

DOI:10.19651/j.cnki.emt.2212110

基于 Efficientnet 的红外目标检测算法^{*}

侯艳丽 王娟

(河北科技大学信息科学与工程学院 石家庄 050018)

摘要: 针对复杂场景下红外目标检测存在准确率低、召回率低的问题,为了提高红外图像中的小目标以及被遮挡目标的检测识别能力,提出基于 Efficientnet 的红外目标检测算法。首先,将高效轻量的 Efficientnet 作为模型的特征提取主干网,降低模型的参数量,提升训练速度。在 Efficientnet 主干网的最后一个输出层引入 SPP 模块,丰富特征图的表达能力,进行多尺度融合,扩大特征图的感受野;在模型特征融合部分,使用 FPN 特征金字塔网络,特征融合后增加 CSPNet 模块和 ECA 注意力机制,加强特征提取。检测部分使用 YOLO Head,对目标进行分类和回归,并用 CIoU Loss 作为边界框回归损失函数,提高对被遮挡目标的识别能力。实验结果表明,基于 Efficientnet 的模型大小仅为 YOLOv3 的 18.8%,并且在 FLIR 数据集上 mAP 达到 80.74%,相比于 YOLOv3 算法提高 10.12%,该模型在减少模型参数数量的同时,提升了检测精度。该模型在 FLIR 数据集上具有良好的泛化能力,提高了对小目标和遮挡目标的检测能力。

关键词: 红外目标检测;Efficientnet;ECA 注意力机制;SPP;CIoU Loss

中图分类号: TP319.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Infrared target detection algorithm based on Efficientnet

Hou Yanli Wang Juan

(School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract: In order to improve the detection and recognition of small targets and obscured targets in infrared images, the Efficientnet-based infrared target detection algorithm is proposed for the problem of low accuracy and low recall of infrared target detection in complex scenes. First, the efficient and lightweight Efficientnet is used as the feature extraction backbone of the model to reduce the number of parameters of the model and improve the training speed. In the last output layer of Efficientnet backbone, SPP module is introduced to enrich the expression capability of feature map, perform multi-scale fusion and expand the perceptual field of feature map; in the feature fusion part of the model, FPN feature pyramid network is used, and CSPNet module and ECA attention mechanism are added after feature fusion to enhance feature extraction. The detection part uses YOLO Head to classify and regress the targets, and uses CIoU Loss as the bounding box regression loss function to improve the recognition ability of the obscured targets. The experimental results show that the Efficientnet-based model is only 18.8% of the size of YOLOv3, and the mAP reaches 80.74% on the FLIR dataset, which is 10.12% better than the YOLOv3 algorithm, and the model improves the detection accuracy while reducing the number of model parameters. The model has good generalization ability on the FLIR dataset and improves the detection of small and occluded targets.

Keywords: infrared target detection;Efficientnet;ECA attentional mechanisms;SPP;CIoU Loss

0 引言

红外成像技术具有抗干扰能力强、测量精度高、穿透能力强等特点,并且不受雨雪恶劣天气影响,以及可以全天候工作^[1],因此在军事和安防领域有着独特的优势,在铁路、煤

矿、消防抢险等领域也有广泛的应用。但红外图像相较于可见光图像,存在对比度低、分辨率差、目标特征模糊、色彩信息不够丰富等缺点,尤其在复杂背景的情况下,红外目标之间存在遮挡并且小目标密集,使得红外场景下的目标检测效果不佳,因此针对红外图像目标检测的研究具有重要意义。

收稿日期:2022-11-17

^{*} 基金项目:河北省重点研发计划(21355901D)项目资助

基于深度学习的检测算法被提出以来,在一些领域取得了不错的成绩。基于深度学习的检测算法分为两大类:以 R-CNN^[2] 及其改进 Fast R-CNN^[3]、Faster R-CNN^[4]、Mask R-CNN^[5] 等为代表的双阶段算法,和以 SSD^[6]、YOLO 系列^[7-9]、RetinaNet^[10] 等为代表的一阶段算法。已有一些研究工作将基于深度学习的算法引入红外目标检测问题。刘冬等^[11] 提出一种 MCA-YOLO 的红外目标检测算法,以 YOLOv4 网络为基础,采用 Mobilenet 作为主干网,引入 CA 注意力模块等方法,减少模型参数量,保证准确率又具有实时性。李北明等^[12] 提出一种 Ghost-YOLOv5 红外目标检测算法,使用特征蒸馏方法以及 Mosaic 和 Copy-paste 两种数据增强的方法,在网络模型仅有 1.9 M 的情况下,提高了检测精度。黄磊等^[13] 提出 FS-YOLOv5 轻量化模型,使用 FS-MobileNetV3 作为主干特征提取网络, α -CIoU 损失函数替换 CIoU 损失函数,K-means++ 聚类重新生成 Anchor,最后利用 DIoU-NMS 后处理方法对红外目标进行检测。mAP 仅降低 0.37%,模型参数量减少了 29%,满足移动端部署的需求。李强龙等^[14] 使用混合池化模块 MPM 改进快速空间池化金字塔 SPPF,利用条形池化解决特征丢失问题,在注意力模块中加入水平和垂直方向上的全局池化操作,使用无批次归一化(BFN)解决算法性能退化问题,提高算法检测性能。楼哲航等^[15] 将 YOLOX 与 Swin Transformer 结合,对 YOLOX 中 Neck 和 Head 部分的激活函数以及标准层进行减枝,优化网络结构,改进后的网络更加适合红外图像的目标检测。

现有的深度学习算法在红外目标检测中取得了一些成绩,但其中一些算法对复杂场景适应性不足,并且红外目标尺度差异较大,缺乏纹理结构等特征,导致检测精度较低。另外一些算法在检测精度有了一定的提升,但由于模型参数量过大,不利于实时检测。因此该算法以 Efficientnet 作为 YOLOv3 的主干网,减少模型参数量;依次在网络中增加空间金字塔池化模块^[16] (spatial pyramid pooling, SPP)、跨阶段局部网络^[17] (cross stage partial network, CSPNet) 以及高效通道注意力^[18] (efficient channel attention, ECA),并引入完整交并比损失函数^[19] (complete IoU loss, CIoU Loss),提高检测的准确率。

1 YOLOv3 算法简介

YOLOv3 网络是目标检测中经典的一阶段网络。YOLOv3 检测的过程是将输入进来的图片划分成不同大小的网格,每个网格内部包含 3 个先验框,判别先验框内部是否真实地包含物体,判断物体的种类,对先验框进行回归调整变成预测框,从而实现端到端的目标检测。YOLOv3 网络分为 3 大部分,Darknet53 主干网络、FPN 特征金字塔、YOLO Head 检测头。YOLOv3 网络使用 Darknet53 作为基础网络,包括 52 个卷积层和 1 个全连接层,Darknet53

中的残差结构可以缓解深层网络训练时梯度消失的问题。主干网输出 13×13 、 26×26 、 52×52 这 3 种尺度,利用 FPN 特征金字塔网络进行多尺度融合,主干网输出的 3 个不同尺度的特征共进行两次融合,加强了网络对不同尺度目标的检测精度。经过 FPN 特征融合获得 3 个有效特征层,将这 3 个有效特征层传入 YOLO Head 进行预测,YOLO Head 本质上是进行两次卷积,一次 3×3 卷积进行分类预测,一次 1×1 卷积进行回归预测。YOLOv3 算法不仅实现了多尺度训练,对小目标检测有了较大的进步。

2 基于 Efficientnet 红外目标检测算法

针对红外数据集色彩单一,特征难提取,检测精度较低等问题,提出一种基于 Efficientnet 红外目标检测算法,将 Efficientnet 主干网替换 YOLOv3 中的 Darknet53 主干网,在不损失精度的情况下,网络模型仅有原模型的 18%,加速网络的计算,缩短训练时间。在 Efficientnet 的最后一个尺度卷积特征图之后加入 SPP 模块,对于多尺度特征的输入,网络始终可以提取出固定大小的特征向量,避免网络过拟合;通过添加 CSPNet 模块,提高模型表达能力和泛化能力;在 YOLOv3 检测头之前加入 ECA 注意力机制,进一步加强特征提取,提升网络检测精度。为了解决网络预测框回归速度慢和精度低的问题,边界框回归损失函数选用 CIoU Loss。基于 Efficientnet 的网络结构图如图 1 所示。

2.1 轻量级网络 (Efficientnet)

经典的神经网络通过放大深度、宽度、输入分辨率 3 个维度中的一个维度提升网络性能。增加网络的深度是指增加网络的层数,通过更深的神经网络实现特征提取,例如常用的残差网络 ResNet^[20],从 ResNet18 到 ResNet200,将网络深度从 18 层增加到 200 层来提高网络准确率,但网络过深会面临梯度消失、训练困难的问题。增加宽度是指增加卷积核的个数,能够获得更高细粒度的特征,实现更多层的特征提取,得到更多的特征,也更容易训练。增加输入图片的分辨率也可以使得网络学习更深层次的特征,提高精确度。

Google2019 提出的轻量级网络 Efficientnet^[21] 将三者结合,对网络的深度、宽度、输入图片的分辨率进行同时缩放,综合优化 3 个维度,在减少了模型计算量的同时,又提升了检测精度。文献[22]对 Efficientnet 进行改进,取得了不错的检测效果。Efficientnet 网络在不同的缩放比例条件下,可得到 EfficientNet-B0~B7 这 8 个不同的模型结构。Efficientnet-B2 网络即高效又轻量,选其作为模型的主干网络。Efficientnet-B2 由 1 个 Stem 层和 7 个 Block 层组成,Block 层由 MBConv (Mobile inverted Bottleneck Convolution) 残差模块堆叠而成,MBConv 包括扩展比例为 1 和 6 的 MBConv1 以及 MBConv6 模块。EfficientNet-B2 模型结构如表 1 所示。

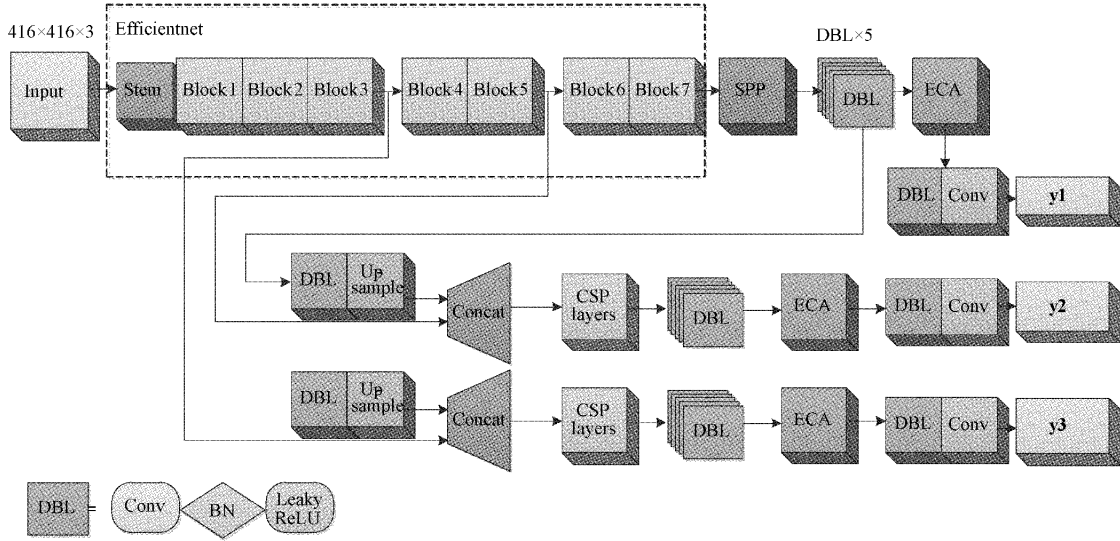


图 1 基于 Efficientnet 的网络结构

表 1 Efficientnet-B2 主干特征提取网络

卷积阶段	参数	分辨率	通道数	层数	步长
Stem	Conv 3×3	208×208	32	1	2
Block1	MBCov1, k3×3	104×104	16	1	1
Block2	MBCov6, k3×3	104×104	24	2	2
Block3	MBCov6, k5×5	52×52	48	2	2
Block4	MBCov6, k3×3	26×26	88	3	2
Block5	MBCov6, k5×5	26×26	120	3	1
Block6	MBCov6, k5×5	13×13	208	4	2
Block7	MBCov6, k3×3	13×13	352	1	1

MBCov 为移动翻转瓶颈卷积模块,其作为 Efficientnet-B2 的基础模块,利用深度可分离卷积和逐点卷积提取特征,还包含了 Swish 激活函数以及 Dropout 连接,并且使用 SENet (squeeze and excitation networks)模块^[23]来引入注意力思想,

SENet 模块更加关注不同通道且信息量最大的通道特征,抑制不重要的通道特征,SENet 注意力的引入有助于提高模型的精度。相对于常规卷积操作,MBCov Block 具有更少的参数数量和运算量。MBCov 网络模型结构如图 2 所示。

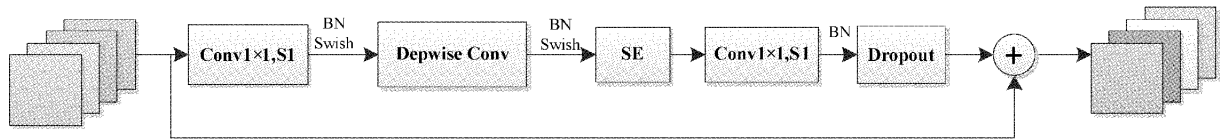


图 2 MBCov 网络结构

SENet 注意力模块由一个全局平均池化和两个全连接层组成。利用 SENet 模块,可以让网络关注需要被关注的通道。其具体实现方式,首先对输入进来的特征层进行全局平均池化,然后进行两次全连接,再取一次 Sigmoid 将值固定在 0~1 之间,获得输入特征层每一个通道的权值,将这个权值乘上原输入特征层。SENet 模块结构如图 3 所示。

2.2 空间金字塔池化 (SPP)

在红外目标检测中,目标尺寸跨度较大,在网络中引入空间金字塔池化 SPP 模块,进行多尺度融合,从不同大小的感受野提取特征,解决大,中,小 3 种尺寸跨度大,识

别难的问题。在 Efficientnet 主干网的最后一个尺度卷积特征图后引入 SPP 模块,SPP 模块由不同的池化层组成,池化核大小分别为 1×1、5×5、9×9 和 13×13,对特征图进行池化,特征图经过 SPP 池化后重新融合传到下一层网络中。SPP 结构如图 4 所示。对于输入图像的不同纵横比和不同尺寸,SPP 可以产生固定大小的输出,不仅提高了图像的尺度不变,并且避免了模型过拟合,模型训练单一尺寸的图像更容易收敛。

2.3 跨阶段局部网络 (CSPNet)

CSPNet 通过将梯度的变化从头到尾地集成到特征图

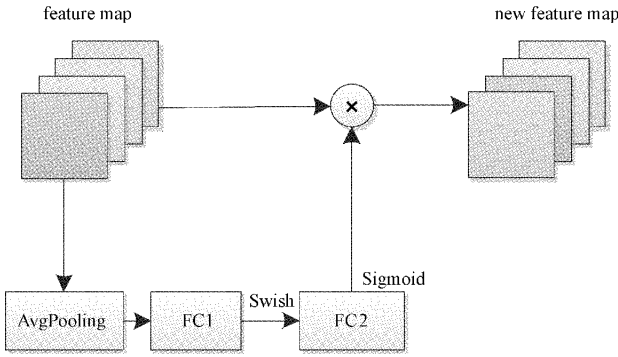


图 3 SENet 网络结构

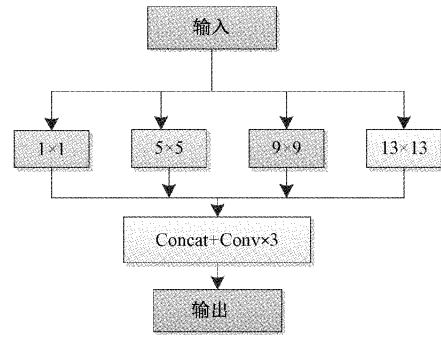


图 4 SPP 网络结构

中,避免过多的重复梯度信息,降低网络的计算量,提升推理速度。CSPNet 被广泛用于目标检测中,在减少计算量的同时,又可以提取丰富的特征信息,以此作为 FPN 网络后的加强特征提取模块。CSPNet 结构有两个输入分支,

其中一个分支进行通道数减半的卷积操作;另一个分支进行 $2 \times X$ 个 CBL 以及卷积操作,将两个分支 Concat,使得输入和输出大小一致,让模型学习到更多的特征,其网络结构如图 5 所示。

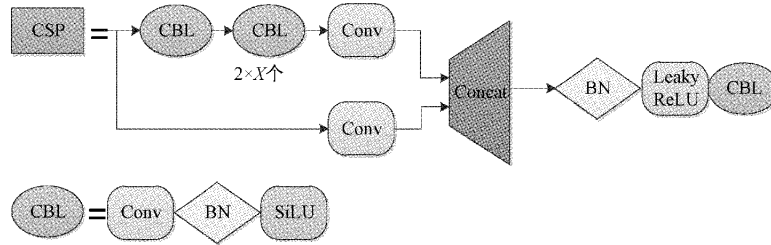


图 5 CSPNet 网络结构

2.4 ECANet 注意力机制

ECANet 是通道注意力机制的一种实现形式,可以看作是 SENet 的改进版。SENet 结构中的全连接层对特征矩阵做降维处理,降维处理会改变特征矩阵的通道信息,对通道注意力机制的预测带来了副作用,捕获所有通道的依赖关系是低效且不必要的。ECANet 去除了原来 SENet 中的全连接层,直接在全局平均池化之后的特征上通过一个一维卷积进行学习,其中一维卷积的卷积核大小为 k ,并经过 Sigmoid 激活函数得到各个通道的权重,将权重与原始输入特征图对应元素相乘,得到输出特征图。ECANet 注意力机制在几乎不增加权重大小的情况下,提升了网络的性能。ECANet 其网络结构如图 6 所示。

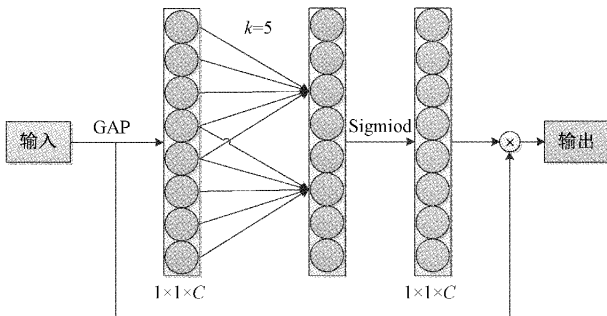


图 6 ECANet 注意力机制网络结构

2.5 损失函数 CIoU Loss

在 YOLOv3 中不同部分的损失项使用不同的损失函数,置信度和分类的损失采用交叉熵函数进行计算,利用均方误差作为边界框回归的损失函数。然而,当目标尺寸误差较大时,均方误差损失函数会产生过大的梯度,并且均方误差损失函数无法反映候选框与真实框相交程度的大小和好坏,从而导致定位不准确。在初始训练过程中会出现梯度消失的现象,回归收敛速度缓慢。为使模型收敛更快更准确,采用 CIoU Loss 作为边界框回归的损失函数,CIoU Loss 考虑了真实框与预测框之间的重叠面积以及中心点距离,还考虑了预测框回归时宽高比的尺度信息,解决了回归重叠一致性问题,使模型收敛效果更佳,检测更精准。

CIoU Loss 的表达式为:

$$L_{IoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (1)$$

其中, IoU 的表达式为:

$$IoU = \frac{A \cap B}{A \cup B} \quad (2)$$

α 的表达式为:

$$\alpha = \frac{v}{(1 - IoU + v)} \quad (3)$$

v 的表达式为:

$$v = \frac{4}{\pi^2} (\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h})^2 \quad (4)$$

式中： b 、 b^{gt} 分别代表了预测框和真实框的中心点， $\rho^2(b, b^{gt})$ 为两个中心点的欧几里得距离， c 为同时覆盖预测框和真实框的最小矩形的对角线长度； ω^{gt} 和 h^{gt} 分别为真实框的宽和高； ω 和 h 分别为预测框的宽和高。

3 仿真实验

3.1 FLIR 数据集

本实验采用 FLIR 公开红外数据集^[24]，数据集包括日间、夜间、雾天、雨天等多种情况，图像背景包括山脉、隧道、建筑等，背景相对复杂。数据集的主要标注类别为人、汽车、自行车，删减一些无效数据后剩余 9 220 张图片，将数据集分为训练集、验证集、测试集，其中训练集和验证集占比 80%，测试集占比 20%。部分图像中的目标较集中并且小尺寸目标居多，存在不同程度的遮挡，增加了检测难度。

3.2 实验配置与训练

实验训练平台采用 Intel(R) XeonW-2245 CPU, GPU 为 NVIDIA RTX 3080, 显存 10 G, 内存大小为 32 GB。在 Windows 10 操作系统中，搭建 Pytorch 1.7 框架，CUDA 版本为 11.2, 用 Python 进行实验。训练过程用 GPU 进行加速。

在实验过程中，所涉及到的训练超参数均保持一致。图像输入尺度为 416×416 ，初始学习率为 0.001，使用 SGD 优化器，学习率下降方式采用余弦退火算法，共训练 300 个 epoch，其中冻结阶段训练 50 个 epoch，Batch size 设置为 16，非冻结训练 250 个 epoch，Batch size 设置为 8。

3.3 实验结果对比与分析

1) 模型的评价指标

使用以下指标对模型进行客观评价。准确率 (Precision) 表示预测为正的样本中有多少是对的；召回率 (Recall) 表示样本中有多少正样本被预测正确了。AP (average precision) 表示每个类别的准确率；mAP 表示所有类别的 AP 值取均值，本文算法的 mAP 值是当 $IoU = 0.5$ 时的平均检测精度；FPS 表示网络模型每秒钟可以处理多少帧图片；F1 值是一个综合指标，由准确率和召回率两者共同决定；模型大小给出了每个模型的参数量。

2) 基于 Efficientnet 算法的消融实验

Efficientnet 算法在 YOLOv3 算法的基础上引入了 5 个改进点，采用轻量高效的 Efficientnet 主干网络，加入 SPP 模块、CSPNet 模块以及 ECANet 注意力机制，采用 CIoU Loss 作为边框位置损失函数。在 FLIR 数据集上进行五类方法的消融实验，消融实验结果如表 2 所示。

3) 基于 Efficientnet 算法与现有 4 种算法的对比

为了说明基于 Efficientnet 算法的有效性，进一步将该算法与 YOLOv3 算法、Faster-RCNN 算法、YOLOv4-Tiny

表 2 在 FLIR 数据集上的消融实验结果

模型	YOLOv3 及其改进	mAP/%	FPS	模型大小/MB
A	Darknet53	70.62	39	236.0
B	Efficientnet	67.15	27	60.0
C	B+SPP	74.14	28	44.4
D	C+CSP Net	78.80	26	44.7
E	D+ECA	80.03	24	44.7
F	E+CIoU Loss	80.74	24	44.7

算法以及 EfficientDet 算法进行对比。在进行实验对比时，采用相同的实验配置，在 FLIR 数据集上实现这些算法，得到的结果如表 3 所示。

表 3 5 种算法的实验结果对比

模型	特征提取网络	mAP/%	FPS	模型大小/MB
YOLOv3	Darknet53	70.62	39	236.0
Faster-RCNN	Resnet50	51.66	16	108.0
YOLOv4-Tiny	CSPDarknet53	46.35	97	22.6
EfficientDet	Efficientnet	75.23	13	31.4
本文算法	Efficientnet-B2	80.74	24	44.7

由表 3 可知，在检测精度方面，改进后的算法 mAP 达到了 80.74%，比 YOLOv3 算法高出 10.12%，高出 EfficientDet 算法 5.51%，高于 YOLOv4-Tiny 算法 34.39%，高出 Faster-RCNN 算法 29.08%。在模型大小方面，本文算法参数量比 YOLOv3 低 191.3 M，比 Faster-RCNN 算法低 63.3 M。实验结果表明，Eff-YOLOv3 算法对红外目标检测有更好的检测性能。

4) 5 种算法 F1 曲线对比

为了更好地判断模型的泛化能力，同时改善准确率和召回率不能同时达到最佳的问题，选取 F1 值作为评价 5 种模型性能的指标。F1 值是准确率与召回率的调和平均数，如式(5)所示。F1 值综合考虑了准确率和召回率，让两者同时取最高值，找到两者之间的平衡点。

$$F1 = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

F1 曲线越平稳并且越接近 1，则表示该算法的检测性能越好。图 7 是 5 种算法 Bicycle、Car、Person 3 个类别的 F1 曲线图。从图 7 可以看出，改进后的算法的 3 个类别 F1 曲线趋于平缓，综合考虑准确率与召回率，相比其他 4 种算法，均表现最优。

5) 5 种算法检测可视化结果

为了更直观地看到 5 种算法的检测结果，图 8 给出 3 种场景下的检测可视化结果图。场景 1 和场景 2 目标较少，检测难度较小。场景 3 目标密集，存在目标遮挡的情

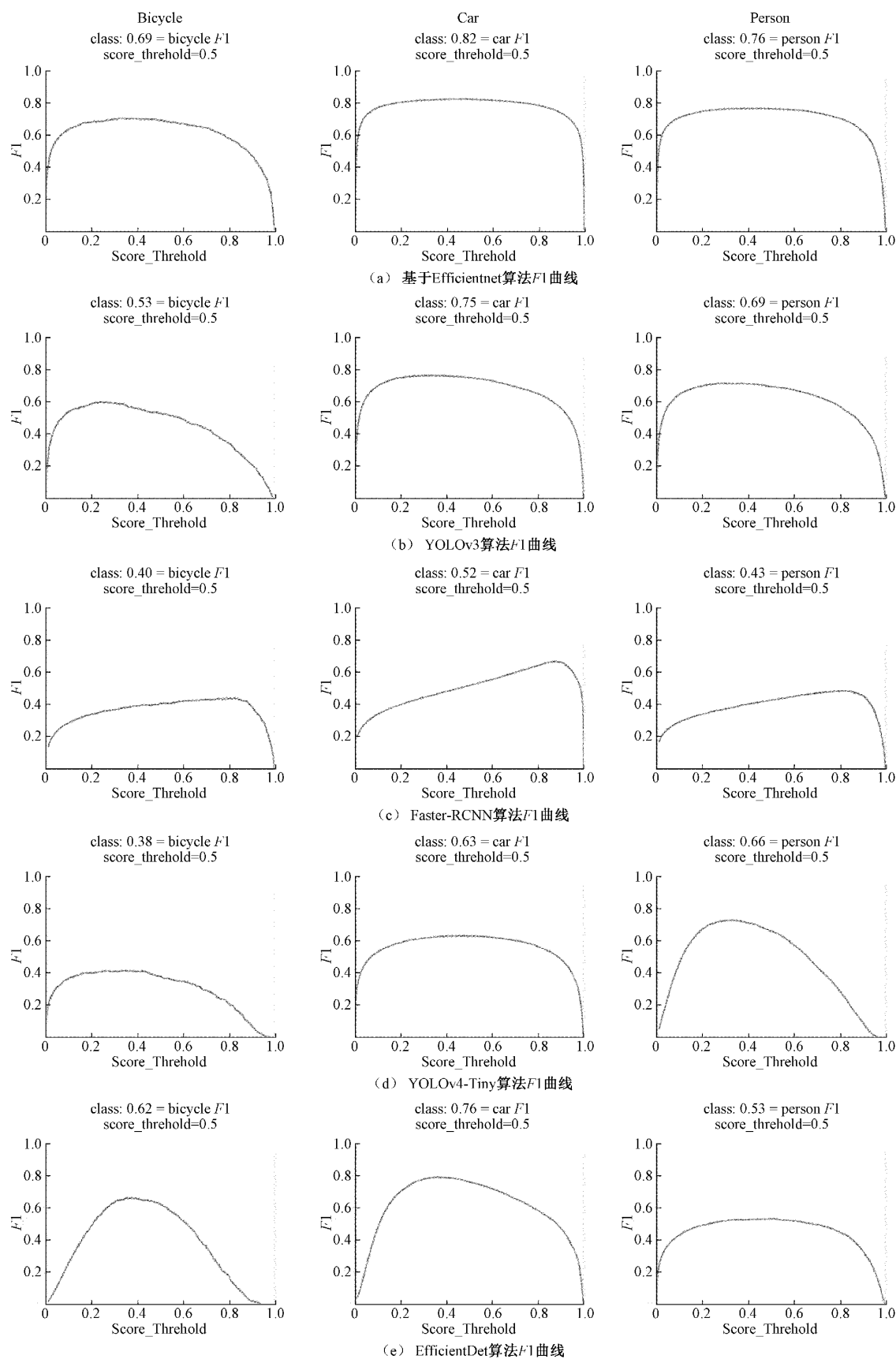


图7 5种算法F1曲线对比

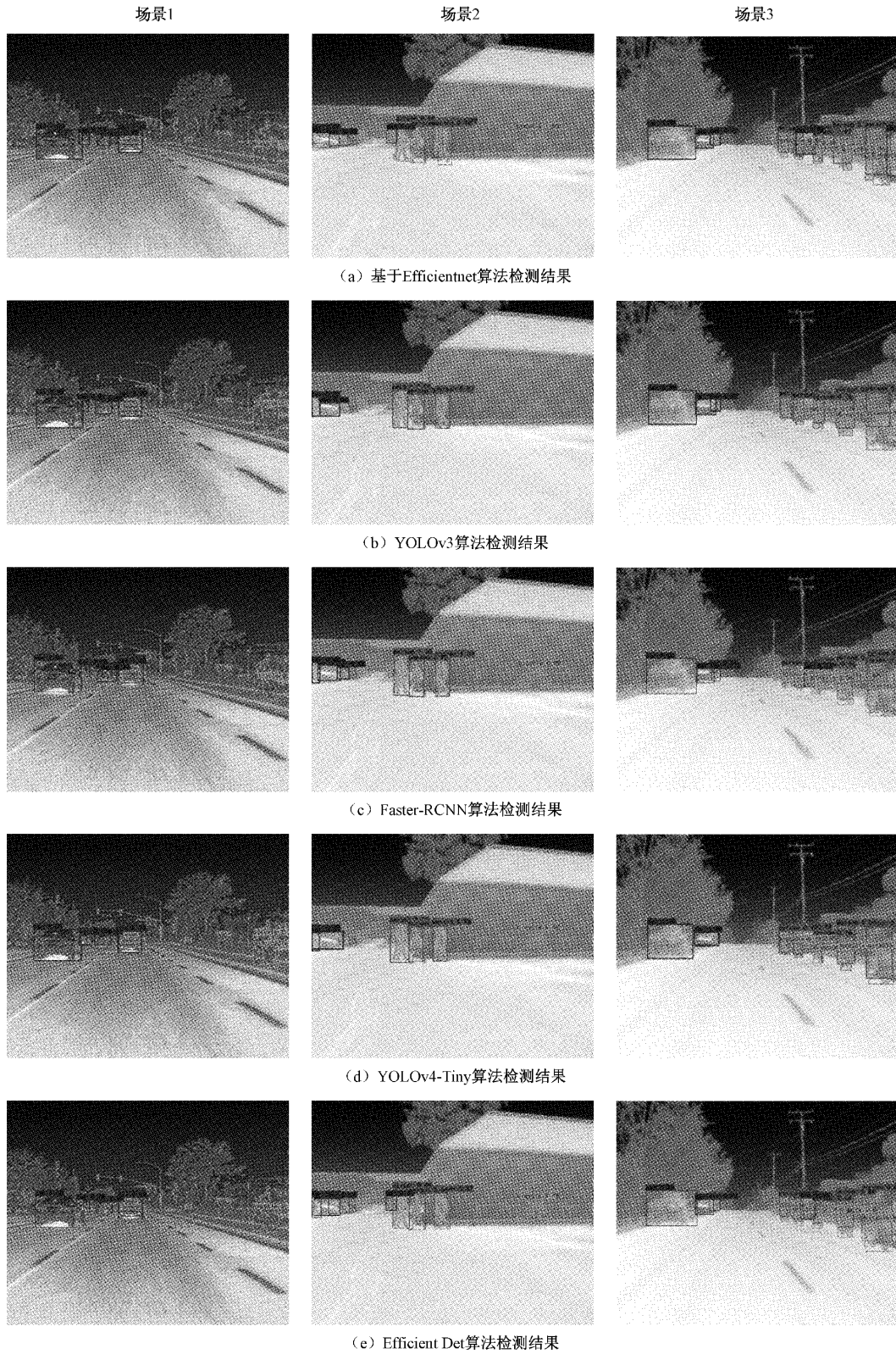


图 8 5 种算法检测可视化图像

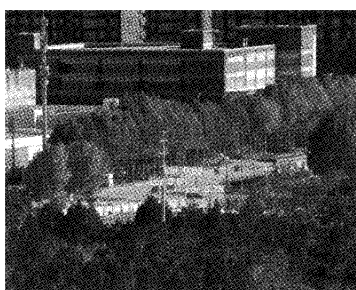
况,有一定的检测难度。从图 8 的可视化结果可以看出,本文算法在 3 种场景下目标检测准确率最高,且没

有漏检误检情况;从场景 1 的检测结果看出, YOLOv3 算法和 EfficientDet 算法对车辆的检测出现漏检, Faster-

RCNN 算法对车辆检测出现误检,目标框出现偏差, YOLOv4-Tiny 算法对“人”的类别误检成“自行车”;在场景 2 中,对于被遮挡的“汽车”YOLOv3 算法、Faster-RCNN 算法及 YOLOv4-Tiny 算法均出现漏检,且 YOLOv3 算法对“自行车”也出现漏检的情况;场景 3 中,除本文算法,其他 4 种算法对右侧被遮挡车辆均出现漏检。表 4 给出 YOLOv3 算法和本文算法的“大中小”3 种类别的 AP 值,大目标和中目标的 AP 值分别提升 5.8% 和 3.2%,小目标提升 7.1%,进一步说明本文算法检测小目标更加精准。综上,从可视化结果可以看出,本文算法在 5 种算法中表现最佳。从“大中小”3 种类别 AP 值可知,相比 YOLOv3 算法,本文算法在保证大、中目标检测准确率的基础上,极大提高了小目标的检测性能。



(a) 两辆车误检为三辆车



(b) 目标全部检出



(c) 检出但置信度较低

图 9 真实场景的检测结果

本文算法针对复杂场景下 FLIR 数据集,虽然改善了错检漏检情况,但错检漏检情况仍然存在。由于应用到实际生活中的红外图像背景更加复杂,本文算法还需进行多种场景下的训练和测试,仍需对目标的细节特征进行精准提取。

4 结 论

红外目标图像检测具有全天候工作,不受雾天、雨天等恶劣天气的影响等优点,在目标检测中具有重要意义。虽然现有的红外目标检测算法有了很大的进步,漏检、误检等问题仍然需要进一步优化。针对红外目标检测现有技术的不足,提出基于 Efficientnet 的红外目标检测算法,轻量高效的 Efficientnet 作为主干网络,在此基础上,引入 SPP 模块、CSPNet 模块、ECA 注意力机制以及 CIoU 损失函数等方法。改进后的算法在公开 FLIR 数据集上进行实验,相比 YOLOv3 算法 mAP 提高了 10.12%,模型大小是原来的 18%,在轻量的同时提高模型的检测准确率。与已有的 4 种经典算法比较,基于 Efficientnet 算法表现最佳,在小目标、被遮挡目标的检测过程中准确率最高,改进后的模型更具泛化能力。

表 4 “大中小”3 类目标 AP 值 %

模型	AP (IoU=0.5;0.95)		
	大目标	中目标	小目标
YOLOv3	56.1	51.6	30.6
本文算法	61.9	54.8	37.7

6) 真实场景下的检测结果

为了进一步验证本文算法更好的性能,将本文算法对真实场景下的红外图像进行检测,真实场景的数据集是在高架桥上对道路路口进行拍摄。因为拍摄角度高,距离目标远,道路上有其他物体干扰,导致检测难度大。针对真实场景检测可视化结果如图 9 所示,图 9(a)中的两辆车误检为 3 辆车,图 9(b)中的目标全部检出,图 9(c)中虽然检测出目标,但是置信度较低。

参考文献

- [1] 张汝榛,张建林,祁小平,等. 复杂场景下的红外目标检测[J]. 光电工程, 2020, 47(10): 128-137.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [3] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [4] 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 and Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [6] LIU W, ANGUOLOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham,

- 2016; 21-37.
- [7] 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.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [11] 刘冬,李庭鑫,杜宇,等.基于MCA-YOLO的轻量级红外实时目标检测算法[J/OL].华中科技大学学报(自然科学版): 1-7 [2022-09-03]. DOI: 10.13245/j.hust.239405.
- [12] 李北明,金荣璐,徐召飞,等.基于特征蒸馏的改进Ghost-YOLOv5红外目标检测算法[J].郑州大学学报(工学版),2022,43(1):20-26.
- [13] 黄磊,杨媛,杨成煜,等.FS-YOLOv5:轻量化红外目标检测方法[J].计算机工程与应用,2023,59(9):215-224.
- [14] 李强龙,周新文,位梦恩,等.基于条形池化和注意力机制的街道场景红外目标检测算法[J].计算机工程,2023,49(8):310-320.
- [15] 楼哲航,罗素云.基于YOLOX和Swin Transformer的车载红外目标检测[J].红外技术,2022,44(11):1167-1175.
- [16] 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.
- [17] 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.
- [18] WANG Q, WU B, ZHU P, et al. Supplementary material for ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2020: 13-19.
- [19] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [21] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, PMLR, 2019: 6105-6114.
- [22] 王敏,王康,李晟,等.基于改进EfficientDet的药丸检测算法[J].电子测量技术,2022,45(19):136-142.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [24] 适用于算法训练的免费FLIR热数据集[EB/OL]. <https://www.flir.cn/oem/adas/adas-dataset-form/>.

作者简介

侯艳丽,副教授,主要研究方向为无线通信、图像处理、人工智能。

E-mail:houyanli26@126.com

王娟(通信作者),硕士研究生,主要研究方向为深度学习目标检测。

E-mail:wxc090816@163.com