

DOI: 10.13382/j.jemi.B2407708

多尺度反向校正增强和无损下采样的毫米波图像 目标检测方法*

叶学义 韩卓 蒋甜甜 王佳欣 陈华华

(杭州电子科技大学通信工程学院 杭州 310018)

摘要:针对毫米波图像中隐匿目标局部信噪比低导致检测障碍的问题,提出了一种基于多尺度反向校正增强和无损下采样的检测方法。首先设计了一种多尺度反向校正特征增强模块,在提取多尺度特征的多卷积核 Res2Net 上融合反向校正操作,实现大感受野区域对区域内相关小感受野区域卷积计算的反向校正,使得深度模型不仅能够获取更细粒度的特征,而且使宏观判别性表示贯穿多个尺度的特征信息;其次,利用非跨步卷积层的 SPD-Conv 实现无损下采样,缓解卷积下采样导致的信息丢失;最后,采用 K-means++ 聚类算法生成适合隐匿目标检测任务的新锚框。实验在 YOLO 系列中选择了各方面性能都适中的 YOLOv5s 作为基础框架,针对现有的两种毫米波图像数据集(阵列图像集和线扫图像集)平均精度均值(mAP) mAP@0.5 分别达到了 96.21%和 97.97%,相较于原版 YOLOv5s 以及 YOLO 其他系列等性能有显著提升。实验结果表明,该方法在不明显增加参数数量和推理时间的同时,能够有效提升深度模型的检测性能。

关键词: 隐匿目标检测;主动毫米波图像;多尺度反向校正特征增强;无损下采样;K-means++

中图分类号: TP391;TN919

文献标识码: A

国家标准学科分类代码: 510.4050

Multi scale reverse correction enhancement and lossless downsampling for millimeter wave image object detection method

Ye Xueyi Han Zhuo Jiang Tiantian Wang Jiixin Chen Huahua

(School of Communication Engineering, Hangzhou University of Electronic Science and Technology, Hangzhou 310018, China)

Abstract: A detection method based on multi-scale inverse correction enhancement and lossless downsampling is proposed to improve the detection of hidden targets in millimeter wave images with low local signal-to-noise ratio. Firstly, a multi-scale reverse correction feature enhancement module was designed, which integrates the reverse correction operation on the Res2Net multi convolution kernel. This achieves the reverse correction of convolution calculation between large receptive field regions and related small receptive field regions, enabling finer-grained features across multiple scales. Secondly, utilizing non-step convolutional layers of SPD-Conv to achieve lossless downsampling and preserve more information. Finally, the K-means++ clustering algorithm generates new anchor boxes suitable for hidden object detection tasks. The experiment selected YOLOv5s, which has moderate performance in all aspects, as the basic framework, targeting two existing millimeter wave image datasets (array image dataset and line scan image dataset) mAP@0.5 reaching 96.21% and 97.97% respectively. Compared to the original YOLOv5s and other YOLO series, the performance has significantly improved. The experimental results show that this method can effectively improve the detection performance of deep models without significantly increasing the number of parameters and inference time.

Keywords: hidden target detection; active millimeter wave image; multi scale inverse correction feature enhancement; lossless downsampling; K-means++

0 引言

当前公共场所中传统的安检技术如金属探测器^[1]、人工检测^[2]、X射线探测器等,在应用场景、检测效率和人体健康等方面存在一定不足之处。随着人们出行需求与安全意识的不断提高,对应用在车站、机场等大型公共场所中的安检技术的要求愈加严格,冲击着这一领域去寻找更加安全高效的安检手段。近些年来,毫米波凭借非电离辐射^[3]、能穿透衣物且不伤害人体^[4]、对人体、金属、陶瓷等不同类型的物体成像结果有较为显著的差异等特点,逐渐在安检领域受到研究者的青睐。随着毫米波成像、计算机视觉以及目标检测技术的不断提高,将其结合的毫米波检测技术有望取代这些传统的安检方式,提高安检的效率和安全性。

毫米波是一类波长处于1~10 mm之间的电磁波^[5]。根据是否主动发射毫米波信号,毫米波成像方式可以分为被动式毫米波(passive millimeter-wave, PMMW)^[6]和主动式毫米波(active millimeter-wave, AMMW)^[7]。由于被动式毫米波仅通过反射人体及环境中的热辐射信号生成图像,具有信号强度弱、信噪比低、易受环境影响等不足,成像质量一般低于主动式毫米波图像。因此,基于主动式毫米波成像技术进行隐匿目标检测是更为主流的研究方向。

当前对毫米波图像隐匿目标检测的研究可以分为传统方法和深度学习方法两种。传统方法主要通过对毫米波图像进行边缘检测、去噪、阈值分割以及传统机器学习等方法提取目标,虽然能够有效检测目标,但在不同数据集之间不具备良好的泛化性。而深度学习方法相较于传统方法表现更为出色,因而采用卷积神经网络(convolutional neural network, CNN)检测毫米波图像隐匿目标逐渐成为了主流。基于深度卷积神经网络的目标检测方法主要分为基于区域提取的多阶段模型和基于一步回归的单阶段模型两类。相对于以YOLO(you only look once)^[8]系列为代表的单阶段检测模型,虽然多阶段检测模型如Faster-RCNN^[9],往往能拥有更高的检测精度,但在实时性方面存在明显劣势,难以满足安检场景下的实际应用需求。因此,近年来基于YOLO模型对毫米波图像隐匿目标检测的研究逐渐增多,旨在保证安检实时性要求的同时提高检测性能。Wang等^[10]提出一种归一化积累图(normalized accumulation map, NAM)和YOLOv2^[11]结合的方法,利用NAM计算每个目标位置二值掩膜的平均值,在训练过程中计算置信度损失时,将NAM作为不同位置的不同权重,从而揭示经常出现的隐藏物体的位置。针对毫米波图像中隐匿物体清晰度不一致、相似物体和复杂的背景干扰等问题,Yuan等^[12]提出

一种将YOLOv3^[13]和多路径特征金字塔(multi-path feature pyramid, MPFP)相结合以及修改残差块分布的方法,提高抓取浅层网络细节的能力,使相似物体更容易区分。此外,Fan等^[14]提出用CGAN^[15]增强毫米波图像特征,然后用YOLOv4^[16]对增强后的数据集进行检测,为小数据集提供了一种解决方案。

然而,由于毫米波图像成像质量差,分辨率低,噪声严重等问题,给目标检测任务带来了很大的阻碍,尤其是毫米波图像中存在大量的隐匿目标,这些隐匿目标冗余信息少,目标尺寸小,与背景纹理相似,导致特征不易提取,难以检出。特征提取的充分与否与目标检测性能息息相关,设计更加有效的特征提取模块可以提高模型的学习能力,改善模型对于隐匿目标的检测性能。因此,本文针对毫米波图像中隐匿目标局部信噪比低导致检测障碍的问题,提出了一种基于多尺度反向校正特征增强和无损下采样的检测方法,并选择YOLOv5s作为基础框架构建目标检测模型。

1 本文方法

与以往的YOLO系列相比,YOLOv5^[17]具有较为优越的检测性能和实时性。YOLOv5模型按照参数量从小到包含n、s、m、l、x等多种不同尺寸的基本模型。考虑到安检实时性与准确性的要求,选择以两方面都较为均衡的YOLOv5s作为毫米波图像隐匿目标检测的基础框架,提出并构建多尺度反向校正特征增强和无损下采样的毫米波图像目标检测模型。

1) 多尺度反向校正特征增强模块(multi-scale reverse correction feature enhancement, MSRCFE),在多卷积核Res2Net^[18]上采用反向校正操作,让大感受野区域反向校正该大感受野区域内相关小感受野区域的卷积变换,提取具备更细粒度和更多宏观判别性表示的多尺度特征信息,生成多尺度反向校正特征增强模块,增强对隐匿目标的特征提取能力。

2) 无损下采样,将深度模型中通过池化或者跨步卷积而实现下采样的过程替换为具有从空间到深度层(space-to-depth, SPD)和非跨步卷积层的SPD-Conv^[19]来实现无损下采样,以保留更多的有效信息,使模型对于冗余信息较少的隐匿目标更加敏感。

3) K-means++聚类生成锚框,采用K-means++聚类算法,针对毫米波图像数据集重新聚类,生成更加适合隐匿目标检测任务的新锚框。

1.1 多尺度反向校正特征增强

对于目标检测模型来说,特征提取充分与否决定着目标检测性能的好坏。因此,改善特征提取方法是目标检测乃至整个深度学习领域的重要研究方向之一。然

而,采用单一的卷积核并不能够保证提取到充足的特征信息。通常,一张图像中可能存在不同尺寸的不同目标,在检测过程中,提取一个目标的关键上下文信息可能在比当前目标更大的区域范围内,如果可以获得不同尺度的感知信息,将有助于提升对目标特征的捕捉能力。目前,提取多尺度特征已经被广泛地应用到深度学习中,例如通过设计不同尺寸的卷积核或者不同数量的卷积核组,分别对同一层特征图进行卷积运算,再将输出特征图拼接起来以获得多尺度信息。但是,这种方式会明显提高网络的运算量和运算复杂度,这对于实时性要求很高的任务场景并不友好。不同于此,Res2Net 利用堆叠 CNN 卷积运算能够扩大感受野的特点,在 ResNet 基础上用具有等级制的类似残差连接替代单个残差块,在更细粒度级别上表示多尺度特征,并增加每个网络层的感受野。ResNet 模块和 Res2Net 模块如图 1 所示。

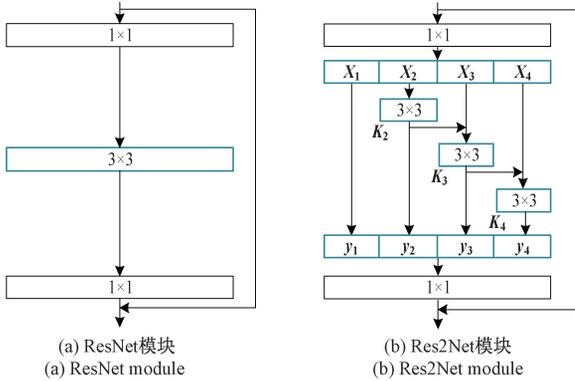


图 1 ResNet 模块和 Res2Net 模块

Fig. 1 ResNet module and Res2Net module

Res2Net 将 ResNet 瓶颈块中的单个 3×3 卷积核替换成了一系列 3×3 卷积核组,然后将一系列卷积核组以等级制类似残差的方式连接,来增加输出特征图可以表示的尺度数量。具体来说,首先将输入特征图经过 1×1 卷积后平均分成 s 个子特征图 X_i ,其中 $i \in \{1, 2, \dots, s\}$, s 作为控制尺度的维度参数,一般 s 越大就越可能学习到感受野尺度更丰富的特征。每个子特征图 X_i 与输入特征图有相同的空间尺寸,但通道数量是输入特征图的 $1/s$ 。除了 X_1 外,其余 X_i 都有一个相应的用来提取特征的 3×3 卷积 $K_i(\cdot)$,并将前一个卷积输出 $K_{i-1}(\cdot)$ 与当前输入 X_i 相加后再输入 X_i 对应的 $K_i(\cdot)$ 中,重复这个过程,直到所有的子特征图都处理完毕。将每个 $K_i(\cdot)$ 输出定义为 y_i ,则 y_i 表示为:

$$y_i = \begin{cases} X_i, & i = 1 \\ K_i(X_i), & i = 2 \\ K_i(X_i + y_{i-1}), & 2 < i \leq s \end{cases} \quad (1)$$

可以看出,每一个子特征图的输出都获取了不同尺

度的信息,残差连接越深的输出特征图,其感受野也就越大。最后,按通道拼接所有的输出特征图 y_i ,再由 1×1 卷积融合信息并减少通道数量。在 Res2Net 中,当子特征图通过 3×3 卷积操作后,其感受野就会增加,由于最后的拼接组合效应就会产生等效的多尺度特征,从而获得丰富的特征信息。

自校正卷积网络^[20]也是一种能够有效提高网络特征提取能力的设计,它通过内部的自校准操作巧妙地扩展了每个卷积层的视野,从而丰富了输出特征,增强了网络的学习能力。自校正卷积网络的结构如图 2 所示,它将输入特征图均匀地划分成为两个部分,并分别送入不同的分支。第 1 个分支只进行简单的卷积操作。而在第 2 个分支中,输入特征图在两个不同的尺度空间中进行了卷积特征变换:一个是原始尺度空间;另一个是通过下采样后得到的小的潜在空间,该空间相比于原始空间拥有更大的感受野。由于视野较大,在小的潜在空间中变换后的低维嵌入被用来校准在原始尺度空间的卷积变换,以融合两个不同尺度的空间信息。

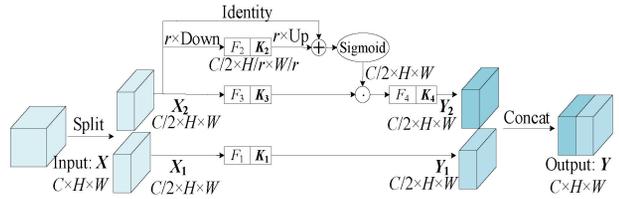


图 2 自校正卷积网络结构

Fig. 2 Self-calibrated convolutions network structure

对于第 2 个分支,具体来说,首先对输入特征图 X_2 采用池化窗口为 $r \times r$,步长为 r 的平均池化操作,得到低维特征空间 T :

$$T = AvgPool_r(X_2) \quad (2)$$

将 T 基于 K_2 做卷积变换,再通过上采样将低维特征空间还原到原始尺度空间:

$$X_2' = Up(F_2(T)) = Up(T \times K_2) \quad (3)$$

式中: $F_2(\cdot)$ 表示 c 对应的卷积操作; $Up(\cdot)$ 表示双线性插值操作;通过上采样使得 X_2' 与原始输入 X_2 空间尺寸相同。

然后,将 X_2' 作为大感受野下的校准权重,指导原始尺度空间下的卷积变换:

$$Y_2' = F_3(X_2) \cdot \sigma(X_2 + X_2') \quad (4)$$

式中: $F_3(\cdot)$ 表示 K_3 对应的卷积操作; $\sigma(\cdot)$ 表示 Sigmoid 激活函数;“ \cdot ”为按元素相乘操作。

最后,将校正后的结果与 K_4 卷积,得到输出 Y_2 :

$$Y_2 = F_4(Y_2') = Y_2' \times K_4 \quad (5)$$

虽然 Res2Net 通过类似残差连接的方式有效提高了网络的多尺度特征提取能力,但其传播过程是单方向的,

并且是由浅入深的,这使得宏观表征信息不能得到充分利用。在此基础上,受到自校正卷积网络启发,在 Res2Net 的 $K_2(\cdot)$ 、 $K_3(\cdot)$ 、 $K_4(\cdot)$ 上融合反向校正操作,利用大感受野区域特征返回校正该区域内的相关小感受野区域的卷积变换,实现多尺度的反向校正特征增强,让其不仅能够获取更细粒度的特征,而且使宏观判别性表示贯穿多个尺度的特征信息。

改进后的多尺度反向校正特征增强模块如图 3 所示,其中红色阴影部分为改进部分,其余部分沿用了 Res2Net 模块。改进部分的核心思想就是利用视野较大的特征空间来指导相关小感受野空间的卷积变换,但与自校正卷积网络不同的是,自校正卷积网络通过下采样的方式生成大感受野区域,而在这里,由于受到 Res2Net 的残差连接的作用,每个输出子特征图本身就具有不同的感受野,无需再通过下采样操作变换尺度空间。因此对于大感受野区域,直接通过步长为 1 的二维最大池化操作,以提取出显著的宏观特征,然后将其作为校正权重反向校正相应小感受野区域的卷积变换。下面将对改进部分的操作过程进行详细阐述,且 Res2Net 部分仍沿用前面的介绍,但将除 X_1 外,其余输入经过 $K_i(\cdot)$ 卷积操作后的输出定义为 y_i' 。

首先,将 X_4 经过 $K_4(\cdot)$ 卷积后得到的 y_4' 进行最大池化操作,然后与 X_4 基于元素相加,再进行 Sigmoid 激活操作,得到对应的校正权重:

$$y_4'' = \sigma(X_4 + \text{MaxPool}(y_4')) \quad (6)$$

式中: $\text{MaxPool}(\cdot)$ 为窗口尺寸为 3×3 , 步长为 1 的最大池化操作; $\sigma(\cdot)$ 为 Sigmoid 激活操作。

然后,用在 X_4 上的校正权重来校正 X_3 上的输出:

$$y_3 = K_5(y_3' \cdot y_4'') \quad (7)$$

式中: $K_5(\cdot)$ 为 3×3 卷积操作。

同样的,在得到 y_3 后,将 y_3 作为校正 X_2 的大感受野信息,得到对应的校正权重:

$$y_3'' = \sigma(X_3 + \text{MaxPool}(y_3)) \quad (8)$$

然后,用在 X_3 上的校正权重来校正 X_2 上的输出:

$$y_2 = y_2' \cdot y_3'' \quad (9)$$

最后,与 Res2Net 相同,将 y_1 、 y_2 、 y_3 、 y_4 拼接后经过 1×1 卷积融合信息并减少通道数量。

改进的 MSRCFE 模块在 Res2Net 基础上,通过反向的校正特征增强操作,巧妙地扩展了前向卷积层的视野,从而使得输出不仅包含多尺度特征,而且具备更丰富的宏观表征信息,有效增强了网络的特征提取能力。并且,该模块可以很方便地嵌入到深度神经网络中,同时也不会增加网络过多的参数量。以 YOLOv5s 为例,原始 YOLOv5s 模型的参数量约为 7.0×10^6 , 而将 YOLOv5s 中用于特征提取的单一卷积核 C3 模块替换为 MSRCFE 模块后,模型整体的参数量约为 7.7×10^6 , 相比原始

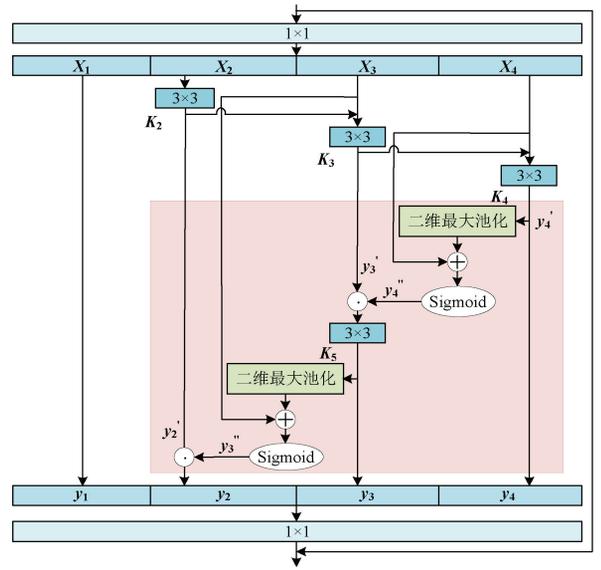


图 3 MSRCFE 模块

Fig. 3 MSRCFE module

YOLOv5s 基准模型只提高了 0.7×10^6 , 依然能够保证检测过程的实时性。

1.2 非跨步卷积

CNN 架构中,一般通过池化层或者步长为 2 的卷积层实现下采样以减少特征图尺寸,获取高级语义信息。这种设计,对于具有大量冗余信息的高分辨率图像或者中大尺寸目标适用,但对于冗余信息较少的低分辨率图像和小尺寸目标而言,这种操作会导致细粒度丢失和学习特征不足的情况。在 YOLOv5s 模型中,下采样正是通过步长为 2 的卷积实现的,对于具有较低分辨率的毫米波图像和较小尺寸的隐匿目标而言并不友好。针对这一问题,采用可以减少信息损失的 SPD-Conv 模块替换步长为 2 的卷积模块,实现下采样功能的同时保留更多的细粒度特征,使模型对于冗余信息较少的隐匿目标更加敏感。

SPD-Conv 由从空间到深度层 (space-to-depth, SPD) 和非跨步卷积层组成。SPD 由图像转化技术启发而来,原先用于对送入到网络之前的输入图像重构尺寸,后来被应用到网络内部实现特征图下采样。它可以保留通道维度上的所有信息,不存在信息丢失的问题。SPD 下采样的倍数与采样步长 $scale$ 有关,对于输入尺寸为 $S \times S \times C_1$ 的特征图 X , SPD 会将其切片得到一系列子特征图:

$$\begin{aligned} f_{0,0} &= X[0:scale;S,0:scale;S], \\ f_{1,0} &= X[1:scale;S,0:scale;S], \quad \dots \\ f_{scale-1,0} &= X[scale-1:scale;S,0:scale;S] \\ f_{0,1} &= X[0:scale;S,1:scale;S], \\ f_{1,1} &= X[1:scale;S,1:scale;S], \quad \dots \end{aligned}$$

$$\begin{aligned}
 f_{scale-1,1} &= \mathbf{X}[scale-1:scale:S,1:scale:S] \\
 &\vdots \\
 f_{0,scale-1} &= \mathbf{X}[0:scale:S,scale-1:scale:S], \\
 f_{1,scale-1} &= \mathbf{X}[1:scale:S,scale-1:scale:S], \dots
 \end{aligned}$$

$$f_{scale-1,scale-1} = \mathbf{X}[scale-1:scale:S,scale-1:scale:S] \quad (10)$$

其中,子特征图 $f_{x,y}$ 是由特征图 \mathbf{X} 横向从索引 x 开始每隔 $scale$ 个元素、纵向从索引 y 开始每隔 $scale$ 个元素采样得到的。假设采样步长为 2,则采样过程如图 4 所示。

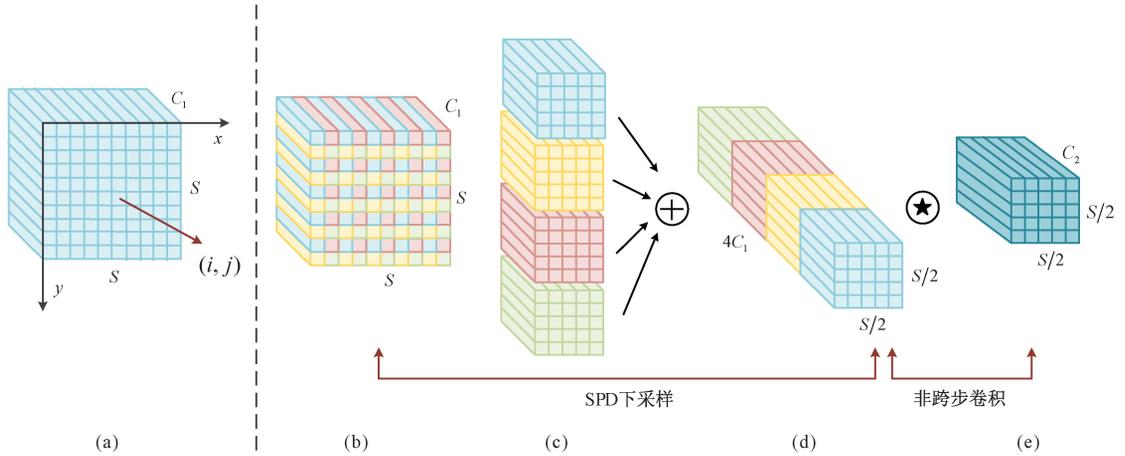


图 4 采样步长为 2 的 SPD-Conv

Fig. 4 SPD-Conv with a sampling step size of 2

由图 4 和式 (10) 可知,当采样步长为 2 时,可以得到 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}$ 这 4 个子特征图,且每个子特征图尺寸为 $(S/2, S/2, C_1)$,然后将 4 个子特征图按通道拼接成尺寸为 $(S/2, S/2, 4C_1)$ 的特征图 \mathbf{X}' ,此时 SPD 下采样已经完成。

SPD 下采样完成后,后面需要采用非跨步卷积来减少通道数量,从而减少计算参数量。非跨步卷积是一种步长为 1 的卷积操作,它不会改变特征图的宽度和高度,不会发生类似步长为 2 时产生的奇数和偶数的行或列采样次数不同的非对称采样情况,可以保留更多特征信息。在图 4 中,经过非跨步卷积后特征图尺寸为 $(S/2, S/2, C_2)$ 。

1.3 锚框改进

在目标检测中,锚框用来拟合真实标签框,选择适合的锚框尺寸是影响目标检测结果的重要因素之一。YOLOv5s 原版中提供的锚框是采用 K-means 聚类算法在 COCO 数据集中聚类而成的,而毫米波图像数据集中隐匿目标相对人体较小,通常在 COCO 数据集的中小尺寸目标范围内,因此,YOLOv5s 原版中提供的锚框并不适合

应用于毫米波图像中的隐匿目标检测。

生成锚框最常用的方法是 K-means 聚类算法,它将数据集中的目标分为 K 个簇,每个簇都有一个簇类中心,通过迭代分配 K 个簇内的点,更新簇内中心,直到 K 个簇类中心的变化很小或者达到迭代次数。然而,该算法在初始化时需要一次性选择 K 个簇类中心,如果簇类中心选择不合适,最终聚类生成的结果未必最优。K-means++ 聚类算法则对其进行了改进,它在初始化时只会选择一个簇类中心,然后根据轮盘法选择与已有簇类中心距离最远的那个作为下一个簇类中心,直到选满 K 个簇类中心,其余迭代更新方法与 K-means 聚类算法相同。K-means++ 可以避免初始聚类中心选择不佳导致的局部最优解问题,提高聚类效果和收敛速度。因此,本文选择采用 K-means++ 聚类算法,分别对阵列图像和线扫图像两种毫米波图像数据集重新聚类生成新锚框,结果如表 1 所示。由表 1 可以看出,经过 K-means++ 聚类算法重新聚类生成的新锚框与 YOLOv5s 原版锚框相比差异较大,新锚框的尺寸皆在原锚框的中小目标范围内。

表 1 K-means++ 聚类新锚框尺寸与 YOLOv5s 原锚框尺寸对比结果

Table 1 Comparison results of K-means++-clustering new anchor box size and YOLOv5s original anchor box size

	(pixels)								
名称	锚框 1	锚框 2	锚框 3	锚框 4	锚框 5	锚框 6	锚框 7	锚框 8	锚框 9
阵列图像	15×21	16×35	21×28	33×23	22×37	19×52	30×34	27×47	37×52
线扫图像	15×12	23×21	11×28	15×30	28×19	14×42	21×40	35×37	35×50
原锚框	10×13	16×30	33×23	30×16	62×45	59×119	116×90	156×198	373×326

1.4 模型整体结构

多尺度反向校正增强和无损下采样的毫米波图像目标检测模型的整体结构如图5所示。图5中,CBS模块由卷积 Conv、批量归一化 BN 和激活函数 SiLU 组成;MSRCFE 模块采用多卷积核来提取更细粒度和更多宏观判别性表示的多尺度特征;SPD-Conv 模块在实现特征图下采样的同时保留更多的细粒度信息。模型整体主要分为主干、颈部和头部3个部分。主干部分共有5级特征层,除第1级特征层仅有CBS模块外,其余特征层皆由SPD-Conv模块和不同数量的MSRCFE模块组成。颈部由SPPF和聚合路径网络(path aggregation network, PANet)^[21]模块组成,用来获取不同尺度的特征信息。头部包含3个检测头,每个检测头都分配了3种不同尺寸的锚框,分别用来检测大、中、小3种尺寸的目标。

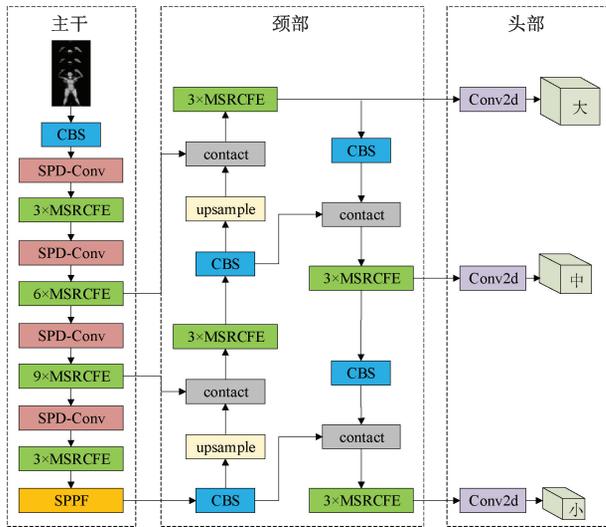


图5 模型整体结构

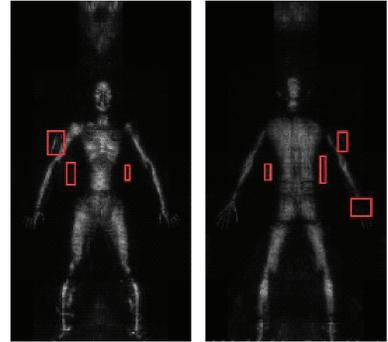
Fig. 5 Overall structure of the model

2 相关实验

2.1 数据集准备

为了验证本文方法的有效性,使用的实验数据集包含了两种从目前实际应用的主动毫米波图像采集设备上得到的数据集,阵列图像数据集和线扫图像数据集,用来测试本文方法对不同成像设备的毫米波数据集中隐匿目标的检测能力。其中,阵列设备通过多个传感器并行扫描,能够获得分辨率较高的毫米波灰度图像,适用于较为精细的目标检测;而线扫设备通过旋转传感器的方式逐步扫描目标,可以获得在多个视角下的图像数据,能更好地应对目标的遮挡和不同姿态的变化,但综合考虑成像效率和设备成本等因素,线扫图像的分辨率一般会比较低,难以取得和阵列图像相当的成像效果。

阵列图像数据集中阵列设备通过扫描每个模特产生正面和背面各1张毫米波灰度图像,图像分辨率为1024×2048,如图6所示,真实目标位置由图中红色框表示。实验选取24700张图像,采用8:1:1的方式随机划分训练集、验证集和测试集。



(a) 正面 (a) Front (b) 背面 (b) Back

图6 阵列图像示例

Fig. 6 Example of array image

线扫图像数据集中线扫设备通过旋转扫描每个模特产生不同角度的正面和背面各9张图像,图像分辨率为400×762,如图7所示,其中图7(a)为模特正面顺时针成像的9张图,图7(b)为模特背面逆时针成像的9张图,真实目标位置由图7中红色框表示。实验选取38250张图像,将每个模特的正面9张图像作为一组,背面9张图像作为一组后,再采用8:1:1的方式按组随机划分训练集、验证集和测试集。这种划分方式能够确保每个模特的正面和背面图像都能公平地用于训练和评估,避免了数据偏倚对实验结果的影响。

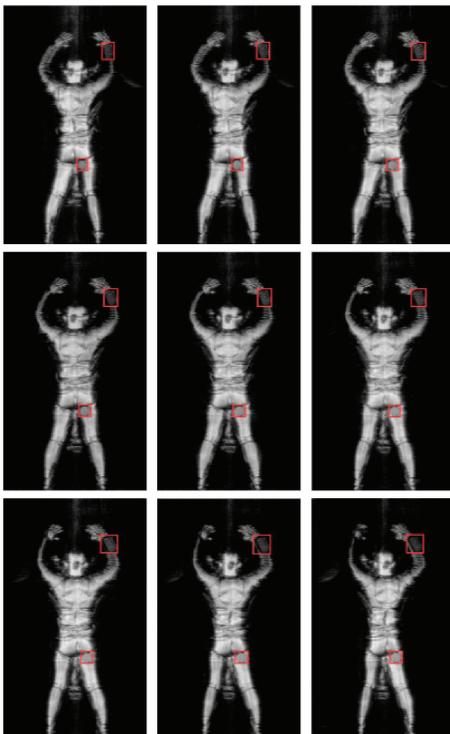
2.2 实验设置和评价指标

实验训练环境为Ubuntu 22.04操作系统,所用GPU为单张NVIDIA RTX 3090。实验采用训练轮次为180,批次大小为32,优化器采用随机梯度下降法。在训练过程中输入图片大小重新调整为640×640,并且所有图像在输入模型之前都进行了标准化处理,将每个像素值归一化到[0,1]区间。同时,为了模拟实际场景中的各种变化,采用了一些数据增强技术,包括随机裁剪、旋转、平移、拼接等,这样可以使得训练集更加多样化,从而进一步提升模型的检测能力和泛化能力。

实验中所用的评估指标为精度(precision, P)、召回率(recall, R)和平均精度均值(mean average precision, mAP)。P表示所有预测框中预测正确的数量所占的比例。R表示成功预测的数量占所有真实标签数量的比例。平均精度(AP)表示在不同置信度阈值下精度和召回率围成的面积,mAP则表示所有类别的平均精度均值,反映了模型的综合性能。



(a) 正面顺时针9张图
(a) 9 clockwise images on the front



(b) 背面逆时针9张图
(b) 9 counterclockwise images on the back

图 7 线扫图像示例

Fig. 7 Example of line scan image

2.3 阵列图像数据集相关实验

1) 对比实验

为了验证本文所提方法的有效性,将本文方法与现有主流目标检测模型做对比,包括 Faster R-CNN、YOLOv3、YOLOv4、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x、YOLOv6s^[22]、YOLOv7^[23]、YOLOv8s^[24]、YOLOv10s^[25]以及 YOLOv11s^[26],实验采用 P 、 R 、 $mAP@0.5$ 、参数量以及推理时间作为评价指标,实验结果如表 2 所示。

表 2 阵列图像数据集不同模型对比实验结果

Table 2 Comparison of experimental results of different models in array image dataset

模型	$P/\%$	$R/\%$	$mAP@0.5/\%$	参数量/ ($\times 10^6$)	推理时间/ms
Faster R-CNN	90.04	86.65	92.38	—	380.0
YOLOv3	89.22	85.14	90.84	61.5	3.8
YOLOv4	90.50	87.63	92.94	52.9	4.0
YOLOv5m	92.09	90.29	94.93	20.8	1.8
YOLOv5l	93.18	90.02	95.34	46.1	3.9
YOLOv5x	92.28	91.05	95.4	86.2	6.4
YOLOv6s	91.46	89.69	94.37	16.3	1.4
YOLOv7	91.85	89.47	94.54	37.2	2.1
YOLOv8s	91.63	89.94	94.81	11.1	1.2
YOLOv10s	92.03	89.35	94.63	8.1	1.0
YOLOv11s	91.94	88.87	94.32	9.4	1.1
YOLOv5s	91.37	89.56	94.20	7.0	0.8
本文	93.52	91.51	96.21	8.7	1.5

注:加粗字体为每列最优结果,下同

由表 2 可知,本文方法相比于现有主流模型在阵列图像数据集中隐匿目标检测上表现最优。与 YOLOv5 之前的 YOLO 系列模型和 Faster R-CNN 相比,基础的 YOLOv5s 模型在阵列图像数据集上的检测性能更胜一筹。与基础的 YOLOv5s 模型相比,本文方法的检测性能进一步提升,在增加 1.7×10^6 参数量和 0.7 ms 单张图像推理时间基础上, P 、 R 和 $mAP@0.5$ 分别增加了 2.15%、1.95% 和 2.01%。相比于 YOLOv5 其他大模型, P 、 R 和 $mAP@0.5$ 依然最高,参数量和单张图片上的推理时间比其中最大的模型 YOLOv5x 分别少 77.5×10^6 和 4.9 ms,比其中最小的 YOLOv5m 分别少 12.1×10^6 和 0.3 ms。相比于 YOLOv5 之后提出的 YOLOv6s、YOLOv7、YOLOv8s、YOLOv10s 以及 YOLOv11s 模型,其检测性能仍然优异。上述结果表明,本文方法在增加少量参数量和推理时间的基础上,其检测性能最好,有效提高了对阵列图像数据集中隐匿目标的检测能力。

2) 消融实验

为了验证所提各个模块单独作用的有效性,将基础模型 YOLOv5s、各个模块单独融合 YOLOv5s 以及整体融合 YOLOv5s 进行对比形成消融实验,实验结果评价指标为 P 、 R 和 $mAP@0.5$,对比结果如表 3 所示。

表 3 消融实验对比结果

Table 3 Comparison results of ablation experiments

模型	P	R	$mAP@0.5$
YOLOv5s	91.37	89.56	94.20
+MSRCFE	92.92	91.07	95.54
+SPD-Conv	93.2	90.36	95.27
+K-means++	91.97	90.49	94.97
+MSRCFE+SPD-Conv+K-means++	93.52	91.51	96.21

在 YOLOv5s 模型上分别融合 MSRCFE 模块、SPD-Conv 模块和 K-means++ 聚类算法时,其检测效果皆比基础 YOLOv5s 模型好。与基础 YOLOv5s 模型相比,当单独融合 MSRCFE 模块时, P 、 R 和 $mAP@0.5$ 分别增加了 1.55%、1.51% 和 1.34%;当单独融合 SPD-Conv 模块时, P 、 R 和 $mAP@0.5$ 分别增加了 1.83%、0.8% 和 1.07%;当单独融合 K-means++ 聚类算法时, P 、 R 和 $mAP@0.5$ 分别增加了 0.6%、0.93% 和 0.77%;当三者共同作用在 YOLOv5s 模型上时, P 、 R 和 $mAP@0.5$ 分别增加了 2.15%、1.95% 和 2.01%。与融合单个模块相比,多个模块共同作用时检测性能最好。

实验结果在阵列图像数据集上的表现如图 8 所示。其中,行表示不同样本,列分别表示标签图、采用 YOLOv5s 模型、采用 YOLOv5s 融合 MSRCFE 模块、采用 YOLOv5s 融合 SPD-Conv 模块、采用 K-means++ 聚类方法生成新锚框以及采用本文方法时所得检测结果。在标签图上的红色方框表示标签框,在其余图上的红色方框表示预测框。由样本 1、样本 2、样本 3 和样本 4 在不同模型上的检测结果可以看出,YOLOv5s 单独融合某个模块与基础 YOLOv5s 相比,都会在一定程度上提高对隐目标检测的检出,且单独融合某个模块检测的结果依然会在整体融合上表现出来。由样本 5 结果可以看出,当单独融合某个模块都不能提高对隐目标检出时,整体融合的共同作用会提高对隐目标检出。综上,本文方法与基础 YOLOv5s 相比提高了对隐目标的检出,且当 YOLOv5s 融合所有模块时对隐目标的检出结果表现最优。

3) 复杂场景下的测试实验

为了验证本文方法在不同复杂场景下的鲁棒性,选取了 5 100 张难以检出的低信噪比图库用于测试。其中,有 2 630 张是由毫米波图像采集设备自然成像生成

的。受到成像设备、环境干扰以及衣物遮挡等因素影响,该图库中普遍存在大面积虚影、衣物褶皱、人体部位模糊甚至丢失等情况。这些图例成像质量差,噪声干扰严重且复杂,目标融于背景,导致特征难以辨别,造成大量漏检和误检的情况。此外,通过在原图上人工添加不同程度的高斯随机噪声生成了额外 2 470 张低信噪比图像,以模拟更多的复杂场景,加上设备自然成像生成的 2 630 张,共计 5 100 张。该图库仅用于测试,并没有参与到模型训练中,既保证了实验结果的有效性,同时又能检验不同模型的鲁棒性和泛化能力。

表 4 所示即为部分模型在上述低信噪比图库下的测试结果。从结果对比可以明显发现,本文方法相比于 YOLOv5s 基准模型在低信噪比测试图库中的性能有较大提升, P 、 R 和 $mAP@0.5$ 分别增加了 1.00%、1.37% 和 2.56%,说明通过 MSRCFE 模块增强网络的特征提取能力和宏观表征能力,以及应用 SPD-Conv 模块改善下采样过程,保留更多细粒度信息,可以有效提高模型对于隐目标的检测能力和在复杂场景下的抗干扰能力。相比于后续 YOLO 系列 YOLOv8s、YOLOv10s 以及 YOLOv11s 模型,本文提出的方法依然能达到最优的性能,进一步验证了本文提出的改进模型优越的鲁棒性和泛化能力。

表 4 部分模型测试结果对比

Table 4 Comparison of partial model test results (%)

模型	P	R	$mAP@0.5$
YOLOv5s	85.73	65.11	72.06
YOLOv8s	85.78	64.80	72.71
YOLOv10s	82.33	64.28	71.93
YOLOv11s	81.89	63.41	71.86
本文	86.73	66.48	74.62

2.4 线扫图像数据集相关实验

为了验证本文方法在线扫图像数据中的有效性,将本文方法在线扫图像数据集中与现有主流目标检测模型做对比,采用的现有主流目标检测模型以及评价指标均与阵列图像数据集相同,实验结果如表 5 所示。实验结果表明,本文方法与现有主流目标检测模型相比,在线扫图像数据集上检测性能最优。与基础 YOLOv5s 相比,本文方法的检测性能能进一步提升,在增加参数量 1.7×10^6 和 0.7 ms 推理时间的基础上, P 、 R 和 $mAP@0.5$ 分别增加了 0.46%、1.57% 和 0.92%。相比于 YOLOv5 其他更大的模型, P 、 R 和 $mAP@0.5$ 依然最高,参数量和单张图片上的推理时间比其中最大的模型 YOLOv5x 分别少 77.5×10^6 和 4.9 ms,比其中最小的 YOLOv5m 分别少 12.1×10^6 和 0.3 ms。相比于后续提出的 YOLOv6s、YOLOv7、YOLOv8s、YOLOv10s 以及 YOLOv11s 模型,其检测性能仍然优异。上述结果表明,本文方法在增加少

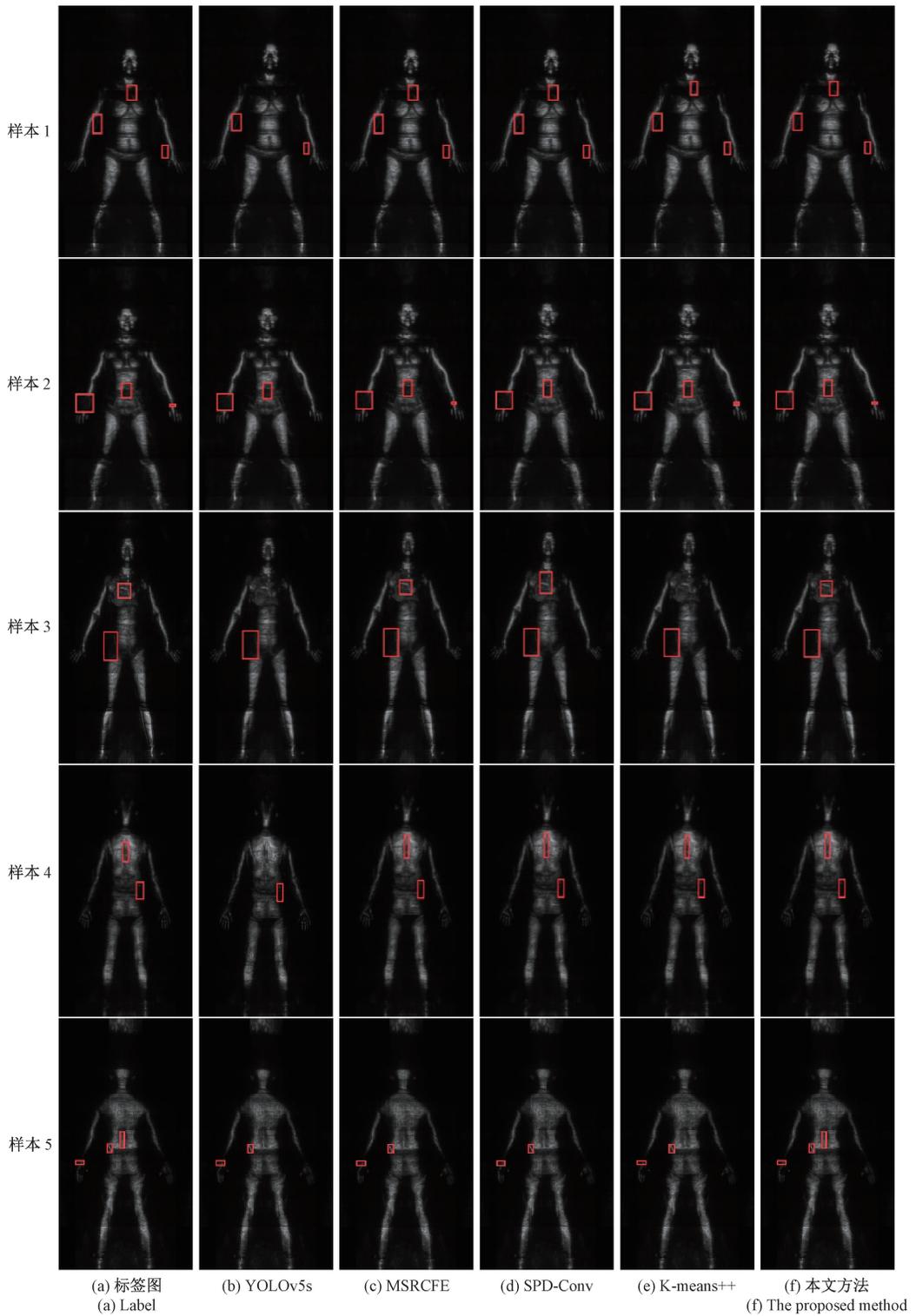


图 8 阵列图像数据集检测结果

Fig. 8 Detection results of array image dataset

量参数数量和推理时间的基础上,其检测性能最好,有效提高了对线扫图像数据集中隐匿目标的检测效果。

本文方法实验结果在线扫图像数据集中的表现如图 9

所示,其中,行表示不同样本,列分别表示标签图、采用 YOLOv5s 模型以及采用本文方法时所得检测结果。在标签图上的方框表示标签框,在其余图上的方框表示预测框。

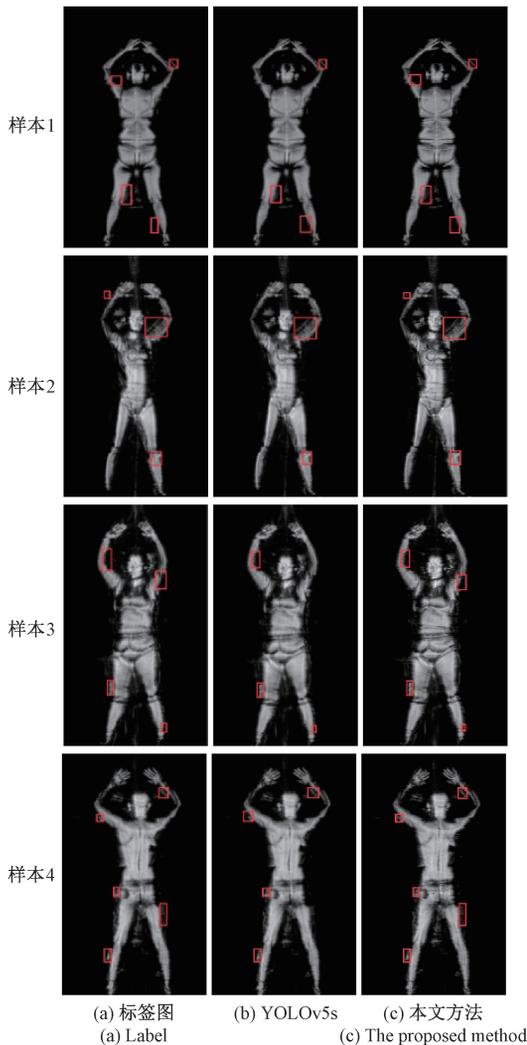


图 9 线扫图像数据集实验结果

Fig. 9 Experimental results of line scan image dataset

表 5 线扫图像数据集不同模型对比实验结果

Table 5 Comparison experimental results of different models on the line scan image dataset

模型	<i>P</i> /%	<i>R</i> /%	mAP@0.5/ %	参数量/ ($\times 10^6$)	推理时间/ ms
Faster R-CNN	91.44	89.86	95.93	—	380.0
YOLOv3	92.72	87.00	94.34	61.5	3.8
YOLOv4	93.68	91.52	96.61	52.9	4.0
YOLOv5m	93.90	93.04	97.53	20.8	1.8
YOLOv5l	93.75	93.70	97.82	46.1	3.9
YOLOv5x	93.73	93.72	97.82	86.2	6.4
YOLOv6s	93.39	92.47	97.13	16.3	1.4
YOLOv7	93.37	92.07	97.30	37.2	2.1
YOLOv8s	93.43	92.54	97.34	11.2	1.2
YOLOv10s	93.30	92.88	97.07	8.1	1.0
YOLOv11s	93.65	93.19	97.49	9.4	1.1
YOLOv5s	93.57	92.29	97.05	7.0	0.8
本文	94.03	93.86	97.97	8.7	1.5

由样本 1~4 在不同模型上的检测结果可以看出,与基础 YOLOv5s 相比,本文方法分别增加了对人体肩膀、手腕、腋下和大腿外侧的检出,显著提升了对隐匿目标的检测能力。

3 结 论

针对当前毫米波图像成像质量差、隐匿目标局部信噪比低、冗余信息少等原因导致目标检测困难的问题,本文提出了一种多尺度反向校正增强和无损下采样的毫米波图像目标检测方法。首先设计了多尺度反向校正特征增强模块,在 Res2Net 上采用反向校正操作,让大感受野区域特征返回校正该大感受野区域内相关小感受野区域的卷积变换,使得在训练过程中提取更细粒度和多尺度特征的同时具备更加丰富的宏观表征信息,有效增强了对隐匿目标的特征提取能力,从而使得模型在复杂背景和低信噪比环境下仍能精准识别隐匿目标;其次采用无信息损失的下采样 SPD-Conv 模块替代深度模型中通过池化或者跨步卷积而实现下采样的过程,有利于保留更多的有效信息,提升了模型对于小尺寸、低对比度目标的检测能力,尤其是在毫米波图像中目标冗余信息普遍较少的情况下,展现出其独特的优势;最后,采用 K-means++ 聚类算法对相对人体较小的毫米波图像数据集中的隐匿目标重新聚类生成新锚框,使得锚框的初始化更加贴合隐匿目标的分布,从而提升了目标检测的精度。实验选择具有优越实时性的 YOLOv5s 作为基础框架构建目标检测模型,在两种从目前实际应用的主动毫米波图像采集设备上得到的数据集——阵列图像数据集和线扫图像数据集上的实验结果均表明,本文方法在增加了少量参数的情况下,有效提高了对毫米波图像中隐匿目标的检出能力,并且与现有主流方法相比,实时性仍然优异,且模型在 *P*、*R* 和 mAP@0.5 的性能均是最好。综上所述,本文提出的多尺度反向校正增强与无损下采样的毫米波图像目标检测方法,通过优化特征提取、增强信息保留以及改进锚框生成等策略,在保证检测实时性的同时,能够有效提升毫米波图像中隐匿目标的检测精度与鲁棒性,具有重要的实际应用价值。

参考文献

[1] LIU CH Y, 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.

[2] MENG ZH CH, ZHANG M, WANG H X. CNN with pose segmentation for suspicious object detection in MMW security images[J]. Sensors, 2020, 20(17): 4974.

[3] 金颖康, 王璞. 毫米波成像技术在人体安检领域的应

- 用探讨[J]. 中国安防,2019,26(9):76-80.
- JIN Y K, WANG P. Exploration of the application of millimeter wave imaging technology in the field of human security inspection[J]. China Security,2019,26(9):76-80.
- [4] 洪伟,余超,陈继新,等. 毫米波与太赫兹技术[J]. 中国科学信息科学,2016,46(8):1086-1107.
- HONG W, YU CH, CHEN J X, et al. Millimeter wave and terahertz technology[J]. China Science and Information Science,2016,46(8):1086-1107.
- [5] 杜琨,王威,年丰,等. 主动毫米波图像的人体携带危险物检测研究[J]. 系统工程与电子技术,2016,38(6):1462-1469.
- DU K, WANG W, NIAN F, et al. Concealed objects detection in active millimeter-wave images[J]. Systems Engineering and Electronics,2016,38(6):1462-1469.
- [6] MENG Y, QING AN Y, LIN CH, et al. Passive millimeter wave imaging system based on helical scanning[J]. Scientific Reports,2018,8(1):7852.
- [7] WANG ZH M, CHANG T Y, CUI H L. Review of active millimeter wave imaging techniques for personnel security screening[J]. IEEE Access,2019,7:148336-148350.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unifed, real-time object detection [C]. Computer Vision & Pattern Recognition. IEEE,2016.
- [9] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.
- [10] WANG CH, SHI J, ZHOU Z N, et al. Concealed object detection for millimeter-wave images with normalized accumulation map [J]. IEEE Sensors Journal, 2020, 5(99):6468-6475.
- [11] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [12] YUAN M H, ZHANG Q SH, LI Y W, et al. A suspicious multi-object detection and recognition method for millimeter wave SAR security inspection images based on multi-path extraction network [J]. Remote Sensing, 2021, 13(24):4978.
- [13] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804. 02767,2018.
- [14] FAN L, YANG Q, DENG B, et al. Concealed object detection for active millimeter wave imaging based CGAN data augmentation [C]. 2021 14th UK-Europe-China workshop on Millimetre-waves and Terahertz Technologies. IEEE,2021:1-3.
- [15] MIRZA M, OSINDERO S. Conditional generative adversarial nets [J]. ArXiv preprint arXiv. 1411. 1784,2014.
- [16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv. 2004. 10934,2020.
- [17] 赵梓杉,桑海峰. 基于改进的 YOLOv5 的交通锥标检测系统[J]. 电子测量与仪器学报,2023,37(2):56-64.
- ZHAO Z S, SANG H F. Traffic cone marker detection system based on improved YOLOv5 [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(2):56-64.
- [18] GAO SH H, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021(2):43.
- [19] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects[C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2022: 443-459.
- [20] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:10093-10102.
- [21] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [22] LI C Y, LI L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209. 02976,2022.
- [23] WANG C Y, BOCHKOVSKIY A, LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. ArXiv preprint arXiv: 2207. 02696,2022.
- [24] VARGHESE R, SAMBATH M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness [C]. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024:1-6.
- [25] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405. 14458,2024.

- [26] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural Enhancements [J]. ArXiv preprint arXiv:2410.17725, 2024.

作者简介



叶学义, 2006年于中国科技大学获得博士学位, 现为杭州电子科技大学副教授, 主要研究方向为图像处理、模式识别与信息隐藏。

E-mail: xueyiye@hdu.edu.cn

Ye Xueyi received his Ph. D. degree from University of Science and Technology of China in 2006. He is now an associate professor at Hangzhou University of Electronic Science and Technology. His main research interests include image processing, pattern recognition, and information hiding.



韩卓(通信作者), 现为杭州电子科技大学硕士研究生, 主要研究方向为深度学习与图像处理。

E-mail: 221080050@hdu.edu.cn

Han Zhuo (Corresponding author) now is a M. Sc. candidate at Hangzhou University of Electronic Science and Technology. His main research interests include deep learning and image processing.



蒋甜甜, 现为杭州电子科技大学硕士研究生, 主要研究方向为图像处理、计算机视觉。

E-mail: 1253933054@qq.com

Jiang Tiantian now is a M. Sc. candidate at Hangzhou University of Electronic Science and Technology. Her main research interests include image processing and computer vision.