文章编号:1001-5078(2024)02-0281-08

·图像与信号处理·

基于 YOLOv5s 的改进实时红外小目标检测

谷　雨[1],张宏宇[2],彭冬亮[1]
(1. 杭州电子科技大学自动化学院,浙江 杭州 310018;2. 杭州电子科技大学圣光机联合学院,浙江 杭州 310018)

摘　要:针对红外图像分辨率低、背景复杂、目标细节特征缺失等问题,提出了一种基于 YOLOv5s 的改进实时红外小目标检测模型 Infrared-YOLOv5s。在特征提取阶段,采用 SPD-Conv 进行下采样,将特征图切分为特征子图并按通道拼接,避免了多尺度特征提取过程中下采样导致的特征丢失情况,设计了一种基于空洞卷积的改进空间金字塔池化模块,通过对具有不同感受野的特征进行融合来提高特征提取能力;在特征融合阶段,引入由深到浅的注意力模块,将深层特征语义特征嵌入到浅层空间特征中,增强浅层特征的表达能力;在预测阶段,裁减了网络中针对大目标检测的特征提取层、融合层及预测层,降低模型大小的同时提高了实时性。首先通过消融实验验证了提出各模块的有效性,实验结果表明,改进模型在 SIRST 数据集上平均精度均值达到了 95.4 %,较原始 YOLOv5s 提高了 2.3 %,且模型大小降低了 72.9 %,仅为 4.5 M,在 Nvidia Xavier 上推理速度达到 28 f/s,利于实际的部署和应用。在 Infrared-PV 数据集上的迁移实验进一步验证了改进算法的有效性。提出的改进模型在提高红外图像小目标检测性能的同时,能够满足实时性要求,因而适用于红外图像小目标实时检测任务。

关键词:红外小目标检测;YOLOv5s;注意力机制;特征融合
中图分类号:TP753　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2024.02.018

Improved real-time infrared small target detection based on YOLOv5s

GU Yu[1], ZHANG Hong-yu[2], PENG Dong-liang[1]
(1. School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China;
2. HDU-ITMO Joint Institute, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: In this paper, an improved infrared small target detection model, infrared-YOLOv5s, based on YOLOv5s is proposed to address the problems of low resolution, complex background and lack of detailed features of infrared images. In feature extraction stage, SPD-Conv is used for down-sampling, which divides the feature map into feature submaps and concatenate them by channel to avoid the loss of features caused by down-sampling in the process of multi-scale feature extraction. And an improved atrous spatial pyramid pooling module is designed to improve feature extraction capabilities by fusing features with different receptive fields. Then, in feature fusion stage, a deep-to-shallow attention module is introduced to embed deep semantic features into shallow spatial features to enhance the expression of shallow features. Moreover, in prediction stage, the prediction layers, feature extraction layers and feature fusion layers for large target detection in the network are cut down to reduce the model size and improve real-time performance at

基金项目:浙江省自然科学基金项目(No. LZ23F030002)资助。
作者简介:谷　雨(1982 -),男,博士,副教授,主要从事遥感图像目标检测、识别和跟踪等方面的研究。E-mail:guyu@hdu.edu.cn
收稿日期:2023-03-30

the same time. The effectiveness of each module is verified by ablation experiments, and experimental results show that the proposed model achieves 95.4 % $mAP_{0.5}$ of on SIRST dataset, which is 2.3 % higher than that of original YOLOv5s. The model size is reduced by 72.9 % to 4.5 MB, and the inference speed on Nvidia Xavier reaches 28 f/s, which is conducive to the actual deployment and application. Therefore, the effectiveness of the proposed model is further verified by transfer experiments using Infrared-PV dataset, and the proposed model can meet the real-time requirements while improving the performance of small target detection in infrared images, and is suitable for the task of real-time small target detection in infrared images.

Keywords: infrared small target detection; YOLOv5s; attention mechanism; feature fusion

1 引 言

红外成像系统具有全天候、抗干扰能力强、探测距离远等优势,因此基于红外成像的目标检测技术在军事侦查、红外制导、自动驾驶等领域得到了广泛应用[1]。与可见光图像不同,红外图像分辨率低、背景复杂,目标多呈现为弱小目标状态,严重影响了检测精度,因此如何提高红外小目标检测性能成为亟待解决的问题。

传统的红外小目标检测方法主要有三种[2],基于滤波器的红外小目标检测算法思路简单、计算量小,但其对于复杂背景的抑制较差,检测精度低;基于人眼视觉系统的检测方法易于实现,但其检测精度依赖于分割阈值,有一定局限性;基于矩阵分解的方法对于复杂背景有较高的可靠性,但由于计算复杂,检测实时性较差。

随着深度学习理论的发展,基于深度学习的目标检测取得了远超传统方法的性能。基于深度学习的通用目标检测算法可以分为基于候选框的两阶段检测算法和基于回归的单阶段检测算法[3]。直接将上述通用目标检测算法用于红外小目标检测时,由于红外图像分辨率低、目标尺寸小、缺乏细节纹理特征的特性,增加了红外目标的检测难度,检测性能会降低,因此研究学者针对深度学习红外图像弱小目标检测进行了一系列优化。Wu 等人[4]基于 YOLOv3[5]算法,将网络预测层扩展到 4 个特征尺度,通过引入 GIoU[6]改进了损失函数,提高了检测性能,在 FLIR 红外数据集上平均准确率提高了 3.4 %。Zheng 等人[7]针对红外小型无人机目标检测,设计了一个特征增强模块以增强“目标特征”,同时将自适应特征融合方法引入特征融合中,以解决跨尺度特征融合中特征表达弱化的问题。Zhao 等人[8]结合 DenseNet[9]和 YOLOv5s[10],将 YOLOv5s 的部分 C3 模块替换为 DenseBlock 模块,并且在主干网络中加入 SENet[11]模块,提高了特征提取能力同时降低参数量,并且使用简化的 BiFPN 取代了原始 PANet[12]结构,增强了网络提取不同尺度特征的能力,在夜间道路场景下对行人和车辆检测的平均准确率提高了 3.49 %。MFSSD[13]重新设计了特征融合网络,加强了不同层次网络之间的信息交互,实现了深层特征和浅层特征的有效融合。现有的卷积神经网络受感受野限制,无法获取大范围内目标和背景的差异性,后续的研究学者开始尝试将 Transformer[14]用于目标检测,TPH-YOLOv5[15]通过探索自注意力机制使用 Transformer 预测头,提升了密集场景和遮挡情况下小目标的检测性能。Xin 等人[16]使用 SwinTransformer 替换 YOLOv5s 中的部分 C3 模块,在 FLIR 数据集上平均准确率较初始 YOLOv5s 提高了 5.6 %。Liu 等人[17]为了获取红外图像中的全局依赖,提出了一种基于 Transformer 的红外弱小目标检测方法,利用 Transformer 的自注意力机制,在全局范围内学习目标特征。同时为了避免目标丢失,使用了类似 U-Net[18]的网络结构来融合不同尺度的特征,在两个公共数据集上取得了更好的检测结果。

结合红外图像的特性和 YOLO 系列算法的优势,本文提出了一种基于改进 YOLOv5s 的实时红外小目标检测模型,主要的创新点如下:

(1)在特征提取阶段,使用 SPD-Conv[19]进行下采样,避免小目标特征丢失,同时通过串联多个不同空洞率的空洞卷积增强多尺度特征提取能力。在主干网络中加入了 CBAM[20]空间和通道混合注意力模块,以提高模型的表征能力,提升小目标的检测性能。

(2)在特征融合阶段,引入由深到浅的注意力模块,将深层语义特征嵌入到浅层空间特征中,提高浅层特征的表达能力。

(3)在预测阶段,裁剪网络中针对大目标检测的预测层及相关特征提取和特征融合层,降低了模

型大小,提高了检测实时性。

(4)最后采用Infrared-PV和SIRST[21]数据集验证了提出算法的有效性。

2 基于改进YOLOv5s的红外小目标检测

2.1 YOLOv5网络结构

根据网络深度和宽度不同,YOLOv5模型由小到大可分为YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x。由于红外图像分辨率较低,样本数量少,复杂的网络会导致过拟合,因此本文选择YOLOv5s作为红外小目标检测基准模型。YOLOv5主要分为输入、特征提取、特征融合和预测输出四个部分。输入模块使用Mosaic进行数据增强以增加样本数量。特征提取模块主要由CBS、C3和SPPF模块组成,CBS采用步长为2的卷积对特征图进行下采样。C3模块借鉴了CSPNet[22](Cross Stage Partial Network)的设计,将CSP-BottleNeck中的卷积减少到3个,在不降低检测精度的前提下减少了模型参数,提高了实时性。SPPF模块在空间金字塔池化[23](Spatial Pyramid Pooling,SPP)的基础上使用多个小尺寸池化核级联代替SPP模块中单个大尺寸池化核,进一步提高了检测速度。在特征融合阶段,YOLOv5s采用特征金字塔网络[24](Feature Pyramid Network,FPN)和PANet(Path Aggregation Network)的多尺度特征融合策略,增强多尺度特征的融合能力。预测模块主要用于检测目标,当输入图像分辨率为640×640时,分别输出20×20、40×40和80×80大小的特征图,对应大、中、小目标检测层。

2.2 基于改进YOLOv5s的红外小目标检测模型

尽管YOLOv5s性能优异,但其在红外场景下的检测精度仍有待提高,故本文从特征提取、特征融合、预测输出三个方面改进YOLOv5s,提出了一个实时红外小目标检测模型Infrared-YOLOv5s,以提高红外小目标检测精度,其结构如图1所示,图中改进模块用不同颜色标识。

2.2.1 基于SPD-Conv和IASPP的改进特征提取网络

现有卷积神经网络通常使用步长为2的卷积或最大池化进行下采样,由于红外图像分辨率低、目标细节特征缺失,这种下采样方式会导致细节信息丢失。因此,本文引入了SPD-Conv来替换YOLOv5s中的下采样模块。

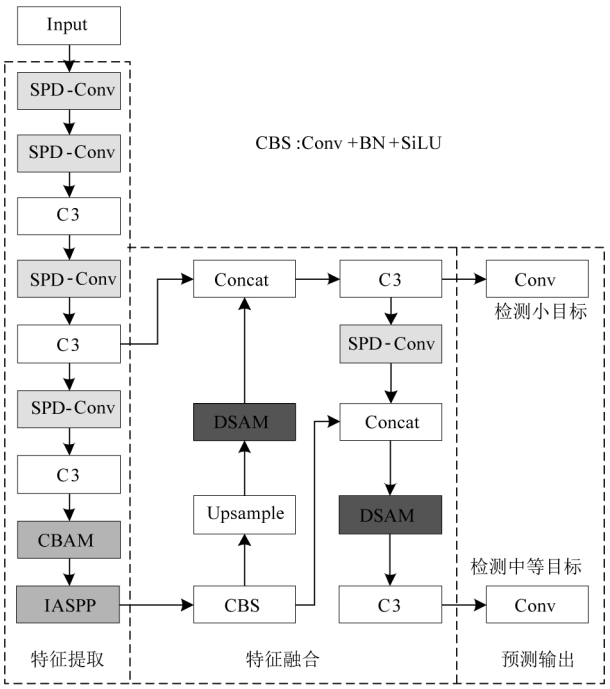


图1 Infrared-YOLOv5s网络结构

Fig. 1 The structure of Infrared-YOLOv5s network

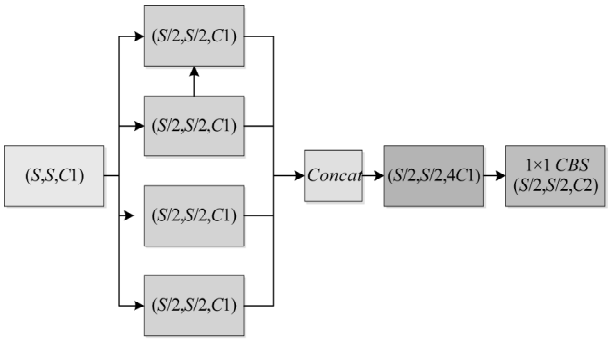


图2 SPD-Conv结构图

Fig. 2 The structure of SPD-Conv

SPD-Conv由空间到深度转换层(Space-to-depth,SPD)和一个1×1卷积的CBS模块组成。SPD层对特征图进行下采样时保留了通道维度中的所有信息,因此没有信息丢失。在SPD层后添加1×1卷积降低通道数。如图2所示,给定$S\times S\times C1$的特征图,将其切片为四个$S/2\times S/2\times C1$的特征子图,将这些子图按通道拼接,得到$S/2\times S/2\times 4C1$的特征图,最后使用1×1卷积调整通道数。使用SPD-Conv进行下采样可以最大程度保留小目标的细节特征,有利于后续的特征提取操作。

针对YOLOv5采样过程中小目标容易丢失的问题,如图1所示,本文在特征提取阶段加入CBAM[20]注意力模块,使网络更专注于对小目标的检测。在目标检测任务中,较大的感受野可以获得

更为全局、语义层次更高的特征，但下采样操作增大感受野的同时会带来分辨率的降低，导致小目标丢失。为了解决这个矛盾，引入空洞卷积[25]（Atrous-Convolution），在减少分辨率损失的前提下，增大感受野。空洞空间金字塔池化[26]（Atrous SpatialPyramidPooling，ASPP）将不同感受野特征图融合，使得像素点分类更准确。然而，随着采样率的增加，空洞卷积的效果会变差。为了在融合多尺度特征信息的同时获得更大的感受野，本文重新设计了 ASPP 模块，提出了改进空洞空间金字塔池化（Improved AtrousSpatialPyramidPooling，IASPP）模块。如图 3 所示，ISAPP 包含三个分支，输入特征图经过 1×1 卷积得到输出 out；经过自适应全局平均池化得到输出 pool；在空洞卷积分支中，经过 3×3 的普通卷积得到输出 out1，然后将 out1 输入采样率为 2 的空洞卷积得到 out2，并将其与 out1 拼接得到 add1，输入到采样率为 3 的空洞卷积得到输出 out3，将 out3 与 add1 拼接得到 add2，串联的空洞卷积结构可以在不同采样率的特征图间共享特征，从而改善 ASPP 因采样率变大导致效果变差的问题，增大感受野的同时又能获取多尺度信息。ISAPP 的最终输出为 Cat（pool，out，add2）。

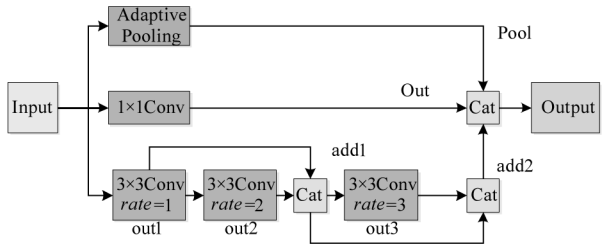


图 3 IASPP 模块结构图

Fig. 3 The structure of IASPP

2.2.2 基于由深到浅注意力的多尺度特征融合

浅层特征感受野小，分辨率高，包含更多细节信息，对于目标定位较为重要；深层特征可以提供更好的语义信息和对场景上下文的理解，有助于解决目标和背景干扰物之间的歧义，但随着分辨率的降低细节信息丢失严重。因此实现浅层特征和深层特征的有效融合，可以提高检测性能。如图 4（a）所示，YOLOv5 通过 Concat 操作将浅层特征和深层特征直接按通道拼接，不能反映不同尺度特征的重要性。在多尺度特征融合阶段，引入由深到浅的注意力模块（Deep-to-Shallow Attention Module，DSAM）如图 4（b）所示，将深层语义特征嵌入到浅层空间特征，可以帮助处理歧义，提高分类准确率。

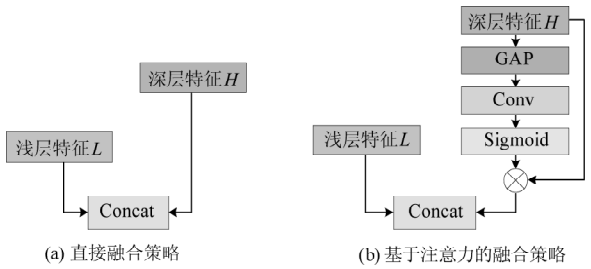


图 4 深层特征与浅层特征融合策略

Fig. 4 Fusion strategy of deep and shallow features

对深层特征 H 首先通过全局平均池化（Global Average Pooling，GAP）编码得到全局特征 U，然后利用 1×1 卷积学习各通道之间的关系，再经过 Sigmoid 激活函数将输出限制在 0 ~ 1 之间，得到权重 w：

$$w = \sigma(\mathrm{SiLU}(\mathrm{Conv}(U))) \tag{1}$$

式中，σ 表示 Sigmoid 激活函数，Conv 为 1×1 卷积，SiLU 为卷积层激活函数；将权重 w 与原始深层特征 H 相乘即可得到加权后的特征 H'。最后，将加权后的深层特征 H' 和浅层特征 L 拼接，得到融合特征图。该模块可以将深层特征更有效地传递给浅层特征，提高了特征融合效果。

2.2.3 适用于红外小目标检测的预测层结构

本文检测对象为红外图像中的弱小目标，在 YOLOv5 中，大目标检测层的特征图是对输入图像进行 32 倍下采样得到的，当目标尺寸小于 32×32 像素时，会出现目标采样不到的现象。因此，对于检测红外小目标，YOLOv5 中的大目标检测层属于冗余层，会增加模型大小但对于小目标检测没有帮助。基于上述结论，如图 1 所示，本文裁减了 YOLOv5 网络中的大目标检测层及其相应特征提取和特征融合层，只进行 4 次下采样，仅保留 8 倍和 16 倍下采样的特征图进行红外小目标检测，改进后的网络结构去除了大量冗余计算，在保证检测精度的前提下，降低了模型大小，防止出现过拟合，提高了检测实时性。

3 实验及结果分析

3.1 红外小目标检测数据集

本文使用 SIRST 红外数据集进行实验，该数据集是南京航空航天大学发布的一个不同场景下的单帧红外小目标数据集[21]。共有 427 张红外图像，包

含500多个目标。图5为SIRST数据集中的部分红外图像及标注信息。目标所处的环境复杂多变,且目标尺寸多样且亮度差异较大。数据集标注信息使用SIRST数据集的分割真值图像利用最小包围盒算法得到,标注为VOC格式,保存为XML文件。其中训练集256张图片,验证集85张图片,测试集86张图片。

图5 SIRST数据集示例图像及标注信息

Fig. 5 SIRST dataset sample images and annotation information

3.2 训练环境和配置

本文模型实现采用Pytorch1.7.1,实验所用的计算机配置如下:CPU为i7-8700k,主频为3.70 GHz,GPU为1080Ti,内存为32 G,操作系统为Windows10。实验代码基于YOLOv5-6.1版本改进,训练次数(epoch)为100次,批大小为16,初始学习率为0.01,采用SGD梯度下降优化器,采用one-cycle学习率衰减,输入的红外图像分辨率为640×640,其他为默认参数设置。

3.3 评价指标

为准确评估模型在红外图像上的检测性能,本文采用平均精度值(mean Average precision,mAP)和F_1值(F1-Score)作为评价指标。数据集中每个类别根据准确率(Precision,P)和召回率(Recall,R)可绘制一条PR曲线,曲线与坐标轴围成的面积即为AP值。其中准确率和召回率计算如式(2),其TP为真正例,FP为假正例,FN为假反例:

$$R = \frac{TP}{TP + FP}; P = \frac{TP}{TP + FN} \tag{2}$$

当检测框与真值框的交并比(Intersection over Union,IoU)大于0.5时认为目标被准确预测,在IoU取0.5时计算每个类别的平均精度和总平均精度,记为$mAP_{0.5}$。

F_1值是分类问题的一个评价指标,同时兼顾了分类模型的精确率和召回率,可以看作是模型精确率和召回率的一种调和平均值,计算方法如式(3):

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

3.4 消融实验

为验证各模块的有效性,本文以YOLOv5s为基准,设计了如下消融实验:A为采用2.2.1节的基于SPD-Conv和IASPP的改进特征提取网络,B为采用2.2.2节的基于由深到浅注意力的多尺度特征融合,C为采用2.2.3节的适用于红外小目标检测的预测层结构。实验结果如表1所示,其中实验1为YOLOv5s基准模型实验结果。

表1 不同模块消融实验结果

Tab. 1 Results of ablation experiments with different modules

编号	A	B	C	$mAP_{0.5}$	模型大小/MB
1	×	×	×	93.1	16.6
2	√	×	×	94.5	16.8
3	×	√	×	93.5	16.6
4	×	×	√	92.7	**4.1**
5	√	√		95.1	16.8
6	√	×	√	94.8	4.3
7		√	√	93.7	4.1
8	√	√	√	**95.4**	4.5

(1)实验2和6证明,在不同的预测层结构下,采用基于SPD-Conv和IASPP的改进特征提取网络,$mAP_{0.5}$分别提高了1.4 %和1.7 %。SPD-Conv下采样模块最大程度地保留了小目标特征,改进空洞空间金字塔池化模块可以融合多尺度特征,增大了感受野,以适应不同目标尺寸大小的变化,CBAM注意力模块同时利用了空间注意力和通道注意力,使网络更关注目标,进一步提高了检测精度。

(2)实验3和7证明,在特征融合阶段引入由深到浅的注意力模块,在不增加额外的内存占用的情况下,$mAP_{0.5}$分别提高了0.4 %和0.6 %。DSAM注意力模块将深层特征丰富的语义信息嵌入到浅层特征,提高了浅层特征的表达能力,相比原始YOLOv5直接通道相加的方法,可以获得更丰富的语义和空间特征,因此可以提高检测精度。

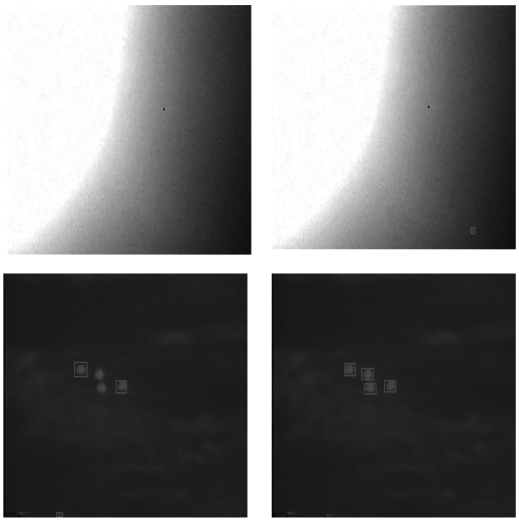
(3)实验4证明,与基准YOLOv5s比较,使用适用于红外小目标检测的预测层结构,$mAP_{0.5}$仅降低0.4 %,模型大小仅为原始的四分之一左右,提升了检测实时性。

综上所述,使用所有改进策略的实验8,提出

的 Infrared-YOLOv5s 较基准 YOLOv5s，$mAP_{0.5}$ 提高了 2.3 %，且模型大小仅为原始的 27.1 %，验证了改进算法在红外小目标图像数据集上的有效性。

3.5 SIRST 数据集算法验证

为验证本文算法的有效性，本文以 YOLOv5s 模型为基准，并与文献[17]和[21]提出的算法进行了对比。实验结果如表 2 所示，在 SIRST 数据集上，较基准模型 YOLOv5s，改进模型 $mAP_{0.5}$ 提高了2.3 %，F_1 分数提高了 3.18，验证了改进算法在红外数据集上的有效性。虽然 F_1 分数比文献[17]提出的模型低，但由于文献[17]使用了 Transformer 结构，使模型参数增加，训练和检测速度较慢，本文算法检测实时性更好，检测时间仅为文献[17]的十分之一，实现了检测性能和检测速度的平衡。由图 6 的检测结果图像可知，Infrared-YOLOv5s 模型在低对比度和复杂多目标场景下的红外小目标检出率优于 YOLOv5s，虚警率更低。



(a)YOLOv5s (b)Infrared-YOLOv5s

图 6 不同算法在 SIRST 数据集的检测结果图像

Fig. 6 Detection result images of different algorithms on SIRST dataset

表 2 不同算法在 SIRST 数据集上的检测性能

Tab. 2 Detection performance of different algorithms on SIRST dataset

算法	$mAP_{0.5}$	F_1-Score	Times(s/100 images)
文献[17]	/	**98.62**	6.41
ACM[21]	/	96.78	1.61
YOLOv5s	93.1	93.43	0.67
Infrared-YOLOv5s	**95.4**	96.61	**0.58**

3.6 基于 Infrared-PV 数据集的迁移实验

为充分验证本文算法的有效性和鲁棒性，在自建的 Infrared-PV 数据集上进行了迁移实验。Infrared-PV 数据集包括行人(Person)和车辆(Vehicle)两类目标，总计包 2138 张图片，其中白热图 1000 张，黑热图 838 张，热力图 300 张，采用 VOC 格式进行标注，保存为 XML 文件。平均单张图片包含 20 个左右目标，单张图片最多目标数超过 100 个，目标比较密集，且目标占整幅图像的 10 % 以下，以中小目标为主，适合迁移验证本文算法的有效性。图 7 为 Infrared-PV 数据集的示例图像。

(a)黑热目标场景 (b) 白热目标场景 (c)热力图场景

图 7 Infrared-PV 数据集示例图像

Fig. 7 Sample images of Infrared-PV dataset

实验结果如表 3 所示，YOLOv7 采用高效的 ELAN 主干网络并结合多种训练优化策略，检测精度比 YOLOv5s 提高了 2.3 %。本文提出的模型针对红外小目标检测任务改进特征提取网络并结合基于注意力的特征融合，较 YOLOv5s 基准模型，检测精度提高了 2.8 %，达到 84.5 %，优于 YOLOv7 算法和两阶段的 CascadeRCNN 算法。由于采用了适用于红外小目标检测的预测层结构，在 PC 机上推理速度可达 172.5 f/s，实时性更好。由图 8 中检测结果图像可知，改进模型在密集和遮挡场景下表现优于 YOLOv5s 模型。实验表明，本文算法对于尺度差异较大、重叠目标和密集目标实现了较好的鲁棒性。

表 3 不同算法在 Infrared-PV 数据集上的检测性能

Tab. 3 Detection performance of different algorithms on Infrared-PVdataset

算法	$AP_{0.5}$ (vehicle)	$AP_{0.5}$ (Person)	$mAP_{0.5}$ (All)	FPS
Scaled-YOLOv4[27]	87.3	73.1	81.1	48.0
YOLOv5s	89.6	74.4	81.8	150.4
YOLOv5s-STF	90.7	73.5	82.1	134
CascadeR-CNN[28]	90.1	74.4	82.3	11.80
YOLOv7[29]	92.0	76.1	84.1	161.0
Infrared-YOLOv5s	**92.2**	**77.1**	**84.5**	**172.5**

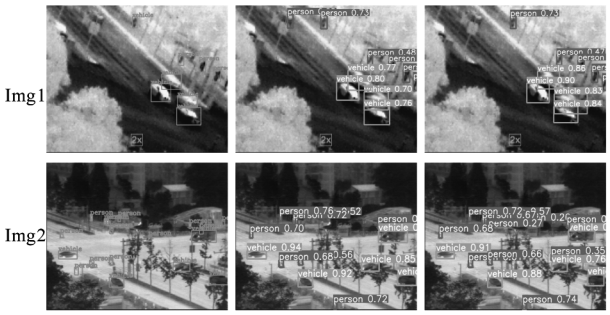


图8 不同算法在Infrared-PV数据集的检测结果图像

Fig. 8 Detection result images of different algorithms on Infrared-PV dataset

3.7 Nvidia Xavier设备部署实验

为验证本文算法在移动设备上的性能，在Nvidia Xavier设备上进行了部署实验。Xavier是一款高性能AI边缘设备，拥有一颗8核心ARM架构CPU，16GB、256位LPDDR4x内存，其GPU含有8个流式多核处理器，拥有512个CUDA核、64个张量核(Tensor-Core)、两个深度学习加速器(Deep Learning Accelerator，DLA)和其他硬件资源，最高算力可达32万亿次每秒(TeraOperationsPerSecond，TOPS)，功耗在10W到30W之间，拥有强大的计算能力且功耗较低，其软硬件资源如表4所示。

表4 NvidiaXavier软硬件资源

Tab. 4 Nvidia Xavier software and hardware resources

硬件	CPU	GPU	内存	操作系统	编译器	外部设备
配置	Aarch64	Volta架构512CUDA核	16 GB	Linux	g++、gcc	USB 3.0、HDMI等

图9为NvidiaXavier设备部署实物图，界面使用QT搭建，集成了模型训练、图像及视频检测、性能测试等功能。使用PC机训练得到的权重文件在Xavier设备上进行测试，推理速度可达28 f/s，达到边缘设备部署的实时性要求。

图9 NvidiaXavier设备部署实物图

Fig. 9 NvidiaXavier device deployment diagram

4 结 论

本文研究了YOLOv5网络结构及其各个模块的作用，通过分析红外小目标图像的特性，提出了一种基于YOLOv5s的改进实时红外小目标检测模型Infrared-YOLOv5s。首先在特征提取阶段采用SPD-Conv下采样避免了红外小目标特征丢失，设计了改进空洞空间金字塔池化模块，增强多尺度特征提取能力，以适应目标尺寸变化；其次在特征融合阶段引入由深到浅的注意力模块，将深层语义特征嵌入到浅层空间特征中，提高浅层特征的表达能力；预测阶段裁剪了针对大目标检测的特征提取、融合以及预测层，降低了模型大小，提升了检测实时性。最后基于SIRST数据集对各个改进模块设计了消融实验和不同算法之间的对比实验。实验结果表明，改进后的算法在SIRST数据集上$mAP_{0.5}$提高了2.3 %，保证检测精度的同时，在NvidiaXavier设备上推理速度达到28 f/s，能够满足实际部署需求。在Infrared-PV数据集上的迁移实验表明，改进模型较YOLOv5s基准模型，$mAP_{0.5}$提高2.8 %，进一步验证了改进算法的有效性和鲁棒性。

参考文献：

[1] Han Jinhui, Wei Yantao, Peng Zenming, et al. Infrared dim and small target detection: a review[J]. Infrared and Laser Engineering, 2022, 51(4): 438-461. (in Chinese)
韩金辉，魏艳涛，彭真明，等. 红外弱小目标检测方法综述[J]. 红外与激光工程，2022，51(4)：438-461.

[2] Zhao M J, Li W, Hu J, et al. Single-frame infrared small-target detection: asurvey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10: 87-119.

[3] Wang Henghui, Cao Dong, Zhao Yang, et al. Survey of infrared dim small target detection algorithm based on deep learning[J]. Laser & Infrared, 2022, 52(9): 1274-1279. (in Chinese)
王恒慧，曹东，赵杨，等. 基于深度学习的红外弱小目标检测算法研究综述[J]. 激光与红外，2022，52(9)：1274-1279.

[4] Wu Y F, Pan F, An Q C, et al. Infrared target detection based on deep iearning[C]//2021 40th Chinese Control Conference (CCC), Shanghai, China, 2021: 81175-8180.

[5] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J]. arXiv: 2018, 21804.02767.

[6] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 658 – 666.

[7] Zheng L, Peng Y P, Ye Z C, et al. Infrared small UAV target detection algorithm based on enhanced adaptive feature pyramid networks [J]. IEEE Access, 2022, 10: 115988 – 115995.

[8] Zhao H X, Liang Z R, Cai D H, et al. An improved method for infrared vehicle and pedestrian detection based on YOLOv5s[C]//2022 International Conference on Machine Learning, Cloud computing and Intelligent Mining (MLCCIM), Xiamen, China, 2022: 377 – 383.

[9] Huang G, Liu S C, Maaten L, et al. CondenseNet: an efficient DenseNet using learnd group convolutions [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 2752 – 2761.

[10] Jocher G. YOLOv5 [EB/OL] Https://github. com/ultralytics/yolov5, 2020.

[11] Hu J, Shen L, Albanie S, et al. Squeeze and excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42: 2011 – 2023.

[12] Wang K, Liew J H, Zou Y, et al. Panet: few-shot image semantic segmentation with prototype alignment[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9197 – 9206.

[13] Wang Fang, Li Qiang, Wu Bo, et al. Infrared small target detection method based on multi-scale feature fusion[J]. Infrared Technology, 2021, (43): 688 – 695. (in Chinese)
王芳, 李强, 伍博, 等. 基于多尺度特征融合的红外小目标检测方法[J]. 红外技术, 2021, (43): 688 – 695.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, 2017: 5998 – 6008.

[15] Zhu X K, Lyu S C, Wang X, et al. TPh-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021: 2778 – 2788.

[16] Xin X L, Pan F, Wang J C, et al. SwinT-YOLOv5s: improved YOLOv5s for vehicle-mounted infrared target detection [C]//2022 41st Chinese Control Conference (CCC), Hefei, China, 2022: 7236 – 7331.

[17] Liu F C, Gao C Q, C F, et al. Infrared small-dim target detection with transformer under complex backgrounds [J]. arXiv: 2021, 2109. 14379.

[18] Nneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234 – 241.

[19] Raja Sunkara, Luo T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2022: 443 – 459.

[20] Sanghyun Woo, Jongchan Park, Joon-Yong Lee, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[21] Dai Y M, Wu Y Q, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection[C]//IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 949 – 958.

[22] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390 – 391.

[23] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904 – 1916.

[24] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 – 2125.

[25] Yu F, Koltun V, Funkhouser T. Dilated residual networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 636 – 644.

[26] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834 – 848.

[27] Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-YOLOv4: scaling cross stage partial network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13029 – 13038.

[28] Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154 – 6162.

[29] Wang C Y, Alexey B, Liao M H. YOLOv7: trainable bag of freebies sets new state of art for real time object detectors[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 1 – 15.