

# 基于深度学习的红外目标检测综述

张睿,李允臣,王家宝\*,李阳,苗壮

(陆军工程大学 指挥控制工程学院,江苏南京 210007)

**摘要:**红外图像由于分辨率低、纹理细节不足,且缺乏颜色信息,导致目标成像模糊,检测难度大。基于深度学习的红外目标检测技术,通过运用神经网络自动提取复杂的目标特征,大大提高了检测精度和检测效率,在自动驾驶、安防监控、军事侦察等领域得到了非常广泛的应用。该文对红外目标检测面临的困难和挑战进行了详细分析,并从数据增强、迁移学习、视觉注意力机制、多尺度特征融合、多模态图像融合和轻量化改进等六个方面,对基于深度学习的红外目标检测研究改进方向进行了系统阐述。针对红外目标检测数据集缺乏的问题,梳理汇总了11个红外目标检测数据集。同时,结合当前发展现状,对红外目标检测的未来发展方向进行了展望,可为其他研究者提供参考借鉴。

**关键词:**深度学习;红外目标检测;迁移学习;注意力机制;特征融合;图像融合

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2023)11-0001-08

doi:10.3969/j.issn.1673-629X.2023.11.001

## Survey on Infrared Object Detection Based on Deep Learning

ZHANG Rui, LI Yun-chen, WANG Jia-bao\*, LI Yang, MIAO Zhuang

(College of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China)

**Abstract:** Infrared images have low resolution, insufficient texture details, and lack of color information, resulting in blurred object imaging and difficult detection. Deep learning-based infrared object detection technology uses complex neural networks to automatically extract object features, has greatly improved detection accuracy and detection efficiency, and has been widely used in the fields of autonomous driving, security surveillance, and military reconnaissance. We present a detailed analysis of the difficulties and challenges facing infrared target detection, and systematically describe the improvement directions of deep learning-based infrared object detection research in six aspects, including data enhancement, transfer learning, visual attention mechanism, multiscale feature fusion, multimodal image fusion and lightness improvement. To address the problem of lack of infrared object detection datasets, we compose and summarize 11 infrared object detection datasets. At the same time, the future development direction of infrared object detection is prospected with the current development status, which can provide reference for other researchers.

**Key words:** deep learning; infrared object detection; transfer learning; attention mechanism; feature fusion; image fusion

## 0 引言

目标检测是指获得图像或视频中物体的位置坐标,并判断其类别。红外目标检测是目标检测的一个重要分支,在自动驾驶、安防监控、军事侦察等领域都有广泛的应用。尤其是在军事领域,红外目标检测可用于对武装人员、军事车辆等各类目标进行侦察监视,有着非常重要的军事价值。

2014年以来,基于深度学习的目标检测算法通过构建复杂的神经网络,并运用反向传播算法更新网络参数,自动高效地提取物体的复杂特征,在目标检测领域取得了巨大成功。基于深度学习的目标检测算法通

常可分为两阶段目标检测算法和单阶段目标检测算法。两阶段目标检测的代表算法主要有R-CNN(Region-based Convolutional Neural Network)系列,包括:R-CNN<sup>[1]</sup>、Fast R-CNN<sup>[2]</sup>、Faster R-CNN<sup>[3]</sup>、FPN<sup>[4]</sup>、Cascade R-CNN<sup>[5]</sup>等。其检测流程为先将图像输入神经网络提取特征,再采用一定算法生成大量候选区域并筛选,然后对筛选的区域内可能包含的目标进行分类和定位,如图1(a)所示。两阶段检测算法精度较高,但速度较慢。为了提高检测速度,出现了以YOLO(You Only Look Once)系列<sup>[6-13]</sup>和SSD(Single Shotmulti-box Detector)系列<sup>[14-15]</sup>为代表的单阶段目

收稿日期:2022-12-31

修回日期:2023-05-04

基金项目:江苏省自然科学基金项目(BK20200581)

作者简介:张睿(1977-),男,教授,博士,研究方向为信息融合;通信作者:王家宝(1985-),男,副教授,博士,研究方向为模式识别和机器学习。

标检测算法。单阶段目标检测算法省去了候选区域生成阶段,用神经网络对图像提取特征后,直接通过回归分析得到目标的类别和位置坐标,如图 1(b)所示。单阶段目标检测算法精度虽然略低于两阶段目标检测算法,但具有明显的速度优势,可进行实时检测。

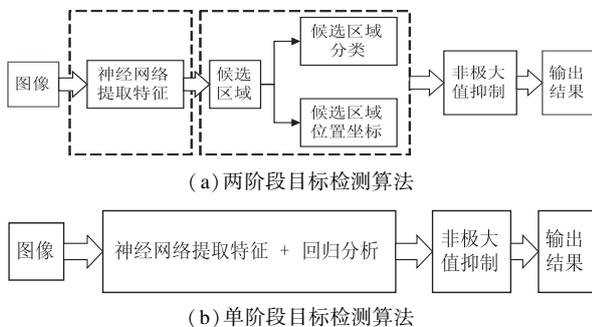


图 1 基于深度学习的目标检测算法

上述深度学习算法最初均应用于可见光领域。近年来,人们开始将深度学习算法引入红外目标检测领域,并取得了不少成果。虽然在可见光目标检测领域,已有较多文章对深度学习目标检测算法进行了系统综述,但在红外目标检测领域,相关综述还比较缺乏。因此,该文主要围绕针对红外目标的深度学习检测算法研究进展进行综述,以促进红外目标检测技术的发展。

## 1 红外目标检测面临的困难和挑战

红外图像是由红外成像设备通过测量物体向外辐射的热红外线得到的,与可见光图像的成像原理差异较大。受红外图像的成像机制影响,红外目标检测面临的困难和挑战主要有五个方面:

(1) 数据集资源缺乏。由于红外成像设备价格较为昂贵,且红外图像数据集大多用于特殊领域,导致研究者可获取的红外图像数据集资源较少。近年来,虽然有一些红外目标检测数据集公开发布,但数据规模与可见光数据集相比,差距还很大。

(2) 背景干扰大。在城市环境中,背景热源多,红外目标与背景热源难以区分,而在野外环境中,树木植被的红外影像较为杂乱,红外目标的形状轮廓等特征易被破坏,红外目标很容易淹没在背景中而难以辨别。

(3) 图像分辨率低。可见光成像设备大多采用高清摄像头,拍摄的可见光图像分辨率较高。而红外图像的分辨率一般在  $640 \times 512$  像素甚至更低,导致红外目标的像素分辨率较低,占有的像素面积小,可利用的目标特征少。

(4) 纹理细节不足。由于红外图像类似于灰度图,缺乏颜色信息,且物体的温度分布大多是渐变的,导致红外目标纹理细节信息不足,成像较模糊。

(5) 弱小目标检测难。在高空或远距离拍摄条件下,红外目标尺寸非常小,有时呈现为点状或斑点状,

不仅缺乏颜色、纹理信息,还缺乏形状信息,特征信息极少,导致漏检及虚警率较高。

## 2 红外目标检测研究改进方向

基于深度学习的目标检测算法在诞生之初主要用于可见光目标检测,而红外目标与可见光目标成像差异较大,因此,相应的深度学习检测算法并不完全适用。所以,在将可见光领域深度学习目标检测算法用于红外目标检测时,应针对红外数据集和红外目标特点进行改进。

### 2.1 数据增强

深度学习方法高度依赖大规模数据集,而红外数据集规模一般较小,可供训练的目标样本少。对此,可进行数据增强处理,通过水平/垂直翻转、旋转、缩放、随机裁剪、改变对比度、增加噪声干扰、图像混合和利用 GAN 生成伪红外图像等方法,设法增加图像数量和目标样本数量。

赵晓枫等<sup>[16]</sup>、Zhang X X 等<sup>[17]</sup>在自建的小规模红外目标数据集上,采用旋转、裁剪、改变对比度等方法来扩充样本数量。李北明等<sup>[18]</sup>则引入 Mosaic 和 Copy-paste 两种方法进行数据增强。Mosaic 数据增强是使用多张图片经过随机裁剪、缩放,拼接在一起, Copy-paste 数据增强则是将原图中的目标经过随机尺度的缩放后,重新粘贴到图像的任意位置,两种方法均可有效增加训练样本数量。吴晗等<sup>[19]</sup>则尝试使用 CycleGAN 通过风格迁移生成伪红外图像,并在 CycleGAN 中添加了通道和空间注意力,从而实现了红外数据样本的有效扩充。但基于 GAN 的图像生成方法,还存在模型崩塌的问题,训练较为困难,且生成的伪红外图像与真实红外图像存在一定差异。

数据增强的方法虽然可提升网络模型的泛化能力和鲁棒性,但可能改变数据的真实分布,引入数据噪声,造成目标的误检。

### 2.2 迁移学习

对于规模较小的红外数据集,即使经过数据增强,可能仍然无法满足复杂的神经网络对大量训练数据的需求。对此,可采用迁移学习的方法来进一步弥补训练数据的不足,提高模型的泛化能力和检测精度。

迁移学习是将其他领域的知识转移到新的领域的一种方法。通过将模型在大规模的数据集上预训练,使模型具备一定的先验知识,而后再将预训练模型在其他领域的数据集上进行再学习,进一步调整网络参数达到最优,即可使模型适应新的任务。Zhang X X 等先在大规模的航拍可见光车辆数据集上进行预训练,然后将预训练模型在航拍红外车辆数据集上进行再训练。由于所利用的数据均为航拍车辆数据,目标相似

度较高,取得了较好的检测效果。王悦行等<sup>[20]</sup>利用计算机仿真生成大量的边缘、轮廓、纹理等特征相似的仿真红外舰船,然后利用特征自适应迁移学习方法,实现从仿真红外舰船到真实红外舰船的跨域知识迁移。李维鹏等<sup>[21]</sup>在利用大规模的可见光数据集预训练的基础上,还进一步利用大量的未标注红外图像对网络进行半监督学习调优,并提出了特征相似度加权的伪监督损失函数,从而更充分地利用了未标注红外图像数据。

迁移学习的方法能够利用其他领域的的数据,辅助提升模型在红外领域的检测性能。但在将其他领域的的数据知识迁移到红外领域时,需要考虑跨域知识的匹配问题,否则,可能将错误的知识引入模型,给模型带来不确定性。

### 2.3 视觉注意力机制

针对红外目标特征较弱且背景干扰大的问题,引入视觉注意力机制,可较好地抑制背景噪声干扰。同时,注意力机制还能在一定程度上弥补数据样本的不足,即使只有少量的数据样本,也能提取到有效的红外目标特征。

人类视觉在观察整幅图像时,会将注意力聚焦于某些关键目标区域,而忽略大部分背景区域。深度学习中的注意力机制通过模仿人类视觉,将重要的通道

或位置区域赋予更大的权重,从而获得通道或空间位置注意力,相关注意力模块如图 2 所示。

Hu J 等<sup>[22]</sup>提出 SE (Squeeze and Excitation) 注意力模块,也称通道注意力模块,如图 2(a) 所示。其通过全局平均池化将特征图压缩,并通过全连接层和激活函数学习得到不同特征通道的权重,再对各通道加权,从而增强有效通道的信息。李向荣等<sup>[23]</sup>在 YOLOv4 的特征金字塔融合过程中加入 SE 模块,来增加有效特征通道的融合比重。代举等<sup>[24]</sup>则在 YOLOv5 的骨干网中添加 SE 模块,在特征提取过程中抑制森林等复杂背景对红外弱小目标的干扰。

SE 注意力模块只考虑了通道的重要程度,没有考虑空间位置的重要程度,而位置信息是捕获目标结构的关键。Woo S 等<sup>[25]</sup>提出 CBAM 模块 (Convolutional Block Attention Module),如图 2(b) 所示。其在通道注意力中增加了一个并行的最大池化层,进一步增强了通道注意力,而后在通道维度进行最大池化和平均池化压缩,从而获得空间注意力信息。陈举等<sup>[26]</sup>在 YOLOv3 骨干网 DarkNet-53 中加入 CBAM 模块,在仅使用小规模红外数据集训练的情况下,检测精度超过了大规模数据预训练的模型。Du S J 等<sup>[27]</sup>将 CBAM 中的  $7 \times 7$  普通卷积改为空洞卷积,设计了 Dilated CBAM,减少了对目标纹理细节的依赖。

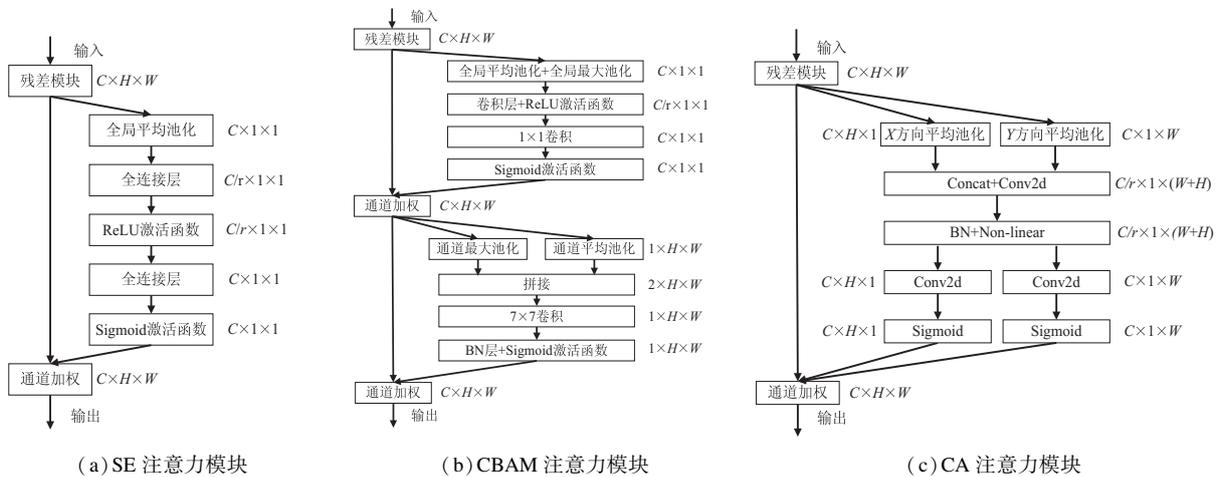


图 2 注意力模块

Hou Q B 等<sup>[28]</sup>认为 CBAM 模块无法获得长程依赖关系,于是将通道注意力分解为水平和垂直注意力,并沿两个方向获取特征图的远程依赖关系,将空间坐标信息整合到通道注意力中,得到坐标注意力 (Coordinate Attention)。胡焱等<sup>[29]</sup>在 YOLOv5s 中特征信息丰富的 Dark5 阶段引入坐标注意力模块,对低分辨率的红外行人目标检测取得较好效果。

除以上注意力机制外,杨其利等<sup>[30]</sup>模仿人类视觉,采用滑动窗口对图像进行扫描采样,以缓解弱小点状红外目标检测中的正负样本不平衡问题。杨子轩

等<sup>[31]</sup>提出解耦注意力机制 (Factor Decoupled Attention),其先通过分子因解机获取全局通道注意力,再通过局部平均池化和局部最大池化获取更细粒度的局部注意力,最后通过多尺度卷积获取不同感受野的空间注意力。FDA 方法通过将全局和局部注意力机制融合,对红外弱小目标检测效果较好。

相关注意力模块可灵活嵌入网络模型中,对提升模型性能有较好效果,而且其参数量和计算量成本也不大,性价比较高。但在模型不同位置,注意力模块带来的性能提升并不相同,有时效果提升并不明显,还可

能降低检测速度。

### 2.4 多尺度特征融合

由于红外图像本身分辨率低,红外弱小目标特征少,而且神经网络在提取特征的过程中,还会对图像不断进行下采样处理,导致目标信息出现丢失,给红外弱小目标检测带来不利影响。对此,采用多尺度特征融合的方法,可以有效增强红外弱小目标的特征,提高对弱小红外目标的检测精度。

多尺度特征融合方法可分为两类:一是将网络模型不同层次的多尺度特征图进行融合。在深度神经网络中,深层次的特征图包含丰富的全局语义信息,而浅层次的特征图包含丰富的局部细节信息。通过将深层与浅层的特征图融合,可兼顾获取目标的全局语义信息和局部细节信息。二是在网络同一层次中构建具有多感受野的多分支结构,提取目标的不同尺度的局部特征,然后将多尺度特征进行融合,从而更全面地捕获目标特征。

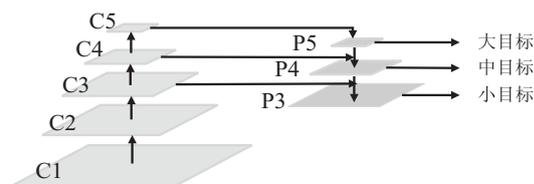


图 3 特征金字塔网络

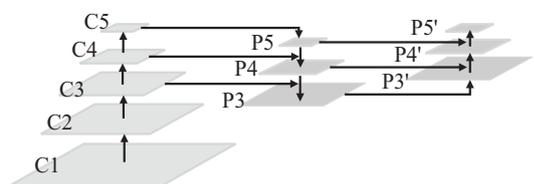


图 4 路径聚合金字塔

Lin T Y 等提出特征金字塔网络 FPN (Feature Pyramid Network),如图 3 所示。其将深层特征图自上而下逐层上采样后与浅层特征图融合,并在高分辨率的浅层特征图上检测小目标,有效提高了小目标的检测精度。Liu S 等<sup>[32]</sup>提出路径聚合网络 PANet (Path Aggregation Network),其在 FPN 自上而下融合的基础上,增加了自下而上的融合路径,将底层的细节位置信息进一步传至上层特征图。YOLOv4 借鉴 PANet 结构,设计了 PAFPN (Path Aggregation Feature Pyramid Network) 结构,如图 4 所示。刘杨帆等<sup>[33]</sup>、林健等<sup>[34]</sup>进一步将金字塔层数扩展为 4 层,以检测红外弱小目标。舒朗等<sup>[35]</sup>将特征金字塔中的残差连接改为 DenseNet 式的密集连接,以保留更丰富的特征信息。朱子健等<sup>[36]</sup>则将特征金字塔的上采样操作改为反卷积,以更好地还原红外弱小目标的细节特征。盛大俊等<sup>[37]</sup>针对红外装甲车辆边缘轮廓较清晰的特点,设计了语义特征提取模块、上下文聚合模块、边缘感知融合

模块,对特征金字塔的多尺度特征进一步融合,以提取装甲车辆的轮廓细节信息和边缘语义信息。

Liu S T 等<sup>[38]</sup>通过采用不同尺寸的卷积核和不同膨胀系数的空洞卷积构成多分支结构,提出感受野网络 RFBNet (Receptive Field Block Net),如图 5 所示,有效增强了网络的特征提取能力。顾燕等<sup>[39]</sup>在 Faster R-CNN 骨干网中使用膨胀系数为 1、2、3 的空洞卷积,从而获得 3×3、5×5、7×7 三种不同大小的感受野。朱子健等在 YOLOv3 骨干网的残差模块中并行添加一个 3×3 最大池化分支和一个 1×1 卷积分支,提出了 PaRNet (Parallel Residual Network)。蒋昕昊等<sup>[40]</sup>则在特征金字塔融合之后引入 RFB 模块,并减少膨胀系数,以适应红外弱小目标检测任务。高凡等<sup>[41]</sup>则借鉴 Inception 网络的多分支结构,设计了 PMFPSNet (Parallel Multi Feature Path Network)。楼哲航等<sup>[42]</sup>将 YOLOX 的骨干网改为 Swin Transformer,以获取全局感受野特征,并在颈部和检测头采用卷积模块提取局部特征,通过融合全局和局部特征,对红外小目标检测取得较好效果。

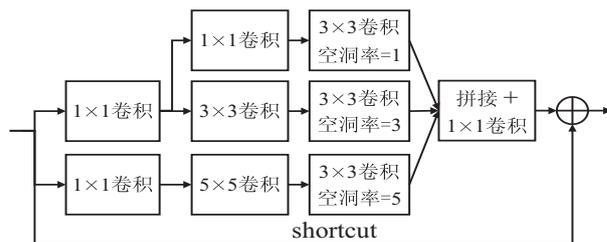


图 5 感受野网络

多尺度特征融合的方法,虽然能够有效增强红外弱小目标特征,提高检测精度,但多层金字塔结构和多分支结构都会增加模型的参数量和计算量,降低训练和推理速度。

### 2.5 多模态图像融合

红外目标受红外成像机制的制约,其本身的特征信息不足。对此,通过融合可见光图像或其他模态图像,利用不同模态图像的互补信息增强红外目标的特征,可使红外目标检测更加准确可靠。

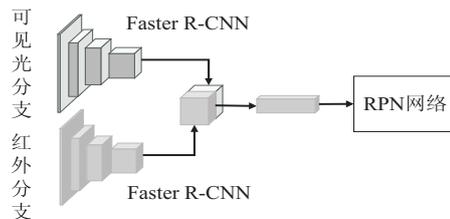


图 6 双分支融合检测网络

Geng K K 等<sup>[43]</sup>基于 Faster R-CNN 进行改进,设计了一个由红外分支和可见光分支组成的双分支融合检测网络,如图 6 所示。其采取特征级的融合方式,将两种模态图像的特征图拼接后,通过卷积模块融合,有

效提高了低辨识度条件下的红外目标检测精度。Liu J Y 等<sup>[44]</sup>采用基于 GAN 的方法,设计了一个目标感知对抗学习网络(TarDAL Network),实现了对红外与可见光图像的自适应融合。该网络由一个生成器、两个目标感知鉴别器以及 YOLOv5 检测网络组成。其中,生成器用于对红外与可见光图像融合,两个鉴别器分别用来鉴别红外图像的目标信息和可见光图像的纹理信息,实现了红外特征和可见光特征的自适应融合。Sun Y M 等<sup>[45]</sup>基于 Faster R-CNN 提出了一个由可见光分支、红外分支和融合分支组成的三支检测网络 UA-CMDet。其同时采用特征融合和决策结果融合的方法,同时利用单模态分支和特征融合分支进行检测,并根据检测结果进行联合决策,取得了最优效果。赵明等<sup>[46]</sup>则采用基于 CycleGAN 的方法,先利用红外图像生成伪可见光图像,再对红外和伪可见光图像分别提取特征,并构建特征金字塔,而后对特征金字塔进行加权融合,在自动驾驶红外目标检测上取得了较好效果。赵兴科等<sup>[47]</sup>则利用 BASNet 生成红外图像的显著图,再将显著图与红外图像融合。由于显著图有较清晰的边界,融合后的图像使红外目标更加清晰,对复杂背景环境下的红外目标检测效果较好。

多模态图像融合的方法,能够有效弥补红外图像特征信息的不足,但成对的多模态数据集的采集、筛选、标注等工作量大,数据资源获取成本较高。同时,多模态图像融合检测网络一般由多个单模态检测网络构成,整体参数量和计算量较大,检测速度较慢,在实际应用时存在一定局限性。

### 2.6 轻量化改进

深度神经网络虽然性能优异,但其模型复杂,参数

量、计算量大。为了将红外目标检测模型部署在资源受限的移动端,就必须对模型进行轻量化改进。目前,研究者主要是借鉴 MobileNet<sup>[48]</sup>、EfficientNet<sup>[49]</sup>、GhostNet<sup>[50]</sup>等轻量级网络的设计方法,设法减少网络深度和模型大小。

针对航拍红外车辆实时检测任务,Liu X F 等<sup>[51]</sup>基于 YOLOv3 进行改进,除最大池化下采样层外,Dark 1/2/3/4 只保留 1 层卷积,Dark 5 保留 5 层卷积,将 53 层的 DarkNet 压缩为只有 15 层的极简模型。秦鹏等<sup>[52]</sup>用 EfficientNet 代替 YOLOv3 骨干网,提出 Effi-YOLOv3 模型,其在 FLIR 数据集上的检测精度超过了 YOLOv3,而参数量仅为 YOLOv3 的 1/3。李北明等则用 GhostNet 代替 YOLOv5 的骨干网,并运用特征知识蒸馏的方法,使用 Scaled-YOLOv4<sup>[53]</sup> 指导该网络学习。其仅有 1.9 M 的参数量,但在红外数据集上的精度和速度都超过了 YOLOv5-s 模型。赵兴科等借鉴 MobileNetv2 网络,提出了轻量的 YOLOv3-MobileNetv2 模型。

目前,相关轻量化改进方法虽然减小了模型规模,提高了检测速度,但在模型精度和泛化能力上,与大模型仍有一定差距。

## 3 红外目标检测数据集

现有的目标检测数据集多为可见光图像数据集,例如 PASCAL VOC 数据集、MS-COCO 数据集、UA-DETRAC 数据集等,而红外数据集资源相对较少,制约了红外目标检测的研究开展。为便于其他研究者开展研究,现将可公开获取的红外目标检测数据集资源梳理汇总如下,如表 1 所示。

表 1 红外目标检测数据集

数据集	拍摄时间	模态	图像/视频帧数量	目标样本数量	目标类别数	分辨率	发布时间
OSU Thermal Pedestrian Database	白天	红外+可见光	284 对	不详	1	320×240	2007
VEDAI	白天	红外+可见光	1,210 对	3,699	9	512×512 / 1 024×1 024	2014
KAIST	白天+夜晚	红外+可见光	95,328 对	103,128	3	640×480	2015
CVC-14	白天+夜晚	红外+可见光	8,519 对	31,074	3	640×512 / 1 280×1 024	2016
SCUT FIR Pedestrian Dataset	夜晚	红外	211,011 张	477,907	6	4 500×3 600	2017
FLIR ADAS	夜晚	红外	26,442 张	520k	15	640×512	2018
NPU_CS_UAV	白天	红外	1,199 张	不详	1	640×512	2018
LLVIP	白天+夜晚	红外+可见光	15,488 对	89,174	1	1 080×720	2020
Dim-small Aircraft	白天	红外	16,177 张	16,944	1	256×256	2021
DroneVehicle	白天+夜晚	红外+可见光	28,439 对	953k	5	840×712	2021
M <sup>3</sup> FD	白天+夜晚	红外+可见光	4,200 对	33,600	6	1 024×768	2022

(1) OSU Thermal Pedestrian Database<sup>[54]</sup>:是美国俄亥俄州立大学制作的道路监控红外行人检测数

据集。

(2) VEDAI<sup>[55]</sup>:是卫星航拍目标检测数据集,提供

了 512×512 和 1 024×1 024 两种分辨率的图像。

(3) KAIST<sup>[56]</sup>: 是韩国科学技术高级研究院下属实验室制作的自动驾驶红外行人检测数据集。

(4) CVC-14<sup>[57]</sup>: 是西班牙巴塞罗那自治大学发布的道路红外行人检测数据集。

(5) SCUT FIR Pedestrian Dataset<sup>[58]</sup>: 是由华南理工大学制作的夜间道路行人检测数据集。

(6) FLIR ADAS: 是由生产红外热成像仪系统的 FLIR SYSTEMS 公司制作的夜间道路红外目标数据集。

(7) NPU\_CS\_UAV<sup>[51]</sup>: 是西北工业大学制作的无人机航拍车辆检测数据集。

(8) LLVIP<sup>[59]</sup>: 是北京邮电大学制作的微弱光线条件下的道路行人检测数据集。

(9) Dim-small Aircraft<sup>[60]</sup>: 是国防科技大学制作的小型无人机目标检测数据集。

(10) DroneVehicle<sup>[45]</sup>: 是天津大学制作的无人机航拍车辆检测数据集。

(11) M<sup>3</sup>FD<sup>[44]</sup>: 是大连理工大学制作的自动驾驶道路目标检测数据集。

## 4 未来发展方向

基于深度学习的红外目标检测研究虽取得一定进展,但距离满足人们需求仍有差距,未来还有很大的发展空间。结合当前发展现状,就未来发展方向展望如下:

### 4.1 大规模的红外目标检测数据集基准

深度学习的发展高度依赖大规模数据集,而目前公开的红外目标检测数据集规模还比较小,且种类较少,与可见光领域相比差距还很大,难以支持大型模型的学习训练。因此,从红外目标检测的长远发展来看,建立红外目标检测的大规模数据集基准是必不可少的基础性工作,尤其是在自动驾驶、视频监控、军事侦察等重要应用领域。

### 4.2 面向特定领域的跨域知识迁移检测方法

在军事等某些特殊领域,获取红外图像的难度较大,可供训练的数据样本比较有限,采用迁移学习的方法可有效弥补数据的不足。但不同领域的数据通常有不同的域知识,如何实现跨域的知识迁移是迁移学习面临的重要问题。因此,寻找不同数据域之间的某些共同特征,以及不同域知识自适应迁移的方法,将是未来应用迁移学习的关键。

### 4.3 多模态图像融合的红外目标检测方法

多模态图像融合的红外目标检测方法,可有效提高目标检测的准确性和鲁棒性,也是未来的一个重要发展方向。未来可关注以下几个方面:(1)多模态图

像的跨域自适应融合。可见光图像虽然可以为红外图像提供互补信息,但当可见光图像质量较差时,如果盲目的融合,可能对检测产生负面影响。未来可基于 GAN 的方法,或借鉴图像融合领域的一些先进做法,探索多模态图像的自适应融合检测方法。(2)基于 GAN 的伪多模态图像融合检测。多模态图像数据不仅获取成本较高,而且在军事等特殊领域,获取目标的多模态图像难度较大。对此,基于 GAN 的方法生成伪多模态图像,而后进行融合检测,有很大的发展潜力。(3)单分支的融合检测网络。多分支的融合检测网络,难以在移动端部署,单分支的融合检测网络在未来更有发展潜力。

## 4.4 面向终端设备的轻量化红外目标检测模型

在实际应用中,目标检测模型往往需要部署在存储及计算资源受限的边缘终端设备上。目前,模型轻量化研究虽取得一定进展,但检测精度较低,还有很大提升空间。因此,如何在保持较高检测精度的前提下实现红外目标检测模型的轻量化,是未来应用部署的一个重要发展方向。对此,可借鉴可见光领域的轻量化改进方法,例如网络剪枝、参数量化、重参数化方法、知识蒸馏等。

## 5 结束语

红外目标检测可适应弱光、无光、逆光等复杂光照环境,能够有效弥补可见光目标检测的不足,在民用、军事的诸多领域都有非常重要的应用价值。该文对红外目标检测面临的困难和挑战进行了详细分析,并从数据增强、迁移学习、视觉注意力机制、多尺度特征融合、多模态图像融合和轻量化改进等方面,系统分析了基于深度学习的红外目标检测改进方向。在实际任务中,应根据任务需求和红外数据集特点,灵活采用多种方法,以达到模型大小、精度和速度的均衡。针对红外数据集资源获取困难的问题,对现有的公开数据集资源进行了系统梳理汇总。最后,结合当前发展现状和未来实际需求,对基于深度学习的红外目标检测发展方向进行了展望,可为其他研究者提供参考。

### 参考文献:

- [1] GIRSHICK R B, 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. Columbus: IEEE, 2014:580-587.
- [2] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015:1440-1448.
- [3] REN S Q, HE K M, GIRSHICK R B, et al. Faster R-CNN;

- towards real-time object detection with region proposal networks[C]//Conference and workshop on neural information processing systems. Montreal:MIT Press,2015:91-99.
- [4] LIN T Y, DOLLÁR P, GIRSHICK R B, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu:IEEE,2017:936-944.
- [5] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:6154-6162.
- [6] 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. Las Vegas:IEEE,2016:779-788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Honolulu:IEEE,2017:6517-6525.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767,2018.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y. YOLOv4: optimal speed and accuracy of object detection [J]. arXiv:2004.10934,2020.
- [10] JOCHER G. YOLOv5 [EB/OL]. [2022-12-10]. <https://github.com/ultralytics/yolov5>.
- [11] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[J]. arXiv:2107.08430,2021.
- [12] LI C, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv:2209.02976,2022.
- [13] WANG C Y, ALEXEY B, MARK L, et al. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv:2207.02696,2022.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European conference on computer vision. Amsterdam:Springer,2016:21-37.
- [15] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[J]. arXiv:1701.06659,2017.
- [16] 赵晓枫,徐明扬,王聃漂,等.基于改进 SSD 的特种车辆红外伪装检测方法[J].红外与激光工程,2019,48(11):116-125.
- [17] ZHANG X X, ZHU X. Vehicle detection in the aerial infrared images via an improved Yolov3 network [C]//2019 IEEE 4th international conference on signal and image processing. Wuxi:IEEE,2019:372-376.
- [18] 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.
- [19] 吴晗,张志龙,李楚为,等.小样本红外图像的样本扩增与目标检测算法[J].控制理论与应用,2021,38(9):1477-1485.
- [20] 王悦行,吴永国,徐传刚.基于深度迁移学习的红外舰船目标检测算法[J].空天防御,2021,4(4):61-66.
- [21] 李维鹏,杨小冈,李传祥,等.红外目标检测网络改进半监督迁移学习方法[J].红外与激光工程,2021,50(3):243-250.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:7132-7141.
- [23] 李向荣,孙立辉.融合注意力机制的多尺度红外目标检测[J/OL].红外技术:1-9[2022-12-18].<http://kns.cnki.net/kcms/detail/53.1053.tn.20221013.1703.002.html>.
- [24] 代攀,赵旭,李连鹏,等.基于改进 YOLOv5 的复杂背景红外弱小目标检测算法[J].红外技术,2022,44(5):504-512.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of the European conference on computer vision. Munich:Springer,2018:3-19.
- [26] 陈皋,王卫华,林丹丹.基于无预训练卷积神经网络的红外目标检测[J].红外技术,2021,43(4):342-348.
- [27] DU S J, ZHANG P, ZHANG B F, et al. Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4 [J]. IEEE Access, 2021(9):25671-25680.
- [28] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Virtual:IEEE,2021:13713-13722.
- [29] 胡焱,胡皓冰,赵宇航,等.红外热成像中低分辨率行人小目标检测方法[J].红外技术,2022,44(11):1146-1153.
- [30] 杨其利,周炳红,郑伟,等.基于全卷积网络的红外弱小目标检测算法[J].红外技术,2021,43(4):349-356.
- [31] 杨子轩,肖嵩,董文倩,等.一种引入注意力机制的红外目标检测方法[J].西安电子科技大学学报,2022,49(3):28-35.
- [32] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:8759-8768.
- [33] 刘杨帆,曹立华,李宁,等.基于 YOLOv4 的空间红外弱小目标检测[J].液晶与显示,2021,36(4):615-623.
- [34] 林健,张巍巍,张凯,等.基于 YOLOv5 的红外目标检测算法[J].飞控与探测,2022,5(3):63-71.
- [35] 舒朗,张智杰,雷波.一种针对红外目标检测的 Dense-Yolov5 算法研究[J].光学与光电技术,2021,19(1):69-75.
- [36] 朱子健,刘琪,陈红芬,等.基于并行融合网络的航拍红外车辆小目标检测方法[J].光子学报,2022,51(2):190-202.

- [37] 盛大俊,张 强. 基于边缘感知的深度神经网络红外装甲目标检测[J]. 红外技术,2021,43(8):784-791.
- [38] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection [ C ]//Proceedings of European conference on computer vision. Munich: IEEE, 2018:404-419.
- [39] 顾 燕,李 臻,杨 锋,等. 基于改进 Faster R-CNN 的复杂背景红外目标检测算法[J]. 激光与红外,2022,52(4):614-619.
- [40] 蒋昕昊,蔡 伟,杨志勇,等. 基于 YOLO-IDSTD 算法的红外弱小目标检测[J]. 红外与激光工程,2022,51(3):502-511.
- [41] 高 凡,杨小冈,卢瑞涛,等. Anchor-free 轻量级红外目标检测方法[J]. 红外与激光工程,2022,51(4):135-143.
- [42] 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.
- [43] GENG K K, ZOU W, YIN G D, et al. Low-observable targets detection for autonomous vehicles based on dual-modal sensor fusion with deep learning approach[J]. Journal of Automobile Engineering, 2019,233(9):2270-2283.
- [44] LIU J Y, FAN X, HUANG Z B, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection [ C ]//Proceedings of the IEEE conference on computer vision and pattern recognition. New Orleans: IEEE, 2022:5792-5801.
- [45] SUN Y M, CAO B, ZHU P F, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning [ J ]. IEEE Transactions on Circuits and Systems for Video Technology, 2021,32(10):6700-6713.
- [46] 赵 明,张浩然. 一种基于跨域融合网络的红外目标检测方法[J]. 光子学报,2021,50(11):339-349.
- [47] 赵兴科,李明磊,张 弓,等. 基于显著图融合的无人机载热红外图像目标检测方法[J]. 自动化学报,2021,47(9):2120-2131.
- [48] SANDLER M, HOWARD A, ZHU M L, et al. Mobilenetv2: inverted residuals and linear bottlenecks [ C ]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018:4510-4520.
- [49] TAN M X, LE Q. EfficientNet: rethinking model scaling for convolutional neural networks [ C ]//Proceedings of the 36th international conference on machine learning. Long Beach: PMLR, 2019:6105-6114.
- [50] HAN K, WANG Y H, XU C, et al. GhostNets on heterogeneous devices via cheap operations [ J ]. International Journal of Computer Vision, 2022,230(4):1050-1069.
- [51] LIU X F, YANG T, LI J. Real-time ground vehicle detection in aerial infrared imagery based on convolutional neural network [ J ]. Electronics, 2018,7(6):78-86.
- [52] 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.
- [53] WANG C Y, ALEXEY B, LIAO H Y. Scaled-YOLOv4: scaling cross stage partial network [ C ]//Proceedings of the IEEE conference on computer vision and pattern recognition. Virtual: IEEE, 2021:13029-13038.
- [54] DAVIS J W, SHARMA V. Background-subtraction using contour-based fusion of thermal and visible imagery [ J ]. Computer Vision and Image Understanding, 2007,106(2-3):162-182.
- [55] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: a small target detection benchmark [ J ]. Journal of Visual Communication and Image Representation, 2016,34:187-203.
- [56] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: benchmark dataset and baseline [ C ]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015:1037-1045.
- [57] GONZÁLEZ A, FANG Z J, SALAS Y S, et al. Pedestrian detection at day/night time with visible and FIR cameras: a comparison [ J ]. Sensors, 2016,16(6):820.
- [58] ZX A, ENVELOPE J, ENVELOPE Q, et al. Benchmarking a large-scale FIR dataset for on-road pedestrian detection [ J ]. Infrared Physics & Technology, 2019,96:199-208.
- [59] JIA X Y, ZHU C, LI M Z, et al. LLVIP: a visible-infrared paired dataset for low-light vision [ C ]//IEEE international conference on computer vision workshop. Montreal: IEEE, 2021:3489-3497.
- [60] HUI B W, SONG Z Y, FAN H Q, et al. A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background [ DS/OL ]. Science Data Bank, 2019 [ 2022-12-25 ]. <https://cstr.cn/31253.11>. sciencedb. 902. DOI:10.11922/sciencedb.902.