文章编号:1001-9014(2021)06-0738-11 DOI:10. 11972/j. issn. 1001-9014. 2021. 06. 006

利用截面序列多级特征全局关联性的毫米波图像隐匿物检测

何婉婷[1,2], 张 铂[1,2], 王 斌[1,2], 孙晓玮[3], 杨明辉[3], 吴晓峰[1,2*]
(1. 复旦大学 电磁波信息科学教育部重点实验室,上海,200433;
2. 复旦大学 信息学院智慧网络与系统研究中心,上海,200433;
3. 中国科学院上海微系统与信息技术研究所 中科院太赫兹固态技术重点实验室,上海,200050)

摘要:基于毫米波图像的隐匿物检测技术在无接触式人体安检中具有重要意义。目前,毫米波设备已实现三维成像,但隐匿物检测算法通常将其简单压缩为二维图像进行目标检测,未能充分利用图像深度方向的信息。针对这一问题,提出一种毫米波图像隐匿物检测框架,将三维图像视为截面序列并充分利用其截面内特征沿序列(即深度方向)的内在逻辑关系。该框架由卷积神经网络与长短时记忆网络构成,前者用于提取截面的粗细粒度特征,后者用于提取上述特征沿深度方向的全局关联性,实现特征级信息融合,从而提高隐匿物二维定位准确率。实验结果表明,与现有主流毫米波图像隐匿物检测方法相比,所提模型能大幅提高检测精度。

关 键 词:毫米波图像; 三维图像; 目标检测; 深度学习; 卷积神经网络; 长短时记忆网络
中图分类号:TP751 **文献标识码**: A

Concealed Object Detection in Millimeter Wave Image Based on Global Correlation of Multi-level Features in Cross-section Sequence

HE Wan-Ting[1,2], ZHANG Bo[1,2], WANG Bin[1,2], SUN Xiao-Wei[3], YANG Ming-Hui[3], WU Xiao-Feng[1,2*]
(1. Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China;
2. Research Center of Smart Networks and Systems, School of Information Science and Technology, Fudan University, Shanghai 200433, China;
3. Key Laboratory of Terahertz Technology, Shanghai Institute of Microsystem and Information Technology, Shanghai 200050, China)

Abstract: The concealed object detection in millimeter wave (MMW) image is of great significance in non-contact body inspection. At present, MMW radar has been able to obtain 3D images, which are simply compressed into 2D images in current methods in general. However, such a rough processing does not take the information along the depth direction into account which results in a bottleneck of detection accuracy. To address this issue, a novel framework for MMW image concealed object detection is proposed, in which a 3D image is regarded as a sequence of 2D cross-sectional images and the most of the internal logic relations of features in the crosss-sectional images can be explored along the sequential direction, i. e. the depth direction of the 3D image. The framework consists of a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network. The former is used to extract the multiscale features in each 2D cross-sectional image while the latter is used to explore the global correlation of the above features along the depth direction to achieve feature-level information fusion and improve the accuracy of 2D location prediction. Experimental results show that the proposed method achieves remarkable results comparing to the known detection method based on 2D MMW images.

收稿日期:2021- 03- 08,**修回日期**:2021- 04- 26 **Received date**:2021- 03- 08,**Revised date**:2021- 04- 26
基金项目:国家自然科学基金(61731021)
Foundation items:Supported by National Natural Science Foundation of China (61731021)
作者简介(Biography):何婉婷(1995-),女,广东揭阳人,硕士研究生,主要研究领域为毫米波图像目标检测。 E-mail: 18210720030@fudan. edu. cn
*通讯作者(Corresponding author): E-mail: xiaofengwu@fudan. edu. cn

Key words: millimeter wave image, three-dimensional image, object detection, deep learning, convolutional neural network, long short-term memory network
PACS:84. 40. Xb

引言

毫米波是指波长在1~10 mm之间的电磁波[1],介于可见光和微波之间,毫米波目标检测技术具有非接触、非电离、可探测多种物质等优点,非常适合人体安检任务,有逐步取代传统安检技术的趋势[2]。由于机场、火车站等人流密集的场所对安检效率要求较高,因此,为毫米波图像设计自动检测算法,准确、快速地检测人体违禁物是非常迫切且必要的[3]。按照工作方式,毫米波成像可分为被动毫米波成像与主动毫米波成像。被动毫米波成像设备不发射毫米波,只对目标辐射出的毫米波进行成像,通常成像质量较差;主动毫米波成像设备依靠系统自身发射毫米波,并接收目标的回波信号进行成像,不易受环境因素的影响,不仅成像质量好,还能实现三维成像,获得更加丰富的被测物信息。

(a)　(b)

图1　毫米波成像设备(a) Sim-Image系统,(b)违禁物示例

Fig. 1　MMW imaging system (a) prototype of Sim-Image, (b) examples of contraband

本文所用数据全部由主动式设备——毫米波全息成像系统Sim-Image采集,仪器原型以及实验所采用的违禁物如图1所示。该系统由中国科学院上海微系统与信息技术研究所研发[4-5],成像过程如图2所示:系统的收发单元构成平面阵列,记为xy平面(x为水平方向,y为垂直方向),阵列发射单元向被测人体发射波段在28 GHz到33 GHz的毫米波信号,随后借助成像算法对接收单元检测到的回波信号进行二维图像重建,二维重建图像平行于阵列平面,沿深度z方向形成等间隔序列,从而得到人体三维毫米波散射强度分布。由于实际检测任务中只需获取隐匿物在xy平面的二维坐标,无需进行三维坐标定位,因此为了降低检测难度与标注成本,通常采用最大值投影法将原始三维数据投影到二维空间,对二维图像进行目标检测。

目前,基于毫米波图像的隐匿物检测仍是一项艰巨的任务。一方面,安检任务中的隐匿物具有纹理形状各异、空间尺度较小等特点;另一方面,毫米波图像相比光学图像空间分辨率较低、本底噪声较大,使得隐匿物特征难以提取。对于以上难点,传统毫米波图像目标检测方法依赖手工设计特征[6-8],但受限于设计者先验知识与特征参数量的不足,手工设计的特征不足以有效地表达隐匿物特性。而随着深度学习的发展,许多基于卷积神经网络(Convolutional Neural Network,简称CNN)的模型取得了良好的性能[9-11]。CNN是一种强大的特征提取器,能编码图像从低阶到高阶的各级语义特征,使得所提取到的特征更具区分力。最近,针对毫米波图像中隐匿目标物尺寸较小的特点,主要从以下两个方面设计模型:1)扩大子图正负样本的特征差别以应对其在训练中出现的样本不均衡问题,如采用IoG(Intersection over Ground-truth)指标划分样本[12]、引入Focal loss设计损失函数[13]等;2)采用不同扩展比例的空洞卷积[14],获得目标的多尺度特征与上下文信息,从而提高对小尺寸违禁物的检测精度。

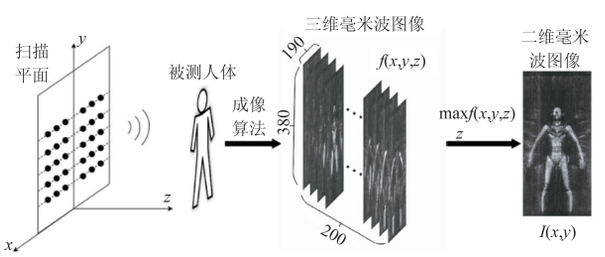


图2　毫米波成像过程示意图

Fig. 2　Diagram of MMW imaging system

然而,主流的毫米波图像隐匿物检测方法均没有充分利用毫米波原始数据的三维空间信息。相比三维毫米波图像,二维图像各像素只保留了三维空间中同一平面位置、不同深度上的最大回波强度值,导致隐匿物空间纹理特征的扭曲,更丢失了深度方向上回波波形蕴含的目标信息,不利于隐匿物的检出。为提高隐匿物检测的准确性,必须充分考虑三维毫米波图像的深度方向上提供的信息。其

中一个解决方案是采用三维卷积核进行特征提取，但该方法不适用于三维毫米波图像，主要原因在于三维毫米波图像在深度方向的物理有效分辨率低于其他两个方向[15]，由于不同维度的信息密度不同，三维卷积核难以从具有各向异性分辨率的体素中学习到有效的特征。此外，三维卷积的计算量与内存量都极高，模型所需计算时间长，与安检任务对检测速度的要求相悖。

最近，视频内容分类中的研究表明[16]，长短时记忆网络（Long Short-Term Memory，简称LSTM）[17]能够提取图像帧在时序上的相关性，可以融合帧级特征得到视频级特征描述。受这一思路启发，考虑三维毫米波图像的分辨率各向异性与沿深度方向强度分布相关性，我们引入长短时记忆网络来设计面向三维毫米波图像的检测方法。具体地，我们的模型将原始图像看作z轴（深度方向）上的xy截面序列，首先利用卷积神经网络对各二维截面进行特征提取，获得包含丰富的纹理-语义信息的粗细粒度特征。随后，利用长短时记忆网络沿深度方向整合上下文信息，获得具有深度方向全局关联信息的特征描述，并基于该特征预测隐匿物在xy平面的二维坐标。

综上所述，本文提出一种面向三维毫米波图像的隐匿物检测框架，用较小的代价尽可能地充分利用毫米波图像的三维空间信息，以提高隐匿物二维坐标预测的准确率。相较于现有的毫米波图像隐匿物检测方法，所提议模型避免了在特征提取之前进行图像信息融合的做法，同时引入长短时记忆网络提取截面图像在深度方向上的逻辑关系，从而充分利用三维毫米波图像的空间信息。更进一步地，所提议模型实际上是将xy平面与z轴向的特征提取进行了分解，不仅解决了毫米波图像分辨率各向异性的问题，同时相比采用三维卷积核进行三维目标检测的方法具有更小的计算代价。实验结果表明，该框架能有效地利用三维毫米波图像沿深度方向的全局关联信息，检测性能相比现有方法有了大幅度的提升。

1 三维毫米波图像隐匿物检测框架

1.1 毫米波图像分析

目前，毫米波设备已可以实现三维成像，但随后通常将三维成像结果投影到二维空间进行检测。上述操作实际上是进行了像素级的信息融合，只保留了三维空间中同一平面位置、不同深度方向上的

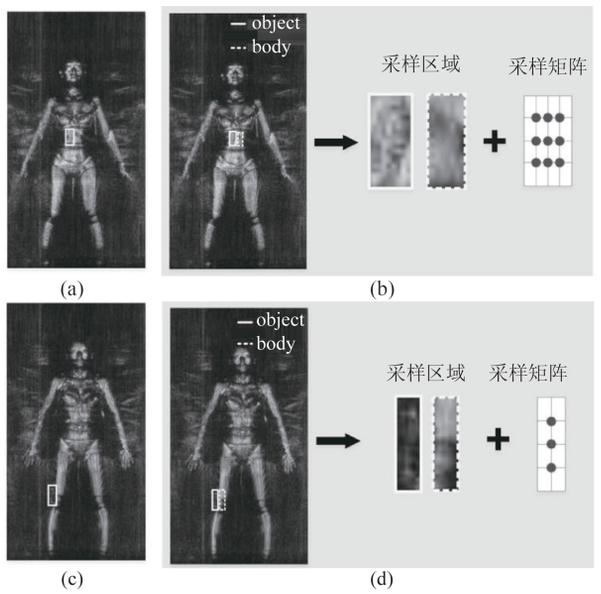


图3 毫米波图像分析示例（a）、（c）第一、二类困难样本二维毫米波图像及真值框，（b）、（d）采样过程示意图

Fig. 3 Examples of MMW image processing (a), (c) 2D MMW image with ground truth of difficult case 1 and 2, (b), (d) illustration of the sampling process

最大回波强度值，但丢失了深度方向上回波所蕴含的目标信息，使得部分隐匿物变得模糊、难以辨认，称为困难样本，具体可分为两类：1）当隐匿物被携带于人体正面，且纹理、亮度与人体特征十分接近时，视觉上难以区分目标与背景，如图3（a）所示；2）当细长类隐匿物（如枪支、铁棍等）被携带于体侧时，正视图中隐匿物空间尺寸较小，难以分辨，如图3（c）所示。为了进一步分析上述困难样本，我们分别对目标物及其周围背景区域进行采样，观察二者在深度方向的强度分布，采样方式如图3所示，由于第二类困难样本中隐匿物宽度较小，故只进行3点采样。

图4展示了第一类困难样本的分析结果。其中，图4（a）和（b）展示了相应采样区域内不同采样点的沿深度方向强度分布曲线；图4（c）为含目标区域与背景区域内所有像素点的平均强度沿深度方向的分布曲线，由图可知二者的最大强度值差距较小，因此投影到二维空间后难以从视觉上区分目标与背景。但是，由于目标物与人体躯干处于不同的深度平面，因此含目标区域与背景区域沿深度方向的强度分布具有差异。如图4（c）所示，含目标区域与背景区域的最大强度出现在不同深度处，视觉上表现为二者在不同深度平面上被“激活”，如图4（d-f）所示。图5展示了第二类困难样本的分析结果。

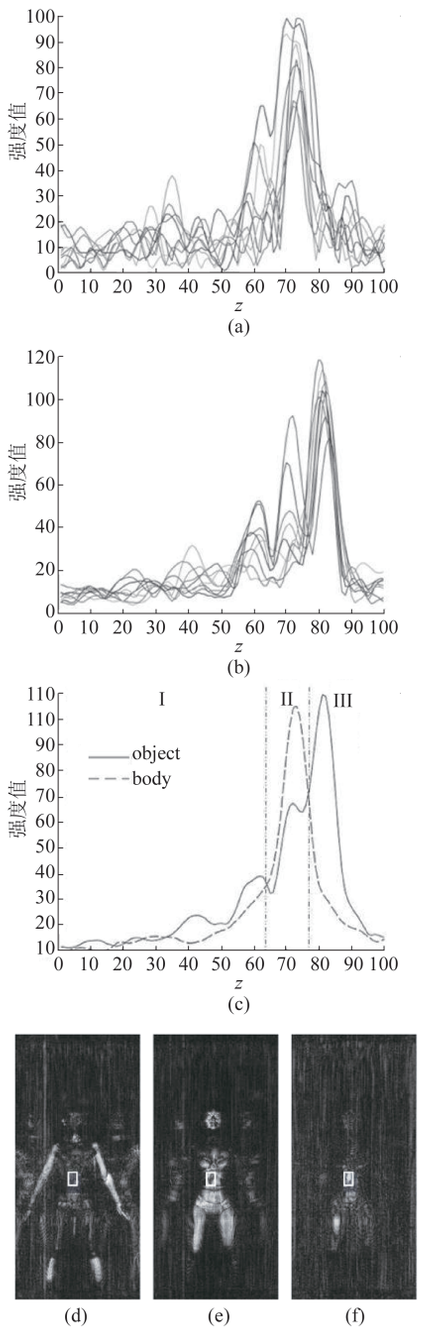


图4　第一类困难样本分析结果（a）含目标区域采样点沿深度方向的强度分布曲线，(b）背景区域采样点沿深度方向的强度分布曲线，(c）背景与含目标区域沿深度方向的平均强度分布曲线比较，(d-f）I、II、III深度区间对应截面示例

Fig. 4　Analysis results of difficult case 1 intensity distribution curve of sampling point along depth direction of (a) object and (b) background, (c) comparison of the average intensity distribution curve of the object and the background area along the depth direction, (d-f) example of MMW cross-section corresponding to interval I, II and III

其中，图5(a)和(b)为采样区域内不同采样点的沿深度方向强度分布曲线。对于这类样本，其正视图中目标物尺寸往往较小，例如，图3(c)中的目标物宽度仅为6个像素，检测难度较大；而在深度方向，目标物在48–68区间内都有比较显著的强度水平，如图5(a)所示，因此，其侧视图的视觉可探测性远大于正视图。此外，通过对比图5(b)，可以发现目标物与周围人体躯干在深度方向的分布不同，这是检测算法区分目标物与周围背景的内在原因。因此，对于上述两类困难样本，充分利用毫米波图像深度方向所提供的的信息有利于区分目标与背景。

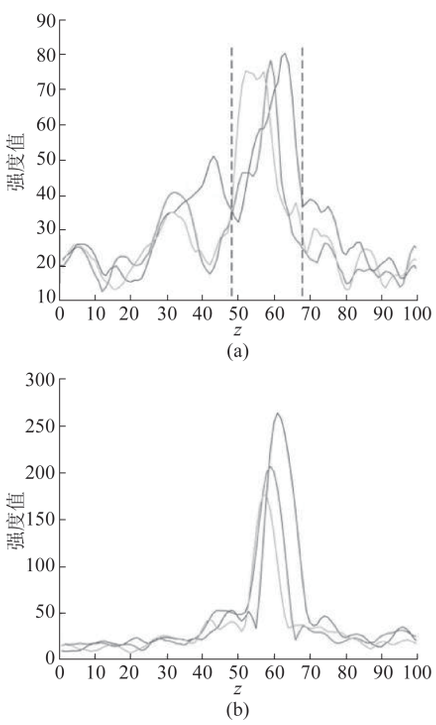


图5　第二类困难样本采样点**沿深度方向**的强度分布曲线（a）含目标区域，(b）背景区域

Fig. 5　Analysis results of difficult case 2 intensity distribution curve of sampling point along depth direction of (a) object and (b) background

除人体躯干外，我们对其他背景区域进行了采样分析，这些区域的主要强度响应为成像噪声。图6展示两组图像的分析结果。其中，图6(a)和(e)为示例图像及其采样情况，对图示采样区域进行9点采样；图6(b–d, f–h)分别为相应采样区域内不同采样点的沿深度方向强度分布曲线。由结果可见，不同于目标物，噪声在深度方向不具有相关性。

综上所述，三维毫米波图像深度方向所包含的

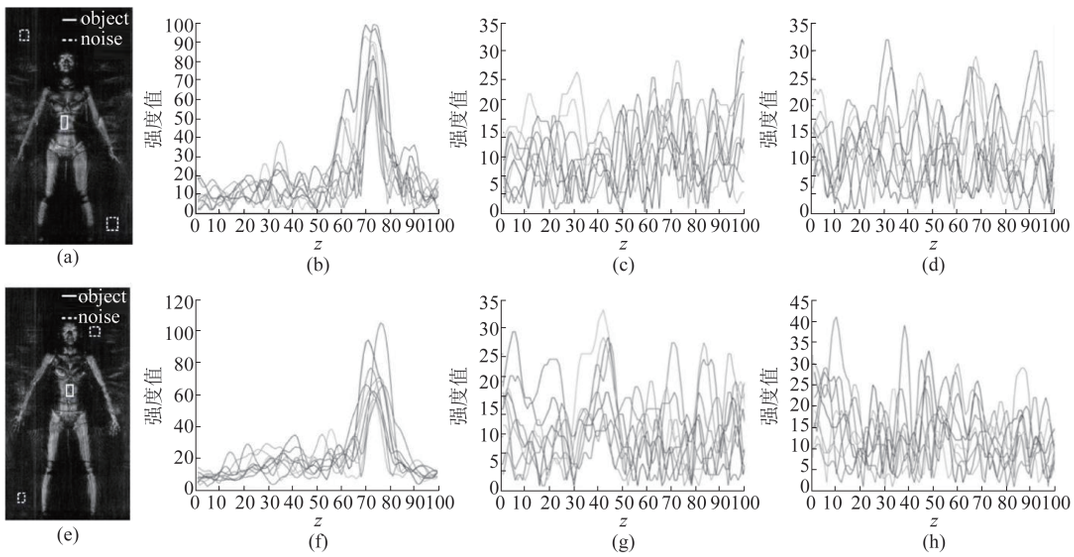


图6 目标物与噪声强度分布对比(a)、(e)二维毫米波图像及采样区域,(b)、(f)含目标区域采样点在深度方向的强度分布曲线,(c-d)、(g-h)噪声区域采样点在深度方向的强度分布曲线

Fig. 6 Comparison of intensity distribution between object and noise (a) (e) 2D MMW image with sampling area, intensity distribution curve in depth direction of (b) (f) object, and (c-d)、(g-h) noise

信息对于隐匿物的检测具有重要作用,由于目标物与人体躯干、噪声等背景在深度方向的分布不同,对深度方向信息的充分利用是本文提高检测精度的主要思想。

1.2 模型构建

1.2.1 框架整体结构

目前,基于深度学习的目标检测模型通常由两个部分构成:第一,特征提取模块,负责从原始图像提取具有一定表达能力的特征描述;第二,预测模块,负责根据所提取特征以及监督信息,进行目标类别与坐标的预测。对于毫米波图像隐匿物的检测,一方面,对原始图像的像素级信息融合无法充分利用深度方向上回波波形蕴含的目标信息,导致隐匿物空间纹理的扭曲,因此需要设计新的特征提取模块,以充分考虑毫米波图像的三维空间信息;另一方面,由于隐匿物三维边界框(bounding box)的预测难度大、标注成本高,加之实际应用中三维边界框并非必要,因此本任务仍然属于二维检测问题,模型所预测与优化的对象仍为二维边界框。我们考虑,毫米波图像的三维空间信息可以理解为二维空间结构沿深度方向的变化关系,故引入长短时记忆网络以挖掘上述变化关系,实现特征级信息融合,从而提高在二维空间的检测精度。

基于以上观察与分析,提出了一个新颖的三维毫米波图像目标检测框架,该框架由三个部分组成:1)截面内特征提取模块;2)截面间上下文提取模块(Bi-CLSTM);3)预测模块。其中,模块一、模块二构成特征提取器,获得毫米波截面序列全局性特征;模块三基于上述特征,进行毫米波图像违禁物品的检测。模型的整体结构如图7所示。设给定三维毫米波图像$I^{N_x \times N_y \times N_z}$,本文方法将该图像作为一个长度为$N_z$的$N_x \times N_y$二维截面序列输入模型,具体流程为:首先,对于每个截面,由模块一进行二维空间特征提取,得到大小为$M_x \times M_y \times C$的特征图(Feature Map)。其中,$M_x \times M_y$为特征图尺寸,C为其通道数。随后,上述特征图序列送入模块二,该模块由N_z个Bi-CLSTM单元(Unit)构成,能实现截面间的上下文信息提取与特征级信息融合,最终得到大小为$M_x \times M_y \times C$的特征图。最后,模块三基于上述特征进行预测,输出隐匿物的二维边界框(中心坐标+尺寸大小)、置信度与类别概率。

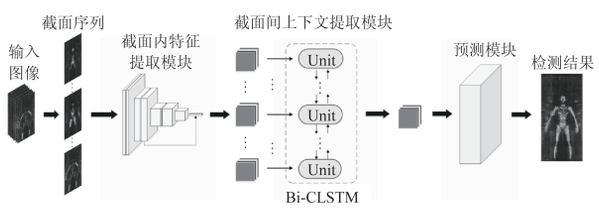


图7 所提议方法的整体框架

Fig. 7 The overall framework of the proposed method

1.2.2 截面内特征提取模块

对于毫米波图像截面的特征提取,由于安检任务中隐匿物的种类繁多、形状各异,且尺寸往往较

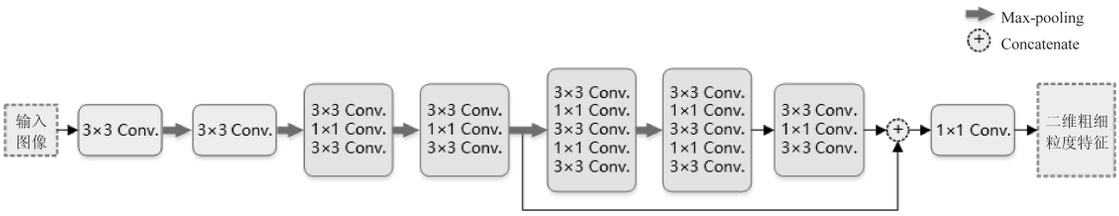


图8　截面内特征提取模块结构示意图

Fig. 8　Structure of intra-section context extraction module

小,因此基于手工设计的方法难以提取目标特征。针对以上特点,借鉴YOLO-v2[19]中的特征提取网络来设计深度卷积网络,实现毫米波二维截面的特征提取。具体地,通过堆叠2个、3个的3×3卷积层,分别获得5×5、7×7的等效感受野,上述操作不仅能降低计算量,同时可以增加网络深度[20];此外,在上述堆叠层中插入1×1卷积,用以压缩参数,并增强网络的非线性表达能力[21]。我们将上述结构作为子卷积网络块,通过组合子卷积网络块来构建深度卷积网络,具体结构如图8所示,网络总层数为21。此外,针对毫米波图像中目标物体尺寸较小的特点,考虑网络浅层特征的细节信息更加丰富,我们抽取浅层特征与最后的深层特征进行拼接(Concatenate),使得最终得到的特征同时具有细粒度纹理信息与粗粒度语义信息,从而提高特征的表达能力。

1.2.3　截面间上下文提取模块

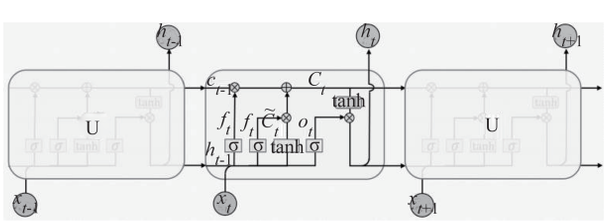


图9　LSTM的基础结构[17]

Fig. 9　Commom structure of LSTM[17]

基于毫米波图像的特点,我们引入长短时记忆网络提取图像深度方向的逻辑关系。LSTM是一种特殊的循环神经网络,被证明可以很好地处理序列问题[18,22]。网络中隐含层的节点之间互相连接,信息得以在节点间进行传递,能高效地提取序列间的相关性。图9展示了LSTM的结构,其中,x为输入的序列,U代表LSTM中的基础单元,每个单元结构相同。可以看到,LSTM中有两个运输信息的“传送带”,分别是细胞状态c与隐藏状态h,其中,细胞状态c保留大部分的主干信息,解决了经典循环神经网络中的长依赖问题。LSTM有三个特殊的门,用来控制信息的通过与否,分别是遗忘门f、输入门i与输出门o。其中,遗忘门控制着细胞状态C信息的遗忘,从而释放不必要的资源,使网络能更好地学习;输入门控制着细胞状态c上新信息的输入;更新了细胞状态后,由输出门输出结果h_t。这三个门的门控由当前时刻的输入x_t与前一时刻的隐藏状态h_{t-1}共同决定。整个过程可由下式表示:

$$\begin{cases} h_t = o(x_t,h_{t-1})\cdot\tanh(c_t) \\ c_t = f\left(x_t,h_{t-1}\right)\cdot c_{t-1} + i\left(x_t,h_{t-1}\right)\cdot\tanh\left(x_t,h_{t-1}\right) \end{cases} \tag{1}$$

然而,传统LSTM模型的输入为一维向量,而模块一输出的特征图具有二维空间结构。为了保证上述二维特征的空间信息不丢失,本文将LSTM改进为卷积LSTM(Convolutional LSTM, CLSTM)网络,即采用卷积操作替换经典LSTM中的点乘,使得步骤一中得到的二维特征可以直接输入LSTM,而不用转换为一维向量。同时,由于步骤一所提取的特征已经足够高阶,CLSTM的卷积操作均采用1×1卷积核。此外,在经典的LSTM中,信息只往一个方向传递,即只能提取一个方向的上下文信息,为了更充分地利用毫米波图像的深度方向信息,本文采用双向的CLSTM(Bi-Directional CLSTM, Bi-CLSTM),以提取深度方向双向的上下文信息,有利于模型性能的提升。设输入为序列$z_i, i=1,2,\cdots,N_z$,则对于Bi-CLSTM模块,第i个单元的输出Y_i可表示为:

$$Y_i = h_i^+ \oplus h_i^- \tag{2}$$

其中,h_i^+表示z^+方向计算的CSLTM隐层输出,h_i^-表示z^-方向计算的CSLTM隐层输出,$\oplus$表示对上述两者进行拼接。h_i^+与h_i^-的具体计算过程与式(1)相同。

通过以上方式,本文将毫米波截面序列全局性特征描述的提取分解为两个步骤:首先,采用深度卷积网络提取各个截面的二维截面内特征表达;随后,利用Bi-CLSTM模块提取截面间上下文关系表

达。由于模块一所提取的特征具有丰富的空间纹理信息与高阶语义信息，不仅能提高Bi-CLSTM模块对深度方向的建模能力，更因为前者的特征向量已经编码了邻域信息，因此Bi-CLSTM模块通过融合上述特征向量，从而获得可以表征毫米波三维空间关系的特征描述。

1.2.4 预测模块及损失函数

目前，深度学习检测方法中的预测模块可分为两大类：1)两阶段结构，其预测通常为串行结构，即先完成背景/前景的分类，再进行进一步分类与定位；2)单阶段结构，即不预先进行背景/前景的分类，而是并行地进行分类与定位两个任务。其中，两阶段结构的设计带来了大量的空间开销与重复计算，无法满足实时性的要求；而单阶段结构只需进行一次回归计算，模型中卷积运算的共享程度更高，内存占用小，具有明显的速度优势，因此本文基于单阶段结构来设计预测模块。

预测模块采用并行两分支结构，同时进行类别预测与边界框定位两个任务，通过构建多任务损失函数实现端到端的优化，仅需一次计算即可进行目标物体的分类与定位。采用卷积层替代传统分类网络中的全连接层，从而更好地保留目标物体的空间位置信息。此外，为提高召回率，引入anchor机制进行预测，即预先设定5个不同尺寸的先验框(anchor box)，在特征图的每个位置上对应地预测5个边界框。其中，先验框不是手工挑选得来，而是针对本研究所用的数据集，采用k-means算法进行聚类，以确定先验框的尺寸。

对于模型输入的毫米波截面序列，设模块一的特征提取函数为$F(\cdot)$，首先对每个截面z进行特征提取，得到截面粗细粒度特征序列：$F(z_i), i=1,2,\cdots,N_z$。随后，Bi-CLSTM模块提取截面间的上下文信息，其参数为$G(\cdot)$，则整个特征提取过程可表示为：$G\left(F(z_1),F(z_2),\cdots,F(z_{N_z})\right)$。基于毫米波截面序列全局性特征，预测模块在特征图的每个位置上进行回归计算，得到边界框、置信度与类别概率。针对以上预测，构建多任务损失函数，对上述任务进行联合训练。因此，模型的损失函数L_{det}由三个部分构成，分别是边界框损失L_{bbox}、置信度损失L_{conf}以及分类损失L_{class}，整体损失函数的表达式如下：

$$L_{\text{det}}(F,G)=\frac{1}{N_z}\sum_{i=1}^{N_z}\left[L_{\text{bbox}}\left(G\left(F\left(z_i\right)\right)\right)+L_{\text{conf}}\left(G\left(F\left(z_i\right)\right)\right)+L_{\text{class}}\left(G\left(F\left(z_i\right)\right)\right)\right], \quad (3)$$

$$L_{\text{bbox}}=\sum_{i=0}^{S^2}\sum_{j=0}^{k}\left(2-w_{ij}*h_{ij}\right)\left[\left(x_{ij}-\hat{x}_{ij}\right)^2+\left(y_{ij}-\hat{y}_{ij}\right)^2+\left(w_{ij}-\hat{w}_{ij}\right)^2+\left(h_{ij}-\hat{h}_{ij}\right)^2\right], \quad (4)$$

$$L_{\text{conf}}=\sum_{i=0}^{S^2}\sum_{j=0}^{k}\left(C_{ij}-\hat{C}_{ij}\right)^2, \quad (5)$$

$$L_{\text{class}}=\sum_{i=0}^{S^2}\sum_{j=0}^{k}\sum_{c\in\text{classes}}\left(p_{ij}(c)-\hat{p}_{ij}(c)\right)^2, \quad (6)$$

其中，S^2表示特征图尺寸，k表示先验框数量。对于第i个网格的第j个预测框，$(\hat{x}_{ij},\hat{y}_{ij},\hat{w}_{ij},\hat{h}_{ij})$表示其中心点坐标与宽/高度，$\hat{p}_{ij}(c)$表示类别概率，$\hat{C}_{ij}$表示置信度；相对应地，$(x_{ij},y_{ij},w_{ij},h_{ij})$、$p_{ij}(c)$与$C_{ij}$表示真值框的各项参数。针对本任务目标尺寸较小的特点，在计算边界框损失L_{bbox}时，乘以修正项$\left(2-w_{ij}*h_{ij}\right)$，即根据真值框的大小对权重系数进行修正，使得对于尺寸较小的框权重更大。训练时，为了降低优化问题的复杂性，采用解耦的优化方式对模型参数F与G进行优化。首先，优化F来保证二维空间特征的有效提取；其次，固定参数F，优化Bi-CLSTM的参数G，保证得到的毫米波截面序列全局性特征描述是毫米波图像三维空间关系的有效表达。

2 实验结果与分析

2.1 实验设计

2.1.1 实验数据

实验采用的三维毫米波数据集由毫米波全息成像系统Sim-Image采集，共包含204张图像，尺寸为190 × 380 × 200像素，图像示例如图10(a)。样本均为人体正面图，每张图像上均有1~2件违禁物，包含枪支、陶瓷刀、手机等多个种类，以各种角度携带于身上。

所提议模型将毫米波三维数据处理为z轴向的二维截面序列，具体为200个尺寸为190 × 380的二维毫米波截面。由于采集数据时人体站立位置固定，因此只有一些特定的深度通道存在有效信息。为了保证算法性能、降低计算量，实验中对不包含人体信息的截面进行舍弃。为此，我们对数据集进行统计分析，图10(d)展示了毫米波散射强度在深度方向的分布曲线，以及分布曲线的平均值。我们考虑，高于平均值的区间为可能包含有效信息的深度通道。可以发现，图10(d)中曲线存在3个强度水平显著的区间，结合三维毫米波图像沿x、y方向的投影图(图10(b)(c))，可以发现区间II、III为设备成像伪影，可以舍弃。因此，区间I(即第40~90截面)为检测算法所需的实际有效深度通道。

考虑到截面之间存在信息冗余，为降低模型复

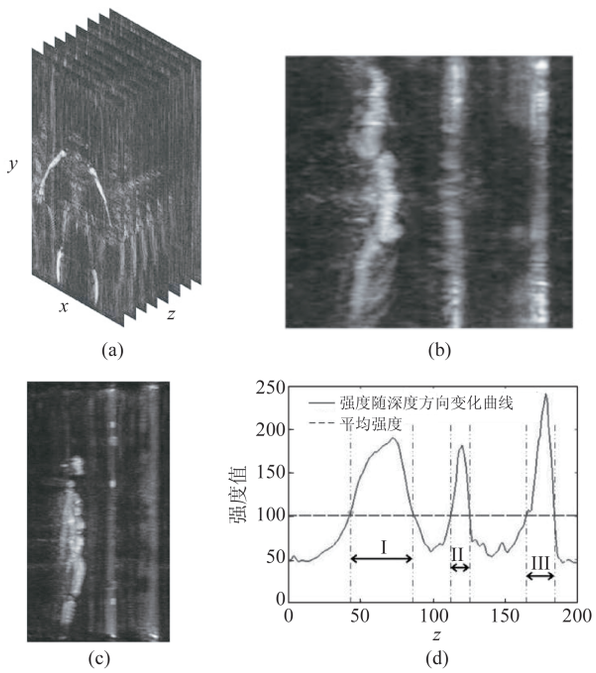


图 10　实验数据集示意图（a）三维毫米波图像,（b）沿 y 方向投影图,（c）沿 x 方向投影图,（d）数据集统计分析结果

Fig. 10　Illustration of our dataset (a) 3D MMW image, (b) the projection in the y-direction, (c) the projection in the x-direction, (d) statistical analysis result of the dataset

杂度与计算成本,采用等间隔抽样对区间 I 进行降 2 倍采样,每个三维毫米波图像抽取 25 个截面,作为算法输入。此外,由于数据集样本数量较小,我们采用 5 折交叉验证法进行实验,以保证实验结果的可靠性;训练与测试数据的划分比例为 4∶1。对于每组数据,均重复 5 次实验。

2.1.2　实验设置

实验基于 Pytorch 框架构建模型,并使用 NVIDIA TITAN XP 的单个 GPU 对模型进行训练与测试,该 GPU 具有 12GB 的内存。训练时,采用解耦训练的方式,首先训练截面内特征提取模块,以获得最优性能,其次训练 Bi-CLSTM 模块。对于模块一,使用在 ImageNet 数据上预训练的权值进行网络参数初始化,从而提高训练效率、缓解过拟合。模型采用带动量的随机梯度下降法(Stochastic Gradient Descent, SGD)进行权值更新,网络初始学习率为 0.000 1。为了进一步防止过拟合,设置权值衰减,系数为 0.000 5。对于 Bi-CLSTM 模块,采用 Adam 优化器进行权值更新,设置初始学习率为 0.000 1,权值衰减系数设置为 0.000 1。

2.2　评价指标

实验采用平均查准率均值(Mean Average Precision, mAP)作为评价指标,以评估算法的性能。其中,平均查准率(AP)表示查准率(Precision)对查全率(Recall)所取的平均。查准率与查全率是一对相互矛盾的指标,通过对预测结果的置信度设置不同阈值可得到不同的结果对,形成查准率-查全率曲线(P-R 曲线),AP 实际上是 P-R 曲线下方的面积,能更好地反映全局性能。mAP 表示各类别平均查准率(AP)的均值,由于本实验中把所有的违禁物归为同一类,故此时 mAP 与 AP 相同。

其中,查准率与查全率的定义分别如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (8)$$

TP(True Positive)、FP(False Positive)、FN(False Negative)分别表示“正确检出”、“虚警”与“漏检”。其中,所预测的边界框是否为正样本取决于其与真值框的重叠程度,用 IoU 度量(Intersection over Union),其计算方法如公式(9)。若 IoU 大于阈值,认为预测结果为正样本,否则为负样本。由于毫米波图像中的目标尺寸较小,故设置 IoU 阈值为 0.1。

$$\text{IoU} = \frac{\text{Pred} \cap \text{Truth}}{\text{Pred} \cup \text{Truth}}. \quad (9)$$

2.3　实验结果

2.3.1　与毫米波图像隐匿物检测方法结果对比

在本节,将所提议方法与常用毫米波图像隐匿物检测方法进行了对比。具体地,对基于二维毫米波图像的 YOLO-v2 和 SSD 模型[23]进行了实验。YOLO-v2 与 SSD 均属于单阶段目标检测模型,能一次性地进行边界框定位与分类预测,在自然图像数据集中取得了优越的效果。其中,SSD 采用特征金字塔(Pyramidal Feature Hierarchy)结构进行预测,从而利用多尺度特征信息,在图像分辨率较低时同样具有较好的检测性能,但其检测速度逊色于 YOLO-v2。为维持对比实验的公平性,采用相同的样本进行训练,区别只在于输入图像的处理方式不同。此外,在应用 5 折交叉验证法时,各个实验的分组相同;测试时对于正负样本划分准则也完全一致。

表 1 展示了不同方法检测准确率的结果对比。可以看出,相比于现有的面向二维毫米波图像的检测算法,本文模型的准确率有了大幅度的提高,其中,与基于二维图像的 YOLO-v2 算法相比,本文的方法提高了约 21% 的 mAP,说明三维毫米波图像深度方向信息的利用有助于检测效果的提升;除此之

表1 与常用毫米波图像隐匿物检测方法的结果对比

Table 1 Accuracy comparison with mainstream method of MMW image object detection

Detection Framework	输入图像类型	mAP/(%)	漏检率/(%)	检测时间/ms
YOLO-v2	二维毫米波图像(投影)	61.20±6.62	14.6	**9**
SSD	二维毫米波图像(投影)	73.92±2.20	11.3	52
所提议模型	三维毫米波图像	**82.34±1.43**	**5.9**	126

外,漏检率下降了8.7%,说明本文方法可以缓解由于深度方向信息未充分利用而导致的检测困难问题,从而提高检测准确率;与检测模型SSD相比,本文方法的mAP同样有将近10%的领先,进一步验证了所提议方法的优越性。此外,各模型所需的检测时间见表1最末列,其中,所提议模型的检测时间为126 ms,增加的计算开销主要由于本文采用深度截面序列毫米波图像作为输入,每个序列包含25个毫米波截面。

2.3.2 消融实验

保持所提议模型的其他结构不变的前提下,移除Bi-CLSTM模块,得到消融模型。其具体结构为:将三维毫米波图像以截面形式依次输入YOLO-v2模块,得到各个截面的特征图;对于上述特征图进行拼接(Concatenate)后使用1×1卷积,实现特征融合,并进行后续的预测,模型具体结构如图11所示。

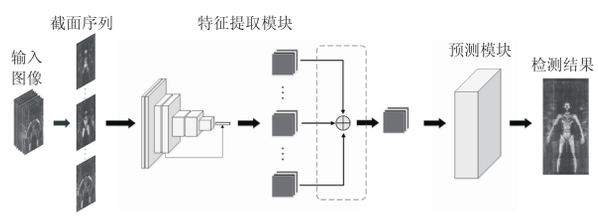


图11 消融模型的结构

Fig. 11 Structure of the ablation model

具体的检测结果如表2所示。可以看出,相比于简单融合多个截面特征,Bi-CLSTM模块的引入使得mAP提高了约6.8%。该实验说明简单地采用拼接来融合各截面特征无法很好地利用沿深度方向的空间相关性,而Bi-CLSTM能够十分有效地聚合截面间的上下文信息得到毫米波截面序列全局性特征描述,从而提高了检测性能。说明所提议模型对于毫米波图像三维空间信息的利用是十分有效的。

2.3.3 模型可解释性分析

为观察所提议模型准确率提高的原因,利用类激活映射(Class Activation Mapping,CAM)[24]分别对两种模型最后一层特征图进行可视化,观察网络进行预测时所关注的区域。为了便于比较,将二者的CAM结果以伪彩色图的形式进行表示,颜色越红的区域表示预测时所占权重越高,越蓝表示权重越低。

表2 Bi-CLSTM模块有效性验证

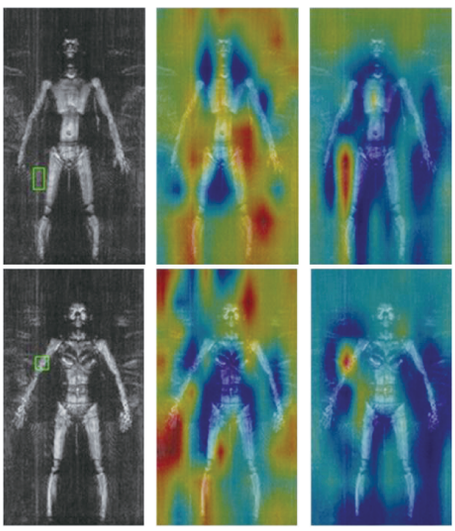
Table 2 Validation of Bi-CLSTM

Detection Framework	mAP/(%)
消融模型	75.52±2.28
所提议模型	**82.34±1.43**

图12是两组图像的可视化结果。图12(a)中,放置于身体侧面的违禁物品十分模糊,这是由于三维毫米波图像投影到二维空间时,损失了深度方向上回波波形蕴含的目标信息。因此,基于二维投影图像的YOLO-v2在进行预测时,无法很好地关注到目标所在的区域(如图12(b)所示),因此模型无法检测到这一类物体。相反,从图12(c)可以看出,由于引入的Bi-CLSTM模块能有效地提取三维上下文信息,本文所提议的模型能很好地注意到目标所在的区域,从而检测到这类物体。

为进一步解释Bi-CLSTM在所提议框架中所起的作用,对所提议模型经Bi-CLSTM处理前、处理后的截面特征图进行CAM可视化,分别观察其显著区域,如图13所示。图13(a)为待测三维图像的yz平面投影图,由于该图难以辨认目标物,选取其中一个yz截面进行观察,如图13(b)所示。图像显示该人体腰间携带有枪支,即图中黄色框线所标注区域;实验采用25个二维截面作为输入图像,其在z轴方向上的所在区域如图中蓝色框线所示。由图13(c-d)可见,经Bi-CLSTM模块处理前,各截面特征图的显著区域易受噪声干扰,无法准确地关注到待测目标;而Bi-CLSTM模块能够有效地排除噪声等干扰,从而提高预测准确性。这是由于Bi-CLSTM模块能有效地挖掘截面图像沿深度方向的逻辑关系,正确地将违禁物与人体躯干进行区分,并滤除图像中的噪声。

为了更深入地探究所提议模型的作用机制,设



(a)　(b)　(c)

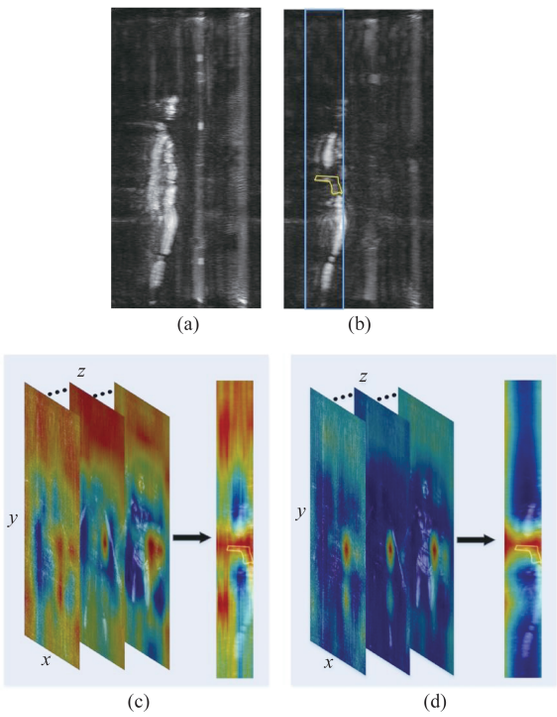
图 12　特征图CAM可视化结果（a）待测图像及真值框,（b）传统方法特征图显著区域,（c）所提议方法特征图显著区域

Fig. 12　The visualization results of the CAM of feature map (a) the image to be measured with ground truth, (b) the salient region of the feature map obtained by the traditional method, (c) the salient region of the feature map obtained by the proposed method

计了截面预测合成模型,与基于二维图像的检测模型相比,该模型舍弃了在特征提取前对图像进行像素级融合的做法。具体地,截面预测合成模型分别对各二维截面进行特征提取与预测,随后采用非极大值抑制法对各个预测结果进行融合。我们统计了不同方法所预测的候选框数量,如表3所示。由结果可知,相比基于二维毫米波图像的检测模型,截面预测合成模型所得到的候选框数量有所上升,说明相比于像素级融合,先进行特征提取再进行融合的策略能提高检测敏感性,这是所提议模型降低漏检率的前提。此外,图14比较了所提议模型与截面预测合成模型的P–R曲线(前者为实曲线,后者为虚曲线),可以发现在同一查全率下,所提议模型的查准率整体上高于截面预测合成模型,说明Bi–CLSTM模块的引入能提取截面图像在深度方向上的逻辑关系,从而降低虚警率。因此,所提议模型通过引入长短时记忆网络实现特征级信息融合,从而充分利用毫米波图像的三维空间信息,提高了检测准确率。

3　结论

提出了一种面向三维毫米波图像的隐匿违禁物品检测框架,该框架能有效利用毫米波截面序列



(a)　(b)　(c)　(d)

图 13　各截面特征图CAM可视化对比（a-b）待测三维图像侧视图,（c）经Bi-CLSTM处理前CAM结果及其侧视图,（d）经Bi-CLSTM处理后CAM结果及其侧视图

Fig. 13　Comparison of CAM visualization of feature maps of cross-section (a-b) the side view of 3D MMW image to be measured, CAM results and the side view (c) before, and (d) after Bi-CLSTM

表3　不同模型生成候选框数量对比

Table 3　Number of candidate bounding boxes predicted by different method

Detection Framework	候选框数量
基于二维图像的检测模型	6
截面预测合成模型	10
所提议模型	4

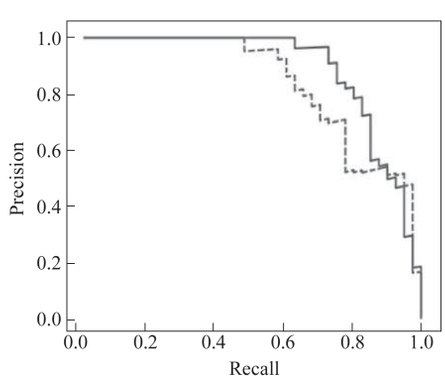


图 14　P-R曲线结果对比

Fig. 14　Comparison P-R curve

间的全局关联性，缓解了以往无法充分利用三维毫米波图像空间信息的问题。所提议框架由卷积神经网络与长短时记忆网络构成，前者用于进行二维截面的粗细粒度特征提取，后者沿深度方向整合截面间的全局关联性，实现特征级信息融合。由于隐匿物与人体躯干、成像噪声等背景区域沿深度方向具有不同分布，通过对深度方向逻辑关系的建模，可以降低漏检与虚警情况，更加准确地实现隐匿物的定位。所提议模型实际上是将毫米波图像xy平面与z轴向的特征提取进行了分解，由于长短时记忆网络中各单元间参数共享，因此能够在参数量固定的前提下实现变长序列的建模。换言之，模型可以针对不同的毫米波成像设备灵活调整长短时记忆网络的长度，以适应毫米波图像分辨率各向异性的问题。此外，相比采用三维卷积核进行三维目标检测的方法，所提议框架不仅具有更小的计算代价，而且得以在训练阶段充分利用基于二维检测任务的预训练模型，从而减少训练与标注成本。因此，我们的方法提供了一种新的思路，用较小的代价尽可能地充分利用毫米波图像的三维空间信息，以提高隐匿物二维坐标预测的准确率，相比现有的毫米波隐匿物检测方法具有更好的检测性能。

References

[1] Xiang J, Zhang M. Millimeter-Wave radar and its applications[M]. National Defense Industry Press, 2005.

[2] Ghasr M T, Ying K P, Zoughi R. Wideband millimeter wave interferrometer for high-resolution 3D SAR imaging [C]. 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2015: 925-929.

[3] Peng F, Fang W, Wen X, *et al*. State of the art and future prospect of the active millimeter wave imaging technique for personnel screening[J]. *Journal of Microwaves*, 2015, **31**(2):91-96.

[4] Zhu Y, Yang M, Wu L, *et al*. Millimeter-wave holographic imaging algorithm with amplitude corrections [J]. *Progress In Electromagnetics Research M*, 2016, **49**: 33-39.

[5] Zhu Y, Yang M, Wu L, *et al*. Practical millimeter-wave holographic imaging system with good robustness[J]. *Chinese Optics Letters*, 2016, **14**(10):43-47.

[6] Yeom S, Lee D S, Son J Y, *et al*. Concealed object detection using passive millimeter wave imaging[C]// Universal Communication Symposium. IEEE, 2010.

[7] Mu S, Shan H, Zhou J, *et al*. A method for detecting hidden objects of human body in passive millimeter wave image. *Science & Technology Information*, 2014, 36: 202-203.

[8] Nian F, Chen W, Wang W, *et al*. Concealed objects detection in active millimeter-wave images[J]. *Systems Engineering and Electronics*, 2016, **38**(6): 1462-1469.

[9] Yao J, Yang M, Zhu Y, *et al*. Using convolutional neural network to localize forbidden object in millimeter-wave image[J]. *Journal of Infrared and Millimeter Waves*, 2017, **36**(003): 354-360.

[10] Zhang B, Chen T, Wang B, *et al*. Densely semantic enhancement for domain adaptive region-free detectors[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021,DOI: 10.1109/TCSVT.2021.3069034.

[11] Zhang B, Wang B, Wu X, *et al*. Domain adaptive detection system for concealed objects using millimeter wave images [J]. *Neural Computing and Applications*, 2021, **33**:11573-11588 .

[12] Luo S, Wu X, Yang M, *et al*. Convolutional neural network based human concealed object detection for millimeter wave images[J]. *Journal of Fudan University*(*Natural Science*), 2018, **57**(4): 442-452.

[13] Liu C, Yang M, Sun X. Towards robust human millimeter wave imaging inspection system in real time with deep learning [J]. *Progress In Electromagnetics Research*, 2018, **161**: 87-100.

[14] Liu T, Zhao Y, Wei Y, *et al*. Concealed object detection for Activate millimeter wave image[J]. *IEEE Transactions on Industrial Electronics*. 2019, **66**(12): 9909-9917.

[15] Sheen D M, McMakin D L, Hall T E. Three-dimensional millimeter-wave imaging for concealed weapon detection [J]. *IEEE Transactions on microwave theory and techniques*, 2001, **49**(9): 1581-1592.

[16] Ng Y H, Hausknecht M, Vijayanarasimhan S , *et al*. Beyond short snippets: Deep networks for video classification [C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[17] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, **9**(8):1735-1780.

[18] Shahzadi I, Tang T B, Meriadeau F, *et al*. CNN-LSTM: Cascaded framework for brain tumour classification[C]// 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). IEEE, 2018, DOI: 10.1109/IECBES.2018.8626704.

[19] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *Computer Science*, 2014, arXiv:1409.1556v6 .

[21] Lin M, Chen Q, Yan S. Network In Network[J]. *Computer Science* , 2013, arXiv:1312.4400 .

[22] Patraucean V, Handa A, Cipolla R. Spatio-temporal video autoencoder with differentiable memory[J]. *Computer Science*, 2015, **58**(11):2415 - 2422.

[23] Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector[J]. *Computer Science* , 2015: 21-37.

[24] Zhou B , Khosla A , Lapedriza A , *et al*. Learning deep features for discriminative localization[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,2016:2921-29 .