

〈图像处理与仿真〉

## Infrared-PV: 面向监控应用的红外目标检测数据集

陈旭<sup>1</sup>, 吴蔚<sup>2</sup>, 彭冬亮<sup>1</sup>, 谷雨<sup>1</sup>

(1. 杭州电子科技大学 自动化学院, 浙江 杭州 310018; 2. 中国电子科技集团第28研究所, 江苏 南京 210007)

**摘要:** 红外摄像机虽然能够全天候 24 h 工作, 但是相比于可见光摄像机, 其获得的红外图像分辨率和信杂比低, 目标纹理信息缺乏, 因此足够的标记图像和进行模型优化设计对于提高基于深度学习的红外目标检测性能具有重要意义。为解决面向监控应用场景的红外目标检测数据集缺乏的问题, 首先采用红外摄像机采集了不同极性的红外图像, 基于自研图像标注软件实现了 VOC 格式的图像标注任务, 构建了一个包含行人和车辆两类目标的红外图像数据集 (Infrared-PV), 并对数据集中的目标特性进行了统计分析。然后采用主流的基于深度学习的目标检测模型进行了模型训练与测试, 定性和定量分析了 YOLO 系列和 Faster R-CNN 系列等模型对于该数据集的目标检测性能。构建的红外目标数据集共包含图像 2138 张, 场景中目标包含白热、黑热和热力图 3 种模式。当采用各模型进行目标检测性能测试时, Cascade R-CNN 模型性能最优,  $mAP_{0.5}$  值达到了 82.3%, YOLO v5 系列模型能够兼顾实时性和检测精度的平衡, 推理速度达到 175.4 帧/s 的同时  $mAP_{0.5}$  值仅降低 2.7%。构建的红外目标检测数据集能够为基于深度学习的红外图像目标检测模型优化研究提供一定的数据支撑, 同时也可以用于目标的红外特性分析。

**关键词:** 红外图像; 数据集; 监控应用; 深度学习; 基准测试

中图分类号: TP391.9 文献标识码: A 文章编号: 1001-8891(2023)12-1304-10

## Infrared-PV: an Infrared Target Detection Dataset for Surveillance Application

CHEN Xu<sup>1</sup>, WU Wei<sup>2</sup>, PENG Dongliang<sup>1</sup>, GU Yu<sup>1</sup>

(1. School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China;

2. No. 28 Research Institute of China Electronic Technology Group, Nanjing 210007, China)

**Abstract:** Although infrared cameras can operate day and night under all-weather conditions compared with visible cameras, the infrared images obtained by them have low resolution and signal-to-clutter ratio, lack of texture information, so enough labeled images and optimization model design have great influence on improving infrared target detection performance based on deep learning. First, to solve the lack of an infrared target detection dataset used for surveillance applications, an infrared camera was used to capture images with multiple polarities, and an image annotation task that outputted the VOC format was performed using our developed annotation software. An infrared image dataset containing two types of targets, person and vehicle, was constructed and named infrared-PV. The characteristics of the targets in this dataset were statistically analyzed. Second, state-of-the-art target detection models based on deep learning were adopted to perform model training and testing. Target detection performances for this dataset were qualitatively and quantitatively analyzed for the YOLO and Faster R-CNN series detection models. The constructed infrared dataset contained 2138 images, and the targets in this dataset included three types of modes: white hot, black hot, and heat map. In the benchmark test using several models, Cascade R-CNN achieves the best performance, where mean average precision when intersection over union exceeding 0.5

收稿日期: 2021-01-15; 修订日期: 2021-02-24。

作者简介: 陈旭 (1997-), 男, 硕士研究生, 主要从事图像目标识别、检测与模型优化研究。

通信作者: 谷雨 (1982-), 男, 博士, 副教授, 主要从事多源信息融合、遥感图像目标检测与识别方面的研究。E-mail:guyu@edu.hdu.cn。

基金项目: 浙江省自然科学基金资助项目 (LY21F030010)。

(mAP<sub>0.5</sub>) reaches 82.3%, and YOLOv5 model can achieve the tradeoff between real-time performance and detection performance, where inference time achieves 175.4 frames per second and mAP<sub>0.5</sub> drops only 2.7%. The constructed infrared target detection dataset can provide data support for research on infrared image target detection model optimization and can also be used to analyze infrared target characteristics.

**Key words:** infrared image, dataset, surveillance application, deep learning, benchmark test

## 0 引言

相比于可见光摄像机,红外摄像机能够全天时全天候工作,因此在城市交通、军事侦查、视频监控等领域有着广泛的应用<sup>[1-2]</sup>。

虽然红外图像的目标指示性较强,但受传感器探测性能、目标成像距离、目标几何形状等因素影响,红外图像通常分辨率低,缺乏丰富的边缘和纹理信息,在复杂背景下目标信杂比低<sup>[3]</sup>,故早期的红外图像目标检测算法研究主要集中在弱小目标检测方面<sup>[4-5]</sup>。主要方法包括基于单帧图像的跟踪前检测方法(detecting before tracking, DBT)和基于序列图像的检测前跟踪方法(tracking before detecting, TBD)。DBT方法<sup>[6]</sup>主要利用单帧图像中目标-背景的差异信息进行去噪、增强等,通过分割得到目标位置,但是在低信杂比和复杂背景下虚警率较高。TBD方法利用运动目标连续性和相似性进行跟踪检测,可通过目标信息的长时间积累实现目标检测,但是在帧间差异较大时目标检测精度会下降,同时该类算法实时性较差<sup>[2]</sup>。

随着大数据和深度学习理论和技术的快速发展,基于深度卷积神经网络(convolution neural network, CNN)的目标检测模型在可见光图像目标检测与识别方面取得了远超传统方法的性能。典型的目标检测模型可分为两阶段模型和单阶段模型<sup>[7]</sup>。以Faster R-CNN<sup>[8]</sup>为代表的两阶段模型,首先通过区域建议网络筛选候选目标区域,然后在候选目标区域进行目标分类和位置精调。以SSD(single shot multibox detector)<sup>[9]</sup>和YOLO(You Only Look Once)系列<sup>[10]</sup>为代表的单阶段模型,根据输出的特征图进行回归,输出目标位置、置信度和类别等信息。虽然单阶段算法具有实时性好的优势,但检测精度比双阶段算法略低。

将深度学习算法和模型应用于红外图像目标检测时,王文秀等人<sup>[11]</sup>针对传统红外检测算法检测率低、实时性差等问题,采用改进的AlexNet<sup>[12]</sup>深度网络进行红外船只目标识别,可快速准确地识别出红外船只目标,但该方法并不是一种端到端的方法,需通过分水岭方法提取感兴趣船只区域。针对该问题,蒋

志新<sup>[13]</sup>提出了一种基于改进损失函数的Faster R-CNN海上红外小目标检测方法,其是一种端到端的方法,通过样本分类的难易程度来调节权重,在加快模型训练收敛速度的同时提高了检测精度,但该方法实时性较差,不利于工程实际应用。针对红外目标检测准确率和实时性不足的问题,陈铁明等人<sup>[14]</sup>提出了一种基于改进YOLO v3的红外末制导目标检测方法。通过半实物仿真方式获得特种车辆、舰船和飞行器3类目标的红外图像,在自适应学习率与动量法联合优化下,对设计的红外目标数据集进行验证,平均准确率达到了77.89%,检测速度达到25帧/s。针对红外弱小目标检测难题,赵琰等人<sup>[15]</sup>在对红外目标特性进行分析基础上,通过简化YOLO v3网络的处理流程,提升了对红外弱小目标检测精度,实验使用的数据为包含无人机的红外视频序列。吴双忱等人<sup>[16]</sup>将对小目标的检测问题转化为对小目标位置分布的分类问题,检测网络由全卷积网络和分类网络组成,其中分类网络通过引入注意力模块(squeeze-and-excitation networks, SENet)<sup>[17]</sup>,致力于解决复杂背景下低信噪比和存在运动模糊的小目标检测难题,取得了较好的检测效果。李慕锴等人<sup>[18]</sup>将SENet模块引入到YOLOv3骨干网络中的残差模块,提高了骨干网络特征提取能力,实现了复杂场景下红外行人小目标的有效检测,相比原算法实时性不变,误检率显著减低。

基于数据驱动的深度学习方法很大程度上依赖于均衡完备的样本数据,目前公开的数据集以可见光图像为主,主要包括VOC(PASCAL visual object classes)数据集<sup>[19]</sup>、COCO(microsoft common objects in context)数据集<sup>[20]</sup>、DOTA(dataset for object detection in aerial images)数据集<sup>[21]</sup>、DIOR(object detection in optical remote sensing images)<sup>[22]</sup>、UCAS-AOD<sup>[23]</sup>等,针对不同数据集的特点均提出了很多有效的改进模型<sup>[24]</sup>。考虑到红外图像的特点,将应用于可见光图像领域的深度目标检测模型应用于红外图像时,需要做出合理的改进,但首先要解决的就是面向具体应用场景的红外数据集构建问题。

韩国科学技术研究院构建的多光谱行人数据集(KAIST)<sup>[25]</sup>每张图像包含可见光图像和红外图像两个版本,图像分辨率为640×480,分为person、people

和 cyclist 三个类别, 该数据集的主要问题是标签质量较低, 需对数据进行重新标注。菲力尔 (FLIR) 红外数据集<sup>[26]</sup>图片分辨率为  $640 \times 512$ , 只包含水平视角下的街道和高速路场景, 主要面向自动驾驶汽车应用。OTCBVS (object tracking and classification in and beyond the visible spectrum) 红外数据集<sup>[27]</sup>用于测试和评估先进的计算机视觉算法, 包含人脸、车辆、行人和舰船等 14 种不同场景的图像数据, 该数据集中目标相对比较少, 且场景较理想化。上述 3 个数据集的示例图像如图 1 所示。

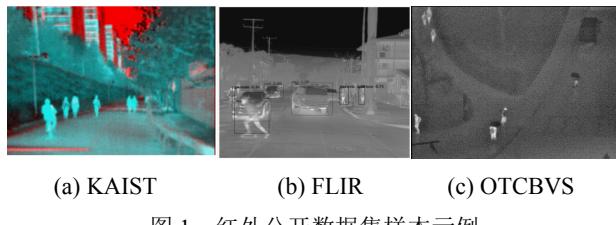


图 1 红外公开数据集样本示例

Fig.1 Samples of infrared public datasets

针对监控领域的实际应用需求, 本文构建了一个包含行人和车辆的红外目标检测数据集, 将其命名为 Infrared-PV。该数据集包含十字路口、横向马路两大主场景, 包含地铁口、广场、公交站台等不同背景区域。采集的红外图像类型包含白热目标、黑热目标和热力图 3 种模式, 总计包含 2138 张图像, 采用自研图像标注软件实现了 VOC 格式的标注任务。同时结合主流的深度目标检测模型给出了在该数据集上的检测性能测试基准。该数据集能够为面向监控应用的红外图像目标检测模型优化研究提供一定的数据支撑。

## 1 Infrared-PV 数据集

### 1.1 数据集构建方式及目标统计信息

使用 FLIR 公司的 Tau2 336 长波非制冷红外机芯配备 25 mm 镜头采集视频数据, 对该机芯极性进行控制, 包括黑热和白热两种模式, 此外还提供热力图模式。采集高度 30 m 左右, 聚焦距离大于 250 m, 拍摄扫描视角范围为  $0^\circ \sim 120^\circ$ , 分别在雨天、晴天、白天、夜晚等环境下采集多个场景的视频图像, 然后通过 3 帧/s 的下采样率得到分辨率为  $720 \times 576$  或  $640 \times 480$  的图像, 构建的数据集总计包含图片 2138 张, 主要考虑场景中的行人和车辆 2 类目标, 故将该数据集命名为 Infrared-PV。不同极性下的样本数量分布如表 1

所示, 目标示例图像如图 2 所示。从图 2 可以看出, 在不同极性下目标的显著性不同, 在黑热极性下, 行人和车辆的轮廓信息更加明显。从热力图模式看, 行人、车辆目标的底部和发动机部分红外热辐射更高。

表 1 Infrared-PV 数据集不同极性图像统计

Table 1 Statistics of the number of images with different polarities in infrared-PV dataset

Images	White-hot target	Black-hot target	Heat map
2138	1000	838	300



(a) 白热目标 (b) 黑热目标 (c) 热力图  
(a) White-hot target (b) Black-hot target (c) Heat map

图 2 不同极性目标图像

Fig.2 Images of targets with different polarities

### 1.2 标注软件及标注信息

使用自研的图像标注软件对构建的 Infrared-PV 数据集进行人工标注, 标注格式为扩展 PASCAL VOC 格式, 保存本地文件为 XML 格式。设计的软件与 Labelme 软件类似, 支持图片、视频和 VOC 格式数据集的标注, 同时实现了对数据集的统计和数据增强功能, 支持目标类别统计、长宽分布统计等, 并可检测标注过程中可能存在的漏标、错标等情况。图 3 为设计的标注软件界面图, 界面上侧为菜单栏, 左侧为待标记图像列表, 右侧为目标信息显示区域。与 Labelme 软件不同, 研制的标注软件支持以下两方面功能: ① 通过修改配置文件支持额外语义信息的添加与保存。其中图 3 右下角区域为语义信息修改和显示区。某一幅图像对应的目标标注内容如表 2 所示, 包括文件名、图像分辨率、目标位置和类别信息, 以及拍摄日期 (date)、拍摄时间 (time)、相机极性 (polarity)、天气信息 (weather) 等语义信息。② 增量标注功能。在已获得少量标注样本的情况下, 通过训练一个基于深度模型的目标检测器, 辅助人工标注, 降低标注工作量。为测试各检测模型的检测性能, 本数据集对部分遮挡度小于 50% 的目标进行了标注。

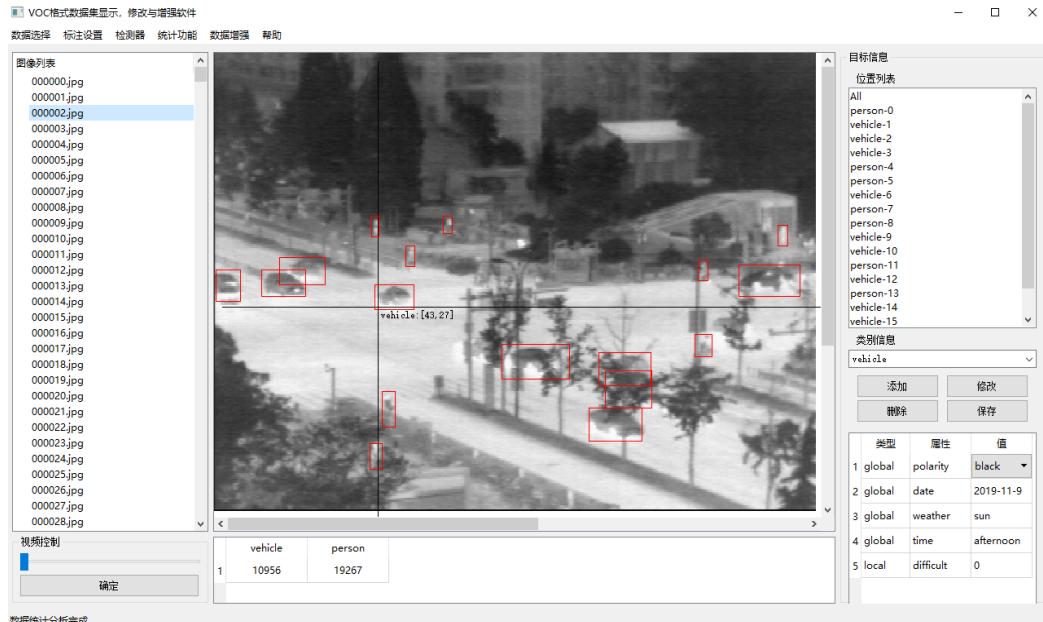


图3 标注软件界面

Fig.3 Interface diagram of labeling software

表2 标注信息

Table 2 Annotation Information

```

<annotation>
<folder>D:/infrared_all/DVD_162025_0/X2</folder>
<filename>img00009754.jpg</filename>
<path>D:/infrared_all/DVD_162025_0/X2/JPEGImages
/img00009754.jpg</path>
<source>
<database>Unknown</database>
</source>
<size>
<width>720</width>
<height>576</height>
<depth>3</depth>
</size>
<segmented>0</segmented>
<polarity>black</polarity>
<date>2019-11-9</date>
<weather>sun</weather>
<time>afternoon</time>
<object>
<name>vehicle</name>
<xmin>328</xmin>
<ymin>402</ymin>
<xmax>488</xmax>
<ymax>521</ymax>
</bndbox>
</object>
</annotation>
<pose>Unspecified</pose>
<truncated>0</truncated>
<difficult>0</difficult>
<bndbox>

```

### 1.3 目标特性统计分析

图4为数据集典型场景的示例图像。从图2和图4可以看出,黑热极性下行人目标的轮廓更加明显,此时目标对比度高于白热极性情况。热力图图像是场景中目标热辐射的伪彩色显示,在该模式下

很难得到目标的轮廓信息。车辆目标包括小轿车、公交车、货车等几种车型,目标尺寸大于行人目标,虽然车辆目标的尺度变化较大,但由于与背景差异显著,故比较容易区分,部分遮挡和目标重叠问题是影响车辆目标检测精度的主要问题。由于是室外场景,数据集中的图像背景比较复杂,包括地铁出入口、交通指示杆和建筑物等其他目标,这会给目标检测算法和模型带来一定挑战,主要体现在虚警方面。

数据集中行人和车辆对应的类别分别为 person 和 vehicle,在整个目标中的占比分别为 60.3% 和 39.7%,如表3所示。平均单张图片包含 20 个左右目标,最多单张图片目标数超过 100 个,说明本红外数据集目标相对比较密集。从图 5(a)和图 5(b)目标尺寸和占比分布可以看出,目标主要分布在  $150 \times 200$  大小范围内,其中 person 类别目标长宽分布在  $10 \times 20$  像素~ $20 \times 50$  像素不等,车辆分布在  $40 \times 23$  像素~ $200 \times 150$  像素不等,集中在整幅图像尺寸的 10% 以下,说明本红外数据集多以中小目标为主。从图 4 可以看出,目标相对比较密集,目标间、背景与目标间存在遮挡情况,因此可以用于评估各检测算法或模型的检测性能和适用性,为红外弱小目标检测、红外行人检测、车辆检测与跟踪等

表3 Infrared-PV 数据集类别统计

Table 3 Category statistics of Infrared-PV dataset

Class	Person	Vehicle
Number	27910	18372

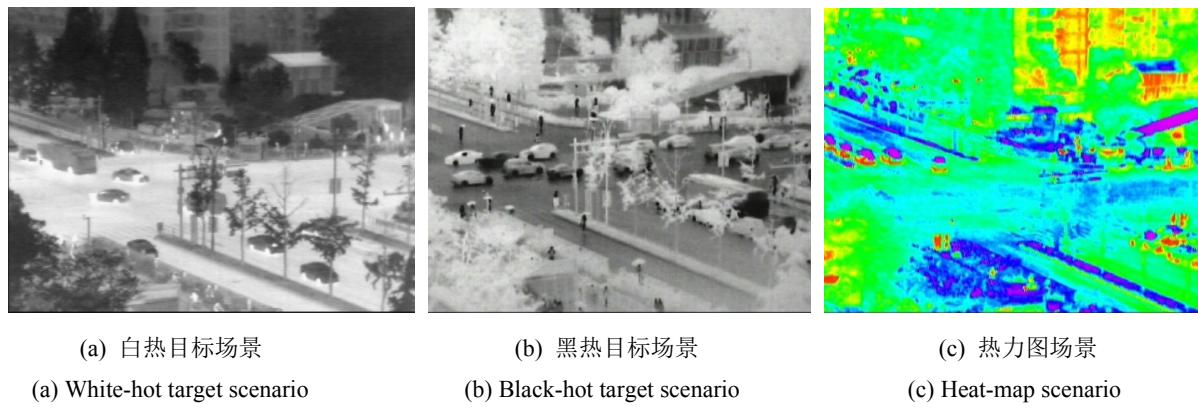


图4 Infrared-PV 数据集图像  
Fig.4 Images in Infrared-PV dataset

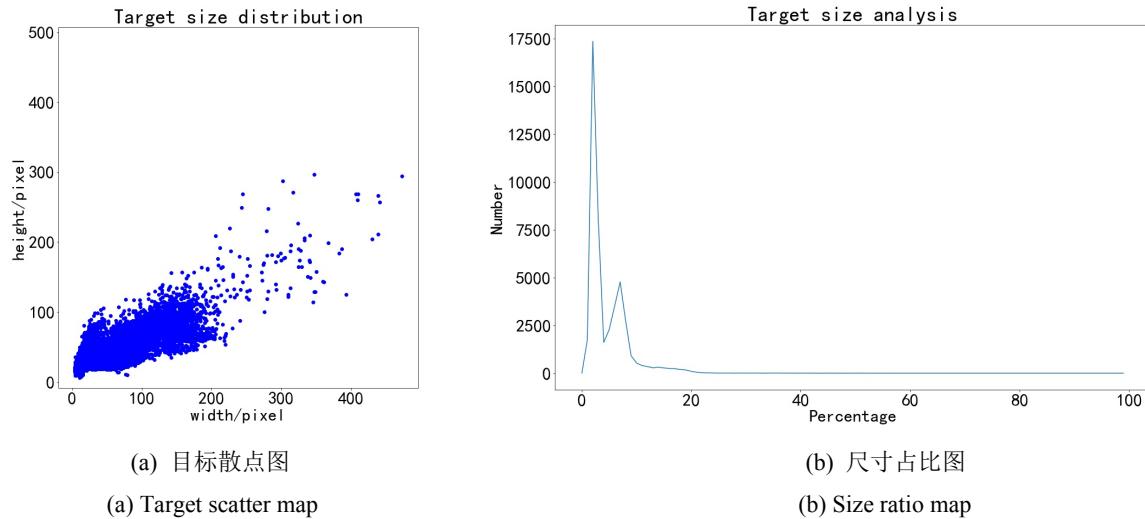


图5 目标统计分析  
Fig.5 Statistical analysis diagrams of targets

研究方向提供一定的数据支撑。Infrared-PV 红外目标检测数据集的获取地址为 <https://pan.baidu.com/s/1j0gqBrtTjI89s2Mdp6xpEA> (提取码: 4 ftz)。

## 2 主流目标检测模型

### 2.1 两阶段目标检测模型

以 Faster R-CNN 为代表的两阶段模型主要由 3 部分组成, 即特征提取网络、候选区域建议网络 (region proposal network, RPN) 网络和分类回归网络。特征提取网络避免了传统手工设计特征的局限性, 能够学习到更有利于目标检测与分类的特征表示。候选区域建议网络用于判别目标和背景区域, 这种端到端模型极大提高了目标检测的速度。分类回归网络对候选目标区域进行类别细分类和位置预测的精调。在 Faster R-CNN 基础上, 研究学者提出了很多改进模型, 包括 Cascade R-CNN<sup>[28]</sup>、Libra R-CNN<sup>[29]</sup>、DoubleHead R-CNN<sup>[30]</sup>等, 这些模型显著地提高了通用目标检测精度。

### 2.2 单阶段目标检测模型

舍弃 RPN 模块, 直接对目标类型和位置进行回归的单阶段模型, 能够实现检测精度和推理速度的平衡, 逐渐成为研究的热点, 其中最具代表性的模型为 SSD、YOLO 系列。其中 YOLOv3<sup>[31]</sup>已被成功应用于红外目标检测, 并取得不错的效果。YOLO v3 是单阶段端到端全卷积网络, 模型组成如下: 以基于 ResNet 残差网络<sup>[32]</sup>的 darknet53 架构作为特征提取网络, 拓展了网络的深度, 降低了训练难度; 采用类似特征金字塔网络 (feature pyramid networks, FPN)<sup>[33]</sup>的思想, 加强了高层语义信息对浅层细节特征的指导; 多尺度输出层根据预设先验框对像素点进行回归、分类、后处理, 提高了中小目标检测率。

基于注意力机制以及优化的特征融合策略, 文献 [34] 提出了 YOLO v4 模型。v4 版本相比于 v3 版本改进如下:

① 特征提取网络设计方面, 使用了加权残差连接、跨阶段局部网络 (cross stages partial network,

CSPNet), 通过加强网络特征融合能力, 设计了更优的特征提取网络 CSPDarknet53;

②检测分支设计方面, 使用了 SPP<sup>[35]</sup>、FPN、PANet<sup>[36]</sup>等特征融合策略, 通过融合不同尺度的语义信息, 提高了中小目标的检测和分类性能;

③数据增强方面, 综合使用了 Cutmix & Mosaic 数据增强、DropBlock 正则化、类标签平滑、Ciou 损失<sup>[37]</sup>、对抗训练等 BoF & BoS (Bags of Freebies & Specials) 手段, 进一步提高了检测精度。

最新的 YOLO v5 网络根据网络深度和宽度不同, 共包含 s、m、l 和 x 四种模型, 其中 s 网络模型结构如图 6 所示。YOLO v5 模型改进了特征提取网络和特征融合模块, 用 CSPNet 模块代替多层混合卷积, 显著降低模型复杂度, 联合 Mosaic 数据增强、自适应图片缩放等手段, 不仅保证了检测精度, 还消除了模型冗余, 大大提高了推理的实时性。其中使用 GTX 1080Ti 显卡的 s 模型推理时间低至 2.5 ms, 在边缘设备上也有很好的检测效果。

### 3 实验结果分析

为评估基于深度学习的红外目标检测性能, 本文在 Infrared-PV 数据集上进行了对比实验, 构建了用于深度目标检测模型评估的指标, 定性和定量分析了各

模型在 Infrared-PV 红外数据集上的性能。

### 3.1 实验准备

采用的服务器配置如下: CPU 为 Intel(R) i7-6850K, 64 G 内存, 两块 NVIDIA GeForce GTX 1080Ti 图形处理器, 操作系统为 Ubuntu 18.04。所有模型使用双卡分布式训练。

Infrared-PV 红外数据集以 7:3 比例随机划分, 其中训练集共 1496 张图像, 测试验证集 642 张图像。

### 3.2 检测模型及训练参数

本文使用的基准算法为 Faster R-CNN, YOLO v3, YOLO v4, YOLO v5 目标检测算法, 此外还和 Cascade R-CNN、FCOS<sup>[38]</sup>模型的检测性能进行了对比分析。训练采用的目标检测框架和配置文件如表 4 所示。当输入网络的图像分辨率为  $608 \times 608$  时, YOLO 系列各模型的锚框大小设置如表 5 所示。

YOLO v3 和 YOLO v4 系列基于 darknet 的 C 语言官方实现, 其中 YOLO v3 模型初始学习率为 0.01, 批大小为 64, 子批大小 (subdivisions) 由显存大小调整, 训练总步长为 12500, 在步长为 5000 和 10000 时退火为前一步长的 10%; YOLO v4 模型初始学习率为 0.00001, 批大小为 8, 退火步长为 18000 和 26000; 其他为默认设置。

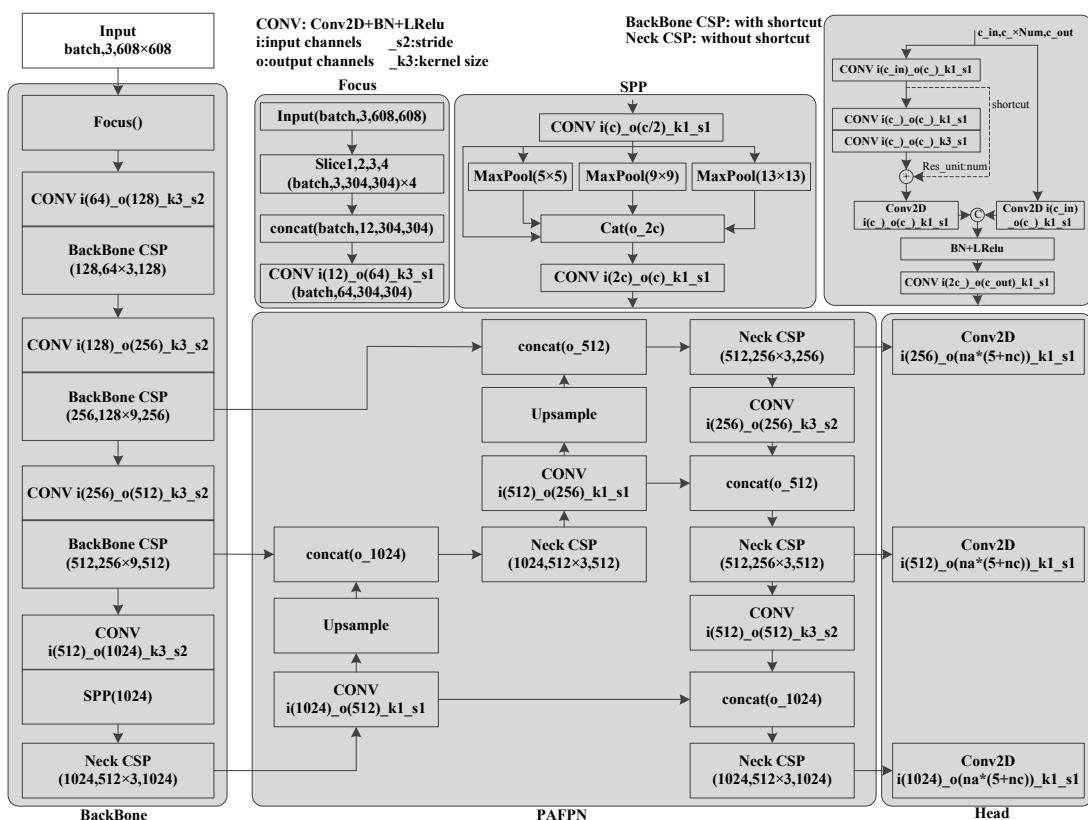


图 6 YOLO v5 模型架构

Fig.6 Models architecture of YOLO v5

表4 模型配置

Table 4 Model configuration

Algorithm	Framework	Configuration file
YOLO v3	Darknet-53	yolov3.cfg
YOLO v4	Darknet-53	yolov4.cfg
YOLO	yolov5	yolov5l.yaml
Faster R-CNN	mm detection	faster_rcnn_r50_fpn_1x.py

表5 608×608分辨率下 Infrared-PV 目标锚框

Table 5 Anchors of targets in Infrared-PV dataset under 608×608 resolution

Scale	Small	Middle	Large
Feature map	76×76	38×38	19×19
Receptive field	165×165	437×437	725×725
Anchors	6,23 50,3 10,25	9,31 12,36 11,87	23,54 67,59 237,179

YOLO v5 系列基于 ultralytics 的 YOLO v5 实现, 训练轮次 (epoch) 为 200, 批大小为 64; 采用 SGD 梯度下降优化器, 动量为 0.9; 初始学习率 0.01, 预热为 3 epoch, 采用余弦退火; 其他为默认设置。

Faster R-CNN、FCOS、Cascade R-CNN 算法基于 mm detection<sup>[39]</sup>实现。其中超参数设置如下: epoch 为 50, 批大小为 8; SGD 优化器, 动量为 0.9; 初

始学习率为 0.02, 在 epoch 为 25 和 38 时退火为前一步长的 33%; 其他为默认设置。

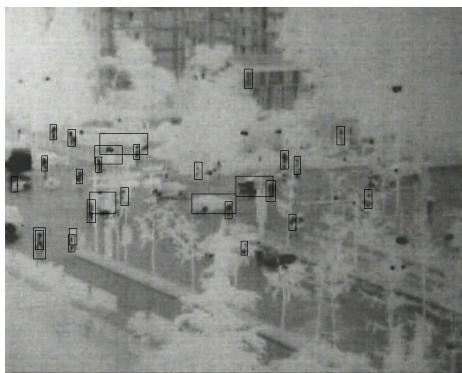
### 3.3 结果分析

对检测结果进行分析, 当目标检测结果与真值的交并比 (intersection over union, IOU) 大于 0.5 认为该目标被检测到, 以单位时间推理图片张数 (frame per second, FPS) 评估模型实时性, 以各类别的平均精度 (average precision, AP)、召回率 (recall) 以及平均精度均值 (mean average precision, mAP) 来评估模型的检测性能。

在基准实验中, 各模型的检测性能如表 6 所示, 检测结果如图 7 所示。由于没有对 Anchor 进行优化, 对行人等小目标检测效果不佳, 故 Faster R-CNN 模型的 mAP<sub>0.5</sub> 只达到了 65.6%, 推理速度仅为 16.08FPS; 当模型输入尺寸为 608×608 时, YOLO v3 模型的检测精度较 416×416 时提升 5.8%。由图 7(c)和(d)可见, 网络输入尺寸的增加有利于小目标检测和位置框的精准回归, 同时降低了虚警率, 但实时性有所降低。YOLO v4 模型的 mAP<sub>0.5</sub> 为 78.97%, YOLO v5 模型的 mAP<sub>0.5</sub> 最高, 达到了 80.7%。从图 7 对比看出, YOLO v4 模型的召回率高于 YOLO v5 模型, 但是虚警率略高。总体而言, 一阶段算法对于遮挡和弱小目标情况检测率相对较高, 虽然存在少量虚警, 但能够满足检测算法的实时性需求。



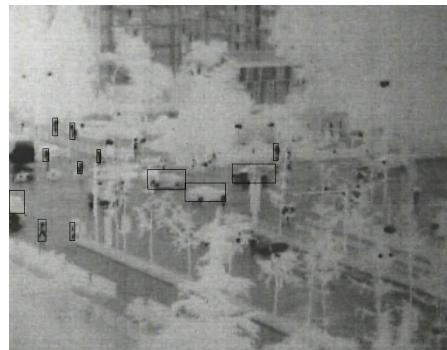
(a) Ground truth



(b) Faster R-CNN



(c) YOLO v3(416×416)



(d) YOLO v3(608×608)

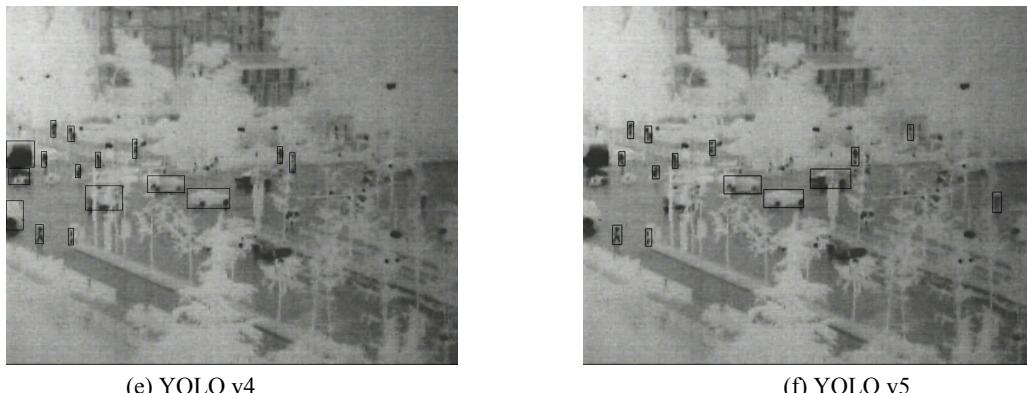


图7 各基准模型检测效果对比 (置信度: 0.25)

Fig.7 Comparison of detection results for benchmark algorithms(confidence: 0.25)

表6 Infrared-PV 数据集各检测模型性能基准 (置信度: 0.05)

Table 6 Performance benchmark of several detection algorithms for Infrared-PV dataset (confidence:0.05)

Algorithm	Rating	Backbone	Size	FPS	AP <sub>0.5</sub> (vehicle) /Recall	AP <sub>0.5</sub> (person) /Recall	mAP <sub>0.5</sub> (%)
Faster R-CNN	5	Resnet50	1333×800	16.08	0.775/0.868	0.536/0.792	65.6
YOLO v3	4	Darknet53	416×416	48.9	0.797/0.8588	0.581/0.806	68.9
YOLO v3	3	Darknet53	608×608	32.8	0.8666/0.9049	0.6278/0.81676	74.7
YOLO v4	2	CSPDarknet53	608×608	29.0	<b>0.87/0.93</b>	0.71/ <b>0.889</b>	78.97
YOLO v5-l	1	-	608×608	<b>99.22</b>	<b>0.882/0.911</b>	<b>0.732/0.853</b>	<b>80.7</b>

注: 加粗字体为该列最优值 Note: Bold font is the optimal value

### 3.4 其他目标检测算法实验

除基准实验外, 本文进行了扩充实验以验证其他模型在 Infrared-PV 数据集上的性能。本文使用先进的骨干网络 ResNest<sup>[40]</sup>替换 Faster R-CNN 和 Cascade R-CNN 模型的主干网络, 进行了对比实验。

各模型的检测指标结果具体如表 7 所示。由表 7 可以看出, 采用 ResNest 的两阶段检测模型相比传统 ResNet 特征提取网络性能上有显著提升, Cascade R-CNN 模型 mAP<sub>0.5</sub> 提升了 15.3%, 达到了最高的检

测精度 82.3%。由图 7(b)可见, 对于遮挡、小目标上的实际效果最佳, 优于所有单阶段算法, 但是该模型实时性较差; 在 YOLOv4 的基础上裁剪模型为原大小的 1/16, 精度提高的同时实时性提高了一倍, 说明针对特定数据集, 尤其是目标类别较少的情况下, 通过裁剪卷积通道数能够在不降低目标检测率的情况下, 提高模型推理速度; FCOS 模型对于红外目标检测结果差强人意, 主要是因为红外图像纹理信息缺失, 边缘信息不明显, 像素框定位目标困难, 故 mAP<sub>0.5</sub> 较低。

表7 其他模型在红外数据集的测试结果 (置信度:0.05)

Table 7 Test results of other models on Infrared-PV datasets (confidence: 0.05)

Algorithm	Backbone	Size	FPS	AP <sub>0.5</sub> (vehicle) /Recall	AP <sub>0.5</sub> (person) /Recall	mAP <sub>0.5</sub> /%
Cascade R-CNN	ResNet50	1333×800	11.80	0.788/0.870	0.552/0.803	67.0
Cascade R-CNN	ResNest50	1333×800	10.18	<b>0.901/0.944</b>	<b>0.744/0.914</b>	<b>82.3</b>
Faster R-CNN	ResNet50	1333×800	16.08	0.775/0.868	0.536/0.792	65.6
Faster R-CNN	ResNet50	1333×800	15.09	0.888/0.935	0.700/0.884	79.4
YOLOv5s	-	608×608	<b>175.4</b>	0.87/0.899	0.722/0.854	79.6
YOLOv5m	-	608×608	125.36	0.889/0.921	0.728/0.849	80.0
YOLOv5x	-	608×608	37.32	0.8754/0.9029	0.7297/0.8541	80.3
YOLOv4-kd	CSPDarknet53	608×608	66.7	0.882/0.9433	0.7380/0.8987	81.0
FCOS	ResNet50	1333×800	18.79	0.755/0.893	0.494/0.787	62.4

注: 加粗字体为该列最优值 Note: Bold font is the optimal value

## 4 结论

本文构建了一个包含行人和车辆2类目标的红外图像检测数据集,对于采用的红外视频图像采用自研的标注软件进行了目标标注,然后对数据集的各项统计信息进行了总结分析。结合目前主流基于深度学习的目标检测模型,给出了各模型在该数据集的检测性能。该数据集的构建能够为视频监控场景下的红外目标检测,以及红外图像目标特性分析提供一定的数据支撑。

## 参考文献:

- [1] 陈钱,隋修宝. 红外图像处理理论与技术[M]. 北京:电子工业出版社,2018.
- CHEN Qian, SUI Xiubao. *Infrared Image Processing Theory and Technology*[M]. Beijing: Electronic Industry Press, 2018.
- [2] 刘让,王德江,贾平,等. 红外图像弱小目标探测技术综述[J]. 激光与光电子学进展, 2016, **53**(5): 050004.
- LIU Rang, WANG Dejiang, JIA Ping, et al. Overview of detection technology for small and dim targets in infrared images[J]. *Progress in Laser and Optoelectronics*, 2016, **53**(5): 050004.
- [3] 武斌. 红外弱小目标检测技术研究[D]. 西安: 西安电子科技大学, 2009.
- WU Bing. Research on Infrared Dim Target Detection Technology[D]. Xi'an: Xidian University, 2009.
- [4] Rawat S S, Verma S K, Kumar Y. Review on recent development in infrared small target detection algorithms[J]. *Procedia Computer Science*, 2020, **167**: 2496-2505.
- [5] 李俊宏,张萍,王晓玮,等. 红外弱小目标检测算法综述[J]. 中国图象图形学报, 2020, **25**(9): 1739-1753.
- LI Junhong, ZHANG Ping, WANG Xiaowei, et al. Infrared small-target detection algorithms: a survey[J]. *Journal of Image and Graphics*, 2020, **25**(9): 1739-1753.
- [6] 谷雨,刘俊,沈宏海,等. 基于改进多尺度分形特征的红外图像弱小目标检测[J]. 光学精密工程, 2020, **28**(6): 1375-1386.
- GU Yu, LIU Jun, SHEN Honghai, et al. Infrared image dim target detection based on improved multi-scale fractal features[J]. *Optics and Precision Engineering*, 2020, **28**(6): 1375-1386.
- [7] LIU L, OUYANG W, WANG X G, et al. Deep learning for generic object detection: a survey[J]. *International Journal of Computer Vision*, 2020, **128**(2): 261-318.
- [8] REN S Q, HE K M, 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.
- [9] LIU W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector[C]//*European Conference on Computer Vision*, 2016: 21-37.
- [10] 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.
- [11] 王文秀,傅雨田,董峰,等. 基于深度卷积神经网络的红外船只目标检测方法[J]. 光学学报, 2018, **38**(7): 0712006.
- WANG W X, FU Y T, DONG F, et al. Infrared ship target detection method based on deep convolutional neural network[J]. *Acta Optics*, 2018, **38**(7): 0712006.
- [12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, **60**(6): 84-90.
- [13] 蒋志新. 基于深度学习的海上红外小目标检测方法研究[D]. 大连: 大连海事大学, 2019.
- JIANG Z X. Research on the Detection Method of Marine Infrared Small Target Based on Deep Learning[D]. Dalian: Dalian Maritime University, 2019.
- [14] 陈铁明,付光远,李诗怡,等. 基于YOLO v3的红外末制导典型目标检测[J]. 激光与光电子学进展, 2019, **56**(16): 155-162.
- CHEN T M, FU G Y, LI S Y, et al. Infrared terminal guidance typical target detection based on YOLOv3[J]. *Progress in Laser and Optoelectronics*, 2019, **56**(16): 155-162.
- [15] 赵琰,刘荻,赵凌君. 基于Yolo v3的复杂环境红外弱小目标检测[J]. 航空兵器, 2020, **26**(6): 29-34.
- ZHAO Y, LIU D, ZHAO L J. Infrared small target detection in complex environment based on Yolo v3[J]. *Aviation Weaponry*, 2020, **26**(6): 29-34.
- [16] 吴双忱,左峥嵘. 基于深度卷积神经网络的红外小目标检测[J]. 红外与毫米波学报, 2019, **38**(3): 371-380.
- WU S C, ZUO Z G. Infrared small target detection based on deep convolutional neural network[J]. *Journal of Infrared and Millimeter Waves*, 2019, **38**(3): 371-380.
- [17] 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.
- [18] 李慕锴,张涛,崔文楠. 基于Yolo v3的红外行人小目标检测技术研究[J]. 红外技术, 2020, **42**(2): 176-181.
- LI M K, ZHANG T, CUI W N. Research on infrared pedestrian small target detection technology based on Yolo v3[J]. *Infrared Technology*, 2020, **42**(2): 176-181.
- [19] Everingham M, Eslami S A, Van Gool L, et al. The pascal visual object classes challenge: a retrospective[J]. *International Journal of Computer Vision*, 2015, **111**(1): 98-136.
- [20] LIN T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C]//*European Conference on Computer Vision*, 2014:

- 740-755.
- [21] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 3974-3983.
- [22] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: a survey and a new benchmark[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, **159**: 296-307.
- [23] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//*2015 IEEE International Conference on Image Processing (ICIP)*, 2015: 3735-3739.
- [24] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10781-10790.
- [25] 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*, 2015: 1037-1045.
- [26] Teledyne FLIR Systems. FLIR ADAS Dataset[DB/OL] [2023-11-27]. <https://www.flir.com/oem/adas/adas-dataset-form/>.
- [27] Davis J W, Keck M A. A two-stage template approach to person detection in thermal imagery[C]//*2005 Seventh IEEE Workshops on Applications of Computer Vision*, 2005, **1**: 364-369.
- [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] PANG J, CHEN K, SHI J, et al. Libra r-cnn: Towards balanced learning for object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 821-830.
- [30] WU Y, CHEN Y, YUAN L, et al. Rethinking classification and localization for object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10186-10195.
- [31] Redmon J, Farhadi A. Yolov3: an incremental improvement [EB/OL] [2018-04-08]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [32] 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.
- [33] LINT 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.
- [34] Bochkovskiy A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL] [2020-04-22]. <https://arxiv.org/pdf/2004.10934.pdf>.
- [35] 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.
- [36] WANG K, LIEW J H, ZOU Y, et al. PaNet: Few-shot image semantic segmentation with prototype alignment[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2019: 9197-9206.
- [37] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//*Proceedings of 2020 Association for the Advancement of Artificial Intelligence*, 2020: 12993-13000.
- [38] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]//*Proceedings of the 2019 IEEE International Conference on Computer Vision*, 2019: 9627-9636.
- [39] CHEN K, WANG J Q, PANG J M, et al. Mmdetection: open mmlab detection toolbox and benchmark[EB/OL][2019-06-17]. <https://arxiv.org/pdf/1906.07155.pdf>.
- [40] ZHANG H, WU C R, ZHANG Z Y, et al. Resnest: Split-attention networks[EB/OL] [2020-04-19]. <https://arxiv.org/pdf/2004.08955.pdf>.