

A Method for Small Object Detection of Passive Terahertz Image in Human Body Security Check—Subtitle as Needed

Cheng Zhang, Xiucong Shi, Zihao Huang, Jigao Fu

Faculty of Computer, Guangdong University of Technology, Guangzhou Guangdong
Email: 2529091204@qq.com

Received: May 6th, 2020; accepted: May 21st, 2020; published: May 28th, 2020

Abstract

Passive terahertz imaging equipment is constantly used in human body security check. In the specific security situation such as airport, it is necessary to detect small object contraband quickly and accurately. Due to the low resolution of passive terahertz image, the traditional detection algorithm and simple deep learning algorithm cannot meet the needs of accuracy for small object detection in human body security check. In this paper, according to the low resolution of passive terahertz image, we preprocess the image before object detection. To improve detection accuracy for small objects, we optimize the architecture and default boxes of SSD model. For small object detection, our method, the optimized SSD with preprocessing, can improve accuracy to 73%. For common object detection, our method can achieve 76% mAp.

Keywords

Human Security Check, Passive Terahertz Image, Image Preprocess, Small Object Detection, Optimization SSD

基于被动式人体安检太赫兹图像的小目标物体检测方法

张 骋, 史秀聪, 黄子豪, 符基高

广东工业大学, 广东 广州
Email: 2529091204@qq.com

收稿日期: 2020年5月6日; 录用日期: 2020年5月21日; 发布日期: 2020年5月28日

摘要

被动式太赫兹成像设备不断应用于人体安检，而在特定安检场合(例如机场)需要快速准确检测出人体安检图像中的小目标违禁品。由于被动式太赫兹图像低分辨率的限制，人体安检对小目标检测精度和速度的需求，传统检测方法和单纯的深度学习算法都难以满足。根据被动式太赫兹图像低分辨率的特性，本文在目标检测前对图像进行预处理。针对小目标检测，我们对SSD算法的网络结构和预选框进行了优化。在小目标检测方面，我们的方法可以提高检测精度到73%。而在一般目标检测方面(包含小目标)我们的方法可以达到76% mAP。

关键词

人体安检，被动式太赫兹图像，图像预处理，小目标物体检测，优化SSD

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

太赫兹成像(Terahertz imaging)技术是通过太赫兹波作为信号源成像。太赫兹波频率在 0.1~10 THz 之间的电磁波，波长 0.03 至 3 mm，电磁波光谱的位置在可见光和微波之间。由于太赫兹波频率低，能量小，对人体不存在任何伤害；太赫兹波能穿透纸张、塑料和各种纺织布料等，因此能发现隐藏于人体衣物下的物品；相对于微波而言其波长短，成像精度高。同时，太赫兹成像设备既能显示出金属物品也能显示非金属违禁品，这些特性使得太赫兹成像技术非常适合于进行人体安检。

人体安检的太赫兹成像技术可以分为主动式和被动式。其中，主动式成像技术是通过对人体发出特定频率的太赫兹波，通过人体和人体携带的物体对太赫兹波的反射和散射性质的不同来判断物体的位置。而被动式人体的太赫兹成像主要是以人体自身散发的辐射和携带物体对人体散发辐射遮挡或吸收来判断物体的位置。因此，被动式人体太赫兹成像设备对人体没有任何辐射，并且不需要太赫兹波发射装置，与主动式成像设备相比被动式成像设备的价格要低，所以被动式太赫兹设备更加广泛应用于安检。但由于人体散发出太赫兹波强度较弱，被动式太赫兹图像(见图 1)相较于主动式太赫兹图像分辨率会更低，因此对被动式人体安检图像进行物体检测具有较大的难度。

随着被动式太赫兹设备成本的降低，被动式太赫兹成像技术已逐渐开始用于机场、地铁等密集人流的安检，目前被动式太赫兹成像速度已达到 10 帧/秒，按照成像技术的发展趋势，未来成像速度将会持续提高。安检过程中仍存在很多体积较小的物体属于违禁品，例如机场安检中，子弹、打火机等都属于违禁品。而在被动式人体安检图像中，小目标物体检测难度会更大。因此提高检测小目标物体的准确度，对被动式太赫兹设备在安检场景的应用，具有很大的意义。

近年来，随着深度学习技术的迅猛发展和广泛应用，图像物体检测和识别的准确度和速度都得到了很大提高。但针对被动式太赫兹人体安检图像中的目标检测，深度学习算法的应用仍处于起步阶段，而目前深度学习应用于太赫兹小目标物体检测存在以下问题：

- 1) 太赫兹图像的分辨率较低；
- 2) 小目标物体的检测准确度问题；

3) 小目标物体在深层网络特征丢失的问题。

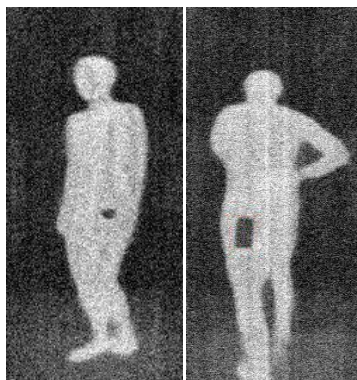


Figure 1. The passive Terahertz body image
图 1. 被动式太赫兹人体安检图像

2. 相关工作

针对太赫兹图像分辨率低的特点,在研究太赫兹图像预处理方面,主要采用的是滤波和增强的算法,如高斯滤波[1]、中值滤波[2]等滤波算法和直接灰度变换、直方图灰度变换等增强算法,以及两者的综合应用。文献中[3]通过分析法获得太赫兹扫描成像系统的点扩展函数,进而使用重构算法了重构图像,再通过灰度变换增强图像对比度和目标分辨力,提高图像质量。文献[4]提出一种改进的滤波算法,在原有算法的基础上加入了阈值的概念,提高了图像质量,增加了对隐匿物的辨别能力。文献[5]提出了一种基于自适应的滤波、中值滤波等降噪算法对太赫兹图像进行去噪,再采用边缘增强算法不断图像进行去噪增强。文献[5]采用改进优化的灰度变换算法,对太赫兹波图像进行去噪,然后再使用增强算法对图像进行锐化增强,以提升图像中物体的对比度和分辨率。这些方法主要集中于通过预处理提升图像质量,而没有考虑提升图像质量和检测之间的联系。

在太赫兹图像小目标物体检测方面,传统的检测算法主要采用的是分割和轮廓提取检测的方法。如文献[6]采用分割算法将物体从背景中分割出来,再采用边缘特征提取实现对太赫兹图像中小目标物体的检测。文献[7]通过对灰度直方图分析并对图像的直方图曲线拟合,通过选定生长阈值对图像进行区域生长,实现人体太赫兹图像中的小目标物体分割和检测。文献[8]根据太赫兹图像特点采用阈值分割、轮廓跟踪法进行轮廓提取,实现人体太赫兹图像中的小目标物体检测。文献[9]采用基于模糊局部信息 C 均值的聚类算法用于人体太赫兹图像中的小目标物体检测。这些传统的检测方法都是先从图像中分离出物体,再进行物体检测,检测过程复杂且速度慢,满足不了在安检过程中实际需求。

近年来,深度学习算法在目标检测方面取得了突破性进展。根据目标检测的体系结构,将网络划分为 VGG16 [9], ResNet-101 [10] [11], ResNeXt-101 [12] [13]和 sharglass-104 [14] [15]等类别。由于行业的保密性的限制,太赫兹图像的数量往往较少。当网络层数太多时,网络将出现较难拟合的情况,因此,我们将 VGG16 作为我们的训练网络。因此,我们将 VGG16 作为我们的训练网络。

目前,深度学习的目标检测器主要分为两类:一种是两阶段检测器,如 R-CNN 系列算法,如 R-CNN [16]、SPP-net [17]、Fast R-CNN [18]和 Faster R-CNN [19]。这些方法在小目标检测中取得了很高的检测精度。但由于检测速度太慢,不符合人体安检的发展趋势。另一种单阶段检测器,如 SSD [20], YOLO [21] 算法。其中, YOLO 算法在小目标检测速度上有很大的提高。但该算法对小目标没有优化。该方法对小目标的检测精度较低。另一种算法是 SSD,它通过多层次、多尺度提取图像特征。与两阶段法相比,该

方法在检测速度上有较大提高，准确度损失较小。因此，具有广阔的发展前景。然而，深度学习算法应用于太赫兹图像中的目标检测，特别是小目标物体的检测，仍处于起步阶段。

因此，针对太赫兹人体安检图像中小目标物体的检测的问题，本文提出了一种集预处理和结构优化为一体的 SSD 检测方法，提高了小目标物体检测精度，满足了人体安检过程中对检测速度的要求。

3. 被动式太赫兹人体安检图像中小目标检测的模型

3.1. 被动式太赫兹人体安检图像的预处理

太赫兹图像预处理的主要目的是提高图像的质量，从而有利于下一步的检测。在被动太赫兹安检中，影响成像的主要因素有两个：1) 由于光波的衍射和干扰，被动式太赫兹图像中会引入随机噪声，如波纹噪声等；2) 空气中的水分吸收太赫兹波，导致了图像的对比度低，边缘模糊。为了解决上述两个因素带来的问题，我们采用图像去噪和增强算法对随机噪声进行滤波，提高了图像的对比度、边缘模糊度和图像质量。

3.1.1. 去噪预处理

为了消除被动式太赫兹图像中存在的随机噪声，在目标检测前我们对其进行了去噪处理。其中，去噪算法大体可分为两类：一类是应用广泛的滤波算法，如中值滤波、均值滤波和高斯滤波。另一类是复杂的滤波方法，如各向异性扩散滤波。由于后者的原理比较复杂，计算量很大。因此不能满足人体安检对于实时性的需求。因此，本文我们主要分析了中值滤波、均值滤波和高斯滤波算法的原理：

中值滤波：中值滤波是一种非线性滤波，广泛用于数字图像处理中。中值滤波的主要思想是逐像素地对图像进行滤波，用相邻的数个像素值的中值代替每个像素值。其中，相邻的数个像素值的模式被称为“窗口”，它在整个图像上逐像素滑动。所有像素值将在窗口中排序。为了避免图像模糊，我们通常使用 3×3 窗口来处理图像

均值滤波：均值滤波是一种线性滤波算法。其过程与中值滤波非常相似。在窗口中，均值滤波没有用中值来替换图像中的每个像素，而是用窗口的平均值来替换它。平均值的计算方法如下：

$$g(x, y) = \frac{1}{n} \sum f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n) \quad (1)$$

其中， $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$ 是窗口中所有像素值的和， n 表示像素的总数， $g(x, y)$ 表示像素的平均值。

高斯滤波：高斯滤波是另一种线性滤波算法。它适用于消除高斯噪声，并广泛应用于图像处理中，不同于均值滤波，高斯滤波使用窗口中的加权均值代替每个像素值。它也贯穿整个图像

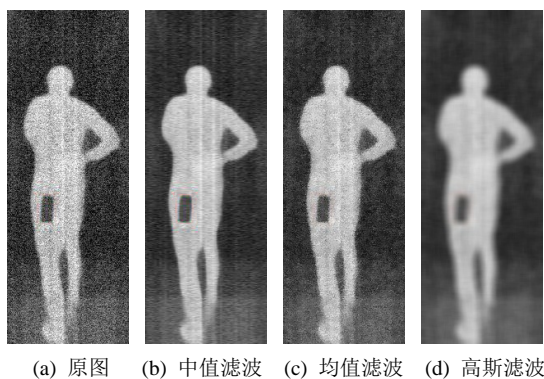


Figure 2. The Diagram of de-noise algorithm effects
图 2. 各种去噪算法的效果图

如图 2 所示, 高斯滤波可以去除被动式太赫兹人体安检图像中的随机噪声, 但是它也使得图像过于模糊, 无法顺利完成目标检测。而均值滤波没有使被动太赫兹图像变得模糊, 但是随机噪声也并没有被完全消除。中值滤波不仅可以去除图像的随机噪声, 而且使图像模糊的程度也较小。因此我们采用中值滤波算法来去除被动式太赫兹人体图像中的随机噪声。

3.1.2. 增强预处理

经过去噪处理后, 被动式太赫兹人体安检图像中的随机噪声被去除。但是, 图像中目标物体的对比度和边缘模糊度没有得到改善, 因此我们采用图像增强的方法来改善目标物体的对比度和边缘模糊度。增强算法可分为空域增强、频率增强和空频结合增强算法。其中, 空域算法是通过直接修改像素的灰度值来获得增强效果, 它包括线性变换、非线性变换; 频率增强算法通过傅立叶变换将图像变换到频域, 然后所有的增强算子对图像进行处理, 最后进行逆傅立叶变换得到增强图像。该算法具有复杂度低、转换和应用性好的优点, 但不能增强局部细节。而小目标物体的特征属于太赫兹人体安检图像中的局部特征; 空频结合增强算法在图像边缘信息分析方面具有一定的优势, 但它的计算量大, 不能满足安全检查的需要。因此, 我们将空域增强作为我们的增强方法。而空域增强算法可以分为线性变换和指数非线性变换, 下面我们分别对这两种算法进行分析。

线性变换: 线性变换算法是一种基于图像灰度的变换, 主要通过调整图像的灰度分布, 使图像更加清晰, 从而使得检测目标更易于识别。由于各种外界环境的影响, 图像的灰度范围往往被限制很窄。这些影响导致图像边缘变得模糊, 同时对比度不够明显。而通过线性变换可以扩大图像的灰度范围, 从而实现增强效果。图像的线性变换公式是

$$g(x, y) = \frac{d-c}{b-a} \times [f(x, y) - a] + c \quad (2)$$

其中, $f(x, y)$ 表示的是原始图像, $g(x, y)$ 表示的是线性变换后的图像。 $[a, b]$ 是原始图像的灰度范围, $[c, d]$ 是线性变换后的图像灰度范围。该变换公式将原先的范围 $[a, b]$ 转换成了 $[c, d]$ 。

指数非线性变换: 指数非线性变换是通过指数函数进行变换的算法。它可以扩展原始图像中较亮像素的灰度值, 增强图像的明暗区域的对比度。其转换公式为:

$$g(x, y) = b^{f(x, y)} \quad (3)$$

其中 $f(x, y)$ 表示原始图像, $g(x, y)$ 变换后的图像, b 表示常数, 在这由于太赫兹图像是灰度图, 我们设置为 2。然而, 我们无法从图像中分辨出哪个变换更好。为了找出哪种算法更好, 我们使用直方图来比较线性变换和指数非线性变换的像素分布。直方图的峰值越高, 变换效果越好。图 3 表明指数非线性变换优于线性变换。

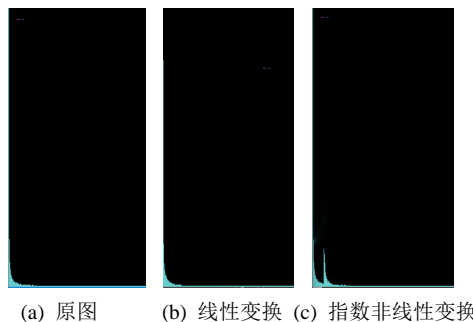


Figure 3. The histogram of images

图 3. 图像的灰度直方图

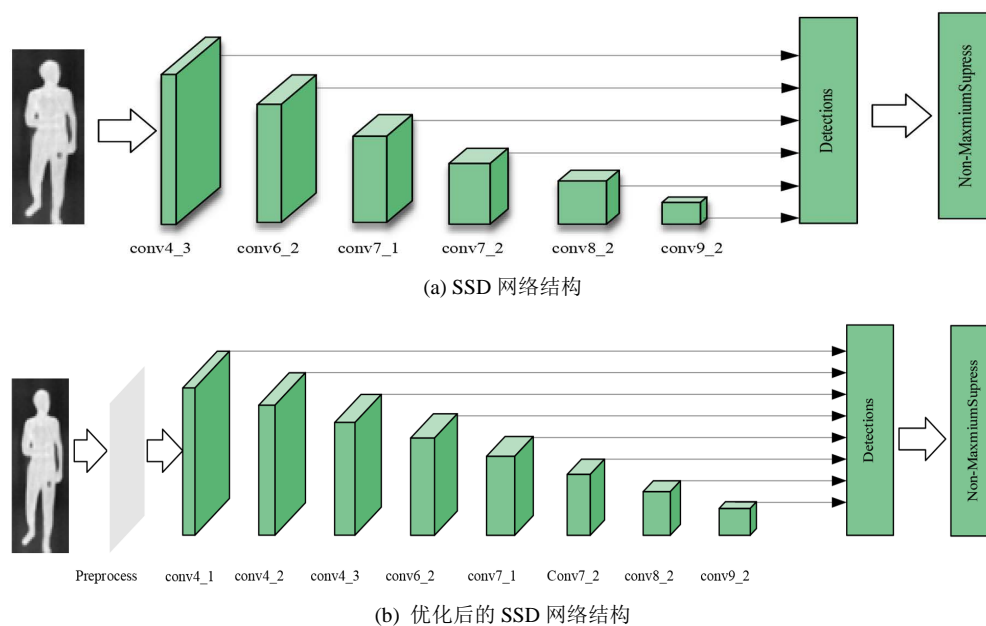


Figure 4. The Comparison of SSD network and the optimized SSD network
图 4. SSD 网络结构和优化后的 SSD 网络结构的比较

3.2. 针对小目标检测优化的 SSD 模型

SSD 是前馈卷积网络。它产生一个默认框的集合，并为这些框中对对象类实例的存在评分，然后通过非最大抑制算法来完成最终检测。

SSD 模型的核心主要是特征提取层。特征提取层主要利用 3×3 大小的两个卷积核进行等位卷积。并行卷积用于边界回归和对象分类。并行卷积可以实现浅层特征图中小目标(具有多个特征)和深层特征图中大目标(具有明显特征)的检测。特征提取网络主要在卷积层 conv4_3, conv6_2, conv7_1, conv7_2, conv8_2, conv9_2 and conv11_2 (如图 4(a)所示)。为了提高被动太赫兹图像中小目标检测的准确性，我们在 SSD 模型中做了以下更改：

针对深层网络中小目标可能丢失的特点，对 SSD 网络结构进行了优化。在 SSD 网络的 conv4_3 层的基础上，在 SSD 中增加了两个卷积层，分别是 conv4_1 和 conv4_2 (额外的特征提取层)，这两个卷积层能够更好地提取图像的低维特征。由于 conv4_1 和 conv4_2 在网络中的位置相对靠前，因此特征层的参数会较大，可能导致网络在训练过程中难以收敛。为了解决这个问题，我们为 conv4_1 和 conv4_2 层设置了 L2 归一化操作，该操作主要是通过归一化使得参数大的网络层更易收敛。

在 SSD 的模型中，先验框的个数需要随着模型特征层数的加深而不断地增加。先验框尺寸的计算公式如下所示：

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1}(k-1), k \in [1, m] \quad (4)$$

其中， S_k 表示先验框相对于比例图的比例，而 S_{\max} 和 S_{\min} ，表示先验框所占比例的最大和最小值，SSD 模型直接将其设置为 0.2 和 0.9。其中， k 表示是第 k 个特征提取层， m 根据先验知识人为地设置为 5，为了更好的提取特征来进行小目标物体的检测，第一层卷积 conv4_3 的先验框比例单独设置为 $S_{\min}/2$ 。

然而，在 SSD 中先验框尺寸和比例的选择和设置都完全取决于模型训练者的经验，这样的设置太过于主观，没有广泛的通用性，因此，我们在先验框的选择上采用了 k-mean 的聚类，从而能够根据检测图

像的性质，自动选择先验框的尺寸和比例。K-mean 聚类的算法步骤如下所示：

- 1) 初始化 K 个检测目标中心点；
- 2) 通过样本和中心点之间的欧几里得距离确定样本的种类；
- 3) 根据样本的种类更新中心点的位置；
- 4) 反复执行第 2,3 步直达样本的位置不再变化为止。

数据增强：为了使得模型在检测过程中更加健壮，每张用于训练的图像我们都会随机选择一下一种处理方式：

- 使用整张原始图像作为输入；
- 对图像按照 0.2,0.4,0.6 和 0.8 的比例随机采样；
- 对图像进行水平或垂直位移；
- 镜像翻转图像。

4. 实验结果与分析

本章实验模型的网络都是基于 VGG16 网络框架。同时，我们在 `ssd_inception_v2_coco` 数据集上预训练了网络。本章将学习率开始至设置为 10-4，权重衰减率为 0.0005，训练的批处理量为 64。根据上文的 k-mean 聚类算法结果，我们将先验框的尺寸设置为：([64,64], [31,31],[15,15],[4,4])，将先验框的长宽比例设置为：([2, 0.5],[2, 0.5, 3, 1./3],[2, 0.5, 3, 1./3])。

实验环境：

硬件配置：16 核 i9-7900 处理器，主频为 3.30Hz，内存容量 32GB，GPU：GTX1080Ti。

软件环境：操作系统：Ubuntu16.04 应用软件：OpenCV-2.4.11 + Python-3.6 + TensorFlow

成像系统配置：本文中的被动式太赫兹成像系统包括光学扫描器、多探头、聚焦透镜和信号处理单元。人体散发的太赫兹波被扫描器传送到探测器上，经过探头把太赫兹波转换成电信号，再由计算机将信号处理成太赫兹图像。成像系统的工作频率为 0.2 Thz。空间分辨率为 2 cm，成像速度为 10 fps。被动太赫兹图像的大小为 408 × 108。

数据集：为了证明我们的模型对小目标物体检测的性能，我们设置了一个仅包含小目标物体的被动式太赫兹人体安检图像数据集。同时，在现实的安检过程中，我们还需要将小目标物体从各种常用的物体中分别出来，因此，我们设置了另一个具有一般检测物体的数据集。以下是对两个数据集的具体情况说明：

仅包含小目标检测物体的数据集：在该数据集中的被检测的物体是小型的打火机。其中，训练集为：200 张被动式太赫兹人体安检图像。测试集：50 张被动式太赫兹人体安检图像。

包含一般目标检测物体的数据集：在该数据集中的被检测的物体除了小型的打火机，还有手机、匕首。其中，训练集：200 张被动式太赫兹人体安检图像，包含了 200 张检测物体是打火机，198 张检测物体是匕首，202 张检测物体是手机。测试集：总共有 150 张被动式太赫兹人体安检图像，包含 60 张打火机图像，58 张匕首图像和 62 张手机图像。

4.1. 仅包含小目标检测物体的数据集

在该数据集中，我们将本章的算法模型分别与 SSD、传统的检测算法相比较。其中传统的检测算法，本文选择了 Haar 特征提取结合支持向量机(SVM)的传统检测算法作为对比，在传统算法中该算法的性能是最好的。Cascade-RCNN [22]是最新的目标检测算法，在小目标检测方面取得了良好的效果，我们选择它作为另一个比较算法。同时，在训练模型时，我们首先使用 10-4 的学习率迭代训练 40,000 次，再使

用 10⁻⁵ 和 10⁻⁶ 的学习率分别迭代训练 10,000 次。由于在本次实验中检测的目标只有小目标物体，本文的目标检测的准确度为：正确检测到的物体总数占训练集中所有物体总数的比例。其结果如表 1 所示：

Table 1. Comparison of accuracy and speed of various network for small object detection in the passive terahertz image
表 1. 各种检测算法被动式太赫兹人体安检图像中小目标物体检测的准确度与速度的对比

模型	准确度	速度(ms)
SSD	67%	42
Haar + SVM	57%	60
Cascade-RCNN	74%	380
本文模型	73%	58

在只包含小目标的数据集上，我们模型的检测精度达到 73%，比 SSD 模型提高了 6%。每幅图像的检测时间为 60 ms。与传统算法相比，该模型具有更高的精度，检测速度比传统算法快 2 ms。虽然 Cascade-RCNN 模型的检测精度为 74%，但检测速度太慢，无法满足人身安全检测的需要。我们的模型比 Cascade-RCNN 快 322 ms，能很好地满足人体安检的需求。

4.2. 包含一般目标检测物体的数据集

由于该数据集包含一般目标物体，我们采用检测算法中常用的评价标准 mAP 来衡量各个算法的性能。在训练模型时，我们首先使用 10⁻⁴ 的学习率迭代训练 60,000 次，再使用 10⁻⁵ 和 10⁻⁶ 的学习率分别迭代训练 10,000 次。实验结果如表 2 所示。

Table 2. Comparison of accuracy and speed of algorithm for passive terahertz image
表 2. 各种检测算法被动式太赫兹人体安检图像中包含一般目标物体检测的准确度与速度的对比

模型	mAP	速度(ms)
SSD	71%	45
Haar + SVM	59%	74
Cascade-RCNN	78%	385
本文模型	76%	60

在包含一般目标物体的数据集上，我们的模型可以达到 76% 的 mAP，比 SSD 模型多 5%，并且我们的检测时间比 cascade-RCNN 少 325 ms。与传统算法相比，该模型的检测速度不仅超过 17%，而且比传统算法快 14 ms。因此，我们的模型对于人身安全检查中的小目标检测更具实用性和更好的性能。在图 5 中，我们展示了在包含公共对象的数据集上不同模型的检测示例。

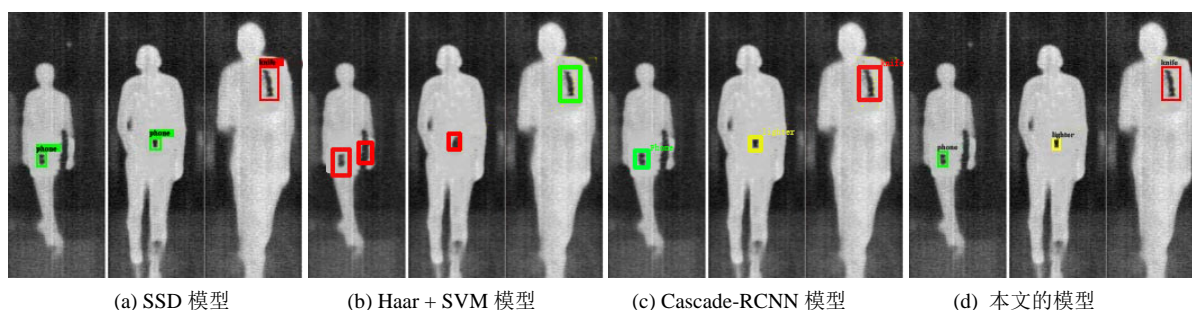


Figure 5. The detection examples with different model
图 5. 不同检测算法的效果图

5. 总结

本文提出了一种结合预处理的优化 SSD 网络结构。根据实验结果,该模型能有效地提高太赫兹人体安检图像中小目标检测的准确性。在只包含小对象的数据集上,模型的精度达到 73%。在包含其他一般目标的数据集上实现了 76% mAP。该模型的平均检测时间约为每幅图像 60 ms。检测精度和速度完全可以满足安检的要求。虽然我们在太赫兹人体图像中的小目标检测方面取得了一些进展。但是还有一些问题没有解决。这些问题主要是被动太赫兹图像分辨率较低,因此很难提高小目标的检测精度导致的。为了解决这些问题,我们考虑用超分辨率重建算法来提高图像的分辨率,从而进一步提高目标检测的精度。

参考文献

- [1] Si, S., H. F. and Gu, Y. (2014) A Gaussian Filtering Algorithm Based on Irregular Region. *Computer Science*, **41**, 313-316.
- [2] Huang, T., Yang, G. and Tang, G. (1979) A Fast Two-Dimensional Median Filtering Algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **27**, 13-18. <https://doi.org/10.1109/TASSP.1979.1163188>
- [3] 赵然, 赵源萌, 张存林. 被动式人体太赫兹图像中的目标辅助识别[J]. 强激光与粒子束, 2014, 26(7): 132-136.
- [4] 朱维文, 赵跃进, 朱慧时, 李冰, 孔令琴, 董立泉. 改进的均值滤波算法在太赫兹成像中的应用[J]. 红外与激光工程, 2013, 42(5): 1241-1246.
- [5] 朱书进, 李跃华. 基于自适应流形高维滤波的太赫兹图像滤波算法[J]. 太赫兹科学与电子信息学报, 2016, 14(4): 497-501 + 507.
- [6] 孙建忠, 章乐, 高飞, 李九生. 太赫兹连续波成像的质量改进与降噪[J]. 光谱学与光谱分析, 2017, 37(11): 3343-3346.
- [7] 冯艳平, 王徽. 基于阈值分割和边缘检测的枪支 THz 图像识别[J]. 红外, 2011, 32(12): 23-26.
- [8] 江天, 沈会良, 杨冬晓, 刘建军, 邹哲. 基于模糊局部信息 C 均值的太赫兹图像目标检测[J]. 激光技术, 2015, 39(3): 289-294.
- [9] Xie, L.L., Liu, Y.L., Jin, L.W. and Xie, Z.C. (2019) DeRPN: Taking a Further Step toward More General Object Detection. <https://arxiv.org/abs/1811.06700>
- [10] Lu, X., Li, B., Yue, Y., Li, Q. and Yan, J. (2019) Grid R-CNN. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-21 June 2019, 7355-7364. <https://arxiv.org/abs/1811.12030>
<https://doi.org/10.1109/CVPR.2019.00754>
- [11] Li, J., Wang, C., Wang, Y., Tai, Y., Qian, J., Wang, C., Yang, J., Li, J. and Huang, F. (2019) Dual Shot Face Detector. <https://arxiv.org/abs/1810.10220>
- [12] Li, Y., Chen, Y., Wang, N. and Zhang, Z. (2019) Scale-Aware Trident Networks for Object Detection. <https://arxiv.org/abs/1901.01892>
- [13] Zhu, C., He, Y. and Savvides, M. (2019) Feature Selective Anchor-Free Module for Single-Shot Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-21 June 2019, 840-849. <https://arxiv.org/abs/1903.00621>
<https://doi.org/10.1109/CVPR.2019.00093>
- [14] Chen, Y., Yang, T., Zhang, X., Meng, G., Pan, C. and Sun, J. (2019) Neural Architecture Search on Object Detection. <https://arxiv.org/abs/1903.10979>
- [15] Gupta, A., Dollar, P. and Girshick, R. (2019) Lvis: A Dataset for Large Vocabulary Instance Segmentation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-21 June 2019, 5351-5359. <https://arxiv.org/abs/1908.03195>
<https://doi.org/10.1109/CVPR.2019.00550>
- [16] Girshick, R., Donahue, J. and Darrell, T. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://arxiv.org/abs/1905.02285>
<https://doi.org/10.1109/CVPR.2014.81>
- [17] He, K., Zhang, X. and Ren, S. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916. <https://arxiv.org/abs/1406.4729>
<https://doi.org/10.1109/TPAMI.2015.2389824>

-
- [18] Girshick, R. (2015) Fast R-CNN. <https://arxiv.org/abs/1504.08083>
- [19] Ren, S., He, K., Grishick, R. and Sun, J. (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C. and Reed, S. (2015) SSD: Single Shot Multi-Box Detector. *European Conference on Computer Vision*, Amsterdam, 8-16 October 2015, 21-37. <https://arxiv.org/abs/1512.02325>
https://doi.org/10.1007/978-3-319-46448-0_2
- [21] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. <https://arxiv.org/abs/1506.02640>
- [22] Cai, Z. and Vasconcelos, N. (2017) Cascade R-CNN: Delving into High Quality Object Detection. <https://arxiv.org/abs/1712.00726>