文章编号：1672-8785(2020)02-0013-13

主动太赫兹成像中的多目标分割与检测识别方法

薛　飞[1]　梁　栋[1*]　喻　洋[2,3]　潘家兴[1]　吴天鹏[1]

(1. 南京航空航天大学计算机科学与技术学院，江苏 南京　211106；
2. 中国工程物理研究院电子工程研究所，四川 绵阳　621900；
3. 中国工程物理研究院微系统与太赫兹研究中心，四川 成都　610200)

摘　要：针对主动太赫兹成像中存在的图像品质差以及藏匿物品类别多样、训练样本稀缺且类别不平衡等问题，提出了基于用条件生成对抗网络构建的 Mask-CGANs 模型的目标分割网络和基于 RetinaNet 的目标检测识别网络，实现了太赫兹图像中藏匿物品的多目标分割和检测识别。针对分割任务提出的约束损失函数和网络结构，使模型在召回率和虚警率之间达到平衡且降低了对训练样本规模的要求。针对检测任务采用的损失函数提高了训练样本不平衡条件下的检测精度。

关键词：太赫兹成像；条件生成对抗网络；目标分割；目标检测

中图分类号：TP391.4　**文献标志码**：A　**DOI**：10.3969/j.issn.1672-8785.2020.02.003

Multi-object Segmentation, Detection and Recognition in Active Terahertz Imaging

XUE Fei[1], LIANG Dong[1], YU Yang[2,3], PAN Jia-xing[1], WU Tian-peng[1]

(*1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;*
2. Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China;
3. Microsystem and Terahertz Research Center, China Academy of Engineering Physics, Chengdu 610200, China)

Abstract: Aiming at the problems in the active terahertz (THz) imaging such as the poor image quality, the variety of hidden objects and the scarcity and imbalance of training samples, the objects segmentation networks based on the conditional generative adversarial networks' model Mask-CGANs and the objects detection and recognition networks based on the RetinaNet are built, which realizes the multi-object segmentation, detection and recognition of hidden objects in the THz imaging. The constraint loss functions and the networks structures proposed for the segmentation task make the model keep the balance between the recall rate and the false alarm rate, and the requirement of training sample size is reduced. The loss functions used for the detection task improve the detection accuracy under the condition of unbalanced training samples.

Key words: terahertz imaging; conditional generative adversarial networks; object segmentation;

收稿日期：2020-01-16

作者简介：薛飞(1995-)，男，甘肃酒泉人，硕士生，主要研究方向为模式识别与计算机视觉。

*通讯作者：E-mail：liangdong@nuaa.edu.cn

object detection

0 引言

检测和识别人体携带的藏匿物品是公共安检中的一项关键任务。现行人工安检的筛查效率低下，且漏检率高。太赫兹(THz)波的波长介于微波与红外光之间，它是一种频率位于0.1～10 THz范围内的电磁波。由于具有高穿透性、低能性、相干性等特点且大多数物质在THz波段都有指纹谱，THz波可穿透衣物，通过合成孔径成像的方式对人体进行成像，从而获取其携带的藏匿物品的相关信息。

根据THz信号源辐照的有无，THz成像系统可分为主动式(有源)和被动式(无源)两类[1]。其中，被动式THz成像系统不使用信号源辐照，仅依靠被测物品自身THz辐射的能量差异进行对比成像；主动式THz成像系统则利用THz信号源辐照被测物品，对反射或透射的信号进行处理后实现成像。尽管被动成像系统的结构相对简单，但由于人体自身辐射微弱，该系统对探测器灵敏度的要求极高，并且受环境的影响较大；而主动成像则不存在上述问题。在合成孔径成像的过程中，信号的频率线性度及相噪、发射机功率、接收机噪声系数等指标对成像结果有较大影响。而且THz射频前端技术尚有技术瓶颈，系统关键指标无法达到接近传统微波雷达成像品质的要求，导致主动THz图像的对比度和信噪比较低，藏匿物品与人体在像素数值上的差异不明显，且物品边缘和人体混淆而难以分辨，并伴有较多背景噪声。

在THz安检系统中，首先需要自动检测当前图像中是否有藏匿物品。由于藏匿物品类别的开放性，通常难以一次性实现精确自动筛查。所以该系统需提取目标外轮廓并识别目标的参考类别，以便于人工进行二次确认筛查。因此，基于THz波的藏匿物品分割和检测识别技术具有重要的应用价值。

在THz成像分割方面，早期的研究工作都是聚焦于THz成像场景的统计模型。Shen X等人[2]提出了一种利用混合高斯模型对辐射温度进行建模的多层次阈值分割算法。该方法使用各项异性扩散算法去除噪声，随后通过修正随阈值变化的分割边界来完成分割。Lee D等人[3]同样使用混合高斯模型，通过期望最大化算法求取贝叶斯分布的条件概率以完成高斯分量的参数估计。Yeom S等人[4]使用基于多尺度混合高斯模型的分割算法，并在期望最大化算法之前通过采用矢量量化技术来实现实时性。由于THz成像条件复杂，上述方法只能得到粗略的分割结果，在恶劣成像条件下的分割结果欠佳。

近年来提出的基于深度卷积神经网络的分割算法具有良好性能[5-8]。该方法大致分为两类：一类是基于R-CNN[9]推荐的方法，另一类是基于语义分割的方法。其中，前者使用自下而上的流程，得到依赖于R-CNN推荐区域的分割结果并对其进行分类。第二类方法先得到语义分割的结果，再将像素分类为不同实例。比如，Mask-RCNN实例分割算法[10]通过共享卷积层的特征，能够同时完成检测分类和分割任务。但是这类方法需要大量的数据才能训练出可用的模型。姚家雄[11]和王崇剑等人[12]使用卷积神经网络检测THz图像中的藏匿物品，实现了单目标的藏匿物品检测。但该方法未能实现多目标物品分割和检测，因此在实际应用中还存在一定的局限性。

本文针对主动THz成像中存在的图像品质差以及藏匿物品类别开放、训练样本稀缺且类别不平衡等问题，提出将基于条件生成对抗网络的Mask-CGANs模型作为目标分割网络，并基于RetinaNet和Focal Loss损失函数构建了目标检测识别网络，完成了对THz图像中的藏匿物品进行多目标分割和检测识别的统一算法框架(见图1)。针对分割任务提出的损失

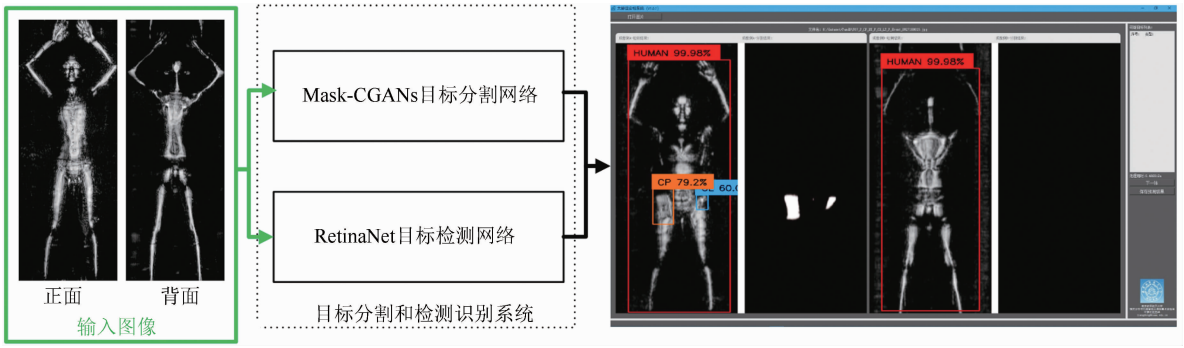


图 1　基于 THz 图像的目标分割和检测识别网络结构

函数和网络结构，使模型在召回率和虚警率之间平衡。针对检测任务采用的 Focal Loss 损失函数提高了训练样本不平衡条件下的检测精度。由于网络结构简单且模型处理速度快，该方法便于在实时安检系统中应用。

1　模型设计

1.1　条件生成对抗分割网络

生成对抗网络(Generative Adversarial Network，GAN)在图像分割、超分辨率、风格迁移等领域获得了应用[13]。它通过对抗博弈机制训练生成模型，避免了对复杂的隐含概率分布的估计，有利于在训练样本量较少的情况下应用。当类别开放且标记训练数据有限时，通过用 GAN 框架对 THz 图像进行目标分割，可以降低算法对训练样本的苛刻要求，从而有利于提升模型的泛化性能。GAN 学习一个从输入噪声 z 到输出空间样本 g 的映射函数 $f: z \rightarrow g$ 时，无需隐含分布的先验信息。它包含一个生成器和一个判别器。其中，生成器用于生成样本，而判别器则用于接收生成器生成的样本和从训练集中采样的真实样本，并通过二分类将这两类样本区分开来，直到难以区分生成器生成的样本和真实样本时结束训练过程。基于条件的生成对抗网络(Conditional GANs，CGANs)[14]通过在 GAN 的输入中加入条件约束来指导生成器的生成过程，最终生成能够符合具体任务要求的结果。将 THz 图像作为条件并将分割结果作为生成器的目标，是本文提出的 Mask-CGANs 网络的基本思路。采用 U-Net[15]作为生成器，并针对当前任务对其进行了改进。结合改进的损失函数，构成了完整的 Mask-CGANs 网络。

U-Net 生成器通过直接连接对称的下采样层和上采样层来实现不同层之间的信息融合，有效保留了低层特征。然而该方法的局限性在于，直接将对应特征层相互连接的方式无法灵活剔除对分割有害的特征。由于低层图像特征易包含噪声，引入过多的低层特征会造成分割结果的虚警率较高。因此，本文提出一种选择性连接的 U-Net 模型。如图 2 所示，生成器的剔除部分靠近输出层的连接。图 2 中，空白层表示 Conv2d 操作(decoder 部分为转置卷积)，卷积核为 5×5，步长为 2×2；灰色层表示 BatchNorm(BN)操作(decoder 部分在 BN 后增加一层 Dropout)；蓝色下对角线层表示 Concat 操作，后接 1×1 卷积将通道减半。后续实验表明，选择性连接的 U-Net 模型在保证召回率不受很大影响的前提下显著地降低了虚警率。

在训练过程中，将 THz 图像 x 输入到生成器中。生成器连同噪声 z 输出对应的分割结果 $G(x,z)$。此时，判别器的输入包含两个部分：生成的样本对 $(x,G(x,z))$ 和真实的样本对 (x,y)。其中 y 是人工标注的分割掩膜。判别器需要给真实的样本对较高的分数，相反给生成的样本对较低的分数。对生成器而言，它需要生产越来越真实的样本 $G(x,z)$ 来逼近真实样本，从而获得更高的分数。该过程直到判别器无法分辨这两对输入为止，至此生成器便学习到了从 THz 图像到分割图像的映射关系。为了在训练过程中提供收敛依

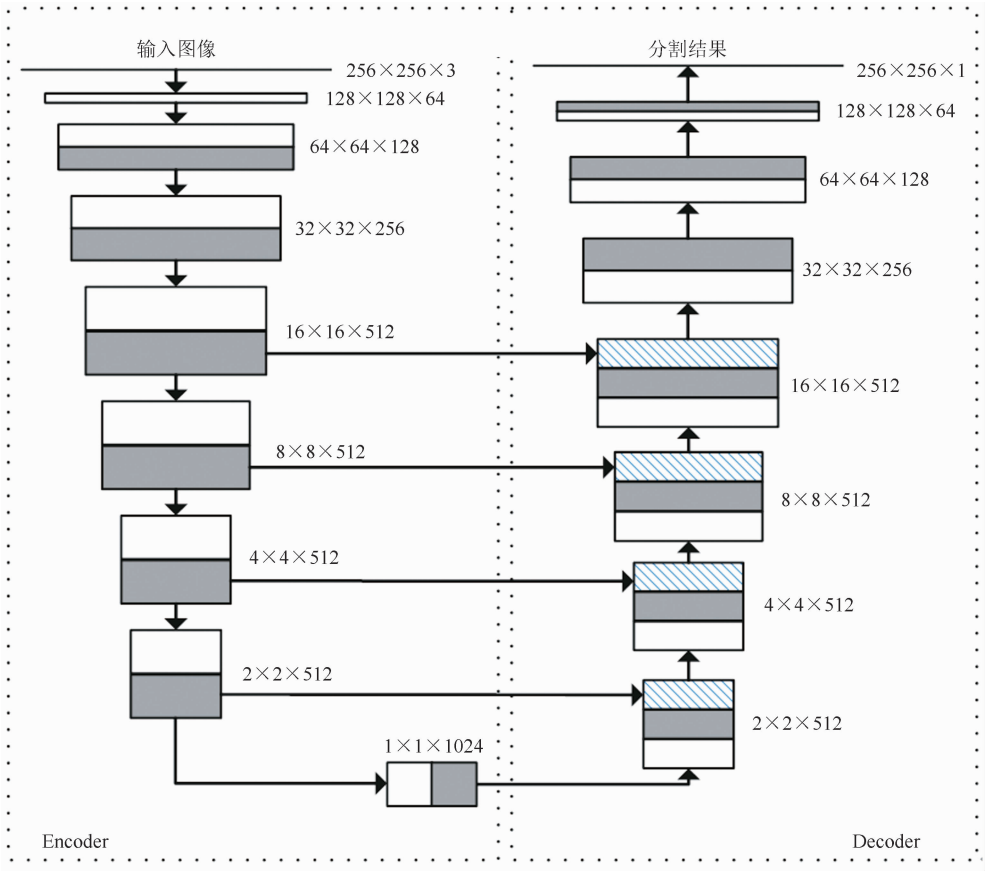


图 2 选择性连接的 U-Net 模型的细节与参数

据，首先引入条件生成对抗网络的损失函数作为 Mask-CGANs 损失函数的第一项：

$$L_{CGANs}=E_{x,y\sim data(x,y)}[\log D(y|x)]+E_{x\sim data(x),z\sim p(z)}[\log(1-D(G(z|x)|x))] \quad (1)$$

其中，判别器通过让 $D(y|x)$尽可能接近 1，并让 $D(G(z|x)|x)$尽可能接近 0，使期望最大化。相反地，生成器则企图让$D(G(z|x)|x)$ 尽可能接近 1，使期望最小化。噪声 z 被表示为生成器中的 Dropout 操作，使生成器的输出更加多样化，进而增强了模型的泛化性能。x 作为条件被分别加入到两个网络中，构成条件生成对抗网络，用于对约束生成器的输出与输入进行配对。式(1)从概率角度考虑了条件生成对抗网络的损失函数，其目标是生成判别器无法判别真伪的图像。但它缺乏对生成的重构图像与人工标注的分割掩膜之间外观相似性的直接约束。因此，引入曼哈顿距离来衡量重构误差，并将其作为损失函数的第二项：

$$L_{L_1}=\|y-G(x,z)\|_1 \quad (2)$$

式(2)要求生成的图像与正确的分割结果足够接近。另外，类似于许多目标分割任务，藏匿物品在整张图像中只占据很小的部分。故据此先验知识引入稀疏误差作为损失函数的第三项：

$$L_S=\|G(x,z)\|_1 \quad (3)$$

式(3)要求生成的图像应该是稀疏的。使用 L_1范数作为稀疏性约束，而不是理论上的 L_0 范数和常用的 L_2范数，原因在于前者需要求解 NP 难的问题，而后者只能将各项逼近到很小的数值，并不能保证大部分元素是 0。最终的损失函数如下：

$$G_{oal}(G,D)=\arg\min_G\max_D L_{CGANs}+\lambda L_{L_1}+\beta L_S \quad (4)$$

式中，λ 和 β 为损失函数中不同约束的比例因子。实验中取 $\lambda=850$，$\beta=50$。通过实现条件生成对抗网络损失函数的最小化来监督生成

过程，以保证生成的图像与输入的 THz 图像一一对应。基于最小化曼哈顿距离重构误差，维持了精确的分割结果。作为先验知识，稀疏性约束进一步降低了分割的虚警率。

1.2　目标检测识别网络

目标检测识别的目的是在图像上标出目标的位置和范围，并且识别出目标的类别。常用的基于深度卷积神经网络的目标检测识别算法大致可以划分为两类：一阶段和两阶段方法。后者采用分步走的方式，即先通过推荐候选区域过程，在待检测目标周围生成候选区域，然后对这些候选区域进行分类[9-10,16-17]。而一阶段方法[18-20]则不需要推荐候选区域过程，可一次性完成检测和分类任务，从而具有更快的检测速度。为了在分割的同时完成检测任务，本文使用 RetinaNet[21]作为目标检测分类模块。它是一种一阶段的目标检测算法，可直接进行检测框的回归和分类，符合快速检测的要求。

图 3 为 RetinaNet 目标检测识别网络的架构图。该网络选用深度残差卷积神经网络 ResNet[22]中的 ResNet-50 网络结构作为特征提取网络，采用高低层特征融合的特征金字塔结构，并在最后的多层特征图后接目标分类和检测框回归子网络，使模型具有多尺度检测能力。

深度残差网络通过用残差块将相邻层的输入输出连接起来而构成自身映射。由于改善了网络训练时存在的“梯度消失”和“梯度爆炸”问题，它已成为很多网络模型的特征提取网络。低层特征语义信息稀缺，但包含丰富的图像细节；高层特征语义信息较丰富，但缺乏图像细节信息。特征金字塔将高低层的特征进行融合，改善了模型的检测性能。分类子网络和检测框回归子网络的结构大部分相同，即从每个特征金字塔级别获取输入特征，然后通过四个卷积核为 3×3、输出通道为 256 的卷积层来解码特征。两者的区别在于最后的输出层：对于分类子网络的输出，通过一个卷积操作将输出通道转换成同目标类别数量 K 与检测框数量 A 的乘积一致(即输出通道数为 KA)，最后利用 Sigmoid 函数将每个通道的预测结果二值化，以判断其是否为该通道对应的类别；对于检测框回归子网络的输出，通过线性操作将其转换为 $4A$ 的向量，即预测每个检测框的中心位置 (X,Y)和大小(W,H)。分类子网络不与检测框回归子网络共享参数，而不同级别的特征金字塔子网络共享参数。

目标检测问题是典型的类不平衡问题。由于场景的多样性，训练所需的负类样本(场景背景)数量往往远大于正类样本(目标)数量。然而，在以损失函数最小化为目的的模型训练过程中，大量负类样本的引入往往会造成训练过程中负类权重过大(即分类器更倾向于将样本预测为负类)，导致检测精度下降。RetinaNet 网络通过在损失函数中加入与样本相关的权重来改变样本对损失函数的影响，并将 Focal Loss 作为损失函数。该函数是在二分类交叉熵函数的基础上发展而来的，其表达式为

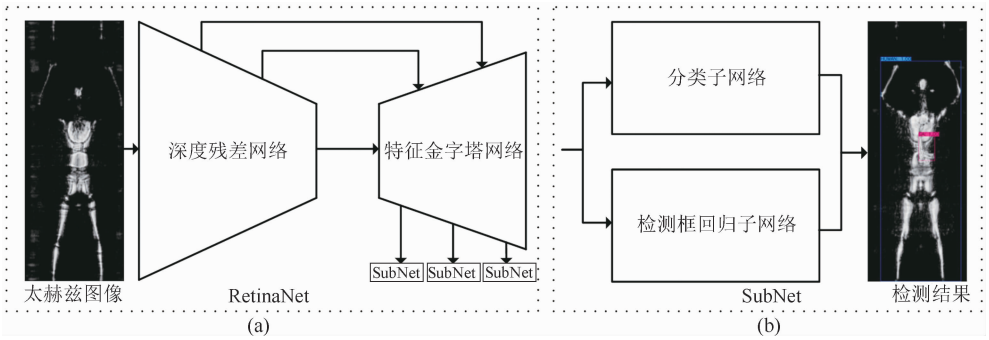


图 3　RetinaNet 架构图：(a)总体结构图；(b)特征金字塔后接的子网络

$$CE(p,y)=\begin{cases}-\log(p), y=1\\ \log(1-p), y=-1\end{cases} \tag{5}$$

式中，p 为模型预测的概率，正类标签 $y=1$，负类标签 $y=-1$。定义 P_t 如下：

$$P_t=\begin{cases}p, y=1\\ 1-p, y=-1\end{cases} \tag{6}$$

可以得到 $CE(p,y)=CE(P_t)=-\log P_t$。式(6)中，$p$ 表示模型预测的概率，正类标签 $y=1$，负类标签 $y=-1$。正类的预测值应该越大越好，负类的预测值应该越小越好。这等同于将所有样本的 P_t 优化为最大值。二分类交叉熵函数可较好地计算分类误差，但是样本类不平衡的问题依然存在。因此引入了平衡交叉熵，其表达式为

$$CE(P_t)=-\alpha\log P_t \tag{7}$$

式中，α 是正负样本参数矩阵。在 THz 图像数据集中，负类样本数量大于正类样本。对于正类样本，$\alpha=1$；对于负类样本，$\alpha=0.25$。这使得正类样本对模型损失函数的影响大于负类样本，可用于处理正负失衡的问题。在此基础上，加入了一项调整因子$(1-P_t)^{\gamma}$，最终得到 Focal Loss 函数：

$$FL(P_t)=-\alpha(1-P_t)^{\gamma}\log P_t \tag{8}$$

式中，γ 是调节难易样本对损失函数贡献的参数。对于困难样本，模型较难预测，并得到较低的 P_t。P_t越小，$(1-P_t)^{\gamma}$ 越大，从而产生相对较大的损失。而对于简单样本，模型可以较好地预测并得到较高的 P_t。P_t越大，$(1-P_t)^{\gamma}$ 越小，从而产生小的损失。因此将模型训练的重点放在困难样本上。

2 实验

2.1 数据集制备

数据集的制备采用了中国工程物理研究院微系统与太赫兹研究中心研制的 THz 快速安检门。该系统采用阵列式扫描成像方式，工作于 140 GHz，成像分辨率为 5 mm×5 mm。采集数据时，模特站立于阵列前端，并根据实验的需要将不同的目标物品藏匿在衣物中。将检测的物品归为表 1 所示的五种类别。

表 1 物品分类表

物品类别	包含物品
枪支等武器	枪支
管制器具	厨刀、金属剪刀、金属匕首、陶瓷水果刀
危险物品	装有水的矿泉水瓶、打火机
随身携带物品	皮质钱包、钥匙串、手机
不明确物品	包括但不限于上述人体携带的其他物品

表 2 列出了实验中制备的两个数据集(单目标数据集 I 和多目标数据集 II)的各项统计值，包括样本、模特、物品类别的数量，训练集、验证集、测试集的划分以及最大和最小目标像素尺寸等。该数据集的训练集全部经过人工标注，并获得了目标物的分割掩膜和分类框，可用于模型训练。

图 4 所示的统计结果展示了数据集 II 中各类目标的像素面积分布情况和训练样本数量。其中，红色实线表示面积均值和误差，蓝色虚线表示样本数量。

2.2 实验结果

2.2.1 Mask-CGANs 分割性能评估

为了评估 Mask-CGANs 的分割性能，采用 Li W[23]工作中的评价标准。具体说来，P 和 N 分别表示样本中的正类和负类。当且仅当样本标签是正类，且输出的分割结果与

表 2 数据集的各项统计值

数据集	样本数量	模特数量	物品类别	训练集	验证集	测试集	多目标	最大目标像素	平均目标像素	最小目标像素	图像尺寸
I	720	4	4	480	0	240	否	5124	2075	268	335×880
II	3166	10	11	2400	300	466	是	6359	1371	113	335×880

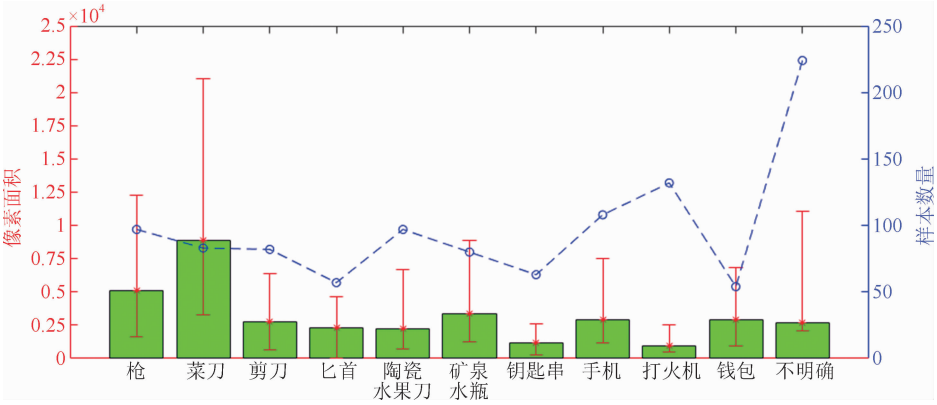


图 4 各类目标的像素面积分布情况与数量统计

ground truth 中的目标区域的重叠比例大于设定的阈值时，该样本才是真正类(TP)，否则它就是假负类(FN)；当且仅当样本标签是负类，且输出的分割结果中所有的像素值都为 0 时，该样本才是真负类(TN)，否则它就是假正类(FP)。评估指标定义如下：召回率为 $\frac{TP}{TP+FN}$，准确率为 $\frac{TP+TN}{P+N}$，精确率为 $\frac{TP}{TP+FP}$，假正率(虚警率)为 $\frac{FP}{N}$。

对比实验中加入 Mask-RCNN 作为参照。它基于 ResNet-101 并且在 COCO 数据集上进行预训练。而 Mask-CGANs 和其他对比模型则没有使用任何预训练模型。

表 3 所示的量化结果表明了网络结构和损失函数改进的合理性和有效性。引入曼哈顿距离 L_{L1}后，召回率有显著的提升，虚警率也得到了抑制。而选择性连接 U-Net 使模型在保证召回率不受较大影响的前提下保持较低的虚警率。L_S 稀疏性约束能够有效地降低虚警率。

表 3 同时列出了 Mask-CGANs 和目前先进的分割算法 Mask-RCNN 在样本中存在多个目标区域的情况下的性能指标。结合图 5 中的可视化结果可以看出，Mask-CGANs 比 Mask-RCNN 的分割结果更为精确。

图 6 展示了 Mask-CGANs 的部分分割结果，其中包含了不同的个体样本、人体姿态、位置和物品类别。可见，Mask-CGANs 能够生成较为准确的物品轮廓细节。

Mask-CGANs 训练过程中使用单张 GTX1080，并采用 Windows 10 操作系统和 TensorFlow 开源框架。在随机初始化的条件下训练了 7.6 h，其中 batch size 为 16 个样本，且所有样本归一化为 256×256。在测试过程中，Mask-CGANs 实现了 69.7 fps 的分割速度，而在相同条件下 Mask-RCNN 的分割速度为1.6fps。这样的速度使Mask-CGA-

表 3 Mask-CGANs 与其他方法的量化结果

数据集	模型	生成器结构	损失函数	召回率	准确率	精确率	假正率
I	Mask-RCNN	—	—	0.8563	0.7583	0.7965	0.4375
	Mask-CGANs	U-Net	L_{CGANs}	0.7500	0.5000	0.6000	1.0000
		U-Net	$L_{CGANs}+L_{L1}$	0.9125	0.7750	0.7849	0.5000
		选择性连接 U-Net	$L_{CGANs}+L_{L1}$	0.9125	0.8000	0.8111	0.4250
		选择性连接 U-Net	$L_{CGANs}+L_{L1}+L_S$	***0.9125***	***0.8208***	***0.8343***	***0.3625***
II	Mask-RCNN	—	—	0.7582	0.8125	0.7625	0.1525
	Mask-CGANs	U-Net	$L_{L1}+L_S$	0.4840	0.6266	0.7289	0.2083
		选择性连接 U-Net	$L_{CGANs}+L_{L1}+L_S$	***0.8846***	***0.8858***	***0.8342***	***0.1135***

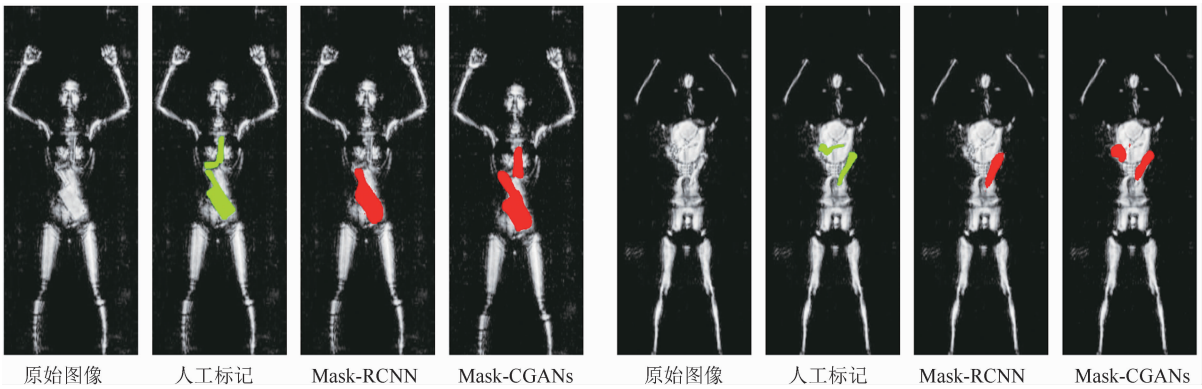


图 5 Mask-CGANs 与 Mask-RCNN 对于多目标分割的可视化结果比较

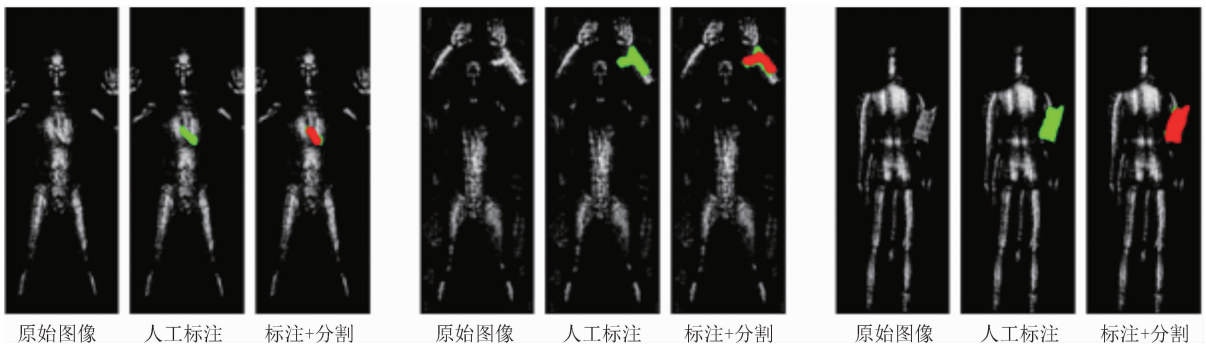


图 6 分割结果的可视化展示(每组包含 3 张图像，从左到右分别是原始 THz 图像、人工标注(绿色)图像、人工标注区域与分割结果(红色)同时叠加到原始图像上的结果)

NS 可应用于实时安检系统。

2.2.2 RetinaNet 检测识别性能评估

为了评估 RetinaNet 在目标检测任务中的性能，将它和 YOLOv3 一阶段目标检测算法[19]进行对比。模型训练时，使用 ImageNet 上的预训练模型并在 THz 数据集上微调。另外还加入了 FRCN-OHEM[24]这种二阶段目标检测算法作对比。它使用了不同于 Focal Loss 的在线困难样本挖掘方法。使用 AP(Average Precision)作为评价标准。具体来说，每个类别的 AP 等于其 PR(Precision-Recall)曲线下面积，mAP (mean Average Precision)是多个类别的平均 AP。

表 4 列出了多目标数据集 II 上的测试结果对比数据。其中 Backbone 表示模型的特征提取网络，ClassLoss 表示训练时分类的损失函数。所有的检测方法都使用了相同的训练集和测试集，训练不同的 RetinaNet 时使用了相同的迭代次数和学习率。其中 RetinaNet 的分类损失函数为交叉熵的实验说明两点：一是在都使用交叉熵时 RetinaNet 的检测精度高于 YOLOv3；二是 RetinaNet 使用 Focal Loss 损失函数时的检测精度要高于使用交叉熵损失函数时。RetinaNet 对多个类别的检测性能优于 YOLOv3。特别是对于“剪刀”、“陶瓷刀”、“打火机”和“手机”这些难以分辨的类别，使用 Focal Loss 的 RetinaNet 的检测效果更好。本实验中，YOLOv3 的检测速度低于 RetinaNet，FRCN-OHEM 的检测精度虽然高于 YOLOv3，但是仍低于 RetinaNet，并且它的检测速度过慢。

注意到，当 Backbone 为 ResNet-101 时，RetinaNet 检测 mAP 反而没有 Backbone 为 ResNet-50 时高。造成这种现象的根本原因是训练样本不足。在数据集 II 的训练集中，虽然训练样本数量超过两千，但平均到 11 类后的每类训练样本数有限(见图 4)。在训练数据有限的情况下，引入深层网络不但不会

表 4 RetinaNet 和其他方法的检测结果对比

模型	方法		每类 AP/%											mAP/%	检测速度/fps
	Back-bone	Class-Loss	人体	枪	菜刀	剪刀	匕首	陶瓷刀	矿泉水瓶	钥匙串	手机	打火机	钱包	不明确		
YOLO-v3	Dark-net-53	交叉熵	**100**	67.15	81.69	45.56	0	16.39	11.29	25	60.74	18.57	**75.43**	30.35	44.35	3.85
FRCN-OHEM	VGG 16	交叉熵	**100**	83.39	81.69	66.67	0	16.39	31.56	37.38	63.21	18.57	44.23	22.38	47.12	2.5
RetinaNet	ResNet－50	Focal Loss	**100**	82.05	88.56	**66.67**	0	37.16	**64.86**	78.57	68.75	**62.87**	65.17	**34.34**	**62.42**	**15.05**
	ResNet－50	交叉熵	**100**	**83.39**	**100**	45	0	**48.73**	57.14	**83.33**	63.21	42.28	44.23	33.02	58.36	**15.05**
	ResNet－101	Focal Loss	**100**	72.31	88.89	44	0	45.57	57.78	53.7	**69.78**	56.51	62.05	24.67	56.27	11.7
	ResNet－101	交叉熵	**100**	73.33	88.89	39.5	0	21.08	47.22	80.56	67.62	47.61	40	22.38	52.35	10.19

提升模型性能，反而会导致模型过拟合，造成检测阶段的模型性能下降，从而出现更多分类错误。

为了从应用角度评估系统性能，还采用个体计数法来评估检出率和误报率。将检出率定义为在正常操作条件下检测出人体携带物品的比率，即检出率＝正确报警次数/实际携带物品测试次数(物品所在位置被标记即为检测出物品)。将误报率定义为在正常操作条件下检测算法出现误报的比率，即误报率＝错误报警次数/总测试次数。数据集 II 的检出率和误报率统计结果分别是 74.68% 和 14.87%。

图 7 展示了用 RetinaNet 检测 THz 图像中藏匿物品的多目标检测结果。图 8 展示了人体不同角度的 THz 成像的 RetinaNet 检测结果。如图 8 左下图第一行所示，在一个特定角度下可得到某物品的检测结果，但由于尺寸微小，在其他角度下未能检测出该物品。

2.2.3　THz 图像中目标分割和检测的尺寸分析

为了评估分割算法在目标不同像素面积下的性能，统计了测试集的目标分割像素，如图 9 所示。其中绿色表示测试集中人工标注掩膜的目标像素面积的数量统计结果，而红色表示 Mask-CGANs 分割掩膜结果的目标像素的面积。当前图像中有目标时，如果分割结果目标的面积和标注目标的面积接近，则认为分割成功，并在标注结果对应的分割结果面积上累计；若分割失败，则只累计标注结果的面积。图 9 中某一目标像素面积对应的柱状图都是红色，就表示该面积范围内的所有目标都被成功地分割出来。从统计结果可以看出，实验所用数据集的目标面积在 100～4700 像素范围内，而 Mask-CGANs 分割最小目标的面积为 300 像素，它在 1300～4700 像素范围内的分割效果较好，在 100～1300 像素范围内的分割结果较差。考虑到表 2 列出的图像尺寸为 335×880(共 294800 像素)，所以该算法在目标占比为整个图像平面的 1‰(约 300 像素)时可以得到分割结果，而在占比为 4‰以上时分割结果较为可靠。图 10 展示了两组成功分割的最小像素目标。对比图 4 中的各类别像素面积统计结果，除钥匙串和打火机以外，所有类别物品的像素面积均值和最小面积均在 1000 像素以上。所以这两个类别的分割结果出现漏检的概率较高。

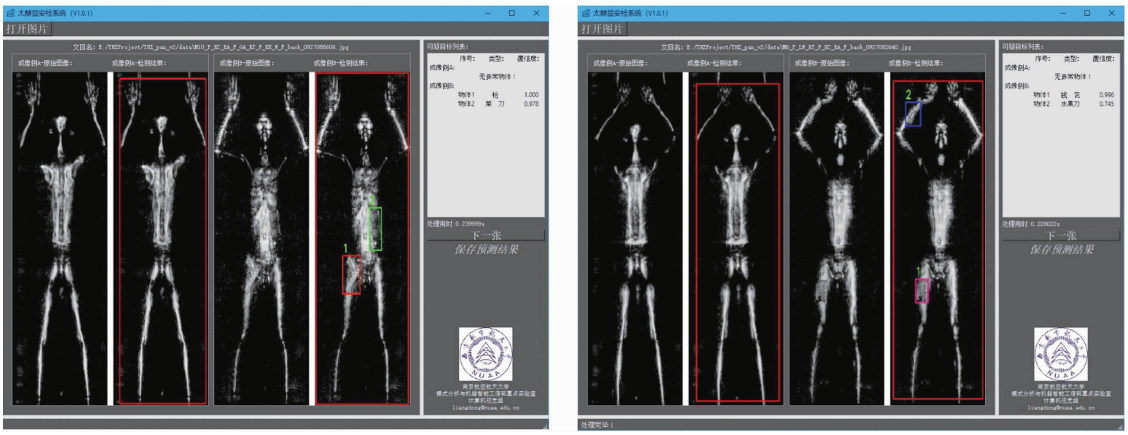


图 7 RetinaNet 的检测结果

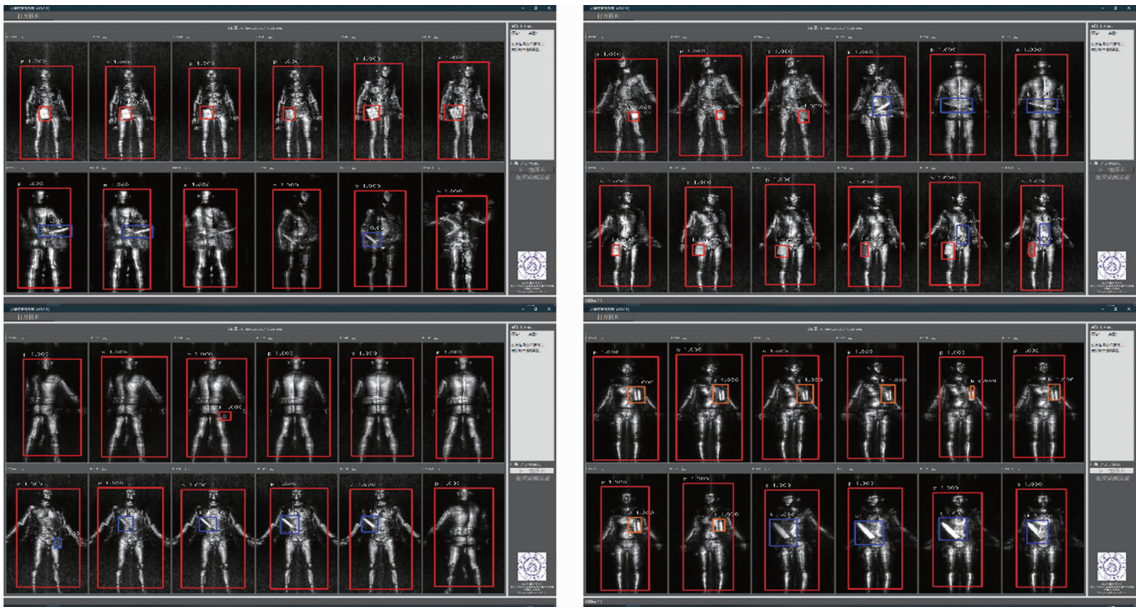


图 8 不同视角 THz 成像的检测结果

为了评估检测算法在目标不同像素面积下的性能，统计了测试集的检测框像素面积的分布情况，如图 11 所示。其中绿色表示测试集人工标注分类框的像素面积的个数统计，红色表示 RetinaNet 检测的目标框的像素面积的个数统计。当前图像的目标类别与检测结果相同且两者检测框的面积接近时，认为检测成功，并在标注结果对应的检测结果面积上累计；若检测失败，则只累计标注结果的分类框面积。图 11 中某一目标包围框像素面积对应的柱状图都是红色，则表示该面积范围内的所有目标都被成功地检测出来。从统计结果可以看出，虽然数据集目标的检测框面积分布离散，但是 RetinaNet 对各个尺度的目标都可以进行有效检测。

在单次成像中，可能存在藏匿物品垂直于成像平面的情况，使得物品成像面积变小或不完整，如图 12 所示。由于物品放置在成像垂直面的训练样本较少，且人工准确标注困难，分割和检测精度会降低。由此可见，采用多视角检测在应用中是十分必要的。这样能有效避免垂直于成像平面放置的物品的漏检，有利于在实际应用中提高检测精度。

3 总结

本文基于条件生成对抗网络机制构建了 Mask－CGANs 目标分割网络，并基于 Reti-

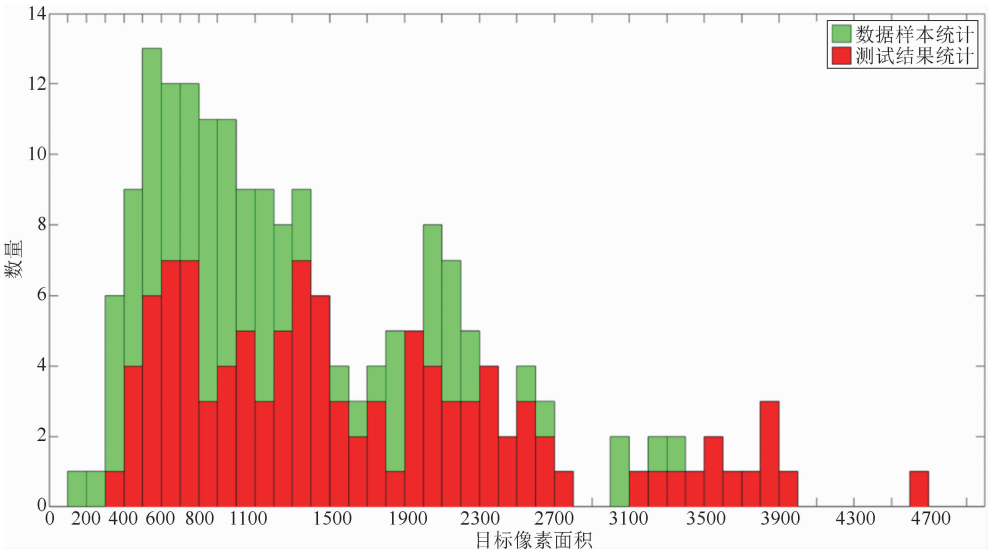


图 9 测试集的目标分割像素统计

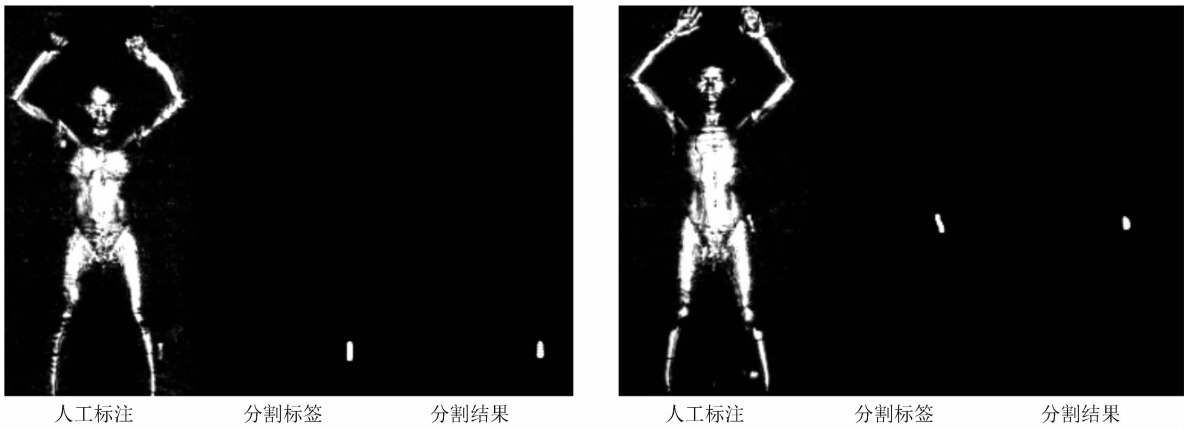


图 10 小目标分割示例(300 像素)

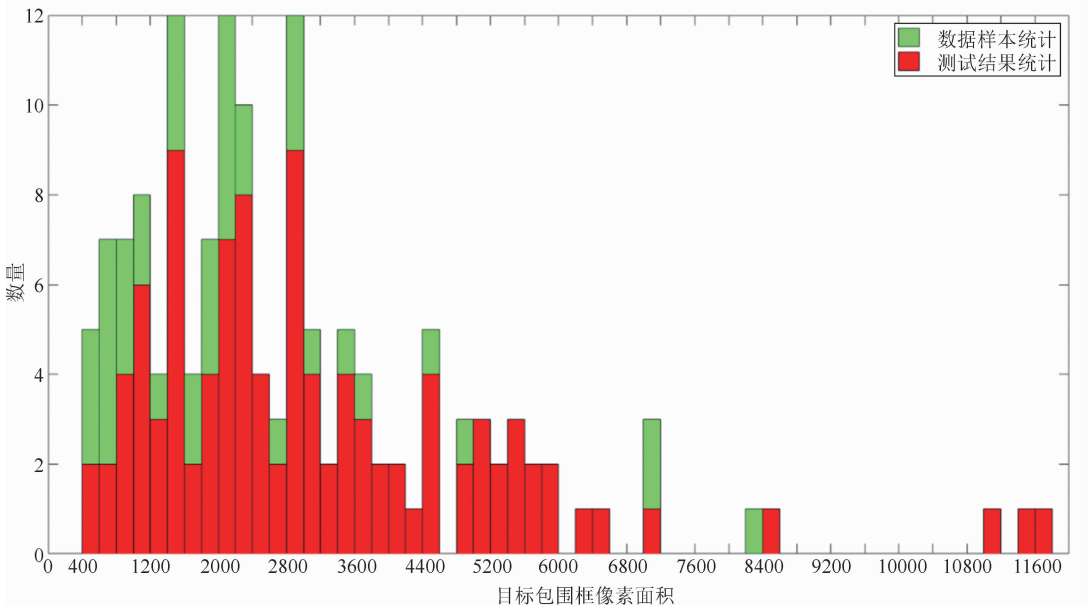
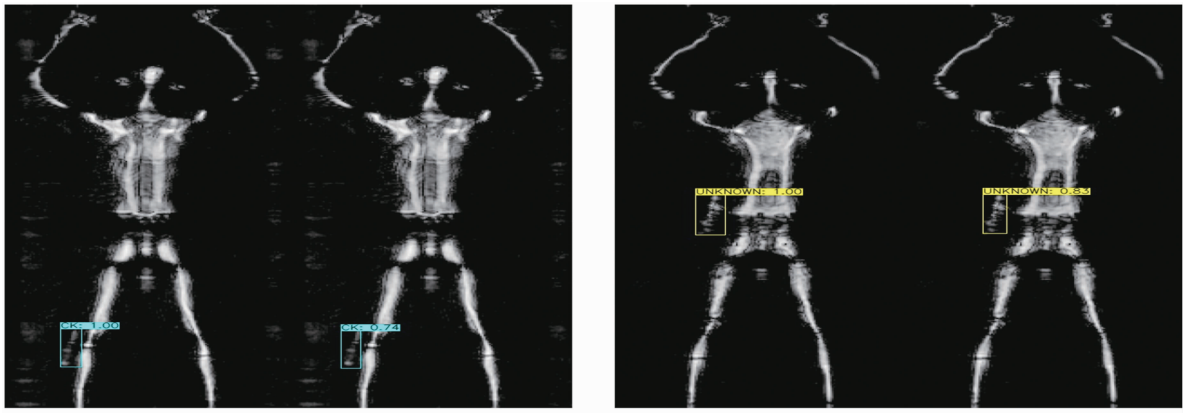


图 11 测试集的目标检测像素统计



检测1 (腿部侧面) 检测2 (腰部侧面)

图 12 小目标检测示例

naNet 构建了目标检测识别网络，做到了端到端的训练和预测，实现了 THz 图像中的多目标分割和检测识别。结果表明，针对该任务特点设计的损失函数和网络结构，不仅能够实现较为精确的分割和检测识别，而且还具有较快的处理速度，使这种方法能够应用于安检系统。未来的工作重点将是进一步探索主动 THz 图像中的微小物品分割和检测方法。

参考文献

[1] 那宏越，周德亮，姜寿禄，等. 一种改进的太赫兹主动成像系统 [J]. **南京大学学报(自然科学)**，2016，**52**(5)：23-30.

[2] Shen X，Dietlein C，Grossman E N，et al. Detection and Segmentation of Concealed Objects in Terahertz Images [J]. *IEEE Transactions on Image Processing*，2008，**17**(12)：2465-2475.

[3] Lee D，Yeom S，Son J，et al. Automatic Image Segmentation for Concealed Object Detection Using the Expectation-Maximization Algorithm [J]. *Optics Express*，2010，**18**(10)：10659-10667.

[4] Yeom S，Lee D，Son J，et al. Real-Time Outdoor Concealed-Object Detection with Passive Millimeter Wave Imaging [J]. *Optics Express*，2011，**19**(3)：2530-2536.

[5] Dai J，He K，Sun J，et al. Instance-Aware Semantic Segmentation via Multi-Task Network Cascades [C]. Las Vegas：29th IEEE Conference on Computer Vision and Pattern Recognition，2016.

[6] Pinheiro P O，Lin T，Collobert R，et al. Learning to Refine Object Segments [C]. Amsterdam：14th European Conference on Computer Vision，2016.

[7] Bai M，Urtasun R. Deep Watershed Transform for Instance Segmentation [C]. Honolulu：30th IEEE Conference on Computer Vision and Pattern Recognition，2017.

[8] Arnab A，Torr P H. Pixelwise Instance Segmentation with a Dynamically Instantiated Network [C]. Honolulu：30th IEEE Confernce on Computer Vision and Pattern Recognition，2017.

[9] Girshick R，Donahue J，Darrell T，et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C]. Columbus：27th IEEE Conference on Computer Vision and Pattern Recognition，2014.

[10] He K，Gkioxari G，Dollar P，et al. Mask R-CNN [C]. Venice：2017 IEEE International Conference on Computer Vision，2017.

[11] 姚家雄，杨明辉，朱玉琨，等. 利用卷积神经网络进行毫米波图像违禁物体定位 [J]. **红外与毫米波学报**，2017，**36**(3)：354-360.

[12] 王崇剑，孙晓玮，杨克虎. 一种用于主动式毫米波图像的低复杂度隐匿物品检测方法 [J]. **红外与毫米波学报**，2019，**38**(1)：32-38.

[13] Goodfellow I，Pougetabadie J，Mirza M，et al. Generative Adversarial Nets [C]. Montreal：28th Annual Conference on Neural Information Processing Systems，2014.

[14] Isola P , Zhu J Y , Zhou T , et al. Image-to-Image Translation with Conditional Adversarial Networks [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[15] Ronneberger O, Fischer P, Brox T, et al. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]. Munich: 18th International Conference on Medical Image Computing and Computer Assisted Intervention, 2015.

[16] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]. Montreal: 29th Annual Conference on Neural Information Processing Systems, 2015.

[17] Girshick R. Fast R-CNN [C]. Santiago: 2015 IEEE International Conference on Computer Vision, 2015.

[18] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[19] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [M/OL]. https://arxiv.org/abs/1804.02767v1, 2018.

[20] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [C]. Amsterdam: 14th European Conference on Computer Vision, 2016.

[21] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [C]. Venice: 2017 IEEE International Conference on Computer Vision, 2017.

[22] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[23] Li W, Mahadevan V, Vasconcelos N, et al. Anomaly Detection and Localization in Crowded Scenes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(1): 18-32.

[24] Shrivastava A, Gupta A, Girshick R, et al. Training Region-Based Object Detectors with Online Hard Example Mining [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.