

改进 RT-DETR 的密集行人检测算法^{*}

李青云 魏佳

(陕西理工大学数学与计算机科学学院 汉中 723001)

摘要: 针对密集行人检测中因高遮挡和尺度变化较大而导致的易漏检和精度低问题,本文提出了一种高效的面向复杂场景密集行人检测的 RT-DETR 改进算法 RSH-RTDETR。首先提出 Regocn 模块改进 backbone, 使用有限的普通卷积进行特征提取, 随后进行线性变换操作, 同时在梯度流通分支上使用 RepConv 弥补舍弃残差块所带来的性能损失并增强特征提取和梯度流通的能力, 在降低计算量和参数量的同时更好的对不同尺度的目标进行检测; 其次在颈部引入 160×160 的 S2 检测层, 通过优化特征融合过程, 增强对小尺度行人目标的检测性能; 最后采用 Haar 小波下采样模块 (HWD), 以扩大感受野、降低模型复杂度、提升对遮挡行人目标的检测精度。在 CrowdHuman 数据集做消融和对比实验, mAP50 达到了 86.6%, mAP50-95 达到了 57.8%, 相比于原算法 mAP50 提高了 1.2%, mAP50-95 提高了 1.9%, 参数量下降 40%。在 Widerperson 数据集上也优于 RT-DETR 算法。实验结果显示, RSH-RTDETR 相比 RTDETR-R18 模型在降低参数量的同时提高了密集行人检测的准确率, 也优于其他算法。本文改进算法在保证高精度的同时也实现了轻量化, 在复杂场景下的密集行人检测任务中具有优异的性能。

关键词: 密集行人检测; RT-DETR; 深度学习; 多尺度; 轻量化

中图分类号: TP391.41; TN98 文献标识码: A 国家标准学科分类代码: 520

Improvement of the dense pedestrian detection algorithm of RT-DETR

Li Qingyun Wei Jia

(School of Mathematics and Computer Science, Shaanxi University of Technology, Hanzhong 723001, China)

Abstract: To address the issues of missed detections and low accuracy caused by high occlusion and large scale variations in dense pedestrian detection, this paper proposes an efficient improved RT-DETR algorithm, RSH-RTDETR, for complex scene dense pedestrian detection. Firstly, the Regocn module is proposed to improve the backbone, using limited ordinary convolutions for feature extraction, followed by a linear transformation operation. Meanwhile, RepConv is used on the gradient flow branch to compensate for the performance loss caused by discarding residual blocks and enhance the feature extraction and gradient flow capabilities, achieving better detection of targets of different scales while reducing the computational load and parameter count. Secondly, a 160×160 S2 detection layer is introduced in the neck to enhance the detection ability of small-scale pedestrian targets during the feature fusion stage. Finally, the Haar wavelet downsampling module (HWD) is adopted to expand the receptive field, reduce model complexity, and improve the detection accuracy of occluded pedestrian targets. Ablation and comparison experiments were conducted on the CrowdHuman dataset, achieving an mAP50 of 86.6% and an mAP50-95 of 57.8%. Compared with the original algorithm, mAP50 was improved by 1.2% and mAP50-95 by 1.9%, with a 40% reduction in parameters. It also outperformed the RT-DETR algorithm on the Wider person dataset. Experimental results show that RSH-RTDETR improves the accuracy of dense pedestrian detection while reducing the parameter count compared to the RTDETR-R18 model, and outperforms other algorithms. The improved algorithm in this paper achieves lightweight while maintaining high accuracy, demonstrating excellent performance in dense pedestrian detection tasks in complex scenes.

Keywords: pedestrian detection in crowded areas; RT-DETR; deep learning; multi-scale; lightweight

0 引言

行人检测是计算机视觉中的一个关键任务, 其目标是

从图像或视频中快速识别出行人并精准定位其位置。行人检测是行人跟踪、识别、搜索和人体姿态估计等任务的重要前置步骤, 其算法性能直接关系到后续任务的表现。因此,

行人检测算法的研究具有极其重要的价值。

随着人工智能领域的快速发展和计算机硬件性能的不断提升,基于深度学习的目标检测方案已经取得显著进展,代表性的有两阶段检测算法:R-CNN, Faster RNN, Mask R-CNN;单阶段检测算法:单激发多框探测器(single shot multibox detector, SSD), YOLO(you only look once)^[1-5], DETR(detection transformer)^[6] 和 RT-DETR(real-time detection transformer)^[7] 算法等。这些研究工作不仅推动了行人检测技术的进步,也展示了深度学习在实际应用中的巨大潜力。然而,在密集行人检测中面临着背景环境复杂、因聚集而导致的遮挡等问题,使得模型需要通过极少的特征来获取整个图像的信息,极大增加了检测难度。

针对上述挑战,不少研究者提出多种优化算法模型。为了应对行人检测任务中的遮挡目标的检测难题,Zhang 等^[8]设计了一种基于变分贝叶斯原理的自动编码算法用以提升检测器的性能。Ge 等^[9]借助损失感知机制以及标签分配方法,提升了检测器对行人目标的特征表征能力。但是双阶段算法虽然有较高的检测精度,检测速度较慢,并不适用于有较高实时性要求的场景。目前 YOLO 系列由于已成为实时目标检测中应用较为广泛的主流框架。魏润辰等^[10]针对 YOLO 模型进行了改进,通过在主干网络中融合浅层和深层特征,并引入多尺度融合策略,从而降低对小目标和遮挡目标的误检率,同时提高检测速度。王泽宇等^[11]在 YOLOv8 基础上通过融合 MobileViT 骨干网络、排斥损失函数和高效多尺度注意力机制(exponential moving average attention, EMA),提高了密集场景下行人检测的精确度。陈高宇等^[12]在 YOLOv5 基础上,通过一种像素差异度注意力机制,引入新的轻量化特征提取模块 AC3 等方式去改善遮挡、形态姿势不同的行人目标的漏检问题。李宝兵等^[13]在主干网络添加 GCNet 模块、融合 SPDConv 和 Conv,在颈部网络添加 GAM 注意力机制等方式来改善行人与背景对比度低,行人目标尺度小检测精度低的问题。然而,YOLO 模型的速度和准确性受到非极大值抑制(non-maximum suppression, NMS)的限制。

近年来,Transformer 模块在多个领域取得了显著的进展。其中,DETR 方法将检测任务与 Transformer 模块相结合,避免了传统方法如 R-CNN 和 YOLO 所需的复杂后处理和参数调整。DETR 采用查询向量作为软锚框来定位目标,而非预设的锚框。然而,这种方法使得软锚框收敛较慢,导致训练时间较长。为了解决这一问题,研究者提出了诸如 Deformable DETR 等改进方法。RT-DETR 采用高效的混合编码器以提高速度,同时保持准确性,并引入了不确定最小化的查询选择方法,通过减少不确定性来提高准确性。该模型还支持灵活的速度调节,以适应不同的场景。随着问题的不断解决,RT-DETR 等算法的出现标志着 DETR 类算法已经成为一种新颖且成熟的检测方法。尽管如此,目前很少有研究者将其应用于密集行人检测任

务中。

本文针对行人数据集中多尺度、背景环境复杂、目标小和被遮挡等问题,在 RT-DETR 基础上,设计了一种针对密集行人检测的多尺度轻量化算法 RSH-RTDETR(Regocn-S2-HWD-RTDETR),保证高检测准确性和速度的同时降低参数量,为实际场景中的密集行人检测提供了有利的支持。主要研究工作如下:

1) 提出全新的 Regocn 模块改进 backbone,首先采用 GhostNet^[14] 分支思想,使用限的普通卷积进行特征提取,随后进行线性变换操作,同时在梯度流通分支上使用重参数化卷积(re-parameterizable convolution, RepConv)^[15] 弥补舍弃残差块所带来的性能损失,并增强特征提取和梯度流通的能力,可以在显著降低计算量和参数量同时更高效的处理跨尺度信息。

2) 在颈部引入 160×160 检测层,将 S2 检测层融入检测体系,在特征融合阶段增强对小尺度行人目标的检测能力,从而提高对检测难度较大的小尺度行人目标的检测精度与定位准确性。

3) 采用基于 Haar 小波下采样模型^[16] (haar wavelet downsampling, HWD) 来优化特征提取过程,并作为整体网络的下采样方法,降低模型复杂度,实现感受野的扩展,改善了对遮挡行人目标的检测精度。

1 RT-DETR 模型

RT-DETR 作为一款基于 DETR 的高实用性实时检测算法,RT-DETR 克服了 NMS 对实时检测推理速度的延迟以及对精度的影响。如图 1 所示,RT-DETR 整体结构由三大部分组成:主干网络、颈部网络和解码器。

在本研究中,ResNet18 被选为基准模型。利用其主干网络提取图像的特征信息,并将最后 3 个阶段的输出特征(分别标记为 S3、S4 和 S5)输入到后续处理流程中。这种设计不仅能够维持较高的准确率,还能有效提升处理速度并降低计算成本。

在颈部网络设计中,RT-DETR 引入了一种高效混合编码器(efficient hybrid encoder)对主干网络输出的 S5 特征,采用了 AIFI (attention-based intra-scale feature interaction)解耦处理,实现了尺度内交互的优化,并将其生成的特征表示为 F5。随后,利用跨尺度特征融合模块(cross-scale feature fusion module, CCFM),将 F5 特征与 S3 和 S4 特征进行融合处理,最终生成图像特征序列。这种设计在大幅减少计算量的同时,显著提升了处理速度,并且有效保障了模型性能的稳定性。

头部网络采用了一种集成辅助预测头的 Transformer 解码器架构。借助 IoU 感知查询选择(iou-aware query selection)技术,从编码器生成的特征序列中筛选出固定数量的图像特征,这些特征随后被用作解码器的初始对象查询。

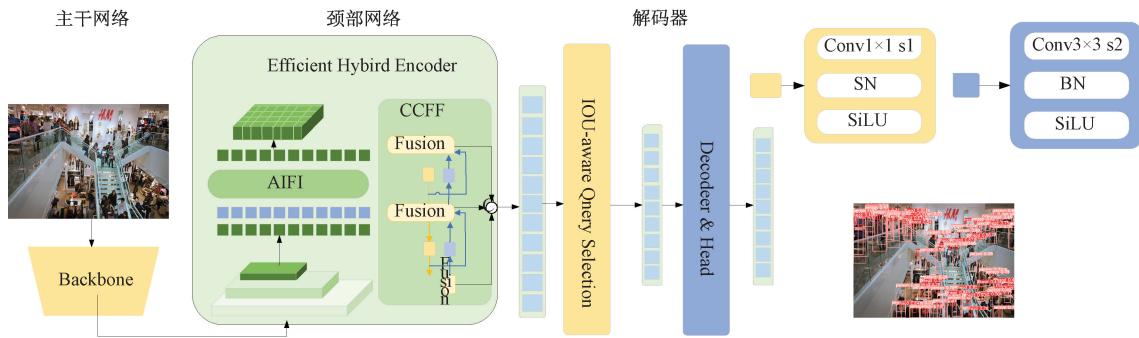


图 1 RT-DETR 网络结构

Fig. 1 Network structure of RT-DETR

此外,通过引入了 DINO (detr with improved denoising anchor boxes for end-to-end object detection) 的“去噪策略”,不仅有效提升了样本匹配的精度,还显著加快了模型的收敛速度。在解码器中,这些对象查询通过迭代优化,逐步生成边界框及其对应的得分,从而提高目标检测的准确性,并加速网络的收敛进程。

2 改进的 RT-DETR

本文从以下 3 个方面对 RT-DETR 模型进行了改进:对于主干网络,提出轻量级 Regocn 模块来改进改进,降低计算量和参数量并且提高多尺度背景下行人的检测能力;在颈部引入 160×160 检测层,在特征融合阶段增强对小尺度行人目标的检测能力;为了解决局部区域内重要空间信息的丢失问题,引入了 HWD 模块,对于下采样部分进行改进,扩大模型感受野,增强对遮挡行人目标的检测能力。

2.1 主干网络的改进

GhostNet 是一种轻量级网络,旨在压缩网络结构并简化模型设计,同时保持一定的准确性。GhostNet 在轻量化网络设计领域得到了广泛应用。如图 2 所示,GhostNet 模型在降低计算量和参数量的同时,保持了通道尺寸。它最初采用有限的普通卷积进行特征提取,随后进行线性变换操作,这一过程比常规卷积更具计算效率。最终,特征图通过张量拼接生成。

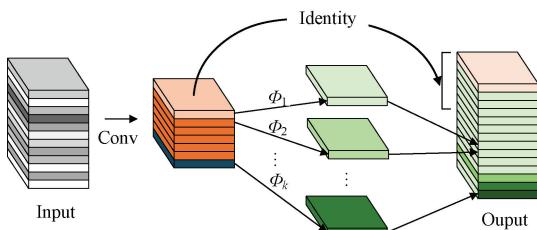


图 2 GhostNet 模块网络

Fig. 2 GhostNet module network

为了弥补舍弃残差快所带来的性能损失,本文引入 RepConv 模块。Ding 等通过结构重参数化技术,设计了一种单路极简架构的重参数化网络——RepVGG,其核心模

块为 RepConv。如图 3 所示,RepConv 由 3×3 卷积、 1×1 卷积和 BatchNorm 归一化层三部分构成。

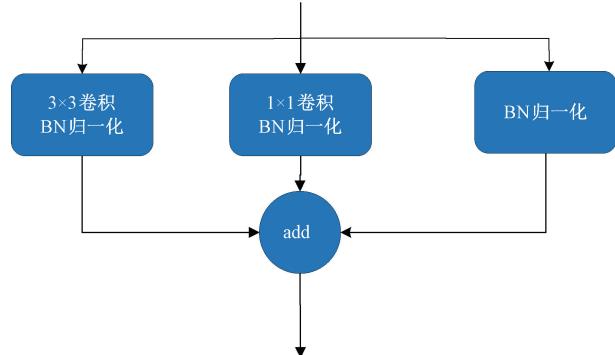


图 3 RepConv 模块网络

Fig. 3 RepConv module network

本文提出全新的 Regocn 模块改进 backbone,首先采用 GhostNet 分支思想,使用限的普通卷积进行特征提取,随后进行线性变换操作,同时在在梯度流通分支上使用 RepConv 弥补舍弃残差块所带来的性能损失,并增强特征提取和梯度流通的能力,在显著降低计算量和参数量同时更高效的处理多尺度的行人目标信息。如图 4 所示,Regocn 使用一个 1×1 的卷积,利用廉价的操作生成一部分冗余特征图来降低计算量和参数量,同时在在梯度流通分支上使用 RepConv 弥补舍弃残差块所带来的性能损失,并增强特征提取和梯度流通的能力,而且 RepConv 可以在推理的时候进行融合,同时可以通过缩放因子控制 Regocn 的大小,使 Regocn 可以兼顾小模型和大模型。

2.2 改进颈部网络

RT-DETR 的原始结构包括 3 个探测层 (S3、S4、S5),输入尺寸为 640×640 的密集行人图像经下采样后,可得到 80×80 、 40×40 以及 20×20 三个有效特征图^[17],以用于对不同尺度目标的检测。不过,在密集行人检测任务当中存在众多小尺度目标,导致模型在较深的特征图中很难精准定位和识别这些目标信息。通过选取恰当的感受野,抑制背景噪声等无关信息的干扰,助力网络掌握小尺度目标的特征,从而减轻网络过拟合状况,增强网络泛化性能,最

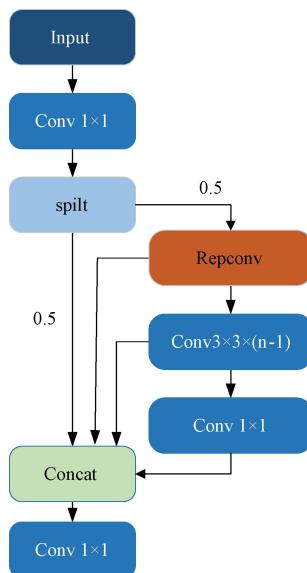


图 4 Regocn 模块网络

Fig. 4 Regocn module network

终提升模型对于小尺度行人目标的检测精确性。如图 5 所示,增设 160×160 的小目标检测层 S2,通过在更早的阶段获取高分辨率的小目标特征信息,并融合不同尺寸目标的高分辨率特征图,模型能够更好地感知小目标的细节信息,从而提高网络捕捉和学习小尺度目标的能力。

2.3 下采样的改进

为应对局部区域中关键空间信息的缺失问题,本文引入了基于 Haar 小波的下采样模块 HWD。通过 Haar 小波变换技术,模型在降低特征图空间分辨率的同时,可以更有效的留存重要信息。相较于传统的下采样方法,HWD 在保留边缘特征、纹理和细节信息上表现出显著优势,能够有效地保留行人的关键信息。Haar 下采样方法能够更有效地保留图像细节,使后续层在学习表征特征时表现得更优。

如图 6 所示,HWD 模块包含两个重要组成部分,其一为无损特征编码模块,其二为特征表示学习模块。无损特征编码模块致力于实现特征的转换以及空间分辨率的降低。为达成这一目标,引入了 Haar 小波变换。该变换能

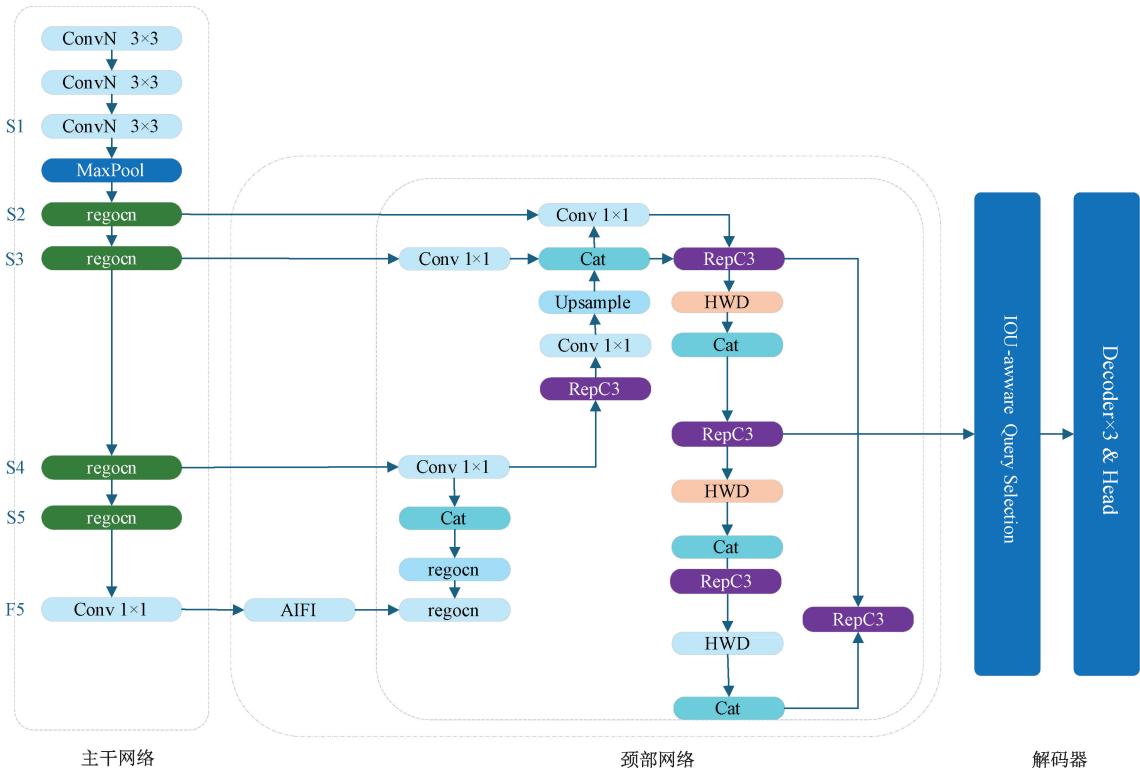


图 5 RSH-RTDETR 展开网络结构

Fig. 5 RSH-RTDETR unfolded network structure

够有力地降低特征图的分辨率,同时确保信息最大程度地保留下来。

1 阶段一维 Haar 变换的小波基函数和尺度函数可以定义如下:

$$\begin{cases} \phi 1(x) = \frac{1}{\sqrt{2}}\phi 1,0(x) - \frac{1}{\sqrt{2}}\phi 1,1(x) \\ \phi 1(x) = \frac{1}{\sqrt{2}}\phi 1,0(x) + \frac{1}{\sqrt{2}}\phi 1,1(x) \end{cases}$$

$$\phi j,k(x) = \sqrt{2^j} \phi(2^j x - k), k = 0, 1, \dots, 2^j - 1 \quad (1)$$

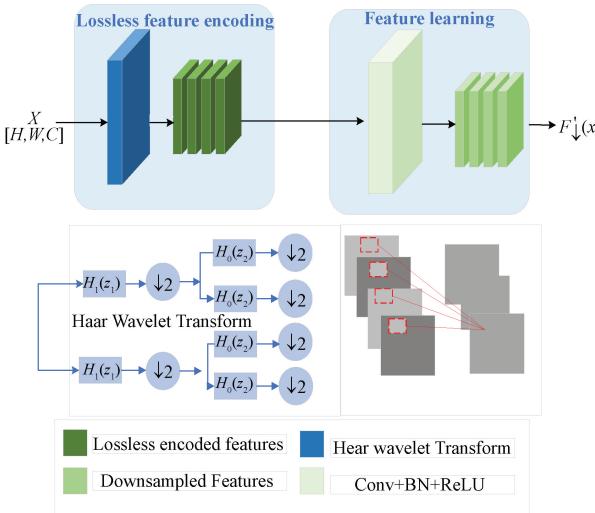


图 6 HWD 网络结构

Fig. 6 HWD network structure

其中, j 与 k 用以表明基函数的阶数。

特征表示学习模块由一个标准的 1×1 卷积层、一个批量归一化层以及 ReLU 激活函数组成。此块核心目的在于调适特征映射的通道数量,使其与后续层相匹配,并且最大程度滤除冗余信息,以便后续层能更高效地学习代表性特征。

3 实验设计与结果分析

3.1 实验环境设置

本文实验所使用的操作系统为 Linux, GPU 型号为 NVIDIA A800, 显存大小为 80 GB, 实验中的模型是使用 Python 编程语言 PyTorch 深度学习框架实现的。CUDA 版本 12.1, OpenCV 版本 4.10.0.84, Python 版本 3.8.16, PyTorch2.4.1。

本文使用的数据集为 CrowdHuman、Wider person 数据集, 输入图像大小为 640×640 , 训练次数为 300 轮, 初始学习率设为 0.0001。

3.2 数据集

为了证明本文提出的方法的有效性, 本文在公开的数据集 CrowdHuman^[18] 上进行对比和消融实验, CrowdHuman 数据集是旷世发布的用于行人检测的数据集。数据集规模庞大, 包含 15 000 张训练图像、4 370 张用于验证的图像和 5 000 张用于测试的图像。数据集涵盖了各种人群密集、背景复杂、人体遮挡和不同亮度环境下的图片, 其中包含超过 470 000 个人体标注, 约每张图片包含 23 个人, 数量较多, 行人密集同时存在各种各样的遮挡, 对行人的检测有极大的挑战性。

为了深入评估模型的泛化能力, 本文选择 WiderPerson^[19] 这一户外密集行人数据集开展对比实验。数据集 WiderPerson 共包含 13 382 张图像, 有标签的训练集图片 8 000 张, 验证集图片 1 000 张, 涵盖约 40 万个遮挡

标注实例。本文选取数据集中给出标签的 9 000 张图片按 7:2:1 的比例划分出训练集、验证集、测试集进行实验。

3.3 评价指标

本文采用平均精度均值 (mean average precision, mAP)、模型参数量 (Parameters) 以及每秒处理图像帧数 (frame per second, FPS) 等标准来评估模型的性能表现, 采用两个不同的交并比 (iou) 阈值: 其中一个是固定的 0.5, 命名为 AP50; 另一个交并比 (iou) 阈值在 0.5~0.95 之间变化, 称为 AP50:95。这两个指标均用于评估模型性能。准确率 (precision, P) 用于评估误检程度, 而召回率 (recall, R) 则评估漏检情况。准确率与召回率的定义分别如式(2)和(3)所示。

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

其中, TP 代表真正例, FN 代表假反例, FP 则为假正例。

单类标签的平均精确度 (AP) 定义如式(4)所示。

$$AP = \int_0^1 P(R) dR \quad (4)$$

mAP 表示所有标签平均精确度总和与所有类别总数 K 的商。mAP 定义如式(5)所示。

$$mAP = \frac{\sum_{k=1}^K AP_k}{K} \quad (5)$$

3.4 对比实验

为了检验 RSH-RTDETR 算法的优越性, 在相同的实验环境下进行了多次对比实验。针对 CrowdHuman 数据集, 把改进后的算法与其他多种算法进行比较, 包括 RTDETR-R34、RTDETR-R50、YOLOv5m、YOLOv5l、YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、其他先进的改进算法。同时设置了 480×480 、 640×640 、 800×800 三种不同尺寸的输入图像改进前后的对比, 进一步验证 RSH-RTDETR 算法的有效性, 表 1 展示了实验结果。

根据表 1 的实验结果, RSH-RTDETR 算法在 CrowdHuman 数据集上展现出了明显优于其他方法的性能。与其他模型算法相比, RSH-RTDETR 算法在检测精度方面更加出色, 在控制计算量和大幅降低参数量的前提下实现了高精度, 其 mAP 值分别达到 86.6% 和 57.8%。因此, RSH-RTDETR 算法在检测精度、速度和轻量化之间实现了更好的平衡, 展现出较强的通用性和实用价值。

为了评估本文提出的改进算法在不同行人目标数据集上的检测效果和泛化能力。在 Wider person 数据集上进行进一步的对比实验。如表 2 的实验结果所示, 改进后的模型比 RT-DETR-R18 在 mAP0.5 和 mAP0.5:0.9 上分别提高了 1.0%、1.2%。

3.5 消融实验

为了评估各个改进部分对整体 RT-DETR 算法性能

表 1 各模型在 CrowdHuman 数据集上检测效果

Table 1 Detection performance of each model on CrowdHuman dataset

模型	Input size	mAP50	mAP50-95	Parameters/ 10^6
RTDETR-R34	640×640	0.858	0.566	31.1
RTDETR-R50	640×640	0.865	0.577	41.7
YOLOv8n	640×640	0.787	0.491	3.0
YOLOv8s	640×640	0.834	0.544	11.1
YOLOv8m	640×640	0.847	0.565	25.8
YOLOv8l	640×640	0.865	0.586	43.6
YOLOv5m	640×640	0.821	0.523	20.9
YOLOv5l	640×640	0.832	0.541	46.1
文献[11]	640×640	0.821	0.529	3.1
文献[20]	640×640	0.832	0.51	11.4
文献[21]	640×640	0.847	0.536	3.6
RTDETR-r18(baseline)	640×640	0.854	0.559	19.9
Ours	640×640	0.866	0.578	11.8
RTDETR-r18(baseline)	480×480	0.83	0.522	19.9
Ours	480×480	0.85	0.554	11.8
RTDETR-r18(baseline)	800×800	0.868	0.558	19.9
Ours	800×800	0.874	0.594	11.8

表 2 各模型在 Wider person 数据集上检测效果

Table 2 Detection performance of each model on the Wider person dataset

模型	mAP50	mAP50-95	Parameters/ 10^6
RTDETR-r18(baseline)	0.726	0.433	19.9
Ours	0.736	0.445	11.8

的提升程度,本文在 CrowdHuman 数据集上执行了多组消融实验,结果如表 3 所示。

从表 3 的数据来看,本文提出的改进方法有效提升了 RT-DETR 算法的性能。本文提出的 RSH-RTDETR 算法与原始的 RTDETR-R18 算法相比,在各项评估指标中均表现出了显著的提升。

相比 RT-DETR,采用 Regocn 作为骨干网络的模型在 mAP50 上提升 0.3, mAP50-95 上有 0.5% 的提升,计算量

表 3 消融实验结果

Table 3 Results of ablation experiment

序号	Rtdetr-r18	Regocn	S2	HWD	P	R	mAP50	mAP50-95	Parameters/ 10^6	FPS/fps
1	√				0.866	0.781	0.854	0.559	19.9	182.9
2	√	√			0.866	0.782	0.857	0.564	13.9	190.8
3	√		√		0.868	0.787	0.86	0.573	18.6	187.8
4	√			√	0.863	0.782	0.855	0.561	19.2	172.8
5	√	√	√		0.865	0.789	0.862	0.579	12.6	192.2
6	√	√	√	√	0.872	0.789	0.866	0.578	11.8	189.4

下降 22%,参数量下降 30%。在降低模型复杂度的同时,该方法在面对目标形变时表现出更强的灵活性,能够更精确地提取重要特征,从而提高目标位置的准确度。此外, RSH-RTDETR 算法还能灵活应对不同场景和目标密集程度,显著提升主干网络的特征提取效率。颈部引入 160×160 检测层 S2 后模型的 mAP50 和 mAP50-95 分别提升了 0.6%、1.4%,验证了模型在捕捉小尺度行人目标特征信

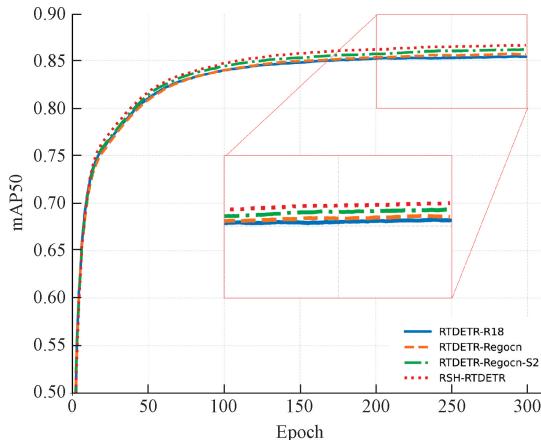
息方面的有效性。颈部下采样操作采用 HWD 模块模型计算精度、计算量、参数量都有优化,使用 Haar 小波变换保留图像细节信息,扩大特征图感受野。

最后同时添加 3 个模块,即本文所提 RSH-RTDETR 算法与原始 RTDETR-R18 相比,参数量下降 40% 对行人的 mAP50 和 mAP50-95 分别提高了 1.2%、1.9%。消融实验结果表明, RSH-RTDETR 算法能够在检测的精度、

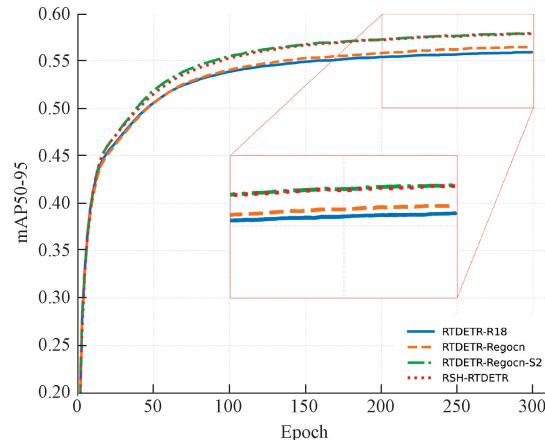
处理速度和模型轻量化之间实现良好的平衡。

图 7 展示了在训练集上本文原模型、各个改进点以及融合后模型的关键指标变化状况，横坐标为训练轮次，纵

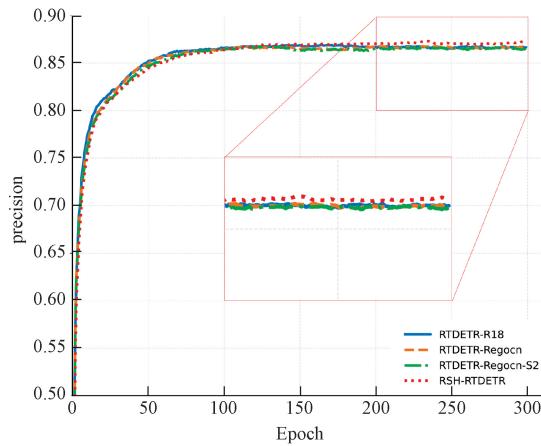
坐标为具体的指标值，可以发现各改进点对模型整体性能均有显著的增益效果，也验证了本文提出的 RSH-RTDETR 算法的有效性。



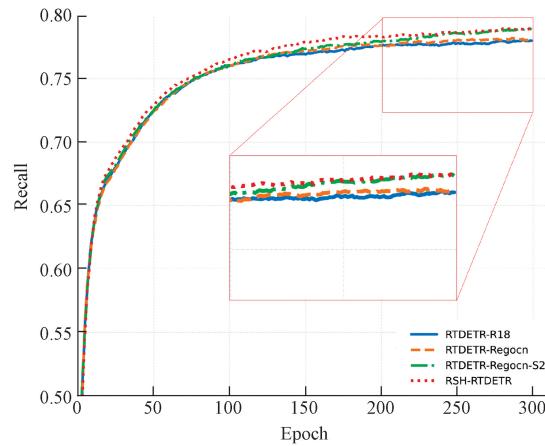
(a) mAP50 曲线图
(a) mAP50 curve graph



(b) mAP50-95 曲线图
(b) mAP50-95 curve graph



(c) 精度曲线图
(c) Precision curve graph



(d) 召回率曲线图
(d) Recall curve graph

图 7 改进效果对比图

Fig. 7 Comparison of improvement effects

3.6 可视化展示

图 8 分别展示了原始 RTDETR-R18 算法与本文设计的 RSH-RTDETR 算法在相同的实验条件下对 CrowdHuman 数据集在白天和黑夜条件下的检测效果。左侧是 RT-DETR 算法的检测效果，右侧是 RSH-RTDETR 算法的检测效果。本文所提出的改进算法在密集高遮挡情况下对优化类似图 1 中的误检、图 2 中的漏检问题见效显著。

此外为了能够更加直观的看到模型对密集行人目标在各种环境下的检测性能，本文采用了 GradCAM++ 技术进行可视化操作，展示改进模型的性能。分别对基线模型和改进模型生成相应的可视化热力图。当图像中包含多个物体时，GradCAM++ 可以为每个物体生成一个对应的热力图。这些热力图通过颜色的深浅来反映特征点的

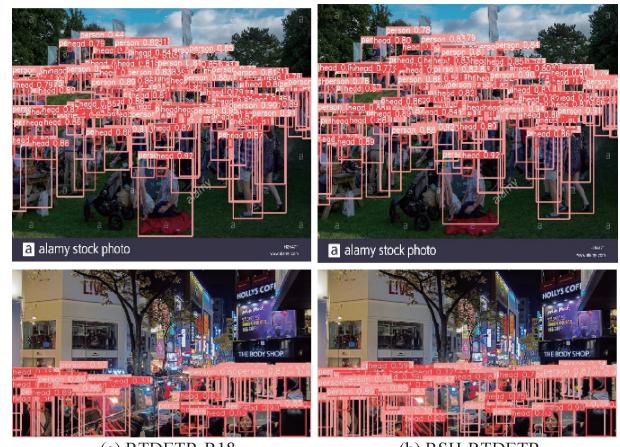


图 8 检测效果对比

Fig. 8 Comparison of detection effects

密集程度,其中红色代表特征点最为密集的区域^[22]。

由图 9 可视化结果可知右侧 RSH-RTDETR 算法对行人目标有更好的检测效果。在密集行人检测任务中,改进后的模型优势明显,可检测出原始 RT-DETR 模型遗漏与误检的密集遮挡行人目标与小尺度行人目标,有力地证实了该改进方法对提升原始模型在密集行人检测场景的有效性。

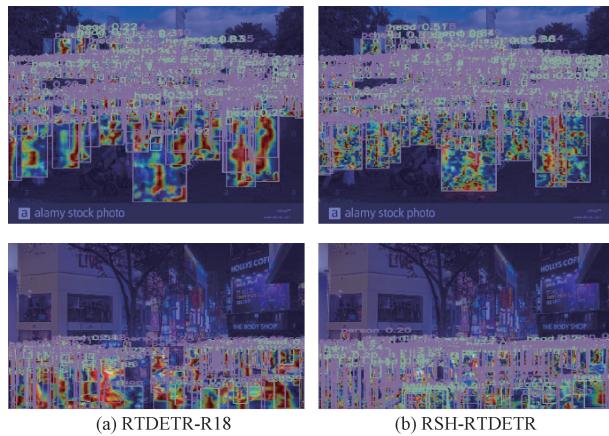


图 9 热力图检测效果对比

Fig. 9 Comparison of thermogram detection effects

4 结 论

本文提出了一种基于 RT-DETR 的密集行人检测改进算法 RSH-RTDETR。针对目前行人检测效率低、特征提取能力差、密集遮挡环境中精度不高且存在漏检误检等问题,提出 Regocn 模块对主干进行改进,在降低计算量和参数量同时增强对不同尺度信息的获取能力,其次引入 S2 小目标检测头,增强对小目标的检测能力。并且采用 Haar 小波下采样模块(HWD),进一步以扩大感受野提高检测精度,解决漏检误检问题。

RSH-RTDETR 在 CrowdHuman 数据集上最终 mAP50 达到了 86.6%,mAP50-95 达到了 57.8%,相比于原算法 mAP50 提高了 1.2%,mAP50-95 提高了 1.9%,参数量下降 40%。此外,在 Widerperson 数据集上 mAP50 提高了 1%,mAP50-95 提高了 1.2%。改进后的模型检测精度、速度、模型复杂度都有优化,综合检测效能有了显著提升。RSH-RTDETR 算法不但检测精准度较高,还可达到实时检测的要求。在众多对比模型中能取得较好的效果。但实际行人场景中的目标分布会更加复杂,接下来将会以实际密集行人场景中的研究为基础在轻量化方面继续着手研究,同时提高在密集场景小尺度行人目标的检测能力和应用能力。

参考文献

- [1] 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.
- [2] GE Z. YOLOx: Exceeding YOLO series in 2021[J]. ArXiv preprint arXiv:2107.08430, 2021.
- [3] 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]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [4] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Cham: Springer, 2025: 1-21.
- [5] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
- [6] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [7] CAI X, LAI Q, WANG Y, et al. Poly kernel inception network for remote sensing detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27706-27716.
- [8] ZHANG Y, HE H, LI J, et al. Variational pedestrian detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11622-11631.
- [9] GE Z, WANG J, HUANG X, et al. LLA: Loss-aware label assignment for dense pedestrian detection [J]. Neurocomputing, 2021, 462: 272-281.
- [10] 魏润辰,何宁,尹晓杰. YOLO-Person: 道路区域行人检测[J]. Journal of Computer Engineering & Applications, 2020, 56(19): 197-204.
- [11] WEI R CH, HE N, YIN X J. YOLO-Person: Pedestrian detection in road areas [J]. Computer Engineering and Applications, 2020, 56 (19): 197-204.
- [12] 王泽宇,徐慧英,朱信忠,等. 基于 YOLOv8 改进的密集行人检测算法: MER-YOLO[J]. 计算机工程与科学, 2024, 46(6): 1050.
- [13] WANG Z Y, XU H Y, ZHU X ZH, et al. An improved dense pedestrian detection algorithm based on YOLOv8: MER-YOLO[J]. Computer Engineering & Science, 2024, 46(6): 1050.
- [14] 陈高宇,王晓军,李晓航. 基于像素差异度注意力机制的轻量化 YOLOv5 行人检测算法[J]. 计算机工程与

- 应用,2025,61(1):291-299.
- CHEN G Y, WANG X J, LI X H. Lightweight YOLOv5 pedestrian detection algorithm based on pixel difference attention [J]. Computer Engineering and Applications, 2025,61(1):291-299.
- [13] 李宝兵,符长友. 基于改进 YOLOv8 的低光行人检测算法[J]. 西华大学学报(自然科学版),2025,44(3):29-36.
- LI B B, FU CH Y. Low light pedestrian detection algorithm based on improved YOLOv8[J]. Journal of Xihua University (Natural Science Edition), 2025, 44(3):29-36.
- [14] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [15] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733-13742.
- [16] XU G, LIAO W, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.
- [17] 冯哲,王彬,黄鹏程,等. 小天体表面着陆区岩石目标检测算法[J]. 仪器仪表学报,2024,45(4):194-205.
FENG ZH, WANG B, HUNG P CH, et al. Surface of the small landing zone rock target detection algorithm [J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 194-205.
- [18] SHAO S, ZHAO Z, LI B, et al. Crowdhuman: A benchmark for detecting human in a crowd[J]. ArXiv preprint arXiv:1805.00123, 2018.
- [19] 袁姐,王嘉丽,张晟翀. 多分支细化的拥挤行人检测算法[J]. 计算机工程与应用,2024,60(22):230-239.
- YUAN H, WANG J L, ZHANG SH CH. Multi-branch refinement of crowded pedestrian detection algorithm[J]. Computer Engineering and Applications, 2024, 60(22):230-239.
- [20] 胡倩,皮建勇,胡伟超,等. 基于改进 YOLOv5 的密集行人检测算法[J]. 计算机工程,2025,51(3):216-228.
HU Q, PI J Y, HU W CH, et al. Dense pedestrian detection algorithm based on improved YOLOv5[J]. Computer Engineering, 2025,51(3):216-228.
- [21] 胡伟超,皮建勇,胡倩,等. 面向复杂场景密集行人检测的 YOLOv8 改进模型[J]. 电子测量技术,2024,47(14):159-169.
HU W CH, PI J Y, HU Q, et al. YOLOv8 improved model for dense pedestrian detection in complex scenes[J]. Electronic Measurement Technology, 2024, 47 (14): 159-169.
- [22] 汤文权,陈灼廷,王东桥,等. 基于改进 YOLO v5 的皮蛋裂纹在线检测方法[J]. 农业机械学报,2024,55(2):384-392.
TANG W Q, CHEN ZH T, WANG D Q, et al. Online detection method for cracks in preserved eggs based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 384-392.

作者简介

李青云,硕士研究生,主要研究方向为图像处理、目标检测。

E-mail:544122320@qq.com

魏佳(通信作者),硕士,副教授,主要研究方向为数字图像处理、机器学习和计算机视觉。

E-mail:weijia@snut.edu.cn