

文章编号:1673-9469(2024)04-0103-10

DOI:10.3969/j.issn.1673-9469.2024.04.014

基于深度学习的密集行人检测场景算法研究

马明杰^{1,2}, 马小陆^{1,2*}, 唐得志^{1,2}, 赵远^{1,2}, 齐晶晶³, 瞿元⁴

(1. 安徽工业大学 电气与信息工程学院, 安徽 马鞍山 243002; 2. 安徽工业大学 芜湖技术创新研究院, 安徽 芜湖 241002; 3. 安徽达尔智能控制系统股份有限公司, 安徽 芜湖 241002; 4. 奇瑞汽车股份有限公司, 安徽 芜湖 241006)

摘要: 针对密集行人检测场景存在目标尺度过小以及目标遮挡等问题, 提出一种基于改进YOLOv7的密集行人检测算法。首先在特征提取网络引入 MobileNet 注意力模块, 减少模型计算量和增强特征提取能力; 其次在特征融合网络加入 BepC3 模块, 提升了行人多尺度特征融合的能力; 最后采用 WD-Loss 作为定位损失函数, 提高模型检测的定位精度。在 Wider-Person 拥挤行人检测数据集上进行训练和验证, 实验结果表明改进后的算法模型 AP₅₀ 精度达到了 0.784, 领先原 YOLOv7 算法 0.031。

关键词: 密集行人检测; YOLOv7; MobileNet; BepC3; WD-Loss

中图分类号: TP391

文献标识码: A

Algorithm Research on Dense Pedestrian Detection Scene Based on Deep Learning

MA Mingjie^{1,2}, MA Xiaolu^{1,2*}, TANG Dezhi^{1,2}, ZHAO Yuan^{1,2}, QI Jingjing³, QU Yuan⁴

(1. School of Electrical and Information Engineering, Anhui University of Technology, Maanshan, Anhui 243002, China; 2. Wuhu Technology and Innovation Research, Anhui University of Technology, Wuhu, Anhui 241002, China; 3. Anhui Dar Intelligent Control System Co., Ltd., Wuhu, Anhui 241002, China; 4. Chery Automobile Co., Ltd., Wuhu, Anhui 241006, China)

Abstract: A dense pedestrian detection algorithm based on the improved YOLOv7 is proposed to address the issues of small target scale and occlusion in dense pedestrian detection scenarios. Firstly, the MobileNet attention module is introduced into the feature extraction network to reduce the model computation and enhance feature extraction capabilities. Secondly, the addition of the BepC3 module in the feature fusion network enhances the ability of pedestrian multi-scale feature fusion. Finally, WD-Loss is used as the localization loss function to improve the localization accuracy of the model detection. Training and validation were conducted on the Wider-Person crowded pedestrian detection dataset, and the experimental results showed that the improved algorithm model AP50 achieved an accuracy of 0.784, leading the original YOLOv7 algorithm by 0.031.

Key words: dense pedestrian detection; YOLOv7; MobileNet; BepC3; WD-Loss

行人检测是目标检测方向的一个重要分支, 主要使用深度学习的方法对目标行人进行识别和定位, 多应用于智慧交通^[1]、无人驾驶^[2]、人流量监测^[3]等领域。现有的目标检测主要分为两类:

以 R-CNN 和 Faster R-CNN 为代表的二阶段检测算法和以 YOLO 和 SSD 为代表的一阶段检测算法。其中一阶段检测算法是直接对图像预测目标的位置和类别, 具有较快的检测速度, 相较于二阶段检

收稿日期: 2023-08-22

基金项目: 国家自然科学基金资助项目(62172004, 61872004); 安徽省科技重大专项(202003a05020028); 安徽省高校协同创新项目(GXXT-2023-020); 芜湖市核心技术攻关科技计划项目(2022hg10); 芜湖市科技计划项目(2023kx17)

作者简介: 马明杰(1999-), 男, 安徽六安人, 硕士研究生, 从事嵌入式和智能网联方面的研究。

* 通信作者: 马小陆(1979-), 男, 安徽芜湖人, 博士, 教授, 从事嵌入式、车联网和服务机器人方面的研究。

测更适用于实时检测任务。但是密集行人检测场景较为复杂,往往会出现行人遮挡、姿态多样和尺度多变等问题,因此对一阶段检测算法提出了更高的要求。

在密集行人检测中,行人目标尺度多变的问题尤为明显,尤其是小目标包含的特征信息较少,难以进行特征提取,并且在训练阶段难以覆盖不同尺度的行人导致模型泛化能力下降,会对较小的行人目标产生错误预测。针对此问题, Li 等^[4]在 YOLOv3 网络加入改进的 HOG 特征提取方法,突出小目标行人轮廓特征,减少背景信息干扰。Li 等^[5]在 YOLOv4 结构中利用深度可分离卷积减少了模型参数量,又引入 CBAM 注意力模块增强网络特征表达能力。另外在实际检测中出现行人目标被遮挡的现象,这导致行人丢失部分特征信息,使模型无法精确定位完整的边界框。针对此问题,李翔等^[6]基于 YOLOv3 提出了一种聚拢损失函数和一种高分辨率特征金字塔,利于区分不同深度的重叠目标。Li 等^[7]在 YOLOv7 加入高分辨率的特征金字塔结构,加强相邻子特征之间的区分度,增强对小尺寸行人的特征提取能力。

综上所述,针对密集行人检测场景中目标尺度较小、局部遮挡等现象,导致误检和漏检的问题,本文以 YOLOv7^[8]为基础模型框架,首先在特征提取网络融入 MobileNet 注意力模块,减少网络模型参数量且增加小目标信息权重;其次在颈部网络引入 BepC3 模块,增强对多尺度行人特征的融合能力;最后采用 Wasserstein Distance Loss^[9](WD-Loss)作为定位损失函数,增强对检测边框的预测回归。为验证本文改进模型的性能,在 WiderPerson 数据集^[10]下进行了消融实验和横向对比实验,最终结果表明,本文改进模型在密集行人目标检测任务中表现出更好的性能。

1 YOLOv7 算法介绍

为提升复杂场景中密集行人检测的精度与速度,本文选取 YOLOv7 算法作为基础目标检测模型框架。YOLOv7 框架主要由进行图片预处理的输入模块(Input)、进行特征提取的骨干网络(Backbone)、进行特征融合的颈部网络(Neck)和输出检测结果的头部预测网络(Head)四个部分构成。YOLOv7 整体网络架构如图 1 所示。

其中 Input 部分使用 Mosaic 数据增强、自适应

描框计算、自适应图片缩放对图像进行预处理;处理后的特征图进入 Backbone 网络(主要有 ELAN 模块、MP 模块和 SPPCSPC 模块)进行特征提取,得到三个大小不同的特征矩阵;随后进入 Neck 部分(主要有 MP 模块和 ELAN-W 模块)对提取的特征进行特征融合得到大、中、小三种尺寸的特征;最后,在 Head 网络中对融合的特征进行处理并生成三个大小不同的预测特征图,可预测九种不同尺寸的目标。

YOLOv7 损失函数分为分类损失、目标损失和定位损失三个部分,其中定位损失使用 Complete Intersection over Union-loss(CIoU-loss)来计算,CIoU 将目标与锚框之间的距离、重叠率、尺度以及惩罚项都考虑进去,其损失函数如式(1)所示:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha\nu \quad (1)$$

其中 IoU 为预测框和真实框之间的交并比, $\rho^2(b, b^{\text{gt}})$ 代表了预测框和真实框的中心点的欧式距离, gt 代表真实值(ground truth), c 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离, α 是权重参数, ν 是用来衡量宽高比一致性的参数。

2 本文改进

密集人群场景下,行人检测会面临部分被遮挡、目标较小等问题,基于深度学习的行人检测是根据提取到的特征信息进行检测,而遮挡现象会对目标行人的特征进行干扰,同样较小的目标行人分辨率低,特征信息较少,都导致特征信息无法较好地提取,不利于后续的检测。针对此复杂场景中行人检测的精度与速度的提升,本文提出改进 YOLOv7 算法包含三个部分:

首先借鉴了 MobileNetV3^[11] 网络结构,将 YOLOv7 的骨干网络引入深度可分离卷积模块(Depthwise Separable Convolution, DSC),大大减少了计算量,并且融入了注意力机制(Squeeze and Excitation Networks, SE)模块^[12],能够优化特征图信息的权重关系,增加对小目标行人的关注度,减少了模型参数,加快了计算速度。

其次借鉴了 Rep-PAN 结构^[13]与 CSPDarknet53 模块^[14]的设计思路,将 BepC3 模块融入 YOLOv7 的颈部网络,增强对多个尺度的语义信息和位置信息进行融合,能够加快推理速度的同时,

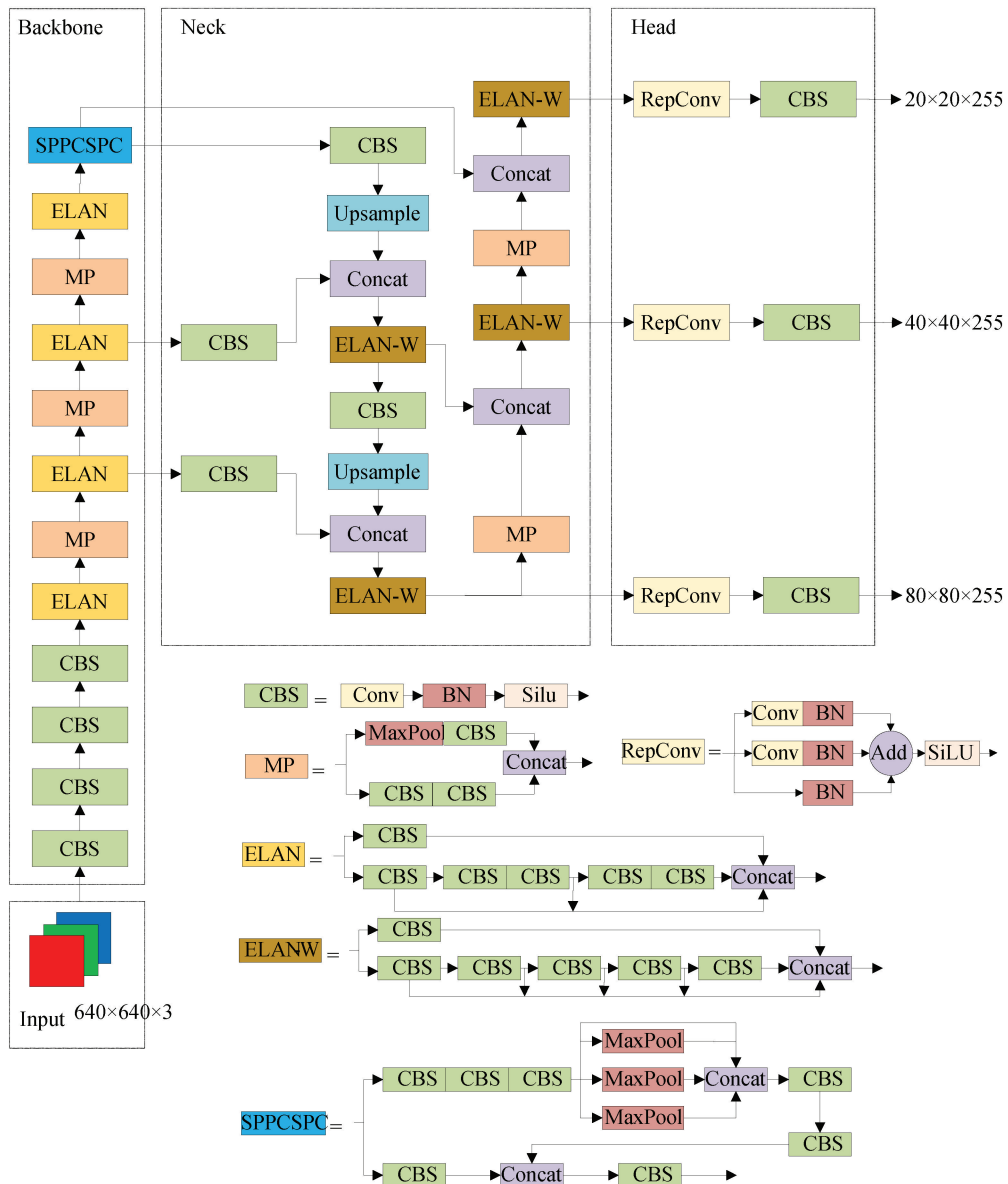


图 1 YOLOv7 网络架构图

Fig. 1 YOLOv7 network architecture diagram

保持较好的多尺度特征融合能力。

最后借鉴 Normalized Wasserstein Distance (NWD) 度量方式,采用 WD-Loss 损失函数作为定位损失函数,减小对微小目标位置偏差的敏感度,优化目标框的回归准确度。

2.1 DSC 深度可分离卷积模块

针对复杂场景的行人目标检测,优化网络模型结构的同时,训练和运行速度也受到限制,而卷积层占据了大部分的推理时间。轻量化网络主要对卷积层进行优化,如图 2 所示深度可分离卷积^[15]由深度卷积 (Depthwise Convolution, DW) 和逐点卷积 (Pointwise Convolution, PW) 组成。

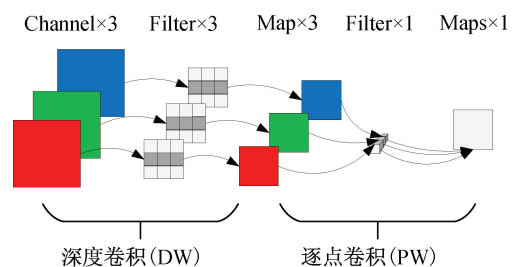


图 2 深度可分离卷积示意图

Fig. 2 Schematic diagram of depthwise separable convolution

图 2 表示先进行深度卷积,每个输入通道都有一个对应的卷积核进行卷积操作,不会改变特征图像的深度,再进行拼接得到输出特征图。第二步进行逐点卷积,将上一层的特征图在深度上进

行卷积操作,生成新的特征图。这种结构可以实现跨通道的信息融合并减少输出通道,提升计算速度。假设输入特征图的大小为 $D_w \times D_h \times M$, 输出特征图的大小为 $D_w \times D_h \times N$ (M 和 N 表示不同的深度), 卷积核大小为 $D_k \times D_k$, 默认卷积步距 1, 则深度可分离卷积与普通卷积的计算量之比如式(2)所示:

$$\frac{D_k \times D_k \times M \times D_w \times D_h + M \times N \times D_w \times D_h}{D_k \times D_k \times M \times N \times D_w \times D_h} = \frac{1}{N} + \frac{1}{D_k^2} \quad (2)$$

由式(2)可见,与标准卷积相比,深度可分离卷积的运算量下降为原来的 $\frac{1}{N} + \frac{1}{D_k^2}$, 使用了更少的空间代价和更少的时间代价实现了一样的特征提取效果。

2.2 SE 注意力模块

SE 模块包含压缩和激励两部分,把输入特征图的信息压缩,加大重要信息特征权重,缩小其他信息权重,再与原特征图融合得到最终的特征信息,该过程能够有效地优化特征图权重关系。图 3 为 SE 模块示意图。

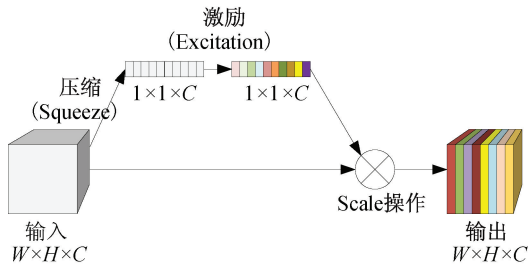


图 3 SE 模块示意图

Fig. 3 Schematic diagram of SE module

输入特征图大小为 $W \times H \times C$, 首先第一步是压缩,进行一个全局平均池化操作得到 $1 \times 1 \times C$ 的向量;接着就是第二步激励,经过两个全连接层并用 ReLu 激活函数和 h-swish 激活函数得到有关特征图权重的向量;最后进行第三步 Scale 操作,将原有的特征图乘上各通道权重向量,得到新的特征图数据。通过 SE 模块,可以给每个特征图不同的权重,增加对小尺度特征的关注度,由此提升针对小目标行人的检测效果。

2.3 MobileNet 注意力模块

MobileNetV3 是 MobileNet 系列网络的第三个版本,相较于前两个版本,MobileNetV3 引入了注意

力机制、深度可分离卷积以及倒残差结构^[16],进一步提升了模型在前向传播效率与性能。因此,本文参考 MobileNetV3 网络,利用 DSC 模块、SE 模块构建了 MobileNet 注意力模块。MobileNet 注意力模块结构如图 4 所示。

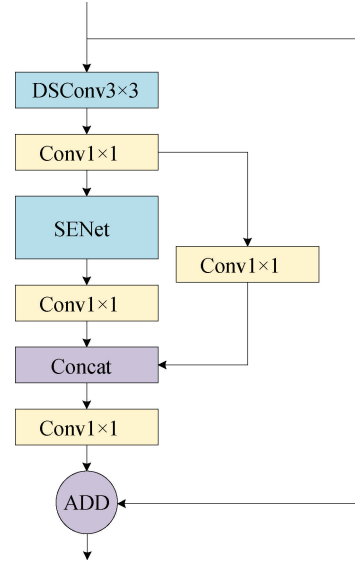


图 4 MobileNet 模块结构图

Fig. 4 MobileNet module structural diagram

2.4 改进骨干网络

YOLOv7 仅采用高效聚合网络 ELAN 模块以及 MP 模块配合常规卷积进行特征提取,导致在特征提取阶段模型难以区分特征重要程度。因此本文采用 MobileNet 注意力模块替换骨干网络中的第一个 ELAN 模块,为骨干网络引入注意力机制,使网络在进行特征提取时可以更多地聚焦于重要小目标特征,改进的骨干网络结构如图 5 所示。

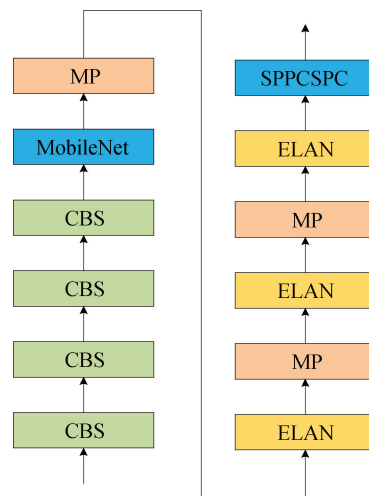


图 5 改进的骨干网络结构

Fig. 5 Improved backbone network structure

2.5 可重参数化卷积

YOLOv7 采用可重参数化卷积^[17]作为输出层之前的信息处理卷积层。可重参数化卷积由一个 1×1 卷积、一个 3×3 卷积以及短接分支分别连接批量归一化模块 (Batch Normalization) 构成,当模型进行训练时,采用该多分支卷积模块进行训练,当模型训练完毕进行前向传播时,多分支卷积被重参数化为一个 3×3 卷积,它允许网络在不牺牲模型性能的前提下,动态地调整卷积核的形状和参数数量,使模型可以有效适应不同的任务和数据特征,提高模型的训练效率与泛化能力。其结构图如图 6 所示。

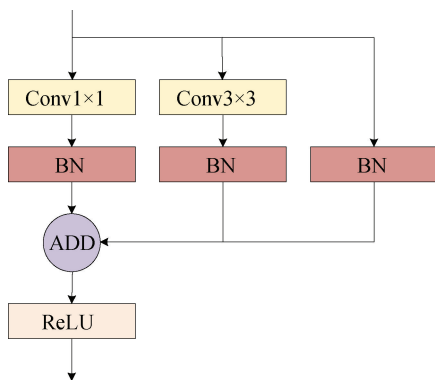


图 6 Repliblock 模块结构图
Fig. 6 Repliblock module structural diagram

2.6 CSPDarknet53 模块

CSPDarknet53 模块网络结构如图 7 所示,该模块由残差结构结合跨层链接网络 CSPNet 模块构成的轻量级卷积神经网络结构,通过采用 3×3 卷积层和残差连接使网络既保持丰富的特征表达又能够快速处理输入数据。

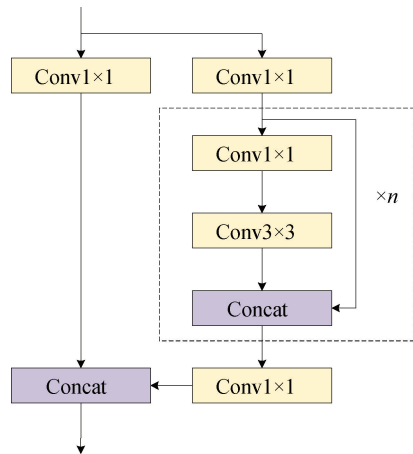


图 7 CSPDarknet53 模块结构图
Fig. 7 CSPDarknet53 module structural diagram

2.7 BepC3 模块

本文借鉴 CSPDarknet53 模块的设计思路,构建了 BepC3 模块,采用多个可重参数化卷积构成的 Repliblock 模块替换 CSPDarknet53 中原有的 Darknet 模块,使模型能够拥有较强非线性拟合能力,能够加快模型推理速度的同时,保持较好的多尺度特征融合能力。BepC3 模块结构如图 8 所示。

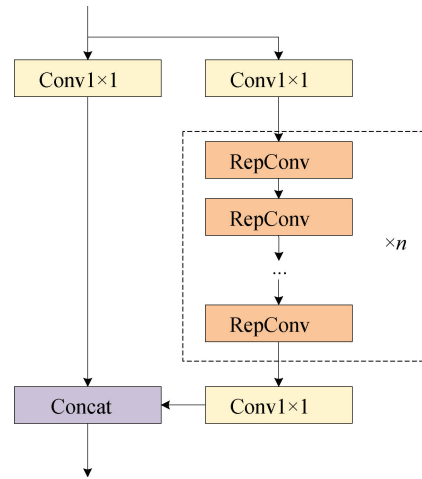


图 8 BepC3 模块结构图
Fig. 8 BepC3 module structural diagram

2.8 改进的颈部网络

原 YOLOv7 模型采用 PANet^[18]的结构,虽然有效提升模型性能,但也导致骨干网络提取的小目标语义信息难以在特征融合时被有效提取,因此本文参考 Rep-PAN 的网络结构,采用 BepC3 替换原有模型中的部分 ELAN-W 模块进行特征拟合,同时采用转置卷积进行上采样操作,有效提升网络的非线性拟合能力,进而提升颈部网络的特征融合效果。改进的颈部网络结构如图 9 所示。

改进后的 Rep-PAN 结构,也是通过上采样和下采样双向操作进行特征信息的融合,其中上采样是通过 torch 官方自带的转置卷积 (Transpose) 实现的,上采样过程中采用 BepC3 模块替换了 ELAN-W 模块,保持精度不变的同时达到更快的推理速度,下采样过程依旧采用 MP 模块,通过最大池化操作和双层卷积操作将保留特征图的主要信息并减少了后续计算的参数量,有效地防止了过拟合现象。特征图通过 Rep-PAN 结构的颈部网络,能够将深层与浅层的特征信息更好地融合,从而提高模型对不同尺度目标的检测能力。

2.9 改进的 YOLOv7 网络结构

综合前节所述,本文为提升 YOLOv7 对遮挡目标和小目标的检测性能,在原模型的基础上,采用

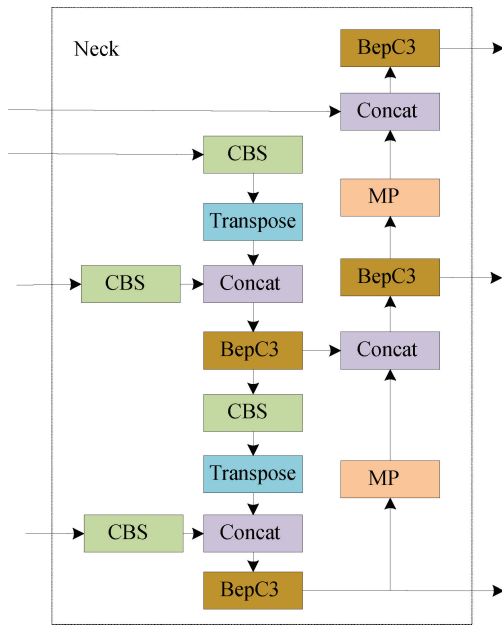


图9 改进的 Rep-PAN 颈部网络
Fig. 9 Improved Rep-PAN neck network

MobileNet 注意力模块替换原 ELAN 模块,不仅可以减少计算量,加快模型推理速度,还增强对小目标行人的特征提取能力;采用 BepC3 模块替换原 ELAN-W 模块,能够使颈部网络融合更多不同尺度的目标信息,提高网络模型泛化能力。完成以上改进后的 YOLOv7 网络结构如图 10 所示。

2.10 损失函数改进

原 YOLOv7 定位损失函数采用 CIoU Loss 来计算,CIoU 虽然考虑了目标与锚框之间的距离、重叠面积、尺度以及惩罚项,但仍是基于 IoU 的度量指标,IoU 对微小行人目标以及遮挡的位置偏差非常敏感。图 11 表示 IoU 对小目标行人和正常目标行人的灵敏度分析,其中每个表格表示 1 个像素,框 A 表示目标实际框,框 B 和框 C 表示预测框,分别具有 1 个像素和 4 个像素对角线偏差。

从图 11 中可以观察到 IoU 对不同尺度物体的灵敏度变化很大。图 11(a)中,对于具有 6×6 像

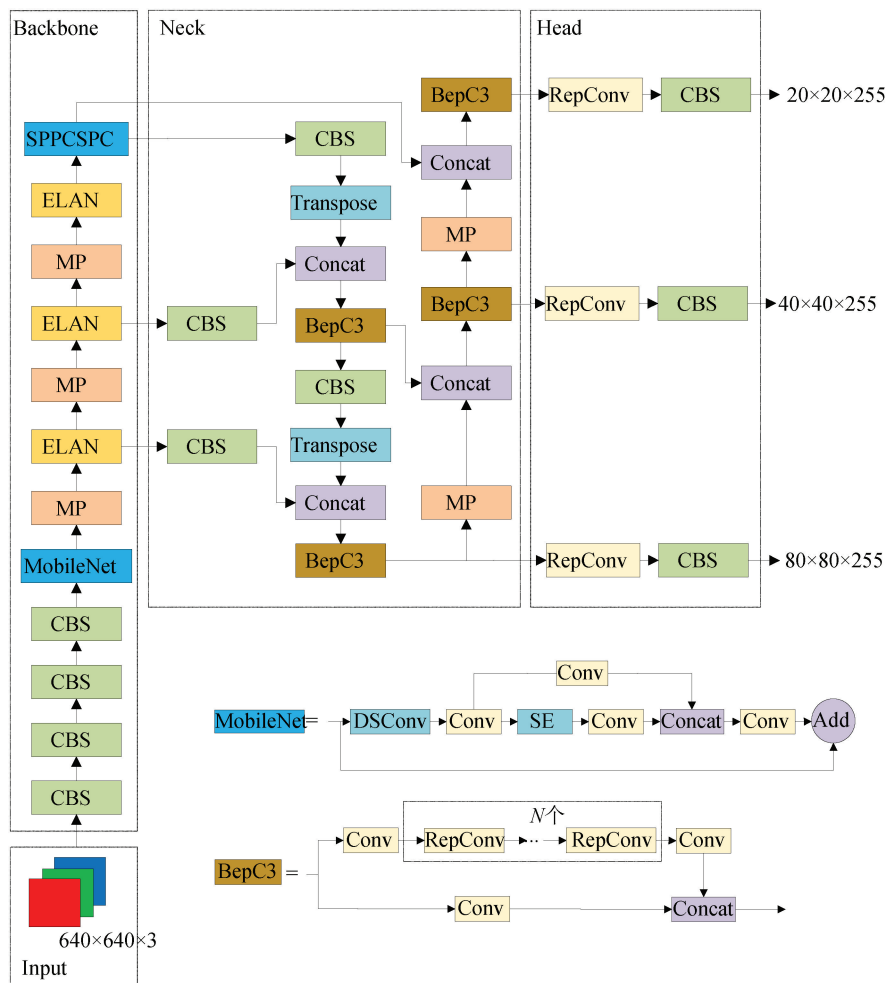


图 10 改进的 YOLOv7 网络架构图
Fig. 10 Improved YOLOv7 network architecture diagram

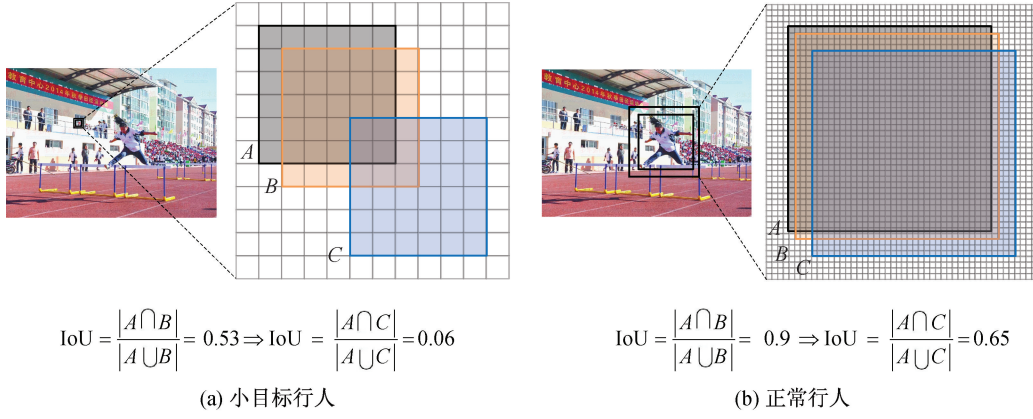


图 11 不同目标 IoU 示意图

Fig. 11 Schematic diagram of IoUs with different targets

素的微小目标行人,微小的位置偏差将导致显著的 IoU 下降(从 0.53 下降到 0.06),导致标签分配不准确。然而在图 11(b)中,对于具有 36×36 像素的正常对象, IoU 在相同的位置偏差下略有变化(从 0.90 下降到 0.65)。为了减小对这种微小目标的位置偏差的敏感度,本文借鉴 NWD 度量方式,提出 WD-Loss 损失函数,替代原 YOLOv7 的 CIoU 的损失函数。

对于微小的物体,它们的边界框中往往有一些背景像素,因为大多数真实物体都不是严格的矩形。在这些边界框中,前景像素和背景像素分别集中在边界框的中心和边界。为了更好地描述边界框中不同像素的权重,可以将边界框建模为二维高斯分布,其中边界框的中心像素具有最高的权重,并且像素的重要性从中心到边界减小。由此,两个边界框之间的相似性可以用这两个高斯分布之间的距离来表示。

对于两个二维高斯分布 $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$ 和 $\mu_2 = \mathcal{N}(m_2, \Sigma_2)$, μ_1 和 μ_2 之间的二阶 Wasserstein 距离定义为

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \left\| \Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}} \right\|_F^2 \quad (3)$$

其中, $\|\bullet\|_F$ 是范数(Frobenius)。此外,根据边界框 $A = (cx_a, cy_a, w_a, h_a)$ 和 $B = (cx_b, cy_b, w_b, h_b)$ 建模的高斯分布 \mathcal{N}_a 和 \mathcal{N}_b , 式(3)可以进一步简化为

$$W_2^2(\mathcal{N}_a, \mathcal{N}_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_2^2 \quad (4)$$

然而 $W_2^2(\mathcal{N}_a, \mathcal{N}_b)$ 是距离度量,不能做 0~1 的相似性度量,故进一步做指数形式归一化,获得 WD 度量:

$$\text{WD}(\mathcal{N}_a, \mathcal{N}_b) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a, \mathcal{N}_b)}}{C}\right) \quad (5)$$

其中 C 是与数据集相关的常数。与 IoU 相比,WD 度量在检测微小物体方面具有以下优势:(1)尺度不变性,(2)位置偏差平滑性,(3)测量不重叠或相互包容的边界框之间的相似性。

基于 WD 度量设计的回归损失如式(6)所示:

$$\mathcal{L}_{\text{WD}} = 1 - \text{WD}(\mathcal{N}_p, \mathcal{N}_g) \quad (6)$$

其中, \mathcal{N}_p 是预测框的高斯分布模型, \mathcal{N}_g 是 gt 框的高斯分布模型。即使在 $|P \cap G| = P$ or G 和 $|P \cap G| = 0$ 这两种情况下,基于 WD 的损失也可以提供梯度。

3 实验结果与分析

3.1 实验数据集

本文实验采用 Wider-Person 户外密集行人检测数据集, Wider-Person 数据集包含 13 382 张图片,共计 40 万个不同遮挡程度的人体目标,并且行人姿态、动作、穿着以及尺度大小丰富多样,户外场景较多,弥补了其他交通场景下行人数据集的不足。本文实验选取数据集中给出 Person 标签的 9 000 张行人图片,并根据 8:2 划分训练集和测试集,其中训练集有 7 200 张图片,测试集有 1 800 张照片。

3.2 实验设备与评价指标

本文使用的操作系统是 Windows10,硬件平台 CPU 采用 AMD Ryzen5-3600X 处理器,运行内存

32 GB, GPU 采用 NVIDIA Tesla P40 显卡, 显存容量 24 GB, 运行库版本为 CUDA11.6, 软件环境为 Pytorch1.12.0 与 python3.9。本文评价实验模型检测精度的指标为平均精准度 (Average Precision, AP), 其中 AP_{50} 指平均精准度 IoU 阈值取 0.5, $AP_{50,95}$ 指平均精准度 IoU 阈值取 0.5~0.95。

3.3 消融实验

本文首先设置了消融实验, 用来验证本文所提模块是否有效改进模型的性能。每组实验均以原 YOLOv7 作基本网络模型, 依次加入 MobileNet、BepC3 和 WD-Loss 等改进模块, 每组实验均采用 3.1 节中的行人数据集, 使用 3.2 节的设备与环境, 并设置训练次数 epoch = 300, 学习率 learning rate = 0.001, 最后使用 AP_{50} 与 $AP_{50,95}$ 表示模型的性能评估。消融实验结果如表 1 所示。

表 1 消融实验结果

Tab. 1 Results of ablation experiments

实验	模型	AP_{50}	$AP_{50,95}$
(1)	YOLOv7	0.753	0.441
(2)	YOLOv7+MobileNet	0.773	0.456
(3)	YOLOv7+BepC3	0.776	0.462
(4)	YOLOv7+MobileNet+ BepC3	0.781	0.466
(5)	YOLOv7+MobileNet+BepC3+WD-Loss	0.784	0.471

由表 1 可知, 实验(2)在原模型中添加了 MobileNet 模块, 模型准确率 AP_{50} 提升了 0.20, 验证了本文 MobileNet 模块为骨干网络引入注意力机制, 使网络模型在特征提取时可以聚焦于更多的小目标特征与遮挡目标特征, 因此提升了重要特征的提取能力; 实验(3)添加了 BepC3 模块, 准确率提升了 0.23, 验证了本文 BepC3 模块为颈部网络增强非线性拟合能力, 提升模型对不同尺度特征的融合效果; 实验(4)对比实验(2)和(3), 验证了同时添加 MobileNet 模块和 BepC3 模块, 使得模型检测精度提升更高; 实验(5)和对比实验(4), 模型准确率提升 0.03, 验证了本文 WD-Loss 损失函数可以降低模型对小目标定位偏差的敏感性, 提高了目标检测的精确度。该消融实验结果表明, 本文所改进模块均可提升 YOLOv7 的检测精度。

3.4 横向对比实验

为了进一步验证本文改进 YOLOv7 模型的性能, 又设置横向对比试验^[19], 将本文改进算法模型与 SSD、RetinaNet、Faster-RCNN、YOLOv3、YOLO-X^[20] 和 YOLOv5 等常见目标检测算法进行对比。

所有实验均采用 3.1 节的数据集与 3.2 节的设备环境, 横向对比实验结果如表 2 所示。

表 2 横向对比实验结果

Tab. 2 Results of cross-comparison experiments

实验	模型	AP_{50}	$AP_{50,95}$
(1)	SSD	0.502	0.210
(2)	RetinaNet	0.520	0.241
(3)	Faster-RCNN	0.545	0.268
(4)	YOLOv3	0.587	0.287
(5)	YOLO-X	0.655	0.362
(6)	YOLOv5	0.679	0.398
(7)	YOLOv7	0.753	0.441
(8)	Ours	0.784	0.471

根据表 2 结果可知, 在密集场景行人检测任务中, 本文改进 YOLOv7 算法的精度均高于以上常见目标检测算法, 其中 AP_{50} 达到 0.784, 较原 YOLOv7 模型提升了 0.031, 较 YOLOv5 模型提升了 0.105; $AP_{50,95}$ 达到 0.471, 较原 YOLOv7 模型提升了 0.030, 较 YOLOv5 模型提升了 0.073。针对密集行人场景, 本文改进模型能体现出更好的检测性能。

3.5 实验结果可视化

为进一步直观体现本文改进算法在密集行人场景的检测效果, 本文将原模型与改进模型的实验结果进行可视化展示, 两者检测效果对比如图 12 所示。

图 12(a)(d) 为 Wider-Person 部分原图, 图 12(b)(e) 为原 YOLOv7 检测效果图, 图 12(c)(f) 为改进 YOLOv7 检测效果图。可见, 原图中存在较多的小目标行人以及遮挡行人, 在原 YOLOv7 模型检测效果图中可以检测大多数行人目标, 但因为行人目标特征较小、身体局部被遮挡等问题, 依旧存在较多的行人漏检与误检。而本文改进 YOLOv7 模型能够检测出更多的小尺寸目标以及遮挡目标, 表现出更好的检测效果。

4 结论

本文针对密集行人场景目标检测任务, 提出了改进 YOLOv7 检测算法, 在原有的基础架构上, 引入 MobileNet 注意力模块, 提升了模型的特征提取能力; 引入 BepC3 模块, 增强了模型多尺度特征融合能力; 又引入 WD-Loss 损失函数, 增强了模型的定位准确性。最后在 Wider-Person 密集行人检测数据集下进行消融实验和横向对比实验, 结果



图12 密集行人场景检测效果对比图

Fig. 12 Comparison of detection effect in dense pedestrian scene

表明本文改进算法模型相较于其他目标检测算法有一定的优势,在密集行人场景下,表现出更好的检测效果,减少了误检和漏检等情况。

参考文献:

- [1] CAI Y, DAI L, WANG H, et al. Pedestrian motion trajectory prediction in intelligent driving from far shot first-person perspective video[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(6): 5298-5313.
- [2] LI C, XU P. Application on traffic flow prediction of machine learning in intelligent transportation [J]. *Neural Computing and Applications*, 2021, 33(2): 613-624.
- [3] 马小陆,王明明,王兵. YOLOv3 在安全帽佩戴检测中的应用研究[J]. *河北工程大学学报(自然科学版)*, 2020, 37(04): 78-86.
- [4] LI A, GAO X, QU C. Pedestrian detection based on improved YOLOv3 algorithm [C]//*Intelligent Life System Modelling, Image Processing and Analysis; 7th International Conference on Life System Modeling and Simulation, LSMS 2021 and 7th International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2021, Hangzhou, China, October 22-24, 2021, Proceedings, Part I 7*. Springer Singapore, 2021: 221-231.
- [5] LI K, ZHUANG Y, LAI J, et al. PFYOLOv4: An improved small object pedestrian detection algorithm [J]. *IEEE Access*, 2023, 11: 17197-17206.
- [6] 李翔,何森,罗海波. 一种面向遮挡行人检测的改进 YOLOv3 算法[J]. *光学学报*, 2022, 42(14): 160-169.
- [7] LI C, WANG Y, LIU X. An improved YOLOv7 lightweight detection algorithm for obscured pedestrians [J]. *Sensors*, 2023, 23(13): 5912.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 7464-7475.
- [9] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [J]. *arXiv preprint arXiv:2110.13389*, 2021.
- [10] ZHANG S, XIE Y, WAN J, et al. Widerperson: A diverse dataset for dense pedestrian detection in the wild [J]. *IEEE Transactions on Multimedia*, 2019, 22(2): 380-393.
- [11] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]//*Proceedings of the IEEE/CVF international conference on computer vision*. 2019: 1314-1324.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 7132-7141.
- [13] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. *arXiv preprint arXiv:2209.02976*, 2022.
- [14] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020: 390-391.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. *arXiv preprint arXiv:1704.04861*, 2017.

- [16] SANDLER M, HOWARD A, ZHU M, et al. Mobile-netv2: Inverted residuals and linear bottlenecks [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.
- [17] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.
- [18] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
- [19] 樊嵘, 马小陆. 面向拥挤行人检测的改进 DETR 算法 [J]. 计算机工程与应用, 2023, 59(19): 159-165.
- [20] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. arXiv preprint arXiv:2107.08430, 2021.

(责任编辑 王利君)

(上接第 102 页)

- [12] 邓昌棉, 张勇. 基于数据驱动的电动出租车充电站规划方法研究 [J]. 森林工程, 2020, 36(03): 77-85.
- [13] KEAWTHONG P, MUANGSIN V, GOWANIT C. Location selection of charging stations for electric taxis: A bangkok case [J]. Sustainability, 2022, 14(17): 11033.
- [14] 任峰, 向月, 雷小林, 等. 基于 GPS 数据的电动出租车充电桩选址定容 [J]. 电力自动化设备, 2022, 42(10): 273-279.
- [15] 罗思杰, 邹复民, 郭峰, 等. 基于轨迹数据的出租车充电站选址方法 [J]. 计算机工程与应用, 2022, 58(08): 273-282.
- [16] 张龙深, 葛星, 刘迪. 基于互联网开放数据的城市实时交通圈提取及分析 [J]. 河北工程大学学报(自然科学版), 2017, 34(04): 94-98.
- [17] 任丹萍, 刘琳, 陈湘国. 基于密度分区的出租车载客热点区域聚类分析 [J]. 计算机应用与软件, 2023, 40(10): 83-89.
- [18] 李文杰, 闫世强, 蒋莹, 等. 自适应确定 DBSCAN 算法参数的算法研究 [J]. 计算机工程与应用, 2019, 55(05): 1-7+148.

(责任编辑 王利君)

《河北工程大学学报(自然科学版)》2024年起由季刊变更为双月刊

为进一步缩短论文出版周期,更快推动科研成果的传播,满足广大专家学者学术交流需要,经主管主办单位同意,申请期刊刊期变更,获得河北省新闻出版局批复(冀新闻出版[2024]-060000001),《河北工程大学学报(自然科学版)》于2024年第1期起由季刊变更为双月刊,每逢双月的25号出版,特此公告。