·自动化技术·

文章编号：1003－1251(2024)04－0001－06

基于 AS-YOLOv7 的红外目标检测算法

王艳梅[a]，李明明[a]，黄　月[b]，张艳珠[b]

（沈阳理工大学 a. 信息科学与工程学院，b. 自动化与电气工程学院，沈阳 110159）

摘　要：为解决传统目标检测算法在红外场景下的目标识别精度低、检测速度差等问题，提出基于 AS-YOLOv7 的红外目标检测方法。首先针对红外图像的特性，对 YOLOv7 模型中的特征提取网络进行重构并构建卷积块 CBH，保证模型的稳定性和计算速度；然后改进金字塔池化（SPPCSPC）模块，使模型在不增加任何参数的同时提高模型性能；最后将原模型中的损失函数替换为具有动态非单调的 WIoU 损失，解决红外数据集中质量较好和质量较差的样本间的不平衡问题，提高模型的泛化性。本文实验在自行采集的数据集上进行训练和验证，与原模型相比，AS-YOLOv7 的平均精度提升了 0.2%，参数量降低了 79%，计算量降低了 85%，推理时间降低了 32 ms。改进后的模型在保持准确度的情况下提高了模型的推理速度。

关　键　词：红外图像；YOLOv7；目标检测；特征提取

中图分类号：TP391.41　**文献标志码：**A　**DOI：**10.3969/j.issn.1003－1251.2024.04.001

Research on Infrared Target Detection Algorithm Based on AS-YOLOv7

WANG Yanmei，LI Mingming，HUANG Yue，ZHANG Yanzhu

（Shenyang Ligong University，Shenyang 110159，China）

Abstract：In order to solve the problems of low target recognition accuracy and poor detection speed of traditional target detection algorithms in infrared scenes，an infrared target detection method based on AS-YOLOv7 is proposed. First，according to the characteristics of infrared images，the feature extraction network in YOLOv7 model is reconstructed and the CBH module is established to ensure the stability and calculation speed of the model. Then，the pyramid pooling（SPPCSPC）module is improved forbetter model performance without adding any parameters. Finally，the loss function in the original model is replaced by a dynamic non-monotonic WIoU loss to solve the imbalance between samples with good quality and poor quality in the infrared data set and improve the generalization of the model. The experiment is validated on the self-collected data set. Compared with the original model，the average accuracy of AS-YOLOv7 is improved by 0.2%，the number of parameters is reduced by 79%，the amount of calculation is reduced by 85%，and the inference time is reduced by 32 ms. The improved model has a higher inference speed while maintaining accuracy.

Key words：infrared image；YOLOv7；target detection；feature-base

基于可见光的图像检测方法在不良环境下检测效果容易受到影响[1]。相比之下，红外成像技术无需借助环境光线，在夜间和烟雾等场景中具有优势，并被广泛应用[2]。红外图像存在模糊成

收稿日期：2023－06－09

基金项目：辽宁省教育厅高等学校基本科研项目（LJKZ0245）；装备预研国防科技重点实验室基金项目

作者简介：王艳梅（1969—），女，副教授，研究方向为图像处理；张艳珠（1971—），通信作者，女，副教授，博士，研究方向为自动控制、深度学习、图像处理。

像、低对比度和少量纹理信息等问题,会严重影响目标检测算法的准确性。因此,研究并设计一种快速、精准的红外图像检测算法具有重要意义。

为了解决红外图像检测所面临的困难,2023年,Shi等[3]利用坐标注意力和特征融合技术构建了名为CAFF-Net的红外目标检测模型,该模型通过深浅特征的融合策略提取结构、纹理和语义特征,避免误检,并利用坐标注意力机制进一步强调目标显著性,抑制背景干扰,从而有效降低复杂场景下红外小目标的误检和漏检率。同年,Li等[4]针对当前深度学习方法处理和融合红外图像重要特征不足所导致的漏检和误检等问题,提出了一种名为YOLOSR-IST的网络,该网络基于YOLOv5模型,使用坐标注意力机制向特征金字塔网络和路径聚合网络添加浅层特征图,并替换原始C3模块为基于移动窗口的层次化视觉自注意力的C3STR模块,使该网络在红外小目标检测方面表现更加精准。2022年,Yu等[5]提出了名为MLCL-Net的端到端全卷积网络,用于红外小目标检测,该网络构建局部对比学习结构,使网络在训练过程中学习局部对比特征图的生成,通过从同级特征图中提取和融合不同尺度的局部对比信息,充分挖掘目标的特征信息。此外,该网络采用双向特征金字塔结构将高级特征图与低级特征图进行融合,以提高对红外小目标检测的准确性。2021年,Ju等[6]构建了一种高效的端到端神经网络ISTDet,该网络先抑制背景信息以增强红外目标响应,再进行类别和位置推测,有效提高了红外图像的检测效率。2019年,Wang等[7]提出了用于红外小目标检测的MNET,限制下采样以获得足够大的特征图,并整合了网络各层输出实现精确的定位和分类。以上改进采用不同方法实现了对红外目标的检测,但仍存在参数量大、模型推理时间长、检测精度不高的问题。

为了使模型在保证检测精度的同时具有较高的运行效率,本文在YOLOv7的基础上提出基于AS-YOLOv7的红外目标检测算法,适应红外图像的检测任务,主要工作如下。

1)为更好地贴合实际情况,使用自行采集的红外图像数据集,并在模型中使用随机翻转等数据增强方法,提高模型的泛化性,同时对YOLOv7模型进行轻量化设计以提升模型的推理速度。

2)改进金字塔池化(SPPCSPC)模块,使改进后的SPPCSPC模块在保持目标感知的同时获得加速。

3)使用WIoU损失函数替换原有损失,提升预测框的回归准确度,增强模型的泛化能力。

1 模型设计

考虑到红外图像的特征,为了能够迅速且准确地进行红外目标检测,本文对YOLOv7网络模型进行了轻量化改进和优化。改进后的网络结构如图1所示,包含骨干网络(Backbone)、颈部网络(Neck)和检测头(Head)。Backbone由MP和ELAN-H模块组成,通过CBH卷积模块进行特征提取。其中ELAN-H将其他卷积层的输出作为输入,并通过多层卷积连接实现全连接操作,以提高模型学习能力。MP由最大池化和步长为2的3×3卷积组成,将输入的特征图下采样并经过卷积处理再进行全连接。在下采样时,同时考虑特征图的最大值和局部信息,以更好地保留重要信息。Neck部分则使用SPPFCSPC和ELAN-H模块聚合多尺度图像特征。SPPFCSPC模块由四个相同大小的池化层和卷积块组成,通过对骨干网络提取的最后一层特征进行卷积和最大池化,增加特征图的感受野并显著地分离出图像的上下文特征。网络模型的Head引入RepVGG改造其网络结构,在训练过程中可以通过多路分支提升性能且通过结构重新参数化实现推理速度的加快。

1.1 模型轻量化

原模型中的CBS卷积块使用Swish激活函数,具备无上界有下界、平滑、非单调的特性,在一定程度上提高了原模型的准确性。但Swish激活函数的计算成本高,求导复杂,在量化时计算较慢。为减轻计算成本的同时保证模型的准确性构建CBH模块,结构如图2所示。将Swish激活函数替换成HardSwish激活函数[8],该函数是在Swish基础上进行的改进,通过用线性的ReLU6函数取代指数类型的Sigmoid函数减少计算量,相较于Swish激活函数,在保证准确性的基础上有更好的稳定性和计算速度,其数学公式见式(1)。

$$\mathrm{HardSwish}(x)=\begin{cases}0, & x\leqslant -3\\ x\dfrac{\mathrm{ReLU6}(x+3)}{6}, & -3<x<3\\ x, & x\geqslant 3\end{cases} \tag{1}$$

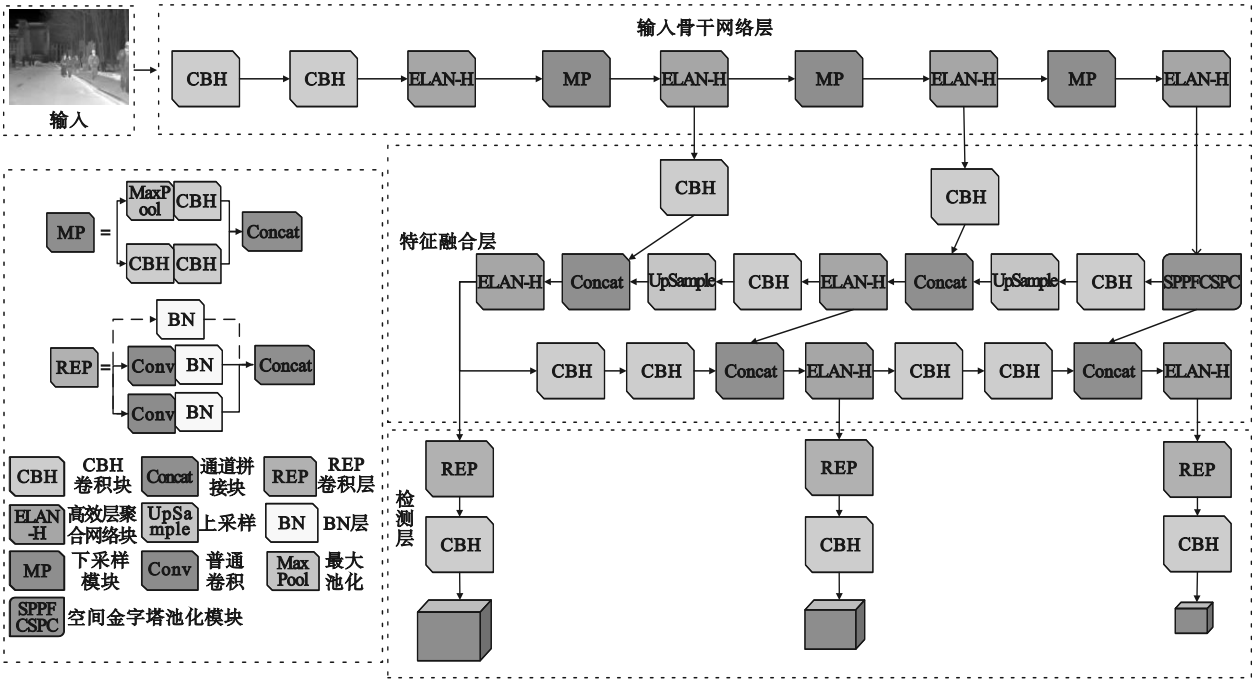

图 1　AS-YOLOv7 网络结构

Fig. 1　Network architecture of AS-YOLOv7

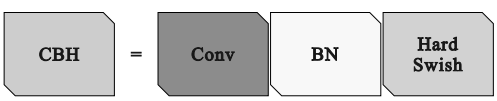

图 2　CBH 模块

Fig. 2　CBH module

YOLOv7 是以可见光图像为基础数据集设计的,由四层 CBS、ELAN 和 MP 模块交替组合而成,旨在提取更深层次的特征。然而,由于红外图像中存在的特征信息较少,直接使用 YOLOv7 进行红外图像检测可能导致特征损失。此外,较深的网络结构会带来更多的计算量和参数量,从而影响网络模型的运行效率。为了能够使模型更适合红外图像的物体检测以及减少网络参数,本文将 ELAN 模块中的 7 个 CBS 卷积缩减为 5 个 CBH 卷积,减少模型中大量的卷积操作,进而缩减参数量和计算量,使网络结构在保持适度深度的同时具有特征提取能力。改进后的 ELAN 模块如图 3 所示。

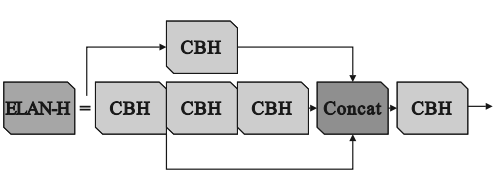

图 3　ELAN-H 模块

Fig. 3　ELAN-H module

MP 模块是一个超级下采样模块,在进行最大池化的同时进行步长为 2 的 3×3 卷积,在下采样过程中同时兼顾特征图的最大值和局部信息。在特征提取部分保留了 MP 模块,该模块中步长为 2 的卷积和最大池化具有相同的目的。因此在 Neck 部分使用一个步长为 2 的 CBH 卷积模块和一个 1×1 的 CBH 卷积模块替代 MP 模块,使其在减少计算量的同时保证模型具有较好的表现。

1.2　改进 SPPCSPC 模块

YOLOv7 中 SPPCSPC 模块通过最大池化获得不同的感受野,使算法适应不同分辨率的图像。SPPCSPC 模块使用三个独立的池化层和不同大小的池化核来获取不同大小的特征图,在此基础上融合跨阶段部分连接(CSP)的思想,并在两者融合之后又增加了一个卷积,增强特征融合的鲁棒性,从而使模型获得更好的检测性能。尽管加入 SPPCSPC 模块后模型性能得到了提升,但是将输入并行通过多个不同大小的池化会增加网络的计算负担。为减少该模块的计算量,本文对该模块进行改进,将三个并行的最大池化操作替换成串行的最大池化,并将所有池化核的尺寸修改为相同大小。通过简化池化过程,避免 SPPCSPC 操作重复,进而提高模型的计算速度。改进后的模块如图 4 所示。

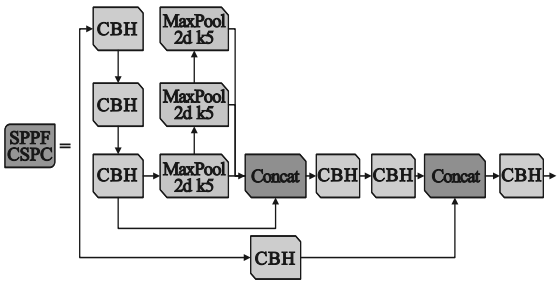

图 4 SPPFCSPC 模块

Fig. 4 SPPFCSPC module

改进后的 SPPCSPC 模块具有相同大小的池化核,在不改变模块对图像场景感知的情况下减少了计算量。首先将输入的特征经过三层卷积进行降维,减少模块的参数量,并另外进行一次卷积。将降维后的特征图进行三次最大池化,且下一层池化在上一层池化输出结果上计算。将第三次卷积与三次池化得到的特征图进行拼接,用于模拟不同尺度的感受野。最后将得到的结果经过两次卷积降维后与之前的一次卷积进行特征融合,并经过一次 1×1 卷积进行降维得到最终的输出结果。

1.3 替换损失函数

原模型使用 CIoU[9] 作为回归定位损失,通过计算预测框和真实框中心点距离构建惩罚项,并添加预测框和真实框的长宽比之间的差异使预测框更加符合真实框,有效解决了预测框和真实框不重合时,由于 IoU[10] 反向传播梯度消失导致重叠区域的宽度在训练期间无法更新的问题。但由于红外图像的特性,数据集中会存在一些低质量的数据,使用距离和纵横比等几何因素的 IoU 作为边框回归(BBR)的损失会加重对低质量数据的惩罚,降低模型的泛化性能。因此选用基于 Focal EIoU 的思想构建的一种具有动态非单调聚焦机制(FM)的 IoU 损失 WIoU,在锚框与目标框重合时削弱几何因素的惩罚项,通过较少的训练干预使模型获得更好的泛化能力。WIoU[11] 损失 L_{WIoUv1} 定义如下。

$$L_{\mathrm{WIoUv1}} = R_{\mathrm{WIoU}} L_{\mathrm{IoU}} \tag{2}$$

$$R_{\mathrm{WIoU}} = \exp\left(\frac{(x - x_{\mathrm{gt}})^2 + (y - y_{\mathrm{gt}})^2}{(W_{\mathrm{g}}^2 + H_{\mathrm{g}}^2)^*}\right) \tag{3}$$

式中:L_{IoU} 为真实框与目标框的交并比损失;R_{WIoU} 为 WIoU 的惩罚项,当 $R_{\mathrm{WIoU}} \in [1, \mathrm{e})$ 时,将对普通质量锚框的 L_{IoU} 进行显著性放大,当 $L_{\mathrm{IoU}} \in [0, 1)$ 时,则对高质量锚框的 R_{WIoU} 进行缩小,并在锚框与目标框重合时,重点关注中心点之间的距离;x、y 为边界框的中心坐标;x_{gt}、y_{gt} 为目标框的中心坐标;W_{g}、H_{g} 为最小封闭框的尺寸;* 表示将 W_{g}、H_{g} 从计算图中分离出来的操作,其目的是为了有效地消除 R_{WIoU} 产生阻碍收敛的因素。

为了使得模型能够聚焦于图像中的低对比度、有遮挡、模糊等困难样例,提高分类性能,基于 Focal 损失函数的交叉熵单调聚焦机制的思想,在 L_{WIoUv1} 的基础上增加单调聚焦系数 $L_{\mathrm{IoU}}^{\gamma *}$,被定义为 L_{WIoUv2}。反向传播的速度可能会随着聚焦系数的增加而发生变化,具体损失计算公式如下。

$$L_{\mathrm{WIoUv2}} = L_{\mathrm{IoU}}^{\gamma *} L_{\mathrm{WIoUv1}}, \gamma > 0 \tag{4}$$

式中 γ 为可调节的聚焦参数,γ 值越大,模型在低质量数据上的聚焦度越高。

在模型训练过程中,$L_{\mathrm{IoU}}^{\gamma *}$ 随着 L_{IoU} 的减小而减小,容易导致训练后期的收敛速度变慢。因此引入 L_{IoU} 的均值作为归一化因子解决此问题,计算如下。

$$L_{\mathrm{WIoUv2}} = \left(\frac{L_{\mathrm{IoU}}^*}{\overline{L_{\mathrm{IoU}}}}\right)^{\gamma} L_{\mathrm{WIoUv1}}, \gamma > 0 \tag{5}$$

式中 $\overline{L_{\mathrm{IoU}}}$ 是动量的滑动平均值,动态更新 $\overline{L_{\mathrm{IoU}}}$ 使 $\left(\frac{L_{\mathrm{IoU}}^*}{\overline{L_{\mathrm{IoU}}}}\right)^{\gamma}$ 整体保持在较高水平,解决了训练后期收敛速度慢的问题。

2 实验与分析

2.1 实验环境

本文实验环境配置如表 1 所示。

表 1 环境配置

Table 1 Environment configuration

名称	环境配置
操作系统	Ubuntu16.04
CPU	Intel(R)Xeon(R)CPU E5-2620
GPU	GTX 1080Ti(12 G/4 块)
编程语言	Python
深度学习框架	Pytorch 1.11.0
集成开发环境	Anaconda,CUDA 11.3

2.2 数据集

为使实验数据更具有真实性,使用 FLIR T640 红外摄像机在街道上拍摄车辆(car)、行人(person)和自行车(bicycle)的视频,将视频通过程序处理成图片,挑选出一部分模糊程度不高的

图片作为数据集。该数据集中共包含 10 116 张分辨率大小为 640×512 的图片,使用 LabelImage 标注工具对该数据集中的目标进行 YOLO 格式的标注,并将数据集通过程序采用 7∶2∶1 的比例划分为训练集、验证集和测试集。红外数据集样本如图 5 所示。

图 5　红外数据集样本

Fig. 5　Infrared dataset samples

2.3　结果分析

本次实验采用精度(AP)、平均精度(mAP)、参数量和推理时间做为评估指标对网络模型的性能进行评估。使用 SGD 优化器对网络模型参数进行优化。首先进行 5 轮的预热学习,学习率为 0.1。然后设置学习率为 0.01,批量大小设置为 32,使用 120 轮对模型进行训练。

2.3.1　AS-YOLOv7 与 YOLOv7 实验结果对比

本文提出的 AS-YOLOv7 模型与 YOLOv7 模型进行性能对比,如表 2 所示。可以看出:AS-YOLOv7 在参数量和模型推理时间上大大减少,其中模型的推理时间比 YOLOv7 降低了 58.2%,模型的参数量比 YOLOv7 减少了 79%;AS-YOLOv7 的平均精度相较于原 YOLOv7 上升了 0.2%。改进后的模型在检测精度和模型推理速度方面取得了较好的效果。

表 2　YOLOv7 和 AS-YOLOv7 性能对比

Table 2　Properties comparisons between YOLOv7 and AS-YOLOv7

模型	参数量/MB	推理时间/ms	平均精度/%
YOLOv7	37.20	55	81.0
AS-YOLOv7	7.95	23	81.2

对比 YOLOv7 和 AS-YOLOv7 在不同场景下的检测效果,如图 6 所示。由三组图可以看出,原 YOLOv7 在红外图像检测时,对小目标、密集严重被遮挡的目标及相似度较高的目标存在误检漏检,而 AS-YOLOv7 有明显改善,且 AS-YOLOv7 的推理时间均小于 YOLOv7。由此可见,本文提出的 AS-YOLOv7 对红外目标的检测精度比 YOLOv7 的检测精度更高,模型的推理速度更快。

(a) YOLOv7的检测效果　　(b) AS-YOLOv7的检测效果

图 6　YOLOv7 与 AS-YOLOv7 检测效果图对比

Fig. 6　Comparison of detection result images between YOLOv7 and AS-YOLOv7

2.3.2　不同模型实验对比

为验证改进模型的性能,分别对 YOLO 系列中的轻量级网络模型进行训练和测试,并以 mAP、AP、网络参数量、推理时间作为评价指标。

不同网络模型的检测结果如表 3 所示。本文提出的 AS-YOLOv7 与 YOLOv3-tiny 和 YOLOv5 相比,虽然在模型的推理时间上有所增加,但针对不同类别的检测精度均优于其他两个模型,且平均精度分别提升了 11.9%、2.8%;与 YOLOv8s 相比较,参数量规模减少了 28.8%,在不同类别上检测精度分别提升了 0.2%、0.2%、6.1%,平均精度提升了 2.1%,模型的推理时间减少了 5 ms,改进后的网络模型在各个方面都优于最新的 YOLOv8s 模型;AS-YOLOv7 与原网络相比,平均精度提高了 0.2%,同时参数量下降了 79%,模型的推理时间得到了大幅度下降。

综上所述,本文所提出的 AS-YOLOv7 模型在红外目标检测中表现出更好的性能。

表 3 不同网络模型的检测结果

Table 3 Detection results of different network models

模型	参数量/MB	AP/%			$mAP_{0.5}$/%	推理时间/ms
		汽车	行人	自行车		
YOLOv3-tiny	8.30	87.1	72.1	48.7	69.3	15
YOLOv5s	7.30	91.3	82.6	61.2	78.4	21
YOLOv8s	11.10	92.3	84.4	60.5	79.1	28
YOLOv7	37.20	91.5	87.5	64.1	81.0	55
AS-YOLOv7	7.95	92.5	84.6	66.6	81.2	23

2.3.3 消融实验

为明确不同模块对模型检测性能的影响,本次实验将以 mAP、AP、模型参数量和推理时间作为评估指标。设置 4 组实验,实验结果如表 4 所示。

通过表 4 可以看出,第一组实验表明 YOLOv7 的精度高但推理速度慢,限制了其在实时应用中的效用。在第二组(A)实验中修改 YOLOv7 网络结构,将网络模型进行轻量化改进,降低了网络模型的深度,使参数量减少了 29.25 MB,以损失了小部分平均精度的代价换取了模型推理速度上的提升。第三组(B)将改进的 SPPCSPC 模块加入到轻量化网络后,在保证参数量不变的同时使模型检测精度比轻量化后的模型上升了 0.6%,模型的推理时间降低了 8%。第四组(C)在 B 组的基础上替换了原网络中的损失函数,将原网络中的 CIoU 替换成 WIoU,有效解决了红外数据集中质量较好和质量较差样本间的不平衡问题,提高了模型的泛化性。加入 WIoU 后,在模型推理速度不变的情况下,各个类别的检测精度都得到了提升。改进后的模型在模型轻量化的同时又保证了模型的精度,更加适合红外场景目标检测。

表 4 消融实验结果

Table 4 Results of ablation experiments

分组	轻量化	SPPFCSPC	WIoU	AP/%			mAP /%	参数量 /MB	推理时间 /ms
				汽车	行人	自行车			
YOLOv7	×	×	×	91.5	87.5	64.1	81.0	37.20	55
A	√	×	×	91.3	82.5	59.3	77.7	7.95	25
B	√	√	×	91.3	83.1	60.4	78.3	7.95	23
C	√	√	√	92.5	84.6	66.6	81.2	7.95	23

3 结论

本文提出了基于 AS-YOLOv7 的红外目标检测方法。该方法通过构建 CBH 卷积块,增加模型的稳定性,同时针对红外图像中存在信息少等特点,采用缩减卷积的方式降低模型的复杂度,避免由于过深的网络结构造成红外图像特征的损失。然后通过简化 SPPCSPC 模块的池化过程,减少模型的计算量,提高模型的运行效率。最后采用 WIoU 损失函数,通过使用较少的训练干预使模型获得更好的泛化能力。AS-YOLOv7 模型的参数量是原模型的 21.5%,并且平均检测精度提升了 0.2%,模型的推理时间降低了 32 ms。综合实验结果表明,本文所提出的改进模型在红外图像检测上具备良好的准确性和运行效率。由于受限于当前条件,实验数据仅基于本地设备测试结果,为进一步验证改进后模型在实际应用中的可行性,下一步将在边缘设备上进行实验,并根据实验结果对网络模型进行进一步优化。

参考文献(References):

[1] 秦鹏,唐川明,刘云峰,等. 基于改进 YOLOv3 的红外目标检测方法[J]. 计算机工程,2022,48(3):211-219.
QIN P,TANG C M,LIU Y F,et al. Infrared target detection method based on improved YOLOv3[J]. Computer Engineering,2022,48(3):211-219. (in Chinese)

[2] 史泽林,冯斌,冯萍. 基于波前编码的无热化红外成像技术综述(特邀)[J]. 红外与激光工程,2022,51(1):32-42.
SHI Z L,FENG B,FENG P. Review on athermalized infrared imaging technology based on wavefront coding(Invited)[J]. Infrared and Laser Engineering,2022,51(1):32-42. (in Chinese)

(下转第 12 页)