

文章编号 1004-924X(2025)14-2278-13

基于头部与整体信息联合的行人检测算法

马晞茗¹, 李 宁^{1*}, 吴 迪¹, 刘一豆², 于祥跃¹, 李 峥¹

(1. 中国科学院 长春光学精密机械与物理研究所 光电对抗部, 吉林 长春 130033;

2. 北京邮电大学 网络空间安全学院, 北京 100876)

摘要:在密集场景中,行人躯体受遮挡和行人尺度不一等因素导致检测器的准确率出现了下降。而行人头部受遮挡程度往往较轻,因此可用于辅助检测。对此,提出一种基于头部与整体信息联合的行人检测算法。首先,设计了一种基于稠密连接与增强融合的特征提取网络,用于强化多尺度特征提取,提升网络对多尺度行人的检测灵敏度。其次,优化区域建议网络模块的采样机制,提出一种基于遮挡重叠率判别的非均匀困难样本挖掘策略,重点关注受遮挡严重的困难样本,提升网络对遮挡情况的适应能力。然后,构建了头部与整体联合检测策略,并对后处理环节进行优化,利用头部检测结果恢复因受遮挡而被错误抑制的整体检测框,从而降低漏检率。同时,结合联合检测框架的特点对损失函数进行优化,进一步缓解了因遮挡造成的误检和漏检情况。最后,通过实验对所提算法的有效性进行了验证。结果表明,所提算法在遮挡程度较高的 CrowdHuman 数据集上对数平均漏检率降低了 5.7%,平均精度提升了 4%,在 TJU-DHD-pedestrian 数据集的两个小尺度子集上对数平均漏检率分别降低了 2.4% 和 2.1%,有效地提升了对遮挡行人目标和多尺度行人目标的检测能力。

关键词:行人检测;联合检测;多尺度特征融合;难例挖掘;后处理优化

中图分类号:TP391.41;TN911.73 **文献标识码:**A

doi:10.37188/OPE.20253314.2278 **CSTR:**32169.14.OPE.20253314.2278

Pedestrian detection algorithm based on joint head and overall information

MA Ximing¹, LI Ning^{1*}, WU Di¹, LIU Yidou², YU Xiangyue¹, LI Zheng¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,

Optical Countermeasures Department, Changchun 130033, China;

2. Beijing University of Posts and Telecommunications, School of Cyber Security,

Beijing 100876, China)

* Corresponding author, E-mail: lining@ciomp.ac.cn

Abstract: In crowded scenes, factors such as occlusion of the pedestrian body and varying pedestrian scales lead to a decrease in the precision of the detector. Since, the pedestrian head tends to be lightly occluded and it can be used to assist detection. In this regard, a pedestrian-detection algorithm based on joint head and overall information was proposed. First, a feature-extraction network built upon dense connections and enhanced fusion was designed to strengthen multi-scale feature extraction and to improve the net-

收稿日期:2025-03-21;修订日期:2025-05-18.

基金项目:国家自然科学基金(No. 61705219);中国科学院青年创新促进会会员(No. 2022216)

work's sensitivity to pedestrians of various scales. Second, the sampling mechanism of the region-proposal-network module was optimized by introducing a non-uniform hard-example-mining strategy that discriminated according to the occlusion-overlap rate; this strategy concentrated on heavily occluded hard samples and enhanced the network's adaptability to occlusion. Next, a joint head-and-body detection strategy was constructed, and the post-processing stage was refined so that head-detection results could recover body boxes that had been erroneously suppressed by occlusion, thereby reducing the missed-detection rate. Meanwhile, the loss function was further optimized by incorporating the characteristics of the joint detection box, so that mis-detections and missed detections caused by occlusion were alleviated. Finally, the effectiveness of the proposed algorithm was verified through experiments. The results show that the proposed algorithm reduces the log-average miss detection rate by 5.7% and improves the average precision by 4% on the CrowdHuman dataset with a higher degree of occlusion, and reduces the log-average miss detection rate by 2.4% and 2.1% on the two small-scale subsets of the TJU-DHD-pedestrian dataset, which effectively enhances the detection capability of both occluded pedestrian targets and multi-scale pedestrian targets.

Key words: pedestrian detection; joint detection; multi-scale feature fusion; hard example mining; post-processing optimization

1 引 言

行人检测是计算机视觉领域中的热门研究方向,在自动驾驶^[1]、视频监控^[2]和智能机器人^[3]等领域具有重要的应用价值。然而在实际检测场景中,行人目标受到的遮挡往往较为严重,具体表现在行人受背景遮挡和行人之间的相互遮挡。目标可见比例下降导致模型训练时提供的信息不足,并且基于随机抽样的方式进行训练会导致检测器难以充分学习遮挡样本等困难样例,进而影响行人检测准确率。此外,场景内不同视角处的行人目标尺度差异较大,远距离行人目标的成像像素和纹理信息均较为匮乏,进而导致检测难度增大。因此,研究复杂人群密集场景中遮挡行人目标和多尺度行人目标检测极具挑战性,对于提升检测器的检测性能具有重要意义。

近年来,基于深度学习的目标检测算法取得了显著进展,相比于传统算法取得了大幅度的性能提升,较为常见的包括以 Faster R-CNN 算法^[4]为代表的两阶段检测算法和以 YOLO 系列算法^[5]为代表的单阶段检测算法。相比于单阶段检测算法,两阶段检测算法的检测精度普遍更高,但检测速度则相对较慢。然而这些通用的目标检测算法均难以高效应对拥挤场景中的行人检测任务,具体表现在遮挡行人和小尺度行人的检

测精度下降和漏检率升高。

针对较远视角处小尺度行人目标漏检率较高的问题, Lin 等人^[6]提出了特征金字塔网络 (Feature Pyramid Network, FPN), 该方法成为多尺度目标检测的重要手段。陈勇等人^[7]提出了一种基于特征增强模块的小尺度行人检测算法,通过对小尺度行人目标特征进行强化和重点聚焦,从而降低其漏检率。艾青林等人^[8]从数据增强角度出发,依据射影几何和消隐点的性质,设计了一种自适应增殖小尺度目标生成方法,从而在检测区域生成更多的尺寸匹配的行人目标,强化了模型在训练过程中对于小尺度目标的学习,进而降低小尺度目标的漏检率。Huang 等人^[9]提出了一种基于融合通道注意力的视觉转换模块并将其嵌入骨干网络中,有效地提取了不同尺度行人目标的特征信息,从而提升算法检测多尺度行人目标的能力。Shen 等人^[10]设计了一种部分空间金字塔池化模块并构建了特征注意力融合模块,能够自适应弱化模态冗余信息,提升多尺度特征融合效率,接着对多尺度特征信息进行级联处理,从而进一步增强网络对小尺度行人目标的敏感度。

针对被遮挡行人目标漏检率较高的问题, Han 等人^[11]提出了一种联合距离交叉损失函数,同时考虑了行人被遮挡区域的位置和面积,并采

用重定位的策略来确立最佳边界框,有效地提升了算法检测遮挡行人目标的能力。Lu 等人^[12]通过非遮挡行人或者部分被遮挡行人附近插入干扰目标来模拟遮挡场景,并提出一种遮挡边缘集中学习策略来预测目标的遮挡边缘偏移量,从而提升遮挡行人目标的检测精度。Huang 等人^[13]在算法的后处理环节利用目标可见部分作为判别筛选条件来抑制冗余预测框,进而提高对遮挡框的非极大值抑制(Non-Maximum Suppression, NMS)效率。Cai 等人^[14]提出了一种 CrowdSAM 网络结构,通过引入高效的提示采样器和背景-目标判别网络,提高了拥挤场景中行人检测的准确率。郝帅等人^[15]提出了一种动态特征优化机制,通过设计亮度感知模块和 EG-Chimp 优化模型,在增强输入图像局部对比度的同时抑制背景信息,并引入了 CIoU 损失函数来加速网络收敛,有效地提升了算法检测遮挡行人的准确性。

针对模型训练过程中因随机采样而造成的困难样本学习不充分、检测器检测性能受限的问题,SHRIVASTAVA 等人^[16]提出了一种在线困难样本挖掘策略,通过计算每一样本的损失并依据损失高低来定义困难样本,并在训练过程中强化对困难样例的学习,从而提升检测器性能。Lin 等人^[17]提出了 Focal Loss 损失函数,通过引入额外参数来优化交叉熵损失,使得网络在训练过程中自主关注困难样本,进而增强检测器的检测能力。以上方法虽能促进检测器性能提升,但实施流程较为繁琐且计算量较高。

上述算法从不同角度对行人检测方案进行了优化,用于解决密集场景中行人遮挡和尺度不一造成的检测器性能下降问题,但忽略了一个重要因素,即行人头部区域往往受到的遮挡较轻甚至不受遮挡,即便行人躯体部分受到了严重遮挡,头部区域仍能提供较为明显的特征信息,这些信息对于检测模块而言极为重要。对此,本文结合对行人头部区域的特定考虑,从优化特征提取、强化困难样本训练、优化后处理环节等方面入手,对 Faster R-CNN 算法进行了改进,提出一种基于头部与整体信息联合的行人检测算法,旨在提升对密集场景中遮挡行人目标和多尺度行人目标的检测精度并降低其漏检率。

2 Faster R-CNN 算法简介

Faster R-CNN 算法由 Girshick 等人提出,在两阶段检测算法中极具代表性,其整体结构如图 1 所示。

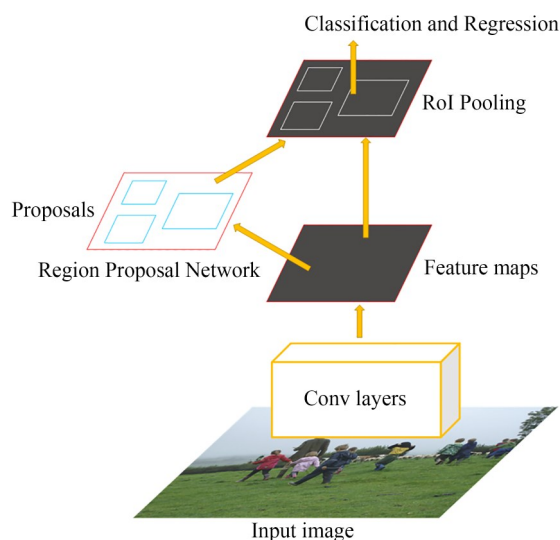


图 1 Faster R-CNN 算法的整体结构图

Fig. 1 Overall structure of Faster R-CNN algorithm

基准 Faster R-CNN 模型采用 VGG16 作为骨干网络,而在实际行人检测任务中,可依据实际需求选择更为合适的骨干网络,从而优化检测器的综合性能,使其更加适用于检测场景。当特征提取网络获取输入图像的特征信息后,将其输入至区域建议网络(Region Proposal Network, RPN),RPN 模块会在每个像素点处生成 k 个候选框,即锚框。利用锚框可以在不同的尺度和比例下生成包含尽可能多目标的区域,从而为后续检测奠定基础,在一定程度上能够提高目标检测成功的概率。

在得到初步的目标候选框集合后,利用 RoI Pooling 模块生成每个候选区域的特征信息,并将其输入至检测头模块,接着通过全连接层和 Softmax 层对其进行分类,然后进行边界框回归环节,使得候选框的参数信息能够更加接近真实框的信息,从而完成行人检测任务。

3 改进算法原理设计

本文针对导致复杂人群密集场景中行人检

测器性能下降的诸多因素,基于 Faster R-CNN 算法进行改进,提出一种基于头部与整体信息联合的行人检测算法,算法的整体结构如图 2 所示。

在本文设计的行人检测算法中,首先在特征提取环节提出了一种基于稠密连接与增强融合的特征提取网络(Dense Connection and Enhanced Fusion Feature Pyramid Network, DF-FPN),旨在获取丰富细致的多尺度目标特征;其次对区域建议网络的采样方式进行了优化,提出一种基于遮挡重叠率判别的非均匀困难样本挖

掘策略,旨在强化训练时对困难样本集中受遮挡严重样本的学习,提升模型对遮挡情况的适应能力;然后在检测模块构建了行人头部与整体联合检测框架,利用头部检测辅助行人检测。在损失函数部分引入了 NMS-Loss 损失^[18],并结合联合检测框架的特点,构建了适用于联合检测的损失函数 Joint NMS-Loss,旨在降低相邻行人目标对检测造成的干扰;最后在后处理环节设计了头部与整体信息的匹配策略,并对未能匹配的头部框进行二次匹配校验,进一步矫正冗余框的筛选结果,使得最终的行人检测结果更加准确。

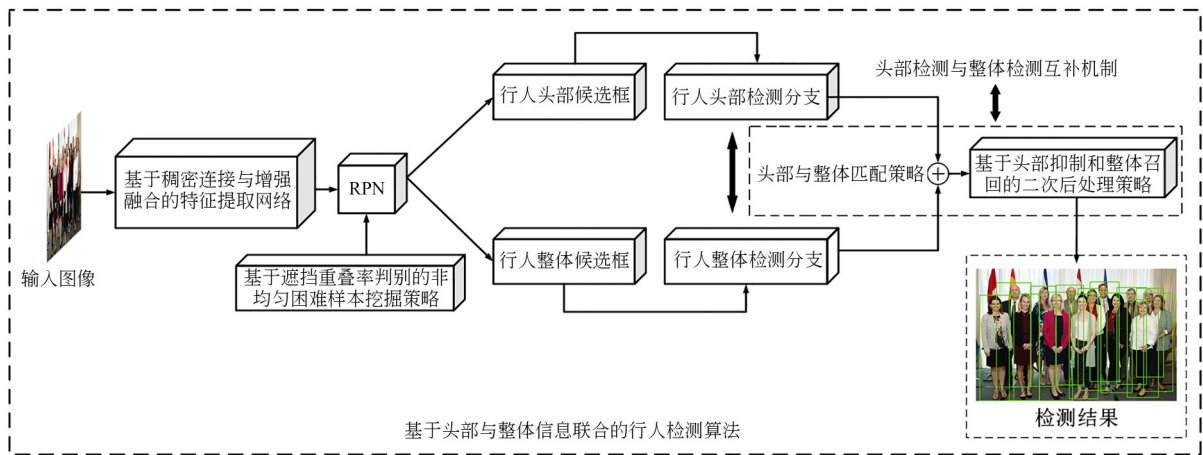


图 2 基于头部与整体信息联合的行人检测算法的整体结构

Fig. 2 Overall structure of pedestrian detection algorithm based on joint head and overall information

3.1 基于稠密连接与增强融合的特征提取网络

在特征提取环节,较为常用的方案是特征金字塔网络(FPN),重点利用不同深度网络对不同尺度特征的学习能力来提升算法的检测性能。然而 FPN 结构中仅对单一尺度图像下采样和上采样后的特征进行融合的策略难以高效应对复杂人群密集场景中的行人检测任务,无法为检测模块提供更加丰富细致的多尺度特征信息。对此,本文对 FPN 结构进行改进,进一步结合稠密连接思想,设计了一种基于稠密连接与增强融合的特征提取网络,其整体结构如图 3 所示。

在 DF-FPN 结构中,采用稠密连接的策略来增强网络对多尺度特征的融合,即令所有尺度的特征均参与计算输出特征信息,从而同时吸取大尺度特征图表达高级语义信息和小尺度特征图表达底层纹理等细节信息的优点。特征提取网络中选用的骨干网络为 ResNet50^[19],当骨干网络

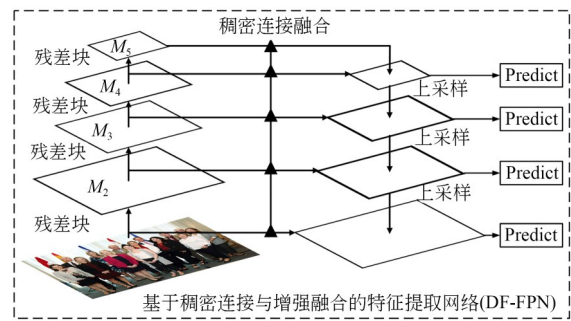


图 3 DF-FPN 整体结构

Fig. 3 DF-FPN overall structure

对输入图像进行下采样操作后,特征图 $M_2 \sim M_5$ 的尺寸也逐渐按倍数缩小。接着通过上采样操作得到与特征图 $M_2 \sim M_5$ 相同尺度的 $M_2' \sim M_5'$ 。在融合不同尺度特征图的过程中,需要提前进行特征图尺寸一致化处理。以 M_4' 处为例,由于特征图 M_5 的尺寸小于 M_4' ,因此需要进行上采样操

作才能得到与输出尺寸相同的特征图。而特征图 M_2 和 M_3 的尺寸均大于 M_4' , 因此需要进行下采样操作来获得与输出尺寸相同的特征图。

以 M_4' 处为例, 不同于 FPN 结构中直接加和方式的特征拼接策略, DF-FPN 结构中采用加权求和的策略将特征图进行叠加求和, 然后拼接至 M_4' 处, 则 M_4' 处对应的预测值 P_4 具体如式 (1) 所示:

$$P_4 = \sum_{j=2}^5 \gamma_j \cdot \text{Translate}_4(M_j), \quad (1)$$

其中: γ_j 均为权重参数, $\text{Translate}(\bullet)$ 为特征映射函数。在具体训练过程中, 参数 γ_j 参与梯度的反向传播, 通过模型的训练来学习和更新至最佳值。该方法使得模型能够学习到哪些连接有助于检测模块进行预测以及哪些连接是无效计算, 进而提升特征提取网络的性能。以此类推, $M_2' \sim M_5'$ 处对应的预测值 $P_2 \sim P_5$ 具体如式 (2) 及式 (3) 所示:

$$P_i = \sum_{j=2}^5 \gamma_j \cdot \text{Translate}_i(M_j), \quad i = 2, 3, 4, 5, \quad (2)$$

$$\text{Translate}_i(M_j) = \begin{cases} 2^{i-j} \text{upsample}(M_j) \cdot \text{conv}(M_j), & i < j \\ \text{conv}(M_j), & i = j \\ 2^{i-j} \text{downsample}(M_j) \cdot \text{conv}(M_j), & i > j \end{cases} \quad (3)$$

从信息流通的角度分析, DF-FPN 结构不仅能够学习到丰富细致的多尺度特征, 并且浅层特征中能够融合深层语义信息, 深层特征中也能融合浅层细节纹理信息, 从而更有助于提升网络对多尺度行人目标的检测灵敏度; 从参数优化的角度分析, 采用稠密连接策略进行多尺度特征融合时, 在实际训练阶段, 梯度能够更加高效地进行反向传播, 进而在加快网络模型收敛速度的同时不过多增加网络复杂度, 更有助于取得高质量的全局最优解。

3.2 基于遮挡重叠率判别的非均匀困难样本挖掘策略

由于在实际训练过程中, 采用随机采样的学习策略会引起样本训练不均衡的问题, 一些困难样本容易被忽略进而造成网络性能下降。而现有困难样本挖掘方法需要计算每一个样本的损失, 然后再依据特定规律筛选出需要重点关注的样本, 这类方案的整体执行流程较为冗杂。对

此, 本文提出一种基于遮挡重叠率判别的非均匀困难样本挖掘策略。通过引入判定阈值 τ , 将样本集合依据其平均遮挡重叠率 \bar{p}^{occ} 划分为困难集 ($\bar{p}^{occ} > \tau$) 和普通集 ($\bar{p}^{occ} \leq \tau$) 两类。然后将每一样本被抽取的概率定义为 R_m , 具体的计算方法如式 (4) 所示:

$$R_m = \frac{\lambda_m n_d}{N}, \quad (4)$$

其中: λ_m 表示第 m 个样本的遮挡系数, 能够反映第 m 个样本受遮挡的程度, 具体如式 (5)~式 (7) 所示:

$$\lambda_m = \begin{cases} 1 + \frac{p_m^{occ} - \bar{p}^{occ}}{\sum_{m=1}^N |p_m^{occ} - \bar{p}^{occ}|}, & \text{if } \bar{p}^{occ} > \tau \\ 1 + p_m^{occ}, & \text{otherwise} \end{cases}, \quad (5)$$

$$p_m^{occ} = \frac{\max_i (\text{Overlap}(m, i))}{\text{Area}(m)}, \quad i \in \{1, 2, \dots, N\}, i \neq m, \quad (6)$$

$$\bar{p}^{occ} = \frac{\sum_{m=1}^N p_m^{occ}}{N}, \quad (7)$$

其中: n_d 表示检测需要的样本数目, N 为候选样本总数。 p_m^{occ} 和 \bar{p}^{occ} 分别表示样本集中第 m 个样本的遮挡重叠率和整个样本集合的平均遮挡重叠率, 阈值 τ 可依据数据集的整体遮挡程度而设置不同值。

上述策略提升了困难样本集合中受遮挡严重样本的权重, 即在实际训练时将对这些样本进行重点关注和学习, 并且随着训练的不断进行, 模型也将逐渐适应并提升检测遮挡行人目标的能力, 从而在复杂人群密集场景中发挥更加出色的性能。

3.3 行人头部与整体联合检测策略

在人群密集场景中, 虽然行人躯体部分受到的遮挡较为严重, 但是由于行人身高差异等因素, 行人头部区域往往受到的遮挡较轻, 因此可用于辅助检测。本文设计的头部与整体联合检测策略为: 在特征提取网络学习到图像的特征信息后, 首先利用区域建议网络对所有行人实例分别生成头部候选框和整体候选框集合并输送至 RoIAlign 模块; 其次基于行人头部和整体双检测分支获得初步的检测结果; 接着利用本文设计的

后处理优化策略对检测结果进行纠正。本文设计的后处理优化策略包括两个环节,分别是头部与整体的匹配环节和基于头部抑制和整体召回的二次后处理环节。头部与整体的匹配环节用于对隶属于同一行人的头部框和整体框进行匹配与绑定。基于头部抑制和整体召回的二次后处理策略用于对未能匹配成功的头部检测结果进行二次匹配校验,进而恢复被错误非极大值抑制的整体框并筛除头部误检框,使得最终的检测结果更加准确。

3.3.1 行人头部框与整体框的匹配策略

在得到初始头部检测框和整体检测框集合后,对其进行初步非极大值抑制处理来剔除冗余框。接着进入头部框与整体框的匹配环节,用于判断其是否属于同一行人。在本文设计的匹配策略中,只有同时满足以下三个约束条件,才能视为成功配对。这三个条件分别为:

(1) 基于头部与整体之间重叠率的判定条件。

设定行人头部框 box_h 的面积为 $area_h$, 则其与整体框 box_b 之间的重叠率 $Occ_{head-body}$ 可表示为:

$$Occ_{head-body} = \frac{Overlap(box_h, box_b)}{area_h}, \quad (8)$$

其中, $Overlap(box_h, box_b)$ 表示头部检测框与整体检测框之间的重叠区域面积。由于隶属于同一行人的头部与整体检测结果之间的重叠比例应当较高,因此若重叠率 $Occ_{head-body}$ 低于预设阈值 t_{occ} 时,则判定无法匹配。参考数据集中对目标的标注情况,本文设置阈值 t_{occ} 为 0.7。

(2) 基于虚拟缩放整体框与行人整体框之间交并比值的判定条件。

对于头部与整体检测结果而言,如果二者之间能够成功匹配,除了位置上需要相互靠近、重叠比例较高之外,还需要在尺度方面满足一定的关系。对此,本文引入一种虚拟缩放整体框的计算策略,来判断头部框与整体框之间的尺度关系。设定头部框 box_h 坐标为 $(x_1^h, y_1^h, x_2^h, y_2^h)$, 其宽度和高度分别为 w_h 和 h_h , 则其对应的虚拟缩放整体框 $box_h^{virtual}$ 的坐标 $(x_1^{vh}, y_1^{vh}, x_2^{vh}, y_2^{vh})$ 的计算方法如式(9)~式(12)所示:

$$x_1^{vh} = x_1^h - \frac{7w_h}{6} = x_1^h - \frac{7(x_2^h - x_1^h)}{6} = \frac{13x_1^h - 7x_2^h}{6}, \quad (9)$$

$$y_1^{vh} = y_1^h, \quad (10)$$

$$x_2^{vh} = x_2^h + \frac{7w_h}{6} = x_2^h + \frac{7(x_2^h - x_1^h)}{6} = \frac{13x_2^h - 7x_1^h}{6}, \quad (11)$$

$$y_2^{vh} = y_2^h + 5h_h = y_2^h + 5(y_2^h - y_1^h) = 6y_2^h - 5y_1^h. \quad (12)$$

在得到虚拟整体框 $box_h^{virtual}$ 的坐标信息后,计算其与整体框 box_b 的交并比值,若交并比值低于所设阈值 t_{iou} ,则判定头部框与整体框无法匹配,本文设置阈值 t_{iou} 为 0.5。若交并比值大于 t_{iou} ,则继续计算两框之间的 L1 距离 D_{L1} ,具体如式(13)所示。如果 D_{L1} 值大于阈值 t_{scale} ,说明头部框与整体框在尺度方面存在较大差异,无法成功配对。对于阈值 t_{scale} 的设置,本文将在后续实验中分析其最优取值。

$$D_{L1}(box_b, box_h^{virtual}) = |W_b - W_{vh}| + |H_b - H_{vh}|. \quad (13)$$

(3) 基于头部与整体关系判别分数的判定条件。

由于复杂人群密集场景中常常出现行人重叠现象,因此周围行人的检测结果也会对头部与整体之间的匹配造成干扰。此时仅依据前两个判定条件难以将头部框与整体框准确地配对,还需要进一步判别二者之间的特征关系,进而提升匹配准确率。对此,本文设计了一个由三个堆叠的全连接层构成的关系判别模块,并为其提取每个{头部~整体}匹配对的特征,进而判定头部框和整体框是否属于同一行人。首先统计能够与当前头部框满足前两个匹配条件的整体框集合,并计算头部与整体之间的关系匹配分数,当关系匹配分数的最大值低于预设阈值 t_{low} 或关系匹配分数集合为空集时,则判定二者之间无法匹配。阈值 t_{low} 在本文中被设置为 0.1。

3.3.2 基于头部抑制和整体召回的二次后处理策略

在对初步 NMS 处理后的行人头部和整体检测结果进行配对后,一些头部框可能没有匹配到对应的整体框。究其原因主要包括两个方面:(1)该头部框本质上是误检结果,因此无法匹配。(2)该头部框对应的整体框由于受到了严重的遮挡,在初步后处理环节已被错误剔除,进而造成匹配失败。对于匹配失败的头部框集合 $head_{fail}$,

需要对其进一步处理而不能直接剔除,因为该集合中不仅包含需要剔除的误检框,还有一些元素需要予以保留,需要利用它们恢复之前被错误抑制的整体检测框,使得最终的检测结果更加准确。

对此,本文设计了基于头部抑制和整体召回的二次后处理策略,用于进一步矫正检测结果。对于集合 $head_{fail}$ 中的元素而言,将其与未经 NMS 处理的整体检测框集合中的元素进行匹配计算,在同时满足匹配条件(1)和条件(2)后,只有在关系判别环节计算得到的匹配分数大于一个较高阈值 t_{high} 时,才能视为匹配成功。阈值 t_{high} 在本文中被设置为 0.9。若此次头部框能够成功匹配,则恢复其对应的整体检测框,因为该整体框属于被遮挡目标的正确预测。如果在关系判别环节计算得到的匹配分数仍低于 t_{low} 或者关系匹配分数集合为空集时,则判定该头部框为误检框。

3.4 损失函数设计

在复杂人群密集场景中,类内遮挡(行人目标之间的相互遮挡)和类间遮挡(行人被其他类别的目标遮挡)均会对行人检测造成干扰,进而造成误检和漏检现象。因此在设计损失函数时,不仅需要考虑分类损失和回归损失,还需要考虑如何降低附近重叠目标对检测造成的干扰以及如何纠正误检和漏检情况。对此,本文在 NMS-Loss 损失的基础上结合头部和整体联合检测网络的具体特点进行进一步改进,提出一种适用于联合检测的损失函数 Joint NMS-Loss,旨在惩罚未被正确 NMS 抑制的误检情况和被 NMS 错误抑制的漏检情况。联合损失函数 L_{Joint} 的计算方法如式(14)~式(18)所示:

$$L_{Joint} = \lambda_{pull}^{body} L_{pull}^{body} + \lambda_{push}^{body} L_{push}^{body} + \lambda_{pull}^{head} L_{pull}^{head} + \lambda_{push}^{head} L_{push}^{head}, \quad (14)$$

$$L_{pull}^{body} = -\ln(1 - N_t^{body} + IoU(box_{max}^{body}, box_m^{body}))s_m^{body}, \quad (15)$$

$$L_{push}^{body} = -\ln(1 - IoU(box_{max}^{body}, box_m^{body}))s_m^{body}, \quad (16)$$

$$L_{pull}^{head} = -\ln(1 - N_t^{head} + IoU(box_{max}^{head}, box_m^{head}))s_m^{head}, \quad (17)$$

$$L_{push}^{head} = -\ln(1 - IoU(box_{max}^{head}, box_m^{head}))s_m^{head}, \quad (18)$$

其中:系数 λ_{pull}^{body} , λ_{push}^{body} , λ_{pull}^{head} 和 λ_{push}^{head} 均为平衡损失的权重。 L_{pull}^{body} 与 L_{pull}^{head} 用于将误检的整体框和头部框“拉近”以便于将其剔除, L_{push}^{body} 与 L_{push}^{head} 用于将被

错误非极大值抑制的漏检整体框和头部框“推远”,使其能够正确对应目标框, N_t^{body} 和 N_t^{head} 分别为整体检测和头部检测对应的 NMS 阈值, box_{max}^{body} 与 box_{max}^{head} 分别表示前一次迭代中置信度分数最高的整体和头部候选框, box_m^{body} 与 box_m^{head} 分别表示第 m 个整体和头部候选框, s_m^{body} 与 s_m^{head} 分别表示 box_m^{body} 与 box_m^{head} 对应的置信度分数。

4 实验结果与分析

4.1 实验环境

本文在进行实验时,所用操作系统为 Ubuntu 16.04,依托的主要软件环境为:PyTorch 1.5.0、CUDA 10.1;硬件平台为:GeForce RTX 3080 Ti GPU×4。在实际训练的过程中,参照基线算法文献中的实验参数设置^[4,6],并结合训练参数经验值以及参数调优处理,采用随机梯度下降优化器进行 35 个 epochs 的训练,每个 GPU 上的 batchsize 设置为 4,动量因子设置为 0.937,学习率采用 warm up 的策略,初始值设置为 1.25×10^{-4} ,在每 1 个 epoch 的第 1 次迭代至第 800 次迭代,学习率线性增加至 1.25×10^{-3} ,然后保持不变。训练至第 24 个 epoch 时,学习率降低至原来的 10%,到第 28 个 epoch 时,学习率衰减为原来的 1%。

4.2 数据集与评估指标

在验证算法的性能时,选用 CrowdHuman^[20], TJU-DHD-pedestrian^[21] 和 CityPersons^[22] 三个行人检测数据集进行实验。其中,Crowd-Human 数据集中共包含 15 000 张训练集和 5 000 张测试集,平均每张图片中包含约 23 人且行人受遮挡现象较为普遍,能够较好还原本文重点研究的复杂人群密集场景,因此非常适合验证算法对遮挡行人目标的检测能力;TJU-DHD-pedestrian 数据集包括两个不同场景下采集和标注的子集 TJU-Ped-traffic 和 TJU-Ped-campus,总共包含 75 246 张图片与 373 241 个标注的行人目标,整体数据规模较大且困难样本较多。此外,该数据集还依据目标尺度大小特点进行了划分,划分结果包括 Reasonable, Reasonable Small 等多个子集,因此非常适合验证算法对多尺度行人目标的检测能力;CityPersons 作为经典的行人检测数据集,也适合验证检测算法对不同遮挡程度行人的

检测能力。在评价算法的综合性能时,本文选用平均精度 AP 和对数平均漏检率 MR^{-2} 作为评估指标,具体计算方法如式(19)及式(20)所示:

$$AP = \int_0^1 Precision(Recall)dRecall, \quad (19)$$

$$MR^{-2} = \int_0^1 MR(FPPI)dFPPI, \quad (20)$$

其中:AP 值越高、 MR^{-2} 值越低,说明算法的检测性能越出色。*Precision*, *Recall* 和 *MR* 分别表示行人检测器的准确率、召回率和漏检率。*FPPI* 表示每张图像中的平均错检数。

4.3 对比实验结果与分析

在对比实验过程中,首先验证阈值 t_{scale} 的选取对算法性能的影响。实验中设置 t_{scale} 的取值区间为 {10, 20, 30, 40, 50, 60, 70, 80}, 其余参数设

置均保持一致,在三个数据集上的实验结果如表 1~表 3 所示。

表 1 CrowdHuman 数据集上 t_{scale} 取值的影响

Tab. 1 Influence of t_{scale} value on CrowdHuman dataset (%)

| Value of t_{scale} | AP | MR^{-2} |
|----------------------|------|-----------|
| 10 | 87.2 | 45.7 |
| 20 | 87.2 | 45.6 |
| 30 | 87.3 | 45.4 |
| 40 | 87.4 | 45.3 |
| 50 | 87.5 | 45.3 |
| 60 | 87.6 | 45.2 |
| 70 | 87.8 | 45.0 |
| 80 | 87.7 | 45.0 |

表 2 TJU-DHD-pedestrian 数据集上 t_{scale} 取值的影响

Tab. 2 Influence of t_{scale} value on TJU-DHD-pedestrian dataset (%)

| Value of t_{scale} | TJU-Ped-campus subset | | | TJU-Ped-traffic subset | | |
|----------------------|--------------------------|--------------------------------|------|--------------------------|--------------------------------|------|
| | Reasonable (MR^{-2}) | Reasonable Small (MR^{-2}) | AP | Reasonable (MR^{-2}) | Reasonable Small (MR^{-2}) | AP |
| 10 | 27.4 | 72.4 | 57.3 | 21.8 | 34.6 | 62.2 |
| 20 | 27.4 | 72.3 | 57.4 | 21.7 | 34.5 | 62.4 |
| 30 | 27.3 | 72.1 | 57.5 | 21.7 | 34.4 | 62.4 |
| 40 | 27.3 | 72.0 | 57.5 | 21.6 | 34.4 | 62.5 |
| 50 | 27.3 | 71.9 | 57.6 | 21.5 | 34.2 | 62.5 |
| 60 | 27.2 | 71.8 | 57.8 | 21.4 | 34.2 | 62.7 |
| 70 | 27.1 | 71.8 | 57.8 | 21.4 | 34.1 | 62.8 |
| 80 | 27.1 | 71.8 | 57.7 | 21.4 | 34.1 | 62.8 |

表 3 CityPersons 数据集上 t_{scale} 取值的影响

Tab. 3 Influence of t_{scale} value on CityPersons dataset (%)

| Value of t_{scale} | Bare (MR^{-2}) | Reasonable (MR^{-2}) | Partial (MR^{-2}) | Heavy (MR^{-2}) |
|----------------------|--------------------|--------------------------|-----------------------|---------------------|
| 10 | 7.6 | 11.3 | 11.0 | 48.4 |
| 20 | 7.6 | 11.2 | 11.0 | 48.3 |
| 30 | 7.5 | 11.0 | 10.9 | 48.1 |
| 40 | 7.4 | 11.0 | 10.7 | 47.9 |
| 50 | 7.3 | 11.0 | 10.6 | 47.9 |
| 60 | 7.3 | 10.9 | 10.4 | 47.7 |
| 70 | 7.2 | 10.8 | 10.4 | 47.6 |
| 80 | 7.2 | 10.8 | 10.4 | 47.6 |

根据表 1~表 3 中的实验结果可知,当阈值 t_{scale} 取值为 70 时,算法在三个行人检测数据集上的综合性能均为最优,因此设置 t_{scale} 为 70。

此外,本文在相同的实验环境下重新训练并测试了融合了 FPN 结构的 Faster R-CNN 算法,并将其作为基线算法,同时选取了一些较为主流的行人检测算法进行对比实验,进而验证所提算法的有效性。对比实验结果如表 4~表 6 所示。

根据表 4~表 6 所示的数据可知,相比于基线算法,本文改进后算法的综合性能取得了明显提升。其中,在 CrowdHuman 数据集上 AP 值提高了 4%, MR^{-2} 值降低了 5.7%,并且与其他主流行人检测算法相比,所提算法也取得了最佳检测效

表 4 CrowdHuman 数据集上的对比实验结果

Tab. 4 Results of comparative experiments on the CrowdHuman dataset (%)

| Algorithm | AP | MR ⁻² |
|------------------------------|-------------|------------------|
| Faster RCNN+FPN (Baseline) | 83.8 | 50.7 |
| Algorithm ^[7] | 87.3 | 50.1 |
| RetinaNet ^[17] | 80.8 | 63.3 |
| YOLOv3 ^[23] | 83.1 | 54.2 |
| YOLOv5 ^[24] | 84.1 | 51.8 |
| FCOS ^[25] | — | 48.3 |
| Adaptive NMS ^[26] | 84.7 | 49.7 |
| Soft-NMS ^[26] | 83.9 | 52.0 |
| JED ^[27] | 85.9 | — |
| CenterNet ^[28] | 79.0 | 53.9 |
| YOLOv8 | 85.6 | — |
| YOLOv10 | 86.2 | — |
| Ours | 87.8 | 45.0 |

果。由此可知,所提算法具备出色的检测遮挡行人目标的能力;在 TJU-DHD-pedestrian 数据集的两个子集中,在由较小尺度行人目标组成的 Reasonable Small 子集上 MR⁻² 值分别降低了 2.4% 和 2.1%,在由正常尺度行人目标组成的 Reasonable 子集上 MR⁻² 值分别降低了 1.6% 和 1.5%,算法的 AP 值分别提升了 3.4% 和 2.8%,并且综合性能也优于对比算法。由此可知,所提算法对多尺度行人目标的检测能力也取得了提升;在 CityPersons 数据集上,所提算法在遮挡情况严重的 Heavy 子集、遮挡情况适中的 Partial 子集和 Reasonable 子集上均取得了最佳检测效果,虽然在遮挡程度较低的 Bare 子集上性能略逊于 OR-CNN 算法,但差距仅有 0.5%。由此可知,所提算法对于不同遮挡程度的行人目标均能表现出较为出色的检测性能,具备较强的适应性。

表 5 TJU-DHD-pedestrian 数据集上的对比实验结果

Tab. 5 Results of comparative experiments on the TJU-DHD-pedestrian dataset (%)

| Algorithm | TJU-Ped-campus subset | | | TJU-Ped-traffic subset | | |
|------------------------------------|--------------------------------|--------------------------------------|-------------|--------------------------------|--------------------------------------|-------------|
| | Reasonable (MR ⁻²) | Reasonable Small (MR ⁻²) | AP | Reasonable (MR ⁻²) | Reasonable Small (MR ⁻²) | AP |
| Faster R-CNN+FPN (Baseline) | 28.7 | 74.2 | 54.4 | 22.9 | 36.2 | 60.0 |
| RetinaNet ^[17] | 34.7 | 83.0 | 50.5 | 23.9 | 37.9 | 58.4 |
| FCOS ^[25] | 31.9 | 81.3 | 51.1 | 24.4 | 37.4 | 58.7 |
| Faster R-CNN+PANet ^[29] | 28.2 | 73.4 | — | 22.4 | 35.5 | — |
| Ours | 27.1 | 71.8 | 57.8 | 21.4 | 34.1 | 62.8 |

表 6 CityPersons 数据集上的对比实验结果

Tab. 6 Results of comparative experiments on the CityPersons dataset (%)

| Algorithm | Bare (MR ⁻²) | Reasonable (MR ⁻²) | Partial (MR ⁻²) | Heavy (MR ⁻²) |
|--------------------------------|--------------------------|--------------------------------|-----------------------------|---------------------------|
| Repulsion Loss ^[30] | 7.6 | 13.2 | 16.8 | 56.9 |
| MFGP ^[31] | 7.9 | 10.9 | 10.9 | 49.9 |
| TLL ^[32] | 10.0 | 15.5 | 17.2 | 53.6 |
| ALFNet ^[33] | 8.4 | 12.0 | 11.4 | 51.9 |
| OR-CNN ^[34] | 6.7 | 12.8 | 15.3 | 55.7 |
| CCFA-Net ^[35] | — | 15.4 | — | 52.9 |
| Ours | 7.2 | 10.8 | 10.4 | 47.6 |

综上所述可知,所提算法适用于复杂人群密集场景中的行人检测任务,能够准确地检测出场景内的遮挡行人目标和多尺度行人目标。为了更加直观地体现所提算法的有效性,列举了算法改进前后对同一场景内行人目标的检测结果,具体如图 4~图 5 所示。其中,图 4 为基线算法的检测结果,图 5 为本文算法的检测结果。

分析图 4 及图 5 所示的检测结果可知,所提算法的检测性能明显优于改进前的基线算法。基线算法检测结果中的漏检目标主要是一些被遮挡的行人和较远视角处的小尺度行人。而本文算法中由于构建了头部与整体联合检测框架,因此即便行人躯体部分被严重遮挡,头部检测也

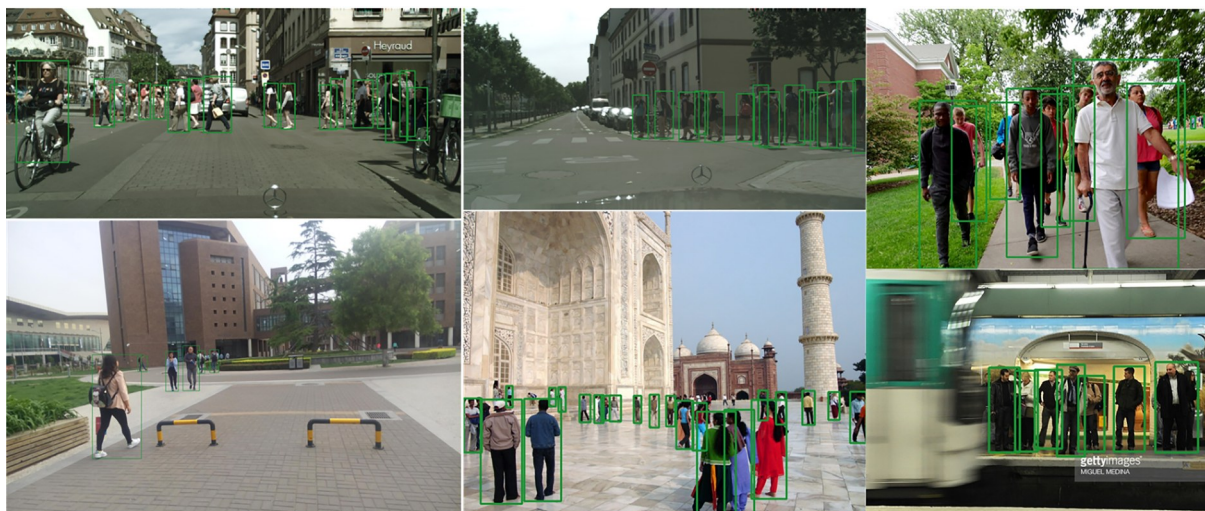


图 4 基线算法检测结果

Fig. 4 Detection results of the baseline algorithm



图 5 本文算法检测结果

Fig. 5 Detection results of the algorithm in this paper

能辅助检测器成功检测出这些目标,并且算法中对后处理环节的优化策略也能有效缓解遮挡行人的漏检现象。同时,所提算法中的DF-FPN模块能够获取到丰富细致的多尺度特征信息,因此能够增强对容易漏检的小尺度行人的检测灵敏度,有效地降低了其漏检率。

4.4 消融实验结果与分析

除对比实验外,本文还在更能反映复杂人群密集场景的CrowdHuman数据集上进行了消融实验,具体方法为在基线算法的基础上逐次添加DF-FPN结构、所提困难样本挖掘策略、头部与整体联合检测策略和联合损失函数,以此来验证

各个改进模块对算法检测性能的提升效果,选用的算法评估指标仍为平均精度AP和对数平均漏检率 MR^{-2} 。消融实验结果如表7所示。

根据表7所示的消融实验结果可知,在基线算法的基础上引入DF-FPN结构后,算法的AP值提升了0.9%, MR^{-2} 值降低了1.2%;进一步引入困难样本挖掘策略后,算法的AP值提升了0.6%, MR^{-2} 值降低了0.9%;接着引入头部与整体联合检测策略后,算法的AP值提升了2.0%, MR^{-2} 值降低了2.9%,算法性能取得了较大幅度的提升;继续引入联合损失函数后,得到了所提算法的完整版,算法的AP值进一步取得了0.5%

表 7 消融实验结果

Tab. 7 Results of ablation experiments

(%)

| Algorithm | DF-FPN | 困难样本挖掘策略 | 头部与整体联合检测策略 | 联合损失函数 | AP | MR ⁻² |
|-----------|--------|----------|-------------|--------|------|------------------|
| Ours | ✓ | ✓ | ✓ | ✓ | 87.8 | 45.0 |
| | ✓ | ✓ | ✓ | × | 87.3 | 45.7 |
| | ✓ | ✓ | × | × | 85.3 | 48.6 |
| | ✓ | × | × | × | 84.7 | 49.5 |
| Baseline | × | × | × | × | 83.8 | 50.7 |

的提升,MR⁻²值也进一步取得了0.7%的下降。由此可见,本文设计的各个改进模块均能促进算法检测性能的提升,有效地缓解了因遮挡和尺度变化造成的目标检测精度下降和漏检率升高。

5 结 论

本文针对复杂人群密集场景中因行人目标受遮挡和行人目标尺度不一等因素导致检测器检测精度下降、漏检率升高的问题,基于Faster R-CNN算法进行改进,提出一种基于头部与整体信息联合的行人检测算法。首先,在特征提取环节设计了一种基于稠密连接与增强融合的特征提取网络,用于获取丰富细致的多尺度特征信息,进而提升网络对多尺度行人目标的检测灵敏度;其次,对区域建议网络的采样方式进行了优化,设计了一种基于遮挡重叠率判别的非均匀困难样本挖掘策略,使得在实际训练时重点关注和学习被严重遮挡的困难样本,增强模型检测遮挡行人目标的能力;然后提出了行人头部与整体联

合检测策略,充分利用了头部区域受遮挡程度较轻的特点进行辅助检测,并对损失函数和后处理环节进行了优化,用于惩罚周围行人目标对检测造成的干扰和矫正最终的检测结果。最后在CrowdHuman, TJU-DHD-pedestrian 和 CityPersons三个数据集上验证了所提算法的有效性。

作者贡献声明:

马晞茗:行人检测算法构建,对比实验与消融实验设计与验证,论文结构构思和内容撰写;

李宁:对算法模型进行推导论证和合理性分析,对论文撰写结构和内容进行指导;

吴迪:论文审核与编辑写作,对论文初稿进行修改和指导;

刘一豆:对比实验和消融实验流程设计并验证;

于祥跃:调研并梳理论文算法相关文献、对算法进行重复性实验以验证其有效性;

李峥:调研并梳理论文相关实验数据集、对算法进行重复性实验以验证其有效性。

参考文献:

- [1] SUKKAR M, JADEJA R, SHUKLA M, *et al.* A survey of deep learning approaches for pedestrian detection in autonomous systems [J]. *IEEE Access*, 2024, 13: 3994-4007.
- [2] 龚安, 李中浩, 梁辰宏. 夜间多场景的邻近感知实时行人检测算法[J]. *中国图象图形学报*, 2023, 28(9): 2693-2705.
GONG A, LI Z H, LIANG C H. NSPDet: real-time nearby-aware pedestrian detection algorithm for multi-scene surveillance at night [J]. *Journal of Image and Graphics*, 2023, 28(9): 2693-2705. (in Chinese)

- [3] 叶洪滨, 林政宽, 程红举. 基于多相机特征融合的行人检测算法[J]. *北京邮电大学学报*, 2023, 46(5): 66-71.
YE H B, LIN Z K, CHENG H J. Pedestrian detection algorithm based on multi-camera feature fusion [J]. *Journal of Beijing University of Posts and Telecommunications*, 2023, 46(5): 66-71. (in Chinese)
- [4] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [5] 吕志轩, 魏霞, 黄德启. 多分支无锚框网络密集行人检测算法[J]. *光学精密工程*, 2023, 31(10):

- 1532-1547.
- LÜ Z X, WEI X, HUANG D Q. Dense pedestrian detection algorithm in multi-branch non-anchor frame network[J]. *Opt. Precision Eng.*, 2023, 31(10): 1532-1547. (in Chinese)
- [6] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* *Feature Pyramid Networks for Object Detection* [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.
- [7] 陈勇, 金曼莉, 刘焕淋, 等. 基于特征增强模块的小尺度行人检测[J]. *电子与信息学报*, 2023, 45(4): 1445-1453.
- CHEN Y, JIN M L, LIU H L, *et al.* Small-scale pedestrian detection based on feature enhancement strategy [J]. *Journal of Electronics & Information Technology*, 2023, 45(4): 1445-1453. (in Chinese)
- [8] 艾青林, 杨佳豪, 崔景瑞. 基于自适应增殖数据增强与全局特征融合的小目标行人检测[J]. *浙江大学学报(工学版)*, 2023, 57(10): 1933-1944, 1976.
- AI Q L, YANG J H, CUI J R. Small target pedestrian detection based on adaptive proliferation data enhancement and global feature fusion[J]. *Journal of Zhejiang University (Engineering Science)*, 2023, 57(10): 1933-1944, 1976. (in Chinese)
- [9] HUANG S C, ZHANG S B, JIAO Y F. Ms-VLPD: a multi-scale VLPD based method for pedestrian detection[J]. *Expert Systems with Applications*, 2025, 271: 126613.
- [10] SHEN Y, XIE X Y, WU J, *et al.* EAFF-Net: Efficient attention feature fusion network for dual-modality pedestrian detection[J]. *Infrared Physics & Technology*, 2025, 145: 105696.
- [11] HAN R, XU M Q, PEI S W. Crowded pedestrian detection with optimal bounding box relocation[J]. *Multimedia Tools and Applications*, 2024, 83(24): 65687-65706.
- [12] LU K Q, ZHU C, LIU M Y, *et al.* OSS-OCL: Occlusion Scenario Simulation and Occluded-edge Concentrated Learning for pedestrian detection[J]. *Pattern Recognition Letters*, 2025, 189: 201-206.
- [13] HUANG X, GE Z, JIE Z Q, *et al.* NMS by representative region: towards crowded pedestrian detection by proposal pairing[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 10747-10756.
- [14] CAI Z, GAO Y J, ZHENG Y Y, *et al.* CrowdSAM: SAM as a smart annotator for object detection in crowded scenes [J]. *arXiv preprint arXiv: 2407.11464*, 2024.
- [15] 郝帅, 何田, 马旭, 等. 动态特征优化机制下的跨尺度红外行人检测[J]. *光学精密工程*, 2022, 30(19): 2390.
- HAO S, HE T, MA X, *et al.* Cross-scale infrared pedestrian detection based on dynamic feature optimization mechanism[J]. *Opt. Precision Eng.*, 2022, 30(19): 2390. (in Chinese)
- [16] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with on-line hard example mining[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 761-769.
- [17] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
- [18] LUO Z K, FANG Z, ZHENG S X, *et al.* NMS-loss: learning with non-maximum suppression for crowded pedestrian detection[C]. *Proceedings of the 2021 International Conference on Multimedia Retrieval, Taipei*. ACM, 2021: 481-485.
- [19] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [20] SHAO S, ZHAO Z J, LI B X, *et al.* CrowdHuman: a benchmark for detecting human in a crowd [J]. *arXiv preprint arXiv:1805.00123*, 2018.
- [21] PANG Y W, CAO J L, LI Y Z, *et al.* TJU-DHD: a diverse high-resolution dataset for object detection [J]. *IEEE Transactions on Image Processing*, 2021, 30: 207-219.
- [22] ZHANG S S, BENENSON R, SCHIELE B. CityPersons: a diverse dataset for pedestrian detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 4457-4465.

- [23] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv preprint arXiv:1804.02767*, 2018.
- [24] GE Z, LIU S T, WANG F, *et al.* YOLOX: exceeding YOLO series in 2021 [J]. *arXiv preprint arXiv:2107.08430*, 2021.
- [25] TIAN Z, SHEN C H, CHEN H, *et al.* FCOS: fully convolutional one-stage object detection [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019. Seoul, Korea (South). IEEE, 2019: 9627-9636.
- [26] LIU S T, HUANG D, WANG Y H. Adaptive NMS: refining pedestrian detection in a crowd [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019. Long Beach, CA, USA. IEEE, 2019: 6459-6468.
- [27] ZHANG Y M, ZHANG S F, ZHUANG C B, *et al.* Feature Enhancement for Joint Human and Head Detection [M]. Biometric Recognition. Cham: Springer International Publishing, 2019: 511-518.
- [28] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as Points [J]. *arXiv preprint arXiv:1904.07850*, 2019.
- [29] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8759-8768.
- [30] WANG X L, XIAO T T, JIANG Y N, *et al.* Repulsion loss: detecting pedestrians in a crowd [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 7774-7783.
- [31] ZHANG Y H. Multi-scale object detection model with anchor free approach and center of gravity prediction [C]. 2020 *IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. June 12-14, 2020. Chongqing, China. IEEE, 2020: 38-45.
- [32] SONG T, SUN L Y, XIE D, *et al.* Small-scale pedestrian detection based on somatic topology localization and temporal feature aggregation [J]. *arXiv preprint arXiv:1807.01438*, 2018.
- [33] LIU W, LIAO S C, HU W D, *et al.* Learning Efficient Single-Stage Pedestrian Detectors by Asymptotic Localization Fitting [M]. Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 643-659.
- [34] ZHANG S F, WEN L Y, BIAN X, *et al.* Occlusion-aware R-CNN: Detecting Pedestrians in A Crowd [M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 657-674.
- [35] ZOU F M, LI X, XU Q M, *et al.* Correlation-and-correction fusion attention network for occluded pedestrian detection [J]. *IEEE Sensors Journal*, 2023, 23(6): 6061-6073.

作者简介:



马晞茗(1999—),男,吉林长春人,硕士,研究实习员,2021年于北京交通大学获得学士学位,2024年于哈尔滨工程大学获得硕士学位,主要从事计算机视觉以及行人检测技术方面的研究。E-mail: ranking1230@163.com

通讯作者:



李 宁(1983—),男,吉林白山人,博士,研究员,2007年于长春理工大学获得学士学位,2012年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事目标识别技术等方面的研究。E-mail: lining@ciomp.ac.cn