ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 36,No. 2,Mar. 2021,pp. 324－333
DOI:10. 16337/j. 1004-9037. 2021. 02. 014

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于雷达与图像数据融合的人体目标检测方法

李文平，袁　强，陈　璐，郑利彪，汤晓龙

（安徽南瑞继远电网技术有限公司，合肥 230088）

摘　要：三维人体目标检测在智能安防、机器人、自动驾驶等领域具有重要的应用价值。目前基于雷达与图像数据融合的三维人体目标检测方法主要采用两阶段网络结构，分别完成目标概率较高的候选边界框的选取以及对目标候选框进行分类和边界框回归。目标候选边界框的预先选取使两阶段网络结构的检测准确率和定位精度得到提高，但相对复杂的网络结构导致运算速度受到限制，难以满足实时性要求较高的应用场景。针对以上问题，研究了一种基于改进型RetinaNet的三维人体目标实时检测方法，将主干网络与特征金字塔网络结合用于雷达点云和图像特征的提取，并将两者融合的特征锚框输入到功能网络从而输出三维边界框和目标类别信息。该方法采用单阶段网络结构直接回归目标的类别概率和位置坐标值，并且通过引入聚焦损失函数解决单阶段网络训练过程中存在的正负样本不平衡问题。在KITTI数据集上进行的实验表明，本文方法在三维人体目标检测的平均精度和耗时方面均优于对比算法，可有效实现目标检测的准确性和实时性之间的平衡。

关键词：三维人体目标检测；多传感器信息融合；深度学习；改进型RetinaNet；聚焦损失函数
中图分类号：TP391.4　　**文献标志码**：A

Human Target Detection Method Based on Fusion of Radar and Image Data

LI Wenping, YUAN Qiang, CHEN Lu, ZHENG Libiao, TANG Xiaolong

(Anhui Narui Jiyuan Power Grid Technology Co Ltd, Hefei 230088, China)

Abstract: Three-dimensional (3-D) human target detection has important application value in intelligent security, robot, automatic driving and other fields. At present, the 3-D human target detection method based on radar and image data fusion mainly adopts two-stage network structure, which respectively completes the selection of candidate boundary boxes with high target probability and the target classification/regression of target candidate boxes. Although the preselection of target candidate bounding box enables the two-stage network structure to achieve higher detection accuracy and positioning accuracy, the complexity of the network structure leads to the limitation of the operation speed, which cannot be applied in scenarios with high real-time requirements. In order to solve the above problem, this paper studies a real-time detection method of 3-D human targets based on improved RetinaNet. The backbone network and feature pyramid network are combined for point cloud and image feature extraction, and the fused feature anchors are input into the functional network to output the 3-D boundary boxes and target

收稿日期：2020-09-09；**修订日期**：2021-01-30

category information. By using the one-stage network structure, the method directly regresses the category probability and position coordinates of the targets, solving the imbalance problem of positive and negative samples in the process of one-stage network training by introducing focal loss function. Experiments on KITTI dataset show that the proposed method outperforms the contrast algorithms in terms of average accuracy and time-consuming, and can effectively balance the accuracy and real-time performance of target detection.

Key words: 3-D human target detection; multi-sensor information fusion; deep learning; improved RetinaNet; focal loss function

引　言

目标检测作为计算机视觉的基础任务之一,其主要目的是在点云或图像序列中精确得出各种目标的类别和位置信息[1-2]。目前,基于深度学习的二维目标检测工作已经取得了显著进展,但由于二维目标检测对真实场景的描述度不够,缺乏目标尺寸、姿态等物理参数信息,在实际应用中受到一定限制,因此结合深度信息的三维目标检测任务引起了广大学者及研究机构的关注[3-4]。基于深度学习的三维目标检测方法具有智能分析、自主检测及泛化能力强等特点[5],已逐渐应用于智能安防、自动驾驶和医疗等诸多领域[6]。

目前,在基于雷达点云与图像数据融合的三维目标检测的研究中,清华大学和百度公司的Chen等[7]提出了MV3D模型,加拿大滑铁卢大学Ku等[8]提出了AVOD模型,瑞典林雪平大学的Gustafsson等[9]提出了A3DODWTDA模型,美国斯坦福大学和Nuro公司的Qi等[10]提出了F-PointNet模型,哈尔滨工业大学的Cao等[11]对F-PointNet模型进行了改进,提出了MVFP模型。其中,MV3D模型融合了视觉和雷达点云信息,其中的三维候选网络将点云表达成具有三维信息的正视图和鸟瞰图,结合图像初步卷积处理后,把特征和候选区域进行融合,然后输出最终的目标边界框。AVOD模型对MV3D模型进行了改进,以雷达点云鸟瞰图和图像数据作为输入,采用特征金字塔网络(Feature Pyramid network,FPN)架构作为主干网络,有利于小目标的检测,提高了整体检测精度。A3DODWTDA网络首先对图像进行二维目标检测,然后将检测到的二维目标投影到三维空间中得到目标可能存在的区域,进一步对目标点云进行分割,最后通过回归得到目标的三维检测结果。F-PointNet模型舍弃了图像与雷达点云数据的融合操作,通过从图像到点云的待检测物体定位过程,实现了逐维(2D到3D)的精准定位,并且该网络直接处理点云数据,避免了点云映射过程中某一维度信息的损失,能够学习更全面的空间几何信息,在小目标的检测上有较好的效果;万鹏[12]在该网络模型的基础上进行了优化,使用不同的参数初始化、L_2正则化和修改卷积核数的方法对模型进行测试,进一步提升了模型的目标检测精度。为了进一步降低F-PointNet模型的漏检率,MVFP网络添加了辅助的鸟瞰图检测部分,实现多视角的F-PointNet模型。通过F-PointNet模型获得初始目标检测结果,同时将雷达点云编码到鸟瞰图的特征图中,并据此预测二维边界框,在漏检判断中采用交并比(Intersection over union, IoU)作为F-PointNet初始目标检测结果与鸟瞰图映射预测结果的匹配准则,将鸟瞰图映射中属于漏检目标的二维边界框投影到F-PointNet通道中,直到鸟瞰图映射中的所有目标在F-PointNet检测结果集中找到匹配的检测结果,有效提高了网络在复杂环境中的目标检测精度。然而,上述网络模型均需要预先选取目标存在概

率较高的候选框，然后再进行目标分类和边界框回归，属于两阶段神经网络模型。由于该类型网络模型的结构和运算过程较为复杂，使得运算速度受到限制，难以满足实时性要求较高的应用场景。

针对以上问题，本文研究并设计了一种基于改进型RetinaNet单阶段卷积神经网络的三维人体目标实时检测方法。该网络结构是在RetinaNet这一单阶段二维目标检测网络结构基础上进行改进：将主干网络与特征金字塔网络两条路径相结合，用于点云和图像的特征提取；设置一系列三维锚框并将其投影到特征图上，将投影的二维锚框裁剪为同样大小并进行融合；设计了适合三维目标检测的功能网络以输出三维边界框和类别信息。上述改进可将RetinaNet扩展为三维多传感器融合检测网络，从而提高三维人体目标检测网络的检测性能，同时保持在运算速度方面的优势。

1 改进型RetinaNet网络架构设计

RetinaNet模型主要由FPN结构[13]和聚焦损失函数构成，以VGG、ResNet等网络作为主干网络有效提取特征，利用FPN网络对提取的多尺度特征进行强化，并将所获得的表达能力更强、包含多尺度目标区域信息的特征图输入到两个子网络中完成目标分类和边界框回归任务。RetinaNet网络的损失函数由L_1损失函数和聚焦损失函数组成，前者用于计算边界框回归误差，后者用于计算分类误差[14]。聚焦损失函数的引入可有效解决正样本和负样本之间的不平衡问题，提高目标检测的精度[15-16]。为了将RetinaNet扩展到三维空间，本文主要做了以下改进：

(1)主干网络包括两条路径的子网络，分别用于点云和图像的特征提取，每条路径采用Resnet网络和FPN结构。

(2)每个锚框由长、宽、高和中心点坐标的六维数组表示，同时采用聚类方法设置锚框的大小。

(3)通过投影方法，利用三维锚框对点云和图像的特征图进行感兴趣区域(Region of interest，ROI)池化和融合。

(4)损失函数中目标分类误差采用聚焦损失函数计算，边界框偏移量回归误差采用L_1损失计算，此外，增加了边界框方位回归误差。

本文所设计的三维目标检测网络结构如图1所示，主要分为主干网络和功能网络。利用主干网络提取点云和图像的特征图，然后将两者的特征图经过ROI池化和融合输入到功能网络，输出目标的分类和边界框回归信息。

2 基于改进型RetinaNet的三维目标检测流程

2.1 输入数据预处理

本文基于KITTI数据集[17]进行网络训练和测试，该数据集包含大量带有标注信息的图像和点云数据，需要对其进行预处理再输入到网络中。本文采用KITTI数据集中其中一侧彩色相机捕捉的图像，同时采用数据集中Velodyne 64线雷达采集的点云数据，以提供准确的目标定位信息。数据集以相机数据采集方向为X轴，向上方向为Y轴，根据右手坐标系原则确定Z轴方向。为了利用深度神经网络来处理点云，本文将点云投影为鸟瞰图(Bird's-eye view，BEV)以避免遮挡问题，并且可以用二维目标检测网络作为主干网络进行处理。为了将点云转换为鸟瞰图，本文在X轴和Z轴方向上将点云划分为0.1 m×0.1 m的网格，将Y轴0～3 m之间的点云等分为6层，对每一个单元格，保存高度特征为该单元格的最高点的高度信息，另外设置2个通道记录点云的强度和密度信息，因此点云鸟瞰图被编码为8个

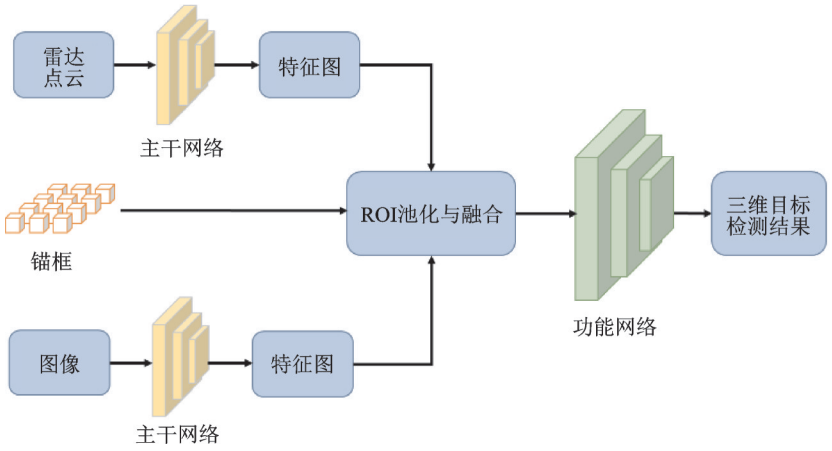


图1 改进型RetinaNet网络架构

Fig.1 Improved RetinaNet network architecture

通道的特征。

2.2 基于主干网络的特征提取

本文在主干网络设计中采用Resnet网络结构[18],利用主干网络进行特征提取。通过将卷积层和池化层逐层叠加,不断增加每层的通道数,以减小特征图的大小。大多数网络只输出最后一层的特征图以用于后续的分类和检测任务,这往往会导致小目标占据较少的像素,不利于对其进行检测。采用FPN结构可以有效地解决这一问题。如图2所示,FPN结构主要包括3个基本过程:自下向上的通路,即自下向上的不同维度特征生成;自上向下的通路,即自上向下的特征补充增强;主干网络层特征与最终输出的各维度特征之间关联表达。改进型RetinaNet在特征金字塔的各个层上进行ROI池化操作,然后将不同尺寸的输出结果输入到功能网络中,从而提高检测效果。

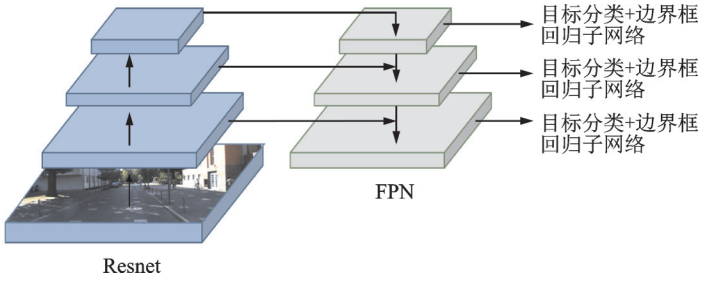


图2 FPN基本架构

Fig.2 FPN basic architecture

2.3 ROI池化和多传感器信息融合

为了提高边界框位置回归的计算效率,本文设置了一定数量的锚框(预先设定的、按一定规则密集排列的边界框),从而通过锚框偏移量确定目标边界框。对于主干网络输出的特征图,首先将三维锚框投影到其中,然后根据投影结果对特征图进行裁剪,得到大量大小相同的特征图二维锚框,从而进一步完成对点云鸟瞰图和图像特征图的ROI池化操作,最后对点云鸟瞰图和图像特征图计算均值进行特征融合。

由于锚框通常不能很好地包围目标，所以需要通过神经网络进行回归，以帮助网络输出更为准确的边界框。在2.4节中将进一步论述将这些融合特征图输入到功能网络中进行最终分类和回归的过程。

2.3.1 三维锚框设置

本文以KITTI数据集中的行人为检测目标，考虑到检测目标的尺寸差异不大，为了降低不必要的计算复杂度，本文设置了3种尺寸的锚框，锚框由其中心坐标(x,y,z)和长度，宽度、高度(l,w,h)六个参数表示。其中，x和y值由点云鸟瞰图中以0.5 m的间隔通过均匀采样获得，z值由传感器距离地面的高度和物体高度来计算。锚框的大小通过对数据集中检测目标的标签信息进行聚类来确定。由于雷达点云稀疏会导致许多空锚，对于不包含点云的空锚，根据锚框中点云的总和是否为零来决定是否将其剔除。

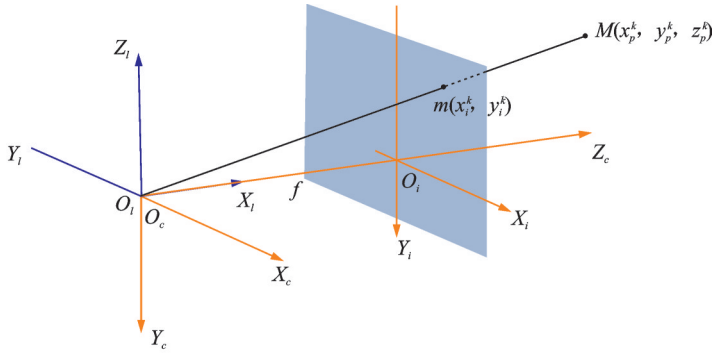
2.3.2 ROI池化和融合

为了实现雷达点云和图像的特征融合，需要对特征图进行ROI池化。因此，本文将三维锚框投影到点云鸟瞰图和图像的特征图上，然后对其进行裁剪和尺寸调整。对于三维锚框(x_p,y_p,z_p,l,w,h)，点云鸟瞰图上投影区域的左上角和右下角可以表示为$(x_{l,\text{left}},z_{l,\text{left}})$和$(x_{l,\text{right}},z_{l,\text{right}})$，即

$$\begin{cases} x_{l,\text{left}}=x_p-\dfrac{l}{2} \\ x_{l,\text{right}}=x_p+\dfrac{l}{2} \\ z_{l,\text{left}}=z_p-\dfrac{h}{2} \\ z_{l,\text{right}}=z_p+\dfrac{h}{2} \end{cases} \tag{1}$$

将三维锚框投影到图像特征图上的计算过程比较复杂。由于KITTI数据集有许多坐标系，如图3所示，其中点云和三维锚框在雷达坐标系O_l-$X_lY_lZ_l$中表示，而图像在坐标系O_i-X_iY_i中表示，因此有必要将三维锚框参数转换到图像坐标系下。

首先，根据锚框中心坐标、长度、宽度和高度计算出其8个顶点的坐标(x_p^k,y_p^k,z_p^k)，$k=1,\cdots,8$，然后



雷达坐标系 O_l-$X_lY_lZ_l$

图像坐标系 O_i-X_iY_i

摄像机坐标系 O_c-$X_cY_cZ_c$

图3 多传感器坐标系

Fig.3 Multi-sensor coordinate system

将顶点坐标转换到图像坐标系。其中,从雷达坐标系转换到摄像机坐标系,需要乘以相应的转换矩阵 $T_{\text{lid}}^{\text{cam}}$。如果锚框中的顶点设置为 $M(x_p^k, y_p^k, z_p^k)$,转换为摄像机坐标系表示为

$$(x_c^k, y_c^k, z_c^k) = T_{\text{lid}}^{\text{cam}}(x_p^k, y_p^k, z_p^k) \tag{2}$$

式中转换矩阵 $T_{\text{lid}}^{\text{cam}}$ 由数据集提供。根据成像投影关系,将摄像机坐标系中的点转换为图像坐标系中的点,数学关系式为

$$\begin{cases} \dfrac{x_i^k}{f} = \dfrac{x_c^k}{z_c^k} \\ \dfrac{y_i^k}{f} = \dfrac{y_c^k}{z_c^k} \end{cases} \tag{3}$$

其矩阵形式为

$$z_c^k \begin{bmatrix} x_i^k \\ y_i^k \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c^k \\ y_c^k \\ z_c^k \\ 1 \end{bmatrix} \tag{4}$$

式中 f 表示摄像机的焦距。利用数学关系式(2,4)将三维锚框从雷达坐标系转换到图像坐标系,得到 (x_i^k, y_i^k),$k=1,\cdots,8$。考虑到遮挡效应,三维锚框的投影区域的顶点可以表示为

$$\begin{cases} x_{i,\text{left}} = \min(x_i^k) \\ x_{i,\text{right}} = \max(x_i^k) \\ y_{i,\text{left}} = \min(y_i^k) \\ y_{i,\text{right}} = \max(y_i^k) \end{cases} \tag{5}$$

根据三维锚框投影区域的参数,对特征图进行裁剪,并将其大小调整为 4×4,使特征图尺寸相同。采用元素平均法实现多传感器特征图的特征融合[19]。

在训练阶段,通过计算锚框与真实边界框之间的交并比,对特征融合锚框进行标记,当 IoU 大于阈值时记录为正样本,反之为负样本。

2.4 基于功能网络的目标检测

将特征融合锚框输入到功能网络中进行目标分类和边界框回归。最终的预测边界框通过基于具有相应偏移量的三维锚框得到,与传统的边界框直接回归方法相比,该方法不仅降低了回归的难度,而且边界框定位更准确。功能网络由 3 个并行的全连接层组成,分别完成分类、边界框偏移量回归和方位回归 3 个任务。

2.4.1 功能网络输出结果

对于 N 个特征融合锚框,分类信息包括目标和背景,分类网络的输出维数为 $2N$;对于边界框偏移量回归网络,为了说明锚框和真实边界框之间的中心坐标、长度、宽度和高度的偏移量,回归结果表示为 $(\Delta x_p, \Delta y_p, \Delta z_p, \Delta l, \Delta w, \Delta h)$,输出维数为 $6N$。对于边界框方位回归网络,本文采用计算边界框在点云鸟瞰图中投影角向量 $(x_{\text{or}}, y_{\text{or}}) = (\cos\theta, \sin\theta)$ 的方法,输出维数为 $2N$。

2.4.2 损失函数设计

为衡量网络模型的三维目标检测性能,本文设计了一个多任务损失函数。用 2 个平滑 L_1 函数计算边界框偏移量和方位回归的误差,用聚焦损失函数计算目标分类误差。其中,聚焦损失函数通过减小

背景样本的权重，可有效解决类别分类不平衡的问题。

$$L(\{p_i\},\{t_i\})=\frac{1}{N_{\text{cls}}}\Sigma_i L_{\text{cls}}(p_i,p_i^*)+\lambda\frac{1}{N_{\text{req}}}\Sigma_i p_i^* L_{\text{reg}}(t_i,t_i^*)+\lambda\frac{1}{N_{\text{ang}}}\Sigma_i p_i^* L_{\text{ang}}(t_i,t_i^*) \tag{6}$$

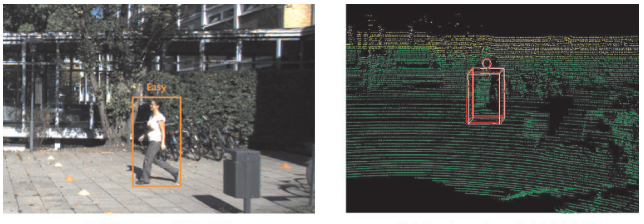
式中：i表示锚框的索引；p_i表示预测目标类别的概率值；p_i^*表示标注目标类别信息；正样本和负样本分别标记为1和0；t_i表示边界框的预测结果；t_i^*表示正样本的边界框的标注信息。等号右侧第1项L_{cls}表示预测类别与真实类别之间的偏差值；第2项L_{reg}表示预测边界框位置与真实值之间的偏差，其中$p_i^*L_{\text{reg}}$表示该项仅与正样本有关，对于那些具有小于0.5的IoU记录为负样本，标记为0；第3项L_{ang}表示边界框方位角预测值与真实值之间的偏差，其中$p_i^*L_{\text{ang}}$表示该项仅与正样本有关，对于负样本标记为0。N_{cls}、N_{reg}和N_{ang}分别对上述3项进行规范化，超参数λ用于平衡3项之间的权重。

3 人体目标检测实验及结果分析

本文在KITTI数据集上对所研究的改进型RetinaNet网络进行训练和测试，将已标注标签的7 481个样本按照3∶1的比例划分为训练集和测试集。根据边界框高度、遮挡程度和截断程度，将样本分为简单、中等、困难3个等级。深度学习计算机的CPU配置为Intel Xeon E5-2678 V3，GPU配置为2套NVIDIA GeForce RTX 2080 Ti。

图4、5显示了本文方法在测试集中的三维行人目标检测可视化结果示例。图4(a)、5(a)分别对应于2幅图像，其索引号分别为000199和000202，图像中的行人目标会根据边界框高度、遮挡程度和截断程度划分为简单、中等、困难3个等级并标注出来。图4(b)、5(b)为在上述2个样本的雷达点云数据图中表示的三维目标检测结果。其中，红色三维边界框为标注的真实边界框，白色三维边界框为本文方法的检测结果。从图中可以看出，所研究的网络对于标记为“简单”和“中等”的目标其检测效果较好，而对于标记为“困难”的目标，可能会出现检测不到的情况。

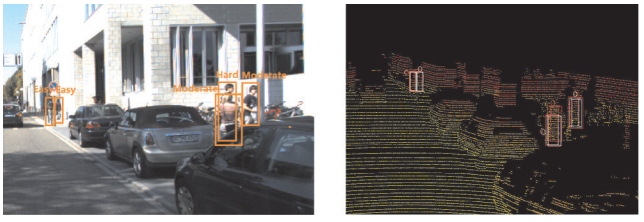
为验证本文所研究的基于雷达点云和图像数据融合的三维人体目标检测性能，开展了分别以图像(主干网络只保留图像处理分支)和雷达点云(主干网络只保留雷达点云处理分支)作为数据来源对网络进行训练和测试的对比实验。三维目标检测结果用平均精度(Average precision，AP)进行评估，其中IoU阈值设置为0.5，对比结果如表1所示。从表1可以看出，雷达点云和图像数据融合作为数据来源的目标检测AP值相比于单独以图像作为数据来源提高了8%左右，相比于单独以雷达点云作为数据来源提高了5%左右，表明将雷达点云与图像数据融合处理可以捕捉到更丰富的行人特征，从而提高三维人体目标检测的精度。



(a) Image 000199　　(b) Point cloud 000199

图4　图像000199三维目标检测可视化结果

Fig.4　Visualization results of 3-D object detection of image 000199



(a) Image 000202 (b) Point cloud 000202

图5 图像000202三维目标检测可视化结果

Fig.5 Visualization results of 3-D object detection of image 000202

表1 基于不同数据来源的改进RetinaNet模型人体目标检测平均精度对比

Table 1 Comparison of human target detection average precision of improved RetinaNet model based on different data sources %

数据来源	简单	中等	困难
图像	46.39	39.86	34.12
雷达	49.04	42.49	37.58
雷达+图像	55.16	47.82	41.91

采用图像与雷达点云数据融合作为数据来源,基于不同深度学习模型的三维人体目标检测结果对比如表2所示。从表2可以看出,本文方法在检测准确性和实时性方面相对于MV3D模型均有大幅提高;相较AVOD模型,本文方法在实时性方面略有提高,运行时间缩短了0.02 s,在准确性方面也有一定程度的提高,对于标记为“简单”和“中等”目标的AP值提高了5%左右;同时,本文方法在实时性方面明显优于F-PointNet和Virtual Multi-View Synthesis方法,在检测精度方面也均有一定程度的提高。综上所述,对于数据集的简单、中等、困难3个等级的行人目标检测,本文方法在检测准确性和实时性的综合性能上优于对比算法,实现了在提高检测精度的同时有效降低网络运行所消耗的时间。但是本文方法对于标记为“困难”的目标,检测精度的提升尚不明显。因此,在严重遮挡情况下提高多目标检测精度的问题,将是本文方法后续研究的重点。

表2 基于不同模型的人体目标检测结果对比

Table 2 Comparison of human target detection results based on different models

方法	结构	数据来源	运行时间/s	平均精度/%		
				简单	中等	困难
MV3D[7]	两阶段	雷达+图像	0.36	40.09	31.14	27.52
AVOD[8]	两阶段	雷达+图像	0.10	50.80	42.81	40.88
F-PointNet[10]	两阶段	雷达+图像	0.17	51.21	44.89	40.23
Virtual Multi-View Synthesis[20]	两阶段	雷达+图像	0.24	53.98	45.01	41.72
改进型RetinaNet	单阶段	雷达+图像	0.08	55.16	47.82	41.91

4 结束语

本文研究了一种基于雷达与图像数据融合的三维人体目标实时检测方法。通过改进现有的RetinaNet,设计了用于三维目标检测的统一体系结构。由于改进型RetinaNet为单阶段卷积神经网络,不涉

及区域候选网络,并且通过引入聚焦损失函数减小负样本的权重以解决正负样本不平衡的问题,有效提高了目标检测的实时性和准确性。在KITTI数据集上进行的实验表明,本文方法在行人目标检测的平均精度和时间消耗方面均优于对比算法,适用于依托巡检机器人、高清摄像头与雷达的自主巡检系统。

参考文献:

[1] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: WA, 2016: 3213-3223.

[2] 黄健, 张钢. 深度卷积神经网络的目标检测算法综述[J]. 计算机工程与应用,2020,56(17): 12-23.
HUANG Jian, ZHANG Gang. Survey of object detection algorithms for deep convolutional neural networks[J]. Computer Engineering and Applications, 2020, 56(17): 12-23.

[3] 张易, 项志宇, 乔程昱, 等. 基于三维点云鸟瞰图的高精度实时目标检测[J]. 机器人, 2020, 42(2): 148-156.
ZHANG Yi, XIANG Zhiyu, QIAO Chengyu, et al. High-precision real-time object detection based on bird's eye view from 3D point clouds[J]. Robot, 2020, 42(2): 148-156.

[4] 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.

[5] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]//Proceedings of the 30th Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2016: 379-387.

[6] 陈娟. 城市智能汽车周围环境的时空行为预测算法研究[D]. 成都:电子科技大学,2020.
CHEN Juan. Research on spatial and temporal behavior prediction algorithm of the surrounding environment of urban intelligent vehicle[D]. Chengdu: University of Electronic Science and Technology, 2020.

[7] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2017: 6526-6534.

[8] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain: IEEE, 2017: 5750-5757.

[9] GUSTAFSSON F, LINDER-NORÉN E. Automotive 3D object detection without target domain annotations[D]. Linköping, Sweden: Linköping University, 2018.

[10] QI C, LIU W, WU C, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 918-927.

[11] CAO P, CHEN H, ZHANG Y, et al. Multi-view frustum Pointnet for object detection in autonomous driving[C]//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). [S.l.]: IEEE, 2019: 3896-3899.

[12] 万鹏. 基于F-PointNet的3D点云数据目标检测[J]. 山东大学学报(工学版),2019,49(5): 98-104.
WAN Peng. Object detection of 3D point clouds based on F-PointNet[J].Journal of Shandong University(Engineering Science), 2019,49(5): 98-104.

[13] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, HI, USA: IEEE, 2017: 936-944.

[14] LI M, HU Y, ZHAO N, et al. One-stage multi-sensor data fusion convolutional neural network for 3D object detection[J]. Sensors, 2019, 19(6): 1434.

[15] 周立旺, 潘天翔, 杨泽曦, 等. 多阶段优化的小目标聚焦检测[J]. 图学学报,2020,41(1): 93-99.
ZHOU Liwang, PAN Tianxiang, YANG Zexi, et al. FocusNet: Coarse-to-fine small object detection network[J]. Journal of Graphics, 2020,41(1): 93-99.

[16] LIN T, GOYAL P, GIRSHICK R, et al. Loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2999-3007.

[17] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2012: 3354-3361.

[18] 蔡强, 李晶, 郝佳云. 基于聚焦损失与残差网络的远程监督关系抽取[J]. 计算机工程, 2019, 45(12): 166-170.
CAI Qiang, LI Jing, HAO Jiayun. Distant supervision relation extraction based on focal loss and residual networ[J]. Computer Engineering, 2019, 45(12): 166-170.

[19] 王立鹏, 张智, 苏丽, 等. 基于多特征融合的自适应权重目标分类方法研究[J]. 华中科技大学学报(自然科学版), 2020, 48(9): 38-43.
WANG Lipeng, ZHANG Zhi, SU Li, et al. Target classification with adaptive weights based on multi-feature fusion[J]. Huazhong University of Science & Technology (Natural Science Edition), 2020, 48(9): 38-43.

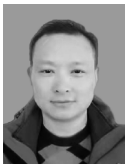
[20] KU J, PON A, WALSH S, et al. Improving 3D object detection for pedestrians with virtual multi-view synthesis orientation estimation[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China: IEEE, 2019: 3459-3466.

作者简介:

李文平(1986-),**通信作者**,男,工程师,研究方向:网络通信及人工智能在智慧城市中的产业应用,E-mail:qwx329@sina.com。

袁强(1992-),男,硕士研究生,研究方向:模式识别,E-mail:1416054179@qq.com。

陈璐(1990-),男,工程师,研究方向:基于视频图像的输变电物联网应用研究,E-mail:cl433x@163.com。



郑利彪(1980-),男,工程师,研究方向:人工智能在视频类在变电站产业应用,E-mail:939251669@qq.com。

汤晓龙(1989-),男,工程师,研究方向:基于机器学习的图像识别,E-mail:tangxiaolong0810@126.com。

(编辑:张黄群)