

DOI:10.19651/j.cnki.emt.2313749

基于先验距离约束的3D卷积毫米波 雷达目标检测方法^{*}

杨文慧 杨宜菩 杨帆 郭亚 张玉博

(河北工业大学电子信息工程学院 天津 300401)

摘要: 用于辅助和自动驾驶系统的各种传感器中,相机和激光雷达的感知性能受天气影响较大,而车载毫米波雷达是一种低成本且几乎不受天气影响的全天候工作器件,对于运动的物体,可提取丰富的多普勒信息。随着雷达技术和开源标注数据集的发展,基于底层雷达数据的目标检测已经成为一个非常有前景的领域。为解决车载毫米波雷达数据的角度分辨率低导致的目标检测和定位不准确的问题,并提升毫米波雷达目标检测的性能,提出了一种基于先验距离约束的3D卷积毫米波雷达目标检测方法,以实现多种动态目标的检测及分类。在本文方法中,通过设计3D ResNet的特征提取器来表征距离-角度-多普勒张量中的目标信息,解决现有的模型因忽略来自原始3D雷达信号的多普勒信息而表征不足的问题;其次,添加了绝对距离损失函数来训练模型,克服距离对目标呈现的影响,提高目标检测的准确性和鲁棒性;此外,还提出了分距离单元区间重新设置先验框的方法,解决现有方法中先验框设计不合理的问题。所提出的模型在RADDet数据集上进行训练以及测试,实验结果表明:与目前的最先进的方法相比,本文模型在IoU阈值为0.1、0.3、0.5、0.7时均达到最优,其中IoU为0.1和0.3时提升最为显著方法,分别提升了6.6%和5.1%。

关键词: 毫米波雷达;目标检测;先验距离约束;3D卷积;自动驾驶

中图分类号: TP2 **文献标识码:** A **国家标准学科分类代码:** 510.1050

3D convolutional millimeter wave radar target detection method based on prior distance constraints

Yang Wenhui Yang Yipu Yang Fan Guo Ya Zhang Yubo

(School of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: Among the varIoUs sensors used in assisted and automatic driving systems, the perception performance of cameras and lidar is greatly affected by the weather, while the automotive millimeter-wave radar is a low-cost and almost weather-free all-weather working device. Moving objects can extract rich Doppler information. With the development of radar technology and open-source labeled datasets, object detection based on underlying radar data has become a very promising field. To address the issue of inaccurate target detection and positioning caused by low angular resolution of onboard millimeter wave radar data, and to improve the performance of millimeter wave radar target detection, this paper proposes a 3D convolutional millimeter wave radar target detection method based on prior box distance constraint to achieve the detection and classification of multiple dynamic targets. In our method, we design a feature extractor for 3D ResNet to characterize target information in Range Azimuth Doppler tensor, solving the problem of insufficient representation in existing models due to ignoring Doppler information from the original 3D radar signal; Secondly, an absolute distance Loss function is added to train the model to overcome the influence of distance on target presentation and improve the accuracy and robustness of target detection; In addition, a method of resetting prior boxes based on distance unit intervals has been proposed to solve the problem of unreasonable prior box design in existing methods. The proposed model is trained and tested on the RADDet dataset. The experimental results show that compared with the current state-of-the-art method, our model achieves the best when the IoU threshold is 0.1, 0.3, 0.5, 0.7, where IoU The improvement is the most significant when it is 0.1 and 0.3, which are increased by 6.6% and 5.1% respectively.

Keywords: millimeter wave radar; target detection; prior distance constraints; 3D convolution; automatic driving

0 引 言

随着自动驾驶时代的来临,自动驾驶技术对周围环境

的感知能力提出了更高的要求。在实现自动驾驶功能时,准确、可靠的目标检测和识别对于车辆的安全性至关重要。而车载毫米波雷达,又称调频连续波(frequency

收稿日期:2023-05-27

^{*} 基金项目:京津冀基础研究合作专项(H2021202008, J210008)、内蒙古自治区纪检监察大数据实验室开放课题(IMDBD202105)项目资助

modulated continuous wave, FMCW) 雷达, 则成为高级辅助驾驶系统(advanced driver assistance systems, ADAS)以及自动驾驶系统中不可或缺的关键传感器。相较于其他机载传感器, 毫米波雷达展现出卓越的鲁棒性和可靠性, 尤其在恶劣天气和复杂光照环境下表现出色。它能够高度估计目标的距离、相对速度和角度, 并通过检测到的目标特征对其进行有效分类。然而, 传统的毫米波雷达在角分辨率方面存在局限, 这导致了目标检测和定位的不准确性。随着毫米波雷达器件工艺水平和数字信号处理技术的不断进步, 其在自动驾驶领域的重要性日益凸显。利用底层毫米波雷达数据, 充分挖掘毫米波雷达的潜力已经成为一种主要趋势。本研究旨在利用基于深度学习的目标检测算法, 以解决汽车雷达数据角度分辨率低导致的目标检测和定位不准确的问题, 并提升毫米波雷达目标检测的性能。

毫米波雷达数据主要有两种解释格式, 一种格式是稀疏点云, 它是模拟数字转换数据(analog-to-digital converter, ADC)经过传统雷达信号处理(距离维度快速傅里叶变换(fast Fourier transform, FFT)、速度维度 FFT、角度维度 FFT 和恒虚警率(constant false-alarm rate, CFAR)检测后的输出格式, 它是轻量级的, 可以使用激光雷达点云算法进行处理和应用, 由于雷达点非常稀疏, 信息缺失, 使得它只适合于目标检测而难以识别。另一种是底层雷达数据, 包含了 ADC、距离-角度(range-angle, RA)频谱、距离-多普勒(range-Doppler, RD)频谱以及距离-角度-多普勒(range-azimuth-Doppler, RAD)张量。虽然这种格式在解释和标注目标方面更加复杂且具有挑战性, 但它包含了丰富的特征, 如速度, 反射强度, 高分辨率距离剖面 and 微多普勒特征。当前基于 RAD 张量的毫米波雷达目标检测方法在处理 RAD 张量时, 常常采用压缩为二维视图^[1-3]或使用 2D 卷积^[4]的方式。Major 等^[1]和 Ouaknine 等^[2]都采用了相似的多视图结构。将 RAD 张量按照 3 个视角投射为 3 个二维视图来实现车辆检测与语义分割。RAMP-CNN 等^[3]也采用了多视图结构, 使用了 3 个编码器-解码器来提取特征。RADDet^[4]将 RAD 张量归一化输入网络, 采用 2D 卷积进行特征提取, 未充分利用多普勒速度的空间分布。这些方法不可避免地导致了信息的丢失, 因为它们未能全面获取 RAD 张量中的信息。为解决这一问题, 本文引入了三维卷积技术, 并将其融合到毫米波雷达目标检测模型中, 显著提升了模型获取信息的能力。通过三维卷积, 模型能够更全面地表达 RAD 数据的特征, 从而有效提高毫米波雷达目标检测的准确性和性能。

在统计和分析数据的标签情况时, 发现目标的各个维度随着距离的变化呈现出不同的变化趋势, 尤其是目标多普勒和高度受到距离影响较大。实际上, 目标的变化可以视作距离变化的映射。基于这一启示, 本文分别引入了绝

对距离损失函数和分距离区间聚类初始化先验框方法, 以减小距离对目标检测性能的影响, 从而显著提高了目标的检测精度。

本文提出了一个基于 RAD 数据立方体的汽车毫米波雷达目标检测管道, 并进一步探讨了使用深度学习技术的目标检测算法在车载毫米波雷达数据中的潜在应用。与以往的方法相比, 本文在如下 3 个方面做出了重要贡献:

1) 提出了一种基于 3D 卷积的毫米波雷达多目标检测模型, 通过多个 3D 卷积残差结构对 RAD 频谱进行特征的提取和表征。相比于传统的 2D 卷积, 3D 卷积不仅能够时间在时间序列上捕捉目标的运动信息, 还能在距离和角度维度上捕捉目标的空间特征。多维度的特征提取使模型能够更全面地捕获毫米波雷达数据中的目标信息, 从而显著提升了模型的目标检测性能。

2) 设计了一种新的损失函数, 即绝对距离损失函数。考虑到距离对目标的呈现影响较大, 针对目标框中心的绝对距离设计了该损失函数, 并将其用于模型训练。实验证明, 这一损失函数的设计显著提升了毫米波雷达目标检测的精度。

3) 提出了一种不同距离单元段生成初始化先验框的方法。在对数据进行统计分析时发现, 随着距离单元的增加, 目标框的 3 个维度呈现出不同的变化趋势。为了获得更能代表真实目标框的先验框, 通过分距离单元区间设置先验框的策略, 设计了更能准确代表目标框的先验框, 并通过实验证明, 这种方法显著提升了目标检测的性能。

1 介 绍

1.1 相关工作

目前, 雷达输出的数据主要包含两种格式。一种是携带目标的距离、角度、速度等信息的点云数据。另一种是底层雷达数据。对于毫米波雷达数据, 最初的输出是中频信号, 经过数字信号处理, 得到 ADC 雷达数据, 然后进行雷达信号处理, 生成 RAD 频谱或点云数据。然而, 在毫米波雷达目标检测及相关任务中, 这两类数据在处理 and 性能上存在一定的区别。

由于雷达点云的稀疏特性, 为获得更好地性能, 在数据处理中往往采用聚类算法。Schumann 等^[5]采用聚类算法得到目标感兴趣区域, 通过提取手工设计的特征, 送入分类器进行分类。随着深度学习的发展, 卷积神经网络逐渐取代了人工设计特征的方式^[6-7]。Danzer 等^[8]首先生成感兴趣区域, 接着利用 PointNet 算法^[8]进行特征提取并进行分类, 边界框回归模块实现车辆 2D 边界框回归。文献[7]采用 PointNet++^[9]对雷达点云进行语义分割。Dreher 等^[10]将雷达点云累积到占用网格映射(occupancy grid map, OGM)中, 然后应用 YOLOv3^[11]进行目标检测。Scheiner 等^[12]比较了两阶

段聚类方法、基于 OGM 的方法和基于 PointNet 的方法在二维检测方面的性能,实验表明,基于 OGM 的方法表现最好。Liu 等^[13]认为结合全局信息可以帮助解决雷达点云的稀疏性问题,在 PointNet++ 中的每个集合抽象层添加了一个 gMLP^[14]块,gMLP 块有望以可承受的计算成本提取全局特征。此外文献^[15-18]利用雷达点云与图像融合实现目标的分类与定位。

虽然深度学习已经在图像和激光雷达方向上取得很大进展^[19],但在雷达领域却发展缓慢。由于多普勒效应,雷达张量数据完全取决于相对速度。因此,使用标准的深度学习预训练网络进行雷达目标检测并不容易。近年来,一些数据集^[4,20-22]的开放,包含了底层雷达数据,为雷达张量的深度学习目标检测方法提供了先决条件,推动了深度学习在雷达领域的更广泛应用。这种趋势鼓励了利用底层雷达数据进行目标检测和相关任务的研究。RADDet^[4]将 RAD 张量归一化输入网络,采用 RadarResnet 进行特征提取,YOLO_head 用于解码,其中存在的缺点是该方法采用 2D 卷积,无法利用多普勒速度的空间分布。RODNet^[23]通过对啁啾序列进行采样,以减少输入维度,降低计算成本。RTCNet^[24]在恒虚警率(constant false-alarm rate,CFAR)检测到的每个点周围裁剪一个小立方体来减少输入张量的大小,再使用三维 CNN 对这些小立方体进行分类,但其检测性能受到 CFAR 检测器的限制。为充分探索 RAD 张量中的信息,Major 等^[1]和 Ouaknine 等^[2]都采用了相似的多视图结构。将 RAD 张量按照 3 个视角投射为 3 个二维视图,3 个解码器分别从 3 个视图中解码特征实现车辆检测与语义分割。不同的是,在特征融合方式上,Ouaknine 等直接串联了 3 个特征图,而 Major 等通过沿缺失维度复制这些二维特征图来恢复张量形状,最后使用三维卷积层来融合它们;另一个区别是 Major 等使用了跳跃连接方式,而 Ouaknine 等采用了 ASPP^[25]模块来编码来自不同分辨率的信息。RAMP-CNN 等^[3]也采用了多视图结构,使用了 3 个编码器-解码器来提取特征,在特征融合方式上与 Major 相似,此外,还提出了几种数据增强的方法。Meyer 等^[26]在雷达原始张量上采用图神经网络,实现了三维目标检测性能上的提升。Li 等^[27]利用连续雷达图像帧的时间信息来识别目标。

1.2 毫米波雷达信号处理

毫米波雷达发射波有脉冲波和连续波两种,本文采用的是 FMCW 雷达采集的数据。毫米波雷达信号产生的原理如图 1 所示,合成器产生连续的线性调频脉冲信号 FMCW,也叫啁啾信号,由发射天线将啁啾信号发射出去,接收天线接收目标对象反射的回波,最后经过混频器得到中频信号。

如图 2 所示,FMCW 信号的特征包含起始频率(也称为载波频率) f_c 、扫描带宽 B 、线性调频持续时间 T_c 以及斜率 $S_f = B/T_c$ 。在一个线性调频持续时间内,频率从 f_c

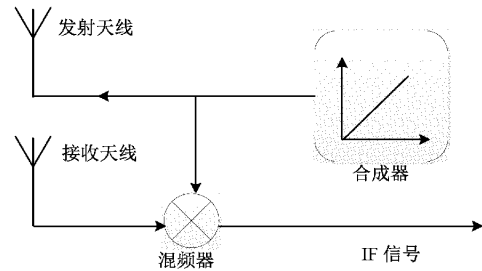


图 1 车载毫米波雷达系统

线性增加到 $f_c + B$,斜率为 S_f 。一个 FMCW 波形称为一个啁啾,毫米波雷达以等时间间隔 T_c 发射 N 个啁啾。接收的回波经过混频器后得到中频信号,如式(1)所示。

$$x_{IF} = \mathbf{A} e^{j(2\pi f_{IF} t + \varphi_{IF})} \quad (1)$$

其中, \mathbf{A} 是振幅, $f_{IF} = f_{T(r)} - f_{R(v)}$ 为差频, φ_{IF} 是 IF 信号的相位,如式(2)、(3)所示。

$$f_{IF} = \frac{2S(r + vT_c)}{c} \quad (2)$$

$$\varphi_{IF} = \frac{2\pi(r + vT_c)}{\lambda} \quad (3)$$

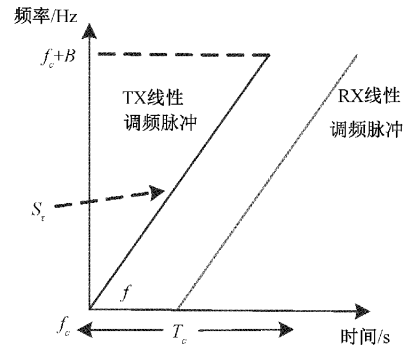


图 2 FMCW 信号

中频信号为模拟信号,经过数字信号处理技术进行离散化后得到 ADC 信号。如图 3 所示,ADC 信号沿着快时间维度(1 个啁啾信号内采样点维度)采用快速傅里叶变换 FFT 来解析信号频率的变化,随后,沿慢时间维度(1 帧内啁啾个数维度)采用 FFT 来解析信号的信号相位的变化。处理后得到 RD 频谱图,通过传统的 CFAR 算法^[28]进行建模,判断目标是否存在同时得到目标对象的距离和速度信息。在此基础上,沿接收天线数量维度应用 FFT 来解析多个天线之间的相位变化,最终得到 RAD 张量。

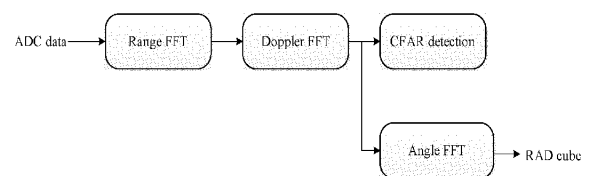


图 3 雷达信号处理流程

2 基于先验距离约束的 3D 卷积毫米波雷达目标检测算法

本章旨在介绍所提出的毫米波雷达目标检测与识别模型,模型的整体架构,如图 4 所示,该模型的阐述主要分为 4 个关键部分:特征提取模块、特征解码模块、损失函数和先验框设置。首先,在特征提取模块中,提出了 3D ResNet 网络,旨在更加全面学习并提取雷达张量特征信息。3D 卷积在提取特征时能够更加准确地表述物体在不同维度(距离、角度、多普勒)上的信息,同时能够探索

物体之间的空间位置关系。其次,特征解码模块借鉴了 YOLO^[29] 网络的解码器部分,通过几层卷积对提取的特征进行解耦并回归目标位置、类别、置信度以及距离信息。接下来,目标检测的损失函数主要涵盖目标对象的定位、类别以及置信度。为减小距离对目标呈现的影响,本文提出了绝对距离损失函数。最后,在先验框设置部分,为设计更加符合数据特征的先验框,提出了不同距离单元段生成初始化先验框的方法,将距离作为聚类生成先验框的重要参数,以更好地适应不同距离下目标的尺寸变化。

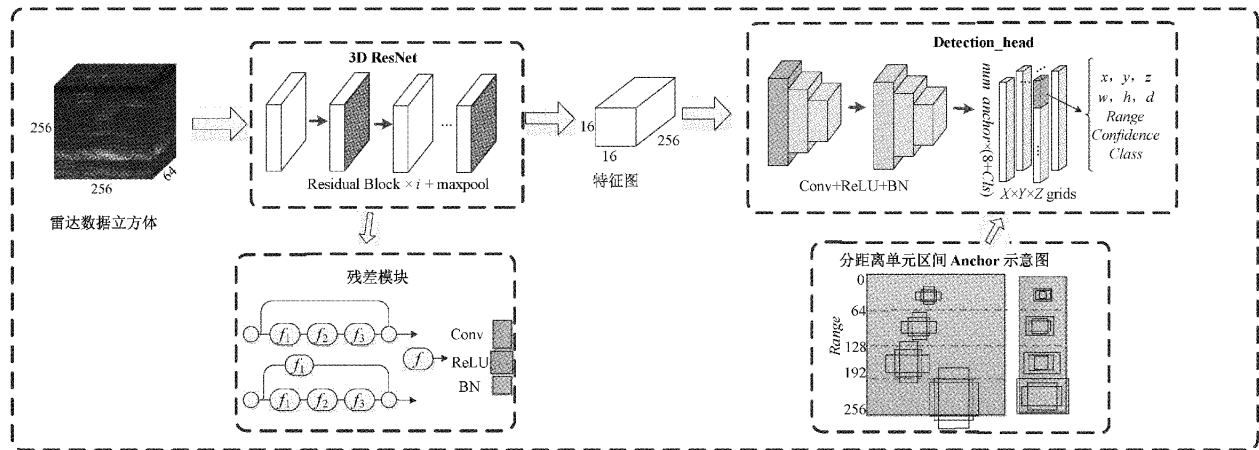


图 4 模型基本流程

2.1 特征提取模块

为解决目前方法对 RAD 雷达数据表征不足的问题,本文提出了 3D ResNet 特征提取网络。该网络结合了三维卷积和 ResNet 网络,以更好地挖掘和利用雷达张量的各个维度特征,同时探索目标在 3D 空间中的相对关系。首先,通过采用三维卷积核,在雷达张量的距离、角度和多普勒三个维度上进行滑动卷积,相比于传统的二维卷积核,能更精确地提取特征。这种设计使得网络能够有效地表征物体在不同维度上的信息,从而更好地理解目标的空间特性。其次,3D ResNet 借鉴了经典的 ResNet 网络,有效地解决了网络退化和梯度消失等问题,并能够提取深层次的特征信息。这种结构使得网络能够更深入地理解雷达数据中的特征,有助于提高目标检测和识别的性能。为了更好地适应雷达数据立方体的特点,本文将 ResNet 网络中的残差结构的卷积核替换为三维卷积核。同时,王龙等^[4] 实验结果表明,使用最大池化层代替残差块进行下采样也能显著提高性能。因此,在残差块后引入最大池化层进行特征压缩,进一步深度解读雷达张量中的目标信息。

如图 5 所示,3D ResNet 特征提取模块主要由残差结构(Block1)以及池化层(Block2)组成,共包含 30 个残差块和 4 个最大池化层,残差有恒等映射与非恒等映射两种结构,在通道数扩张前使用非恒等映射残差结构,残差块中使用的卷积核包含 $3 \times 3 \times 3$ 和 $1 \times 1 \times 1$ 两种,后经 Relu 激

活函数。在分别经过 2、4、8、16 个残差块后进行下采样。提取高噪声下底层雷达中对象的显著特征。通道数扩张分为 2 倍和 4 倍,为降低网络的复杂程度,最后一次通道数扩张 2 倍。特征提取模块的输入数据形状为 $(256, 256, 64, 1)$ 的张量,经过四次通道数扩张和下采样后,得到特征输出为 $(16, 16, 4, 128)$ 。该网络的主要优点在于将 ResNet 与三维卷积结合,更加适用于底层雷达数据。特征提取器在保留 ResNet 网络优点的同时,加强了网络在雷达数据的距离、角度以及多普勒维度的信息提取能力,进一步探索了目标的三维空间分布。

2.2 特征解码模块

基于先验框距离约束的 3D 卷积毫米波雷达目标检测是基于锚框的方法,如图 6 所示,在每个特征点上生成不同比例的锚框,并通过模型的输出对锚框进行调整和筛选,从而获得预测框。特别地,本模型采用了 YOLO^[29] 的检测头进行特征解码。由于研究任务是底层雷达数据下的三维目标检测,为了适应这一任务,对现有的 YOLO_Head 进行了修改。具体而言,将 YOLO_Head 中输出特征图的输出格式从 $(w, h, d, num_of_anchors \times (5 + num_of_classes))$ 修改为 $(w, h, d, num_of_anchors \times (7 + 1 + num_of_classes))$,如图 7 所示。其中, (w, h, d) 代表特征提取模块输出的特征图尺寸; $num_of_anchors$ 表示预设的先验框数量,6 个先验框是通过 K-means 算法聚类产

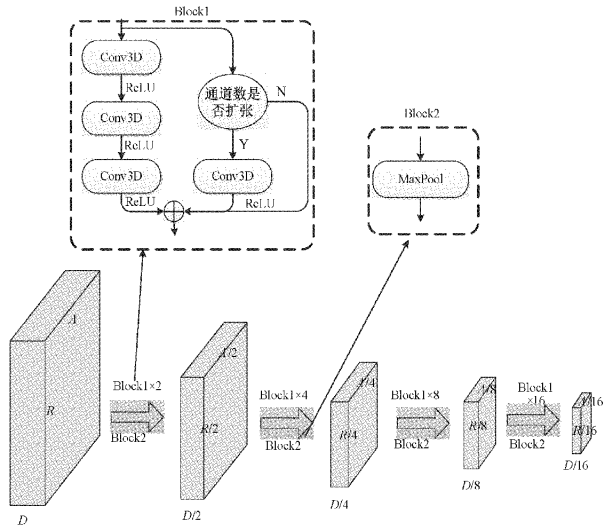
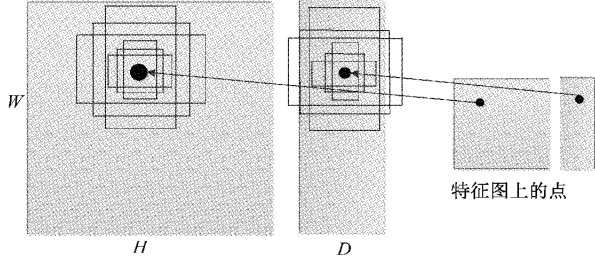


图5 3D ResNet 特征提取器

生;7代表 *objectness* 和 3D *box* 信息,3D *box* 信息由三维框中心 $[x, y, z]$ 和大小 $[w, h, d]$ 组成;1表示预测目标的距离信息 *dst*; *num_of_classes* 表示目标的种类,本文共检测6种类别的目标。经过分解,最终输出格式为 $(w, h, d, num_of_anchors \times (7+1+ num_of_classes))$,通过对先验框进行调整,得到预测框,随后送入非极大值抑制(non-maximum suppression, NMS)以求得最优预测框。



原图上anchor分布

图6 锚框示意图

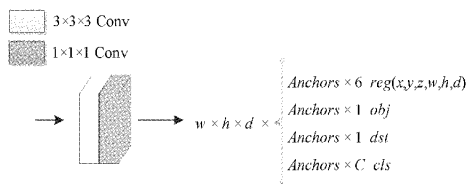


图7 多任务检测头

2.3 损失函数设计

模型输出分为4个任务,即位置回归、目标置信度、类别以及目标距离预测。其中,前3个任务的输出结构与YOLO_Head^[29]相似。由2.4节可知,距离越远标签框越小,预测越不准确。因此,为减小距离对模型感知性能的影响,本文增加了对目标距离的预测,设计了一个绝对距离损失函数来预测目标与雷达之间的距离,使模型在不同距离下的预测更加准确。总损失 L_{total} 如式(4)所示。

$$L_{total} = \beta \times L_{box} + L_{obj} + L_{class} + L_{range} \quad (4)$$

其中, L_{box} 是位置回归损失, β 为系数, L_{obj} 是目标置信度损失, L_{class} 是目标类别损失, L_{range} 是绝对距离损失。

位置回归损失 L_{box} 的损失函数源自^[30],在二维目标框损失的基础上添加多普勒维度的位置损失,以适应三维检测任务。如式(5)所示。

$$L_{box} = \left(\sum_{l \in \{x, y, z\}} (l' - l)^2 + \sum_{l \in \{w, h, d\}} (\sqrt{j'} - \sqrt{j})^2 \right) \times h \quad (5)$$

从式(5)中可以看到,位置损失主要由两部分组成,一部分是中心点 (x, y, z) 的损失, l' 和 l 分别代表预测值和真实值,另一部分是目标框的各维度 (w, h, d) 引起的损失, j' 和 j 分别表示预测结果和真实值。 h 表示真实目标的置信度,只在有目标时进行损失的计算。在训练过程中,由于位置损失对总损失值的影响较大,因此,本文为位置回归损失设置了系数 $\beta = 0.1$ 。

Focal Loss^[31]用于训练目标置信度损失 L_{obj} ,其中,主要用交叉熵来计算正负样本的损失值,然而,为了使模型更加关注正样本,为负样本引入了一个系数,并将其设置为0.01,以对负样本进行惩罚。对于类别损失 L_{class} ,使用交叉熵^[32]来计算损失值,与置信度损失不同,类别损失只关注正样本。

毫米波雷达作为一种距离传感器,由2.4节中距离与标签框的关系可知,目标的呈现受距离的影响较大。因此,为减小距离对预测结果的影响,本文提出了绝对距离损失函数 L_{range} 。该损失函数中计算预测目标与雷达之间的距离借鉴了 *Smooth L1* 损失函数,同时,绝对距离损失函数仅关注正样本的训练。具体如式(6)所示。

$$L_{range} = s \times |r_{pred} - r_{gt}| \quad (6)$$

式中: s 代表目标的置信度, r_{pred} 为预测的目标框中心的距离值, r_{gt} 为归一化的目标真实框的中心距离值。

2.4 聚类初始化先验框

对 *anchor_based* 的目标检测来说,合理的先验框设计是模型实现目标准确检测的前提。目前的方法并没有关注不同距离下的目标呈现差异的问题,设置的先验框的未能全面反映数据特征,为聚类出更加合理的先验框,本文将距离作为聚类范围的标准,提出了分距离单元区间聚类并初始化先验框的方法,提高模型在不同距离下目标检测的性能。

为了解目标真实框在距离单元范围内的分布情况,在RA视图下,在角度维度上的单元数作为目标边框的宽度,记为 $width_cells$;在距离维度上的单元数作为目标边框的高度,记为 $height_cells$; $width_cells$ 和 $height_cells$ 的乘积作为目标框的面积,记为 ra_area ;目标对应的距离单元数记为 $range_cells$ 。如图8(a)~(c)所示,分别以 ra_area 、 $width_cells$ 、 $height_cells$ 为纵轴, $range_cells$ 为横轴,生成三幅统计直方图。观察并分析图8(a)~(c)发

现,随距离的增加,目标真实框的面积值在减小,宽度变化并不明显,高度却下降明显。在 RD 视图下,将多普勒维度上的单元数作为目标框的多普勒维度,记为 $doppler_cells$;将 $width_cells$ 和 $doppler_cells$ 的乘积作为目标框的面积,记为 rd_area 。如图 9(a)~(c)所示,分别以 rd_area 、 $width_cells$ 、 $doppler_cells$ 为纵轴,

$range_cells$ 为横轴,生成 3 幅统计直方图。观察并分析图 9(a)~(c)发现,随距离的增加,目标真实框的面积在减小,宽度变化并不明显,但多普勒值下降明显。总之,距离越远,目标框呈现变小的趋势。通过上述分析可知,目标真实框的高度以及多普勒值的下降是目标框呈现变小趋势的主要原因。

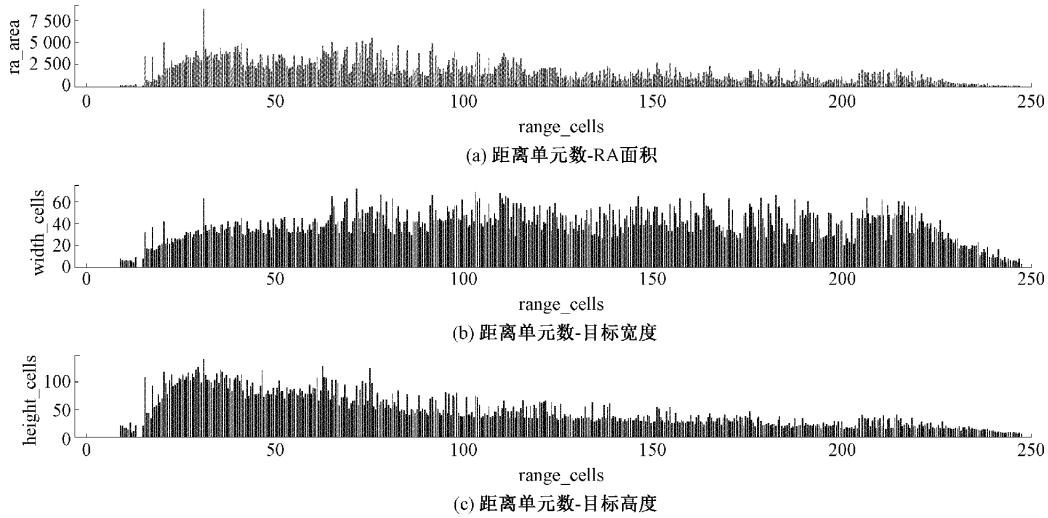


图 8 RA 视角统计直方图

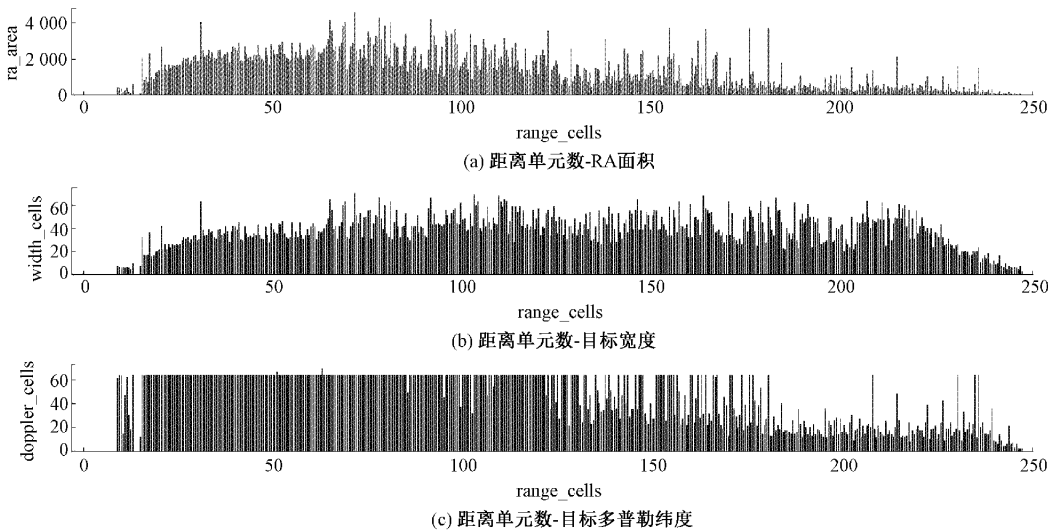


图 9 RD 视角统计直方图

由于目标框随着距离发生变化,为保证先验框设计更符合数据特征,本文提出利用 K-means 算法在不同距离范围内聚类并初始化先验框的方法。聚类算法与 YOLOv3 中先验框聚类的思想相似,具体步骤为:首先,获取目标真实框的长度、宽度和高度,并随机选取 6 个目标框作为初始化的先验框。然后,通过计算每个目标框与每个先验框之间的距离 d ,更新先验框,直到目标框的簇中心不再改变,得到最终所需的先验框。为了减小由距离引起的目标框呈现差异影响,提升目标检测性能,本文将距离单元数

0~255 划分为 4 个区间:0~63、64~127、128~191 和 192~256,并在每个距离单元区间内对其中的目标真实框进行聚类,得到相应的先验框。具体地,目标框与先验框的距离 d 的计算方法如式(7)所示。

$$d = 1 - IoU(boxes, clusters) \quad (7)$$

式中: $clusters$ 代表先验框 (w_i, h_i, d_i) , $i \in (1, 2, \dots, k)$; $boxes$ 代表目标真实框 (w_j, h_j, d_j) , $j \in (1, 2, \dots, N)$; $IoU(boxes, clusters)$ 代表目标框与目前聚类先验框的交并比,所得 IoU 越大,意味着目标框与先验框的距离越近。

本文采用 RADDet 数据集,在划分的 4 个距离单元区间内分别采用 K-means 算法,得到 4 组先验框,每组包含 6

个,如表 1 所示,通过观察每组先验框的宽、高以及多普勒的变化,符合真实框的各个维度随距离变化的趋势。

表 1 不同距离单元区间的先验框

距离区间	Anchor1	Anchor2	Anchor3	Anchor4	Anchor5	Anchor6
0~63	(3 5 2)	(4 7 4)	(4 4 1)	(8 10 5)	(9 9 2)	(19 13 6)
64~127	(3 6 3)	(6 9 4)	(9 13 9)	(16 13 4)	(20 19 7)	(20 21 16)
128~191	(4 8 5)	(8 13 7)	(20 23 10)	(21 28 18)	(22 37 28)	(24 40 64)
192~256	(6 14 9)	(11 25 17)	(13 36 64)	(1742 22)	(2253 35)	(23 69 64)

3 实验及实验结果分析

3.1 实验数据集

本文采用公开数据集 RADDet^[4],数据采集的硬件配置为德州仪器 AWR1843-BOOST2 雷达和一对来自 The Imaging Source 的 DFK 33UX2733 立体相机。采集的环境为晴朗的天气条件,数据标注范围为 5~50 m 的目标,数据集包含 6 种类别,分别为行人、自行车、汽车、卡车、公交车和摩托车。数据集提供经过雷达信号预处理后的

RAD 数据,一共包含 10 158 帧,其中,从雷达捕获的原始 ADC 数据的形状为 (256, 8, 64),在方位角维度上计算 FFT 时使用了零填充,最终生成形状为 (256, 256, 64) 的 RAD 输出。如图 10 所示,数据集提供的标签有 RAD 数据立方体下的三维的标注框以及笛卡尔坐标下 RA 视图的二维标注框。本文使用随机抽样将数据集分为 80% 和 20% 的训练集和测试集,确保训练和测试的数据分布一致。此外,将训练集的 90% 用于训练,10% 用于训练中的验证,以找到最优系数和模型。

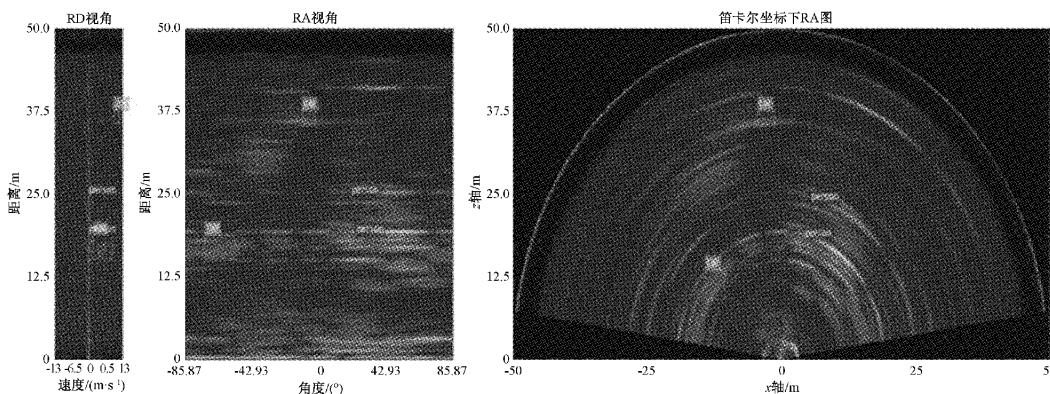


图 10 数据标注

3.2 实验环境及评价指标

本文的模型训练和评估都是基于 RADDet 数据集进行的。训练阶段的参数配置如下:batch_size 设置为 3,初始学习率为 0.000 1,迭代次数为 1000 个 epochs。在训练的前 60 000 个 steps 后,每隔 10 000 个 steps,学习率会衰减 0.96。置信度阈值为 0.5,检测头的 NMS 阈值为 0.3。优化器采用了 Adam 算法^[33]。实验环境为 Ubuntu20.04,使用一张 RTX 1080Ti 显卡,Python 版本为 3.6,训练框架为 TensorFlow 2.3。值得注意的是,所有实验均在相同参数配置下进行训练,以确保结果的可比性和一致性。

在对算法性能进行评估时,本文采用了平均精度 (average precision, AP) 作为评价指标。AP 值反映了查准率 (precision) 和查全率 (recall) 曲线下的面积,是衡量目标检测算法性能的重要指标。查准率和查全率的计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

其中,TP (true positive) 为真正例,即预测正确的正例;FP (false positive) 为假正例,预测错误的正例;FN (false negative) 为假负例,是未能正确预测出来的正例。查准率用来衡量预测正例中正确的正例的比率,查全率则衡量真正例占总正例的比例,两者呈现反比关系。

3.3 实验结果与分析

为了充分评估并验证本文提出的毫米波雷达目标检测算法的有效性和优越性,本文使用衡量目标检测算法的重要指标即 IOU 阈值下的 AP 值,衡量不同模型在 RADDet 测试集上的 3D 目标检测性能。具体而言,为全面评估模型的检测性能与鲁棒性,设置了 4 组评价指标 AP0.1、AP0.3、AP0.5、AP0.7,分别表示 IOU 阈值在

0.1、0.3、0.5 以及 0.7 时的 AP 值。将本文方法 3D ResNet_AL_DA 与先前取得 SOTA 的方法 VGG、RadarResNet(max-pool)、RadarResNet(strided convolution)以及 RadarResNet 的变体(RadarResNet+SA、RadarResNet+SAUNet、RadarResNet+MLP)进行了对比实验,其中,RadarResNet(max-pool)为本文的基线模型,在同等条件(硬件设备、参数配置)下对比了 7 种模型在 4 种 IOU 阈值下的 AP 值。

如表 2 所示,3D ResNet_AL_DA 与其他 6 个模型在 RADDet 测试集上的平均检测精度的对比结果显示,3D ResNet_AL_DA 的平均检测精度在所有评估指标下均达到了最优。在被对比的 6 个模型中,RadarResNet(max-pool)的表现最突出,VGG 和 RadarResNet(max-pool)、

RadarResNet(strided convolution)以及 RadarResNet 的变体模型相比,说明了基于 ResNet 的主干更加适合雷达目标检测任务,同时也反映出使用残差块比顺序卷积层更合适探索雷达目标信息。在对比模型 RadarResNet(max-pool)和 RadarResNet(strided convolution)时发现,使用最大池层而不是残差块进行下采样也可以显著提高模型的检测性能。其次,RadarResNet+SA、RadarResNet+SAUNet 将自注意力层^[34-35]引入模型中,试图探索自注意力层在雷达目标检测任务中的潜力,结果表明,检测精度并没有高于 RadarResNet(max-pool),说明基于图像的自注意力层不太适用于基于雷达的任务。此外,模型 RadarResNet+MLP 在主干末端添加了多层感知机 MLP,但结果表明,MLP 的加入对于提升模型感知性能是有限的。

表 2 不同模型在 RADDet 测试集上 3D 目标检测性能对比

检测网络	AP _{0.1}	AP _{0.3}	AP _{0.5}	AP _{0.7}
VGG ^[4]	0.685	0.499	0.209	0.035
RadarResNet(strided convolution) ^[4]	0.727	0.538	0.243	0.054
Baseline(RadarResNet(max-pool) ^[4])	0.764	0.563	0.251	0.059
RadarResNet+SA ^[34]	0.738	0.498	0.219	0.044
RadarResNet+SAUNet ^[35]	0.752	0.559	0.251	0.053
RadarResNet+MLP ^[4]	0.759	0.521	0.217	0.043
3D ResNet_AL_DA(本文)	0.830	0.614	0.278	0.061

与其他 6 种模型不同,3D ResNet_AL_DA 在特征提取部分引入了三维卷积来探索多普勒维度信息。同时,为减小距离对感知性能的影响,在模型训练中添加了绝对距离损失函数。此外,提出了分距离区间设置先验框方法,优化先验框的设置。实验结果表明,相比基准基线方法,本文的模型在毫米波雷达目标检测任务的效果上取得了显著地提升,平均检测精度 AP_{0.1}、AP_{0.3}、AP_{0.5}、AP_{0.7} 分别提高了 6.6%、5.1%、2.7%和 0.2%。值得注意的是,IoU 越大,对应的 AP 值提升越小,这是由于传统雷达角分辨率较低,基准检测精度本身不高造成的。总的来说,本文模型在不同 IoU 阈值下都表现出优异的检测性能,验证了所提出的网络在雷达目标检测任务中的有效性和先进性。

3.4 消融实验结果及分析

通过上述对比实验结果,已经证明了最终模型在检测精度的性能上具有显著提升。为验证本文提出的三个改进方法的有效性,分析它们分别对模型检测精度的提升效果,以及探索更合理的网络设计,本文进行了消融实验。

1)RadarResNet(max-pool)作为基线方法。

2)3DResNet 在基线模型的基础上引入了 3D 卷积技术,改进了特征提取方式,利用多普勒信息探索目标空间信息。

3)3D ResNet_AL 模型在 3DResNet 的基础上添加了目标框绝对距离损失函数(AL),减小距离对目标呈现的

影响。

4)3D ResNet_AL_DA 在 3D ResNet_AL 模型的基础上添加了分距离区间聚类并初始化先验框方法(DA),聚类出能真实反映不同距离下目标框特征的先验框。

消融实验结果如表 3 所示,与基线方法相比,3D ResNet 模型的平均检测精度 AP_{0.1}、AP_{0.3}、和 AP_{0.5} 分别增加了 3.1%、3.2%和 1.3%,这是因为与 2D 卷积相比,3D 卷积在探索雷达立方体的多普勒信息以及目标的空间关系方面更具优势,但由于传统毫米波雷达角分辨率低的问题,在 IoU 为 0.7 时,提升模型检测精度是困难的。与未增加绝对距离损失的 3D ResNet 模型相比,3D ResNet_AL 在 AP_{0.1}、AP_{0.7} 下分别提升了 3.3%和 0.4%,绝对距离损失函数的添加,使模型关注到了目标的距离,减小了距离对目标呈现的影响。实验结果说明了其对改善模型目标检测性能有效果,却存在一定的局限性。3D ResNet_AL_DA 与未添加分距离区间聚类并初始化先验框方法的 3D ResNet_AL 模型相比,AP_{0.1}、AP_{0.3}、AP_{0.5} 和 AP_{0.7} 分别提升了 0.2%、2.3%、1.4%和 0.9%,在 2.4 节数据分析中得出,距离越远,目标框越小,分距离区间聚类初始化先验框的方法,考量了距离对目标的影响,设置了更加符合数据特性的先验框。实验结果也表明了它对于改善模型的检测性能起着显著作用,进一步证实了预设合理的先验框在目标检测中的重要性。

表3 改进方法的消融实验

改进模型	AP _{0.1}	AP _{0.3}	AP _{0.5}	AP _{0.7}
3D ResNet_AL_DA	0.830	0.614	0.278	0.061
3D ResNet_AL	0.828	0.591	0.264	0.052
3D ResNet	0.795	0.595	0.264	0.048
Baseline(RadarRes.Net(max-pool))	0.764	0.563	0.251	0.059

3.5 检测结果分析

为进一步验证本文提出的模型在毫米波雷达目标检测任务上的有效性,本节对模型的推理效果图进行了对比分析。本文提出的模型3D ResNet_AL_DA与基准模型RadarResNet(strided convolution)在测试集上的可视化效果,如图11所示,左侧为本文模型的推理效果图,右侧为基准模型推理图,6种目标分别由不同的颜色表示,预测的三

维框由虚线构成,对应的真实框填充了目标颜色。通过观察并对比两个模型的效果图可知,两种模型都表现出较好的目标检测效果,但本文所提出的模型在不同距离下目标的三维边界框更加贴近真实框,进一步表明了本文模型对提升毫米波雷达目标检测的有效性;同时,本文的模型对一些小样本类别如行人、自行车以及一些远距离目标的目标类别检出率更高,目标框的回归更加准确,有更好的鲁棒性。

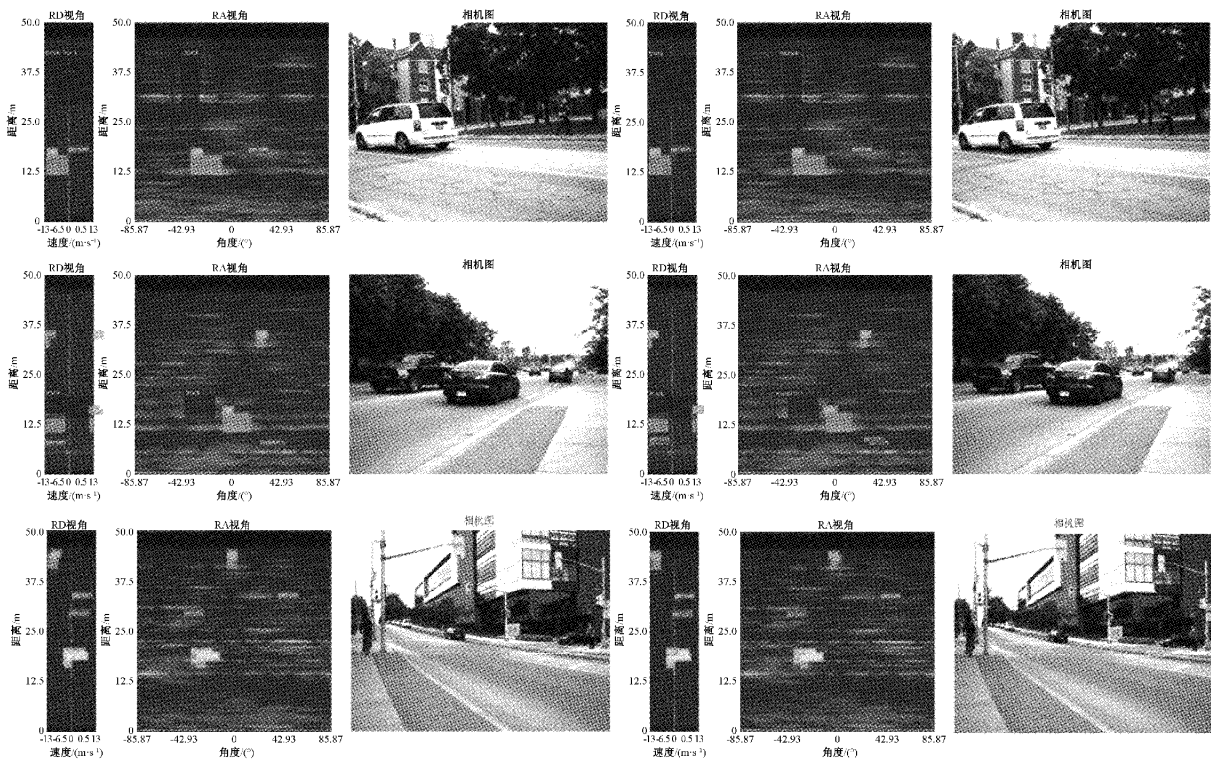


图11 两种模型可视化效果对比

为探索添加绝对距离损失函数与分距离聚类先验框的必要性,进一步对比了消融实验模型:3D ResNet, 3D

ResNet_AL, 3D ResNet_AL_DA在测试集上的推理可视化效果图。如图12~14所示,分别展示了3种模型在同一

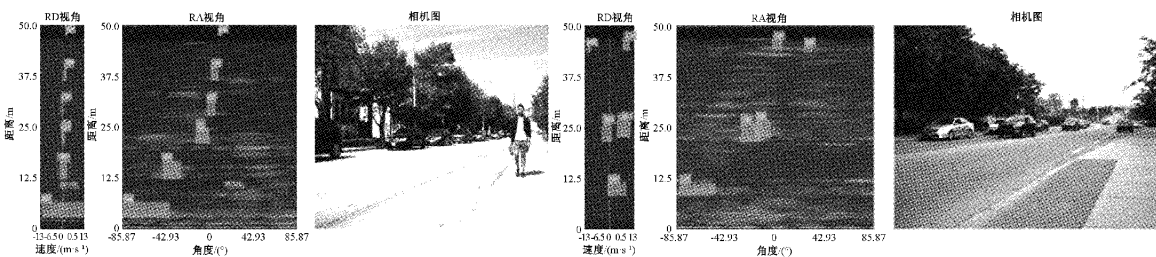


图12 3D ResNet_AL_DA 推理图

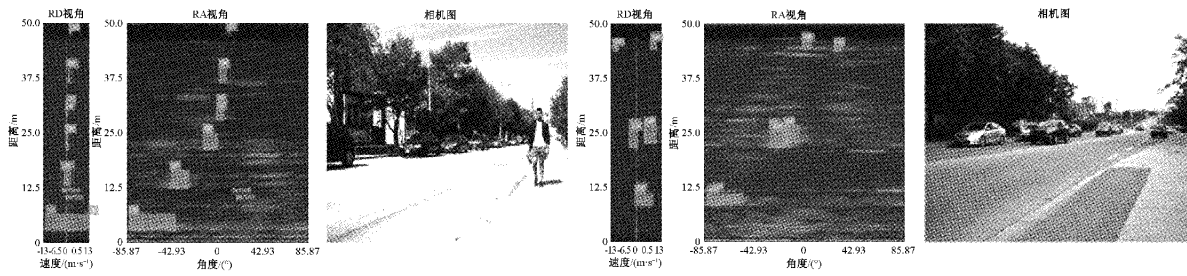


图 13 3D ResNet_AL 推理图

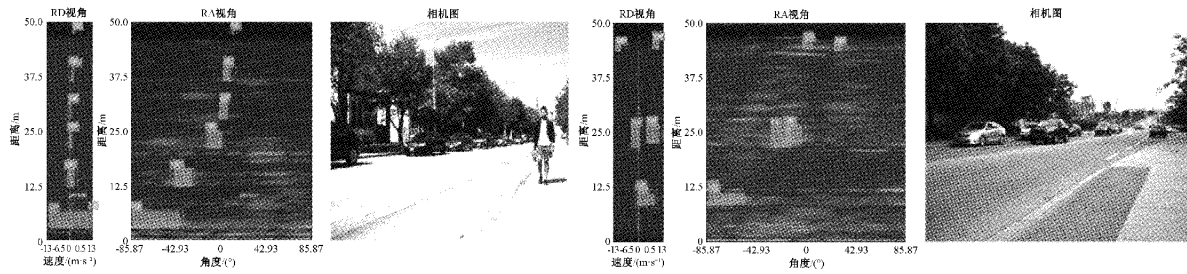


图 14 3D ResNet 推理图

场景下的表现,通过观察并对比以下效果图可知,在不同距离区间内,3D ResNet_AL_DA 预测的三维框更为精准,3D ResNet_AL 对远距离目标定位较差,而 3D ResNet 在不同距离范围内的目标定位效果均较差,充分证实了提出的分距离设置先验框的方法和损失函数的有效性,进一步论证了距离对于雷达检测性能的影响较大,突出了本文改进方法的优势。

4 结 论

针对自动驾驶场景下的毫米波雷达多目标识别与检测任务,本文以底层 RAD 雷达张量为研究对象,提出了基于先验距离约束的 3D 卷积毫米波雷达目标检测模型。在特征提取部分,将 3D 卷积和 ResNet 网络结合,设计了 3D ResNet 特征提取器,有效地解决了当前方法对雷达张量目标信息表征不足的问题。在损失函数部分,为模型增添了目标框绝对对距离损失函数,使其更关注目标距离的变化。此外,本文还提出了分距离单元聚类并初始化先验框的方法,生成更加符合数据特征的先验框,这两种改进极大地减小了距离对检测性能的影响。在公开数据集 RADDet 的测试集上进行评估实验证明,与先前取得 SOTA 的方法相比,本文所提出的模型在所有 IOU 阈值下都取得了显著提升。特别是在 IOU 为 0.3 时,检测精度达到了 61.4%,相比于基线模型提升了 5.1%,充分表明了本文方法的有效性与优越性。

当然,本文的改进也存在一些问题,更适合底层雷达数据的轻量化网络值得进一步探索,以求在速度和精度上实现均衡,未来,雷达目标检测模型研究将朝着轻量化,嵌入式部署方向发展。

参考文献

- [1] MAJOR B, FONTIJNE D, ANSARI A, et al. Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 1-9.
- [2] OUAKNINE A, NEWSON A, PÉREZ P, et al. Multi-view radar semantic segmentation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 15671-15680.
- [3] GAO X, XING G, ROY S, et al. Ramp-CNN: A novel neural network for enhanced automotive radar object recognition [J]. IEEE Sensors Journal, 2020, 21(4): 5119-5132.
- [4] ZHANG A, NOWRUZI F E, LAGANIERE R. RADDet: Range-Azimuth-Doppler based radar object detection for dynamic road users [C]. 2021 18th Conference on Robots and Vision (CRV). IEEE, 2021: 95-102.
- [5] SCHUMANN O, WÖHLER C, HAHN M, et al. Comparison of random forest and long short-term memory network performances in classification tasks using radar [C]. 2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF), IEEE, 2017: 1-6.
- [6] DANZER A, GRIEBEL T, BACH M, et al. 2D car detection in radar data with pointnets [C]. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019: 61-66.
- [7] SCHUMANN O, HAHN M, DICKMANN J, et al.

- Semantic segmentation on radar point clouds[C]. 2018 21st International Conference on Information Fusion (FUSION), IEEE, 2018: 2179-2186.
- [8] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [9] QI C R, YI L, SU H P N, et al. Deep hierarchical feature learning on point sets in a metric space[J]. ArXiv Preprint, 2017, ArXiv:1706.02413.
- [10] DREHER M, ERÇELIK E, BÄNZIGER T, et al. Radar-based 2D car detection using deep neural networks[C]. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2020: 1-8.
- [11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [12] SCHEINER N, KRAUS F, APPENRODT N, et al. Object detection for automotive radar point clouds-a comparison[J]. AI Perspectives, 2021, 3(1): 1-23.
- [13] LIU J, XIONG W, BAI L, et al. Deep instance segmentation with automotive radar detection points[J]. IEEE Transactions on Intelligent Vehicles, 2022, 8(1): 84-94.
- [14] LIU H, DAI Z, SO D, et al. Pay attention to mlps[J]. Advances in Neural Information Processing Systems, 2021, 34: 9204-9215.
- [15] NABATI R, QI H. Centerfusion: Center-based radar and camera fusion for 3D object detection [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 1527-1536.
- [16] CHANG S, ZHANG Y, ZHANG F, et al. Spatial attention fusion for obstacle detection using mmwave radar and vision sensor[J]. Sensors, 2020, 20(4): 956.
- [17] NOBIS F, GEISSLINGER M, WEBER M, et al. A deep learning-based radar and camera sensor fusion architecture for object detection[C]. 2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF), IEEE, 2019: 1-7.
- [18] JOHN V, MITA S. RVNet: Deep sensor fusion of monocular camera and radar for image-based obstacle detection in challenging environments[C]. Pacific-Rim Symposium on Image and Video Technology. Springer, Cham, 2019: 351-364.
- [19] 石欣, 卢灏, 秦鹏杰, 等. 一种远距离行人小目标检测方法[J]. 仪器仪表学报, 2022, 43(5): 136-146.
- [20] OUAKNINE A, NEWSON A, REBUT J, et al. Carrada dataset: Camera and automotive radar with range-angle-doppler annotations [C]. 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021: 5068-5075.
- [21] LIM T Y, MARKOWITZ S A, DO M N. Radical: A synchronized fmcw radar, depth, imu and rgb camera data dataset with low-level fmcw radar signals[J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(4): 941-953.
- [22] DIMITRIEVSKI M, SHOPOVSKA I, VAN HAMME D, et al. Weakly supervised deep learning method for vulnerable road user detection in FMCW radar[C]. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2020: 1-8.
- [23] WANG Y, JIANG Z, LI Y, et al. RODNet: A real-time radar object detection network cross-supervised by camera-radar fused object 3D localization[J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(4): 954-967.
- [24] PALFFY A, DONG J, KOIJ J F P, et al. CNN based road user detection using the 3D radar cube[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 1263-1270.
- [25] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. ArXiv Preprint, 2017, ArXiv: 1706.05587.
- [26] MEYER M, KUSCHK G, TOMFORDE S. Graph convolutional networks for 3D object detection on radar data [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3060-3069.
- [27] LI P, WANG P, BERNTORP K, et al. Exploiting temporal relations on radar perception for autonomous driving[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 17071-17080.
- [28] RICHARDS M A. Fundamentals of radar signal processing[M]. McGraw-Hill Education, 2014.
- [29] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [30] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International Conference on Machine Learning. PMLR, 2015: 448-456.

- [31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [32] NASR G E, BADR E A, JOUN C. Cross entropy error function in neural networks: Forecasting gasoline demand[C]. FLAIRS Conference, 2002: 381-384.
- [33] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. ArXiv Preprint, 2014, ArXiv:1412.6980.
- [34] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C]. International Conference on Machine Learning, PMLR, 2019: 7354-7363.
- [35] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-net: Learning where to look for the

pancreas[J]. ArXiv Preprint, 2018, ArXiv:1804.03999.

作者简介

杨文慧, 硕士研究生, 主要研究方向为计算机视觉下的无人驾驶环境感知。

E-mail: wenhui19980204@163.com

杨宜善, 博士研究生, 主要研究方向为智能雷达多模态数据融合与识别技术。

E-mail: bodhipupu@163.com

杨帆(通信作者), 博士研究生, 教授, 博士生导师, 主要研究方向为电子电路与计算机视觉。

E-mail: yangfan@hebut.edu.cn

郭亚, 博士研究生, 主要研究方向为智能信息处理。

E-mail: 18734166761@163.com

张玉博, 硕士研究生, 主要研究方向为智能信息处理。

E-mail: zhangyubo0722@163.com