

## 近场毫米波三维成像与异物检测方法

师 君\* 阙钰佳 周泽南 周远远 张晓玲 孙铭芳

(电子科技大学信息与通信学院 成都 611731)

**摘 要:** 主动式毫米波阵列3维成像系统是人体安检成像系统的研究热点, 该文对主动式毫米波阵列3维系统工作模式、信号模型和成像算法进行了介绍, 并将深度学习中的卷积神经网络(CNN)热图检测方法和边框回归检测技术应用于人体安检成像异物检测。研究表明, 基于热图的检测方法和基于YOLO的检测方法均可实现异物检测。基于热图的检测方法网络结构简单、易训练, 但由于需要遍历整幅待检测图像, 运算时间长, 且生成的检测框尺寸固定, 无法适应异物尺寸变化。基于YOLO的检测算法网络结构复杂、训练耗时长, 但该方法在检测速度与检测框精度上优势明显, 更利于机场安检等对实时性要求较高的检测应用。

**关键词:** 近场毫米波3维成像; 后向投影; 卷积神经网络; 图像检测; 边框回归

中图分类号: TN957.52

文献标识码: A

文章编号: 2095-283X(2019)05-0578-11

DOI: 10.12000/JR18089

**引用格式:** 师君, 阙钰佳, 周泽南, 等. 近场毫米波三维成像与异物检测方法[J]. 雷达学报, 2019, 8(5): 578-588. doi: 10.12000/JR18089.

**Reference format:** SHI Jun, QUE Yujia, ZHOU Zenan, *et al.* Near-field millimeter wave 3D imaging and object detection method[J]. *Journal of Radars*, 2019, 8(5): 578-588. doi: 10.12000/JR18089.

## Near-field Millimeter Wave 3D Imaging and Object Detection Method

SHI Jun\* QUE Yujia ZHOU Zenan ZHOU Yuanyuan

ZHANG Xiaoling SUN Mingfang

(School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract:** Active mm-wave linear-array 3D imaging system has become one of the active research areas in the field of imaging for human security. In this paper, the operating mode, signal model, and imaging algorithm are introduced. Deep learning algorithms, including the Convolutional Neural Network (CNN) with heat map and You Only Look Once (YOLO) network, were used for the object detection of human security image. The results show that the method based on heat map and YOLO can both detect foreign objects. We find that the CNN with heat map has a simple network construction and can be easily trained, but the detection process needs to traverse the whole image, which is relatively time-consuming, and the size of the detection region cannot adapt to the objects. On the contrary, though with a relatively complex construction, YOLO network has advantages in terms of detection efficiency and accuracy. Furthermore, the size of the detection region can adapt to the objects, which is more suitable for the human security imaging application.

**Key words:** Near-filed millimeter wave 3D imaging; Back projection; Convolutional Neural Network (CNN); Objects detection; Boxing regression

收稿日期: 2018-10-22; 改回日期: 2019-07-02; 网络出版: 2019-08-23

\*通信作者: 师君 shijun@uestc.edu.cn

\*Corresponding Author: SHI Jun, shijun@uestc.edu.cn

基金项目: 国家自然科学基金(61671113)

Foundation Item: The National Natural Science Foundation of China (61671113)

责任主编: 陈杰 Corresponding Editor: CHEN Jie

### 1 引言

近年来, 对于人口流动性大以及人员密集区域, 如车站, 机场, 海关等, 如何有效地预防恐怖袭击事件显得尤其重要。穿透衣物、包裹发现隐藏的危险物品是安检成像中的关键因素和难点, 包括成像时间、清晰度、识别违规物品的准确率等都是安检系统中的重要指标。传统安检安防手段, 如X射

线, 存在电离辐射的风险, 长期暴露在X射线下对于人体有相当大的危害<sup>[1]</sup>。而手持扫描探测器、安检门则存在穿透性、准确性的问题。相比之下, 高频微波, 如毫米波、太赫兹具备良好的安全性和穿透性, 能够对隐匿的可疑危险物品进行较为清晰地成像且对人体健康性危害较小, 已经逐渐成为最主流的人体安检成像方式<sup>[2]</sup>。毫米波成像安检系统由于具备分辨率适中、成像清晰度和对比度合适、对人体影响较小等特点, 成为具有极大潜力的新一代人体安检系统, 近些年以近场毫米波3维成像为代表的主动式毫米波安检成像技术得到了高速发展<sup>[3-5]</sup>。

根据系统工作模式以及原理上的区别, 毫米波成像又分为主动式毫米波成像和被动式毫米波成像系统。主动式毫米波成像系统通过布置收发阵元, 发射电磁波信号到目标场景, 并接收目标反射的回波信号, 最后结合成像算法对目标场景进行3维重建, 进而获取较高精度的成像场景。主动式成像系统具备受环境因素影响小、能获取丰富的3维场景成像信息、成像分辨率高等特点<sup>[6]</sup>, 成为安检成像技术研究的重要方向。主动式毫米波安检成像算法主要包括后向投影(Back Projection, BP)算法以及距离多普勒(Range Doppler, RD)算法等。

在刀、枪等异物检测方面, 由于毫米波安检成像质量受到系统指标、隐私等方面的约束, 无法或不宜获得过于高清晰度的图像, 如何从安检图像中检测敏感目标成为毫米波安检系统的关键问题。传统检测方法通过在图像中提取特征(如, 方向梯度直方图(Histogram of Oriented Gradient, HOG)<sup>[7]</sup>、局部二值模式(Local Binary Patterns, LBP)<sup>[8]</sup>、Harr特征<sup>[9]</sup>等)并结合Adaboost、支持向量机(Support Vector Machine, SVM)等分类器进行异物检测<sup>[10]</sup>。这些技术需要对毫米波图像特点有较深的分析, 且算法的通用性较差, 当成像质量较差时尤其明显。深度卷积神经网络是目前图像识别和检测技术中的热点技术, 该技术通过大量样本训练,

自动寻找图像中的特征, 并进行识别, 具有算法通用性好、开发灵活等优点。

自从AlexNet<sup>[11]</sup>, VGG<sup>[12]</sup>等深度卷积神经网络在图片识别任务取得优异成绩之后, 卷积神经网络开始广泛应用在目标检测算法中<sup>[13-17]</sup>。Girshick等人<sup>[15]</sup>提出了区域卷积网络(Regional Convolutional Neural Network, RCNN)算法, 先用选择性搜索方法(selective search)提取候选区域, 然后使用卷积神经网络对候选区域进行特征提取, 提取的特征由分类器进行分类, 并通过边框回归(box regression)得到目标位置。文献<sup>[14]</sup>在快速区域卷积网络(Fast RCNN)<sup>[13]</sup>的基础上提出了区域生成网络(Region Proposal Network, RPN), 通过交替训练的方式, 该网络可以与Fast RCNN共享网络参数, 提高了训练效率, 节省了训练成本, 检测准确率也得到了进一步的提升。为了进一步减少提取候选框时间损耗, 提高检测速度, YOLO(You Only Look Once)算法<sup>[16]</sup>直接对待检测(大)图的特征图进行边框回归, 用单个神经网络结构同时实现对输入图像中目标的位置和分类的预测。

针对近场毫米波3维成像与异物检测中存在的问题, 本文主要讨论了基于BP算法的近场毫米波3维成像系统以及基于深度卷积网络的异物检测技术。第2节主要介绍了近场毫米波3维成像系统结构和信号模型, 第3节讨论了基于热图和YOLO算法的异物检测方法, 最后在第4节通过实测数据处理对本文所使用的方法进行验证和性能分析。

## 2 近场毫米波3维成像介绍

### 2.1 工作模式

原理上, 近场毫米波安检系统通过控制线阵运动合成大尺寸天线获得微波图像。但在实现过程中, 受到成本等方面的制约, 具体的系统结构有所不同, 主要分为垂直扫描模式和圆周扫描模式如图1所示。

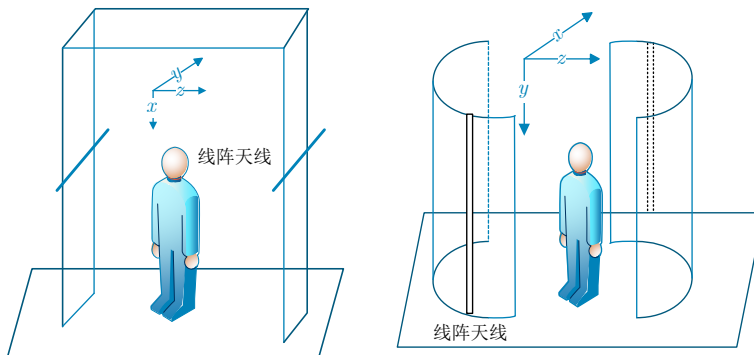


图1 近场毫米波3维成像几何模型

Fig. 1 The geometric model of near field millimeter wave 3D imaging

垂直扫描模式中,其线性阵元沿水平方向布设并以一定的速度往下扫描,分别经历匀加速,匀速以及匀减速运动的过程,同时发射阵元通过一定的脉冲重复频率发射线性调频信号并由接收阵元负责接收场景中的人体反射回波信号,经过一个完整的合成孔径时间,也即完成了一次对目标场景回波的采集工作,进而对目标回波信号实现基于3维BP成像算法便可对目标人体进行3维成像。通过在安检仪的两面各分别安置带有线性阵列的扫描架,这样便可以在一次扫描的过程中分别实现人体前、后两面的成像图。

圆柱扫描模式下,线性阵元沿垂直方向布设,通过控制该天线围绕圆心做水平旋转,合成圆柱形的阵列天线面<sup>[18]</sup>。与垂直扫描模式相比,该模式能够实现对人体水平面内360°范围内的等分辨率成像,且不存在人体侧面无法照射的情况,整体性能优于垂直扫描模式。但垂直阵列尺寸一般远长于水平阵列,因此该系统成本也相对较高。为了降低成本,在天线阵列设计时,可采用多入多出(Multiple-Input Multiple-Output, MIMO)技术,通过MIMO、单发多收(Single-Input Multiple-Output, SIMO)等设计降低系统成本。

## 2.2 回波信号模型

无论其具体工作模式,近场毫米波3维成像安检系统信号模型相同,下面以垂直扫描模式为例进行推导。此时,水平布置天线阵列的扫描架平台沿着 $x$ 方向分别经历匀加速、匀速以及匀减速运动的方式向下运动并对人体全身进行扫描,天线阵列水平安置在扫描架上且平行于 $y$ 轴,其中, $x$ 为方位向即扫描架运动方向, $y$ 为阵列方向即天线布置的水平方向, $z$ 为距离向即检测人体到扫描面阵的垂直距离。

假设每个时刻线阵天线只有1个阵元收发信号, $p_\omega$ 表示为目标场景中某个成像散射目标点位置,则第 $m$ 个慢时刻天线相位中心(Antenna Phase Center, APC)位置矢量 $\mathbf{p}_{pc}(m)$ 为

$$\mathbf{p}_{pc}(m) = \mathbf{p}_c(m) + \hat{\mathbf{y}}g(m) \quad (1)$$

其中, $g(m)$ 代表 $m$ 时刻收发阵元位置, $\mathbf{p}_c(m)$ 为阵列天线运动轨迹, $\hat{\mathbf{y}}$ 为线阵方向矢量。

由此可得目标场景中散射点 $p_\omega$ 到某个阵元的距离历史公式为

$$R(m; p_\omega) = 2 \times \|\mathbf{p}_{pc}(m) - p_\omega\|_2 \quad (2)$$

对于近场毫米波3维成像系统,一般需要较大的信号带宽,如3 GHz,采用匹配滤波技术难以实现,一般采用去斜率处理进行距离压缩。假设近场毫米波3维成像系统发射阵元发射的线性调频信号

(Linear Frequency Modulation, LFM)为

$$S_t(t) = Ae^{j2\pi f_c t + j\pi K t^2}, |t| \in T_p/2 \quad (3)$$

其中, $A$ 为发射信号幅度, $t$ 为快时间, $f_c$ 为雷达信号载频, $K$ 为线性调频信号的调频斜率, $T_p$ 为线性调频信号脉冲宽度。发射的线性调频信号将由目标散射点 $p_\omega$ 反射并由接收阵元接收回波信息,则接收的回波信号经过固定的时延后表达式为

$$S_r(t, m; p_\omega) = \sigma(p_\omega) e^{j2\pi f_c(t - R(m; p_\omega)/c)} e^{j\pi K(t - R(m; p_\omega)/c)^2} \quad (4)$$

其中, $m$ 为慢时间序号, $\sigma(p_\omega)$ 为散射点 $p_\omega$ 的后向散射系数, $c$ 为光速,第1个指数项代表雷达信号载频引起的延迟相位,第2个指数项包含了散射点在距离向的延时信息。

将原始回波信号 $S_r(t, m; p_\omega)$ 进行去斜率处理,得到采集的单目标散射点 $p_\omega$ 回波信号为

$$S_{st}(t, m; p_\omega) = \sigma(p_\omega) e^{-j2\pi f_c R(m; p_\omega)/c} e^{-j2\pi K t R(m; p_\omega)/c} \cdot e^{j\pi K(R(m; p_\omega)/c)^2} \quad (5)$$

其中,第1项对应运动合成的阵列天线,第2项为距离延迟对应的单频信号,第3项为残留相位,由于近场成像中距离的数值远小于光速的数值,此项可以忽略(假设系统工作在Ka波段,最远作用距离 $R=0.5$  m,调频斜率为 $1.5 \times 10^{14}$  Hz/s,残留项产生的最大相位偏差为 $\pi K R^2/c^2 = 0.0013(\text{rad}) = 0.0745^\circ$ ,远小于Ka波段波长的1/8,故可忽略)。

对于面目标,雷达接收回波为目标场景3维空间 $\Omega$ 内所有目标点回波对应的积分,则回波信号表达式为

$$S_r(t, m) = \sum_{p_\omega \in \Omega} S_{st}(t, m; p_\omega) \quad (6)$$

从近场毫米波3维成像模型可以看出,与传统SAR成像相比,近场3维成像系统等效阵列结构更为复杂,可能为2维平面阵列、圆柱阵列等,增加了信号处理难度。成像过程中,当聚焦深度大于一定值后会导致较严重的散焦现象,而安检成像中人体的厚度较宽,这些因素都会导致回波处理和较高精度成像的困难。从系统角度考虑,还可能在不同阵元接收回波串扰问题,且回波信号需要补偿通道误差,进一步增加了成像处理难度。

## 2.3 2维投影图

对于近场3维成像而言,成像算法主要分为频域成像算法和时域成像算法。两类成像算法都在2维成像基础上扩展到3维空间,其中典型近场3维成像算法有3维距离-多普勒算法以及3维后向投影算法。

后向投影算法是典型时域成像算法，通过对同一目标场景像素点距离压缩后的回波进行多普勒相位补偿以及相干累加得到该像素点后向散射系数，进而实现聚焦成像，详细流程可参阅文献[19]。

对于近场安检成像应用，采用BP算法遍历3维成像场景，即可得到目标区域的3维图像。为了更好地呈现结果，需要将该3维图像沿距离方向进行投影，得到符合人眼视觉的2维图像。最典型的方法为“最大值投影法”。该方法遍历方位向 $x$ 和阵列向 $y$ ，选择距离向 $z$ 所有目标像素值中最大的点作为2维投影图像对应点 $(x,y)$ 的像素值。如图2所示为最大值投影后经过质量优化的成像图，其中成像的垂直和水平方向分别对应方位向和阵列方向，且后续章节中所有成像结果横纵坐标轴含义与该图相同，所以不再赘述。假设3维空间每个场景点对应像素值函数为 $f(x,y,z)$ ，则最大值投影至2维图像的公式为

$$g(x,y) = \max_z (f(x,y,z)) \quad (7)$$

其中， $g(x,y)$ 为2维投影图像， $z$ 表示距离向。

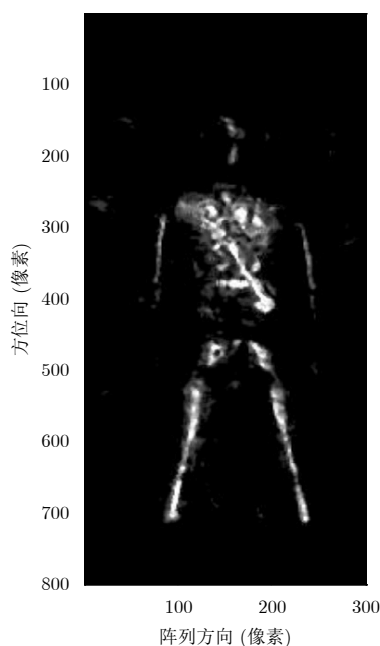


图 2 2维成像结果

Fig. 2 The 2D imaging result

### 3 基于深度神经网络的目标检测

#### 3.1 卷积网络

卷积神经网络(Convolutional Neural Networks, CNN)属于前馈神经网络，通过权值共享和局部感受野的概念，将卷积运算引入到网络结构中，在显著扩展网络结构的条件下，很好地控制了网络参数维度，在图像分类识别、自然语言处理等研究领域获得了良好效果。

卷积神经网络一般由交替出现的卷积层、池化层以及最后的全连接层组成，基本结构如图3所示。

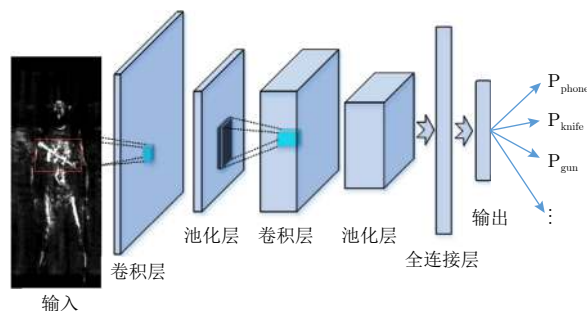


图 3 CNN结构

Fig. 3 The CNN structure

卷积层由多个特征图组成，它的每个神经元通过卷积核(权值矩阵)与上一层特征面的局部区域相连接。卷积层通过卷积操作对输入进行特征提取。第1层卷积层的输入即为输入图像，提取如边缘、线条等低级特征。更高层的输入则为上一层的特征面，从中提取更高级的特征。

池化层紧随在卷积层之后，是一种非线性的下采样。它的每个神经元对局部接受区域进行池化操作，具有二次特征提取作用。最大池化(max pooling)、均值池化(mean pooling)以及随机池化(stochastic pooling)等是常用的池化方法。其中最大池化取局部接受域中最大的点，均值池化取接受域中所有值的平均。

在经过多个卷积层和池化层后，连接着一个或者一个以上的全连接层。全连接层中的每个神经元与其前一层的所有神经元相连接，故而称为全连接层。该层可以整合卷积层或池化层中具有类别区分性的局部信息。最后1个全连接层起到分类器作用，根据前层提取的特征信息对输入进行分类处理，输出目标的类别。

卷积网络的标准模型主要用于分类任务，其输入为标准尺寸图像，输出为该图像的分类编号。但是，在异物检测任务中，异物图像嵌于整幅图像中，在做出类属判断前，还需要从图像中框选出异物可能出现的区域。下面介绍基于热图的异物检测方法和基于YOLO网络模型的异物检测方法。

#### 3.2 基于热图的检测方法

基于热图的目标检测方法的基本策略类似于图像滤波过程。首先，利用目标图像(小图)训练出一个分类卷积网络，其输入为小图，输出为该图像的分类。然后，将该网络在待检测图像中滑动，每个位置得到一个目标类属的输出构成了一张类属/概率的图像，称之为热图。

基于热图的异物检测过程如图4所示。标准分

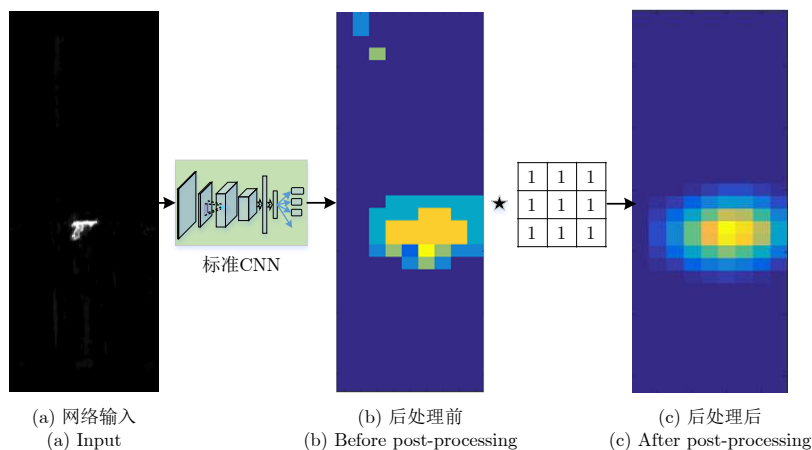


图 4 基于热图的目标检测结构

Fig. 4 The target detection structure based on heat map

类网络按照一定的步长在待检测大图中滑动，每个位置网络运行1次，得到1个概率输出，遍历整幅图像后得到初始热图，如图4(b)所示。

对于最后层以softmax为激活函数的分类网络，其输出在0~1之间。当相邻小图中均包含部分异物时，邻近区域具有相似大小的概率输出，所以通过寻找初始热图中能量最大值定位异物位置时会引入较大误差。例如，热图4(b)橙色区域能量大小基本一致，每个点均可能判定为目标中心。为此，本文对原始输出热图进行了后处理。使用值为1，尺寸大小为 $3 \times 3$ 的模板与热图做卷积运算，使能量进一步聚集在异物中心位置，得到一张处理后的热图，如图4(c)。可以看出，只有一个能量最强点，即为待检测异物中心。

热图检测法原理简单，但是，由于其需要遍历整幅待检测图像，运算量庞大，且生成的检测框尺寸固定，无法适应异物尺寸变化。为此，需要对热图法进行替换，引入更灵活、高效的异物检测算法。

### 3.3 基于YOLO的检测方法

传统的目标检测方法通常分为3个阶段。首先通过不同尺寸的滑动窗口在图像上选择候选区域，其次对候选区域特征进行提取，最后通过分类器进行识别。传统目标检测基于滑窗进行区域选择，由于缺乏针对性，窗口冗余多，时间复杂度高，手工选择的候选框缺乏泛化能力，因此不适用于对实时性、准确率要求高的安检应用。

随着深度学习的高速发展，目标检测算法不断朝着高效率、高性能的方向发展。当前比较流行的算法主要可以分为两大类，一类是基于候选区域(region proposal)的RCNN系算法，主要有RCNN, Fast RCNN和Faster RCNN。RCNN系算法是两阶段(two stage)检测算法，需要先使用选择性检索

(selective search)或RPN网络产生候选区域，然后在候选区域上做分类与回归。该类算法在生成候选区域时消耗较多时间，由输入到检测需要3 s左右。虽然RCNN系算法对于目标检测耗时较多，但是在检测精度上保持领先水平。另一类是YOLO和SSD这类单阶段(one stage)算法，即本文使用的检测算法。YOLO算法是一种端到端的目标检测算法，每次检测时将整张图像输入网络，不同于RCNN系算法生成候选区域的方法，YOLO算法将图像划分为 $m \times n$ 的网格，将目标检测问题作为回归问题解决，直接对网格中的目标中心点位置及边框的高宽进行预测。因此YOLO算法仅仅使用一个CNN网络就能直接预测不同目标的类别与位置。相比于传统检测网络，YOLO算法在检测精度得到了很大的提升；相对于RCNN系列的检测网络，YOLO算法检测时间更短，效率得到了优化。这些特点说明了YOLO算法更适用于人流量多的实时安检应用。

YOLO算法使用单个卷积网络即可实现端到端(end to end)的目标检测，其训练流程如图5所示。YOLO网络使用过程简单，首先将待检测图片(大图)尺寸插值到合适的大小，然后送入训练好的YOLO网络，即可得到目标检测结果。相比RCNN等常用检测算法，YOLO速度更快，而且训练过程也直接在包含标签(ground truth)的待检测(大)图像中操作，过程更为简单。

YOLO网络设计主要包括特征提取、边框回归和非极大值抑制3个部分。

特征提取主要靠标准卷积网络的卷积层实现，一般选择GooLeNet模型，共包含24个卷积层和全连接神经网络，且输出为类属数目。损失函数采用最小二乘准则，其公式为

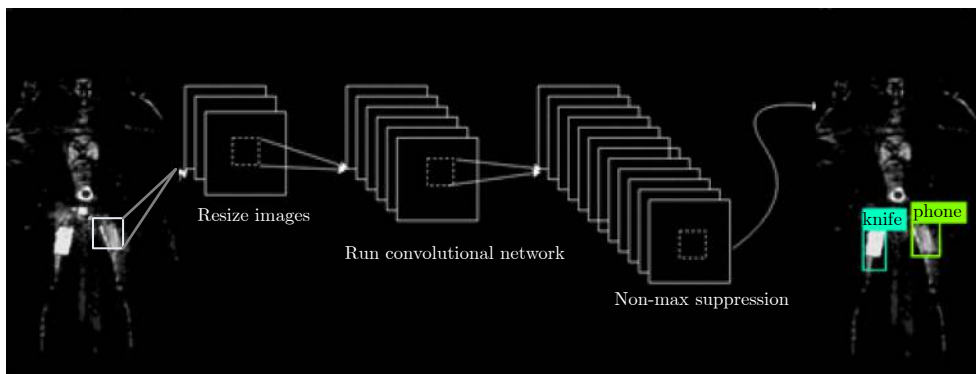


图 5 YOLO的训练流程

Fig. 5 The training process of YOLO

$$L_C = \sum_{i=0}^{S^2} I_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (8)$$

其中,  $I_i^{\text{obj}}$  表示第  $i$  个栅格区域内是否存在目标, 如果存在目标, 则为1, 否则为0。该代价函数对所有包含目标的栅格区域进行训练。对于不包含目标的栅格, 其分类网络识别输出没有意义, 所以通过  $I_i^{\text{obj}}$  将代价函数直接置零且不对其进行优化。

边框回归主要任务是从特征图中预测出目标框的位置和尺寸。为了实现该任务, 其在特征图后串联了一个5输出网络(分别为位置坐标  $x, y$ 、尺寸高宽  $h, w$  和置信度  $C$ ), 并采用最小二乘准则对该神经网络的参数进行训练, 代价函数为

$$\begin{aligned} L_B = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} [(x_{ij} - \hat{x}_{ij})^2 + (y_{ij} - \hat{y}_{ij})^2] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} \left[ \left( \sqrt{w_{ij}} - \sqrt{\hat{w}_{ij}} \right)^2 \right. \\ & \left. + \left( \sqrt{h_{ij}} - \sqrt{\hat{h}_{ij}} \right)^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \frac{1}{2} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{noobj}} (C_i - \hat{C}_i)^2 \end{aligned} \quad (9)$$

其中,  $\lambda_{\text{coord}} = 5$ ,  $x_{ij}, y_{ij}, w_{ij}, h_{ij}$  表示第  $i$  个栅格区域第  $j$  个候选框的位置和宽高,  $\hat{x}_{ij}, \hat{y}_{ij}, \hat{w}_{ij}, \hat{h}_{ij}$  为真实框的位置和高宽,  $I_{i,j}^{\text{obj}}$  表示第  $i$  个栅格区域内第  $j$  个候选框是否存在目标, 如果存在目标, 则为1, 否则为0, 其作用是让只包含目标的候选框参与优化训练,  $C_i$  为第  $i$  个栅格中存在目标的置信度(confidence)。

分类和边框预测使得该网络训练变为一个多目标最优化问题, 为了简化处理, 将两个损失函数合并写为

$$L = L_C + L_B \quad (10)$$

非极大值抑制(Non Maximum Suppression, NMS)<sup>[20]</sup>主要用于解决同一个目标中出现多个候选框的问题。其基本策略是选择所有候选框中, 目标出现概率最大的一个作为目标框, 而丢弃掉其它与该框存在很大重叠, 且类属相同的候选框。为了实现该目标, 首先从所有的检测框中找到置信度最大的框, 然后逐个计算其与剩余框的交并比(Intersection Over Union, IOU), 如果其值大于一定阈值(重合度过高), 那么就将该框剔除; 然后对剩余的检测框重复上述过程, 直到处理完所有的检测框。如图6所示, 当IOU阈值过高时对knife识别了2次(图6左), 降低阈值后去掉了重复的检测框(图6右)。

## 4 实验数据处理与性能分析

### 4.1 图像增强与成像处理

由于成像过程中存在不同阵元接收回波之间相互串扰、扫描元器件需要补偿的固定通道误差提前预估不精确等因素将会导致2维成像中出现较多干扰信息, 这些因素降低了成像质量, 并一定程度上

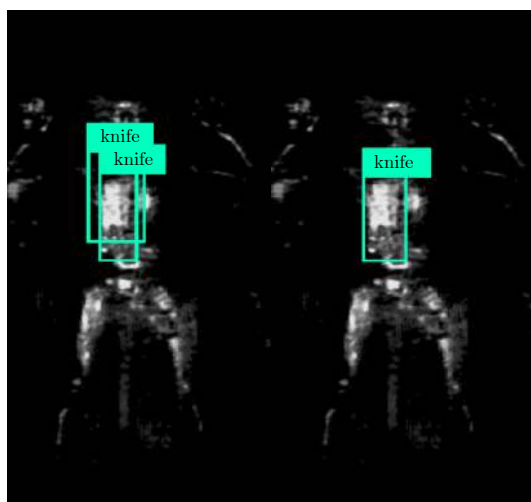


图 6 非极大值抑制算法去除重复预测框

Fig. 6 Removal of repeated prediction box by NMS algorithms

影响了安检识别。因此,需要使用图像处理方法对图像进行增强,如进行图像锐化增加成像对比度,使用中值滤波、图像平滑减小成像干扰噪声等。

图7、图8为实测垂直扫描模式近场毫米波3维成像处理结果,系统工作于Ka波段,信号宽带大于3 GHz,阵列长度大于0.5 m,系统阵列方向分辨率和方位向分辨率约为1 cm,距离向分辨率约为5 cm,但由于采用了最大值投影,所以该方向分辨率未在图像中体现。

其中图7为最大值投影后的2维成像结果,验证了3维BP算法适用于近场安检成像且图像分辨率也较高,并证实了获得3维成像后用最大值投影方法压缩到2维平面的方法是有效的。但观察发现,该成像结果仍存在较为明显的栅瓣,影响图像观感及后续识别的有效性。

图8为经过图像处理方法后的成像效果。对于原始成像图7首先进行中值滤波处理减小图像噪声信息,然后使用图像平滑抑制原始图的重影问题并

使图片显得均匀,最后使用图像锐化处理增加成像对比度使得异物更为明显。经过以上一系列图像后处理能够有效减小干扰信息并增强图像质量,处理后的图像更为匀称平滑,噪声较少,并且成像中的重影问题得到了很好的抑制,对于安检中识别率的提高起到了较好的作用。

#### 4.2 训练过程及检测结果

本文需要检测的异物分别为:枪(gun),手机(phone)和刀(knife),训练集原始图像大小约为 $800 \times 300$ 。对基于热图的检测网络,需要从训练图像中剪裁包含目标(正样本:刀、枪、手机)以及非目标(负样本:噪声背景以及身体各个部位成像结果)的小图用于训练分类网络,小图尺寸为 $128 \times 128$ ,见图9。训练过程中平均损失以及准确率如图10所示,可以看出随着迭代次数的增加,平均损失逐渐减小并趋近于0,而准确率则逐渐提升并接近于百分百准确预测。

而对于YOLO网络,本文在实验中将输入网络



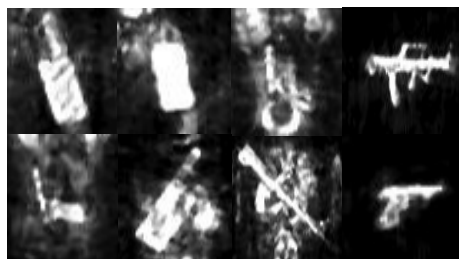
图 7 原始实测成像图

Fig. 7 The original measured image

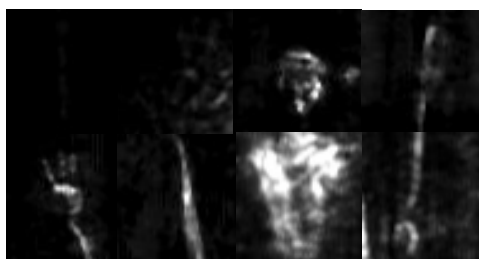


图 8 图像处理后成像图

Fig. 8 The image after processing



(a) 正样本  
(a) Positive samples



(b) 负样本  
(b) Negative samples

图 9 训练分类网络的样本

Fig. 9 Samples of training classification network

的训练图片尺寸由 $800 \times 300$ 插值到 $608 \times 608$ 。训练过程中损失函数的变化过程如图11所示，可以看出20000次训练后已经完全收敛，训练模型已经能够拟合训练样本，此时，训练集样本准确率趋于平稳，测试集准确率开始有所下降。为了避免过拟合，本文使用“早停”(early stopping)技术提前结束训练。对比两种网络的训练过程可以看出，相比基于热图检测的分类子网络，由于复杂性更高，YOLO网络的训练过程需要更多的迭代次数以及更长的训练时间。

在本文安检成像及异物检测的实验过程，本文将异物归结为刀、枪和手机3类并训练YOLO网络去实现异物分类检测。为了得到检测出异物并正确

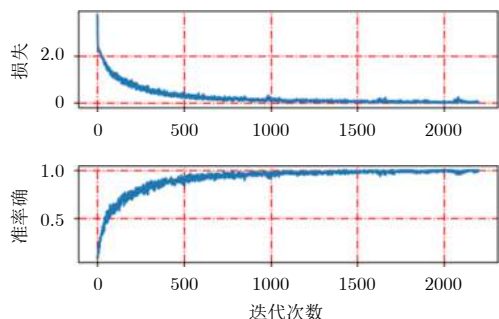


图 10 分类网络训练过程中平均损失和准确率

Fig. 10 The average loss and accuracy in classification network training

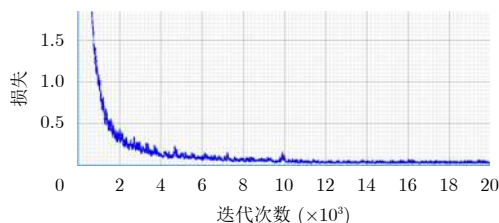


图 11 YOLO网络训练过程中的平均损失

Fig. 11 The average loss in YOLO network training

分类所需合适的尺寸大小，文中将几组测试图片输入到训练好的YOLO网络检测之前，对图片的尺寸做了一次调整。如图12所示为其中一组的检测情况，可以看出随着尺寸从 $416 \times 416$ 增加到 $608 \times 608$ (从左到右尺寸依次为 $416 \times 416$ ,  $480 \times 480$ ,  $544 \times 544$ ,  $608 \times 608$ ), 检测准确率也随之提高并且当尺寸达到 $608 \times 608$ 时不存在漏检的情况，验证了测试图像尺寸在一定范围内越大则检测准确率和预测框精度也随之改善。而且本文在训练YOLO网络时也会将原始训练集每一张图片的尺寸调整为 $608 \times 608$ ，这种情况也说明了将测试图像输入训练好的网络前，调整图像尺寸为训练图像的尺寸 $608 \times 608$ 检测性能最好，因为在同样像素比例的图片中，测试图片中异物形状更加贴近训练集图像中对应类别的异物，得到异物检测准确率也会更高，所以本实验中异物检测最合适的尺寸大小为 $608 \times 608$ 。

图13为连续经过1000, 5000, 10000, 15000和20000(由左至右)次训练后模型的测试结果。从图中可以看出，随着训练次数增多，模型的检测和识别能力逐渐提高，预测框范围不断调整并逼近目标的真实大小。在训练YOLO网络时本文对160张图片进行了数据扩充，包括对图像随机旋转、随机裁剪、随机调整亮度和引入不同噪声等，共生成了3300张图片，其中70%的图片作为训练集图像，30%的图片作为测试集图像，经过实验证明20000次训练后检测平均精确率在90%左右，详细测试结果表1。在评测指标中，精确率可以理解为在预测结果中，正确被预测为正样本的数量在所有预测为正样本中的比例；召回率表示正确被预测为正样本的数量占原始标签正样本中的比例。分别求得每个类别的精确率和召回率，利用PR曲线计算积分面积即为每一类的平均准确率(Average Precision, AP), mAP为各个类别AP的均值。

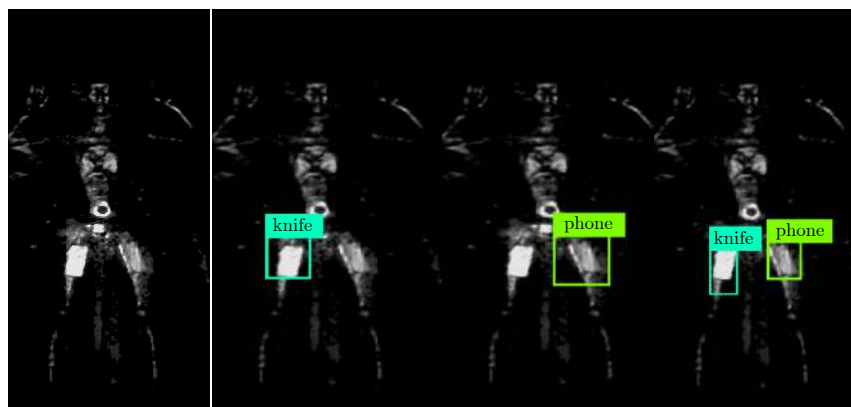


图 12 YOLO网络检测结果随图像尺寸变化情况

Fig. 12 YOLO network detection results of different image size



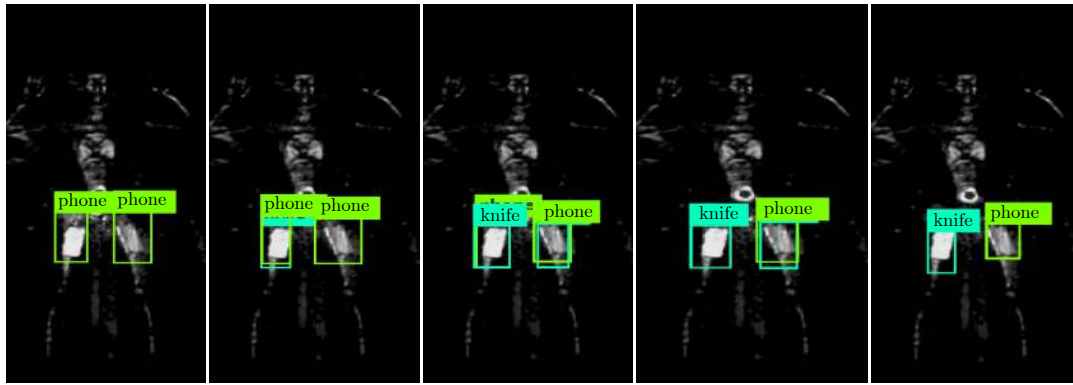


图 13 YOLO测试结果(不同训练次数)

Fig. 13 YOLO test results (different training numbers)

表 1 YOLO网络检测结果(%)

Tab. 1 The YOLO network detection results (%)

类名	平均精确率(AP)
gun	95.43
phone	89.86
knife	90.88
mAP	92.06

为了对比两种方法在检测过程中的效率，本文使用C程序对二者进行测试，并在Intel core i7-7800、GTX1060计算环境下完成。对于800×300的输入图像，基于热图的检测方法用时约200 ms(主要开销为卷积网络部分，对于800×300输入图像，需调用约300次网络)，而YOLO所需耗时不足50 ms，可见YOLO网络在检测效率要优于热图方法。

如图14所示，从随机取出两张图片的不同网络检测方法试验结果来看，在检测精度上，基于热图的检测方法检测框大小固定并且检测范围很大，相

比之下YOLO方法能够自适应调整检测框大小并精准地框出刀、枪等异物，说明了检测精度更高；在检测的准确率上，从图14中可以看出热图方法存在一定程度上的漏检情况，准确率不如YOLO检测方法；在检测时间上，YOLO网络的检测时间远小于热图方法，检测效率更高。对于目前安检应用，YOLO网络的检测时间很短，能够更好地满足人流量较大区域安检实时性的要求，并且该方法较高精度和准确率的优点能够有效地保障安检区域的安全性，相比于热图方法，这些优点使得YOLO网络成为更适合安检的检测算法。

### 5 结论

本文对主动式毫米波阵列3维系统成像及目标检测问题进行了研究。研究表明，后向投影算法由于其灵活性，可用于近场毫米波3维成像处理。基于热图的检测方法和基于YOLO的检测方法均可实现成像中的异物检测。基于热图的检测方法网络结构简单、易训练，但由于需要遍历整幅待检测图

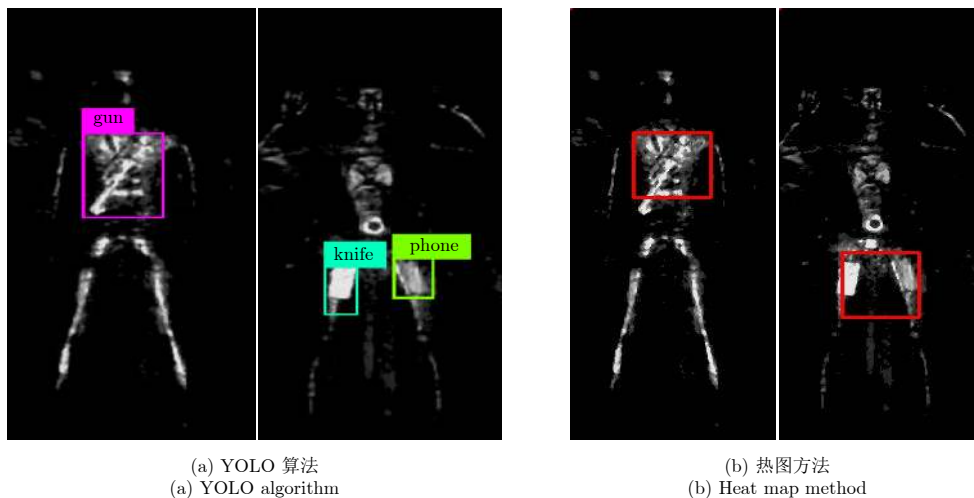


图 14 基于热图和YOLO检测结果

Fig. 14 Test results based on heat map and YOLO

像, 运算时间长, 且生成的检测框尺寸固定, 无法适应异物尺寸变化。基于YOLO的检测算法网络结构复杂、训练耗时长, 但该方法在检测速度与检测框精度上优势明显, 更利于机场安检等对实时性要求较高的异物检测应用。

### 参 考 文 献

- [1] CURRIE N C, DEMMA F J, FERRIS D D JR, *et al.* ARPA/NIJ/Rome laboratory concealed weapon detection program: An overview[C]. Proceedings of SPIE 2755, Signal Processing, Sensor Fusion, and Target Recognition V, Orlando, USA, 1996: 492–502.
- [2] FARHAT N H and GUARD W R. Millimeter wave holographic imaging of concealed weapons[J]. *Proceedings of the IEEE*, 1971, 59(9): 1383–1384. doi: [10.1109/PROC.1971.8441](https://doi.org/10.1109/PROC.1971.8441).
- [3] GONZALEZ-VALDES B, ALLAN G, RODRIGUEZ-VAQUEIRO Y, *et al.* Sparse array optimization using simulated annealing and compressed sensing for near-field millimeter wave imaging[J]. *IEEE Transactions on Antennas and Propagation*, 2014, 62(4): 1716–1722. doi: [10.1109/TAP.2013.2290801](https://doi.org/10.1109/TAP.2013.2290801).
- [4] 成彬彬, 李慧萍, 安健飞, 等. 太赫兹成像技术在站开式安检中的应用[J]. *太赫兹科学与电子信息学报*, 2015, 13(6): 843–848. doi: [10.11805/TKYDA201506.0843](https://doi.org/10.11805/TKYDA201506.0843).  
CHENG Binbin, LI Huiping, AN Jianfei, *et al.* Application of terahertz imaging in standoff security inspection[J]. *Journal of Terahertz Science and Electronic Information Technology*, 2015, 13(6): 843–848. doi: [10.11805/TKYDA201506.0843](https://doi.org/10.11805/TKYDA201506.0843).
- [5] 温鑫, 黄培康, 年丰, 等. 主动式毫米波近距离圆柱扫描三维成像系统[J]. *系统工程与电子技术*, 2014, 36(6): 1044–1049. doi: [10.3969/j.issn.1001-506X.2014.06.05](https://doi.org/10.3969/j.issn.1001-506X.2014.06.05).  
WEN Xin, HUANG Peikang, NIAN Feng, *et al.* Active millimeter-wave near-field cylindrical scanning three-dimensional imaging system[J]. *Systems Engineering and Electronics*, 2014, 36(6): 1044–1049. doi: [10.3969/j.issn.1001-506X.2014.06.05](https://doi.org/10.3969/j.issn.1001-506X.2014.06.05).
- [6] APPLEBY R and WALLACE H B. Standoff detection of weapons and contraband in the 100 GHz to 1 THz region[J]. *IEEE Transactions on Antennas and Propagation*, 2007, 55(11): 2944–2956. doi: [10.1109/TAP.2007.908543](https://doi.org/10.1109/TAP.2007.908543).
- [7] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886–893.
- [8] AHONEN T, HADID A, and PIETIKAINEN M. Face description with local binary patterns: Application to face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037–2041. doi: [10.1109/TPAMI.2006.244](https://doi.org/10.1109/TPAMI.2006.244).
- [9] VIOLA P and JONES M. Rapid object detection using a boosted cascade of simple features[C]. Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001: I-511–I-518. doi: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517).
- [10] Kelly E J. An adaptive detection algorithm[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1986, AES-22(2): 115–127. doi: [10.1109/TAES.1986.310745](https://doi.org/10.1109/TAES.1986.310745).
- [11] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. Imagenet classification with deep convolutional neural networks[C]. Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 1097–1105.
- [12] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.
- [13] GIRSHICK R. Fast R-CNN[C]. Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448.
- [14] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[C]. Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 91–99.
- [15] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580–587.
- [16] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788.
- [17] LIU Wei, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot multibox detector[C]. Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21–37.
- [18] GOMEZ-MAQUEDA I, ALMOROX-GONZALEZ P, CALLEJERO-ANDRES C, *et al.* A millimeter-wave imager using an illuminating source[J]. *IEEE Microwave Magazine*, 2013, 14(4): 132–138. doi: [10.1109/MMM.2013.2248652](https://doi.org/10.1109/MMM.2013.2248652).
- [19] 师君. 双基地SAR与线阵SAR原理及成像技术研究[D]. [博士学位论文], 电子科技大学, 2009.  
SHI Jun. Research on principles and imaging techniques of bistatic SAR & LASAR[D]. [Ph.D. dissertation], University of Electronic Science and Technology of China, 2009.
- [20] ROTHE R, GUILLAUMIN M, and VAN GOOL L. Non-maximum suppression for object detection by passing messages between windows[C]. Proceedings of the 12th Asian Conference on Computer Vision, Singapore, 2014: 290–306.

## 作者简介



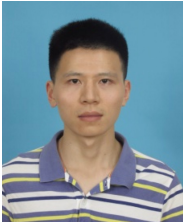
师 君(1979-), 男, 河南南阳人, 博士, 电子科技大学信息与通信工程学院副教授, 主要从事SAR成像技术、雷达信号处理研究, 已发表论文50余篇。

E-mail: shijun@uestc.edu.cn



周远远(1992-), 男, 山东金乡人, 电子科技大学博士生, 主要从事深度学习技术在SAR图像的应用研究。

E-mail: 732156543@qq.com



阙钰佳(1993-), 男, 福建龙岩人, 电子科技大学硕士生, 主要从事阵列3维SAR成像及目标识别技术研究。

E-mail: queyujia1993@163.com



张晓玲(1964-), 女, 四川成都人, 博士, 电子科技大学信息与通信工程学院教授, 主要从事SAR成像技术、雷达探测技术研究, 已发表论文50余篇。

E-mail: xlzhang@uestc.edu.cn



周泽南(1997-), 男, 海南海口人, 电子科技大学硕士生, 主要从事深度学习技术在SAR图像的应用研究。

E-mail: 2942714332@qq.com