之间的关系,对特征图中的关注位置进行动态调整,使模型能够更加有效地感知被遮挡的行人信息。

SEAM的实现流程如图5所示,首先利用不同Patch尺度的通道与空间混合模块(Cross-scale multi-module, CSMM)并行提取多尺度上下文信息并汇聚得到 $F_s = \sum_{p \in \{6, 7, 8\}} \text{CSMM}_p(F)$;随后对 F_s 做全局平均池化并经轻量多层感知器(Multi-layer perceptron, MLP)生成通道权重,采用指数缩放得到 $w = \exp(\text{MLP}(\text{GAP}(F_s)))$;最后将权重与输入特征逐元素相乘得到输出 $F' = F \odot w$ 。其中,CSMM模块利用不同的特征块来提取多尺度特征,并通过深度可分离卷积来学习空间维度与通道之间的相关性。CSMM内部通过Patch embedding与带残差的深度可分离卷积进行空间-通道混合,在保持轻量化的同时强化跨尺度上下文与通道交互。通过上述机制,SEAM能够在遮挡背景下增强与行人相关的判别性响应,并利用上下文关联对遮挡导致的特征缺失进行补偿,从而抑制背景噪声干扰,提高特征判别性与检测稳定性。

1.3 Alpha-Shape IoU 损失函数

YOLOv12原始损失函数采用CIoU作为其边界框回归损失函数,从框的交并比、中心点的距离,以及长宽比的匹配度3个因素进行考量。但是在口岸快捷通道遮挡场景下,行人目标常呈现局部可见、形

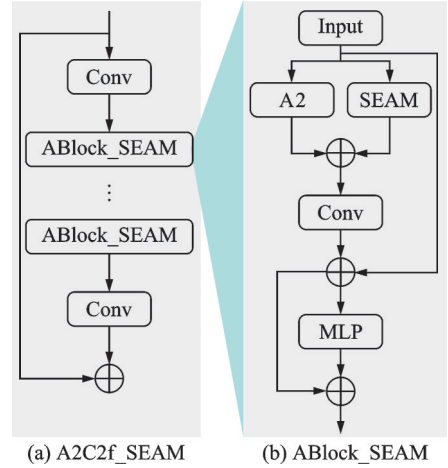


图4 A2C2f_SEAM模块结构图

Fig.4 A2C2f_SEAM module structure

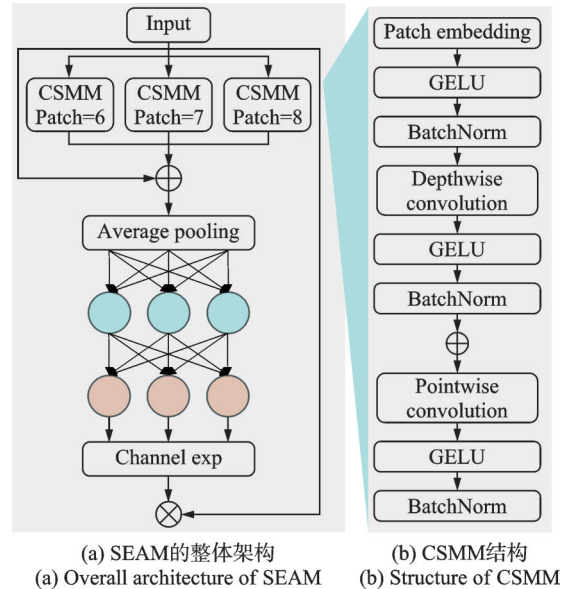


图5 SEAM 的示意图

Fig.5 Illustration of SEAM

态变化明显的特点,CIoU难以有效处理局部遮挡和目标形态差异,导致检测精度下降。

为增强模型对遮挡行人形状差异的感知能力,本文借鉴 Shape-IoU 中的形状约束思想,引入与目标形状相关的距离损失项和形状惩罚项。同时,考虑到 Shape-IoU 中的 IoU 重叠项仍采用线性形式,本文进一步结合 Alpha-IoU 的非线性加权思想,对 IoU 项进行幂次变换,构建 Alpha-Shape IoU 损失函数,并将其用于替代原有的 CIoU 边界框回归损失函数。为说明 Alpha-Shape IoU 的构建过程,首先给出 Shape-IoU 的损失计算公式为

$$L_{\text{Shape-IoU}} = 1 - \text{IoU} + \text{distance}^{\text{shape}} + 0.5 \times \Omega^{\text{shape}} \quad (11)$$

式中:IoU^[20]为传统的交并比,用于衡量预测框与真实框的重叠度;distance^{shape}表示预测框与真实框之间的形状匹配度,即距离损失函数; Ω^{shape} 为形状损失函数,用于加强对框形状的约束。distance^{shape}主要衡量中心位置的偏差,先计算宽、高的形状权重,用于让宽、高的偏差在距离计算中更贴合真实框的形状,再计算中心坐标偏差,乘以上面的形状权重,让位置偏差和形状绑定。

$$ww = \frac{2 \times (w^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (12)$$

$$hh = \frac{2 \times (h^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (13)$$

$$\text{distance}^{\text{shape}} = hh \times (x_c - x_c^{\text{gt}})^2 / c^2 + ww \times (y_c - y_c^{\text{gt}})^2 / c^2 \quad (14)$$

式中:ww和hh分别表示水平方向和垂直方向上的权重系数,其值与gt框的形状有关; w^{gt} 为真实框的宽; h^{gt} 为真实框的高;s为尺度缩放因子,范围为0.5~1.5; (x_c, y_c) 为预测框的中心坐标; $(x_c^{\text{gt}}, y_c^{\text{gt}})$ 为真实框的中心坐标;c为预测框与真实框最小外接矩形的对角线长度。

Ω^{shape} 是对宽、高的形状差异的直接惩罚,确保预测框的宽高更贴近真实框,表达式为

$$\Omega^{\text{shape}} = \sum_{t=w, h} (1 - e^{-w_t})^\theta \quad \theta = 4 \quad (15)$$

$$w_w = hh \times \frac{|w - w^{\text{gt}}|}{\max(w, w^{\text{gt}})} \quad (16)$$

$$w_h = ww \times \frac{|h - h^{\text{gt}}|}{\max(h, h^{\text{gt}})} \quad (17)$$

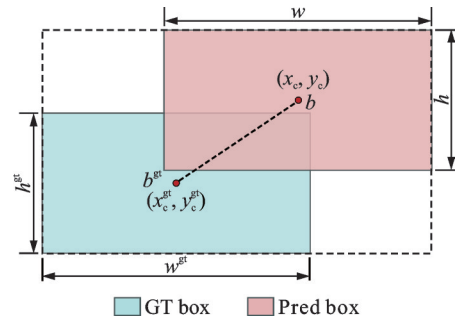
式中: Ω^{shape} 为形状惩罚项; w_w 与 w_h 为形状损失因子,用来衡量两框宽、高方向的相对误差,并进行加权,从而更好地优化模型对真实框形状的匹配; w, h 为预测框的宽、高。

然而,Shape-IoU中的IoU重叠项仍采用线性形式进行优化,在训练中后期对高质量预测框的进一步精细回归能力有限。为进一步优化边界框回归过程,本文在Shape-IoU框架下引入Alpha-IoU的幂次变换思想,对其中的IoU重叠项进行非线性加权,构建Alpha-Shape IoU损失函数,如图6所示。通过调整 α ,使得损失函数的加权更加灵活,从而优化梯度的平滑性,进一步改善边界框回归过程。

Alpha-IoU采用Box-Cox变换^[21],其变换表达式为

$$O(\alpha) = \begin{cases} \frac{O^\alpha - 1}{\alpha} & \alpha \neq 0 \\ \log O & \alpha = 0 \end{cases} \quad (18)$$

式中:O为连续变量,且要求取值为正; α 为指数调节参数,用于控制幂变换过程中的非线性程度,其取值决定变量映射形式。



注:GT box表示真实框;Pred box表示预测框。

图6 Alpha-Shape IoU示意图

Fig.6 Alpha-Shape IoU diagram

通过引入 Box-Cox 变换,对传统 IoU 损失函数 $\text{Loss}_{\text{IoU}} = 1 - \text{IoU}$ 进行推广,构建 Alpha-IoU 损失函数,其表达式为

$$\text{Loss}_{\alpha\text{-IoU}} = \frac{1 - \text{IoU}^\alpha}{\alpha} \quad \alpha > 0 \quad (19)$$

当 α 取固定正值时,式中的 $1/\alpha$ 可视为常数比例因子,该因子不会改变损失函数的优化目标,仅影响损失值和梯度幅值。为简化计算并保持与传统 IoU 损失形式的一致性,本文采用如下未归一化形式,即

$$\text{Loss}_{\alpha\text{-IoU}} = 1 - \text{IoU}^\alpha \quad (20)$$

在多个基准目标检测数据集和模型上实验证明^[14],Alpha-IoU 损失可以一致地超越现有的基于 IOU 的损失。基于此,本文在 Shape-IoU 损失函数的基础上,保留其距离损失项 $\text{distance}^{\text{shape}}$ 和形状惩罚项 Ω^{shape} ,仅将原有线性 IoU 项 $1 - \text{IoU}$ 替换为 Alpha-IoU 项 $1 - \text{IoU}^\alpha$,从而构建 Alpha-Shape IoU 损失函数,其定义为

$$\mathcal{L}_{\alpha\text{-Shape IoU}} = 1 - \text{IoU}^\alpha + \text{distance}^{\text{shape}} + 0.5\Omega^{\text{shape}} \quad (21)$$

2 实验与结果分析

2.1 实验环境及参数配置

为保证实验结果的公平性与可比性,所有模型均在相同的软硬件环境下进行训练与评测。实验硬件配置为 NVIDIA RTX 4090(24 GB)GPU 和 Intel Xeon Platinum 8352V 处理器;软件环境为 Ubuntu 22.04 操作系统,采用 Python 3.10 和 PyTorch 2.1.0 深度学习框架,CUDA 12.1 作为 GPU 并行计算平台。模型训练的主要参数设置如表 1 所示。所有对比模型均采用相同的数据划分和输入图像尺寸;为提高实验结果的可靠性,消融实验和不同损失函数对比实验均采用 3 个不同训练随机种子进行独立重复训练,并给出了各项指标的均值和标准差(Mean $\pm$ Std)。

推理及检测结果可视化阶段,对于采用标准非极大值抑制(Non-maximum suppression, NMS)后处理的模型,置信度阈值统一设置为 0.25, NMS 的 IoU 阈值设置为 0.7,单图最大检测目标数采用框架默认设置,且不采用测试时增强(Test-time augmentation, TTA)。对于具有特定推理或后处理机制的模型,则保持其原有机制,不额外施加标准 NMS。

2.2 实验数据集

为验证本文提出的改进模型在口岸快捷通道复杂背景与遮挡条件下的行人检测性能,采用 Crowd-Human^[22]公开数据集和自采快捷通道翻越行为数据集 CBCD,数据集示例如图 7 所示。

2.2.1 CrowdHuman 数据集

从 CrowdHuman 公开且具有完整标注的训练集中抽取 5 000 张图像作为实验数据,并按照 8:2 的比例划分为训练集和验证集,即训练集包含 4 000 张图像,验证集包含 1 000 张图像。所有模型均在相同的数据划分下进行训练与评测,模型对比实验、消融实验、损失函数对比实验以及不同遮挡程度下的专项评测均基于该验证集完成,本文中的指标均由该验证集统计得到。在数据预处理阶段,对于 Crowd-Human 数据集中标记为 Ignore 的区域,未进行专门的忽略区域建模,而是在标注格式转换过程中将其剔除。对于有效行人实例,仅采用其全身边界框作为模型训练和总体评测的检测标注;可见区域框不参与模型训练,仅用于后续计算人员目标可见比例并进行不同遮挡程度的分组。

表 1 模型参数设置

Table 1 Model parameter settings

参数	数值
输入图像	640 $\times$ 640
批次大小	16
训练次数	150
学习率	0.001
优化器	SGD
线程数	12
预训练权重	不加载

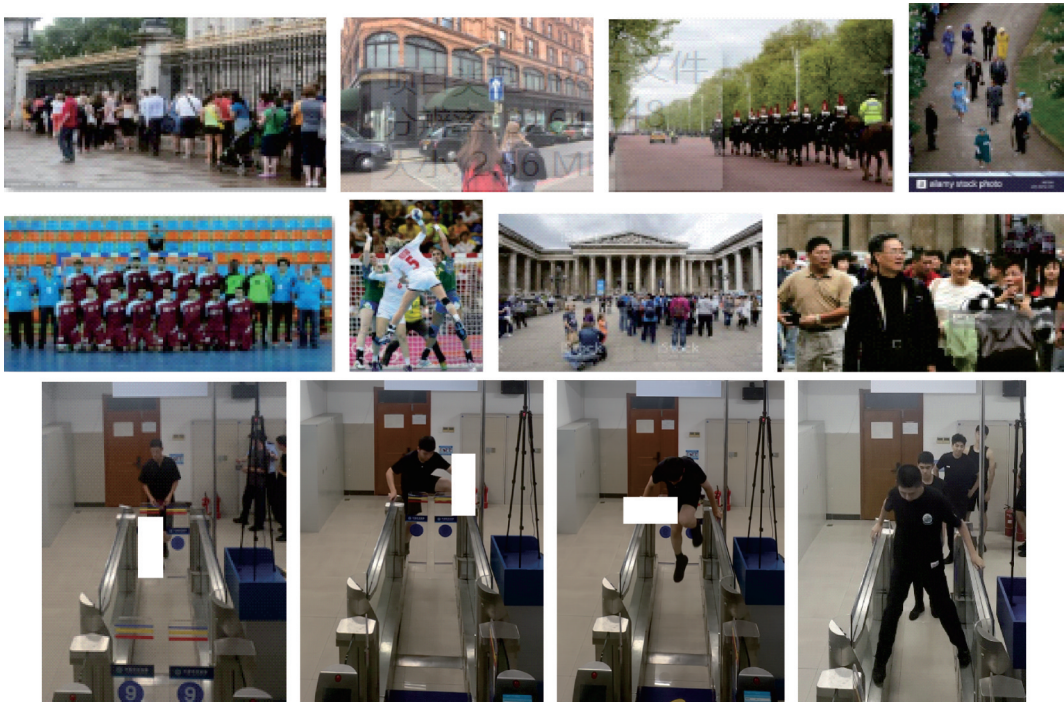


图7 数据集示例

Fig.7 Dataset examples

2.2.2 CBCD数据集

由于真实口岸数据涉及敏感信息而难以获取,因此依托中国人民警察大学口岸通关模拟实验室构建了快捷通道翻越行为数据集CBCD,实验场景为口岸快捷通道模拟环境。由150名志愿者循环模拟正常通关及不同姿势的翻越行为,摄像机安装位置于通道出口上方,以俯视视角对通行区域进行监控,俯角为30°,距地面高度为2 m,与快捷通道出口水平距离为1.2 m。最终采集的原始数据为视频数据,内容包括正常通行行为以及翻越通道行为,同时场景中包含由于人员排队造成的行人遮挡以及通道设备结构带来的局部遮挡情况。

在数据处理阶段,首先从采集的原始视频中按照视频序列抽取有效图像帧,并对模糊及与研究任务无关的图像进行清洗。因研究任务主要关注人员在不同姿态、遮挡位置及翻越方式下的目标特征,同一志愿者在不同独立视频中呈现的姿态、空间位置和遮挡状态存在差异,可形成不同的目标检测样本。因此,本文以完整原始视频序列为基本单位进行互斥划分,确保同一视频及其相邻帧不跨训练集、验证集和测试集,以降低视频抽帧带来的直接数据泄露风险。随后,使用X-AnyLabeling对图像中的人员目标进行人工标注,采用边界框对通关人员的完整身体区域进行标注,类别设置为Person,并对标注结果进行人工复核。模型训练过程中,仅对训练集实施随机遮挡增强,以模拟人员相互遮挡和闸机设备遮挡,验证集和测试集不进行随机遮挡增强。

最终共获得10 000张图像构成CBCD数据集,并按照7:2:1的比例划分为训练集、验证集和测试集。相较于通用行人检测数据集,CBCD数据集在场景布局和设备配置方面更加贴近真实口岸快捷通道场景,具有视角固定、遮挡频繁和行为特征明显的特点,能够较好地反映口岸通关环境下行人的实际检测难点。此外,本研究中行人检测任务为翻越行为识别的前置步骤,因此数据采集中包含翻越行为实例,为后续翻越行为识别实验提供了基础数据支持。

2.3 评价指标

为客观评估改进后的模型的性能,本文采用精确率 P 、召回率 R 、IOU阈值为0.5时计算得到的平均精度均值(mean Average precision, mAP)(mAP@0.5)、IOU阈值为0.5~0.95(以0.05为间隔)时计算得到的平均精度均值(mAP@0.5:0.95)以及参数量(Params)作为综合评价指标。精确率指模型预测为正样本中实际为正样本的比例,召回率指实际为正样本中模型预测为正样本的比例,表达式分别为

$$P = \frac{TP}{TP + FP} \quad (22)$$

$$R = \frac{TP}{TP + FN} \quad (23)$$

式中:TP为真阳性,表示被模型预测为正类的正样本;FP为假阳性,表示被模型预测为正类的负样本;FN为假阴性,表示被模型预测为负类的正样本。

PR曲线所围面积为平均精度(Average precision, AP),多个类别物体检测的AP平均后得到的数值为平均精度均值,综合衡量模型目标检测的性能,计算公式为

$$AP = \int_0^1 P(R) dR \quad (24)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (25)$$

式中:PR曲线所围面积为平均精度AP,多个类别的AP取平均得到mAP。CrowdHuman仅包含Person一个检测类别,因此其AP与mAP数值相同,为保持全文指标名称一致,本文统一采用mAP表示。所有模型均采用统一的验证程序计算 P 、 R 、mAP@0.5和mAP@0.5:0.95等指标。同时引入Params,指模型中所有可训练参数的总量。较小的参数量有助于降低模型存储与参数内存开销,是衡量模型复杂度的重要指标之一。

2.4 消融实验

为了验证本文所提出各改进模块在遮挡目标检测任务中的有效性,基于YOLOv12构建了一系列消融实验,逐步引入A、B、C模块。A表示引入AMFE模块;B为添加的A2C2f_SEAM改进;C表示改进损失函数Alpha-Shape IoU,并在相同训练配置与数据集条件下进行对比分析,实验结果如表2所示。从表2中可以看出,基线模型在未引入任何改进时,其mAP@0.5和mAP@0.5:0.95分别为73.8%和41.8%,在复杂遮挡场景下对行人目标的检测能力仍存在提升空间。当仅引入AMFE模块后,mAP@0.5提升至77.3%,mAP@0.5:0.95提升至45.2%。这表明多尺度卷积与注意力机制的结合能够有效增强网络对局部可见区域和上下文信息的建模能力,提高对遮挡行人的检测性能。进一步引入A2C2f_SEAM模块后,mAP@0.5提升至78.7%,说明在中高层特征中引入SEAM的遮挡关联建模,有助于缓解遮挡导致的特征缺失与目标混淆,减少漏检。当3种改进同时引入,即AMFE +

表2 消融实验

Table 2 Ablation experiments

Improved				$P/\%$	$R/\%$	mAP@0.5/%	mAP@0.5:0.95/%	Params/ 10^6
YOLOv12	A	B	C					
✓				81.8 ± 0.16	61.2 ± 0.29	73.8 ± 0.20	41.8 ± 0.17	2.51
✓	✓			82.9 ± 0.22	65.4 ± 0.25	77.3 ± 0.28	45.2 ± 0.28	2.28
✓		✓		80.8 ± 0.20	63.0 ± 0.18	75.1 ± 0.16	43.4 ± 0.19	2.50
✓			✓	79.5 ± 0.25	62.4 ± 0.18	74.9 ± 0.21	42.6 ± 0.20	2.51
✓	✓	✓		82.4 ± 0.15	68.3 ± 0.17	78.7 ± 0.22	45.6 ± 0.19	2.28
✓	✓	✓	✓	83.2 ± 0.20	68.4 ± 0.15	79.2 ± 0.23	46.3 ± 0.11	2.28

A2C2f_SEAM + Alpha-Shape IoU时,模型性能达到最优,mAP@0.5以及mAP@0.5:0.95分别提升至79.2%和46.3%。其中,Alpha-Shape IoU的引入进一步优化了遮挡条件下的边界框回归质量,使预测框在位置与形状上更加贴合真实行人目标,从而提升整体检测精度。

图8为MSS-YOLO在CrowdHuman上进行训练与验证过程的损失曲线及评价指标的变化趋势,结果表明模型具有较好的训练稳定性和泛化表现。各次实验均采用相同的数据划分和训练配置,仅改变随机种子。3类训练损失与验证损失均在训练初期快速下降,说明模型能够较快完成对目标位置、类别语义以及边界分布的初步建模;随后各项损失进入缓慢下降阶段,反映出网络已由粗粒度拟合转入细粒度参数调整,模型并未在早期陷入次优解,而是在持续学习更稳定、更具判别性的特征表示。训练损失与验证损失变化趋势基本一致,二者间隔始终较小,未观察到明显的过拟合趋势,验证端未出现明显反弹,说明模型在拟合训练数据的同时保持了较好的泛化能力。从检测性能看,4项评价指标整体持续提升并在后期保持稳定,其中mAP@0.5:0.95的稳步提高说明模型在较严格IoU阈值下仍具备较好的边界框回归能力。

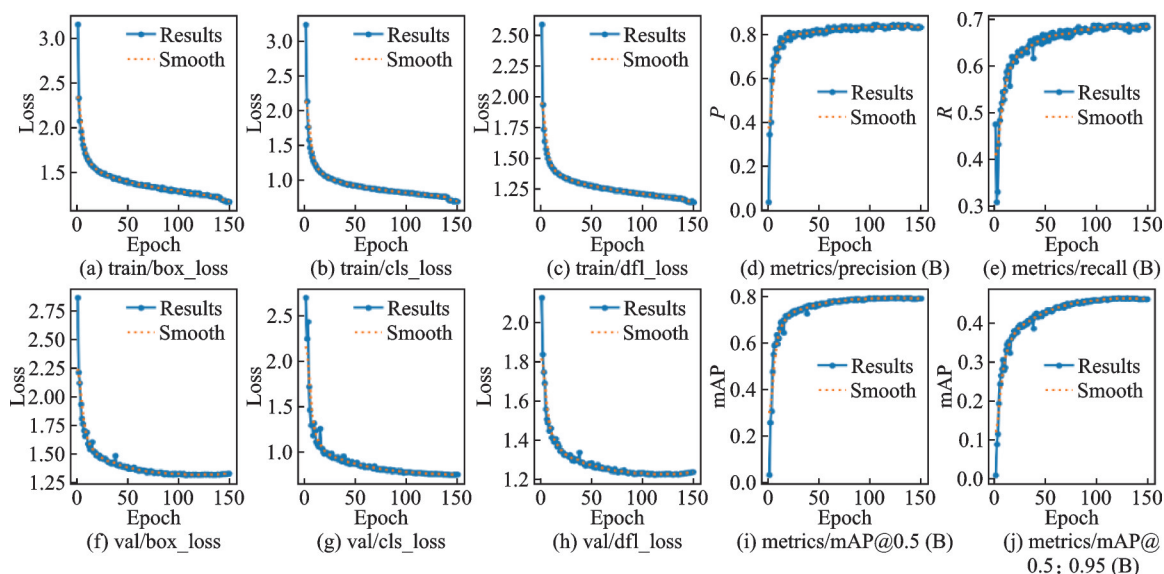


图8 MSS-YOLO在CrowdHuman上进行训练与验证损失曲线及评价指标变化趋势

Fig.8 Loss curves and performance metrics trends of MSS-YOLO during training and validation on CrowdHuman

2.5 对比实验

2.5.1 不同损失函数对比实验结果

为验证 Alpha-Shape IoU 损失函数对模型训练过程的影响,在保持网络结构与训练参数一致的条件下,分别采用 IoU、CIoU、WIoU、Shape IoU 及 Alpha-Shape IoU 在同一数据集下进行对比实验。图9为不同损失函数在训练过程中的收敛曲线。在训练初期均呈现快速下降趋势,但在收敛速度与稳定性方面存在差异。

不同损失函数的具体指标差异如表3所示。Alpha-Shape IoU 在 Shape IoU 的位置与形状联合约束基础上,引入 Alpha-IoU 的幂次变换思想加速训练时的收敛速度。与原始

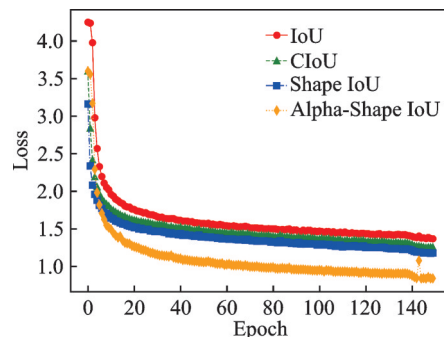


图9 不同损失函数对比

Fig.9 Comparison of different loss functions

损失函数 CIoU 相比, Alpha-Shape IoU 在 mAP@0.5 指标上提升 0.5 个百分点, 同时在 Recall 指标上提升 0.1 个百分点, 说明该方法在减少遮挡目标漏检方面具有一定优势。

表 3 不同损失函数对比实验

Table 3 Comparison experiment of different loss functions			%
Model	P	R	mAP@0.5
IoU	81.8 ± 0.15	67.4 ± 0.18	78.8 ± 0.14
CIoU	82.4 ± 0.15	68.3 ± 0.17	78.7 ± 0.22
WIoU ^[23]	81.0 ± 0.16	67.7 ± 0.20	78.0 ± 0.20
Shape IoU	82.3 ± 0.14	68.1 ± 0.14	79.0 ± 0.17
Alpha-Shape IoU	83.2 ± 0.20	68.4 ± 0.15	79.2 ± 0.23

2.5.2 不同模型对比实验结果

为全面验证所提出的 MSS-YOLO 模型对遮挡目标检测的性能及模型复杂度, 在 CrowdHuman 数据集上与多种主流目标检测模型进行了对比实验, 不同模型对比实验采用固定训练设置下的单次结果, 结果如表 4 所示。实验结果显示, 在检测精度方面, MSS-YOLO 得到了明显提升, 其 mAP@0.5 提升了 5.4 个百分点, mAP@0.5:0.95 提升了 4.5 个百分点; 指标 R 从 61.5% 大幅提升至 68.6% (提升了 7.1 个百分点)。所提出的多尺度特征提取、遮挡特征增强与边界框回归优化策略能够协同改善遮挡人员的检测性能。虽然浮点运算次数 (FLOPs) 由 6.0×10^9 增加至 7.7×10^9 , 但模型仍保持较小的参数规模, 并在检测性能与计算复杂度之间取得一定平衡。

表 4 不同模型的对比实验

Table 4 Comparison experiment of different models						
Model	$P/\%$	$R/\%$	mAP@0.5/%	mAP@0.5:0.95/%	Params/ 10^6	FLOPs/ 10^9
YOLOv5n ^[24]	80.2	60.3	72.0	40.3	2.50	7.1
YOLOv6n ^[25]	81.6	62.3	73.7	41.9	4.23	11.8
YOLOv8n ^[26]	81.2	62.6	74.1	42.0	3.00	8.1
YOLOv9t ^[27]	80.9	62.0	73.9	42.0	1.97	7.6
YOLOv10n ^[28]	77.5	59.0	70.8	39.6	2.69	8.2
RT-DETR-R34 ^[29]	71.0	48.5	70.1	37.5	31.5	41.7
FPN + Adaptive NMS ^[30]	—	65.2	76.5	—	—	—
YOLO11n ^[31]	79.6	60.8	73.0	41.0	2.58	6.3
YOLOv12n	82.0	61.5	74.0	41.9	2.51	6.0
MSS-YOLO	83.4	68.6	79.4	46.4	2.28	7.7

与其他主流目标检测模型进行综合对比, 可以进一步评价 MSS-YOLO 的检测性能。由表 4 可知, MSS-YOLO 在 CrowdHuman 实验数据上的整体检测性能优于 YOLOv5n、YOLOv6n、YOLOv8n、YOLOv9t、YOLOv10n 和 YOLO11n 等轻量化模型, 其 mAP@0.5 达到 79.4%, 较 YOLO 系列中性能较优的 YOLOv8n 和 YOLOv9t 分别提升 5.3 和 5.5 个百分点; R 达到 68.6%, 较上述 YOLO 系列模型至少提升 6.0 个百分点, 表明所提出的多尺度特征提取与遮挡特征增强策略能够在一定程度上缓解遮挡造成的有效特征缺失, 提高人员目标的检出能力。与不同检测框架相比, MSS-YOLO 同样表现出较好的检测性能, 其中 RT-DETR-R34 的 mAP@0.5 为 70.1%, 低于 MSS-YOLO。进一步与针对拥挤场景目标检测的 Adaptive NMS 方法相比, Recall 和 mAP@0.5 分别提高 3.4 个百分点和 2.9 个百分点, 说明本文模型在密集遮挡人员检测任务中具有较好的目标检测能力。

图 10 更直观地对比了精确度、召回率以及 mAP@0.5 核心指标在 MSS-YOLO 与基线模型的变化趋势,由结果可以看出改进模型的整体性能优于基线模型。

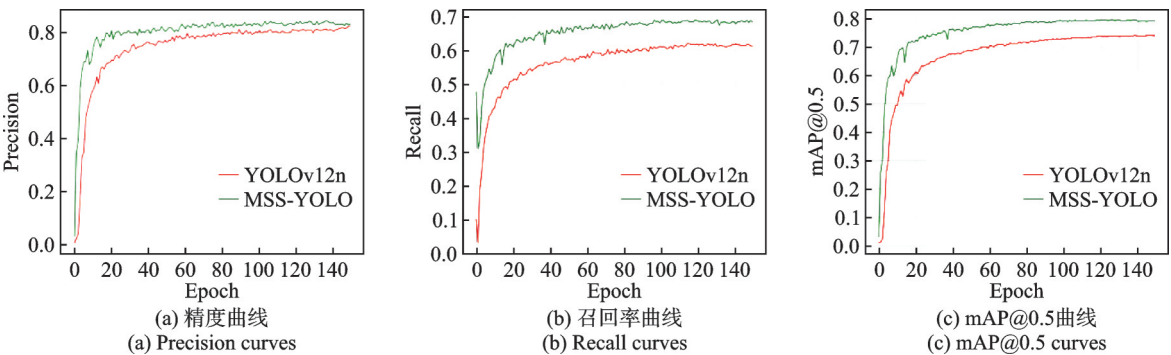


图 10 MSS-YOLO 与 YOLOv12n 的评价指标变化趋势对比

Fig.10 Comparison of the evaluation metric trends between MSS-YOLO and YOLOv12n

2.5.3 不同遮挡程度下的检测性能分析

为进一步验证 MSS-YOLO 在不同遮挡条件下的人员检测能力,基于本文构建的 CrowdHuman 验证集开展遮挡分组实验。CrowdHuman 数据集中每个人员实例均标注头部框(hbox)、可见区域框(vbox)和完整人体框(fbox)。本文采用 fbox 作为模型训练及总体评测的真实框,并利用 vbox 与 fbox 计算人员目标的可见比例;hbox 不参与本文实验。对于每个人员实例,其可见比例定义为

$$VR = \frac{\text{Area}(\text{vbox})}{\text{Area}(\text{fbox})} \quad (26)$$

式中:VR 表示人员目标的可见比例,Area(vbox)表示可见人体区域面积,Area(fbox)表示完整人体区域面积。VR 越小,说明人员受到的遮挡程度越严重。本文以 $VR = 0.5$ 为阈值,将验证集中的人员实例划分为两组:当 $VR > 0.5$ 时,将目标划分为低遮挡组;当 $VR < 0.5$ 时,将目标划分为高遮挡组。该划分仅用于验证阶段的遮挡专项评测,不改变模型的训练数据及训练方式。

不同遮挡程度下的检测性能如表 5 所示。由表 5 可知,随着人员目标可见比例降低,YOLOv12n 和 MSS-YOLO 的检测性能均出现一定程度下降,表明遮挡会减少人体有效判别信息,增加目标定位与识别的难度。随着遮挡程度加重,MSS-YOLO 相较于于基线模型的性能增益进一步扩大。尤其是在高遮挡条件下,MSS-YOLO 的 AP@0.5 和 R 分别较 YOLOv12n 提升 6.8 和 7.5 个百分点,说明所提出模型能够在人体可见信息减少的情况下提取更加有效的目标特征,减弱遮挡造成的特征缺失对检测性能的影响,对复杂遮挡场景具有较好的适应能力。

表 5 不同遮挡程度下的检测性能分析

Table 5 Detection performance analysis under different occlusion levels

Model	低遮挡	低遮挡	高遮挡	高遮挡
	AP@0.5	R	AP@0.5	R
YOLOv12n	80.8	68.5	67.5	52.9
MSS-YOLO	83.4	74.3	74.3	60.4

2.6 自建场景数据集验证

为进一步验证 MSS-YOLO 在口岸快捷通道场景中的适用性,本文在自建 CBCD 训练集上开展实验。YOLOv12n 和 MSS-YOLO 均采用 CBCD 训练集进行重新训练,在验证集上完成模型选择,并在测试集上进行最终性能评测。CBCD 包含正常通行、翻越闸机以及人员和闸机设备造成的不同程度遮挡,能够反映口岸快捷通道场景下的实际检测难点。为保证实验结果的公平性,两种模型采用相同的数据划分、硬件环境、训练参数和推理设置。结果如表 6 所示,MSS-YOLO 在所有评估指标上均优于基线模型,P 值和 R 值分别提升了 6.1% 和 4.3%,说明改进模型对口岸快捷通道存在遮挡的通关人员检出能力

得到改善,能够在一定程度上减少遮挡条件下的漏检现象。同时,能够综合衡量模型性能的mAP@0.5和mAP@0.5:0.95分别提升了5.0%和12.7%,进一步验证了其在遮挡场景下的鲁棒性与准确性。

表 6 自建CBCD数据集验证结果

Table 6 Validation results on the self-built CBCD dataset

%

Model	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv12n	81.5	85.6	90.6	56.1
MSS-YOLO	87.6	89.9	95.6	68.8

为了更直观地观察改进算法在检测效果上的提升,在自采数据集上利用Grad-CAMplus绘制检测能力热力图。图11为热力图对比分析结果,可以清晰地展示MSS-YOLO在遮挡条件下更好地聚焦于真实目标区域,模型关注区域更集中,减少环境干扰。该结果从特征响应角度进一步表明,MSS-YOLO在复杂遮挡环境下能够更加集中地关注与人员目标相关的区域。

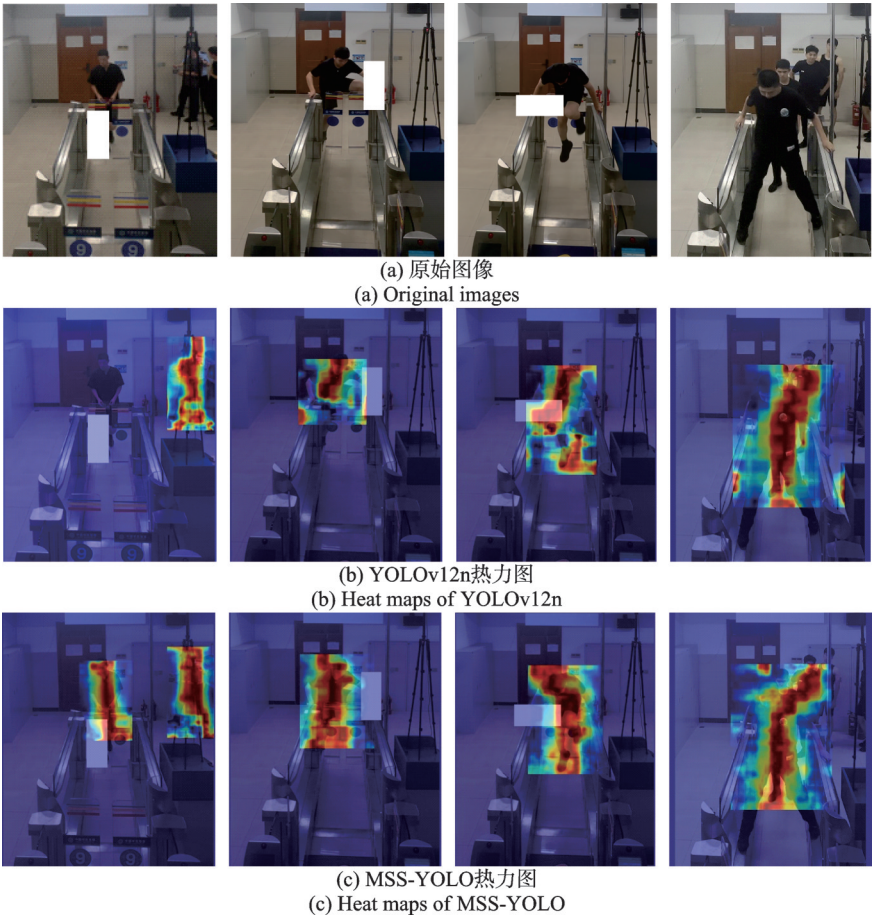


图 11 YOLOv12n 和 MSS-YOLO 热力图比较

Fig.11 Comparison of heat maps between YOLOv12n and MSS-YOLO

3 结束语

为了解决口岸快捷通道中遮挡导致的目标检测精度下降问题,本文基于YOLOv12提出了一种改

进的行人检测算法 MSS-YOLO,采用 CrowdHuman 验证集评价模型在密集遮挡场景下的检测性能,并在自建 CBCD 数据集上分别重新训练和测试 YOLOv12n 与 MSS-YOLO,以验证所提模型在口岸快捷通道场景中的适用性。本文算法分别从特征提取、特征增强,以及提高检测框定位的准确性 3 方面进行以下关键改进:设计了 AMFE 模块,通过多卷积并行有效捕获遮挡目标的特征,并自适应加权关注关键人体特征区域;在 A2C2f 中引入 SEAM 注意力机制,构建了双路径并行注意力架构,实现了互补的特征捕获能力;改进的 Alpha-Shape IoU 损失函数改善了模型在复杂遮挡场景下的检测精度与稳定性。实验结果表明,模型改进后性能表现出色,明显优于基线模型。

在口岸智能安防系统中,通关人员行为检测与异常识别高度依赖于前端目标检测结果的准确性与稳定性。行人检测作为行为分析与风险识别的前置基础任务,其在遮挡、密集与尺度变化条件下的性能表现直接影响后续行为特征提取与时序建模的可靠性。考虑到口岸智能安防系统对实时性与部署成本的严格要求,后续将结合具体边缘硬件平台,进一步开展模型推理延迟、运行帧率和内存占用等部署性能测试,进一步提升系统在实际应用环境中的稳定性与适应性,为构建完整、高效的通关人员智能行为分析系统提供支撑。

参考文献:

- [1] 海关总署,国家发展改革委,工业和信息化部,等. 2024 关于智慧口岸建设的指导意见[EB/OL]. (2024-09-10). https://www.gov.cn/zhengce/zhengceku/202409/content_6975795.htm.
- [2] 徐新智,何宏. 基于局部通道编码的轻量化人体姿态估计算法[J]. 数据采集与处理, 2025, 40(6): 1625-1636.
XU Xinzhi, HE Hong. Lightweight human pose estimation algorithm based on partial channel encoding[J]. Journal of Data Acquisition and Processing, 2025, 40(6): 1625-1636.
- [3] HUANG Y, LI C, ZHANG H, et al. Track any anomalous object: A granular video anomaly detection pipeline[C]// Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2025: 8689-8699.
- [4] 叶枫,董凡可,贾川民. 端到端智能视频压缩技术及其在无人机中的应用[J]. 数据采集与处理, 2025, 40(2): 303-319.
YE Feng, DONG Fanke, JIA Chuanmin. End-to-end video compression technology and its application in unmanned aerial vehicles[J]. Journal of Data Acquisition and Processing, 2025, 40(2): 303-319.
- [5] CHEN N, LI M, YUAN H, et al. Survey of pedestrian detection with occlusion[J]. Complex & Intelligent Systems, 2021, 7(1): 577-587.
- [6] ZHANG T, YE Q, ZHANG B, et al. Feature calibration network for occluded pedestrian detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(5): 4151-4163.
- [7] 刘毅,于畅洋,李国燕,等. UAST-RCNN: 遮挡行人的目标检测算法[J]. 电子测量与仪器学报, 2022, 36(12): 168-175.
LIU Yi, YU Changyang, LI Guoyan, et al. UAST-RCNN: Object detection algorithm for blocking pedestrians[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(12): 168-175.
- [8] 王家辉,王洪玉,郝应光. 联合跟踪的特征聚合: 遮挡场景下的视频目标检测任务[J]. 计算机科学, 2026, 53(8): 165-173.
WANG Jiahui, WANG Hongyu, HAO Yingguang. Feature aggregation of joint tracking: Video target detection task in occlusion scene[J]. Computer Science, 2026, 53(8): 165-173.
- [9] NIKOUEI M, BAROUTIAN B, NABAVI S, et al. Small object detection: A comprehensive survey on challenges, techniques and real-world applications[J]. Intelligent Systems with Applications, 2025, 27: 200561.
- [10] 李承烨,张震,梁哲恒,等. 目标检测模型综述[J]. 计算机研究与发展, 2025, 62(12): 2895-2928.
LI Chengye, ZHANG Zhen, LIANG Zheheng, et al. Survey on object detection models[J]. Journal of Computer Research and Development, 2025, 62(12): 2895-2928.
- [11] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[EB/OL]. (2025-02-18). <https://doi.org/10.48550/arXiv.2502.12524>.
- [12] YU Z, HUANG H, CHEN W, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.
- [13] ZHANG H, ZHANG S. Shape-IoU: More accurate metric considering bounding box shape and scale[EB/OL]. (2024-01-12).

<https://doi.org/10.48550/arXiv.2312.17663>.

- [14] HE J B, ERFANI S, MA X J, et al. Alpha-IoU: A family of power intersection over union losses for bounding box regression[EB/OL]. (2022-1-22). <https://doi.org/10.48550/arXiv.2110.13675>.
- [15] LI Y, KAISER L, BENGIO S, et al. Area attention[EB/OL]. (2020-05-07). <https://doi.org/10.48550/arXiv.1810.10126>.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [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*, 2022, 52(8): 8574-8586.
- [18] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 1800-1807.
- [19] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2016-04-30). <https://doi.org/10.48550/arXiv.1511.07122>.
- [20] YU J, JIANG Y, WANG Z, et al. UnitBox: An advanced object detection network[C]//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, The Netherlands: ACM, 2016: 516-520.
- [21] BOX G E P, COX D R. An analysis of transformations[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1964, 26(2): 211-243.
- [22] SHAO S, ZHAO Z, LI B, et al. CrowdHuman: A benchmark for detecting human in a crowd[EB/OL]. (2018-04-30). <https://doi.org/10.48550/arXiv.1805.00123>.
- [23] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-04-08). <https://doi.org/10.48550/arXiv.2301.10051>.
- [24] JOCHER G, STOKEN A, BOROVEC J, et al. ultralytics/yolov5:v3.1[CP/OL]. (2020-10-29). <https://zenodo.org/records/4154370>.
- [25] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07). <https://doi.org/10.48550/arXiv.2209.02976>.
- [26] VARGHESE R M S. YOLOv8: A novel object detection algorithm with enhanced performance and robustness[C]//Proceedings of 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India: IEEE, 2024: 1-6.
- [27] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [C]//Proceedings of Computer Vision—ECCV 2024. Cham, Switzerland: Springer Nature, 2025: 1-21.
- [28] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[C]//Proceedings of Advances in Neural Information Processing Systems 37. Vancouver, BC, Canada: NeurIPS, 2024: 107984-108011.
- [29] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 16965-16974.
- [30] LIU S, HUANG D, WANG Y. Adaptive NMS: Refining pedestrian detection in a crowd[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 6452-6461.
- [31] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. (2024-10-23). <https://doi.org/10.48550/arXiv.2410.17725>.

作者简介:



陈泊睿(1999-),男,硕士研究生,研究方向:口岸安全管控技术等,E-mail: 2024903046@cpgu.edu.cn。



张梅(1983-),通信作者,女,副教授,硕士生导师,研究方向:信号与系统、口岸风险行为识别与预警,E-mail: 15501000805@163.com。



李昕蕊(2001-),女,硕士研究生,研究方向:口岸安全管控技术等。

(编辑:张黄群)

MSS-YOLO: Occluded Object Detection in Customs Scenarios

CHEN Borui¹, ZHANG Mei^{2*}, LI Xinrui¹

(1. Graduate School, China People's Police University, Langfang 065000, China; 2. School of Immigration Management, China People's Police University, Langfang 065000, China)

Abstract: Object detection is a prerequisite task for smart ports to realize behavior recognition of passengers. To address the problems that occlusion caused by crowded queues and gate equipment in fast-track channels leads to decreased detection accuracy of existing models, while also considering the requirements of edge devices, an improved YOLOv12n-based occluded object detection algorithm, MSS-YOLO, is proposed. First, an adaptive multi-scale feature extraction (AMFE) module is designed in the backbone network to replace the downsampling module, enlarge the receptive field of the model, and adaptively weight and focus on key human body feature regions, thereby enhancing multi-scale feature extraction capability and alleviating feature information loss and occlusion interference. Second, an improved feature enhancement module, A2C2f_SEAM, is introduced into the neck network to learn the relationship between occluded and non-occluded regions and dynamically adjust the focus of attention. Finally, CIoU is replaced with Alpha-Shape IoU, and power transformation and geometric constraints are introduced to further optimize bounding box regression performance. Experiments are conducted on a CrowdHuman experimental subset and the self-collected port fast-track channel crossing behavior dataset CBCD. The results show that, compared with the baseline model, the proposed MSS-YOLO improves precision by 1.4 and 6.1 percentage points, recall by 7.1 and 4.3 percentage points, mAP@0.5 by 5.4 and 5.0 percentage points, and mAP@0.5: 0.95 by 4.5 and 12.7 percentage points on the two datasets, respectively. Meanwhile, the number of model parameters is reduced by 0.23×10^6 . MSS-YOLO improves the detection accuracy of occluded persons in port clearance scenarios while maintaining a relatively small parameter scale, providing a foundation for subsequent edge deployment and optimization.

Highlights:

1. A lightweight occluded object detection algorithm MSS-YOLO is proposed for personnel detection in smart port express channel scenarios.
2. An AMFE module is designed to expand receptive field and enhance multi-scale feature extraction under occlusion interference.
3. The improved A2C2f_SEAM module is introduced to capture the relationship between occluded and non-occluded regions and optimize feature representation.
4. MSS-YOLO achieves significant accuracy improvement with fewer parameters, satisfying edge device deployment in smart ports.

Key words: smart port; occluded object detection; YOLOv12n; multi-scale feature extraction

Foundation item: National Key Research and Development Program of China (No.2023YFC3321602).

Received: 2026-02-06; **Revised:** 2026-08-20

***Corresponding author, E-mail:** 15501000805@163.com.