

景区行人检测 YOLOv5-GSPE 算法模型 研究与实现

何 薇, 陈宇拓

(中南林业科技大学 计算机与信息工程学院, 湖南 长沙 410004)

摘 要:针对景区内高密度行人检测中遮挡与小目标行人漏检率高、模型复杂度高、计算量大的问题,提出一种 YOLOv5-GSPE 改进算法模型,在保证精度的同时改善检测效果,降低模型复杂度。改进算法模型通过 GhostConv 优化主干网络中常规卷积(Conv)降低模型复杂度,并使用空洞卷积改善 SPPF 模块中池化操作带来的特征信息丢失,提升模型检测时效性,增强主干网络特征提取。提出一种增强的特征金字塔网络—PrFPN,使用同层连接进一步丰富原始输入特征的融合,减少特征提取过程中的特征损失。将引入正态分布计算优化后的 EIou 损失函数作为边界框回归损失函数,提高边界框定位精度。实验结果表明,YOLOv5-GSPE 算法模型对比 YOLOv5s 模型在保证检测时效性的情况下整体复杂度降低了 12.51%,基于 Pedestrian 测试集的平均精度提升 4.05%,基于 WiderPerson 测试集的平均精度提升 3.28%,并降低了行人遮挡及小目标漏检率,改善了检测效果,该模型的可行性与有效性得到验证。

关键词:行人检测;YOLOv5-GSPE;GhostConv;PrFPN;EIou

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2023)09-0113-06

doi:10.3969/j.issn.1673-629X.2023.09.017

Research and Implementation of YOLOv5-GSPE for Pedestrian Detection in Scenic Spots

HE Wei, CHEN Yu-tuo

(School of Computer & Information Engineering, Central South University of Forestry and Technology,
Changsha 410004, China)

Abstract: Aiming at the problems of high missed detection rate of occlusion and small target pedestrians, high model complexity and large amount of calculation in high-density pedestrian detection in scenic spots, an improved YOLOv5-GSPE is proposed, which can ensure the accuracy while improving the detection effect. The improved algorithm model reduces the complexity of the model by optimizing the conventional convolution (Conv) in the backbone network through GhostConv, and uses atrous convolution to improve the loss of feature information caused by the pooling operation in the SPPF module, improving the timeliness of model detection and enhancing the feature extraction of the backbone network. An enhanced feature pyramid network, PrFPN, is proposed to further enrich the fusion of original input features using same-layer connections. The optimized EIou loss function is used as the bounding box regression loss function to improve the positioning accuracy of the bounding box, and the proposed method reduces feature loss during feature extraction. The experimental results show that compared with the YOLOv5s model, the complexity of the YOLOv5-GSPE is reduced by 12.51%, the average accuracy based on the Pedestrian test set is increased by 4.05%, and the average accuracy based on the WiderPerson test set is increased by 3.28%, and it reduces pedestrian occlusion and small target missed detection. The feasibility and effectiveness of the model have been verified.

Key words: pedestrian detection; YOLOv5-GSPE; GhostConv; PrFPN; EIou

0 引言

景区中的行人高密度、小目标、易遮挡的特点尤为显著,行人检测作为计算机视觉领域的一个重要分支,

对景区的安全管理有着重要的现实意义。目前基于深度学习的行人检测算法在准确度和检测速度方面都有明显提升,但很多算法只是在单维度方面不断深究,引

收稿日期:2022-10-07

修回日期:2023-02-09

基金项目:国家自然科学基金(31870532)

作者简介:何 薇(1998-),女,硕士研究生,研究方向为计算机视觉、图像处理;通信作者:陈宇拓(1963-),男,博士,教授,研究方向为计算机图形图像处理、智能信息处理。

入多种复杂模块来提升检测精度,很难兼顾算法模型复杂度以及检测速度,同时在行人遮挡、行人小目标检测等方面仍存在诸多挑战。

针对行人检测的现状,目前主流的行人检测算法主要分为两类:单阶段(One-stage)检测算法^[1]和两阶段(Two-stage)检测算法^[2]。One-stage 检测算法包括 SSD^[3]、YOLO^[4]算法等。Huang 等^[5]提出一种基于改进的 YOLOv5 的微小物体检测算法,将提取特征划分为四种特征分类,以增强原始输入图像的特征。郭磊等^[6]将通道注意力和空间注意力结合的 CBAM^[7]注意力模块引入 YOLOv5 网络中,并利用图像增强算法对图片进行预处理后进行迭代训练,提高算法模型的鲁棒性。此类算法主要通过回归预测出目标边界框,能够以较高的速度和精度完成检测。

Two-stage 检测算法包括 Faster-RCNN^[8]、Mask-RCNN^[9]算法等。音松等^[10]基于 Mask R-CNN 网络的 Mack 分支中增加了 CFPN 模块和 CA^[11]注意力模块,降低了遮挡对于行人检测精度的影响。此类算法发展较早,在检测精度方面有一定优势,但是整个过程计算量大、检测速度慢,很难满足实时性需求。由于景区行人检测的应用需要兼顾检测速度和精度,因此,基于 One-stage 检测算法中最新的 YOLOv5 算法模型,该文提出了 YOLOv5-GSPE (GhostConv & SPPF & PrFPN & EIou) 综合改进算法模型,主要解决如下几个方面问题:

(1) 将 GhostConv^[12]和改进后的 SPPF 模块集成到主干网络中,降低模型复杂度,增强提取的目标特征信息,降低行人遮挡及小目标的漏检率。

(2) 针对 YOLOv5 模型特征金字塔网络 FPN^[13]结构忽略原始输入特征融合的缺点,提出一种增强的特征金字塔网络 PrFPN 结构,提升模型的检测精度。

(3) 基于 YOLOv5 模型中 GIoU 损失函数在预测框位于真实框内部情况下不准确定位的问题,采用 EIou^[14]损失函数代替 GIoU,并引入正态分布对 EIou 损失函数进行优化计算,提升边界框的回归精度,进而改善遮挡及小目标行人检测效果。

1 算法模型网络结构

Redmon J 等^[15]提出了 One-stage 检测算法 YOLO,目前 YOLO 系列算法已更新至 YOLOv5, YOLOv5 有多种版本:YOLOv5s、YOLOv5l、YOLOv5m、YOLOv5x 等。其中 YOLOv5s 是最小的算法模型,考虑到检测类别的单一性,该文基于 YOLOv5s 算法模型提出 YOLOv5-GSPE 算法模型,其网络结构如图 1 所示。

在主干网络中使用 GhostConv 取代 Conv 作为卷

积层,降低算法模型的复杂度,提高模型检测速度;将 SPPF 模块(见图 1 中标注为①的区域)中 max pooling 操作改为空洞卷积(Dilated Conv),减少行人特征信息的丢失。提出一种增强的特征金字塔网络 PrFPN 结构(见图 1 中标注为②的区域),丰富多尺度特征融合。

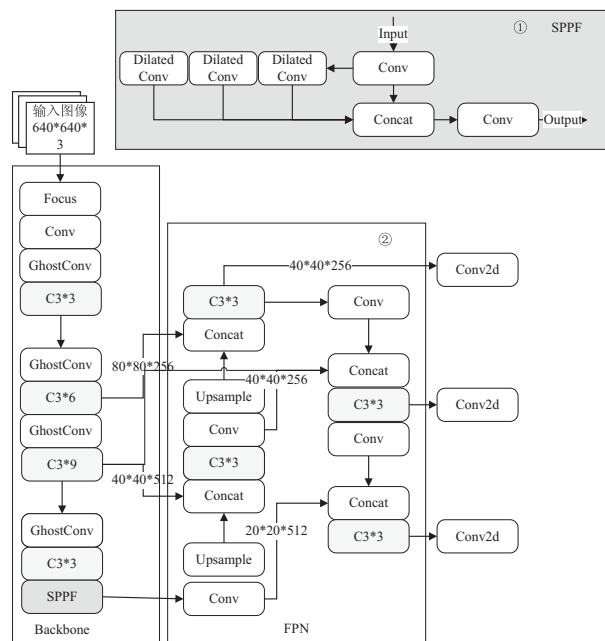


图1 YOLOv5-GSPE 网络结构

2 算法模型改进方法

YOLOv5-GSPE 算法模型主要包括对主干网络中的卷积层、SPPF 模块、特征金字塔网络以及损失函数等。

2.1 主干网络

2.1.1 卷积层优化

YOLOv5 主干网络卷积层为常规卷积 Conv,使用各个通道中的特征图与多个卷积核进行运算。假设 Input 端输入尺寸为 $w \times s$ 的图像,Output 端输出尺寸为 $w' \times s'$ 的图像; c 和 n 分别是 Input 端和 Output 端的通道数,则 Conv 的计算总量 Q_1 如公式(1)所示。

$$Q_1 = w * h * c * n * w' * h' \quad (1)$$

GhostNet^[12]提出的 GhostConv 将卷积流程主要分为常规卷积和 Ghost 卷积,如图 2 所示。

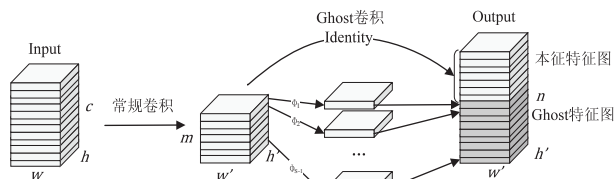


图2 GhostConv 流程示意图

常规卷积计算总量 Q_2 见公式(2)。假设 $s(s > 1)$ 为 $Y_{m \times w' \times h'}$ 通道产生的总映射数,其中包含 1 次恒等变

换 (Identity) 将 $Y_{m \times w' \times h'}$ 映射成本征特征图, 以及 $s-1$ 次单位映射运算 Φ_{ij} 操作生成 Ghost 特征图。Ghost 特征图与本征特征图连接成 n 个通道的输出特征图, 其计算总量 Q_3 见公式(2)。

$$\begin{cases} n = m * s \\ Q_2 = w * h * c * m * w' * h' \\ Q_3 = (s - 1) * m * d * d * w' * h' \end{cases} \quad (2)$$

$d * d$ 为单位映射运算 Φ_{ij} 操作的平均卷积核大小, 与 $w * h$ 具有相似大小, c 为 Input 端输入特征图通道数, 远大于 s , 即 $c \gg s$, 因此, GhostConv 的计算总量为 $Q_2 + Q_3$, 通过公式(3)能近似计算出 Conv 的计算总量 Q_1 是 GhostConv 计算总量的 s 倍, 由 $s > 1$ 可知 GhostConv 具有更小的计算量。

$$\begin{aligned} \frac{Q_1}{Q_2 + Q_3} &= \frac{w * h * c}{\frac{1}{s} * w * h * c + \frac{s-1}{s} * d * d} \approx \\ &\frac{s * c}{c + s - 1} \approx s \end{aligned} \quad (3)$$

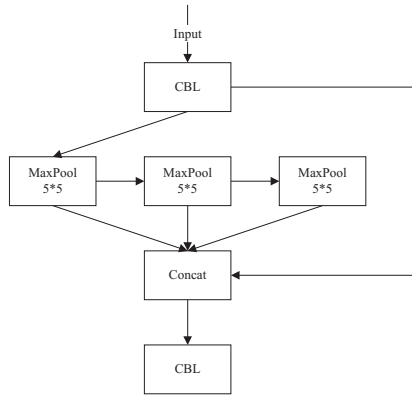
综上所述, GhostConv 相比 Conv 减少了卷积过程

的计算量, 进而表明卷积过程中参数量的减少。因此, 将 GhostConv 代替主干网络中的 Conv 能够有效降低算法模型复杂度。

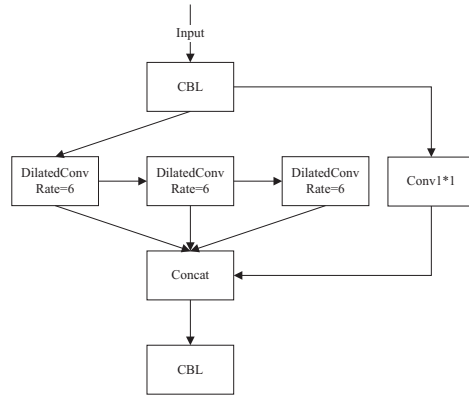
2.1.2 SPPF 模块的改进

YOLOv5 主干网络中 SPPF 模块如图 3(a), 通过串行 max pooling 操作集合不同感受野的特征进行连接, 虽然一定程度上增强了主干网络的感受野, 但 max pooling 操作不可避免地造成一定的特征信息丢失。因此, 改进的 SPPF 模块使用空洞卷积来代替 max pooling 操作, 如图 3(b)。

空洞卷积和池化操作都可以增强感受野, 但池化操作增大感受野的同时不可避免地造成特征信息丢失。空洞卷积可以尽量保留内部数据结构和避免下采样, 增大感受野的同时不改变图像输出特征图的分辨率, 减少了特征信息丢失。通过控制不同的卷积核参数来设置空洞卷积速率, 可以灵活地替换三个串行 max pooling 操作, 丰富主干网络提取的特征。改进的 SPPF 模块降低行人特征信息丢失, 进而增强主干网络的特征提取, 降低遮挡及小目标行人的漏检率。



(a) SPPF 模块



(b) 改进的 SPPF 模块

图 3 SPPF 模块改进前后模块结构对比

2.2 特征金字塔网络

YOLOv5 算法模型中特征金字塔网络为 PANet 结构, PANet 结构中同层相邻连接方式导致输出特征图忽略了原始输入特征的融合。如图 4 所示, 该文基于 BiFPN^[16] 的启发, 提出了增强的特征金字塔网络 PrFPN 结构, 在输出特征图(上采样特征图, 如 N3、N4、

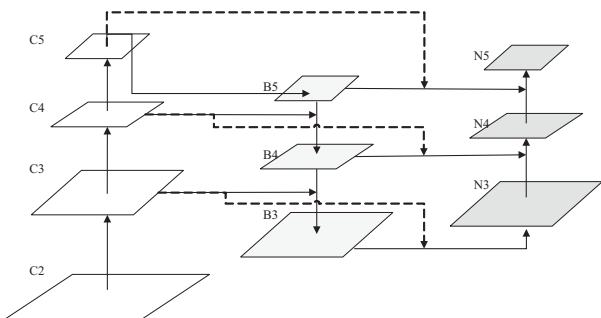


图 4 PrFPN 结构示意图

N5) 中分别利用了同层的原始输入特征图(如 C3、C4、C5)的特征融合, 如图中虚线所示。

PrFPN 结构基每层的输出特征图将上采样特征图、下采样特征图(如 B3、B4、B5)和原始输入特征图进行特征融合, 丰富了特征融合多样性, 更好地利用原始输出特征信息, 增强小目标特征敏感度, 提升算法模型检测精度。

2.3 损失函数

YOLOv5 算法模型采用 GIoU 损失函数, 主要通过预测框 A 与真实框 B 的最小闭包区来计算损失函数值, 当预测框位于真实框内部时, 即 $A \cap B = B$, 就无法精确定位效果最佳的预测框。

针对以上问题, 该文选择了更为出色的 EIoU 损失函数(L_{EIoU}), 将预测框与真实框的重叠区、长宽比以及中心点距离三个参数都融合到计算中, 如式(4)

所示。

$$\begin{cases} L_{\text{EIoU}} = 1 - \text{IoU} \\ \text{IoU} = \frac{A \cap B}{A \cup B} + \left(\frac{r^2}{c^2} - \frac{h^2}{C_h^2} - \frac{w^2}{C_w^2} \right) \end{cases} \quad (4)$$

其中, r 为 A 与 B 中心点之间的欧氏距离, A 与 B 之间的长宽之差分别为 h 和 w , A 与 B 最小闭合区为 C , 其高度为 C_h , 宽度为 C_w , 对角线距离为 c 。 L_{EIoU} 融合多参数能够更加精确地计算损失函数值, 但式 (4) 中的损失函数 L_{EIoU} 是随着定位损失 (IoU) 线性变化的, 不利于模型训练。因此, 引入正太分布函数对 L_{EIoU} 进行新的定义, 用 $L_{\text{EIoU-N}}$ 表示, 见公式 (5)。

$$L_{\text{EIoU-N}} = e^{-\frac{(\text{IoU}-\mu)^2}{2\sigma^2}} - C \quad (5)$$

其中, $\mu = -0.6$, $\sigma^2 = 0.2$, $C = 0.03$ 为调整常数。

如图 5 所示, L_{EIoU} 都随着 IoU 的不断增大线性下降, 但 $L_{\text{EIoU-N}}$ 随着 IoU 的梯度增大加速减小。IoU 在逐步增大的过程中波动对 $L_{\text{EIoU-N}}$ 影响是逐步减少的, 表明 $L_{\text{EIoU-N}}$ 比 L_{EIoU} 的稳定性更好, 更具合理性。

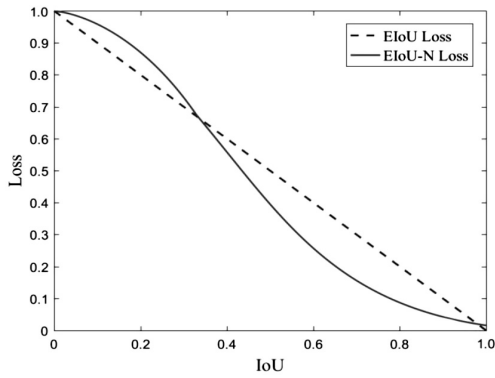
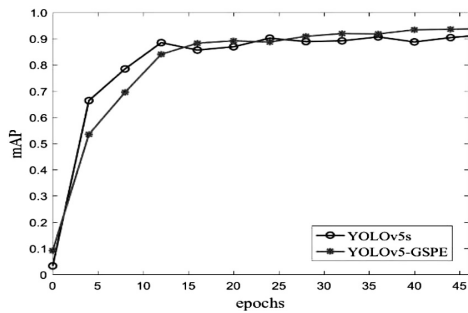
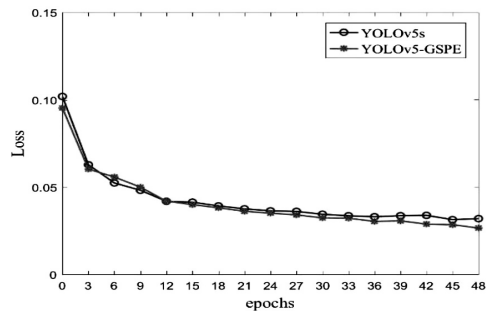


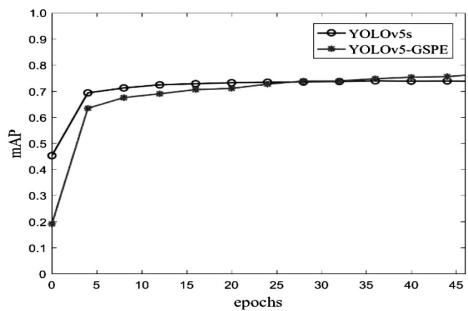
图 5 L_{EIoU} 与 $L_{\text{EIoU-N}}$ 对比示意图



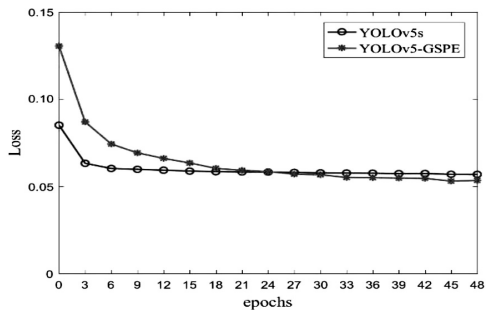
(a) Pedestrian 数据集 mAP 曲线



(b) Pedestrian 数据集 Loss 值下降曲线



(c) WiderPerson 数据集 mAP 曲线



(d) WiderPerson 数据集 Loss 值下降曲线

图 6 两种数据集实验的 mAP 及 Loss 值下降曲线

3 实验结果与分析

实验环境如下: 操作系统为 Windows 11, CPU 为 AMD Ryzen 7 4800H 2.9 GHz, 运行内存为 16 GB。算法模型基于 Pytorch 深度学习框架, 编程语言为 Python。

3.1 实验数据来源

该文使用了两种公开的数据集进行训练及测试, 一种是数据建模和分析竞赛平台 Kaggle (<https://www.kaggle.com/>) 网站公开的行人检测数据集 Pedestrian, 其中训练集 944 张、验证集 160 张和测试集 235 张。另一种是生物识别与安全技术研究中心 (<http://www.cbsr.ia.ac.cn/>) 公开的密集行人基准数据集 WiderPerson, 其中训练集 8 000 张、验证集 1 000 张和测试集 4 382 张。两种实验数据集包含了不同姿态、相互遮挡以及多种场景下的行人, 有效提升了行人检测的鲁棒性。

3.2 检测精度实验结果分析

两个算法模型对比 300 轮迭代训练中的变化明显的前 50 轮迭代的平均精度 (mAP) 及 Loss 值下降曲线, 如图 6 所示, 其中横坐标为训练的迭代轮次 (epochs), 纵坐标为每轮迭代的 mAP。基于 Pedestrian 数据集的实验中, YOLOv5-GSPE 算法模型的 Loss 值下降曲线及 mAP 上升曲线在收敛稳定性上优势明显。基于 WiderPerson 数据集的实验中, 两个算法模型的 Loss 值下降曲线及 mAP 上升曲线收敛都较为稳定。综合来看, 两种数据集在到达稳定迭代轮次后, YOLOv5-GSPE 算法模型在 mAP 上优于 YOLOv5s 算法模型。

在 300 轮迭代中, Pedestrian 数据集实验的最高 mAP 达到了 95.87%, 相比 YOLOv5s 算法模型的 91.82% 提升了 4.05 百分点。WiderPerson 数据集实验的最高 mAP 达到了 77.46%, 相比 YOLOv5s 算法模型的 74.18% 提升了 3.28 百分点。由公式(1)和(2)采用 YOLOv5 和 YOLOv5-GSPE 算法模型计算得到的总参数量分别为 7 012 822 和 6 135 420, 其复杂度降低 12.51%。

将 YOLOv5-GSPE 算法模型与目前主流标准版本的目标检测模型进行对比实验, 在召回率 (Recall)、mAP 和推理时间 (Time) 三个方面进行对比, Pedestrian 数据集实验对比结果如表 1 所示, WiderPerson 数据集实验对比结果如表 2 所示。

表 1 Pedestrian 数据集模型对比实验结果

模型	mAP/%	Recall	Time/s
Faster-RCNN	89.10	0.92	0.5
YOLOv5s	91.82	0.93	0.08
YOLOv5m	94.24	0.94	0.095
YOLOv5-GSPE	95.87	0.94	0.08

表 2 WiderPerson 数据集模型对比实验结果

模型	mAP/%	Recall	Time/s
Faster-RCNN	72.30	0.62	0.5
YOLOv5s	74.18	0.65	0.08
YOLOv5m	75.45	0.65	0.095
YOLOv5-GSPE	77.46	0.67	0.08

综上所述, 通过表 1 和表 2 的对比结果可以看出, YOLOv5-GSPE 算法模型在检测速度和精度方面, 较

目前几个主流目标检测算法模型有一定优势。

3.3 消融实验结果分析

消融实验是针对 YOLOv5-GSPE 算法模型中每个改进合理性验证实验, 主要有 GhostConv 卷积层、SPPF 模块、特征金字塔网络 PrFPN 结构和损失函数 EIou 四个改进。首先对各个改进进行独立的精度实验评估分析, 如图 7 所示。

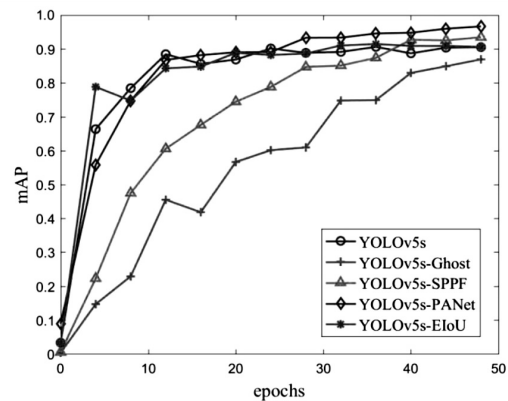


图 7 各模块精度评估对比曲线

如图 7 所示, GhostConv 代替 Conv 作为主干网络卷积层可以有效降低模型的复杂度及计算量, 但对 mAP 的收敛稳定性造成影响。改进的 SPPF 模块能够一定程度上提升模型 mAP, 降低特征信息丢失可以提高模型检测效果, 额外的开销导致 mAP 曲线收敛速度变慢。特征金字塔网络 PANet 结构改进能够大幅度提升模型 mAP, 增强 mAP 曲线稳定性。优化后的 EIou 损失函数代替 GIoU 损失函数, 能够小幅度提升 mAP。基于 Pedestrian 数据集下训练和推理, 实验结果如表 3 所示。

表 3 消融实验结果

模型	GhostConv	SPPF	PrFPN	EIoU	mAP/%	Time/s
YOLOv5s					91.82	0.08
模型 1	✓				89.26	0.075
模型 2	✓	✓			91.55	0.076
模型 3	✓	✓	✓		94.80	0.078
YOLOv5-GSPE	✓	✓	✓	✓	96.87	0.08

如表 3 所示, 模型 1 基于 YOLOv5s 模型进行改进, 检测精度有所下降。模型 2 在模型 1 基础上优化 SPPF 模块, 提升了模型精度, 少量的开销影响推理速度。模型 3 基于模型 2 提出的改进 PrFPN 结构, 可以大幅度提升模型平均精度。YOLOv5-GSPE 算法模型在模型 3 基础上引入优化后的 EIou 损失函数, 进一步提升模型检测精度。

3.4 检测实验效果分析

(1) Pedestrian 数据集的检测效果分析。

如图 8 所示, YOLOv5-GSPE 算法模型更为丰富

的特征提取带来行人目标特征符合度下降, 进而使得锚框置信度相比 YOLOv5s 算法模型稍有下降。由图 8(a) 可以看出, YOLOv5-GSPE 算法模型能够对图中远近大小成像行人信息实现准确检测, 对于遮挡目标也能保证一定的检测精度, 而 YOLOv5s 算法模型(图 8(b))对于较远成像的小目标的行人漏检率较高, 如图 8(c) 与 (d) 局部放大图对比所示。

(2) WiderPerson 数据集检测效果分析。

图 9 为对场景二中遮挡行人检测效果对比, 由图 9(a) 可以看出, YOLOv5-GSPE 算法模型对拥挤遮挡

的行人相比 YOLOv5s 算法模型的检测效果更好,但遮挡行人的特征并不完整,容易导致算法模型重复锚框,如图 9(b)所示,其漏检区域从图 9(c)和(d)局部放大图可以清晰对比。

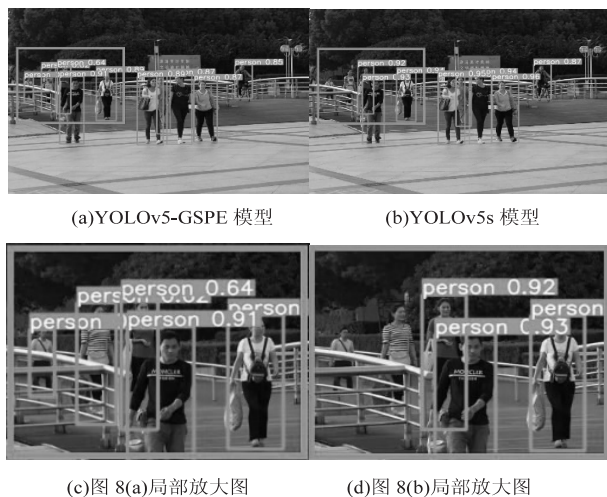


图 8 Pedestrian 数据集检测效果对比

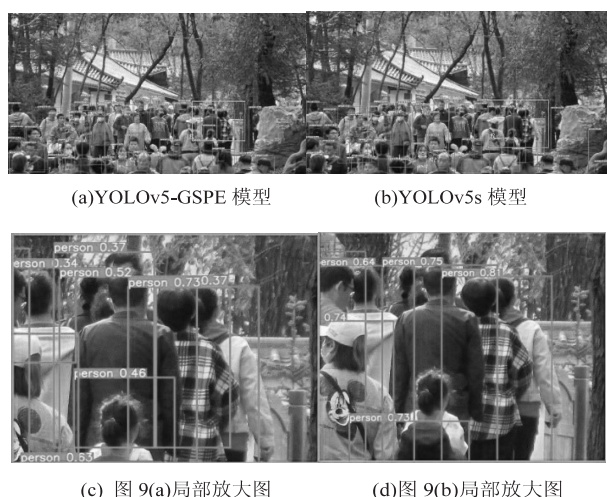


图 9 WiderPerson 数据集检测效果对比

综上所述, YOLOv5 - GSPE 算法模型相比 YOLOv5s 算法模型的检测效果更好,尤其对小目标行人以及遮挡行人的漏检率更低。

4 结束语

针对景区场景下的行人表现特征,目前的行人检测算法在景区场景下应用存在小目标及遮挡行人漏检率高、模型复杂度高、计算量大等问题。该文提出 YOLOv5-GSPE 算法模型,通过 GhostConv 和改进的 SPPF 模块优化主干网络;提出增强的特征金字塔网络 PrFPN 结构,在输出特征图中加入了原始输入特征的融合;用优化后的 EIou 损失函数代替 GIoU 损失函数,提升边界框的回归精度。

实验结果表明, YOLOv5-GSPE 算法模型在均衡检测精度和速度的情况下更具综合优势,同时大幅度

地减少了模型复杂度,对小目标行人和遮挡行人检测漏检率有明显改善,为景区的行人智能检测应用提供一种有效途径。但算法模型中特征提取明显增强,使其在非拥挤场景下的应用中对目标背景的抗干扰度下降。

参考文献:

- [1] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu: IEEE, 2017: 6517 - 6525.
- [2] ANSARI M F, LODI K A. A survey of recent trends in two-stage object detection methods[C]//Renewable power for sustainable growth: proceedings of international conference on renewal power (ICRP 2020). Singapore: Springer, 2021: 669-677.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer vision - ECCV 2016; 14th European conference. Amsterdam: Springer International Publishing, 2016: 21-37.
- [4] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[J]. arXiv: 2004. 10934, 2020.
- [5] HUANG T, CHENG M, YANG Y, et al. Tiny object detection based on YOLOv5[C]//2022 the 5th international conference on image and graphics processing (ICIGP). New York: Association for Computing Machinery, 2022: 45-50.
- [6] 郭磊, 王邱龙, 薛伟, 等. 基于注意力机制的光线昏暗条件下口罩佩戴检测[J]. 电子科技大学学报, 2022, 51(1): 123-129.
- [7] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer International Publishing, 2018: 3-19.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(2): 386-397.
- [10] 音松, 陈雪云, 贝学宇. 改进 Mask RCNN 算法及其在行人实例分割中的应用[J]. 计算机工程, 2021, 47(6): 271-276.
- [11] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Kuala Lumpur: IEEE, 2021: 13708-13717.
- [12] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features

(下转第 125 页)