

改进 YOLOv10n 的密集行人小目标检测方法

陈海秀,陈子昂,宁馨,王思源,陆亚舒,卢海涛

(南京信息工程大学 自动化学院,南京 210044)

E-mail:hxchen@nuist.edu.cn

摘要:在针对固定摄像设备进行的行人检测任务中,由于小尺度行人占比高、分布密集和复杂背景的干扰,模型易受到遮挡、尺度变化等情况的影响,出现误检和漏检问题.为有效解决上述问题,提出了一种基于 YOLOv10n 改进的行人小目标检测算法.该算法在 C2f 模块中融入自适应点移动和卷积门控机制,提出了 C2f-MC 模块,显著提升了局部特征的捕捉能力和特征筛选能力;通过构建 SLC-P2 小目标特征增强金字塔结构,优化了颈部网络,增强模型提取微小特征的能力;引入基于点采样的 Dysample 上采样模块,有效减少了传统卷积方法对高分辨率特征引导的依赖,并显著提高了对遮挡小目标区域的敏感度;设计了一种轻量化检测头 LWD,进一步提升了模型在严重遮挡场景下对人群的识别能力.实验结果表明,与 YOLOv10n 相比,改进后的模型在 Crowd Human 和 WiderPerson 数据集上的精度分别提高了 4.0% 和 1.9%,且参数量仅为 2.0M,计算量为 10.2GFLOPs,能够在计算资源有限的环境中实现高精度部署.

关键词:行人检测;YOLOv10;特征选择;轻量化

中图分类号:TP391

文献标识码:A

文章编号:1000-1220(2026)04-0999-09

Improved YOLOv10n Dense Pedestrian Small Target Detection Method

CHEN Haixiu, CHEN Ziang, NING Xin, WANG Siyuan, LU Yashu, LU Haitao

(School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: In view of the high proportion of small-scale pedestrians, dense distribution and complex background interference in pedestrian detection tasks carried out by fixed camera equipment, the model is susceptible to occlusion, scale change and other problems of false detection and missing detection. An improved pedestrian small target detection algorithm based on YOLOv10n was proposed to solve this problem. Among them, C2F-MC module is proposed by combining adaptive point movement and convolutional gating mechanism in C2f module, which enhances the ability of local feature capture and feature screening. SLC-P2 small target feature enhancement pyramid structure was constructed to optimize the neck network and enhance the ability of the model to extract small features. The upsampling module of Dysample based on point sampling is introduced, which effectively eliminates the dependence of traditional convolutional methods on high-resolution feature guidance, and significantly improves the sensitivity of occluding small target region. A lightweight detection head (LWD) was designed to further improve the model's ability of crowd recognition in severely occluded scenes. The experimental results show that compared with YOLOv10n, the accuracy of the improved model on the Crowd Human and WiderPerson data sets is increased by 4.0% and 1.9%, respectively, and the parameter number is only 2.0M and the computation amount is 10.2GFLOPs, which can realize high-precision deployment in the environment with limited computing resources.

Keywords: pedestrian detection; YOLOv10; feature selection; light weight

0 引言

随着城市化进程加快及公共安全需求的增加,密集行人小目标检测已成为一个重要的研究方向,其核心在于利用先进的图像处理和机器学习算法,从各种环境中精确识别和分析密集人群中的个体,应用于安全监控、人流管理及灾难响应等领域.行人检测技术主要分为两阶段和一阶段检测算法.两阶段算法首先生成目标候选区域,再进行分类和特征提取,如 R-CNN、Faster R-CNN^[1];一阶段算法则实现端到端的行人检测,例如 SSD、YOLO^[2].其中,YOLO 网络因其高检测速度和

准确性而备受关注.但由于密集行人场景中存在目标尺度差异大、目标密集、小目标比例高及复杂背景等问题,常规的 YOLO 检测算法难以取得理想效果.

近年来,许多研究者在密集行人小目标检测中充分利用 YOLO 系列模型,取得了一系列成果:Tang 等人基于 YOLOv5 构建了遮挡筛选模块及位置注意检测层,以抑制背景干扰并提高检测精度^[3];Wang 等人使用 BiFPN 结构替代 YOLOv5 的原有颈部网络,优化了特征提取性能^[4];Zhang 等人在 YOLOv7 中引入 CBAM 注意力机制,并添加 NWD 模块来减少误检率^[5];Ma 等人通过可变形 DCNv2 卷积来改进骨

收稿日期:2025-03-21 收修改稿日期:2025-04-21 基金项目:国家自然科学基金项目(61302189)资助;江苏省研究生科研与实践创新计划项目(SJCX25_0500)资助. 作者简介:陈海秀(通信作者),女,1977年生,博士,副教授,研究方向为目标检测、图像融合;陈子昂,男,2000年生,硕士研究生,研究方向为目标检测;宁馨,女,2001年生,硕士研究生,研究方向为光学图像;王思源,女,2001年生,硕士研究生,研究方向为目标检测;陆亚舒,男,2001年生,硕士研究生,研究方向为电力系统优化调度;卢海涛,男,2001年生,硕士研究生,研究方向为图像融合.

干网络的同时,结合多尺度注意力 EMA 模块于 YOLOv8,成功提高了模型的特征提取能力^[6];Ren 等在 YOLOv8 中用 DBB 模块替换普通卷积,并引入特征选择模块 FSM,增强了模型的目标定位能力^[7];Li 等人在 YOLOv8 中引入了跨尺度特征融合模块(CCFM),以提升特征提取与融合的效果,同时减少计算冗余,此外在检测层增加了动态检测头(Dynamic Head),有效地降低了复杂环境下行人检测的误报率^[8]. Wang 则将 CIFNet 集成到 YOLOv8 的颈部网络,以提高特征信息的利用效率,并且聚焦于特征融合网络中的小尺度行人,从而提升了模型的检测性能^[9]. Chen 等为了改善 YOLOv8 算法在遮挡场景下误检与漏检率较高的问题,提出了引入 SimOTA 动态样本分配策略以及 Soft-NMS 后处理算法的改进方案^[10]. Li 等针对检测过程中复杂背景对小目标检测的干扰问题,设计了 BiP2 检测层,并采用 SE Attention 机制以聚焦特定通道信息,从而提高了复杂背景下小目标的检测精度^[11]. Zhang 等通过结合 Non-Local 和 Convolution 结构,显著增强了 YOLO 算法在特征提取阶段的能力,提升了特征图的细粒度表达效果^[12]. Li 等进一步解决了 YOLO 系列检测算法在多尺度场景下识别行人能力不足的问题,将 MSBlock 模块嵌入骨干网络,增强了检测器对多尺度特征的提取能力,并有效减少了因目标尺寸差异大而导致的漏检与误检问题^[13]. Hemant Kumar 等通过在 YOLOv9 模型中增加一个主干层,以捕捉图像中行人的更复杂特征信息,从而提升了行人检测的精度^[14].

尽管上述改进有效提升了 YOLO 系列模型的检测精度,降低了误检率,但在满足密集行人小目标检测中精确性和实时性的检测要求方面仍有不足.因此,本文在 YOLOv10n 网络基础上提出了改进的 MLD-YOLO(C2f-MC、SLC-P2、Dysample、LWD)网络,主要改进包括:

1) 嵌入 C2f-MC 模块来替换 YOLOv10n 骨干网络中的 C2f 模块.通过自适应点移动及卷积门控机制,增强对局部特征的捕捉能力,并选择性地增强特征提取过程中小目标特征的比重,提高特征提取的灵活性和准确性.

2) 在 YOLOv10n 原有的 PAN-FPN 结构基础上,结合 ASF 的核心思想,构建了 SLC-P2 小目标特征增强金字塔结构,以捕获更多小目标行人的细粒度特征,从而提高检测和定位的准确性,同时避免了传统方法直接引入 P2 检测层导致计算量和参数数量过大的问题.

3) 引入基于点采样的 Dysample 上采样模块,消除传统方法对高分辨率引导特征输入的需求,同时无需额外的 CUDA 支持,有效提升对遮挡小目标区域的敏感性,实现性能与效率间的平衡.

4) 提出了一种轻量化检测头 LWD,采用共享卷积模块替换传统检测头的常规卷积模块.在减少计算资源消耗的同时,显著提升了模型的检测性能.

1 YOLOv10 模型和改进

1.1 模型概述

YOLOv10^[15]于2024年5月23日由清华大学的研究团队发布,整合了此前多个版本的优势,并引入多项新功能和显

著改进.该系列提供6种不同尺寸的模型版本:YOLOv10n、s、m、b、l和x.随着模型深度和宽度的增加,这些版本在参数和计算量上也相应变化,以满足从资源有限设备到高性能计算平台的多种需求.YOLOv10n是最轻量级的版本,尤其适合在移动设备及不需GPU加速的CPU设备上部署.基于此,选择了以高分辨率和高效率著称的YOLOv10n作为行人小目标检测的基线模型,其架构主要分为骨干网络、颈部网络和检测头,分别负责特征提取、特征层融合以及分类和定位预测输出.

在本研究中,提出了一种基于YOLOv10n的改进模型:MLD-YOLO(如图1).该算法通过嵌入C2f-MC模块,替换YOLOv10n骨干网络中的C2f模块,构建新的特征提取网络.在PAN-FPN双向融合金字塔结构上结合ASF的理念,并添加P2检测层,开发了SLC-P2小目标特征强化金字塔结构,用以重构整个颈部网络.嵌入了Dysample上采样模块,利用缩放因子来增强恢复空间细节及保留高级语义信息的能力.在检测头部分采用了轻量化检测头LWD,通过共享卷积策略实现了性能与检测精度的优化平衡.在后续部分,本文将依照流程图的顺序详细介绍各改进模块的工作原理及实现细节,如Conv、C2f、C2fCIB等模块的工作流程与YOLOv10n保持一致,本文不再赘述.

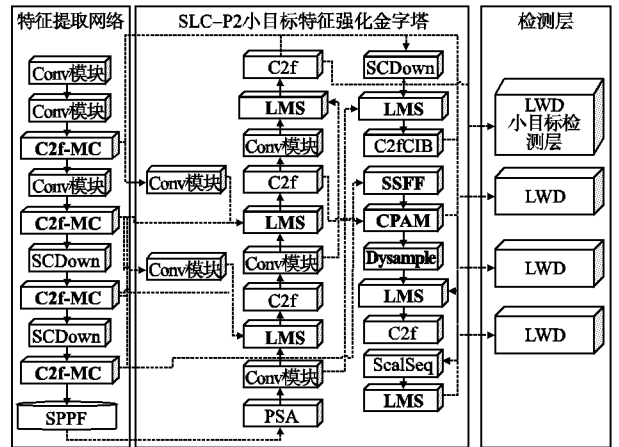


图1 MLD-YOLO架构

Fig. 1 MLD-YOLO architecture

1.2 C2f-MC 模块

在YOLOv10n模型中,C2f模块通过Bottleneck结构逐步提取特征.每个Bottleneck模块包含多个卷积层,但由于这些卷积层依赖于固定的点表示方式来执行常规卷积,这使得其在近似复杂函数时存在局限性.在处理来自固定监控设备的密集行人图像数据时,尤其是面对小目标尺度变化和复杂背景,现有卷积方法未能充分捕捉有效特征,这一点在小目标行人检测任务中表现尤为突出.

为解决这些问题,本文使用SMPR-CGLU模块替换Bottleneck模块,形成了C2f-MC模块(如图2(a)).在SMPR-CGLU模块中(如图2(b)),输入数据首先经过由自移动点构建的连续卷积模块(SMPConv),再使用SiLU激活函数进行激活,对激活后的数据通过CGLU模块来进行特征选择,以增加进入信息通道的小目标信息,最后使用DropPath进行正

则化以防止过拟合情况的发生。

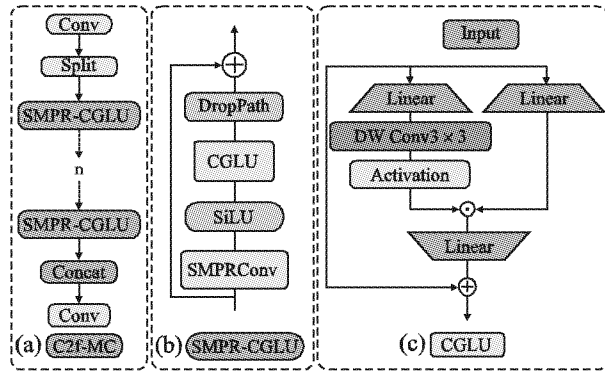


图2 C2f-MC 模块
Fig.2 C2f-MC module

SMPRConv 模块中使用自移动点法 (Self-moving point representation, SMPR) 来表示连续函数. 给定点 $x \in R^d$, 自移动点公式如下:

$$SMP(x; \phi) = \frac{1}{|N(x)|} \sum_{i \in N(x)} g(x, p_i, r_i) w_i \quad (1)$$

其中 $\phi = \{\{p_i\}_{i=1}^{N_p}, \{w_i\}_{i=1}^{N_p}, \{r_i\}_{i=1}^{N_p}\}$ 是一组可学习的参数, N_p 表示函数的点集. $p_i \in R^d$ 通过自动点表示 $w_i \in R^{N_c}$ 坐标, 每个点的表示还包含一个可学习的半径, $r_i \in R^+$ 是一个正实数.

自移动点实现了连续卷积算子. 在一维情况下, 连续卷积公式化为:

$$(f * SMP)(x) = \sum_{c=1}^{N_c} \int_R f_c(\tau) SMP_c(x - \tau) d\tau \quad (2)$$

其中 $f: R \rightarrow R^{N_c}$ 是输入函数, f_c 和 SMP_c 表示输入的第 c 个元素. 卷积运算符生成函数, 通过对整个 N_c 个通道求和得到滤波器响应.

一个自移动点对应一个卷积运算符, 而多个自移动点则用于构建卷积层, 以便生成多个输出通道. 不同于早期依赖于通道多层感知器 (Multilayer Perceptron, MLP)^[16] 的连续卷积, 该方法为层中每个卷积滤波器提供独立的参数, 赋予滤波器更多的自由度, 这增强了学习过程中的滤波器描述能力. 因此, SMPRConv 更擅长捕捉局部特征, 提高了特征提取的灵活性和精确性.

门控线性单元 (Gated Linear Unit, GLU)^[17] 是一种通道混合机制, 已被证明在多种图像处理任务中优于多层感知器 (MLP). GLU 由两个线性投影构成, 并通过逐元素相乘的方式结合, 其中一个投影通过门控函数激活. CGLU (如图 2 (c)) 则通过在 GLU 的门控分支激活函数前添加一个 3×3 的深度卷积, 使其结构符合门控通道注意力的设计理念, 并转化为基于最近邻特征的门控通道自适应感知机制.

在 CGLU 中, 每个特征根据其最近细粒度特征获得独特的门控信号, 这解决了 MLP 中全局平均池化过于粗糙的问题, 同时为特征提取过程中针对小目标特征的比重调整提供了灵活性和可能性. 此外, CGLU 的值分支支持与 MLP 和 GLU 相同的深度, 从而提高了反向传播的友好性. 当保持参数体积与卷积前馈 (Convolutional Feed-Forward Network, ConvFFN) 一致时, 扩展比为 R , 卷积核大小为 $k \times k$, CGLU 的

计算复杂度为 $2RHW C^2 + 2/3RHW C^2$, 小于 ConvFFN 的 $2RHW C^2 + RHW C^2$. 这些特性使得 CGLU 成为一种简单而强大的混合器, 特别适合在资源受限的固定监控设备上部署.

1.3 SLC-P2 小目标特征强化金字塔

通过固定监控摄像机捕获的图像往往面临复杂的背景挑战, 如密集的人群流动导致行人特征被严重遮挡、小目标行人的细节信息模糊、以及光照变化引发的噪声干扰等问题. 因此, 选择合适的颈部网络结构来调整感受野并减少背景噪声至关重要. YOLOv10n 中的 PAN-FPN 特征金字塔结构实现多尺度特征融合, 但由于骨干网络中不同层所提取特征分辨率之间存在差异, 常导致输出融合时对各特征贡献不均, 限制了不同尺度特征的利用. 此外, 骨干网络在特征提取阶段也存在尺度差异: 浅层大尺度特征图分辨率高、空间信息丰富, 利于小目标检测, 但语义信息稀疏; 深层特征图分辨率低、感受野大, 能捕获复杂语义特征, 适合复杂场景的检测.

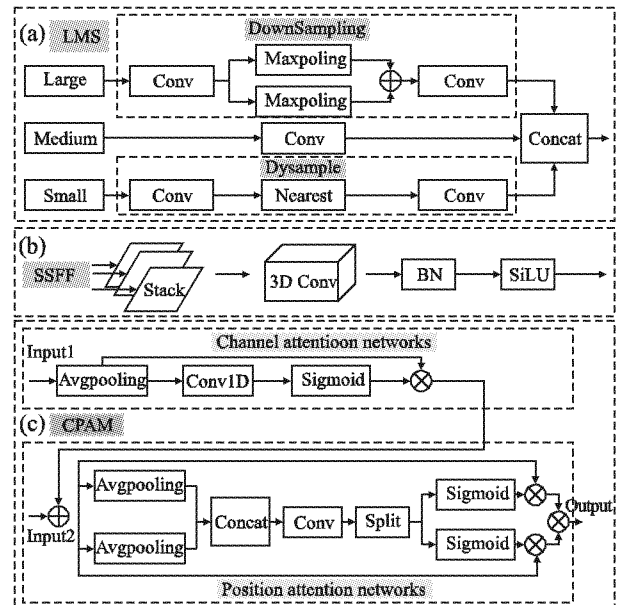


图3 SLC-P2 网络内置模块

Fig.3 SLC-P2 Built-in network module

为此, 本文结合 ASF^[18] 的核心思想, 在 PAN-FPN 双向融合金字塔网络的基础上新增了 P2 层以专门针对小目标检测, 提出了一种新的 SLC-P2 颈部网络结构 (如图 3). 该结构通过整合空间特征和多尺度特征, 从骨干网络的 P2、P3、P4 和 P5 层提取特征, 并由 SSFF (如图 3 (b)) 和 LMS (如图 3 (a)) 两个核心模块构成的特征融合架构, 为小目标检测提供互补信息. 其中, SSFF 模块旨在融合多尺度特征图的全局及高级语义信息, 而 LMS 模块则侧重于捕获小目标的局部细节.

为降低模型计算量与参数量, 实现计算资源的优化分配, 本文设计的 SSFF 模块通过融合 P2、P3 和 P4 的特征图来捕获不同尺寸与形状的行人空间特征, 并在在计算资源与检测精度间实现有效平衡. 在 SSFF 模块中, 各特征图先经归一化与上采样操作, 然后堆叠为三维卷积的输入, 以综合多尺度特征. 在 LMS 模块检测密集小目标人群头部及部分可见肢体时, 通过拼接不同尺寸特征图来提取细粒度信息. LMS 模块

的输出通过 PANet 结构整合至各特征分支,并在 P2 分支中与 SSFF 模块生成的多尺度特征进一步融合.此外,为增强对小目标的感知能力,本文在 P2 分支中引入了通道与位置自适应感知模块(CPAM)以同时利用高级多尺度特征与细节特征.CPAM(如图3(c))通过结合通道与位置自适应机制,优化信息通道的特征捕获并增强空间定位,从而有效提升检测精度.接下来分别介绍 SSFF、LMS 和 CPAM 这3个模块的具体内容.

1) 尺度序列特征融合模块(SSFF):

针对行人小目标检测中的多尺度问题,现有研究多采用特征金字塔结构,通常仅通过简单求和或连接的方式对特征进行融合,难以充分挖掘不同尺度特征图之间的相关性.相比之下,SSFF 模块能够更高效地整合多尺度特征图的高级语义信息与浅层细节信息.本文进一步构建了从主干网络(P2、P3和P4)生成的多尺度特征图的顺序表示,用于捕获具有不同细节与尺度的图像内容.特征图首先经过一系列高斯核卷积处理,生成不同尺度的平滑特征图,其公式如下:

$$F_{\sigma}(i, j) = \sum_u \sum_v f(i-u, i-v) \times G_{\sigma}(u, v) \quad (3)$$

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (4)$$

其中 f 表示二维特征图, F_{σ} 是通过标准差逐渐增加的二维高斯滤波器 G_{σ} 对 f 进行卷积平滑生成的输出特征图.然后水平堆叠这些尺度各异特征图,并通过三维卷积提取其尺度序列.但由于高斯平滑后的特征图分辨率不一致,本文采用 Dysample 的上采样方法将所有特征图对齐至 P2 的分辨率.

2) 三重特征编码器(LMS):

为了有效识别受到严重遮挡的小目标行人,可以通过增加更大尺度特征图,从而比较不同尺度下的行人特征变化.但主干网络中各特征层的尺寸存在差异,传统 FPN 融合机制仅对小尺寸特征图进行上采样,忽略了大尺寸特征图所蕴含的丰富细节信息.而 LMS 模块则通过分别拆分大、中、小特征,并额外添加大尺寸特征图来强化特征放大,提升对细节信息的捕获能力.

LMS 模块在特征编码前通过调整通道数量,保持与中尺寸特征图(Medium)一致,来实现不同尺度特征图的拼接.针对小尺寸特征图(Small),利用卷积调整通道数后通过 Dysample 方法进行上采样.该方法依据输入特征内容调整缩放因子,可以自适应地调整采样点步长,从而增强采样过程的灵活性与灵敏性,避免小目标行人特征的丢失.针对大尺寸特征图(Large),在卷积模块处理后将通道数规范为 $1C$,随后采用混合池化结构(最大池化与平均池化结合)进行下采样,在降低空间维度的同时实现平移不变性,以增强网络应对输入图像空间变化和平移的鲁棒性.最后,在通道维度上拼接大、中、小3个相同尺寸的特征图,计算公式如下:

$$F_{LMS} = \text{Concat}(F_l, F_m, F_s) \quad (5)$$

3) 通道与位置自适应感知模块(CPAM):

为了提取和整合颈部网络中不同尺度的特征,采用 CPAM 来整合 SSFF 捕获的多尺度特征信息和 LMS 捕获的局部信息.CPAM 由一个接收 LMS 输入的通道感知网络(Input1)和一个接收通道感知网络和 SSFF 输出叠加输入(Input2)的位置感知网络组成,下面具体介绍其实现过程:

①在通道自适应感知网络中,引入了保持维度不变的感知机制,以更高效地提取跨通道相互作用.该机制首先通过全局平均池化操作对所有 LMS 输入的通道进行特征汇总,而不降低其维度.随后,利用合适核大小的一维卷积,编码局部跨通道依赖关系并通过两个全连接层与非线性激活函数 Sigmoid 结合,生成每个通道的权重分布.

②在位置自适应感知网络中,首先将通道自适应感知网络的输出与作为位置自适应感知网络的输入 Input2 相结合以达到信息互补的作用,然后将输入的特征图在水平(p_h)方向和垂直(p_v)方向上进行池化,以保留特征图的空间结构信息.这种设计显著增强了 SSFF 在捕捉不同尺度特征图方面的能力,使模型能够灵活适应各类尺度的小目标检测任务,同时显著提高空间分辨率,并精准地定位小目标的位置.

1.4 Dysample 上采样

在 YOLOv10n 网络中,上采样采用最近邻插值法实现(如图4(a)),该方法以计算简便和资源需求低的特点受到青睐.然而,由于其依赖周围像素计算新像素值,这种方式可能导致过度平滑效应,从而丢失小目标特征.此外,这一方法还对特征图的完整性产生不利影响,并因额外的计算需求而减慢模型的推理速度.为克服上述限制,本文引入了 Dysample 方法.该方法不同于传统的基于核的上采样技术,其核心在于采用点采样策略通过生成采样位置并应用固定的双线性内核,消除了对高分辨率引导特征输入的依赖,同时无需额外的 CUDA 支持,在性能与效率之间实现了理想的平衡.

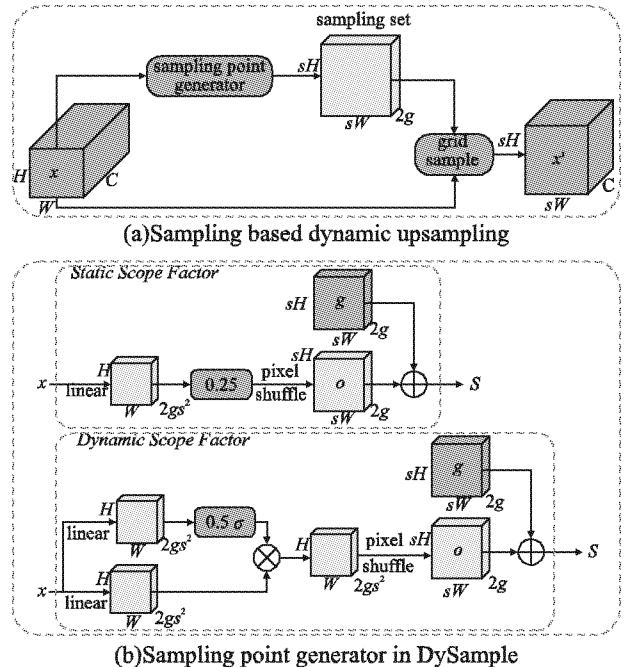


图4 Upsampling 和 Dysample

Fig. 4 Upsampling and Dysample

如图4(b)所示,在 Dysample 模块中,输入尺寸为 $H \times W \times C$ 的特征映射 x ,首先通过采样点生成器进行处理以学习特征偏移量 O .基于这些偏移,生成一个尺寸为 $sH \times sW \times 2g$ 的采样集 δ .随后,网格采样函数利用生成的偏移对 x 进行重新采样,得到尺寸为 $sH \times sW \times C$ 的上采样特征图 x' .

为了更高效地学习特征点的偏移,本研究提出了两种采样点生成器方案:静态缩放和动态缩放,具体如图 4(b)所示。

静态缩放因子通过一个固定值 0.25 对偏移范围进行限制。在线性层的基础上结合像素混洗操作,生成偏移量 O ,并将其添加到原始采样网格 G 以获得采样集合 δ 。这种方法控制了每个上采样点的局部搜索范围,避免采样位置过度重叠,从而减少输出特征图中的模糊边界和潜在误差传播。动态缩放因子通过输入特征动态生成,并为每个采样点提供适应性偏移范围。在这一过程中,输入特征映射 x 通过动态采样点生成器,分别经过两个分支处理:

第一分支通过线性层 L_1 将特征映射的每个通道映射到新的特征空间。随后,利用 Sigmoid 函数将偏移量归一化至 $[0,1]$ 区间,并乘以 0.5,生成以 0.25 为中心的动态缩放因子。

第二分支通过线性层 L_2 直接生成偏移量 O 。最终,第一分支生成的动态缩放因子对 L_2 输出的偏移量进行调整,从而实现偏移的动态调整,相关过程可用公式(6)表示:

$$O = 0.5 \text{Sigmoid}(\text{linear}_1(x)) \cdot \text{linear}_2(x) \quad (6)$$

动态缩放因子的引入使采样点的步长能够根据输入特征内容的变化自适应调整,从而提高采样过程的灵活性与敏感性。这种机制显著增强了模型对特征变化的适应能力,为构建更加鲁棒的小目标行人检测系统提供了支持。

1.5 轻量级检测头 LWD

YOLOv10n 检测头采用将分类和定位任务分离的解耦头结构存在一些局限。首先,每个检测头均通过两个 Conv3×3 和一个 Conv1×1 进行特征提取,导致检测头的计算开销较大。其次,YOLOv10n 解耦头中原有的 BN (Batch Normalization) 层在归一化过程中,其统计信息的计算依赖于批次大小,而在小批次情况下,统计均值和方差的计算可能不稳定,导致较高的误检率。最后,解耦头中将分类和定位任务分离的设计仅能基于单一尺度进行检测,忽略了其余尺度特征对检测性能的贡献,这非常影响人流密集区域对小目标行人的检测准确率。

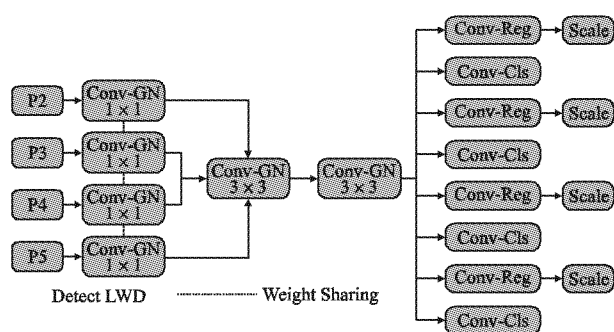


图 5 轻量级检测头 LWD

Fig. 5 Lightweight detection head LWD

为解决上述问题,本文提出了一种全新的 LWD 检测头,其结构如图 5 所示。原检测头中的常规卷积被共享卷积替代,有效减轻了多余的计算开销。同时 LWD 中新设计的 GN (Group Normalization) 层通过将特征通道分组,分别计算每组内的均值和方差并进行归一化,避免了对批次大小的依赖。

此外,共享卷积策略显著降低了模型计算开销,提升了对资源受限设备的适用性。

在检测任务中,LWD 检测头接收来自 P2 至 P5 层的输入特征后,首先采用 1×1 的共享卷积增加通道间的信息交互;随后,利用两个 3×3 的共享卷积聚合信息,减少冗余并增强相邻特征的学习能力。最后,将提取的特征输入分类和回归检测头,通过 Scale 层进行特征缩放,进一步提升多尺度特征的表达能。

2 实验与分析

实验平台的操作系统配置为 Windows 11,硬件采用 Intel® Core™ i5-13400F 处理器,配备 32G 的运行内存,GPU 型号为 Nvidia GeForce RTX 4060 Ti,显存容量为 16G。CUDA 版本为 12.4,而 Torch 版本为 1.17.2。在本文的实验中,模型训练设置保持一致,输入图像的尺寸为 640×640 ,批量大小设定为 12,优化器为 SGD,初始学习率为 0.01,并在相同的条件下执行 300 次训练迭代。

2.1 数据集

2.1.1 Crowd Human 数据集

Crowd Human^[19] 数据集是一个专注于密集场景下的行人检测任务的大规模人群图像数据集。数据集中包含约 24000 张高分辨率图像:其中 15000 张用于训练,5000 张用于测试,另有 4370 张构成验证集,总计标注了超过 47 万个行人实例。该数据集中目标标注密度极高,遮挡现象尤为显著,为复杂环境下的行人检测研究提供了极具挑战性的测试基准。其标注细致、场景丰富,能够帮助评估模型在高密度目标分布和遮挡情况下的性能表现。

2.1.2 WiderPerson 数据集

WiderPerson^[20] 数据集是专为拥挤场景中的行人检测任务设计的重要基准数据集。该数据集包含 13382 张图像,涵盖多样化的背景环境,具备显著的目标遮挡特征和复杂的人群分布。图像分为 8000 张训练集、1000 张验证集和 4382 张测试集,共标注约 40 万个目标实例。数据集的标注类型覆盖广泛,包括行人、骑行者、部分遮挡目标、假人和密集人群。为实现与 Crowd Human 数据集的兼容,去除假人部分,统一将剩余目标归为行人类别。作为一个标注精细、场景丰富的行人检测数据集,WiderPerson 提供了大量具有挑战性的实例,有助于探索模型在处理遮挡现象和密集人群分布时的表现,为行人检测算法的泛化能力和鲁棒性验证提供了宝贵资源。

2.2 评价指标

本文采用参数规模 (Parameters, P)、召回率 (Recall, R)、计算复杂度 (GFLOPs)、mAP50、mAP50-95 作为本次实验的评价指标,以此来衡量改进的有效性。P 和 R 的计算方法如下,分别对应公式(7)和公式(8),其中,TP 表示真正例;FN 表示假负例;FP 表示假正例。

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

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

平均精度 (mAP) 是评估目标检测模型性能的重要指标,用于衡量模型在所有类别上的检测精度。其计算方法是对每

个类别的平均精度进行求和,然后除以类别总数,以获得整体性能的量化结果.具体计算方式见公式(9):

$$mAP = \frac{\sum_{i=1}^k AP_i}{k(class)} \quad (9)$$

在模型性能评估中,交并比(Intersection over Union, IoU)被广泛用于衡量预测框与真实标注框的重叠程度.根据不同的IoU阈值,预测结果可被判定为正确或错误.本研究采用两种平均精度(mAP)指标进行评估:mAP50(IoU阈值固定为0.5时的平均精度)和mAP50-95(IoU阈值从0.5到0.95),以全面反映模型的检测性能.

2.3 实验结果与分析

2.3.1 LWD 对比实验

轻量级检测头(LWD)通过共享卷积减轻了模型的运行开销,显著提高了模型的运行效率的同时,保持了较高的检测性能.为验证其性能,本文将LWD与动态检测头((Dynamic Head, Dyhead)^[21])和基线模型YOLOv10n的检测头(Detect)进行了对比分析,实验结果见表1.实验均在YOLOv10-n模型的基础上进行,从结果来看,添加LWD以后的mAP50、mAP50-95、P和FLOPs分别为77.8%、48.4%、1.9M和6.3G.相比基线模型的Detect,LWD在mAP50和mAP50-95上分别提升了0.3%和0.4%,同时显著降低了参数量.与Dyhead检测头相比,LWD在保持较高检测精度的同时,进一步实现了模型的轻量化.

表2 MLD-YOLO 消融实验

Table 2 MLD-YOLO ablation experiment

C2f-MC	SLC-P2	Dysample	LWD	mAP50/%	mAP50-95/%	Params/M	FLOPs/G
√				78.0	48.4	2.0	6.5
	√			80.9	52.9	2.3	10.7
		√		78.1	49.0	2.3	6.5
			√	77.8	48.4	1.9	6.3
√	√			81.1	53.0	2.2	10.6
√	√	√		81.4	53.2	2.3	10.7
√	√	√	√	81.5	53.4	2.0	10.2

2.3.3 对比实验

为了评估MLD-YOLO在轻量化的同时是否能保持卓越的检测能力,研究中将其与目前主流目标检测算法(包括SSD、Faster R-CNN及YOLO系列)进行了对比,结果汇总于表3.

与SSD算法(特征提取网络采用ResNet)相比,MLD-YOLO在mAP50上提升了14.2%;相比Faster R-CNN,则提升了5.6%.尽管两阶段检测方法通常具有精度优势,但MLD-YOLO通过对网络结构的优化,成功在轻量化设计下实现了对这些复杂模型的精度超越.进一步与单阶段检测算法中的YOLO系列进行对比,结果表明MLD-YOLO在检测精度上超越了YOLOv5n、YOLOv7tiny、YOLOv8n和最新的YOLOv11n,甚至超过了更高性能的YOLOv8s.然而,尽管在精度上优于YOLOv8s,MLD-YOLO的参数量甚至不到其1/5,计算量则降低了近2/3.此外,与当前主流的行人检测的改进算法如:RO-YOLOv9、GSG-YOLOv8、DED-YOLOv8、YOLOv10n-(HG-NetV2,efficientVit)相比,MLD-YOLO的轻量化设计不仅实现了高精度,还显著降低了模型复杂度及计算成本,这使其在计

2.3.2 消融实验

为了验证所提改进模块对检测效果的影响,本文将YOLOv10n作为基线模型,使用C2f-MC模块、SLC-P2小目标特征强化金字塔、Dysample上采样和轻量级检测头LWD在Crowd Human数据集上进行消融实验.通过表2的消融实验结果可以发现,在网络结构中引入SLC-P2小目标特征强化金字塔和Dysample上采样模块后,模型在mAP50/%和mAP50-95/%上分别实现了3.4%和4.9%的提升.然而,这些增强措施虽然有效提升了模型性能,但不可避免地增加了模型的参数量和计算复杂度.

表1 检测头对比实验

Table 1 Detection head comparison experiment

	mAP50/%	mAP50-95/%	Params/M	FLOPs/G
Detect	77.5	48.0	2.2	6.5
Dyhead	78.4	48.7	2.7	7.7
LWD	77.8	48.4	1.9	6.3

为应对上述问题,本文提出用C2f-MC模块替代YOLOv10n骨干网络中的C2f模块,同时以轻量级检测头LWD替换原始的Detect解耦头,从而显著减小了模型参数和计算开销.最终,改进后的MLD-YOLO在Crowd Human数据集上与基线模型相比,mAP50/%提升了4.0%,mAP50-95/%提升了5.4%.尽管模型的计算复杂度略高于YOLOv10n,但显著的性能提升表明改进策略的有效性.

算资源受限的设备上具备良好的部署潜力.

2.3.4 可视化实验

为全面评估MLD-YOLO算法在复杂场景中的检测能力,本研究从Crowd Human数据集中选择了3种典型的小目标场景进行实验分析,这些场景的特点是头部或肢体区域的尺寸不足图像总面积的1%,具体包括严重遮挡场景、远距离拍摄场景和夜晚小目标场景.相应的检测结果已在图6中展示.

在图6(a)中,严重遮挡场景的挑战在于后排队人群存在明显的遮挡情况,导致后排队人群的头部与身体被大幅遮挡.YOLOv10n能够较为准确地识别前排的大目标,但在检测被遮挡的后排小目标时,性能有所下降.相比之下,MLD-YOLO在处理受遮挡的后排目标时,显著提高了对full body和head类别的识别数量,并且其置信度较基线模型有了显著提升,从而增强了对遮挡目标的敏感性,能够较为精准地识别和定位被遮挡的目标.

在图6(b)的远距离拍摄场景中,图像中包含多个小目标人群,这些目标位于距离摄像机较远的地方,且前排人群的遮

挡进一步加大了后排小目标的检测难度. YOLOv10n 在此类场景中出现了漏检和误检现象,尤其是在中后排人群的头部和全身区域. 相较而言,MLD-YOLO 在头部类别的检测中表

现出了明显的优势,成功检测到更多的目标,且检测框的数量和置信度也得到了显著提高,显示出其在远距离小目标行人检测中的敏感性.

表 3 MLD-YOLO 与 YOLOv10n 各类别人检测精度对比

Table 3 Comparison of pedestrian detection accuracy between MLD-YOLO and YOLOv10n

模 型	mAP50/%	mAP50-95/%	Params/M	FLOPs/G
SSD	67.3	33.4	34.3	41
Faster R-CNN	75.9	-	40	207.1
YOLOv5n	74.9	46.1	2.4	7.1
YOLOv7tiny	76.2	48.1	5.9	13.2
YOLOv8n	75.4	47.6	3.0	8.1
YOLOv8s	79.8	52.7	11.1	28.7
YOLOv10n (baseline)	77.5	48.0	2.2	6.5
YOLOv11n	75.1	47.3	2.6	6.3
RO-YOLOv9 ^[22]	75.8	47.7	7.2	26.8
GSG-YOLOv8 ^[23]	79.4	51.8	3.03	32.7
DED-YOLOv8 ^[24]	77.2	47.9	3.4	10.3
YOLOv10n-HGNetV2 ^[25]	75.8	48.9	2.4	6.9
YOLOv10n-efficientVit ^[25]	78.0	48.9	4.0	9.4
MLD-YOLO	81.5	53.4	2.0	10.2

在图 6(c)中的夜晚小目标场景, YOLOv10n 未能准确检测到远处小目标人群,特别是在低光照条件下,其检测效果较差. MLD-YOLO 在此场景中的表现显著优于基线模型,成功

检测到更多目标,且各个类别的目标数量均有所增加,置信度也大幅提升,展现了其在夜间环境下对小目标行人的高效检测能力.

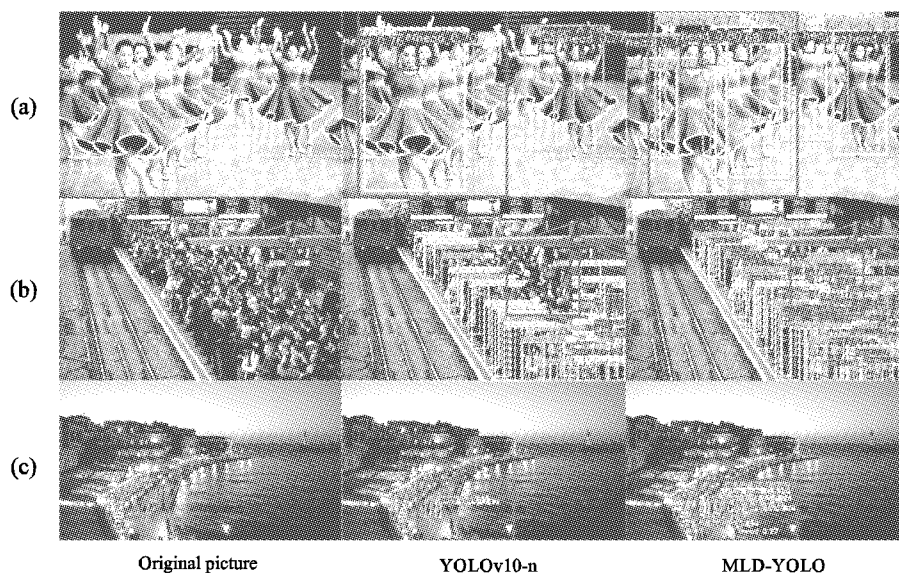


图 6 复杂场景下改进模型与基线模型的检测效果对比

Fig. 6 Comparison of detection effect between improved model and baseline model in complex scenarios

为了更全面地评估 MLD-YOLO 的改进效果,本研究分别对骨干网络、颈部网络、上采样模块、检测头进行了深入分析,并采用 GradCAMPlusPlus^[26]方法生成热力图,以便可视化检测结果. 热力图中的暗区域表示特征图中置信度较低的部分,而亮区域则代表置信度较高的区域. 相关结果如图 7 所示. 在骨干网络方面,MLD-YOLO 显示出比 YOLOv10n 更广泛的特征关注区域,这一改进得益于 C2f-

MC 模块在有效利用通道和空间位置关系方面的优异表现. 对于颈部网络,通过小目标特征强化金字塔结构 SLC-P2 以及 Dysample 模块的高效上采样,MLD-YOLO 在捕捉和整合小目标行人的细粒度特征方面表现突出,进一步提高了对人群聚集区域的关注和覆盖能力. 至于检测头,MLD-YOLO 新增的 LWD 检测头采用共享卷积策略,增强了通道间信息的交互以及相邻特征的学习能力,相比解耦头 Detect, LWD

更加注重被遮挡人群的深层次特征,从而提升了模型对遮挡小目标的识别敏感性.

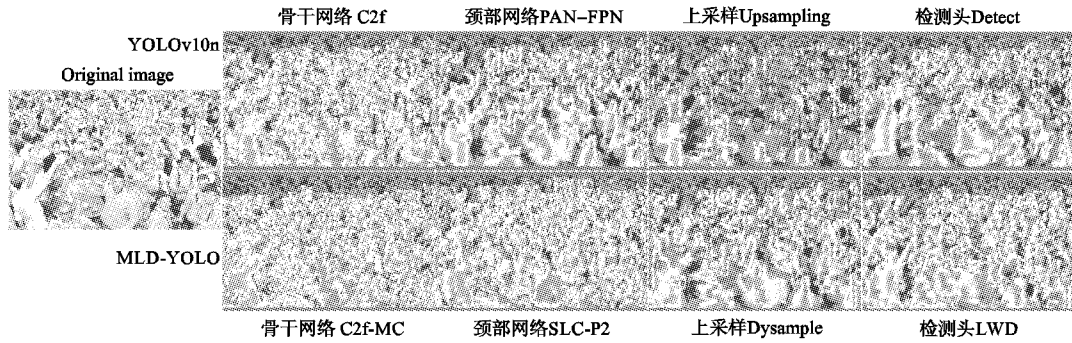


图7 密集小目标人群检测效果热力可视化

Fig. 7 Thermal visualization of the detection effect of dense small target population

2.3.5 模型泛化性实验

为了进一步评估 MLD-YOLO 算法的泛化能力,本研究在公开的 WiderPerson 数据集上进行了全面的测试与分析.实验在与先前相同的硬件平台和参数配置下进行,以确保结果的可比性.实验结果列于表4,MLD-YOLO 在准确率和召

回率方面均显著优于 YOLOv10n,分别提高了1.6%和3.2%.在检测精度上,MLD-YOLO 在 mAP50 和 mAP50-95 上分别提升了1.9%和2.5%.这些结果充分验证了所提出的改进算法在减少误检和漏检方面的优势,并且在密集行人检测任务中表现出更高的精度和鲁棒性.

表4 MLD-YOLO 与 YOLOv10n 各类别行人检测精度对比

Table 4 Comparison of pedestrian detection accuracy between MLD-YOLO and YOLOv10n

模型	P	R	mAP50/%	mAP50-95/%	Params/M	FLOPs/G
YOLOv10n	86.1	78.8	88.5	62.9	2.2	6.5
MLD-YOLO	87.7	82.0	90.4	65.4	2.0	10.2

3 结束语

本文针对复杂场景中因人群遮挡或拍摄距离较远等因素导致小目标行人检测存在漏检与误检问题,提出一种基于 YOLOv10n 改进的密集行人小目标检测算法:MLD-YOLO.采用 C2f-MC 模块实现以替代 YOLOv10n 骨干网络中的 C2f 模块;构建了 SLC-P2 小目标特征强化金字塔,并结合 Dysample 上采样模块对颈部网络进行重构;添加了轻量化检测头 LWD,以提升模型对遮挡小目标区域的响应能力,从而有效降低了漏检率与误检率.

实验结果表明,尽管 MLD-YOLO 在计算复杂度上有所增加,但在性能上显著超越了基线模型.在 Crowd Human 数据集上,MLD-YOLO 的 mAP50 和 mAP50-95 分别提高了4.0%和5.4%,而在 WiderPerson 数据集上,这两项指标的提

升则为1.9%和2.5%.这些结果证明了 MLD-YOLO 在多个数据集上的出色表现,特别是在密集人群场景中,展现了更高的检测精度与鲁棒性.未来研究方向将集中于进一步提升模型的计算效率,增强在极端遮挡和低光照条件下对小目标的适应性,并扩展其应用场景,以更全面地验证泛化能力.

References:

[1] Huang J, Fu M, Tang T, et al. Overview of target detection algorithm based on deep learning [C]//2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE), IEEE, 2024: 1174-1178.

[2] Li B, Jiang W, Gu J. Research on target detection algorithm based

on deep learning technology [C]//International Conference on Power Electronics, Computer Applications (ICPECA), IEEE, 2021: 137-142.

[3] Tang H, Li Z. Pedestrian detection algorithm based on improved YOLO v5 [C]//9th International Conference on Intelligent Computing and Signal Processing (ICSP), IEEE, 2024: 83-86.

[4] Wang X, Wei R, Zou R. Research on pedestrian detection algorithm based on improved YOLOv5 [C]//3rd International Conference on Electronic Information Engineering and Computer Science (EIECS), IEEE, 2023: 1109-1112.

[5] Zhang Z, Zhang X, Tian H, et al. Application of improved YOLOv7 algorithm in pedestrian detection [C]//5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), IEEE, 2024: 956-961.

[6] Li X, Ma L. DED-YOLOv8: dense pedestrian detection algorithm based on YOLOv8 [C]//4th International Symposium on Computer Technology and Information Science (ISCTIS), IEEE, 2024: 545-548.

[7] Ren Y, Li Y, Gao H, et al. A pedestrian detection algorithm based on improved YOLOv8 [C]//IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), IEEE, 2024: 75-81.

[8] Li J, Liu F, Gu M. Pedestrian detection based on improved YOLOv8 [C]//7th International Conference on Machine Learning and Natural Language Processing (MLNLP), IEEE, 2024: 1-4.

[9] Wang N, Huang S, Liu X. CIFE-YOLOv8: a dense pedestrian detection model based on improved YOLOv8 [C]//3rd International

- Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE), IEEE, 2024:399-403.
- [10] Chen T, Chen J, Gao T, et al. A novel scene-aware pedestrian detection in dense scenes [C]//27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2024:1316-1321.
- [11] Li Y, Zhang X, Zhou Z. DBS-YOLO: a vehicle detection model based on improved YOLOv8 for UAV aerial scenes [C]//5th International Conference on Computer Vision, Image and Deep Learning (CVIDL), IEEE, 2024:1432-1438.
- [12] Zhang H, Yang K, Liu H, et al. Small-scale pedestrian detection using fusion network and probabilistic loss[J]. IEEE Access, 2024, 12:42509-42520.
- [13] Li W, Zhao Y, Zhang Y, et al. Pedestrian detection algorithm based on improved YOLOv5 [C]//7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), IEEE, 2024:177-182.
- [14] Kumar H, Mamoria P, Dewangan D K. Optimized pedestrian detection leveraging YOLOv9: a thorough deep learning framework for autonomous vehicle system [C]//IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT), 2024:1-6.
- [15] Liu R. Enhanced motorbike helmet detection using YOLOv10 with attention-based modifications [C]//4th International Conference on Computer Science and Blockchain (CCSB), IEEE, 2024:114-118.
- [16] Gong S, Zhang D, Du S, et al. An empirical analysis and research on the prediction of stock trends based on the MLP neural network model [C]//International Conference on Artificial Intelligence and Blockchain Technology (AIBT), IEEE, 2021:28-33.
- [17] Komatsu T, Togami M, Takahashi T. Sound event localization and detection using convolutional recurrent neural networks and gated linear units [C]//28th European Signal Processing Conference (EUSIPCO), IEEE, 2021:41-45.
- [18] Kang M, Ting C M, Ting F F, et al. ASF-YOLO: a novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing, 2024, 147:105057, doi:10.1016/j.imavis.2024.105057.
- [19] Huang J, Liang J, Cao Y. Pedestrian target detection model based on improved Yolov8 [C]//5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), IEEE, 2024:376-381.
- [20] HUANG K, QI Z J, WANG J M, et al. Dense pedestrian detection model based on improved YOLOv8 [J/OL]. Computer Engineering, 1-11, <https://www.ecice06.com/CN/10.19678/j.issn.1000-3428.0069026>, 2025-03-21.
- [21] GONG Y X, GAO L, ZHANG H, et al. Comparative study on the effectiveness of various improved models of YOLOv8 algorithm [J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2024, 42(4):470-479.
- [22] LIAO Y H, WAN X J, ZHAO Z Z, et al. RO-YOLOv9 vehicle pedestrian detection algorithm [J/OL]. Computer Engineering and Applications, 1-15, <https://cea.ceaj.org/CN/10.3778/j.issn.1002-8331.2409-0128>, 2025-03-21.
- [23] LI B B, FU C Y. Low light pedestrian detection algorithm based on improved YOLOv8 [J/OL]. Journal of Xihua University (Natural Science Edition), 1-8, <https://aj.xhu.edu.cn/xhdxxbzkbcn/article/pdf/preview/10.12198/j.issn.1673-159x.5422.pdf>, 2025-03-21.
- [24] Li X, Ma L. DED-YOLOv8: dense pedestrian detection algorithm based on YOLOv8 [C]//4th International Symposium on Computer Technology and Information Science (ISCTIS), IEEE, 2024:545-548.
- [25] Huang J, Liang J, Cao Y. Pedestrian target detection model based on improved Yolov8 [C]//5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), IEEE, 2024:376-381.
- [26] ZHENG S J, WANG G C. Fine-grained image classification based on ConvNeXt heat map localization and contrast learning [J]. Journal of Computer Science, 2023, 50(10):119-125.
- 附中文参考文献:**
- [20] 黄 昆, 齐肇建, 王娟敏, 等. 基于改进 YOLOv8 的密集行人检测模型 [J/OL]. 计算机工程, 1-11, <https://www.ecice06.com/CN/10.19678/j.issn.1000-3428.0069026>, 2025-03-21.
- [21] 龚宇翔, 高 林, 张 豪, 等. YOLOv8 算法多种改进模型的有效性对比研究 [J]. 湖北民族大学学报 (自然科学版), 2024, 42(4):470-479.
- [22] 廖炎华, 万学俊, 赵周洲, 等. RO-YOLOv9 车辆行人检测算法 [J/OL]. 计算机工程与应用, 1-15, <https://cea.ceaj.org/CN/10.3778/j.issn.1002-8331.2409-0128>, 2025-03-21.
- [23] 李宝兵, 符长友. 基于改进 YOLOv8 的低光照行人检测算法 [J/OL]. 西华大学学报 (自然科学版), 1-8, <https://aj.xhu.edu.cn/xhdxxbzkbcn/article/pdf/preview/10.12198/j.issn.1673-159x.5422.pdf>, 2025-03-21.
- [26] 郑世杰, 王高才. 基于 ConvNeXt 热图定位和对比学习的细粒度图像分类研究 [J]. 计算机科学, 2023, 50(10):119-125.