

文章编号: 2095-4980(2022)08-0810-07

基于 Focal-EIOU 函数的被动式太赫兹图像违禁物品识别

周 敏

(中国铁路设计集团有限公司 电化电信院, 天津 300308)

摘 要: 针对被动式太赫兹安检系统因环境影响导致图像质量波动, 从而影响识别算法, 导致准确率大幅降低的问题, 提出了基于 Focal-EIOU 损失函数的改进 YOLOv4 算法, 并用被动式太赫兹人体安检图像对刀、枪违禁物品进行模型训练获得模型。建立不同环境、不同位置角度携带刀枪嫌疑人人员的太赫兹图像数据库, 采用图像增广的方法构建丰富数据集; 将 YOLOv4 的 CIUO loss 改进为 Focal-EIOU loss, 提高算法对太赫兹图像识别的鲁棒性, 进而经过训练获得较优的模型。在本文的测试集中, 使用改进后的算法训练的模型平均检测精确度(mAP)达到 96.4%, 检测速度在 28 ms 左右, 交并比(IOU)平均值为 0.95, 在同等条件下高于常规算法, 改善了检测识别的效果。实验结果表明, 本文方法能够有效提高被动式太赫兹人体安检系统的嫌疑物识别准确率, 有利于该项技术在人体安检领域的推广应用。

关键词: 太赫兹成像; 人体安检; 目标识别; 损失函数

中图分类号: TN29; TP391.41

文献标志码: A

doi: 10.11805/TKYDA2021269

Detection of prohibited objects in passive terahertz images based on Focal-EIOU loss function

ZHOU Min

(Electrification & Telegraphy Engineering Design Research Department, China Railway Design Corporation, Tianjin 300308, China)

Abstract: Aiming at the problem that the image quality fluctuation of passive terahertz security system is caused by environmental change, which affects the recognition algorithm and leads to a significant decrease in accuracy, this paper proposes an improved YOLOv4 algorithm based on Focal-Efficient Intersection Over Union(EIOU) loss function, and uses passive terahertz human security image to conduct model training for prohibited items of knife and gun. A terahertz image database of people carrying suspected objects in different environments and different positions is established, and a rich data set is constructed by image augmentation method. The Complete IOU(CIUO) loss of YOLOv4 is improved to Focal-EIOU loss to improve the robustness of the algorithm for terahertz image recognition, and then a better model is obtained after training. In the test set of this paper, since YOLOv4 algorithm has low robustness for terahertz image recognition accuracy, CIUO loss of YOLOv4 is modified and adjusted to Focal-EIOU loss, and a better model is finally obtained through training. The mean Average Precision (mAP) of the model trained by the improved algorithm reaches 96.4%, the detection speed is about 28 ms, and the average value of IOU is 0.95, which are higher than those of the conventional algorithms under the same conditions, the detection and recognition effect are improved. The experimental results show that the proposed method can effectively improve the suspect identification accuracy of passive terahertz human security system, which is conducive to the popularization and application of this technology in the field of human security.

Keywords: terahertz imaging; human security; object identification; loss function

近年来, 全球在公共交通人群聚集处的暴恐事件频发, 造成大量人员伤亡和经济损失, 社会影响极其恶劣,

收稿日期: 2021-07-02; 修回日期: 2021-09-22

基金项目: 中国铁路设计集团有限公司科技开发课题资助项目(2020YY240802)

因此安检变得越来越严格。传统安检手段在我国这种大客流量的情景下经常造成拥堵，特别在当前新冠病毒肆虐的时期，民众对这种拥堵的负面反映强烈，在此背景下被动式太赫兹安检因其具有无感知检测、能够检测到衣物遮挡下的物品、检测速度快等特点^[1]，在众多公司科研机构的努力下逐渐在合肥、上海等城市的轨道交通中试用。因太赫兹成像颜色单调且成像效果模糊，并且受环境影响会导致图像质量波动，因此对图像识别提出很大挑战。对标注框的位置偏移、漏标情况或随着温度升高导致图像噪声增加，安检人员能够看到嫌疑物形状，但识别系统并未识别检测到，可能会造成漏检，发生安全隐患。随着卷积神经网络的发展，在太赫兹图像识别领域也有很多突破，其中常用的算法有 Faster R-CNN, YOLO 系列等^[2-5]。2021 年王葛、朱艳等老师提出使用主要基于非极大值抑制 Faster R-CNN 算法进行图像识别^[6]，能够提高黏连图像的处理精确度，整体识别准确率较高，但算法空间和时间资源消耗太大，预测框和真实框偏移较多，对实时性要求高的场景不太适用；2019 年余亚萍提出使用基于数据处理、特征层面的数据增强^[7]，丰富了枪、火药等少数样本数，然后使用 YOLOv3 进行图像识别，不过其速度和识别准确率相对较低一些，对于安检行业的需求满足度不是太高。为进一步优化提高识别准确率和模型的鲁棒性，本文在 YOLOv4 的基础上通过数据集增广、损失函数优化等方法，对目标识别算法进行进一步的调整和改进，从而改善目前被动式太赫兹安检图像的可疑物识别问题。

1 实验环境和算法

1.1 实验框架介绍

本实验采用 PyTorch 1.2 深度学习框架^[8]，Python 3.6.13 编译语言，使用 NVIDIA GeForce GTX 1660 的 GPU，并行计算平台架构采用 CUDA 10.0 等搭建深度学习的开发环境并实现本文算法；硬件成像设备采用被动式太赫兹安检仪(如图 1 所示)，该安检仪成像速度为可选的 8,10,12,16 帧/s，检测距离为 2~3 m，尺寸为 0.85 m×1 m×1.5 m，质量为 200 kg，工作温度为 -20~40 °C，该安检仪可在 0.1 s 内完成一次全身扫描，能够实现无停留安检，有效提升检查效率。

1.2 数据采集

本实验中人体太赫兹图像为 5 月~12 月内在被动式太赫兹成像系统采集而来，实验人员身高 178 cm，体重 76 kg，枪械为等比例的生铁模型，长度约为 15 cm 左右，刀具为常见的金属材质的水果刀、菜刀、砍刀(如图 2 所示)，枪械刀具均有多种型号，实验人员将在上身、裤子口袋中携带这些嫌疑物品，并在太赫兹成像系统前做如下动作：随意走动、模仿放包、双手抬起接受安检。被动式太赫兹系统将采用用户数据报协议(User Datagram Protocol, UDP)的方式传送图像数据，编写程序接收储存到图像数据库中。

1.3 数据处理

将采集的数据按照图 3 所示步骤进行处理：

从图像数据库中读取图像进行人工筛选，删除没有人像的图。通常情况下，数据集过小，会导致模型效果不佳，且消耗大量人力采集图像也难以涵盖所有场景。因此为了增加样本的丰富度，提高训练结果的精确度，本实验设计了两种对太赫兹图像的数据增广方法：嫌疑物区域覆盖方法增广、采集的原始图像数据增广。

1.3.1 嫌疑物区域覆盖方法增广

采集不同类型号规格的刀、枪太赫兹嫌疑物图像，并将嫌疑物裁剪，每种规格各取两张做成嫌疑物数据库，采集不同角度的人像图片，人工将人像前景图截取出制作成人像数据库，从嫌疑物、人像数据库中随机选取图像，对嫌疑物图像做角度旋转、镜像操作，对人像图片进行镜像操作^[9]，所有操作更贴近现实情况，对处理后的嫌疑物图像和人像图进行随机覆盖融合，将生成的新图像整理为融合图像数据集，如图 4 所示。

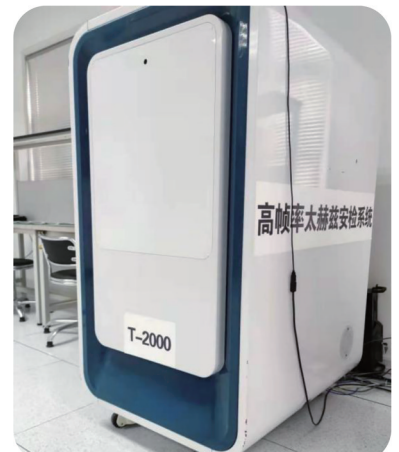


Fig.1 Passive terahertz security detector
图 1 被动式太赫兹安检仪图



Fig.2 Suspected articles for experiment
图 2 部分实验用嫌疑物图

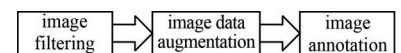


Fig.3 Dataset processing steps
图 3 数据集处理步骤图

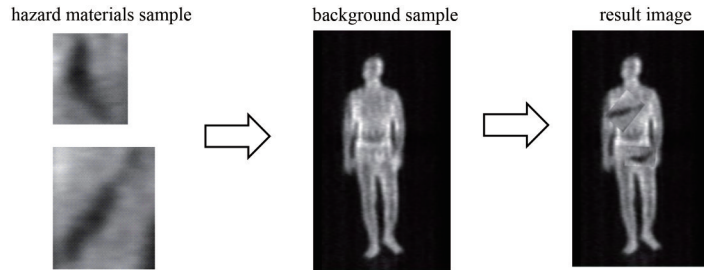


Fig.4 Suspected articles enlargement method
图4 嫌疑物增广方法图

1.3.2 采集的原始图像数据增广

为进一步丰富数据，本实验还采取了第二种增广方法，直接对原始图像进行特定的处理：a) 随机裁剪，将裁剪到的结果再改变大小，筛选出含有嫌疑物的图像；b) 将原始图像水平平移，同样筛选出包含嫌疑物的图像，将上述筛选出的图像放进数据集。

1.3.3 数据标注

数据标注采用 Labelme 软件人工标注的方式，将标注好的数据再转成 VOC 数据集的格式。

1.4 算法描述

1.4.1 YOLOv4 网络结构

YOLO 网络是一种利用回归网络实现目标检测的分类算法，算法的目标检测器主要包括 3 部分：主干网络 (Backbone)、为了更好地利用提取的特征所使用的颈部 (YOLO Neck)、预测对象类别边界的头部 (Head)。相比传统的区域候选网络，YOLO 拥有更快的检测速度，并在检测任务中表现出良好的效果，这些成果与算法的主干网络有很大关系，YOLOv4 的网络结构如图 5 所示。

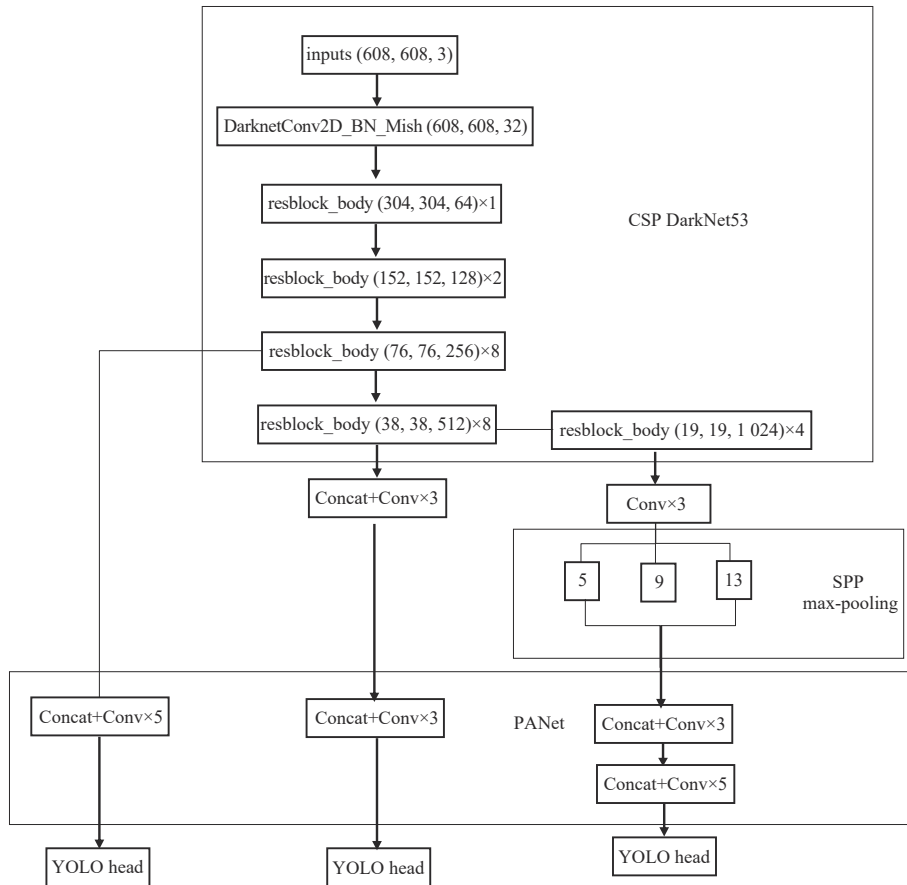


Fig.5 Diagram of YOLOv4 network structure
图5 YOLOv4网络结构图

YOLOv4 对网络进行了很多改进：算法将 DarknetConv2D 的激活函数由 LeakyReLU 修改为 Mish，卷积块也由 Darknet Conv2D_BN_Leaky 修改成了 DarknetConv2D_BN_Mish，并且将 Resblock_body 的结构进行修改，使用了 CSPNet 结构，于是 Darknet53 被修改成了 CSPDarknet53，主干网络包含 5 个 CSP 模块，各模块前的卷积核大小为 3×3 ，步幅为 2，该结构能够起到下采样的作用，进一步增强网络学习能力；对于特征金字塔的部分，YOLOv4 使用了空间金字塔(SPP)附加模块和路径聚合网络(PANet)的结构，其中 SPP 结构可以增加感受野^[10]，能够很好地分离出上下文特征，而 PANet 能够准确地保存空间信息进行实例分割，有助于正确定位像素点；算法还采用了 1×1 ， 5×5 ， 9×9 ， 13×13 最大池化方式，能够增加感受区并分离出更重要的上下文特征；YOLOv4 网络头部沿用了 YOLOv3，网络提取 3 个特征层进行目标检测，分别位于中间层、中下层、底层，提取出的特征供头部进行预测。

1.4.2 损失函数

深度学习算法的损失函数的好坏直接影响训练出的模型性能，YOLOv4 的损失分为 3 个部分：类别损失(class_loss)、置信度损失(confidence_loss)、位置损失(location_loss)。算法的最终损失函数处理方法为：当在某个边界框内不存在目标，则仅计算置信度损失；若存在目标，最终算法的损失函数为它们之和；较上一版本的损失函数算法，YOLOv4 对位置损失改进成 CIOU^[11]，其计算公式如下：

$$Loss_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + av \quad (1)$$

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

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

$$a = \frac{v}{(v - IOU) + v} \quad (4)$$

$$\begin{cases} \frac{\partial v}{\partial w} = \frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{h}{w^2 + h^2} \\ \frac{\partial v}{\partial h} = -\frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{w}{w^2 + h^2} \end{cases} \quad (5)$$

式中： A, B 分别为锚框和真实框； IOU 为锚框和真实框的交并比，即重叠面积的比例，重叠越大，说明锚框越接近真实框，但当锚框距离真实框较远没有重合时，或锚框和真实框完全重合时， IOU_{loss} 不能给出优劣评价； $\rho(b, b^{gt})$ 为预测框和真实框的中心点的欧式距离， b 表示预测框的中心点， b^{gt} 表示真实框的中心点； c 为能够包含预测框和真实框的最小闭包区域的对角线距离； $\frac{\rho^2(b, b^{gt})}{c^2}$ 优化了 IOU 的问题，但若锚框和真实框中心点重合，便无法处理；参数 a 用于平衡比例； v 为描述预测框和真实框的长宽比的比例一致性参数，当中心点重合时，可以通过长宽比进一步评价锚框和真实框的接近情况。式(1)~(4)为 YOLOv4 的 CIOU 完整表达式，它考虑了重叠面积、中心点距离、宽高比 3 个几何因素，但 CIOU 还不够完美，它的缺陷如下：a) 当 $\{(w = kw^{gt}, h = kh^{gt}) | k \in R^+\}$ 时，会造成 $v=0$ ，loss 函数无法做出有效评判；b) 从式(5)可以得出 $\frac{\partial v}{\partial w} = -\frac{h}{w} \times \frac{\partial v}{\partial h}$ ，则锚框的长宽不能同增同减，这会导致预测框无法很好地贴合真实框；c) 使用长宽比例无法有效反映锚框与真实框的差距，如 $w^{gt}=1, h^{gt}=1$ ，经过多次训练后得到 $w=1.64, h=2.84$ ，长宽比例接近，但和实际框还是有一定差距。

为了提高模型性能和精确度，本实验采用 Focal-EIOU^[12]来替代 CIOU，Focal-EIOU 的计算公式如下：

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (6)$$

$$L_{f(x)} = \begin{cases} -\frac{\alpha x^2 [2 \ln(\beta x) - 1]}{4}, & 0 < x \leq 1; 1/e \leq \beta \leq 1 \\ -\alpha \ln(\beta) x + C, & x > 1; 1/e \leq \beta \leq 1 \end{cases} \quad (7)$$

$$L_{Focal-EIOU} = IOU^\gamma L_{EIOU} \quad (8)$$

式中： x 代表真实值与预测值的差值； e 为自然常数； β 用于控制曲线的弧度； C 为一个常数； γ 为控制异常值抑制程度的参数。

EIOU 对 CIOU 存在的缺陷进行了改进，将损失函数分为三块：距离损失、方向损失、IOU 损失，将 CIOU 的

α 和 ν 进行了修改, c_w 和 c_h 代表着覆盖真实框、锚框的最小框的长度和高度, 这样便解决了由 CIOU 使用长宽比例造成的问题。

由于太赫兹安检机会受周围环境因素影响, 导致出图质量波动。当出图噪声多、轮廓不太清晰时, 一般的深度学习模型会出现误报漏报的情况, 这是由于太赫兹人体安检图像背景较多, 前景和背景内容失衡, 会在训练时出现很多易分候选框和负例的现象, 从而造成正负例难易样本不均衡, 由此训练出来的模型能够识别易分样本, 但是难分样本依旧识别效果很差。因此, 为更好地提高模型性能, 使用 FocalL1 loss 设置不同梯度, 如式(7)所示, 将错误率大的地方设置更高梯度, 从而更关注难样本的识别, 可以降低低质量样本对模型性能的影响。通过整合 EIOU loss 和 FocalL1 loss, 得到最终的 Focal-EIOU loss。式(8)为 Focal-EIOU loss 完整的表达式。说明 Focal-EIOU loss 对被动式太赫兹安检场景和 YOLO 算法都能很好匹配对应, 使用该算法改进 YOLOv4 的损失函数会提高训练出的目标识别模型准确率, 精确度更高, 鲁棒性更好。

1.4.3 评价指标

在目标检测领域为了评估一个模型的效果, 会计算查准率(Precision, P)、查全率(Recall, R), 计算公式如下:

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN} \quad (9)$$

式中: TP 为正确检测分类的嫌疑物数量; FP 为错误检测的嫌疑物数量; FN 为漏检的嫌疑物的数量。

查准率为实际预测正确的样本数占预测为正确的总数量的比例, 查全率为实际预测正确的样本数占正样本总数的比例。为了让模型识别更准确、全面, 通常会绘制出查准率-查全率($P-R$)曲线图^[13], 观察平衡点(查准率等于查全率)的位置可用于衡量目标检测算法的性能, 曲线平衡点位于右上方的模型通常来说相对较优。另一种模型评价指标为检测精确度(Average Precision, AP), 是指 $P-R$ 曲线下方的区域面积, 因为单纯用 Precision 或 Recall 作为模型评优的标准都不科学, 因此用 $P-R$ 曲线下方面积做评价标准; mAP 是指模型对多个类, 即刀、枪嫌疑物识别的 AP 的平均值。

为了对目标检测模型的预测框的评价更直观, 使用 IOU 方法, 对测试集进行测试时, 将所有正样本 IOU 取平均, 模型判断为 FN 时取 0, 本文将 PR 图和 mAP、IOU 平均值作为最终的模型结果评价指标。

2 实验结果

本文数据集中共包含了 5 000 张图像, 含有刀、枪嫌疑物的图像各 2 250 张, 剩余 500 张为未含有金属物品或含有手机、金属拉链等物品。随机从数据集中挑选出刀、枪各 1 800 张, 其他图像数据共 400 张, 剩余 1 000 张图像作为测试集, 针对同样的 VOC2007 数据集, 在同等环境下分别对改进的 YOLOv4 算法、YOLOv4 算法、YOLOv3 算法、Faster R-CNN 进行训练调整, 完成训练后用测试集数据, 检测效果如图 6 所示。改进的 YOLOv4 算法训练时比原框架训练收敛速度更快, 比 YOLOv4 算法减少 4 h 的训练时间。本实验采用 CPU 为 Interl i5, 内存为 16 G, GPU 为 NVIDIA GeForce GTX 1660, 加载模型运行会占用 4 G 内存, 模型处理每张 377×754 的太赫兹图像的速度在 28 ms 左右。太赫兹安检仪出图速度设定在每秒 10 帧, 这个速度人眼观察未出现延迟感, 改进的 YOLOv4 算法单图处理时间小于 100 ms, 因此并不会造成延迟, 满足安检的实时性要求。依据测试集图像数据对不同算法训练出的模型进行训练, 再计算不同类别 AP 值求平均得到 mAP, 计算预测框和真实框 IOU 的平均值, 由此得到表 1、表 2 的结果。

表 1 模型结果值对比表

experimental model	Knife AP/%	Gun AP/%	maximum a posteriori/%
improved YOLOv4	96.5	96.3	96.4
YOLOv4	93.3	95.7	94.5
YOLOv3	92.2	94.3	93.25
Faster R-CNN	94.0	94.5	94.25

表 2 模型 IOU 平均值表

experimental model	average IOU values
improved YOLOv4	0.95
YOLOv4	0.82
YOLOv3	0.76
Faster R-CNN	0.84

从模型的结果对比表来看, 对于刀嫌疑物的识别, 改进的 YOLOv4 算法高于 YOLOv4 算法 3.2 个百分点, 高于 YOLOv3 4.3 个百分点, 高于 Faster R-CNN 2.5 个百分点; 对于枪嫌疑物识别, 改进的 YOLOv4 算法高于 YOLOv4 0.6 个百分点, 高于 YOLOv3 2.0 个百分点, 高于 Faster R-CNN 1.8 个百分点; 整体上, 改进的 YOLOv4 算法高于 YOLOv4 算法平均检测精确度 1.9 个百分点, 高于 YOLOv3 3.15 个百分点, 高于 Faster R-CNN 2.15 个百分点。图 7 为改进的 YOLOv4, YOLOv4, YOLOv3, Faster R-CNN 的训练模型对刀、枪嫌疑物检测的 $P-R$ 曲线图, 4 种曲线值均在 90% 以上, 改进的 YOLOv4 算法模型位于其他 3 种曲线的上方, 表 2 中数据显示改进后 YOLOv4 算法的模型比其他模型 IOU 平均值更大, 因此模型的预测框和真实框更加贴近, 改进的 YOLOv4 算法较优。

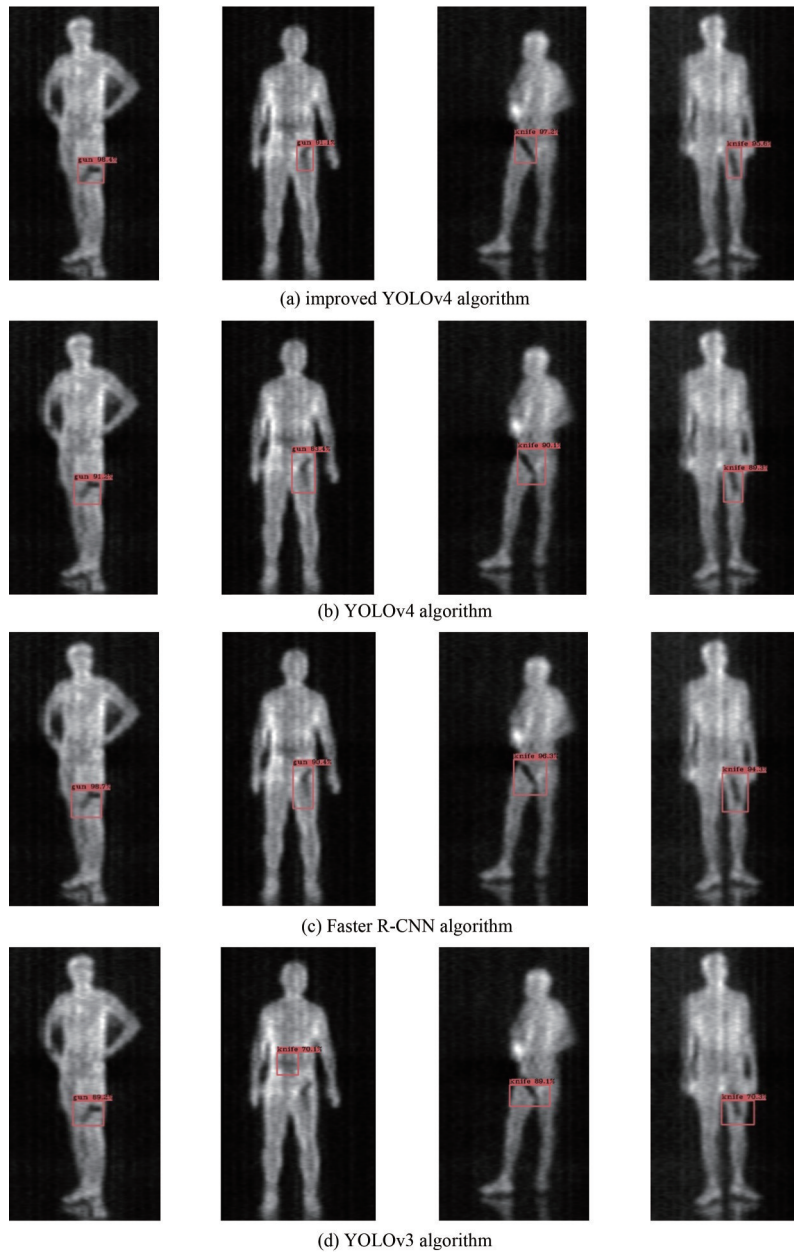


Fig.6 Target detection effect of different models

图 6 不同模型的目标检测效果

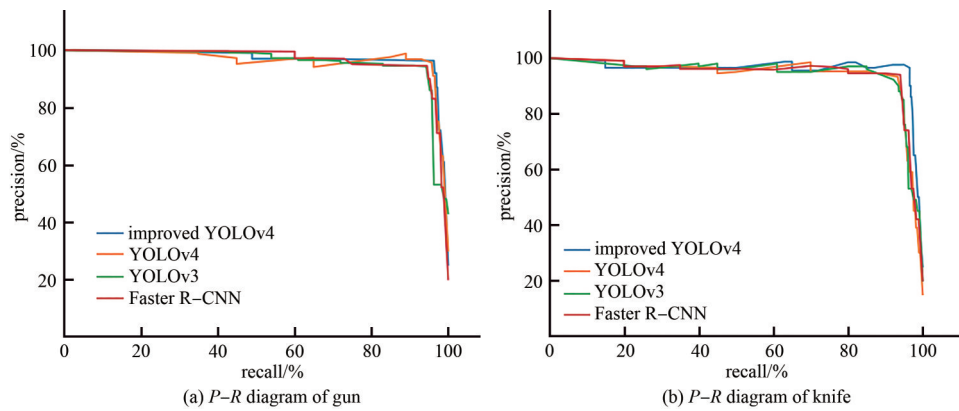


Fig.7 P-R diagrams of suspect detection

图 7 嫌疑物检测 P-R 图

3 结论

从实验结果来看,改进的YOLOv4算法在对被动式太赫兹成像系统图像进行检测时较优于YOLOv4,YOLOv3,Faster R-CNN,在本文的测试集中,改进的YOLOv4的平均精确度达到96.4%,相比YOLOv4高出1.9个百分点,相比于YOLOv3,Faster R-CNN提高了3.15,2.15个百分点,改进的YOLOv4算法模型在*P-R*曲线图中处于YOLOv4,YOLOv3,Faster R-CNN模型的上方,改进损失函数的方法通过EIOU loss将预测框更贴近真实框,提高目标框的命中率,通过FocalL1 loss减少训练时样本不平衡,改善难分样本识别不出的问题。通过实验数据,本文算法相对于其他模型算法较优,能基本满足公共交通安检系统对图像检测疑似物识别算法的速度和准确率的需求,一定程度上提高了安检的效率和准确度。

参考文献:

- [1] 冯辉,涂昊,高炳西,等. 被动毫米波太赫兹人体成像关键技术进展[J]. 激光与红外, 2020,50(11):1395-1401. (FENG Hui,TU Hao,GAO Bingxi,et al. Progress on key technologies of passive millimeter wave and terahertz imaging for human body screening[J]. Laser & Infrared, 2020,50(11):1395-1401.)
- [2] 朱峰. 人体安检太赫兹图像物体检测[D]. 广州:广东工业大学, 2018. (ZHU Feng. Object detection of THz human images for security checkpoints[D]. Guangzhou,China:Guangdong University of Technology, 2018.)
- [3] 余奕晨. 基于深度学习的太赫兹危险物品检测算法的研究[D]. 武汉:华中科技大学, 2019. (YU Yicheng. Research on terahertz dangerous goods detection algorithm based on deep learning[D]. Wuhan,China:Huazhong University of Science and Technology, 2019.)
- [4] SANG J, GUO P, XIANG Z, et al. Vehicle detection based on faster-RCNN[J]. Journal of Chongqing University, 2017,40(7): 32-36.
- [5] 袁建新. 基于深度学习的危险器具检测与识别方法研究[D]. 大连:大连理工大学, 2018. (YUAN Jianxin. Research on detection and recognition of dangerous equipment based on deep learning[D]. Dalian,China:Dalian University of Technology, 2018.)
- [6] 王葛,朱艳,沈韬,等. 基于改进Faster R-CNN算法的太赫兹安检图像识别检测[J]. 太原理工大学学报, 2021,52(2):292-299. (WANG Ge,ZHU Yan,SHEN Tao,et al. Recognition and detection of terahertz security images based on an improved faster R-CNN network algorithm[J]. Journal of Taiyuan University of Technology, 2021,52(2):292-299.)
- [7] 余亚萍. 基于深度学习的非平衡小样本异物检测识别技术研究[D]. 西安:西安电子科技大学, 2019. (YU Yaping. Research on foreign object detection and recognition technology of unbalanced small sample based on deep learning[D]. Xi'an,China: Xidian University, 2019.)
- [8] 张校捷. 深入浅出PyTorch:从模型到源码[M]. 北京:电子工业出版社, 2020. (ZHANG Xiaojie. Easy to understand PyTorch: from model to source[M]. Beijing:Electronic Industry Press, 2020.)
- [9] SHORTEN C,KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019,6(1): 1-48.
- [10] CHETHAN K B,PUNITHA R,MOHANA M. YOLOv3 and YOLOv4:multiple object detection for surveillance applications[C]// 2020 Third International Conference on Smart Systems and Inventive Technology(ICSSIT). Tirunelveli,India:IEEE, 2020.
- [11] ZHENG Z H,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. New York,USA:AAAI, 2020:12993-13000.
- [12] ZHANG Y F,REN W,ZHANG Z,et al. Focal and efficient IOU loss for accurate bounding box regression[J/OL]. arXiv: 2101.08158v1.
- [13] 杜鹏,谌明,苏统华. 深度学习与目标检测[M]. 北京:电子工业出版社, 2020. (DU Peng,SHEN Ming,SU Tonghua. Deep learning and object detection[M]. Beijing:Electronic Industry Press, 2020.)

作者简介:

周 敏(1977-),男,天津市人,本科,高级工程师,主要研究方向铁路通信信息技术 .email:zhoumin@crdc.com.