

文章编号:1001-5078(2024)02-0295-07

· 图像与信号处理 ·

基于 YOLOv5s 的轻量化红外图像行人目标检测研究

胡焱¹, 赵宇航¹, 胡皓冰¹, 巩银², 孙寰宇³

(1. 中国民用航空飞行学院航空电子电气学院, 四川 广汉 618307;

2. 西藏航空有限公司, 四川 成都 610202; 3. 山东航空股份有限公司, 山东 济南 250107)

摘要: 基于红外图像的行人识别是现代安防系统的重要组成部分。在计算资源受限场景, 由于红外行人检测算法中模型尺寸的影响, 检测精度与部署难度往往难以平衡。针对此问题, 本文提出了一种基于 YOLOv5s 的轻量化目标检测算法: 首先引入 MobileNetv3 轻量化特征提取网络, 并使用深度可分离卷积减小模型尺寸, 使其更易部署至 CPU 设备; 其次, 将最近邻插值上采样方式替换为 CARAFE (Content-Aware ReAssembly of FEatures), 明显提升了图像重建效果; 最后使用 EIOU Loss 作为边界框损失函数改善模型回归性能。本文在采样后的 LLVIP 红外行人图像数据集上进行了测试: 对于红外图像下的行人目标, 本文在保持高检测精度 ($AP=95.4\%$) 的同时, 模型大小减少 80.6%, 参数量减少 82.8%; 在使用 CPU 平台进行推理时, 推理速度提升 43.3%, 且检测多尺度目标的性能有所提升。以上两方面结果验证了算法的有效性。

关键词: 红外行人检测; YOLOv5; MobileNetv3; 轻量化

中图分类号: TP391; TN219 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2024.02.020

Research on pedestrian target detection in lightweight infrared images based on YOLOv5s

HU Yan¹, ZHAO Yu-hang¹, HU Hao-bing¹, GONG Yin², SUN Huan-yu³

(1. School of Aviation Electronics and Electrical Engineering, China Civil Aviation Flight Academy, Guanghan 618307, China;

2. Tibet Airlines Co., Ltd., Chengdu 610202, China; 3. Shandong Airlines Co., Ltd., Jinan 250107, China)

Abstract: Pedestrian recognition based on infrared images is an important component of modern security systems. In scenarios with limited computing resources, it is often difficult to balance the detection accuracy and deployment difficulty due to the influence of model size in infrared pedestrian detection algorithms. In response to this issue, a lightweight object detection algorithm based on YOLOv5s is proposed in this paper. Firstly, the MobileNetv3 lightweight feature extraction network is introduced and deep separable convolution is used to reduce the model size, making it easier to deploy to CPU devices. Secondly, the nearest neighbor interpolation upsampling method is replaced with CARAFE (Content-Aware ReAssembly of FEatures) which significantly improves the image reconstruction effect. Finally, EIOU Loss is used as the loss function of the bounding box to improve the regression performance of the model. Additionally, tests are conducted on the sampled LLVIP infrared pedestrian image dataset and the results show that for pedestrian targets in infrared images, the model size is reduced by 80.6% and the number of parameters is reduced by 82.8% while maintaining a high detection accuracy ($AP=95.4\%$); and the inference speed is improved by 43.3% when using a CPU platform for inference, and the performance of detecting multi-scale targets is improved. The above two results validate the effectiveness of the algorithm.

Keywords: infrared pedestrian detection; YOLOv5; MobileNetv3; lightweight

基金项目: 国家自然科学基金项目 (No. 62061003); 四川省科技计划重点研发项目 (No. 2021YFG0192); 校级大学生创新创业项目 (No. S202210624201) 资助。

作者简介: 胡焱 (1973-), 男, 教授, 硕士生导师, 研究方向为航空电子设备维修测控。E-mail: huyan@cafuc.edu.cn

通讯作者: 赵宇航 (2001-), 男, 硕士研究生, 研究方向为深度学习目标检测。E-mail: 1127095630@qq.com

收稿日期: 2023-04-23

1 引言

传统的安防系统采用大量的监控摄像头,结合目前发展迅速的目标检测算法,实现安防智能化。然而,大多数系统使用可见光成像,容易受到温湿度、光照度等环境因素的影响,导致在某些场景下算法性能不佳,如夜间能见度不足等情况,无法满足安防需求。对此,可以采用红外热成像进行行人目标检测,与可见光成像系统形成互补,从而达到较好的性能。红外热成像技术有不受光照等环境因素影响的特点,相比于可见光有着更加突出的检测优势^[1]。但由于红外成像系统的限制,导致其存在目标的纹理信息较弱、对比度不佳、分辨率低等检测弱点^[2]。如何提高红外行人目标检测性能仍是目前研究的热点之一。

目标检测是计算机视觉领域中的一个重要研究方向,其应用范围广泛,包括安防监控、智能交通等领域。近些年来,深度学习相关研究发展迅速,已被广泛应用于如计算机视觉、自然语言处理、语音识别等领域。将深度学习与目标检测相结合是一个热门的研究方向。目前对于可见光成像目标检测研究方面,已经取得了较大的进展,由于深度学习的方式需要大量的数据喂养,借助现有的大型可见光图像数据集如 MSCOCO、VOC 等,以及容易获取的可见光图像,基于深度学习的可见光目标检测已经具有了较好的性能,但是对于红外热成像此类图像,由于现存红外图像公开数据集较少,所需数据较难获得,因此相关的应用研究较少。此外,在许多领域,目标检测算法的应用需要将其部署到嵌入式设备,而此类设备的存储与性能限制了部署模型的尺寸,因此如何有效利用深度学习方式提高红外目标检测性能,以及实现算法轻量化,取得精度与速度的平衡,是现在需要解决的难题。

随着深度学习技术的发展,与目标检测相关的算法也在被不断的提出。目前主流的检测算法分为 Two stage 和 One stage 两种类型^[3]。Two stage 类型算法需要先进行 Region Proposal,也就是目标候选框的生成,再对其进行分类与回归。典型算法如 SPP-Net^[4]、FastR-CNN^[5]、FasterR-CNN^[6]等;One stage 类型算法直接在网络中提取特征,预测目标的类别与位置。相关算法如 SSD^[7]、RetinaNet^[8]以及 YOLO^[9]系列。众多算法中,YOLO 系列凭借其较快的检测速度,成为当前的热门实用级算法。在红外图像行人检测研究方面,蔡伟等人^[10]基于 YOLO 检测框架,通过减少下采样次数,结合跨阶段局部模块、Focus 结构和

空间金字塔池化结构设计了特征提取网络,对红外弱小目标的特征提取性能较好;王晓红等人^[11]提出了一种主要用于检测小尺寸行人的 YOLOv5-p4 模型,通过 BiFPN 特征融合以及增加小目标检测层,避免了小目标的特征信息丢失;张立国等人^[12]在 YOLOv3 算法基础上,通过改进轻量化特征提取网络,特征融合方式以及解耦检测头,实现了模型的轻量化;张印辉等人^[13]通过在 YOLOv4-tiny 结构上增加精细尺度检测层,用残差结构加深网络以获取红外行人位置特征,对于模型参数的增加使用通道剪枝以及知识蒸馏方法进行轻量化,实现了检测精度与速度的平衡。

本文的主要目的是在保持一定的精度要求下,改进原有网络结构使其轻量化。对此,本文提出了一种基于 YOLOv5 算法的轻量化方法,通过更换轻量级的特征提取网络以及引入深度可分离卷积,降低模型参数量及浮点运算量;使用具有更大感受野的轻量级上采样算子进行特征重组,提升模型性能;将 YOLOv5 原有损失函数中的 CIoU 替换为效果较好的 EIoU,提高回归精度。实验结果表明,本文算法在保持较高检测精度的同时,显著降低了模型参数量,具有较好的轻量化性能。

2 YOLOv5 网络模型改进

2.1 网络结构改进分析

由于需要对网络进行轻量化改进,模型复杂度不可过大,因此本文所进行的改进均是在 YOLOv5s 6.0 基础上进行。在 YOLOv5s 6.0 中,其整体网络结构如图 1 所示。

其中,网络结构被分为了 Input 输入端、Backbone 特征提取层、Neck 以及 Detect 检测输出层四部分。Input 部分对图像进行数据增强等预处理以及自适应锚框计算;Backbone 部分主要依靠各种卷积操作对输入图像进行不同层次的特征提取;Neck 部分使用了 FPN (Feature Pyramid Networks) 和 PAN (Path Aggregation Network) 结构来加强语义特征和定位特征,从而丰富特征信息;Detect 部分设计了 large、medium、small 三种检测分支,用于不同尺度下的检测目标,对其预测类型及位置信息,并给出置信度。YOLOv5 模型的检测性能已达到较高水平,但该算法对部署硬件有一定要求,需要较高的硬件成本来实现高精度的检测性能,应用到行人检测等实际情况下仍有一定缺陷。对此本文通过改进特征提取网络以及使用深度可分离卷积计算来降低模型复杂度,对于改进带来的精度损失,通过更换上采样模块以及改进损失函数来提升检测性能。

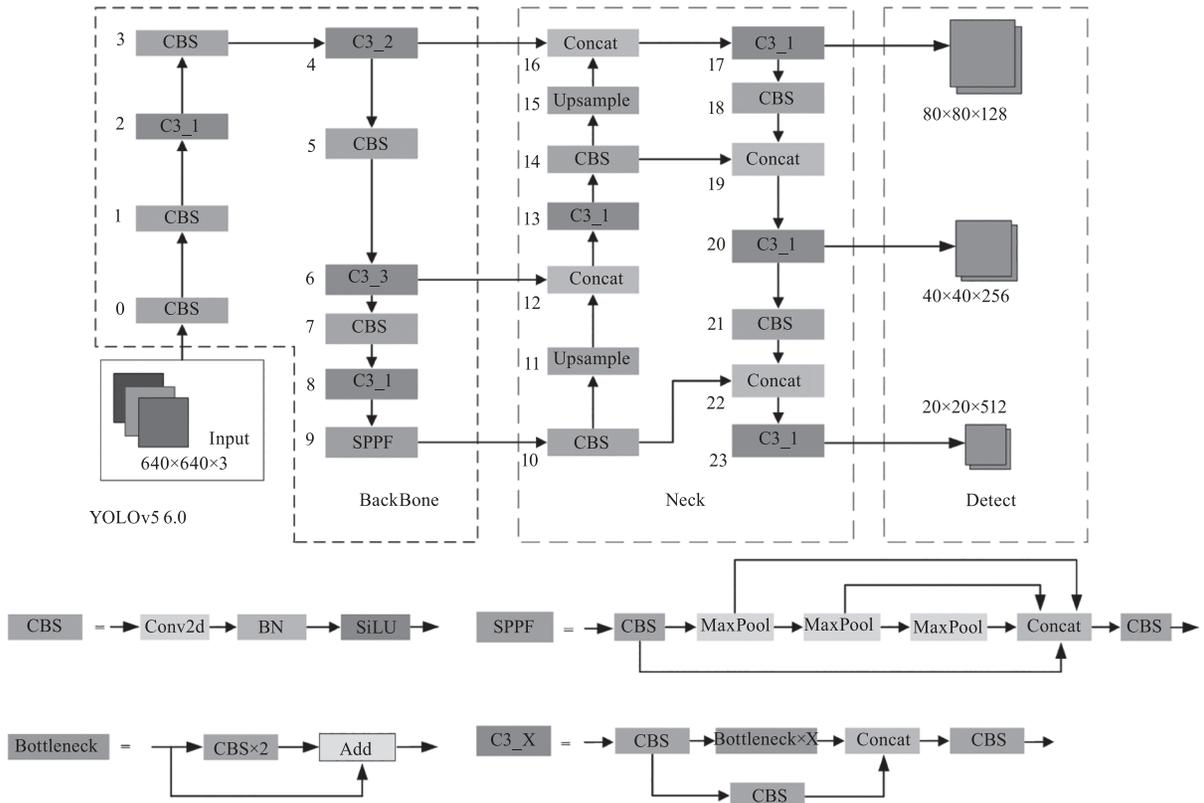


图1 YOLOv5s 6.0 网络结构
Fig. 1 YOLOv5s 6.0 network architecture

2.2 轻量化特征提取网络

YOLOv5s 初始模型的特征提取层主要分为了 Conv 模块、CSPDarkNet-53 和 SPPF 模块。由于其网络模型参数多,对计算性能要求较高,难以用于嵌入式等实际应用平台及设备。对此,本文选择使用 MobileNetv3 特征提取网络代替原有骨干层,从而在保证一定精度的前提下减少运算量,缩减模型大小。MobileNetv3 网络^[14]是一种轻量级的卷积神经网络架构,其设计目的是使得网络更适合移动端部署,在 CPU 设备上实现高精度和效率的平衡。其 Block 结构如图 2 所示,包含了 v1 中的深度可分离卷积和 v2 中具有线性瓶颈的倒残差结构,前者用于降低模型的计算量,后者用于增加模型的代表能力。相较于之前版本,v3 版使用了 h-swish 激活函数代替之前的 swish 函数,使得计算速度加快,量化效果更好,其表达式如下式:

$$h-swish[x] = x \frac{ReLU6(x+3)}{6} \quad (1)$$

同时加入了轻量级的 SE (Squeeze and Excitation) 通道注意力机制来学习通道之间的相关性,为比较重要的通道添加更大的权重,提高了模型的性能。

通过更换 MobileNetv3 特征提取网络,可以在保持一定精度的情况下,显著减少运算量,缩减模型大

小,从而使得 YOLOv5s 更加适合在资源受限的场景中使用。

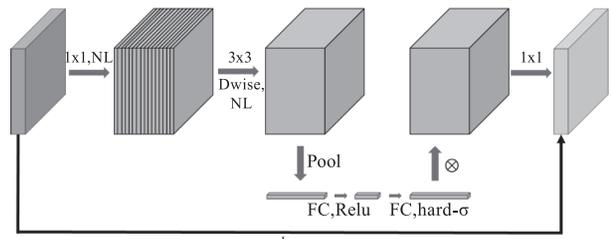


图2 MobileNetv3 Block 结构
Fig. 2 Structure of the MobileNetv3 Block

2.3 深度可分离卷积

深度可分离卷积 (Depthwise Separable Convolution) 相较于标准卷积操作,其参数量以及运算成本显著降低,且精度降低较少。其主要分为两步操作:深度卷积 (Depthwise Convolution) 和逐点卷积 (Pointwise Convolution)。具体流程示意如图 3 所示。对于输入图像,首先在每个通道上分别进行卷积操作,一个通道只被一个卷积核卷积,特征图通道数不变,因此需要后续的 1x1 逐点卷积运算来将上一步的特征图在深度方向上进行加权混合,生成新的特征图。

设输入特征图尺寸为 $C_1 \times H \times W$, 输出时通道数变为 C_2 , 则卷积核尺寸为 $C_1 \times K \times K$, 共有 C_2 个

卷积核,其中 H 、 W 分别为特征图高度及宽度, K 为卷积核尺寸。标准卷积运算与深度可分离卷积运算的参数量及运算量对比如表 1 所示。

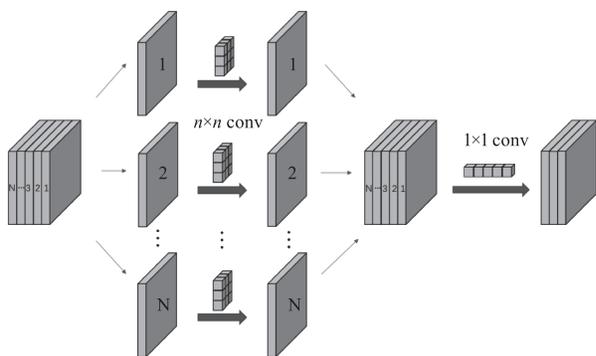


图 3 深度可分离卷积示意图

Fig. 3 Schematic diagram of depth-wise separable convolution

表 1 两种卷积参数及运算量对比

Tab. 1 Comparison of parameters and computational cost between two types of convolutions

类型	Parameters	FLOPS
Conv	$C_1 \times C_2 \times K \times K$	$C_1 \times C_2 \times H \times W \times K \times K$
DWConv	$C_1 \times K \times K + C_1 \times C_2$	$C_1 \times H \times W \times K \times K + C_1 \times C_2 \times H \times W$

根据表中公式,可计算出两者参数量之比:

$$\frac{P_{Dw}}{P} = \frac{1}{C_2} + \frac{1}{K^2} \quad (2)$$

两者计算量之比:

$$\frac{F_{Dw}}{F} = \frac{1}{C_2} + \frac{1}{K^2} \quad (3)$$

显然,式(2)、(3)的值相等且小于 1,说明使用 DWConv 运算代替标准 Conv 运算可以减少参数量及计算量,实现模型轻量化,将算法移植到实用场景具有可行性。本文将 Neck 部分的 conv 层替换为 DWconv 层,从而减小了模型尺寸。

2.4 上采样模块

YOLOv5s 中的 Neck 部分中使用了上采样操作来对上层的特征图进行尺寸扩充,并与浅层信息逐元素相加,从而构建出不同尺寸的特征金字塔结构。原有模型中的上采样操作使用最近邻插值法,该方法所生成的新特征图中的每个像素的值是由原特征图上最近邻的像素值所决定的。这样的操作运算快,但是未考虑到周围像素的影响。会导致特征图出现锯齿状伪影,采样效果较差;对此本文使用 CARAFE (Content-Aware ReAssembly of FEatures) 方法^[15]替代最近邻插值,通过引入一定的计算量换取较高精度,从而提高上采样操作的效果和性能。

CARAFE 是一种基于内容感知的特征重组方法,其主要思想是在上采样过程中,对每个上采样点周围的局部区域进行自适应特征重组,从而使得重组后的特征更能够适应目标区域的细节和纹理信息。首先对于输入的低分辨率特征图,在每个像素点位置上应用卷积操作,得到一组权重;然后使用最近邻插值法将该特征图上采样到目标分辨率,得到一个更大的特征图;接着将该特征图与前一层的高分辨率特征图进行级联,产生一个上下文感知的特征图;最后结合前面的权重和级联特征图,生成一个内容感知的上采样结果。整个过程的目的是提高特征图的语义信息,并且减少上采样过程中的信息损失,提高图像重建性能。将其应用到 YOLOv5s 网络中,可以提高网络的检测性能和效果。

2.5 损失函数

在 YOLOv5 模型中,其损失函数由三部分组成:边界框损失 (bbox_loss)、分类损失 (cls_loss) 以及置信度损失 (obj_loss),其计算公式为:

$$Loss = G_b \times bbox_loss + G_c \times cls_loss + G_o \times obj_loss \quad (4)$$

式中, G_b 、 G_c 、 G_o 为损失权重。YOLOv5s 6.0 版本中,计算分类损失与置信度损失的函数为 BCE Loss,计算边界框损失的函数为 CIOULoss, CIOU 对于目标框的中心点偏移的惩罚较为合理,但其存在宽和高不能同时增减的问题。对此本文使用了 EIOULoss^[16],公式如下:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (5)$$

式中, C_w 、 C_h 为覆盖两个目标框的最小外接框的宽和高, ρ 为 b 和 b^{gt} 之间的欧式距离, w 、 h 为预测框宽高, w^{gt} 、 h^{gt} 为真实框宽高。EIOULoss 包含了重叠、中心距离、宽高三种情况的损失,对目标框及锚框的宽高分别计算,解决了 CIOULoss 中宽和高不能同时增大或者减小的问题,使两框之间的宽高差更小,回归精度更高。

3 数据集及评价指标

3.1 数据集设置

本文实验所采用的数据集为 LLVIP 红外行人检测公开数据集^[17],这是一种弱光条件下的可见光-红外多帧配对数据集。该数据集共包含 30976 张图像,其中的 15488 张红外图像分辨率为 1080×720 ,在 26 个不同场景下采集,采集波段为 $8 \sim 14 \mu\text{m}$,多数采

集场景为弱光条件,且图像为连续采集,图像中包含大量行人且均已打上标记,适合红外场景下的行人检测研究。但该数据集中行人尺寸多为中等大小,对于远距离小目标的行人检测研究尚有缺陷。对此本文选择对原数据集进行稀疏化抽样并加入自行拍摄的远距离小目标行人图像,形成自建优化数据集。具体实现方式是对原数据集中的红外图像随机抽取 3500 张图片,并且加入 356 张自行拍摄的含有远距离小目标行人的红外图像,形成融合自建数据集。图像采集设备为 TiX650 红外热成像仪,为确保一致性,使用与 LLVIP 数据集同样波段进行拍摄。对于拍摄过程中行人较为模糊的图像,为增强目标特征,使用专业软件进行高斯滤波器锐化。最终形成含 3856 张弱光环境下行人特征的红外融合数据集,其具体数据集指标见表 2,其中按 9:1 的比例随机分配训练集和验证集,推理测试采用另行拍摄红外行人视频。

表 2 融合数据集

Tab. 2 Merged dataset

Dataset	Number	Resolution	Camera angle	Pedestrian
Sampling	3500	1080 × 720	surveillance	√
Addition	356	640 × 480	ground	√

3.2 模型评价指标

对于深度学习目标检测来说,常用的模型评价指标为精确率(Precision)、召回率(Recall)、均值平均精度(mean average precision, mAP)等^[3]。此外,模型大小(Model size)以及参数量(Parameters)也是反映模型轻量化性能的重要指标。由于本文仅针对红外图像中的行人这一类目标进行分析,因此最终选择 Precision、Recall、AP@0.5、Model size、Parameters 五类作为实验评价指标。相关计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F_{AP} = \int_0^1 P(R) dR \quad (8)$$

上述公式中,TP 为真正例(True Positive),即检测行人目标正确的次数;FP 表示假正例(False Positive),即将其他类型目标错检为行人目标的次数;FN 表示假负例(False Negative),即未检测到应是行人目标的次数。P(R) 是 Precision-Recall 曲线,即以 Precision 为坐标纵轴,Recall 为坐标横轴的曲线,其与坐标轴所围面积即为 AP 值。

4 实验设计及结果分析

4.1 实验环境

本文所有训练及测试实验均在 PyTorch 深度学习框架下进行,Pytorch 版本为 1.12.1,实验设备操作系统为 Windows 11,CPU 为 AMD Ryzen 7 5800H with Radeon Graphics, RAM 为 16G, GPU 为 NVIDIA GeForce RTX 3060 Laptop,显存 6G。为节省训练时间,训练实验均使用 GPU 加速;为验证模型在移动端的推理性能,模型推理实验均使用 CPU 进行。为确保实验一致性,在模型训练阶段均采用同样的超参数配置。表 3 为相关训练初始化参数。

表 3 训练初始化参数

Tab. 3 Training initialization parameters

Initial Weights	None
Epoch	100
bs	16
Imgsz	640
Lr	0.01
Momentum	0.937
Weight_decay	0.0005

4.2 轻量化改进及性能分析

根据本文改进思想,首先通过更换 YOLOv5s 的特征提取网络以及引入 DWconv 运算来使模型初步轻量化,并对其轻量化后参数及性能进行分析,之后加入 CARAFE 模块,以较少的参数量提升换取检测精度的提高,最后更换 EIOULoss 提升回归性能,并对最终模型进行参数及性能分析。为测试各模块改进效果,对增加各模块前后进行训练测试,得到改进前后算法在自建融合数据集上的训练效果。表 4 及表 5 为训练后评测结果。

表 4 改进前后 PR 值对比

Tab. 4 Comparison of PR values before and after improvement

类型	Precision/%	Recall/%
Baseline	95.1	92.2
Lightweight	93.5	89.8

对表 4 中数据分析可知,轻量化算法相较于原 YOLOv5s 算法,其对验证集中行人检测时的 FP 以及 FN 数量增加,导致其 P、R 值有所降低,精度稍许下降;对表 5 中数据分析可知,原 YOLOv5s 算法对于红外图像下行人的检测精度较高,AP 可达 96.9%,但其模型大小、浮点运算次数以及参数量

也较高,不利于部署;在更换 MobileNetv3 特征提取网络后,其检测精度略有下降,模型大小减小 11.3MB,浮点运算次数减少 13.6 GFLOPs,参数量减少 5.65×10^6 ,显著减小了模型的尺寸;将 Neck 部分的四个标准 Conv 替换为 DWconv 后,模型尺寸进一步减小,且精度有所提升;添加 CARAFE 上采样

模块以及更换 EIOULoss 后,通过增加模型复杂度,提升了部分模型检测精度,在保证模型尺寸较小的情况下恢复了模型的检测性能。

将原 YOLOv5s 算法与轻量化改进后算法用于 LLVIP 红外行人数据集训练,测试其性能差距,所得结果见表 6。

表 5 改进前后性能参数对比

Tab. 5 Comparison of performance parameters before and after improvement

Model				AP/%	size/MB	GFLOPs	Params/ $\times 10^6$
MobileNetv3	DWconv	CARAFE	EIOU				
Baseline(YOLOv5s 6.0)				96.9	14.4	15.9	7.02
√	×	×	×	94.2	3.1	2.3	1.37
√	√	×	×	95.1	2.7	2.0	1.17
√	√	√	×	95.2	2.8	2.1	1.21
Lightweight(all improvements)				95.4	2.8	2.1	1.21

表 6 LLVIP 数据集测试结果

Tab. 6 LLVIP dataset test results

类型	Number of images		Precision/%	Recall/%	AP/%	size/MB
	Train	Val				
Baseline	13939	1549	96.2	95.9	98.4	14.4
Lightweight			94.8	94.2	97.7	2.8

由于训练样本的增加,相较于自建数据集训练结果,模型的 P、R 值以及 AP 均有所提升。测试结果表明改进后算法在 LLVIP 数据集上仍能维持较高检测精度,在显著降低模型尺寸的情况下仍具有较好的检测水平。相较于原算法,本文改进方法在精度略有降低的情况下达到了轻量级水平。

4.3 检测效果分析

为测试改进模型的检测效果,对后续另行拍摄的红外行人视频进行检测。分别使用原算法与改进算法在自建融合数据集上训练完成后的检测性能最好的模型权重,对同一段红外行人视频进行推理测试。推理所用视频采用 TiX650 红外热成像仪拍摄,共包含 88 帧图像。其中包含了多尺度的行人目标,可验证模型对不同大小的行人目标的检测能力。推理实验结果见表 7。

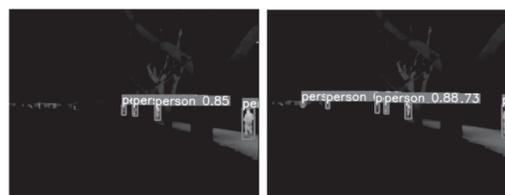
表 7 推理测试结果

Tab. 7 Reasoning test results

类型	Inference/ms	Number of detected targets
Baseline	117.2	473
Lightweight	66.4	596

结果表明改进后算法推理时间较原算法降低

50.8 ms,推理速度提升 43.3%,且对视频中小尺度行人检测效果更佳,增加了模型对多尺度目标的检测能力。为对比检测效果,选择同一时间戳下的视频帧进行对比,其可视化结果如图 4 所示。可以看到,改进后算法可检测到更多小目标行人,模型检测能力有所提升。



(a)原算法 (b)改进后算法

图 4 检测效果对比图

Fig. 4 Comparison of detection effects

5 结论

对于红外图像下行人目标检测模型尺寸过大,导致不易于部署至移动端的问题,本文提出了一种轻量化 YOLOv5s 模型的方法。通过应用轻量化特征提取网络,极大地减小了模型尺寸,同时改进上采样模块以及更换边界框损失函数,降低轻量化所导致的检测精度损失。实验结果表明本文改进方法能在保证较高检测性能的同时,极大降低模型参数量,更有利于实际安防环境的硬件要

求。后续研究将着力于实现模型工程应用,将模型移植至移动端平台等设备,探寻实时检测能力方面的改进。

参考文献:

- [1] YI Shi, NIE Yan, ZHANG Yangyi. Nighttime target recognition method based on infrared thermal imaging and YOLOv3[J]. Infrared Technology, 2019, 41(10): 970 - 975. (in Chinese)
易诗, 聂焱, 张洋溢, 等. 基于红外热成像与YOLOv3的夜间目标识别方法[J]. 红外技术, 2019, 41(10): 970 - 975.
- [2] WANG Chen, TANG Xin-yi, GAO Si-li. Infrared image enhancement algorithm based on human vision[J]. Laser & Infrared, 2017, 47(1): 114 - 118 (in Chinese)
王晨, 汤心溢, 高思莉. 基于人眼视觉的红外图像增强算法研究[J]. 激光与红外, 2017, 47(1): 114 - 118.
- [3] XU Degang, WANG Lu, LI Fan. Review of typical object detection algorithms for deep learning U]. Computer Engineering and Applications, 2021, 57(6): 10 - 25. (in Chinese)
许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10 - 25.
- [4] Purkait P, Zhao C, Zach C. SPP-Net: deep absolute pose regression with synthetic views[J]. arXiv preprint arXiv: 1712.03452, 2017.
- [5] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
- [6] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 1137 - 1149.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [J]//Lecture Notes in Computer Science, 2016: 21 - 37.
- [8] Wang Y, Wang C, Zhang H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery [J]. Remote Sensing, 2019, 11(5): 531.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
- [10] CAI Wei, XU Peiwei, YANG Zhiyong, et al. Dim-small targets detection of infrared images in complex backgrounds [J]. Journal of Applied Optics, 2021, 42(4): 643 - 650. (in Chinese)
蔡伟, 徐佩伟, 杨志勇, 等. 复杂背景下红外图像弱小目标检测[J]. 应用光学, 2021, 42(4): 643 - 650.
- [11] WANG Xiaohong, CHEN Zheqi. Research on pedestrian detection in infrared image based on YOLOv5 algorithm [J]. Laser & Infrared, 2023, 53(1): 57 - 63. (in Chinese)
王晓红, 陈哲奇. 基于YOLOv5算法的红外图像行人检测研究[J]. 激光与红外, 2023, 53(1): 57 - 63.
- [12] ZHANG Li-guo, MA Zi-jian, JIN Mei, et al. An infrared image pedestrian detection method based on YOLO algorithm [J]. Laser & Infrared, 2022, 52(11): 1737 - 1744. (in Chinese)
张立国, 马子荐, 金梅, 等. 基于YOLO的轻量红外图像行人检测方法[J]. 激光与红外, 2022, 52(11): 1737 - 1744.
- [13] ZHANG Yinhui, ZHANG Pengcheng, HE Zifen, et al. Lightweight real time detection model of infrared pedestrian embedded in fine scale [J]. Acta Photonica Sinica, 2022, 51(9): 266 - 276. (in Chinese)
张印辉, 张朋程, 何自芬, 等. 红外行人目标精细尺度嵌入轻量化实时检测[J]. 光子学报, 2022, 51(9): 266 - 276.
- [14] 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.
- [15] Wang J, Chen K, Xu R, et al. Carafe: content-aware reassembly of features [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007 - 3016.
- [16] Zhang Z, Deng Z, Wu Z, et al. An improved EIoU-Yolov5 algorithm for blood cell detection and counting [C]//2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI). IEEE, 2022: 989 - 993.
- [17] Jia X, Zhu C, Li M, et al. LLVIP: a visible-infrared paired dataset for low-light vision [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3496 - 3504.