文章编号:1001-9014(2021)06-0870-08

DOI:10.11972/j.issn.1001-9014.2021.06.023

高定位精度的毫米波全息图像三维目标检测

李怀乾[1,2], 杨明辉[1], 吴 亮[1*]

(1. 中国科学院上海微系统与信息技术研究所 中国科学院太赫兹固态技术重点实验室,上海 200050;
2. 中国科学院大学 材料与光电研究中心,北京 100049)

摘要:投影角度不同导致目标的形状及尺寸变化限制了基于主动式毫米波(AMMW)全息图像投影视图的隐匿物品二维检测方法对小目标检测性能的提升,为此,提出了基于点云的隐匿物品三维检测方法。通过阈值化处理将AMMW全息图像转换为点云输入经空洞卷积及多分支结构改进的SECOND三维目标检测器,提取对目标的三维几何理解及其多尺度上下文信息以提高对小目标的检测能力。实验结果表明,较基于投影的二维检测方法,该方法平均召回率(AR)提升了3.33%,有效提升了定位精度;在交并比(IOU)阈值为0.5时的检出率提升了8.75%,虚警率降低了1.78%,平均精度(AP)提升了7.11%,不同IOU阈值下的平均AP提升了4.30%,有效提升了检测精度;检测速度为17.3 FPS,达实时水平。

关 键 词:信息科学与系统科学;三维目标检测;空洞卷积;小目标;主动式毫米波全息图像

中图分类号:TP751 **文献标识码**:A

High localization accuracy 3D object detection in active millimeter wave holographic images

LI Huai-Qian[1,2], YANG Ming-Hui[1], WU Liang[1*]

(1. Key Laboratory of Terahertz Solid State Technology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China;
2. Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The difference in projection angle leads to changes in the shape and size of objects, which limits the improvement of the detection performance of small objects by the two-dimensional (2D) concealed object detection method based on projected views of active millimeter wave (AMMW) holographic images. For this reason, a three-dimensional (3D) concealed object detection method based on point clouds was proposed for the first time. AMMW holographic images were converted into point clouds by thresholding, and then were input into the 3D object detector SECOND, which was improved by dilated convolution and multi-branch structure, to extract the 3D geometric understanding of the objects and their multi-scale context information to improve the detection ability for small objects. The experimental results showed that compared with the projection-based 2D detection methods, the average recall (AR) of this method was improved by 3.33%, which proved the effective improvement of localization accuracy. The detection rate and the average precision (AP) was relatively improved by 8.75% and 7.11%, and the false alarm was reduced by 1.78% at an intersection over union (IOU) threshold of 0.5. The average AP under different IOU thresholds was improved by 4.30%. The detection accuracy was effectively improved. The detection speed was 17.3 FPS, which reached the real-time level.

Key words: information science and system science, 3D object detection, dilated convolution, small object, active millimeter wave holographic image

收稿日期:2021-01-28,**修回日期**:2021-05-27 **Received date**:2021-01-28,**Revised date**:2021-05-27

基金项目:国家自然科学基金(61731021),中国科学院科技创新重点部署项目(KGFZD-135-18-028),上海市信息化发展专项资金(201901015),上海市科委科研计划项目(19511132400),硅基氮化镓射频及毫米波器件关键工艺技术开发及测试(20DZ1100702)

Foundation items: Supported by National Natural Science Foundation of China (61731021), the Key Research Program of the Chinese Academy of Sciences (KGFZD-135-18-028), the Shanghai Municipal Commission of Economy and Informatization (201901015), the Science and Technology Commission of Shanghai Municipality (19511132400), Development and Testing of Key Process Technologies for Silicon-based GaN RF and Millimeter Wave Devices(20DZ1100702)

作者简介(Biography):李怀乾(1996-),男,山东安丘人,硕士研究生,主要研究领域为毫米波全息图像处理及目标检测
E-mail: net_owl@outlook.com

[*] **通讯作者**(Corresponding author): E-mail: wuliang@mail.sim.ac.cn

PACS:42. 30. Tz,42. 40. My,84. 40. Xb

引言

主动式毫米波(Active Millimeter Wave,AMMW)全息成像技术[1]已广泛应用于机场安检系统,对全息图像中人员携带的隐匿物品进行快速且准确的检测是该系统的关键问题之一。提升定位精度是提高安检系统效率的有效手段,对安全行业具有重要意义。然而,AMMW全息图像往往存在噪声大,分辨率低的问题,而大多数隐匿物品尺寸较小,且易于与人体背景混淆,使隐匿物品的检测与精确定位成为一项艰巨的任务[2]。此外,三维全息图像数据量庞大,难以实现实时检测。

将全息图像投影为二维视图馈入二维目标检测器进行检测可有效提高检测速度[3-5]。但同一物体因放置角度不同,投影所得二维视图的形状与尺寸差别较大,导致网络难以提取统一的特征对目标分类。在网络降采样过程中,小物体的细节逐渐减少甚至消失,导致检测器难以有效辨别小物体。而投影导致目标尺寸变小,且损失了深度信息,进一步加剧了小目标的检测难度。Liu等人[6]使用二维正视图每个点对应的深度构成深度图,联合深度图与正视图输入改进的Faster RCNN[7]进行检测,该方法引入了一定的深度信息,有效提高了检出率。Liu等人[2]引入空洞卷积[8]增大特征图分辨率,将全息图像沿14个角度投影,融合多视图检测结果,在其测试集上取得了较高的检测精度,但该方法耗时较高。

近年来,基于激光雷达点云的三维目标检测器[9-12]取得了重大突破。MV3D[9]首先将点云投影为鸟瞰图与正视图,并分别使用卷积神经网络(Convolutional Neural Network,CNN)对其提取特征,而后融合不同特征进行预测,在车辆检测任务中实现了先进的性能,但多个独立的CNN结构导致极高的计算成本,检测速度极慢。VoxelNet[10]首次将点云特征提取与边界框预测统一为一个端到端的网络框架,避免了手动提取特征,实现了检测性能的大幅提升。SECOND[11]使用稀疏三维卷积取代VoxelNet中的经典三维卷积,使网络具备了实时检测的能力。Zhu等人[12]通过抑制数据样本不均衡实现了检测精度的进一步提升。

为规避投影操作为后续目标检测带来的局限,提高对小目标的检测能力,本文首次提出将AMMW全息图像转换为点云后馈入三维目标检测器检测隐匿物品的方法。通过阈值化处理粗略提取前景图像保存为点云,降低后续处理的计算压力的同时保留物体原始的三维几何形状。由于隐匿物品检测仅涉及二分类,本文将在激光雷达点云的车辆检测任务中实现先进性能的SECOND网络引入隐匿物品检测任务。因绝大多数隐匿物品属于小目标,本文引入空洞卷积及多分支结构对SECOND进行改进,在不降低特征图分辨率的情况下提取多尺度长程上下文信息,以提高对小目标的检测精度。本文使用AMMW全息图像投影所得二维正视图上的边界框作为监督信息及输出,便于标注及可视化。为评估模型性能,本文建立了包含33881张AMMW全息图像的大型数据集。实验结果表明,较基于投影的二维检测方法,本文方法的平均召回率(Average Recall,AR)[13]提升了3. 33%,有效提升了对隐匿物品的定位精度;得益于此,在交并比(Intersection Over Union,IOU)阈值为0. 5时的检出率提升了8. 75%,虚警率降低了1. 78%,平均精度(Average Precision,AP)提升了7. 11%,不同IOU阈值下的平均AP提升了4. 30%,有效提升了隐匿物品的检测精度。检测速度为17. 3 FPS,可实现实时检测。

1　三维目标检测器

针对AMMW全息图像上的隐匿物品检测任务,本文设计的三维目标检测器框架如图1所示,整个系统由三个部分组成:输入模块、三维特征提取器及输出模块。对于给定AMMW全息图像,系统通过输入模块对其进行预处理,而后馈入三维特征提取器提取目标的三维几何特征,将生成的特征图馈入输出模块,输出模块在特征图的每个点预测该点为隐匿物品的置信度及对应的二维边界框。下文将分别对系统的三个部分展开详述。

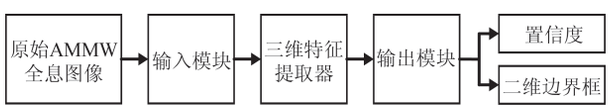

图1　AMMW全息图像隐匿物品三维目标检测器框架

Fig. 1　The structure of our proposed 3D concealed object detector for AMMW holographic images

1. 1　输入模块

本文采集的AMMW全息图像如图2(a)所示,其

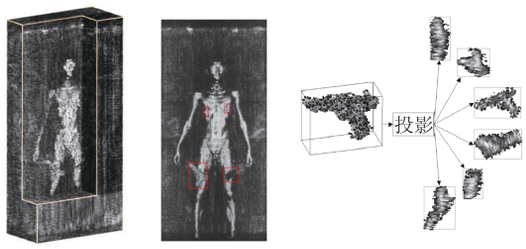

图2 AMMW全息图像投影 (a) AMMW全息图像,(b) 图(a)沿Z方向投影所得二维正视图,(c) 三维物体投影至二维视图导致形状及尺寸变化

Fig. 2 Projection of the AMMW holographic image (a) AMMW holographic image, (b) the resulting 2D front view of performing projection along the Z axis of the holographic image in Fig. 2(a), (c) the shape and size changes caused by projecting a 3D object into 2D views

分辨率为190 × 400 × 100,每个点包含空间位置坐标及反射强度四个特征。因其数据量庞大而信噪比较低,人体及隐匿物品完全淹没于噪声,若使用深度神经网络直接对其提取特征受噪声的影响较大,且计算成本极高。现存方法沿全息图像一个或多个视角投影生成二维图像作为检测系统输入,如图2(b)所示。投影显著降低了数据量,有利于实现实时检测,方便应用二维目标检测算法,但也为目标检测任务带来一定的局限性。如图2(c)所示,沿同一物体的不同角度投影将生成不同尺寸及形状的二维视图,同理,隐匿物品置于人体的角度不同,导致同一类物体投影后生成的二维视图形状及尺寸差异较大,神经网络难以提取统一的几何特征对其分类。部分投影视图与人体部位相似,对特征提取造成干扰,易造成错误检测。此外,二维视图的面积不大于三维物体的最大横截面积,投影后物体尺寸进一步降低,加剧了对小目标的检测难度。

点云是一种非结构化、无序的、稀疏的三维数据表示形式。简单地说,点云即为空间中一组点的集合。为规避投影操作导致的局限,本文将AMMW全息图像转换为点云表示以保持目标原始的三维几何形状。图3展示了本文输入模块的结构,主要包含阈值化处理及体素化操作。在AMMW全息图像中,噪声点的反射强度普遍小于前景图像,本文根据反射强度值对全息图像进行阈值化处理。具体来说,统计全息图像各点的反射强度,选取合适的统计量(如均值、分位数)作为阈值,将反射强度大于阈值的点保存为点云,实现前景图像的粗略提取。该方法滤除大量噪声点,将数据量降低近两个数量级,有效缓解了数据处理压力;同时保留了物体原始的三维几何信息,深度维度的引入为目标检测任务提供了更多信息,有利于准确提取形状特征。阈值化处理破坏了全息图像的结构化特性,为方便后续卷积运算,本文对点云进行体素化处理,即将点云均匀划分为指定尺寸的体素格,每个体素格表征其所包含的点。在SECOND中,为降低后续计算成本,体素格尺寸往往较大,但AMMW雷达点云的分辨率低,而隐匿物品尺寸极小,目标的细节信息对其检出至关重要。本文设置点云X、Y、Z方向有效范围分别为[1,192]×[1,400]×[0,100]。由于人体的正面及背面较为平整,且与电磁波发射方向垂直,因此在点云的X、Y方向保留了更多细节信息,为减少体素化导致的细节信息损失,选取X、Y、Z方向尺寸为1×1×2.5的体素格进行体素化,即仅在Z方向上进行降采样。此时,每个体素格内至多包含3个点,本文使用每个体素格中各点特征的均值作为所提取的体素级特征[12]。体素化后,稀疏、非结构化的AMMW雷达点云被转换为尺寸为192×400×41×4的紧凑、结构化的体素表示。

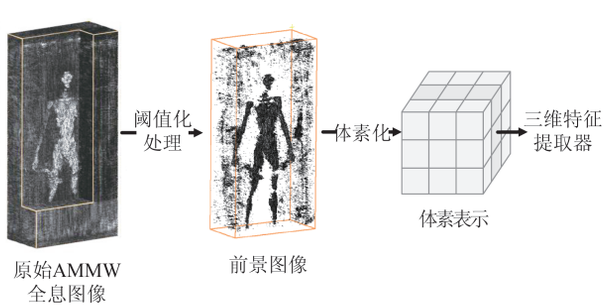

图3 输入模块结构

Fig. 3 The structure of our proposed input module

提升小目标的数据量有利于缓解网络降采样过程中小目标细节丢失的问题。与二维投影视图相比,三维点云保留了物体深度维度的信息,具有更高的数据量,更适合小目标检测任务。图4为本文数据集中隐匿物品在三维点云及二维正视图中的边界框内点的数量分布直方图,其纵坐标为数据集中具有对应点数的边界框的数量。可以看出,与二维正视图相比,三维点云中物体的数据量提升了数倍,可保留小物体更多的细节信息。此外,二维图像边界框内包含部分不属于隐匿物体的像素,而点云仅保留前景图像各点,边界框内背景噪声更少,更有利于精确定位。

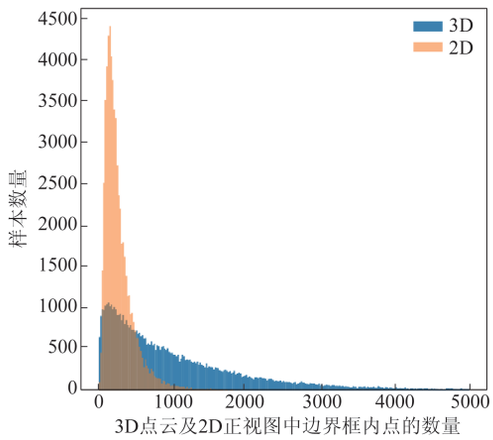

图4　三维点云及二维正视图中的边界框内点的数量分布

Fig. 4　The distribution of the number of points in the bounding box in 3D point clouds and 2D front images

1.2　三维特征提取器

SECOND网络在基于激光雷达点云的车辆检测任务中实现了先进的性能,但激光雷达点云与AMMW雷达点云的数据分布差异较大,后者包含的隐匿物品尺寸更小,形状与尺寸变化更加多样,且存在复杂的人体背景对目标的判定造成干扰。检测小尺寸目标最常用的方式是在浅层高分辨率特征图上进行预测,但浅层特征图缺乏高层次语义信息,难以区分隐匿物品与人体背景;与之相对,深层特征图包含丰富的语义信息但小目标的细节信息随网络降采样逐渐损失。SECOND网络仅根据单层特征图进行预测,难以兼顾空间细节信息与语义信息,对小目标及多尺度目标的检测能力较弱。为此,本文设计上下文信息提取模块嵌入SECOND网络,经改进的三维特征提取器如图5所示,其中降采样模块结构与SECOND结构相同,负责提取低层次空间信息;上下文信息提取模块负责提取高层次语义信息。整个三维特征提取器使用空间稀疏卷积(Spatially Sparse Convolution,SpConv)[11]与子流形空间稀疏卷积(Submanifold Convolution,SubMConv)[11]搭建,每个卷积层后应用批标准化(Batch Normalization,BN)和修正线性单元(Rectified Linear Unit, ReLU)。卷积层具体细节如表1所示,为使表示更简洁明了,当参数XYZ维度数值一致时,仅使用单个值表示。

1.2.1　降采样模块

如图5及表1所示,降采样模块接受体素化后的四维张量,通过级联的SubMConv与SpConv对特征图进行降采样,学习目标的三维空间几何信息。AMMW全息图像分辨率低,且隐匿物品仅占全息图像的极小部分,过高的降采样步长会导致物体细节逐渐丢失,严重损害检测性能。为充分保留物体的空间细节信息,设置X、Y方向降采样步长为4,特征图分辨率较高,有利于提取局部特征,更适合小目标检测。为方便后续二维卷积运算的应用,设置Z方向降采样步长为8,逐步学习Z方向信息,降低Z方向的尺寸。

1.2.2　上下文信息提取模块

在低降采样步长下,高分辨率特征图存在丰富的空间细节信息(如角点、边缘等),但缺乏高层次的语义信息以区分物体、人体与噪声,从而导致大量虚警。为此,本文将空洞卷积引入SubMConv,在

表1　三维特征提取器卷积层细节

Table 1　Details of convolutional layers of the 3D feature extractor

网络组件	卷积类型	通道	卷积核	步长	填充	空洞率	感受野	特征图
降采样模块	SubMConv	16	3	1	1	1	3	(192,400,41)
	SubMConv	16	3	1	1	1	5	(192,400,41)
	SpConv	32	3	2	1	1	7	(96,200,21)
	SubMConv	32	3	1	1	1	11	(96,200,21)
	SpConv	64	3	2	1	1	15	(48,100,11)
	SubMConv	64	3	1	1	1	23	(48,100,11)
	SpConv	128	3	(1,1,2)	(1,1,0)	1	31	(48,100,5)
分支1	SubMConv	128	3	1	1	1	39	(48,100,5)
	SpConv	128	(1,1,3)	(1,1,2)	0	1	39	(48,100,2)
分支2	SubMConv	128	3	1	(2,2,1)	(2,2,1)	47	(48,100,5)
	SpConv	128	(1,1,3)	(1,1,2)	0	1	47	(48,100,2)
分支3	SubMConv	128	3	1	(3,3,1)	(3,3,1)	55	(48,100,5)
	SpConv	128	(1,1,3)	(1,1,2)	0	1	55	(48,100,2)

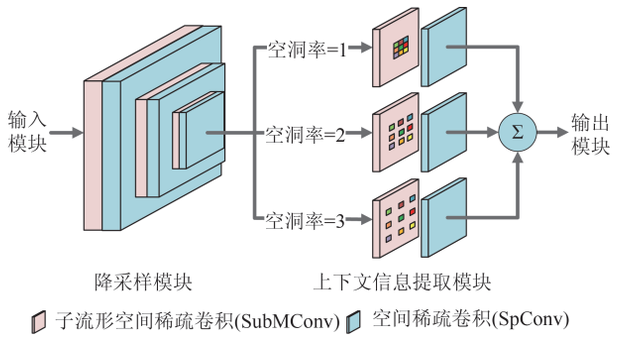

图5 三维特征提取器结构

Fig. 5 The structure of our proposed 3D feature extractor

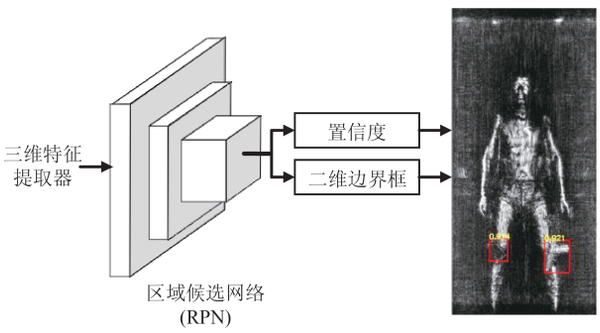

图6 输出模块结构

Fig. 6 The structure of our proposed output module

保持特征图高分辨率的情况下,增大网络感受野,为形态描述提取长程上下文信息,构建小目标与背景间的关系作为区分小目标的特征。另外,隐匿物品的尺寸变化较大,本文设置不同分支以提取目标的多尺度信息。如图5及表1所示,各个分支共享由降采样模块馈入的特征图,并应用空洞率[8]分别为1,2,3的SubMConv生成感受野分别为39,47,55的特征图以提取不同尺度上下文信息,应用SpConv在Z方向进一步降采样至深度为2,其中各分支的空洞率根据实验设置。不同分支的特征图通过逐体素等比例叠加实现特征融合,生成尺寸为48×100×2×128的包含多尺度上下文信息的高分辨率特征图馈入输出模块。

1.3 输出模块

如图6所示,输出模块接受三维特征提取器生成的四维特征图,合并其深度与通道维度,变换为尺寸为48×100×256特征图后馈入区域候选网络(Region Proposal Network,RPN)[10],通过级联的二维卷积运算,实现分类任务及边界框回归任务,即预测特征图各点包含隐匿物品的置信度及其对应的二维边界框,判定置信度高于阈值的边界框作为网络的最终输出。本文使用AMMW全息图像沿Z方向投影所得正视图上的二维边界框作为监督信息,方便标注与可视化,与基于投影的二维检测方法相比,未给数据采集带来额外的工作量。

分类任务中,使用Focal Loss[14]作为损失函数以降低样本不均衡的影响,如式(1)所示:

$$L_{\mathrm{FL}}(p,p^*) = -\alpha p^*(1-p)^{\gamma}\log(p) - (1-\alpha)(1-p^*)p^{\gamma}\log(1-p) \quad , \quad (1)$$

其中,p为模型输出预测的置信度,p^*为真值,α及γ为Focal Loss的超参数。本文取$\alpha=0.25$,$\gamma=2$。边界框回归任务中,对边界框偏移进行编码,如式(2.1)-(2.5)所示:

$$t_x = (x - x_{\mathrm{a}})/d_{\mathrm{a}}, t_y = (y - y_{\mathrm{a}})/d_{\mathrm{a}} \quad , \quad (2.1)$$

$$t_w = \log(w/w_{\mathrm{a}}), t_h = \log(h/h_{\mathrm{a}}) \quad , \quad (2.2)$$

$$t_x^* = (x^* - x_{\mathrm{a}})/d_{\mathrm{a}}, t_y^* = (y^* - y_{\mathrm{a}})/d_{\mathrm{a}} \quad , \quad (2.3)$$

$$t_w^* = \log(w^*/w_{\mathrm{a}}), t_h^* = \log(h^*/h_{\mathrm{a}}) \quad , \quad (2.4)$$

$$d_{\mathrm{a}} = \sqrt{x_{\mathrm{a}}^2 + y_{\mathrm{a}}^2} \quad , \quad (2.5)$$

其中,d_{a}表示锚框对角线长度,t与t^*分别表示预测边界框及对应边界框真值(ground-truth,GT)相对于锚框的偏移,x,y,w及h分别表示边界框中心坐标、宽度及高度,x,x_{a}及x^*分别表示预测值,锚框值及真值(y,w及h亦同)。本文使用SmoothL1作为回归任务的损失函数,如式(3-4)所示:

$$L_{\mathrm{reg}}(t,t^*,p^*) = \sum_{i\in\{x,y,w,h\}} p^*\mathrm{SmoothL1}(t_i - t_i^*), \quad (3)$$

$$\mathrm{SmoothL1}(x) = \begin{cases} 0.5x^2, & if\ |x|<1 \\ |x| - 0.5, & \mathrm{otherwise} \end{cases} \quad , \quad (4)$$

系统最终损失函数为:

$$L_{\mathrm{total}}(t,t^*,p,p^*) = \beta_1 L_{\mathrm{FL}}(p,p^*) + \beta_2 L_{\mathrm{reg}}(t,t^*,p^*), \quad (5)$$

其中,β_1与β_2用于平衡分类任务与回归任务的权重,本文取$\beta_1=1.0$,$\beta_2=0.2$。通过后向传播算法最小化该损失函数,可实现网络对数据的拟合与泛化。

2 实验

2.1 数据集

实验建立了大规模数据集以验证方法的有效性。为模拟真实场景,本文准备了60多件物品,包括多种枪、刀具、打火机、粉末、液体瓶、手机等。模特选择一个或多个物品置于身体各个部位,经毫米波雷达扫描重建其全息图像。本数据集共包含几十名模特,涵盖不同性别与体型。全息图像沿Z方向投影得二维正视图,用于标注隐匿物体边界框及可视化。实验采集了33 881张图像,物体边界框的边长范围为[2, 72],超过60%的边界框面积小于

256像素,超过90%的边界框面积小于1 024,即绝大多数隐匿物品为小目标(面积小于32^2)。实验划分31 609张图像为训练集,由不同模特采集的2 272张图像为测试集。

2.2　实施细节

对于给定的AMMW全息图像,统计各点反射强度的95%分位数进行阈值化处理。本文使用Adam优化器及one-cycle策略[15]对网络进行训练,最大学习率为0.000 5,除法因子为10,权值衰减为0.01,Batchsize为16,共训练30个epoch。设置锚框的宽度及高度分别为w_a=12.32,h_a=15.60。当一个锚框与某个GT具有最高IOU或IOU超过0.1时,判定该锚框为正样本;当一个锚框与所有GT的IOU均小于0.01时,判定该锚框为负样本;否则,判定该锚框为无关样本。整个训练在4块GTX Titan XP上完成。

2.3　评价指标

本文绘制了检出率(Recall)关于IOU阈值的函数曲线并计算AR以评估系统的定位性能[16],使用AP评估系统的检测性能。依据置信度对所有输出降序排序,根据排序结果计算精确度(Precision)-检出率(Recall)曲线(PR曲线),曲线下的面积即为AP。精确度precision及检出率recall计算公式如下:

$$\text{precision} = \text{num}_{TP}/(\text{num}_{TP} + \text{num}_{FP}) \quad , \quad (6)$$

$$\text{recall} = \text{num}_{TP}/\text{num}_{GT} \quad , \quad (7)$$

其中,num_{TP},num_{FP}和num_{GT}分别表示预测正确、预测错误及真实边界框的数量。由于隐匿物品尺寸多数较小,较低的IOU亦可接受[2]。实验计算IOU阈值为0.1、0.2、0.3、0.4及0.5时的AP,分别记为AP_{10}、AP_{20}、AP_{30}、AP_{40}及AP_{50};计算IOU阈值为0.5时的虚警率(即1-precision)及检出率,分别记作FA_{50}及Re_{50},并计算在1块GTX Titan XP上检测单帧全息图像的检测速度。

2.4　检测器性能分析

2.4.1　三维特征提取器组件分析

为验证高分辨率特征图及上下文信息提取模块的重要性,本文进行了消融实验,实验结果如表2所示。实验的基线模型是SECOND网络,其AP_{50}为72.28%;通过降低降采样步长,在高分辨率特征图上提取特征,AP_{50}相对于基线提升了1.59%;通过引入上下文信息提取模块,AP_{50}相对于基线提升了1.57%;当同时引入高分辨率特征图及上下文信息模块时,AP_{50}相对于基线提升了2.16%。证明高分辨率特征图及上下文信息对于隐匿物品检测具有积极作用。

表2　不同网络结构的AP对比

Table 2　Comparison of the AP with different network structures

网络结构	上下文信息提取模块	X、Y方向降采样步长	AP_{50}
SECOND[11]	×	8	72.28
SECOND + HRF	×	4	73.87
SECOND + CIE	√	8	73.85
SECOND + HRF + CIE	√	4	**74.44**

"HRF"表示通过降低降采样步长获得的高分辨率特征图,"CIE"表示本文提出的上下文信息提取模块

"HRF" denotes high resolution features obtained by reducing the down-sampling stride, and "CIE" denotes our proposed context information extraction module

2.4.2　与其他方法的对比

为评估定位性能,本文在图7(a)中绘制了不同网络的召回率关于IOU阈值的函数曲线,并在图例中给出对应的AR。更高的召回率及AR表明网络具有更优的定位性能[15]。可以看出,本文方法在不同IOU阈值下均取得了最高的召回率;与基于投影的方法相比,基于三维点云的方法在不同IOU阈值下均取得了更高的召回率,证明了基于三维点云的方法具有更优的定位性能。本文方法取得了35.98%的AR,较Faster RCNN提升了3.33%,较SECOND提升了1.60%,有效提升了网络的定位精度。

表3为本文方法与Faster RCNN版本的RPN[7]、Faster RCNN、RetinaNet[14]及TridentNet[17]的检测性能对比。图7(b)为不同网络在IOU阈值为0.5时的PR曲线。可以看出,本文方法取得了最高的AP_{50},达74.44%,较Faster RCNN提升7.11%,较SECOND提升2.16%。与TridentNet相比,在IOU阈值为0.1~0.5时,SECOND分别提升了1.18%,1.85%,2.80%,4.87%及7.41%的AP,平均AP提升3.63%;本文方法分别提升了1.59%,2.14%,2.95%,5.24%及9.57%的AP,平均AP提升4.30%。说明相比于基于投影的方法,基于三维点云的方法有效提高了检测精度,且随着IOU阈值的提高,提升效果愈加明显。与SECOND相比,本文方法在不同IOU阈值下AP取得了普遍提升,证明了高分辨率特征图及上下文信息提取模块的有效性。

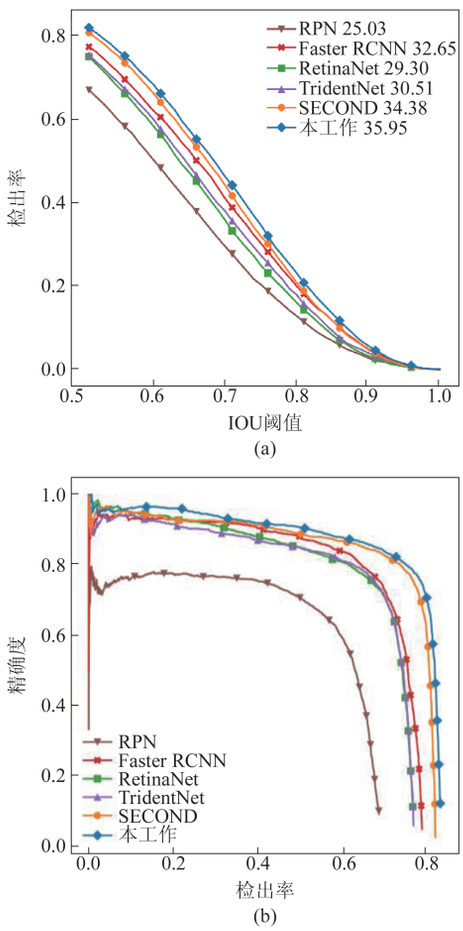

图7 不同网络的定位及检测性能对比 (a) 召回率关于IOU阈值的函数,(b) IOU = 0.5时的PR曲线

Fig. 7 Comparison of localization and detection performance for different networks (a) Recall as a function of IOU threshold, (b) PR curve under IOU = 0.5

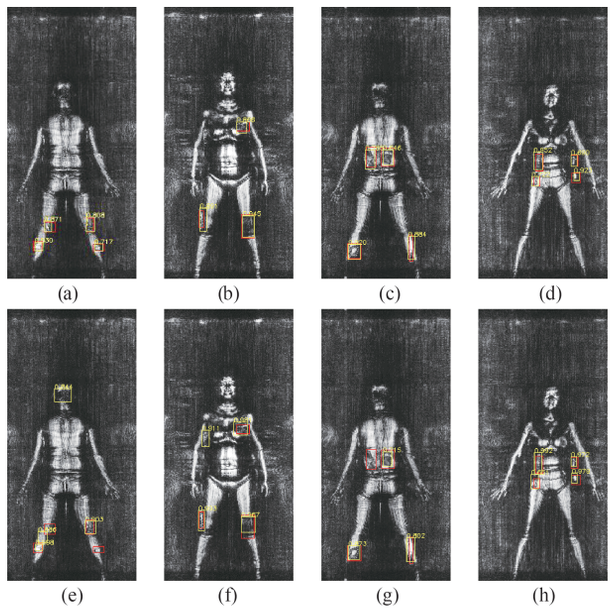

图8 不同网络检测结果示意图,其中红色边界框表示真值,黄色框代表预测值 (a)-(d) 本文方法,(e) RPN,(f) Faster RCNN,(g) RetinaNet,(h) TridentNet

Fig. 8 Qualitative detection results of different networks, where the red bounding boxes denote the ground-truth, and the yellow bounding boxes denote the predicted results (a)-(d) Our proposed method, (e) RPN, (f) Faster RCNN, (g) RetinaNet, (h) TridentNet

本文方法实现了17.3 FPS的检测速度,略慢于RPN及SECOND,但仍可实现实时检测,更好地实现了速度-精度的权衡。部分检测结果如图8所示,其中第一行为本文方法的输出,第二行为其他网络针对对应图像的输出。

2.4.3 检出率与虚警率

置信度阈值用于判断是否保留检测图像输出的边界框[2]。为确定最优置信度阈值,本文绘制了F1-score关于置信度阈值变化的曲线,如图9所示。当设置系统的置信度阈值取0.42时,F1-score取得最大值。在IOU阈值为0.1时,定位精度较低,本文方法在测试集上实现了86.57%的检出率及9.84%的虚警率;在IOU阈值为0.5时,定位精度较高,本文方法实现了76.26%的检出率Re_{50}及20.96%的虚警率FA_{50}。不同网络Re_{50}及FA_{50}的对比如表3所示。本文方法实现了最高的Re_{50}及最低的FA_{50},与SECOND相比,Re_{50}提升了1.61%,有效提高了检出率;与Faster RCNN相比,Re_{50}提升了8.75%,大幅提升了检出率,同时FA_{50}降低了1.78%,有效降低了虚

表3 不同网络的检测性能对比

Table 3 Comparison of detection performance on different Networks

网络	输入	AP_{10}	AP_{20}	AP_{30}	AP_{40}	AP_{50}	平均AP	FA_{50}	Re_{50}	速度(FPS)
RPN[7]	2D	82.35	79.26	74.15	64.26	48.50	69.70	36.68	57.99	**23.0**
Faster RCNN[7]	2D	88.88	87.53	84.76	78.71	67.33	81.44	22.74	67.51	4.7
RetinaNet[14]	2D	89.65	88.76	86.20	80.11	65.44	82.03	24.59	67.18	9.0
TridentNet[17]	2D	91.38	89.90	87.07	80.10	64.87	82.66	23.52	67.04	4.1
SECOND[11]	3D	92.56	91.75	89.87	84.97	72.28	86.29	21.24	74.65	22.9
本工作	3D	**92.97**	**92.04**	**90.02**	**85.34**	**74.44**	**86.96**	**20.96**	**76.26**	17.3

警率,证明了本文方法的有效性。

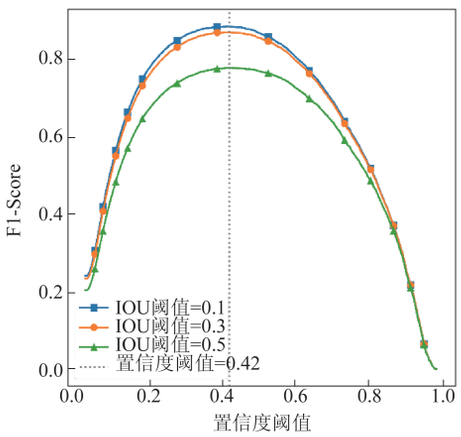

图9　不同置信度阈值下的F1-score

Fig. 9　F1-score under different thresholds of the confidence

3　结论

本文首次提出了基于AMMW三维点云的隐匿物品检测方法,在保留物体原始三维空间几何信息的同时,增大了小目标的数据量,改善了现存方法使用二维投影图像存在的目标特征不一致及小目标细节损失等问题;并引入空洞卷积及多分支结构改进了SECOND网络,在保证特征图分辨率的同时提取长程上下文信息,提高对小目标的检测能力。实验结果表明,该方法的AR提升了3.33%,有效提升了对隐匿物品的定位精度;在IOU阈值为0.5时,检出率提升了8.75%,虚警率降低了1.78%,AP提升了7.11%,不同IOU阈值下平均AP提升了4.30%,有效提升了检测精度;检测速度为17.3 FPS,可实现实时检测,证明了基于三维点云的隐匿物品检测方法的优越性。

References

[1] 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.

[2] 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.

[3] Zheng L, Yingkang J, Zongjun S, *et al*. A synthetic targets detection method for human millimeter-wave holographic imaging system [C]//2016 7th International Conference on Cloud Computing and Big Data (CCBD). IEEE, 2016: 284-288.

[4] YAO Jia-Xiong, YANG Ming-Hui, ZHU Yu-Kun, *et al*. Using convolutional neural network to localize forbidden object in millimeter-wave image [J]. *Journal of Infrared and Millimeter Waves*(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品定位。**红外与毫米波学报**), 2017, **36**(3): 354-360.

[5] Wang C J, Sun X W, Yang K H. A low-complexity method for concealed object detection in active millimeter-wave images [J]. *Journal of Infrared and Millimeter Waves*, 2019, **38**(1): 32-38.

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

[7] Ren S, He K, 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.

[8] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-05-03) [2021-01-26]. https://arxiv.org/abs/1511.07122.

[9] Chen X, Ma H, Wan J, *et al*. Multi-view 3d object detection network for autonomous driving[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 1907-1915.

[10] Zhou Y, Tuzel O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4490-4499.

[11] Yan Y, Mao Y, Li B. SECOND: Sparsely embedded convolutional detection [J]. *Sensors*, 2018, **18**(10): 3337.

[12] Zhu B, Jiang Z, Zhou X, *et al*. Class-balanced grouping and sampling for point cloud 3d object detection [EB/OL]. (2019-08-27) [2021-01-26]. https://arxiv. org/abs/1908.09492.

[13] Hosang J, Benenson R, Dollar P, *et al*. What Makes for Effective Detection Proposals? [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(4): 814-830.

[14] Lin T Y, Goyal P, Girshick R, *et al*. Focal loss for dense object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(2): 318-327.

[15] Smith L N, Topin N. Super-convergence: Very fast training of neural networks using large learning rates[C]//Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. International Society for Optics and Photonics, 2019, **11006**: 1100612.

[16] Gidaris S, Komodakis N. Locnet: Improving localization accuracy for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 789-798.

[17] Li Y, Chen Y, Wang N, *et al*. Scale-aware trident networks for object detection[C]//Proceedings of the IEEE international conference on computer vision. 2019: 6054-6063.