

文章编号 1004-924X(2026)14-2217-15

夜间低光照场景下基于多尺度特征融合的 密集行人检测算法

马晞茗, 李 宁*, 年 伦, 吴 迪, 于祥跃, 李 峥, 王 瑶

(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)

摘要:在复杂夜间场景中,光照变化、行人受遮挡以及行人尺度不一等因素导致行人检测的准确率出现了降低。对此,提出一种夜间低光照场景下基于多尺度特征融合的密集行人检测算法。首先,设计了一种融合自注意力与卷积混合模块的双向路径特征聚合网络,用于增强多尺度特征融合并提升行人与背景的区分度,从而降低目标尺度差异对检测造成的影响。其次,构建了多分支联合检测策略,充分利用夜间低光照环境中行人轮廓、腿部和手臂特征的强辨别性来辅助行人检测,降低弱光照和遮挡对检测造成的影响。然后,在后处理环节引入对检测框之间中心距离的考虑,通过降低遮挡严重检测框的置信度分数并进行重新筛选,从而保留被遮挡目标的正确检测结果,进一步降低行人漏检率。最后,通过实验验证了所提算法的有效性。实验结果表明,在 NightOwls 和 NightSurveillance 数据集上算法的对数平均漏检率分别降低了 2.8% 和 3.3%,在 LLVIP 数据集上算法的平均精度和召回率分别提升了 2.1% 和 2.9%,并且提出的三个改进模块均有助于提升算法的整体检测性能。

关键词:行人检测;多分支联合检测;多尺度特征融合;注意力机制;后处理优化

中图分类号:TP391.41;TN911.73 **文献标识码:**A

doi:10.37188/OPE.20263414.2217 **CSTR:**32169.14.OPE.20263414.2217

Dense pedestrian detection algorithm based on multi-scale feature fusion for low-light nighttime scenes

MA Ximing, LI Ning*, NIAN Lun, WU Di, YU Xiangyue, LI Zheng, WANG Yao

(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Optical Countermeasures Department, Changchun 130033, China)

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

Abstract: In complex nighttime scenes, factors such as light variations, pedestrian occlusions, and varying pedestrian scales lead to a decrease in pedestrian detection accuracy. Addressing this challenge, we proposed a dense pedestrian detection algorithm based on multi-scale feature fusion for low-light nighttime scenes. Firstly, a bidirectional path feature aggregation network integrated with self-attention and convolutional mixing module was designed. This enhanced multi-scale feature fusion and improved the discriminability between pedestrians and background, thus reduced the impact of target scale variations upon detection. Secondly, a multi-branch joint detection strategy was constructed. This strategy effectively leveraged the strong discriminative power of pedestrian contours, legs, and arms in low-light environments to

收稿日期:2026-03-25;修订日期:2026-04-13.

基金项目:国家自然科学基金(No. 62001447);中国科学院青年创新促进会会员(No. 2023224)

assist detection, reduced the impact of weak light and occlusion. Next, we introduced consideration of center distances between detection boxes during post-processing. By reducing confidence scores for severely occluded boxes and performing re-selection, we successfully retained correct detections of occluded pedestrian targets, further decreased the miss detection rate. Finally, the effectiveness of the proposed algorithm was verified through experiments. The results show that the log-average miss detection rate is decreased by 2.8% and 3.3% on the NightOwls and NightSurveillance datasets, respectively. On the LLVIP dataset, the average precision and recall are improved by 2.1% and 2.9%, respectively. Furthermore, all three proposed improvement modules contribute to enhancing the detection performance of the algorithm.

Key words: pedestrian detection; multi-branch joint detection; multi-scale feature fusion; attention mechanism; post-processing optimization

1 引言

行人检测作为计算机视觉领域的研究热点,在无人驾驶^[1]、安防监控^[2]、智能机器人^[3]等领域得到了广泛的应用。然而,多数行人检测方法对光照条件的依赖性较高。当光照条件良好时,这些算法能够发挥较为出色的性能。在面向夜间等低光照场景时,其检测性能出现了明显下降^[4-5],尤其在低光照与密集行人共存的复杂场景下,检测难度会进一步提升。造成该现象的原因主要包括以下几点:(1)在夜间低光照场景中,行人目标与背景的对比度较低,并且背景干扰目标与行人的外观特征高度相似,此时模型对行人和背景的判别能力较弱,容易造成误检;(2)低光照条件导致行人整体特征和纹理信息较为匮乏,并且不同视角下行人目标的尺度差异显著,较小尺度的目标容易漏检;(3)在复杂密集场景中,在单位区域内行人分布较为集中,并且目标之间的间距较小,因此行人容易受到遮挡,具体包括行人相互遮挡与行人受背景遮挡^[6],导致行人全身可见区域所占比例降低,使得特征提取难度加大,并且算法检测模块推理所需的信息也随之减少,从而引起漏检率升高和检测精度下降。因此,研究夜间低光照场景下的行人检测极具挑战性,对于构建普适性计算机视觉检测模型具有重要意义^[7]。

近年来,随着计算机视觉技术的不断发展,基于深度学习的目标检测算法取得了巨大突破,其检测性能相比于传统检测算法取得了显著提

升^[8]。其中,较为常见的包括以 Faster R-CNN 算法^[9]为代表的两阶段检测算法和以 YOLO 系列^[10]为代表的单阶段检测算法。然而,这些通用的目标检测算法均难以高效应对夜间低光照场景中的行人检测任务,具体表现在可见度低的行人目标、遮挡行人目标和小尺度行人目标的检测精度下降和漏检率升高。

针对夜间低光照条件下行人与背景对比度低造成的检测精度下降问题,Li 等人^[11]设计了一种带有卷积门控线性单元的单头自注意力模块,有效地增强了模型在低光照条件下的特征表达能力。Chen 等人^[12]提出了一种融合小波多级提取器和知识调度模块的行人检测网络,用于实现不同光照环境下自适应融合语义层级特征,提升网络对低光照环境的适应能力,进而提升目标检测精度。Liu 等人^[13]面向夜间交通场景中的行人检测任务,引入了特征增强模块来强化提取低可见度行人的特征,并且设计了混合注意力机制来突出场景中的显著区域,以此来增强检测模型的综合性能。

针对远距离视角处小尺度行人目标漏检率较高的问题,Li 等人^[14]将 MetaFormer 架构^[15]与可变形卷积及深度可分离卷积进行融合,以此来增强特征提取模块对细粒度特征的提取能力,进而有效降低了小尺度行人目标的漏检率。Huang 等人^[16]提出了一种基于网格的概率粘贴增强策略,通过在网格区域依据概率粘贴目标信息以增加多尺度训练样本的多样性,同时构建了融合通道注意力的视觉变换器模块并将其嵌入至骨干

网络,旨在融合全局特征和局部信息,从而提升模型对多尺度行人的检测能力。Yu 等人^[17]提出了一种多尺度等高线特征增强模块,用于提升网络在不同尺度下感知行人目标的能力,并对行人目标和复杂背景进行准确区分,以此来降低小尺度行人目标的漏检率。

针对被遮挡行人漏检率较高的问题,Chang 等人^[18]通过融合全局上下文信息和空间信息,旨在提取额外的潜在特征,从而降低因目标受遮挡而造成的特征损失,并引入了对行人姿态信息的学习来进一步抑制行人遮挡对检测精度造成的不利影响。郑开魁等人^[19]首先基于 CBAM(Convolutional Block Attention Module)^[20]注意力模块动态调整特征图各通道与空间位置的重要性,然后顺序级联 Mamba 模块对增强后的特征图进行全局信息建模,旨在提取被遮挡行人目标周围上下文信息,并结合行人可见特征进行全局推理,从而降低遮挡行人的漏检率。Li 等人^[21]构建了一种面向遮挡场景的检测策略,首先引入基于 Transformer 的遮挡感知模块并与主检测器组成多任务学习框架,然后设计了遮挡感知互补机制,将学习到的遮挡信息与检测解码器进行融合,从而形成反馈优化闭环,有效地提升了模型对遮挡行人的检测能力。

上述算法从不同角度对行人检测算法进行

了改进,以缓解遮挡和尺度变化引起的检测性能下降问题,并提升了算法对夜间低光照等复杂场景的适应性,但忽略了一个重要因素,即在夜间低光照场景中,行人目标具有相对固定的人体轮廓特征,如头部、肩部、手部和腿部等,此类特征与背景具有显著区分度。同时,道路行人的腿部和手臂特征视觉辨识度较高,能够为提升检测精度提供重要支撑。对此,本文选取 Faster R-CNN^[9]作为基准算法,该算法为经典的两阶段检测算法,具有结构扩展性强、检测精度高等优势,能够高效适配复杂场景下的行人检测任务,并且便于嵌入改进模块。在此基础上,本文聚焦局部关键特征的挖掘与利用,并从优化特征提取和优化检测框筛选机制方面入手,对 Faster R-CNN 算法进行改进,提出一种夜间低光照场景下基于多尺度特征融合的密集行人检测算法,旨在提升对夜间场景中低可见度、遮挡及多尺度行人的检测能力并降低漏检率。

2 改进算法原理设计

本文针对导致夜间低光照场景下行人检测准确率降低的诸多因素,基于 Faster R-CNN 算法进行改进,提出一种基于多尺度特征融合的密集行人检测算法,算法的整体结构如图 1 所示。

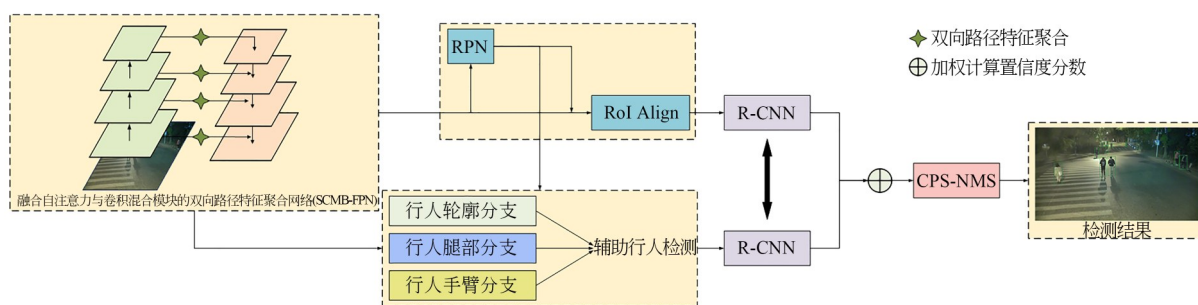


图 1 夜间低光照场景下基于多尺度特征融合的密集行人检测算法的整体结构图

Fig. 1 Overall structure of dense pedestrian detection algorithm based on multi-scale feature fusion for low-light nighttime scenes

在图 1 所示的行人检测算法中,首先在特征提取环节提出了一种融合自注意力与卷积混合模块的双向路径特征聚合网络(Bidirectional Path Feature Aggregation Network Integrated with Self-Attention and Convolutional Mixing Module,

SCMB-FPN),用于获取丰富细致的多尺度特征信息并且更加准确地区分背景和待检测行人目标,提升网络对多尺度行人目标的检测灵敏度,从而降低目标尺度差异对检测造成的影响;其次,在算法的检测模块设计了一种多分支联合检

测框架,利用构建的行人轮廓检测分支、行人腿部检测分支和行人手臂检测分支来辅助行人检测,以此来降低暗弱光照和遮挡对检测造成的干扰;最后在后处理环节设计了一种融合置信度惩罚策略的非极大值抑制算法(Non-Maximum Suppression Algorithm with Integrated Confidence Penalty Strategy, CPS-NMS),通过尽可能保留被遮挡行人的正确检测结果,从而进一步降低漏检率。

2.1 融合自注意力与卷积混合模块的双向路径特征聚合网络

对于多尺度目标检测任务而言,较为常用的策略是在特征提取环节引入特征金字塔网络(Feature Pyramid Network, FPN)^[22],通过融合深层语

义信息和浅层细节信息来应对目标尺度变化问题。然而这种单向融合机制忽略了自底向上路径在传递细节定位信息方面的作用,导致细节信息逐层衰减,难以高效适配复杂场景下的行人检测任务。对此,本文对FPN结构进行改进,新增了自底向上的特征融合路径,构建双向特征聚合机制,旨在实现语义信息与定位信息之间的互补传递。同时,引入ACmix混合自注意力增强模块(Mixed Self-Attention and Convolution)^[23],充分发挥卷积的局部特征提取优势与自注意力的全局特征建模能力,提出一种融合自注意力与卷积混合模块的双向路径特征聚合网络(SCMB-FPN),从而实现更充分、更具判别性的多尺度特征融合。SCMB-FPN的整体结构如图2所示。

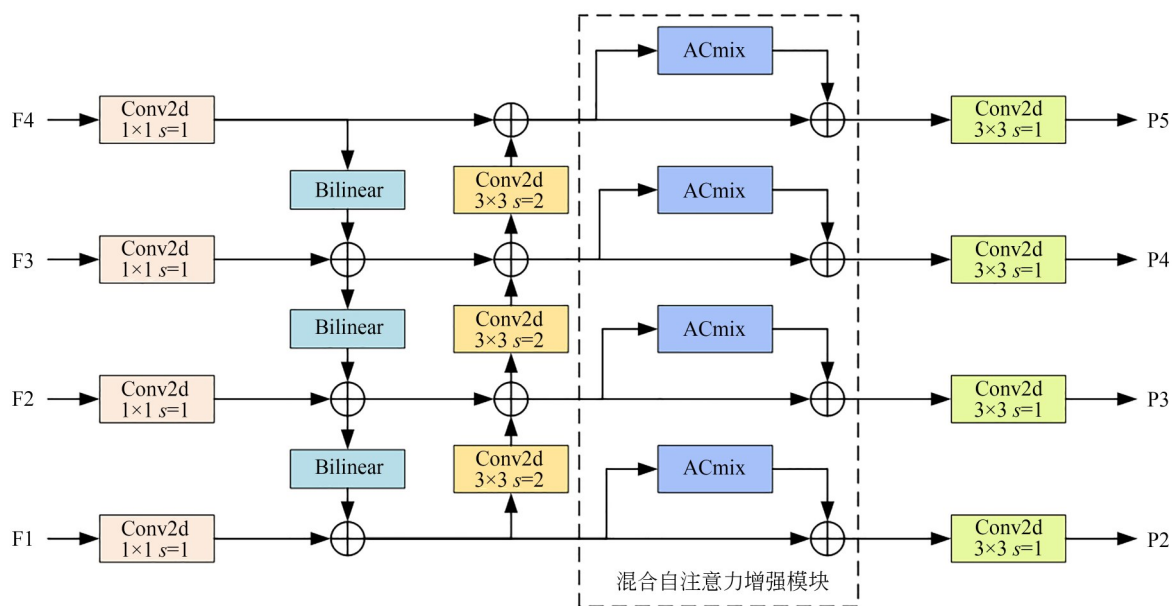


图2 SCMB-FPN整体结构图

Fig. 2 Overall structure of SCMB-FPN

在图2所示的SCMB-FPN结构中,首先利用骨干网络ResNet50学习待检测图像的特征,得到特征图F1,F2,F3和F4,然后对骨干网络获取的特征信息进行融合,具体特征融合的过程中共包括两个阶段:在第一阶段先将4个特征图F1,F2,F3和F4分别进行 1×1 的卷积操作,将其转为256通道数,再对特征图进行2倍上采样操作并将上采样处理后的特征图进行自上而下的相加操作,从而实现特征图的第一次融合。其中,在实际采样过程中,采用双线性插值算法在两个

方向分别进行一次线性插值,从而充分利用目标点周围4个像素点的像素值信息,并降低特征图传递过程中的信息损失。在第二阶段首先利用步长为2、尺寸为 3×3 的卷积核将融合后的特征图进行下采样处理,然后将得到的各层特征图进行自下而上的相加操作,从而完成特征图的第二次融合。

在进行两次特征融合操作后,将得到的特征图传入ACmix混合自注意力模块中进行特征增强,通过对卷积模块和自注意力模块进行特征分解与

整合操作,实现对图像的全局和局部建模,以此来获取更大的感受野和更加丰富的上下文信息,从而

更加准确地区分背景和待检测目标。ACmix 混合自注意力增强模块的结构图如图 3 所示。

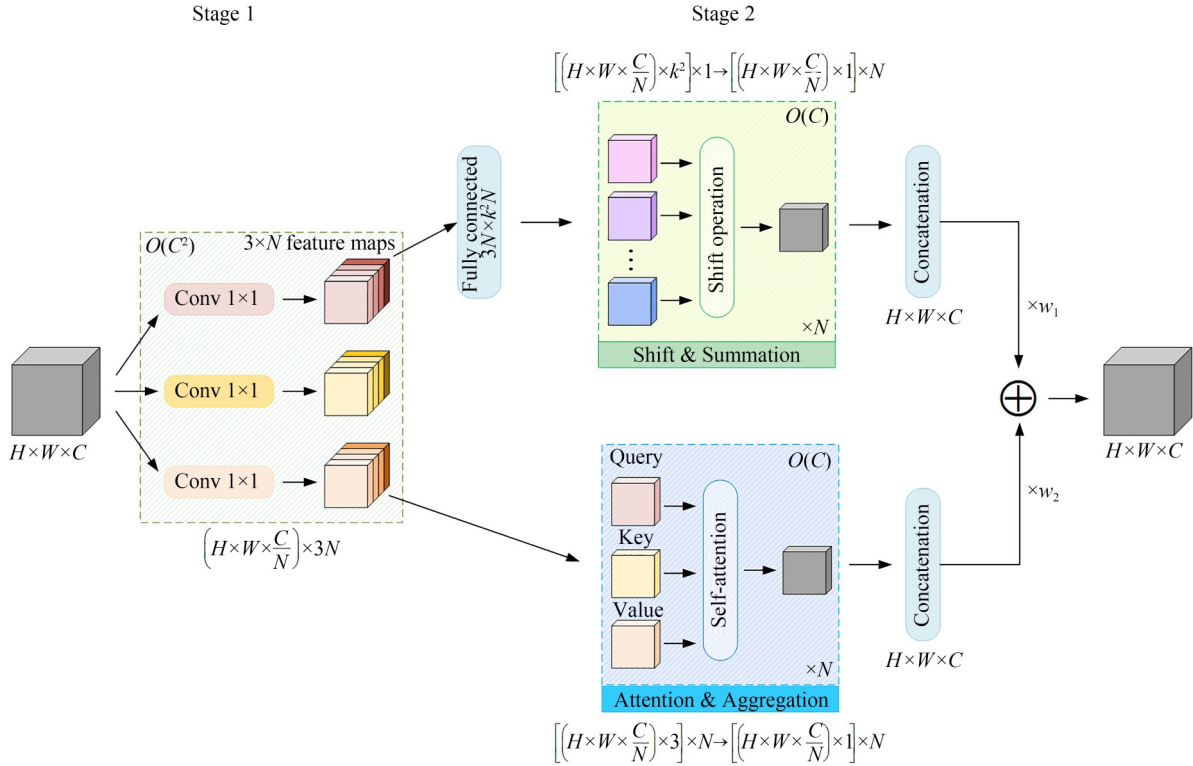


图 3 ACmix 结构图

Fig. 3 Overall structure of ACmix

特征增强环节共包括两个阶段。第一阶段是特征学习阶段,利用 3 个 1×1 的卷积对输入特征进行投影,用于重塑 N 个特征片段并得到包含 $3N$ 个特征图的中间特征集,从而将特征投射到更加深层次的空间。第二阶段为特征聚合阶段,通过两种路径完成信息收集。

在特征增强的过程中,首先对输入的特征图进行特征学习,将中间特征集成 N 组,每组包含 3 个特征,对应的 3 个特征映射分别作为 Query (查询)、Key (键) 和 Value (值),并且遵循自注意力模块的方式来收集信息。其中,Query 表示当前的关注点或信息需求,用于与 Key 进行匹配; Key 包含输入序列中各位置的标识信息,用于被 query 查询; Value 负责存储与 Key 对应的信息内容,并根据 Query 的匹配程度进行加权求和。设 $N_k(i, j)$ 表示以 (i, j) 为中心,空间宽度为 k 的局部像素区域,则 $A(W_q^{(l)} f_{ij}, W_k^{(l)} f_{ab})$ 表示 $N_k(i, j)$ 对

应的权重,具体如式(1)所示:

$$A(W_q^{(l)} f_{ij}, W_k^{(l)} f_{ab}) = \text{softmax}_{N_k(i, j)} \left(\frac{(W_q^{(l)} f_{ij})^T (W_k^{(l)} f_{ab})}{\sqrt{d}} \right), \quad (1)$$

其中: $W_q^{(l)}, W_k^{(l)}$ 分别为 Query 和 Key 的投影矩阵, d 为 $W_q^{(l)} f_{ij}$ 的特征维数, f_{ab} 为 $a, b \in N_k(i, j)$ 的像素张量。

在特征聚合阶段,首先利用 1×1 卷积将输入特征投影为 Query, key 和 Value, 然后进行包含注意力权重的计算以及值矩阵的聚合操作,具体如式(2)及式(3)所示:

$$q_{ij}^{(l)} = W_q^{(l)} f_{ij}, k_{ij}^{(l)} = W_k^{(l)} f_{ij}, v_{ij}^{(l)} = W_v^{(l)} f_{ij}, \quad (2)$$

$$g_{ij} = \parallel \left(\sum_{l=1}^N \sum_{a, b \in N_k(i, j)} A(q_{ij}^{(l)}, k_{ab}^{(l)}) v_{ab}^{(l)} \right), \quad (3)$$

其中: f_{ij} 和 g_{ij} 分别表示像素 (i, j) 输入和输出对应的张量, $W_q^{(l)}, W_k^{(l)}$ 和 $W_v^{(l)}$ 分别表示为 Query, Key 和 Value 的投影矩阵, $q_{ij}^{(l)}, k_{ij}^{(l)}$ 和 $v_{ij}^{(l)}$ 分别表示 Que-

ry, Key 和 Value 矩阵; $\parallel$ 表示 N 个注意力头输出的级联。接着将两个分支的输出结果进行加权融合, 得到输出结果 F_{out} , 具体的计算方法如式 (4) 所示:

$$F_{\text{out}} = w_1 F_{\text{att}} + w_2 F_{\text{conv}}, \quad (4)$$

其中: F_{att} 为自注意力路径的输出结果, F_{conv} 为卷积路径的输出结果, w_1 和 w_2 均为权重参数。在具体训练过程中, w_1 和 w_2 均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。最后, 将得到的特征图用步长为 1、尺寸为 3×3 的卷积核进行卷积操作, 从而得到最终的特征图 P2~P5。

2.2 多分支联合检测策略

在夜间低光照检测场景中, 虽然行人躯体部分受到的遮挡较为严重且行人与背景的对比度

较低, 但行人手部和腿部等特征却具有显著的可辨别性, 因此可用于辅助检测。本文设计的多分支联合检测策略为: 当特征提取网络获取输入图像的特征信息后, 首先利用区域建议网络生成行人候选框集合。然后构建行人轮廓检测分支、行人腿部检测分支和行人手臂检测分支作为辅助检测模块, 与行人整体检测分支联合检测得到初步的检测结果。最后利用本文设计的融合置信度惩罚策略的非极大值抑制算法筛选冗余检测框, 得到最终准确的行人检测结果。

2.2.1 构建辅助检测分支

行人轮廓、行人腿部以及行人手臂辅助检测分支包括三层卷积层和三层全连接层, 三个辅助检测分支的网络结构完全一致, 具体如图 4 所示。

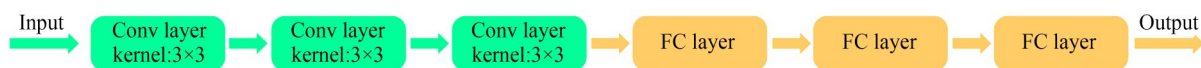


图 4 行人轮廓/腿部/手臂辅助检测分支的结构图

Fig. 4 Overall structure of auxiliary detection branches for pedestrian contour/leg/arm

每个辅助检测分支的输入包括两个部分, 一部分是由特征提取网络(SCMB-FPN)获取的特征图, 另一部分是由 Faster R-CNN 模型内置的区域建议网络 (Regional Proposal Network, RPN) 提出的高质量候选区域框。二者均通过感兴趣区域对齐 (Region of Interest Align, RoI Align) 操作截取候选区域的特征图。由于 RPN 网络提供的候选区域与行人目标的真实位置并非完全对齐, 实际上多数候选区域只能覆盖部分行人轮廓/腿部/手臂信息, 其包含对应特征的比例为 0~1 之间的线性数值。因此, 本文基于候选区域的位置信息在对应层级的特征图上截取相应的候选区域特征, 然后分支网络根据候选区域的特征对其包含的行人轮廓/腿部/手臂的比例进行计算与评估。其中, 对于行人轮廓特征而言, 获取的是一个完整的候选区域特征; 对于行人腿部特征而言, 需要通过截取建议区域的下半部分特征来获得; 对于行人手臂特征而言, 首先截取候选区域的上半部分特征, 然后在上半部分特征中, 按宽度将其平均分成 4 个部分, 接着截取最左边的 1 个部分和最右边的 1

个部分, 从而得到行人手臂特征信息。该提取策略在行人正面、背面姿态下均可有效捕获双臂区域特征, 从而有效促进检测精度的提升; 当行人呈现侧面姿态时, 所截取区域虽然无法精准对应手臂区域, 难以产生额外的性能增益, 但不会引入冗余干扰信息, 也不会对整体检测结果的准确性产生负面影响。在此情况下, 辅助检测分支的有效贡献主要由行人轮廓分支与行人腿部分支提供。

在实际模型训练过程中, 为了保证辅助检测分支的有效训练, 本文基于原始数据集的真实标注构建辅助检测分支对应的标注信息。对于行人轮廓分支, 根据候选区域与行人真实框的覆盖比例生成 $[0, 1]$ 区间的回归标签; 对于行人腿部和手臂分支, 通过几何计算来确定候选区域内对应的像素掩码, 并统计有效特征占比, 从而生成 $[0, 1]$ 区间的连续回归标签。辅助检测分支主要依据候选区域的特征进行辅助检测, 其收敛目标是一个 $[0, 1]$ 范围内的数值来评估候选区域内包含行人轮廓/腿部/手臂特征的比例。因此, 在实际训练过程结束后能够得到对应的检测分数。

而最终的候选检测框的置信度分数由原始检测网络得到的概率分数 P 和辅助检测分支得到的比例分数 C 两部分加权得到。设定辅助检测分支预测得到的行人轮廓/腿部/手臂的比例分数占最终置信度分数的权重为 μ , 则检测框的最终置信度分数 S 具体如式(5)所示:

$$S = (1 - \mu)P + \mu C, \mu \in [0, 1], \quad (5)$$

其中:对于权重参数 μ 的设置,过高或者过低均会影响算法的检测性能:若 μ 设置过低,则难以有效发挥辅助检测分支的作用;若 μ 设置过高,则会对行人整体检测分支输出的预测结果产生过度干扰,进而不利于整体检测精度的提升,因此本文选择通过对比实验来分析其最优取值。

2.2.2 融合置信度惩罚策略的非极大值抑制算法

在得到初步的检测结果集合后,需要对集合中包含的冗余框进行剔除。对于这一环节,传统非极大值抑制算法采用的策略是基于预设阈值来判定当前检测框是需要被保留的正确检测结果还是应当被剔除的冗余框,算法的实际检测效

果也与预设阈值的设置紧密关联。然而,对于夜间复杂场景下的行人检测任务而言,过高或者过低的预设阈值均对检测精度产生影响。若设置的预设阈值较低,则可能导致密集遮挡处的行人检测框被过度抑制,进而引起漏检率升高^[24];若设置的预设阈值较高,则可能因抑制强度减弱而导致冗余框剩余较多,进而引起误检率升高并影响检测精度^[25]。因此,仅依赖预设阈值来筛选检测框的传统非极大值抑制算法存在一定的局限性。对此,本文引入对检测框之间中心距离的考虑,设计了一种融合置信度惩罚策略的非极大值抑制算法(CPS-NMS),目的是在密集遮挡场景中尽可能多地保留目标的正确检测结果,从而为后续筛选环节提供充足的资源。

在 CPS-NMS 算法中,首先计算候选检测框 box_i 与置信度最高的检测框 box_{max} 的中心点之间的欧氏距离 $D(box_i, box_{max})$, 然后计算能够完全覆盖 box_i 与 box_{max} 的最小矩形的对角线长度 $L(box_i, box_{max})$, 则优化后的非极大值抑制过程具体如式(6)所示:

$$Score_i = \begin{cases} Score_i, & IoU(box_{max}, box_i) - \frac{D^2(box_i, box_{max})}{L^2(box_i, box_{max})} < N_i \\ Score_i \left[1 - IoU(box_{max}, box_i) + \frac{D^2(box_i, box_{max})}{L^2(box_i, box_{max})} \right], & IoU(box_{max}, box_i) - \frac{D^2(box_i, box_{max})}{L^2(box_i, box_{max})} \geq N_i \end{cases}, \quad (6)$$

其中: N_i 为非极大值抑制算法的预设阈值,参考复杂场景中主流行人检测模型的经验阈值设置^[26], 本文实验中设置 $N_i = 0.5$ 。 $Score_i$ 为目标分类得分。优化后的惩罚策略增加了对检测框之间中心点距离的考虑,通过降低遮挡严重检测框的置信度分数并进行重新筛选,从而有效地缓解了受遮挡目标的检测结果被错误抑制的问题,有助于降低行人目标的漏检率。

3 实验结果与分析

3.1 实验环境

本文在实验过程中所用的操作系统为 Ubuntu 18.04, 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 $\times 4$ 。在实际训练过程中,结合训练参数经验值以及参数调优处理^[9,22], 初始学习率

设置为 0.001, 动量因子设置为 0.937, 权重衰减设置为 0.000 5, 采用随机梯度下降优化器进行 50 个 epochs 的训练, batch size 设置为 16。

3.2 数据集与评估指标

本文在实验过程中选用 NightOwls^[27], LLVIP^[28] 和 NightSurveillance^[29] 三个行人检测数据集来验证所提算法的检测性能。其中, NightOwls 数据集中所有图片均在夜间采集且涵盖 4 个季节和不同天气的场景。由于图片整体对比度较低且噪声较多, 行人受遮挡现象也较为明显, 因此该数据集能准确地还原本文重点研究的夜间低光照复杂检测场景, 非常适合作为验证算法性能的数据集。LLVIP 数据集是专为低光照视觉检测任务而设计, 包括行人遮挡、尺度差异、光照条件变化等丰富场景, 因此适用于验证算法的综合性能。NightSurveillance 数据集由 16 个十字路口的夜间监控视频组成, 每个视频均逐帧切

割成图像,涵盖的场景多为夜间低光照环境,并且行人之间存在明显的尺度差异和遮挡,因此也可用于验证算法在夜间低光照场景中的检测性能。

在评估算法的检测性能时,本文结合行人检测数据集的实际场景特征与任务侧重差异化选取算法的评估指标: NightOwls 和 NightSurveillance 数据集以夜间监控密集行人、目标遮挡易漏检为主要特点,因此采用对数平均漏检率 MR^{-2} 作为评估指标,能够客观反映算法在低光照、复杂遮挡场景下的漏检抑制能力;对于 LLVIP 数据集,为综合衡量模型的定位精度与目标检出能力,本文选用召回率 Recall 和平均精度 AP 作为评估指标。各评估指标的具体计算方法如式(7)~式(9)所示:

$$AP = \int_0^1 Precision(Recall) dRecall, \quad (7)$$

$$MR^{-2} = \int_0^1 MR(FPPI) dFPPI, \quad (8)$$

$$Recall = \frac{TP}{TP + FN}, \quad (9)$$

其中, AP 值和 Recall 值越高、 MR^{-2} 值越低,说明算法的检测性能越出色。Precision, MR 和 FPPI 分别表示算法的准确率、漏检率和每张图片中的平均错检数, TP 表示预测结果和真实结果均为行人的情况, FN 表示预测结果为背景而真实结果为行人的情况。

3.3 对比实验结果与分析

在对比实验中,首先验证权重参数 μ 的取值对算法性能的影响。由于行人轮廓/手臂/腿部分支均作为辅助检测分支来进一步提升行人检测准确性,因此实验中设置权重参数 μ 的取值集合为 {0.05, 0.1, 0.15, 0.2, 0.25, 0.3}, 其余算法模型参数设置均保持一致。在 NightOwls 及 NightSurveillance 数据集上的实验结果如表 1 所示,在 LLVIP 数据集上的实验结果如表 2 所示。

分析表 1 和表 2 中的实验结果可知,当权重参数 μ 取值为 0.1 时,所提算法在三个数据集上的性能均为最优,因此设置权重参数 μ 为 0.1。

此外,本文基于相同的实验环境重新训练了融合 FPN 结构的 Faster R-CNN 算法并以此作为基线算法,同时将所提算法与一些主流行人检测算法进行性能对比分析,以此来验证所提算法的

表 1 NightOwls 及 NightSurveillance 数据集上权重参数 μ 取值的影响

Tab. 1 Influence of μ value on NightOwls and NightSurveillance dataset (%)

Value of μ	MR^{-2}	MR^{-2}
	(NightOwls)	(NightSurveillance)
0.05	15.6	20.6
0.1	15.5	20.4
0.15	15.5	20.5
0.2	15.7	20.6
0.25	15.8	20.8
0.3	15.8	20.8

表 2 LLVIP 数据集上权重参数 μ 取值的影响

Tab. 2 Influence of μ value on LLVIP dataset (%)

Value of μ	Recall	AP
0.05	86.0	90.5
0.1	86.2	90.8
0.15	86.2	90.6
0.2	86.1	90.5
0.25	86.0	90.3
0.3	86.0	90.3

有效性。本文选取的对照组算法包括基于积分通道特征与提升决策树框架的 ACF^[30]、Checkerboards^[31] 算法;在 Faster R-CNN 基础上通过改进候选区域生成、优化非极大值抑制、结合语义分割或引入困难样本挖掘策略的 RPN+BF^[32], Adapted Faster R-CNN^[33], Adaptive NMS^[26] 与 SDS-RCNN^[34] 算法;通过时序特征融合、局部上下文建模与增强特征学习等策略优化检测能力的 TFAN^[35], TLL^[36], Mamba-YOLO^[44], Cascade R-CNN^[46] 和 EGCL^[37] 等算法;在工业界被广泛应用的 YOLO 系列算法等。在 NightOwls 及 NightSurveillance 数据集上的对比实验结果如表 3 所示,在 LLVIP 数据集上的对比实验结果如表 4 所示。

根据表 3 和表 4 所示的对比实验结果可知,改进后的算法相比于基线算法取得了显著的性能提升。其中,在 NightOwls 和 NightSurveillance 数据集上 MR^{-2} 值分别降低了 2.8% 和 3.3%,在 LLVIP 数据集上 AP 值提升了 2.1%, Recall 值提升了 2.9%。此外,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果。由此可

表 3 NightOwls 及 NightSurveillance 数据集上的对比实验结果

Tab. 3 Results of comparative experiments on nightowls and nightsurveillance dataset (%)

Algorithm	MR ⁻² (NightOwls)	MR ⁻² (NightSurveil- lance)
Adaptive NMS ^[26]	16.4	21.6
ACF ^[30]	51.7	89.3
Checkerboards ^[31]	39.7	—
RPN+BF ^[32]	23.3	86.3
Adapted Faster R-CNN ^[33]	18.8	24.8
SDS-RCNN ^[34]	17.8	23.6
TFAN ^[35]	16.5	—
TLL ^[36]	16.4	22.9
EGCL ^[37]	15.9	—
YOLOv11 ^[38]	16.6	22.1
YOLOv12 ^[39]	16.9	22.3
YOLOv13 ^[40]	16.2	21.8
Faster R-CNN+FPN (Baseline)	18.3	23.7
Ours	15.5	20.4

表 4 LLVIP 数据集上的对比实验结果

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

Algorithm	Recall	AP
Adaptive NMS ^[26]	85.1	89.4
YOLOv11 ^[38]	81.0	88.1
YOLOv12 ^[39]	79.0	86.1
YOLOv5 ^[41]	81.2	88.3
YOLOv9 ^[42]	79.8	87.9
YOLOv10 ^[43]	79.1	87.4
Mamba-YOLO ^[44]	84.3	90.2
MDCFV _{it} ^[45]	83.6	90.4
Cascade R-CNN ^[46]	85.7	90.5
Faster R-CNN+FPN(Baseline)	83.3	88.7
Ours	86.2	90.8

知,所提算法能够有效提升对夜间低光照场景内行人的检测能力。在实际应用方面,对于安防监控、夜间巡视等对安全性要求较高的场景,平均精度和召回率的有效提升以及漏检率的有效降低,能够显著降低关键目标遗漏、危险区域漏判等风险,对于保障夜间行人出行安全、提高智能

监测系统的可靠性与实用性具有重要作用。

同时,为进一步验证所提算法对不同遮挡程度、不同尺度目标的检测性能与泛化能力,本文在对目标类型进行精细划分的 NightOwls 数据集上,分别测试了算法对小尺度子集(Small)和严重遮挡子集(Occlusion)的检测效果,实验结果如表 5 所示。

表 5 各算法在 NightOwls 小尺度与遮挡行人子集上的检测性能对比结果

Tab. 5 Detection performance of different algorithms on small and occlusion subsets in NightOwls dataset (%)

Algorithm	Small MR ⁻²	Occlusion MR ⁻²
Adaptive NMS ^[26]	24.5	46.7
YOLOv11 ^[38]	23.4	47.6
YOLOv12 ^[39]	23.8	48.5
YOLOv13 ^[40]	23.5	47.9
Mamba-YOLO ^[44]	21.8	45.6
Cascade R-CNN ^[46]	23.9	47.0
Faster R-CNN+FPN (Baseline)	25.1	49.3
Ours	21.2	45.1

根据表 5 所示的实验结果可知,本文改进后的行人检测算法在小尺度行人子集和严重遮挡行人子集上均取得了最优效果。相比于基线算法,对数平均漏检率分别降低了 3.9% 和 4.2%。相比于其他主流行人检测算法,所提算法也体现出较为明显的性能优势。由此可见,本文改进后的算法对于不同尺度与不同遮挡程度的行人均具有更优的检测效果,展现出良好的鲁棒性与场景适用性。

此外,为了更加直观地体现所提算法的有效性,实验中还列举了基线算法、一阶段检测算法中主流的 YOLOv11, YOLOv12 和两阶段检测模型中性能优异的 Adaptive NMS, Cascade R-CNN 以及本文改进后算法对于同一场景内行人目标的可视化检测结果,选取的测试图像中涵盖了行人遮挡、目标尺度差异以及行人与背景对比度较低的情况。可视化对比结果具体如图 5 所示。

从图 5 所示的行人检测结果可以看出,本文提出的改进算法取得了最优的检测效果。其余



图 5 所提算法与主流算法的检测结果对比图

Fig. 5 Comparison of detection results between the proposed algorithm and mainstream algorithms

对比算法的检测结果中均存在较为明显的漏检情况,漏检目标主要集中在远距离小尺度行人、严重遮挡行人以及低可见度行人等困难样本。本文算法中由于构建了多分支联合检测策略,利用行人轮廓、腿部、手臂对比度和可辨别性较强的特点来辅助行人检测,有效地缓解了夜间弱光照条件引起的目标漏检率升高问题,提升了算法对可见度较低行人目标的检测能力;同时,本文设计的特征提取网络 SCMB-FPN 能够获取丰富细致的多尺度特征信息,并强化了对行人和背景的区分能力,从而有效增强了算法对多尺度行人目标的检测能力,缓解了小尺度行人目标漏检严重的现象;此外,本文在算法的后处理环节对非极大值抑制策略进行了优化,旨在尽可能保留被遮挡目标的正确检测结果,从而有效地降低了遮挡行人的漏检率。

3.4 消融实验结果与分析

为了验证本文设计的各个改进模块的有效性,本文还在 NightOwls、LLVIP 和 NightSurveillance 三个数据集进行了消融实验,具体方式是在基线算法的基础上逐次添加融合自注意力与卷积混合模块的双向路径特征聚合网络(SCMB-FPN)、多分支联合检测策略和融合置信度惩罚策略的非极大值抑制算法(CPS-NMS),以此来验证各个改进模块对算法整体检测性能的提升效果,选用的评估指标仍为平均精度 AP、对数平均漏检率 MR^{-2} 和召回率 Recall。在 NightOwls 及 NightSurveillance 数据集上的消融实验结果如表 6 所示,在 LLVIP 数据集上的消融实验结果如表 7 所示。基于测试图像得到的消融实验可视化对比结果如图 6 所示。

根据表 6 和表 7 所示的消融实验结果可知,

在基线算法的基础上引入 SCMB-FPN 模块后,算法在 NightOwls 和 NightSurveillance 数据集上的 MR^{-2} 值分别降低了 1.0% 和 1.1%,在 LLVIP 数据集上召回率和平均精度分别提升了 1.1% 和 0.7%;进一步引入本文提出的多分支联合检测策略后,算法在 NightOwls 和 NightSurveillance 数据集上的 MR^{-2} 值分别降低了 1.3% 和 1.6%,在 LLVIP 数据集上召回率和平均精度分别提升

表 6 NightOwls 及 NightSurveillance 数据集上的消融实验结果

Tab. 6 Results of ablation experiments on the NightOwls and NightSurveillance dataset (%)

Algorithm	SCMB-FPN	多分支联合检测策略	CPS-NMS	MR^{-2} (NightOwls)	MR^{-2} (NightSurveillance)
Ours	✓	✓	✓	15.5	20.4
	✓	✓	×	16.0	21.0
	✓	×	×	17.3	22.6
Baseline	×	×	×	18.3	23.7

表 7 LLVIP 数据集上的消融实验结果

Tab. 7 Results of ablation experiments on the LLVIP dataset (%)

Algorithm	SCMB-FPN	多分支联合检测策略	CPS-NMS	Recall	AP
Ours	✓	✓	✓	86.2	90.8
	✓	✓	×	85.8	90.5
	✓	×	×	84.4	89.4
Baseline	×	×	×	83.3	88.7

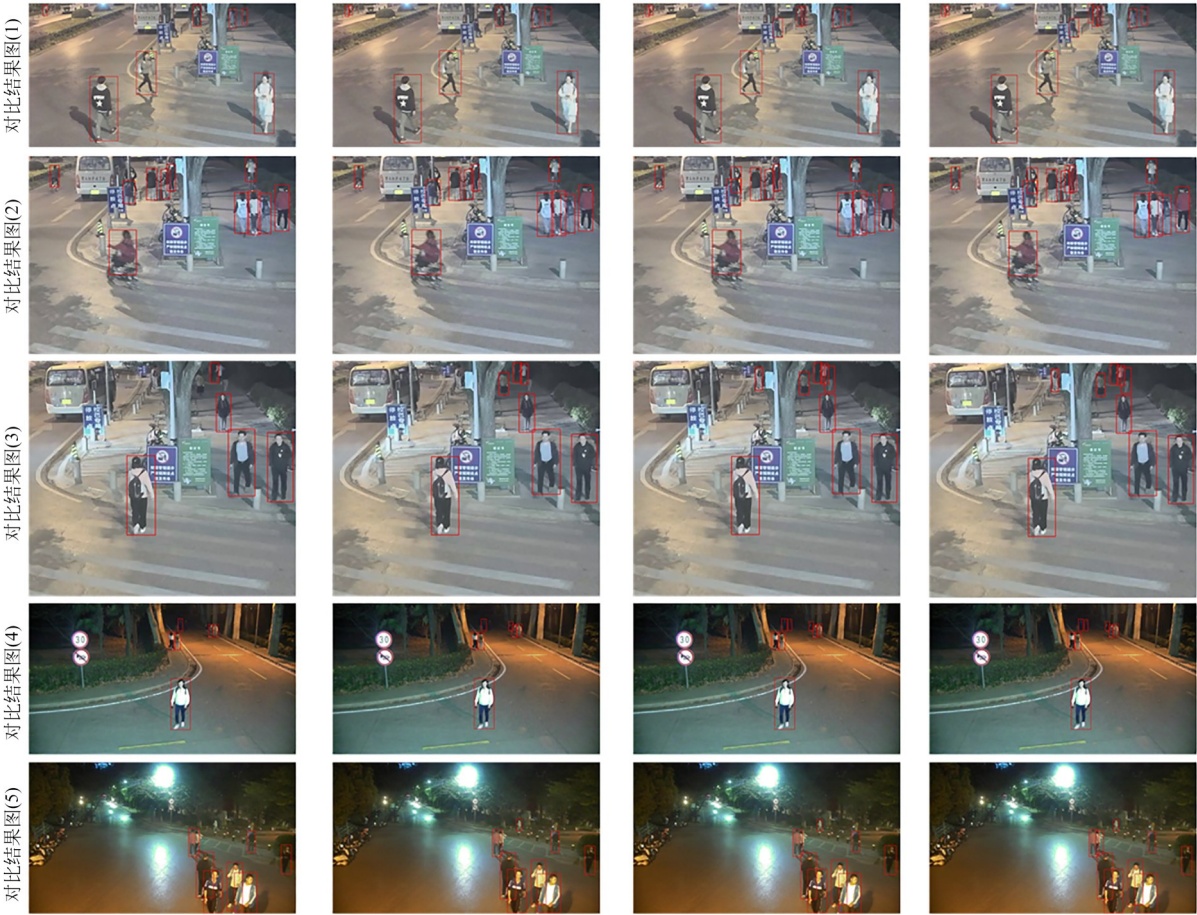




图 6 消融实验可视化结果图

Fig. 6 Visualization results of ablation experiments

了 1.4% 和 1.1%; 继续引入 CPS-NMS 算法后, 得到了算法改进后的完整版, 算法在 NightOwls 和 NightSurveillance 数据集上的 MR^2 值分别降低了 0.5% 和 0.6%, 在 LLVIP 数据集上的召回率和平均精度也进一步提升了 0.4% 和 0.3%。此外, 根据图 6 所示的消融实验可视化结果图也能够看出: (1) 在基线算法的基础上引入 SCMB-FPN 模块能够有效提升算法对小尺度行人目标的检测能力, 并抑制其漏检现象; (2) 进一步引入多分支联合检测策略后, 因光照较弱导致的行人漏检现象得到了有效抑制, 降低了低暗光照对行人检测造成的干扰; (3) 继续引入 CPS-NMS 算法后, 检测器对受遮挡行人目标的检测能力取得了进一步提升, 该类行人的漏检现象得到了有效缓解。由此可知, 本文设计的三个改进模块均对算法整体性能的提升具有增益效果, 能够有效提升算法的检测精度并降低目标漏检率。

4 结 论

本文针对导致夜间低光照场景中行人检测准确率降低的诸多因素, 基于 Faster R-CNN 算法进行改进, 提出一种基于多尺度特征融合的密集行人检测算法。首先, 在特征提取环节设计了一种融合自注意力与卷积混合模块的双向路径特征聚合网络, 通过对多尺度特征进行增强融合并提升行人与

背景的区分度, 以此来提升模型对不同尺度行人目标的检测灵敏度, 降低目标尺度差异对检测造成的影响; 其次, 在算法的核心检测环节提出了一种多分支联合检测策略, 利用构建的行人轮廓、腿部和手臂三个检测分支来辅助行人检测, 旨在降低弱光照和遮挡对检测造成的干扰; 然后, 在算法的后处理环节设计了融合置信度惩罚策略的非极大值抑制算法, 用于进一步降低密集遮挡处行人目标的漏检率。最后, 在 NightOwls, LLVIP 和 NightSurveillance 三个数据集上进行了对比实验和消融实验, 充分验证了所提算法的优越性。

作者贡献声明:

马晞茗: 设计行人检测算法, 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, 对论文撰写的整体结构和内容进行指导;

年伦: 调研并梳理论文算法相关文献, 对实验流程和实验方法进行指导;

吴迪: 论文内容审核与编辑写作, 对论文的初稿进行修改并指导;

于祥跃: 调研相关实验数据集, 对数据进行整理和分析;

李峥: 调研文献并参与对比实验验证与分析;

王瑶: 参与消融实验验证与分析。

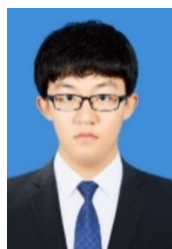
参考文献:

- [1] 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)
- [2] 高德勇,李惠敏,张学秀,等. 融合多尺度特征聚合分配的行人检测算法[J]. 北京邮电大学学报, 2025, 48(6): 13-18, 30.
GAO D Y, LI H M, ZHANG X X, *et al.* Dense pedestrian detection algorithm via feature gather-and-distribute and attention[J]. *Journal of Beijing University of Posts and Telecommunications*, 2025, 48(6): 13-18, 30. (in Chinese)
- [3] LUO S Y, SUN P, ZHU J, *et al.* GSON: a group-based social navigation framework with large multi-modal model[J]. *IEEE Robotics and Automation Letters*, 2025, 10(10): 9646-9653.
- [4] NIE L Z, LU M H, HE Z W, *et al.* Multispectral pedestrian detection based on feature complementation and enhancement[J]. *IET Intelligent Transport Systems*, 2024, 18(11): 2166-2177.
- [5] 王栋,罗亚波,张峰,等. CORM-YOLO: 低照度环境下交通行人检测算法[J]. 液晶与显示, 2025, 40(9): 1369-1380.
WANG D, LUO Y B, ZHANG F, *et al.* CORM-YOLO: Pedestrian detection algorithm in low-light conditions for traffic environments[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(9): 1369-1380. (in Chinese)
- [6] 曹洁,牛瑜,梁浩鹏. 基于优化权重的YOLOv7密集行人检测算法[J]. 液晶与显示, 2025, 40(3): 505-515.
CAO J, NIU Y, LIANG H P. Dense pedestrian detection algorithm based on YOLOv7 with optimized weights[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(3): 505-515. (in Chinese)
- [7] 陆燕,李富,祁铭瑞,等. 基于改进RT-DETR的夜间低光照行人检测算法[J]. 液晶与显示, 2025, 40(12): 1881-1893.
LU Y, LI F, QI M R, *et al.* Improved RT-DETR-based pedestrian detection algorithm for low-light conditions at night[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(12): 1881-1893. (in Chinese)
- [8] 张云佐,李文博,郭威,等. 面向多元场景的轻量级行人检测[J]. 光学精密工程, 2022, 30(14): 1764-1774.
ZHANG Y Z, LI W B, GUO W, *et al.* Light-weight pedestrian detection for multiple scenes[J]. *Opt. Precision Eng.*, 2022, 30(14): 1764-1774. (in Chinese)
- [9] 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.
- [10] XIE Y C, DU D F, BI M J. YOLO-ACE: a vehicle and pedestrian detection algorithm for autonomous driving scenarios based on knowledge distillation of YOLOv10[J]. *IEEE Internet of Things Journal*, 2025, 12(15): 30086-30097.
- [11] LI F, LU Y, ZHAO M, *et al.* BNE-DETR: nighttime pedestrian detection with visible light sensors via feature enhancement and multi-scale fusion[J]. *Sensors*, 2026, 26(1): 260.
- [12] CHEN Z Y, YAO H, LI W T, *et al.* Rethinking low-light knowledge for pedestrian detection in nighttime conditions[J]. *Engineering Applications of Artificial Intelligence*, 2025, 160: 111979.
- [13] LIU H, ZHANG Z Y, FENG B Y. POP-YOLOv8: an object detection framework for partially occluded pedestrians in nighttime traffic environments[J]. *Scientific Reports*, 2026, 16: 4841.
- [14] LI M X, HUANG S C. Fine-grained feature enhancement and occlusion-aware dense pedestrian detection in complex scenes[J]. *Multimedia Systems*, 2025, 32(1): 21.
- [15] YU W H, SI C Y, ZHOU P, *et al.* MetaFormer baselines for vision[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(2): 896-912.
- [16] 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.
- [17] YU Y X, LYU W T, GUO Q, *et al.* EDPDet: Efficient dense pedestrian detectors with multi-scale feature extraction, enhancement and aggregation[J]. *Expert Systems with Applications*, 2026, 298: 129933.

- [18] CHANG H, LI M X, HUANG S C, *et al.* GSPNet: a crowded pedestrian detection network integrating global context, spatial, and pose information [J]. *IEEE Access*, 2025, 13: 149373-149389.
- [19] 郑开魁, 吉康友, 李俊, 等. 基于全局特征聚焦与信息增强的遮挡行人检测算法[J]. *工程科学与技术*, 2025, 57(6): 104-118.
ZHENG K K, JI K Y, LI J, *et al.* Global feature focusing and information enhancement network for occluded pedestrian detection[J]. *Advanced Engineering Sciences*, 2025, 57(6): 104-118. (in Chinese)
- [20] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module[M]. *Computer Vision-ECCV 2018*. Cham: Springer International Publishing, 2018: 3-19.
- [21] LI X X, LI X. Insight any invisible: an occlusion perception method for robust pedestrian detection in crowded scenes[J]. *Expert Systems with Applications*, 2026, 297: 129464.
- [22] 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.
- [23] PAN X R, GE C J, LU R, *et al.* On the integration of self-attention and convolution [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 805-815.
- [24] 李翔, 何森, 罗海波. 面向遮挡行人检测的自适应收缩非极大值抑制方法[J]. *控制与决策*, 2024, 39(7): 2177-2185.
LI X, HE M, LUO H B. Adaptive shrinkage non-maximum suppression for occluded pedestrian detection[J]. *Control and Decision*, 2024, 39(7): 2177-2185. (in Chinese)
- [25] TANG Y, LIU M, LI B P, *et al.* OTP-NMS: toward optimal threshold prediction of NMS for crowded pedestrian detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 3176-3187.
- [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, 2020: 6452-6461.
- [27] NEUMANN L, KARG M, ZHANG S S, *et al.* NightOwls: a pedestrians at night dataset [C]. *Computer Vision-ACCV 2018*. Cham: Springer, 2019: 691-705.
- [28] JIA X Y, ZHU C, LI M Z, *et al.* LLVIP: A visible-infrared paired dataset for low-light vision [C]. 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. October 11-17, 2021, Montreal, BC, Canada. IEEE, 2021: 3489-3497.
- [29] WANG X, CHEN J, WANG Z, *et al.* When pedestrian detection meets nighttime surveillance: a new benchmark [C]. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. July 11-17, 2020. Yokohama, Japan. *International Joint Conferences on Artificial Intelligence Organization*, 2020: 509-515.
- [30] DOLLÁR P, APPEL R, BELONGIE S, *et al.* Fast feature Pyramids for object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(8): 1532-1545.
- [31] ZHANG S S, BENENSON R, SCHIELE B. Filtered channel features for pedestrian detection [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 7-12, 2015, Boston, MA, USA. IEEE, 2015: 1751-1760.
- [32] ZHANG L L, LIN L, LIANG X D, *et al.* Is Faster R-CNN doing well for pedestrian detection [C]. *Computer Vision - ECCV 2016*. Cham: Springer, 2016: 443-457.
- [33] 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.
- [34] BRAZIL G, YIN X, LIU X M. Illuminating pedestrians via simultaneous detection & segmentation [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017. Venice. IEEE, 2017: 4960-4969.
- [35] WU J L, ZHOU C L, YANG M, *et al.* Temporal-context enhanced detection of heavily occluded pedestrians [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*

- (CVPR). June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 13427-13436.
- [36] SONG T, SUN L Y, XIE D, *et al.* Small-scale Pedestrian Detection Based on Somatic Topology Localization and Temporal Feature Aggregation [EB/OL]. 2018: *arXiv*: 1807.01438. <https://arxiv.org/abs/1807.01438>
- [37] LIN Z B, PEI W J, CHEN F L, *et al.* Pedestrian detection by exemplar-guided contrastive learning [J]. *IEEE Transactions on Image Processing*, 2023, 32: 2003-2016.
- [38] KHANAM R, HUSSAIN M. YOLOv11: An Overview of the Key Architectural Enhancements [EB/OL]. 2024: *arXiv*: 2410.17725. <https://arxiv.org/abs/2410.17725>
- [39] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-Centric Real-Time Object Detectors [EB/OL]. 2025: *arXiv*: 2502.12524. <https://arxiv.org/abs/2502.12524>
- [40] LEI M Q, LI S Q, WU Y H, *et al.* YOLOv13: Real-Time Object Detection with Hypergraph Enhanced Adaptive Visual Perception [EB/OL]. 2025: *arXiv*: 2506.17733. <https://arxiv.org/abs/2506.17733>
- [41] GE Z, LIU S T, WANG F, *et al.* YOLOX: Exceeding YOLO Series in 2021 [EB/OL]. 2021: *arXiv*: 2107.08430. <https://arxiv.org/abs/2107.08430>
- [42] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]. *Computer Vision - ECCV 2024*. Cham: Springer, 2025: 1-21.
- [43] WANG A, CHEN H, LIU L H, *et al.* YOLOv10: Real-time end-to-end object detection [C]. *Advances in Neural Information Processing Systems* 37. December 10-15, 2024. Vancouver, BC, Canada. *Neural Information Processing Systems Foundation, Inc. (NeurIPS)*, 2024: 107984-108011.
- [44] WANG Z Y, LI C, XU H Y, *et al.* Mamba YOLO: a simple baseline for object detection with state space model [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39 (8): 8205-8213.
- [45] ZHANG H Y, ZHANG Q H, GONG Y F, *et al.* MDCFvit-YOLO: a model for nighttime infrared small target vehicle and pedestrian detection [J]. *PLoS One*, 2025, 20(6): e0324700.
- [46] CAI Z W, VASCONCELOS N. Cascade R-CNN: high quality object detection and instance segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(5): 1483-1498.

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