

DOI: 10.13382/j.jemi.B2407881

# 采用改进的单发多框检测器探测红外人体目标\*

罗体华 袁丽华 朱笑卢 超

(南昌航空大学无损检测教育部重点实验室 南昌 330063)

**摘要:**针对单发多框检测器(single shot multibox detector, SSD)的目标检测模型计算复杂度较高的问题,以及处理小目标、遮挡等情况存在鲁棒性差的问题,提出了一种改进的 SSD 红外人体目标检测方法,以满足智能监控实时性强的要求。首先,将 MobileNet V2 作为基础特征提取网络,取代了传统 SSD 中的 VGG16(visual geometry group network 16)骨干网络,通过深度可分离卷积降低计算量;然后,引入特征金字塔网络(feature pyramid network, FPN)结构,实现多尺度特征图像融合,增强浅层特征表征能力;最后,引入 SE(squeeze-and-excitation)通道注意力机制,动态学习通道权重以聚焦关键特征,提高了模型对浅层特征的表征能力以及对主要通道信息的关注度。在自建 IR-HD 数据集上的实验表明,改进后的 SSD 模型检测精度提升了 1.3%AP@0.5、14.3%AP@0.75,模型推理速度提高了 3.835 fps。结论表明,该方法通过轻量化设计、特征融合与注意力机制协同优化,在保障检测精度的同时显著提升实时性,对红外小目标、遮挡场景具有强鲁棒性。

**关键词:** 红外目标检测;单发多框检测器;注意力机制;特征融合

**中图分类号:** TP391.41;TN219

**文献标识码:** A

**国家标准学科分类代码:** 520.20

## Infrared human target detection by improved single shot multibox detector

Luo Tihua Yuan Lihua Zhu Xiao Lu Chao

(Key Laboratory of Nondestructive Testing (Ministry of Education), Nanchang Hangkong University, Nanchang 330063, China)

**Abstract:** To address the issue of high computational complexity in the single shot multibox detector (SSD) model and its poor robustness in handling small targets and occlusions, an improved SSD-based infrared human target detection method is proposed to meet the real-time requirements of intelligent surveillance. First, MobileNet V2 is used as the backbone feature extraction network, replacing the traditional visual geometry group network 16(VGG16) network in SSD, which reduces computational cost through depthwise separable convolutions. Then, a feature pyramid network (FPN) structure is introduced to achieve multi-scale feature fusion, enhancing the representation ability of shallow features. Finally, the squeeze-and-excitation (SE) channel attention mechanism is incorporated to dynamically learn the channel weights, focusing on key features and improving the model's attention to important channel information. Experimental results on the self-built IR-HD dataset show that the improved SSD model's detection accuracy is increased by 1.3%AP@0.5 and 14.3%AP@0.75, while the model's inference speed improves by 3.835 fps. The conclusion indicates that this method, through lightweight design, feature fusion, and attention mechanism collaboration, significantly enhances both detection accuracy and real-time performance, demonstrating strong robustness in infrared small target and occlusion scenarios.

**Keywords:** infrared human target detection; single shot multibox detector; attention mechanism; feature fusion

## 0 引言

红外人体目标检测旨在通过计算机技术判断图像或

视频中是否存在行人并精确定位,这对视频监控、自动驾驶和智能机器人等领域至关重要,然而人体兼具刚性与柔性特性,其外观易受穿着、视角等因素影响,成为检测中的主要难题。现阶段特征提取方法主要分为两类:基

于人工设计的传统方法和基于神经网络的方法<sup>[1-3]</sup>。目标检测算法主要分为基于回归的目标检测算法和基于区域的目标检测算法<sup>[4]</sup>两大类。

基于区域的目标检测算法也叫双步检测法,其主要过程就是先对图像的候选区域进行选取,再对候选区域进行分类和定位。Bui 等<sup>[5]</sup>受卷积神经网络<sup>[6]</sup>(convolutional neural network, CNN)的启发,因而提出区域卷积神经网络(regions with CNN features, R-CNN)用于目标检测,该方法主要由候选区域选取模块、特征向量提取模块以及线性支持向量机(support vector machines, SVM)<sup>[7]</sup>的分类模块三大模块构成,分类后的特征向量采用非极大值抑制<sup>[8]</sup>筛选区域候选框。在选取候选框时使用选择性搜索<sup>[9]</sup>导致候选区域框数量过多计算量复杂且对于不同尺寸的候选框直接变为相同尺寸进行后续处理,导致原图像信息丢失。针对图像信息丢失问题, Ma 等<sup>[10]</sup>提出空间金字塔池化网络(spatial pyramid pooling network, SPPNet)通过在最后一层卷积后加入空间金字塔池化(SPP)解决了尺寸限制问题,但存储需求高。为了进一步优化, Maity 等<sup>[11]</sup>再提出将特征提取、区域分类和边界回归 3 个模块整合在一起的 Fast R-CNN,主干网络采用 VGG16<sup>[12]</sup>取代 AlexNet<sup>[13]</sup>,通过特征图映射候选框一次提取特征,再经奇异值(singular value decomposition, SVD)<sup>[14]</sup>分解同时输出分类结果和窗口回归结果,计算量大大减小的同时提升了检测速率。然而,候选区域提取速度和召回率仍是瓶颈, Yi 等<sup>[15]</sup>为了在选取候选区域阶段更省时,提出了 Faster R-CNN,该方法通过引入区域候选网络(region proposal network, RPN)提取候选区域,极大加速候选框生成,并实现端到端检测(集合特征提取、候选提取和回归),但在目标背景区分度低时精度会下降。

基于回归的目标检测算法又称为单步检测法,由于这类算法实时性强,而且处理速度快。为了改善双步检测法中参数多,训练时间长等问题, Kotar 等<sup>[16]</sup>首次提出基于回归的目标检测算法(you only look one, YOLO),它不需要预先选取候选框。具体方法就是将图像划分为网格直接预测边界框和置信度,无需候选框,速度快但易漏检多目标、定位不准且泛化能力弱。YOLOv2<sup>[17]</sup>引入锚框机制并采用 DarkNet19<sup>[18]</sup>作为主干网络,加入批量归一化,改善了定位但小目标召回率不高。针对这一系列问题, Redmon 等<sup>[19]</sup>提出 YOLOv3,该方法将 YOLOv2 中的主干网络改为 DarkNet53,为避免浅层特征的丢失以及梯度消失等问题,引入了残差网络 ResNet<sup>[20]</sup>,提高了小目标的检测精度。该方法检测速度快,背景误检率低<sup>[21]</sup>,但由于在处理过程中进行了多次下采样,对目标特征提取有一定的影响,于是 Cai 等<sup>[22]</sup>提出了 SSD,该模型的主干网络采用 VGG16 并构建多尺度特征金字塔进

行预测,有效检测小目标,但 VGG16 参数多导致计算量大、复杂度高。

尽管 SSD 存在计算量大的缺点,但其检测精度在现有方法中表现优异,因此目前大部分行人检测方法都是基于 SSD 框架进行的改进<sup>[23]</sup>。

本文针对于红外视频流中存在的人体目标对比度低、纹理细节匮乏、目标遮挡、多尺度等问题,同时为了满足人体目标识别的实时性需求,提出一种融合注意力机制的改进 SSD 红外人体目标检测算法,为提升模型检测效率,可先把 FPN 层级特征与 SE 注意力模块融合,再重构模型主干网络以及提高模型对红外图像中人体目标有用信息的专注能力。

## 1 改进的单发多检测器

### 1.1 MobileNet V2 轻量化网络

针对智能红外动作监控的实时检测的任务要求,将 SSD 目标检测架构中的骨干网络 VGG16 更改为轻量化网络 MobileNet V2<sup>[24]</sup>,作为基础特征提取层,既能保证模型精度的也能提升检测速度,保证全辐射热像视频流中红外人体目标的实时检测需求。

深度可分离卷积模块组成了 MobileNet V1 核心架构,涵盖深度卷积(depthwise convolution)与逐点卷积(pointwise convolution), Depthwise 以 3×3 卷积核进行操作,通道间卷积核不共用; Pointwise 采用 1×1 卷积核,可使收敛加快、泛化表现优化,卷积运算之后接入批量归一化层(batch normalization, BN),再经线性整流激活函数(rectified linear unit, ReLU)激活,图 1(a)为深度可分离卷积的主要构成,该架构核心理念是兼顾准确和效率,极大简化计算和参数存储。在输入输出图像尺寸相同情况下,标准卷积的计算量  $A_{Conv}$  表示为:

$$A_{Conv} = D_K \times D_K \times M \times N \times D_F \times D_F \quad (1)$$

深度可分离卷积的计算量  $A_{DConv}$  表示为:

$$A_{DConv} = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (2)$$

式中:  $D_K$  表示卷积核的大小;  $D_F$  表示为输入与输出特征图的尺寸;  $M$  表示输入特征图的通道数;  $N$  为输出特征图通道数。

$$\frac{A_{DConv}}{A_{Conv}} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} \quad (3)$$

通过两类卷积计算量的对比分析可知,深度可分离卷积运算量是标准卷积的  $\frac{1}{N} + \frac{1}{D_k^2}$ , 和标准卷积结构对比,采用这种卷积方式后计算效率得到显著改善。

MobileNet V2 在 V1 基础上进行改进,具体包括在 Depthwise Convolution 之前增加 Pointwise Convolution,引入了反向残差结构,MobileNet V2 的基本单元是瓶颈深度可分离卷积模块如图 1(b)所示。通过引入了反向残差和线性 Bottlenecks 线性瓶颈来实现了 MobileNet V2 网络的更高性能。

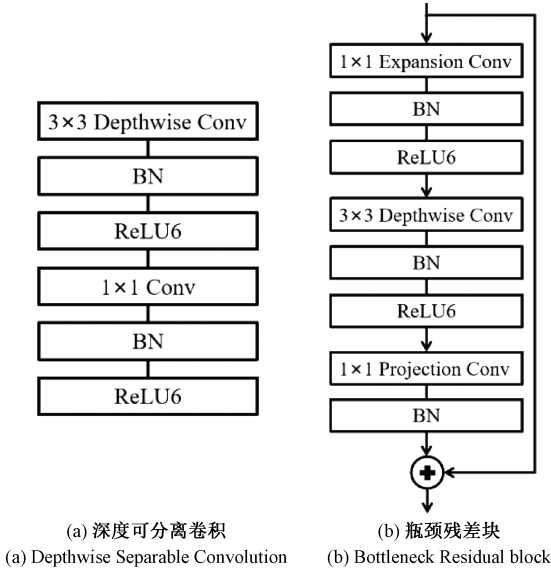
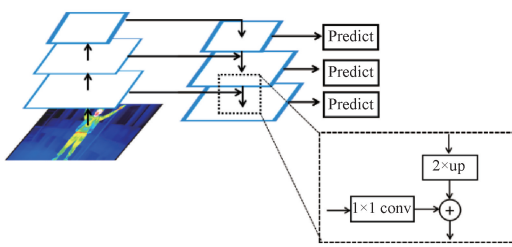
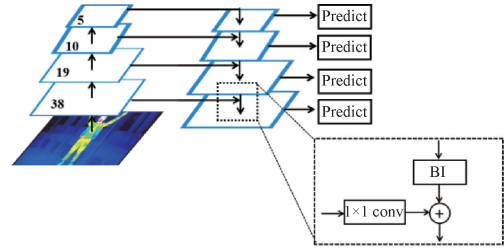


图 1 MobileNet 基本结构

Fig. 1 The basic structure of MobileNet



(a) Schematic diagram of FPN fusion



(b) Schematic diagram of feature fusion in this paper

图 2 特征融合结构示意图

Fig. 2 The schematic diagram of the feature fusion structure

SE 注意力机制结构由激励阶段开始,接着是压缩阶段,最后为重标定阶段。在压缩过程中,全局平均池化只是对特征图空间范围进行压缩,不会影响通道结构,即将输入维度为  $W \times H \times C$  的特征图压缩成  $1 \times 1 \times C$  输出,特征图全局压缩量为:

$$S_c = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H x_c(i, j) \quad (4)$$

式中:  $x_c \in \mathbf{R}^{W \times H}$  为输入的特征映射。

压缩后的特征图输入两个全连接层,为了增加非线性过程首先将特征数据进行降维处理(输出为  $1 \times 1 \times C/r$ ),然后升维(输出  $1 \times 1 \times C$ )的处理方式,再通过 Sigmoid

## 1.2 特征融合模块

多尺度特征融合模块参照特征金字塔网络的思想,如图 2(a)所示,融合分辨率有差异的特征图,在融合的特征图上对人体目标进行预测。高层特征分辨率低但语义信息丰富,在重构特征金字塔时,将分辨率越低的特征图置于底层,使得所有尺度的特征都具有丰富的语义信息。此方案不用显著增添运算负担,就能显著提升模型对小个子人体的识别精准度和检测成效。

为了对主干网络中提取的四层特征图进行融合,如图 2(b)所示,本文使用  $1 \times 1$  卷积层统一特征通道数量,并采用双线性插值(bilinear interpolation, BI)的方式对特征图进行上采样来统一特征图大小,然后使用线性叠加融合特征。

## 1.3 SE 注意力机制

红外视觉下人体特征语义信息匮乏,为提升目标检测精度,引入 SE 通道注意力模块,提升网络模型提取图像有用特征信息的能力。通道注意力机制,如 SE 模块通过动态调整通道权重来增强特征提取。它针对默认通道间权重均等的局限,建模通道间的相互依赖关系,并对特征进行重标定。该机制让网络拥有自动识别关键通道特征的本事,提高关键特征的分辨率,减少非明显通道的输出状况,由此提升网络对整体特征的提取能力,最终提升目标检测的精准性。

激活函数得到  $1 \times 1 \times C$  的特征图。

$$e_c = \sigma(w_2 \delta(w_1 s_c)) \quad (5)$$

式中:  $\delta(x) = \max(0, x)$  表示 ReLU 激活函数;  $\sigma(x) = \frac{1}{1 + e^{-x}}$  表示 Sigmoid 函数, ( $w_1 \in \mathbf{R}^{C/r \times C}, w_2 \in \mathbf{R}^{C \times C/r}$ )。

将输入的  $W \times H \times C$  的特征图与池化后的  $1 \times 1 \times C$  的特征图进行矩阵全乘,得到新的  $W \times H \times C$  特征图。通道间的相关性由两个全连接层与相应的激活函数构建,公式如下:

$$y_c = e_c \cdot x_c \quad (6)$$

式中:  $y_c = [y_1, y_2, y_3, \dots, y_c]$  表示输入特征映射与其对应通道权重参数相乘的结果。SE 注意力机制结构如图 3 所示。

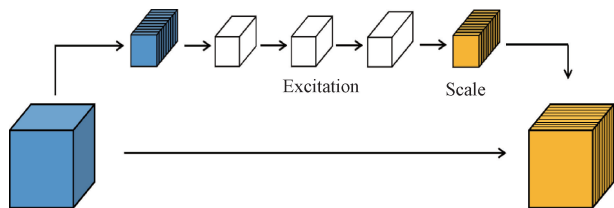


图 3 SE 注意力机制结构

Fig. 3 The structure of SE attention mechanism

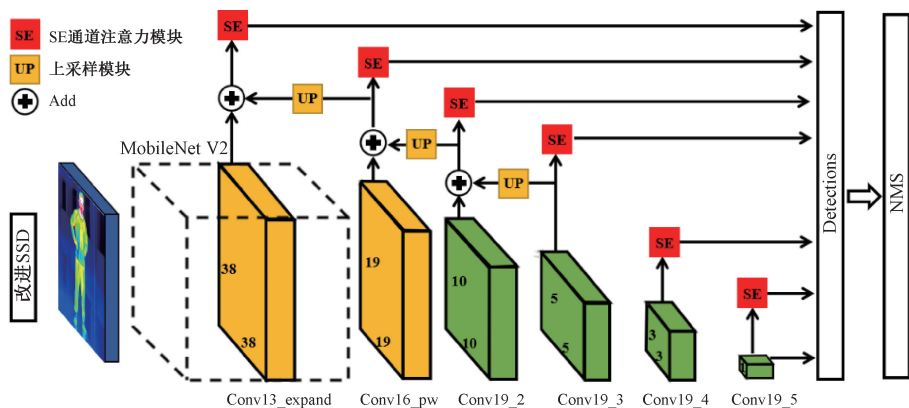


图 4 改进 SSD 网络模型结构

Fig. 4 The structure of modified SSD network model

## 2 实验与结果分析

### 2.1 数据采集

采用 VarioCAM<sup>®</sup> HD980 型长波红外 (long-wave infrared spectral range, LWIR) 红外热成像系统对人体数据集进行采集,构建红外人体目标检测数据集 (IR-HD 数据集)。为了保证数据集的多样性,反映真实的应用场景,系统考量影像采集角度、距离参数、环境温变及衣着厚度等外场变量,且涉及目标遮挡、多尺度特征及动作变化等识别瓶颈,在不同环境下对目标人体进行数据采集。IR-HD 红外人体数据集共 200 个全辐射热像视频,包含多目标和单目标,共 6 000 个红外图像,1 000 个红外人体目标。自制数据集在数量上依然不足,在模型训练过程中,为了获取更多的人体目标数据,对 IR-HD 数据集进行数据增强,采用水平旋转、垂直旋转、镜像旋转、旋转、噪声、模糊策略等数据增强策略,为训练提供丰富训练数据集,从而达到提升网络性能的目的。

### 2.2 评价指标

采用平均精度 (average precision, AP) 以及图像处理速度 (frames per second, FPS) 作模型评价指标。FPS 这

### 1.4 改进 SSD 模型架构

图 4 所示为改进后的 SSD 红外人体目标检测模型结构,采用 Top-Down 的行为估计模型。在特征提取时,使用 MobileNet V2 网络结构获取图像空间特征,鉴于末端特征图 Conv19\_4 与 Conv19\_5 仅包含单层级语义信息,浅层特征图上采样后获得的语义信息有限,还会额外消耗计算资源,就在 Conv13\_expand、Conv16\_pw、Conv19\_2、Conv19\_3 等 4 个特征层进行多尺度融合处理,给 6 个特征提取环节添入 SE 通道注意力单元,完成特征图的边界框回归和类别判断,借助级联式 NMS 处理模块,达成单阶段红外人体目标识别。

一参数可直接展现模型检测的及时性,体现网络模型实时处理的水准,首先说明交并比 (intersection over union, IoU),它被定义成预测框和真值框相交部分面积与二者联合面积的商值。

$$IoU = \frac{A \cap B}{A \cup B} \quad (7)$$

目标检测算法采用精确度 (precision)、召回率 (recall) 和 F 值 (F-measure) 以及平均精确率均值衡量算法结果。

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

式中:  $TP$  表示正确检测目标数;  $FP$  表示错误检测数目;  $FN$  表示漏检目标数;  $AP$  表示召回率在  $[0, 1]$  范围内的平均精值。

### 2.3 消融实验

为证明改进 SSD 在特征提取网络引入 MobileNet V2 网络和通道 SE 注意力机制的有效性,在 IR-HD 数据集上进行消融实验,分析改进 SSD 的性能,试验结果如表 1 所示。



表 1 改进 SSD 模型有效性验证

Table 1 Validity validation of modified SSD model

SSD	MobileNet v2	SE +FPN	AP@0. 5/%	AP@0. 75/%	帧率/fps	模型大小/( $\times 10^6$ )
√			95. 3	71. 0	7. 949	100. 3
√	√		89. 1	87. 6	16. 83	18. 4
√	√	√	96. 8	85. 6	11. 784	19. 7

为了验证 MobileNet V2 模型的有效性,将标准 SSD 模型的骨干网络更改为 MobileNet V2,模型的 mAP 下降了 0.65%,模型检测速度提高了 8.881 fps,结果表明 MobileNet V2 得益于瓶颈深度可分离卷积模块极大地提升了 SSD 模型的检测速度。

为了验证通道注意力机制与特征融合模块的有效性,在特征检测层上添加 SE 注意力机制模块并结合 FPN 结构,模型的 mAP 上升了 7.7%@0.5,帧率下降了 5.046 fps。表明 SE 注意力机制缓解了特征图之间的红外人体

目标信息不平衡的问题,提升了模型对主要通道信息的关注度。

综上所述,通过改变主干网络和融入注意力机制与特征融合模块,模型相较于 SSD 模型其 mAP 上升了 1.5%,帧率上升了 3.835 fps,证明了改进算法的有效性。因此在整个识别过程中,引入 MobileNet V2 和注意力机制对于提升模型检测效率和检测精度是可行的,对模型综合性能有较大提升左右。改进前后模型在 IR-HD 数据集上的人体目标检测结果如图 5 所示。

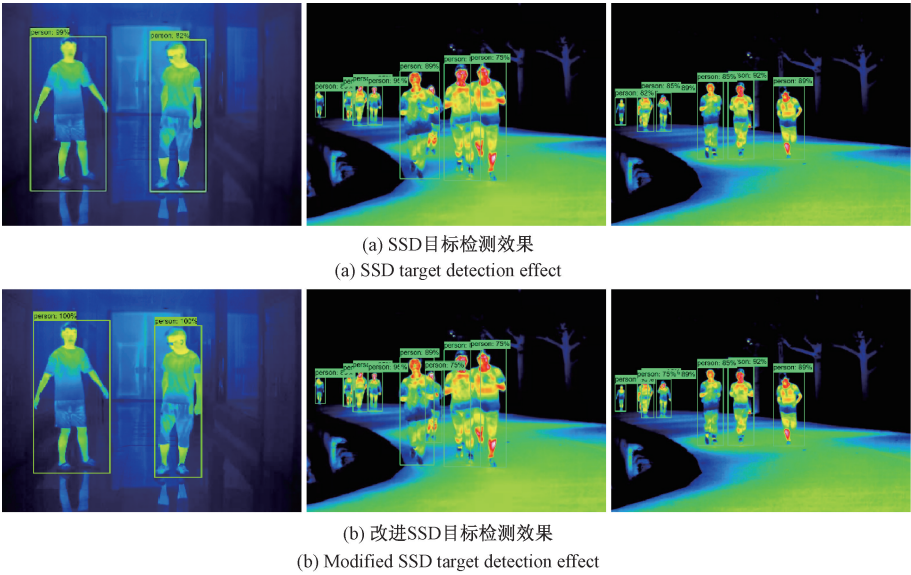


图 5 红外人体目标检测结果

Fig. 5 Results of infrared human target detection

图 5(a)、(b)3 个子图分别对应 3 种情况。对比存在人体遮挡以及多尺度复杂场景时改进前后模型的目标检测结果,可发现优化后的模型对常规目标检测置信度提升明显;对于人体目标遮挡情况改进后的目标识别精度提升,对于存在遮挡情况下的目标识别更为准确;对于多尺度目标检测问题,改进后的模型检测精度提升明显。综上,改进后的算法检测状态良好,并且在 CPU 上的检测速度达到了 11.78 fps,基本达到了实时检测的要求。模型检测率较高,对于多尺度、小目标、遮挡等情况下的人体目标检测结果都有较大提升。增加注意力机制增强了模型的检测精度,提升了模型检测置信度,由此可见注

意力机制是对红外图像信息表征匮乏、图像纹理细节少等问题有明显精度提升效果。

2.4 对比实验

为定量分析改进后模型的效果,与 Faster R-CNN 以及 YOLO V5 网络模型进行对比,相关模型检测指标如表 2 所示。由表 2 红外目标检测结果定量评估结果可知, Faster R-CNN 模型检测精度最高,在交并比为 0.5 时 AP 达到了 98.8%,但模型检测速在 CPU 上仅有 0.995 fps。其次,相较于 YOLOV5 模型,改进算法的检测精度在交并比为 0.5 时低了 0.7%,在交并比为 0.75 时上升了 2.8%,同时检测速度提高了 5.04 fps。与 CenterNet 和

EfficientDet 模型相比,在交并比为 0.5 时,模型检测精度分别提高了 0.8%、1.3%。在交并比在 0.75 时,检测精度提高了 7.6%、9.6%。模型的检测速度分别提高了 9.843 fps、8.58 fps。

表 2 红外人体目标检测结果检测精度分析

Table 2 Infrared human body target detection results detection accuracy analysis

模型	骨干网络	AP@ 0.5/%	AP@ 0.75/%	帧率/fps	模型大小/(×10 <sup>6</sup> )
Faster R-CNN	ResNet50	98.8	91.2	0.995	216
YOLO V5	CSPDarknet53	97.6	82.8	6.835	192
CenterNet	ResNet50	96.0	78.0	2.032	341
EfficientDet	EfficientNet	95.8	76.0	3.295	42.5
本文	MobileNet V2	96.8	85.6	11.875	19.7

3 结 论

针对应用场景实时检测的需求与红外图像空域表征信息匮乏的问题,通过改变骨干网络,引入注意力机制两种方式,提升目标检测精度与速度以及解决人体遮挡、小目标等问题。本文对 SSD 模型主要做了 3 个方面进行了改进以实现红外人体目标检测:1)为提升速度,将骨干网络 VGG-16 替换为轻量化的 MobileNet V2;2)为缓解红外图像特征匮乏及应对小目标、遮挡问题,采用特征金字塔融合方法,实现浅层特征图与深层特征图融合达到信息互补,提升小目标识别精度;3)针对不同通道表征能力差异,引入 SE 通道注意力机制,通过特征重标定自动学习通道重要性,增强关键特征并抑制弱表征通道,提升检测精度。

研究结果表明,改进后模型在精度与速度上综合优势显著,相较于传统 SSD,检测精度提升 1.3% AP@ 0.5 和 14.3% AP@ 0.75,推理速度提高 3.835 fps;虽然精度在 0.5 和 0.75 交并比下较 Faster R-CNN 低 2%和 5.6%,但检测速度优势明显,更适用于对速度要求高的红外人体行为识别任务,为后续红外人体姿态估计研究奠定了基础。

参考文献

[ 1 ] PATEL C I, LABANA D, PANDYA S, et al. Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences [ J ]. Sensors, 2020, 20(24): 7299.

[ 2 ] YANG J, GE H, YANG J, et al. Online multi-object tracking using multi-function integration and tracking simulation training [ J ]. Applied Intelligence, 2022, 52(2): 1268-1288.

[ 3 ] DING X, JI D. Specific patches decorrelation channel feature on pedestrian detection [ J ]. The International Arab Journal of Information Technology, 2021, 18(4): 493-503.

[ 4 ] 范丽丽,赵宏伟,赵浩宇,等. 基于深度卷积神经网络的目标检测研究综述 [ J ]. 光学精密工程, 2020,

28(5):1152-1164.

FAN L L, ZHAO H W, ZHAO H Y, et al. A review of object detection based on deep convolutional neural network [ J ]. Optics and Precision Engineering, 2020, 28(5):1152-1164.

[ 5 ] BUI Q H, SUHR J K. CNN-based two-stage parking slot detection using region-specific multi-scale feature extraction [ J ]. IEEE Access, 2023, 11: 58491-58505.

[ 6 ] ALJUAID H, ALTURKI N, ALSUBAIE N, et al. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning [ J ]. Computer Methods and Programs in Biomedicine, 2022, 223: 106951.

[ 7 ] LI X, DONG S, GUO S, et al. Applying support vector machines to a diagnostic classification model for polytomous attributes in small - sample contexts [ J ]. British Journal of Mathematical and Statistical Psychology, 2025, 78(1):167-189.

[ 8 ] AL-FURAIJI O J, TUAN N A, YUREVICH T V. A new fast efficient non-maximum suppression algorithm based on image segmentation [ J ]. Indonesian Journal of Electrical Engineering and Computer Science, 2020, 19(2): 1062-1070.

[ 9 ] JI X, YAN Q, HUANG D, et al. Filtered selective search and evenly distributed convolutional neural networks for casting defects recognition [ J ]. Journal of Materials Processing Technology, 2021, 292: 117064.

[ 10 ] MA X, GUO J, SANSOM A, et al. Spatial pyramid attention for deep convolutional neural networks [ J ]. IEEE Transactions on Multimedia, 2021, 23: 3048-3058.

[ 11 ] MAITY M, BANERJEE S, CHAUDHURI S S. Faster r-cnn and yolo based vehicle detection: A survey [ C ]. 2021 5th international conference on computing methodologies and communication ( ICCMC ). IEEE, 2021: 1442-1447.

[ 12 ] HRIDAYAMI P, PUTRA I K G D, WIBAWA K S. Fish species recognition using VGG16 deep convolutional

- neural network [J]. Journal of Computing Science and Engineering, 2019, 13(3): 124-130.
- [13] MINHAS R A, JAVED A, IRTAZA A, et al. Shot classification of field sports videos using AlexNet Convolutional Neural Network [J]. Applied Sciences, 2019, 9(3): 483.
- [14] COLACE F, CONTE D, DE SANTO M, et al. A content-based recommendation approach based on singular value decomposition [J]. Connection Science, 2022, 34(1): 2158-2176.
- [15] YI D, SU J, CHEN W H. Probabilistic faster R-CNN with stochastic region proposing: Towards object detection and recognition in remote sensing imagery [J]. Neurocomputing, 2021, 459: 290-301.
- [16] KOTAR K, MOTTAGHI R. Interactron: Embodied adaptive object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14860-14869.
- [17] LIU Z, CHEN Z, LI Z, et al. An efficient pedestrian detection method based on YOLOv2 [J]. Mathematical Problems in Engineering, 2018, 2018(1): 3518959.
- [18] HUANG Y C, HUNG K C, LIU C C, et al. Customized Convolutional Neural Networks Technology for Machined Product Inspection [J]. Applied Sciences, 2022, 12(6): 3014.
- [19] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [20] YI Z, YONG S, JUN Z. An improved tiny-yolov3 pedestrian detection algorithm [J]. Optik, 2019, 183: 17-23.
- [21] 徐印赞, 江明, 李云飞, 等. 基于改进 YOLO 及 NMS 的水果目标检测 [J]. 电子测量与仪器学报, 2022, 36(4): 114-123.
- XU Y Y, JIANG M, LI Y F, et al. Journal of Electronic Measurement and Instrumentation, 2022, 36 (4): 114-123.
- [22] CAI L, ZHANG Z, ZHU Y, et al. Bigdetection: A large-scale benchmark for improved object detector pre-training [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4777-4787.
- [23] 胡鑫, 欧阳华, 侯卓成, 等. 特征增强的 SSD 算法及在仪表读数识别上的应用 [J]. 电子测量技术, 2021, 44 (17): 154-159.
- HU X, OUYANG H, HOU ZH CH, et al. Feature-enhanced SSD Algorithm and Its Application in Meter Reading Recognition [J]. Electronic Measurement Technology, 2021, 44 (17): 154-159.
- [24] 孙曙光, 张婷婷, 王景芹, 等. 基于连续小波变换和 MTL-SEResNet 的断路器故障程度评估 [J]. 仪器仪表学报, 2022, 43 (6): 162-173.
- SUN SH G, ZHANG T T, WANG J Q, et al. Evaluation of Fault Degree of Circuit Breaker Based on Continuous Wavelet Transform and MTL-SEResNet [J]. Chinese Journal of Scientific Instrument, 2022, 43 (6): 162-173.

## 作者简介



**罗体华**, 2022 于福建江夏学院获得学士学位, 现为南昌航空大学硕士研究生, 主要研究方向为红外图像处理。

E-mail: 1597669860@qq.com

**Luo Tihua** received his B. Sc. degree from Fujian Jiangxia University in 2022. Now he is a M. Sc. candidate at Nanchang Hangkong University. His main research interest includes infrared image processing.



**袁丽华** (通信作者), 1998 年于南昌大学获得硕士学位, 2009 年于南京航空航天大学获得博士学位, 现为南昌航空大学教授, 主要从事红外无损检测技术方面的研究。

E-mail: lihuayuan@nchu.edu.cn

**Yuan Lihua** (Corresponding author), received her M. Sc. degree from Nanchang University in 1998 and Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 2009. Now she is a professor at Nanchang Aeronautical University. Her main research interest includes infrared non-destructive testing technology.