

# 基于多模态融合的室内人体跟踪技术研究

于翔,周波

(东南大学自动化学院,江苏南京 210096)

**摘要:**传统的室内人体跟踪一般基于单独的相机或者二维激光雷达进行,基于二维点云的人腿检测在面对桌/椅腿较多的环境时准确度明显不足,而基于图像的人体检测在室内光线较暗时也会出现鲁棒性不足的问题,这些单模态的方法往往难以在干扰嘈杂的环境工作。因此,该文提出了一种基于多模态融合的室内人体跟踪方法,训练一个 Adaboost 分类器将点云段分为人腿/非人腿,使用成熟的 YOLOv3 网络对图像中的人体进行检测,结合点云段分类结果和视觉人体边界框完成基于概率数据关联的模态融合,最后使用扩展卡尔曼滤波器完成人体跟踪。对该方法在嘈杂的实验室环境下和晦暗的走廊环境下进行测试,结果表明多模态融合的室内人体检测取得了优于单模态的效果,验证了该系统的鲁棒性和有效性,适用于移动机器人的室内人体跟踪任务。

**关键词:**移动机器人;多模态融合;室内人体跟踪;人体检测;数据关联

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2023)02-0038-06

doi:10.3969/j.issn.1673-629X.2023.02.006

## Research on Indoor Human Tracking Technology Based on Multi-modal Fusion

YU Xiang, ZHOU Bo

(School of Automation, Southeast University, Nanjing 210096, China)

**Abstract:** Traditional indoor human tracking is usually based on a separate camera or two-dimensional LiDAR. The accuracy of two-dimensional point clouds-based human leg detection is significantly inadequate when facing the environments having more table/chair legs. Image-based human detection also has the problem of insufficient robustness when the light is dim. These single-modal methods are often difficult to work in noisy environments. Therefore, we present an indoor human tracking method based on multi-modal fusion. An Adaboost classifier is trained to divide the segments of point clouds into human legs and non-human legs. The mature YOLOv3 network is used to detect human body in the image. The classification results of segments and the visual human bounding boxes are combined to complete the modal fusion based on probability data association. Finally, the extended Kalman filter is used to complete human tracking. The experiments in noisy laboratory environment and dark corridor environment show that the proposed algorithm achieves better results than three single-modal methods. The robustness and effectiveness of the proposed system are verified and it is suitable for indoor human tracking tasks of mobile robots.

**Key words:** mobile robots; multi-modal fusion; indoor human tracking; human detection; data association

### 0 引言

过去,机器人主要应用于将人类从工业现场中的简单重复劳动中解放出来,随着经济的发展和人工智能领域的快速推进,人们对于机器人的需要不再局限于工业中的应用,开始希望机器人能够在人们的日常生活提供帮助,像清洁机器人、服务机器人、物流机器人等移动机器人逐渐火爆起来。

在日常生活中,移动机器人避免不了要跟人“打交道”,这就要求移动机器人必须有分辨主人位置的

能力,具有检测环境中人类的能力,移动机器人才能实现更好的人机交互。移动机器人在室内应用时,有时需要对主人进行跟踪,而受制于传感器的位置限制,机器人往往只能“看到”人的后背下半身,因此在室内人体跟踪的背景下,需要对人体的腿部、背部进行检测。而由于室内人体跟踪相比室外检测和跟踪难度较低,且用于室内跟踪的机器人携带的计算资源一般不会太好,因此,室内移动机器人也一般不会搭载较为贵重的三维激光雷达传感器,主要以二维激光雷达传感器和

收稿日期:2022-03-10

修回日期:2022-07-13

基金项目:国家自然科学基金资助项目(62073075)

作者简介:于翔(1997-),男,硕士研究生,通信作者,研究方向为移动机器人;周波,博士,副教授,研究方向为工业机器人、移动机器人等。

RGB-D 相机为主。

现有的室内人体跟踪技术大多基于单独的相机或者二维激光雷达传感器。

传统的视觉人体跟踪<sup>[1]</sup>基于 HOG (Histogram of Oriented Gradient, 方向梯度直方图) 特征和 SVM (Support Vector Machines, 支持向量机) 或者其他一些机器学习的人工特征和分类器进行人体检测, 这种方法无法取得很高的精度。基于深度学习的视觉人体检测可以采用行人检测领域的相关技术, 行人检测领域的研究不止能够检测到正面的人体, 其他朝向的人体 (如侧面、背面) 也能够取得较好的检测效果, 使用行人检测领域的相关技术进行室内人体检测往往绰绰有余, 但一些室内场景经常存在光照严重不足的状况, 这可能会导致性能大幅缩水。行人检测算法可划分为二阶段检测框架和单阶段检测框架, 两阶段检测方法生成大量的窗口, 对窗口中的图像进行检测和分类; 一阶段检测方法将输入图像进行分格化, 每个格子固定预设一个锚点, 将图像中的物体分配到不同的格子之后再进行检测和分类。Fast RCNN<sup>[2]</sup>、Faster RCNN<sup>[3]</sup>、Mask RCNN<sup>[4]</sup> 是两阶段检测方法的代表, YOLO<sup>[5]</sup>、SSD<sup>[6]</sup>、RetinaNet<sup>[7]</sup> 算法则是单阶段检测方法的代表。

传统的二维激光雷达的人体跟踪同样使用机器学习的方式检测环境中的人腿, Arras 等人<sup>[8]</sup> 提出使用 Adaboost 分类器进行单一人腿的检测, 并使用腿部信息判断人的运动状态。Aguirre 等人<sup>[9]</sup> 使用支持向量机学习环境中不同模式 (动态/静态) 的人腿。近年来, Guerrero-Higueras 等人<sup>[10]</sup> 提出将二维激光雷达采集的环境点云转化为占用图, 并以生物医学中经常应用的 U-Net<sup>[11]</sup> 检测占用图中成对出现的人腿, 这种方法通过将二维点云转化为图像, 由此可以使用深度学习的卷积神经网络进行人腿检测, 这类似于三维点云的俯视图操作, 将三维点云转化为俯视图再使用图像卷积神经网络对俯视图进行特征提取, 本质上是一种模态之间的相互转化。点云人体跟踪方法缺乏对于环境中类似人腿物体的鲁棒性, 极易把室内细小的桌子腿、椅子腿检测为人腿。

由于单模态的方法在室内人体跟踪应用时各自存

在一些问题, 多模态融合来提高室内跟踪的鲁棒性就成为了一种自然而然的想法, 利用两种模态的检测结果进行决策方面的融合有望取得更好的效果。

因此, 为增加室内人体跟踪的鲁棒性, 该文提出了一种二维激光雷达点云和相机图像融合的室内人体跟踪方法。使用目标检测领域表现出色的 YOLO v3 网络<sup>[12]</sup> 检测相机中出现的行人, 利用 Adaboost 分类器对分段二维点云进行训练和分类来检测二维点云中的人腿, 结合视觉和点云的检测位置最终判定人的位置, 并基于 EKF<sup>[13]</sup> 进行人体跟踪。

## 1 整体流程

提出的基于多模态融合的室内人体跟踪框架可以分为以下四个部分: 二维点云聚类、分段点云分类、视觉人体检测和多模态融合的 KCF 跟踪, 整体流程如图 1 所示。

二维点云聚类: 由于完整的二维点云数据难以通过机器学习或深度学习直接分辨其中每个点所属的类别, 因此, 需要将原始二维点云分割为数个易于分类的点云段。根据激光雷达传感器的扫描线角度和扫描点距离传感器的距离进行二维点云的聚类, 将原始点云分割成数个小点云段。

分段点云分类: 二维激光雷达采集到的二维点云中点数较少, 不适合使用类似 PointNet 的三维点云处理网络进行训练。鉴于此, 该文依然沿用了传统的机器学习方法, 使用 Adaboost 分类器对点云段的几何和统计特征进行训练, 将点云段分类为人腿和非人腿。

视觉人体检测: 深度学习流行之后, 行人检测领域取得了极大突破, 使用经典的 YOLO v3 网络进行视觉人体检测, 输出图像坐标系中的人体边界框, 完成视觉图像的行人定位。

多模态融合的 EKF 跟踪: 根据点云的分类结果将人腿点云段投影到二维图像坐标系。基于概率数据关联对点云人腿段和视觉人体检测框进行匹配, 并基于 EKF (Extended Kalman Filter, 扩展卡尔曼滤波器) 进行跟踪, 实现后融合的多模态人体跟踪。

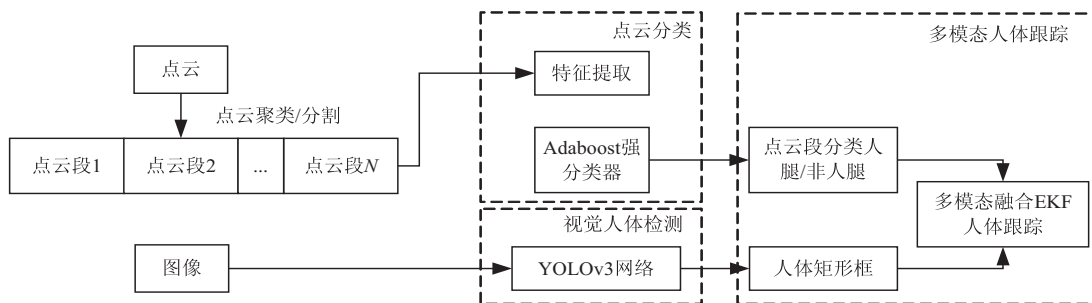


图1 多模态融合的室内人体跟踪整体流程

## 2 模型和方法

### 2.1 基于 Adaboost 的点云段分类

为了能够在二维点云中检测出人腿,需要将激光雷达扫描到的数据点分割成不同的段。激光扫描一周发射 270 条射线,最大得到 270 个点的有序点云,假设原点云  $P$  中的点由极坐标表示,即  $p = (\varphi, r)$ , 其中  $r$  和  $\varphi$  分别表示扫描点距离激光雷达的距离和角度值。根据邻近点的跳变距离,将原始点云分割为不同的段,如果相邻两点之间的距离小于某阈值,则视其与上个点同属于一个点集,如果距离大于某阈值,则后点属于一个新的点集;遍历整个原始点云后,剔除一些数据点数过少的分段使  $i \geq 3$ ,最后形成多个可提取特征点云段  $S_i (i = 1, 2, \dots, N)$ , 二维点云  $P = \{S_1, S_2, \dots, S_N\}$ 。

构建  $S_i$  的 8 种特征<sup>[14]</sup>的弱分类器用于之后的训练过程,10 种特征代表点云段的几何属性和统计特性:

(1) 点的数目。

$$n = \text{len}(S_i) \quad (1)$$

(2) 标准差。

$$\sigma = \sqrt{\frac{1}{n-1} \sum_j \|x_j - \bar{x}\|^2} \quad (2)$$

其中,  $x_j$  代表点云段中的点,  $\bar{x} = (\frac{1}{n} \sum_j x_j, \frac{1}{n} \sum_j y_j)$  代表该点云段的质心。

(3) 平均偏差。

$$\zeta = \frac{1}{n} \sum_j \|x_j - \tilde{x}\| \quad (3)$$

其中,  $\tilde{x}$  为点云段的中值,其定义方式为:

$$\tilde{x} = \begin{cases} x_{(K+1)/2} & \text{if } K \text{ is odd} \\ \frac{1}{2}(x_{K/2} + x_{K/2+1}) & \text{if } K \text{ is even} \end{cases} \quad (4)$$

(4) 点云段宽度。

$$d_{1,n} = \|x_1 - x_n\| \quad (5)$$

其中,1 和  $n$  分别表示点云段的开始点和末尾点,该宽度衡量了点云段开始点和结束点之间的欧氏距离。

(5) 边界长度。

$$l = \sum_j d_{j,j-1} \quad (6)$$

其中,  $d_{j,j-1} = \|x_j - x_{j-1}\|$  表示点云段中相邻两点之间的欧氏距离,边界长度是所有相邻点距离的总和。

(6) 边界标准差。

$$\sigma_l = \sqrt{\frac{1}{n-2} \sum_j \|d_{j,j-1} - \bar{d}\|} \quad (7)$$

其中,  $\bar{d} = \frac{1}{n-1} \sum_j d_{j,j-1} = \frac{l}{n-1}$  为平均边界距离。

(7) 平均曲率:假设点云段  $S_i$  的三个相邻点为

$x_a, x_b$  和  $x_c, d_{ab}, d_{bc}$  和  $d_{ac}$  分别表示相邻两点之间的距离,  $A$  表示三角形  $x_a x_b x_c$  的面积,则  $x_b$  处的曲率可表示为:

$$\hat{k} = \frac{4A}{d_a d_b d_c} \quad (8)$$

$A$  可由海伦公式得出:

$$A = \sqrt{p(p-d_a)(p-d_b)(p-d_c)} \quad (9)$$

其中,  $p = (d_a + d_b + d_c)/2$ 。则平均曲率为:

$$\bar{k} = \sum_j \hat{k}_j \quad (10)$$

(8) 平均角度变化:记相邻两点之间的连线的角度变化值为:

$$\beta_j = \angle(x_{j-1}x_j, x_jx_{j+1}) \quad (11)$$

则平均角度变化:

$$\bar{\beta} = \frac{1}{n-2} \sum_j \beta_j \quad (12)$$

对点云分割得到的点云段训练集进行人工标注,正样本(人腿)标签为+1,负样本(非人腿)标签为-1;将训练集样本输入至 Adaboost 最终得到强分类器用以对点云段进行分类。

### 2.2 基于 YOLO v3 的视觉行人检测

YOLO 是一种单阶段的目标检测架构,在目标检测领域取得了极大的成就。因此,选择较为新颖的 YOLO v3 网络进行行人检测。网络结构上,YOLO v3 使用 DarkNet53<sup>[15]</sup>作为主干网络,整合了残差思想和特征金字塔思想,DarkNet53 的网络结构如图 2 所示。

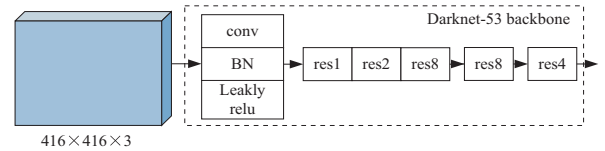


图 2 DarkNet53 网络结构

网络的输入图像被裁剪成  $416 \times 416 \times 3$  的 RGB 图像,在 Darknet53 骨架网络中,输入首先经过一个卷积层,其后跟随着 Batch Norm(批正则化)和 Leaky ReLU 激活函数,随后经过数个残差块,输出三个维度的特征图。三个维度的特征图再经过卷积操作,得到  $8 \times 8 \times 255, 16 \times 16 \times 255$  和  $32 \times 32 \times 255$  三个维度上的输出特征图。不同维度的特征图对应的感受野不同,负责检测不同尺寸的目标,小尺寸的  $8 \times 8 \times 255$  特征图用于检测大尺寸的目标,大尺寸的  $32 \times 32 \times 255$  特征图用于检查较小的目标。

YOLO v3 使用聚类算法,每张输出特征图生成 3 个先验框,先验框的大小如表 1 所示。

有了先验框和输出特征图之后,就可以对先验框进行解码,判断类别和置信度,得到真正的检测框。

最终,对于每个输入图像,整个 YOLO v3 可以解



码出  $8 \times 8 \times 3 + 16 \times 16 \times 3 + 32 \times 32 \times 3 = 4\,032$  个矩形框以及对应的类别和置信度。

表1 特征图对应的不同尺度先验框

特征图维度	先验框
8×8	116×90, 156×198, 373×326
16×16	30×61, 62×45, 59×119
32×32	10×13, 16×30, 33×23

### 2.3 基于概率数据关联的融合人体检测

为了提高室内人体检测的准确性和鲁棒性,该文从二维点云中检测到人腿点云段,从视觉图像中获得人体边界框后,对两种传感器获得的结果进行融合处理。

首先,将人腿点云段投影到图像上,原始二维点云基于激光雷达坐标系,通过激光雷达和相机的位置关系先将二维点云转换至相机坐标系,再从相机坐标系转换到图像坐标系。激光雷达和相机的联合标定过程由 ROS 开源的 Autoware Calibration Toolkit 完成,通过选取相同一点在点云和图像中的位置来自动计算激光雷达坐标系到图像坐标系的变换矩阵。

点云段和人体边界框统一到图像坐标系下,就可以对室内人体的位置进行数据关联。为了评估点云段和人体边界框的关联性,即多大程度上二者为同一人,使用概率数据关联方法来评估两种模态目标之间的关联性。

假设能够包含图像人腿点的最小矩形框为  $R_{Li}$ , 视觉传感器检测到的人体边界框为  $R_{Gj}$ ,  $i$  和  $j$  分别表示人腿点矩形区域和人体矩形区域的序号。则  $R_{Li}$  与  $R_{Gj}$  的交并比可以衡量人腿和人体属于同一人的概率:

$$P_{iou} = \frac{S(R_{Li} \cap R_{Gj})}{S(R_{Li} \cup R_{Gj})} \quad (13)$$

其中,  $S$  表示各区域的面积,人体和人腿矩形框的重叠区域占总区域比重越大,说明二者的关联性越高,属于同一人的概率越大;人体和人腿的重叠区域占总区域比重越小,说明二者关联性越小,倾向于属于不同的人或者某种模态的数据出现假阳性。图像中检测到的人体边界框遍历所有被分类为人体的点云段,其概率

$P_{iou}$  最大且超过一定阈值的即为同一人。

### 2.4 基于 EKF 的人体跟踪

结合多模态信息检测出人体后,需要对被检测的目标进行状态估计。假设被检测人体的状态量:

$$X = (x, y, z, v_x, v_y, v_z) \quad (14)$$

其中,  $x, y, z$  分别表示某时刻目标人体在  $x, y, z$  轴的坐标,  $v_x, v_y, v_z$  表示此时机器人在三个轴的速度。则  $t$  时刻扩展卡尔曼滤波的预测方程为:

$$\hat{x}_t^- = f(\hat{x}_{t-1}^-) \quad (15)$$

$$P_t^- = A_t P_{t-1} A_t^T + W_t Q_{t-1} W_t^T$$

其中,  $f(\bullet)$  代表非线性函数,  $A_t$  和  $W_t$  为过程协方差矩阵,  $Q_{t-1}$  为上一时刻的噪声协方差矩阵。 $t$  时刻 EKF 的修正方程为:

$$K_t = P_t^- H_t^T (H_t P_t^- H_t^T + V_t R_t V_t^T)^{-1} \quad (16)$$

$$\hat{x}_t = \hat{x}_t^- + K_t (z - h(\hat{x}_t^-, 0))$$

$$P_t = (I - K_t H_t) P_t^-$$

其中,  $h$  表示状态量与观测量之间的非线性关系,  $H_t$  和  $V_t$  为  $h$  的雅可比矩阵,  $R_t$  为观测噪声协方差矩阵。

如果得到融合人体目标之后没有 EKF 的预测值与之对应,则初始化一个新的目标 ID,根据此人的状态生产新的预测值,继续进行迭代匹配。如果 EKF 预测的目标在一定时间内得不到与之对应的人体目标,则删除此人的 ID。

## 3 实验测试

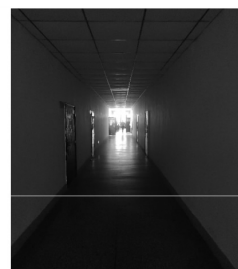
为验证上述多模态人体跟踪系统的有效性,基于 ROS 系统使用 python 语言搭建了以上多模态融合的室内人体跟踪系统,通过 ROS 节点之间的发布和接收消息完成传感器数据的传输。该系统的机器人平台选择 Turtlebot2 移动机器人, Turtlebot2 是 ROS 官方打造的软硬件科研平台,计算平台使用 ThinkPad E580,搭载处理器 Intel Core i7-8550 CPU@1.80 GHz,显卡为 AMD Radeon RX 550 4G,安装 Ubuntu 16.04 和 ROS Kinetic Kame。二维激光雷达选择日本北洋(Hokuyo)公司的 URG-04LX-UG001 传感器,视觉传感器使用 Kinect V2,平台如图 3(a)所示。



(a)实验平台



(b)实验室环境



(c)走廊环境

图3 实验平台及实验场景

为了测试多模态融合室内跟踪方法的效果,分别在嘈杂的实验室环境、晦暗但安静的走廊环境采集共 400 帧点云和图像数据并进行标注,实验环境如图 3(b)、(c)所示。

对图像数据,在图中标注出人所在的边界框,实验

表 2 四种算法在自建数据集中的 mAP(平均检测精度)及 IoU(交并比)比较

检测方法	实验室环境 mAP/%	实验室环境 IoU	走廊环境 mAP/%	走廊环境 IoU
SSD	63.9	53.2	65.5	57.7
YOLOv3	67.6	57.4	69.9	60.5
YOLOv4	70.3	58.9	73.2	62.5
文中融合算法	72.4	60.2	76.6	65.0

从表 2 可以看出,所提多模态融合室内人体检测方法在与其他视觉行人检测方法对比时取得了最优效果,由于室内环境的光照往往会发生忽明忽暗的变化,不像室外环境一样近似平行光,图像人体检测总会出现一段时间检测不到的现象,此时就需要点云模态进行补充。

对点云数据,首先对整体点云进行分割,分割后标注出各个分段属于人腿还是非人腿。实验时分别使用单独的 Adaboost 人腿检测器和多模态融合的人腿检测器对人腿段进行分类,最后统计所有点的平均分类准确度,结果如表 3 所示。

时使用 SSD、YOLO v3、YOLO v4 和文中算法输入数据进行人体检测,评估指标为平均检测精度(mean Average Precision, mAP)和检测框与真值框之间的平均交并比(Intersection over Union, IoU),实验对比结果如表 2 所示。

表 3 两种算法在自建数据集中的平均准确率比较

实验方法	实验室环境的平均准确率/%	走廊环境的平均准确率/%
Adaboost	83.5	87.7
文中融合算法	86.0	88.9

由表 3 可以看出,推理人体位置时引入了多模态决策后,实验室环境和走廊环境的检测平均准确率均有所提高,证明了多模态融合决策的有效性和可行性。室内点云人腿检测由于数据量较小,使用机器学习训练的分类器准确率不会很高,因此,图像模态可以为其提供补充。



图 4 实验室环境的人体检测结果

图 4 展示了实验室环境中的人体检测结果,点群为检测到的人腿段,方框则为视觉检测到的人体。由于环境中存在椅子腿和隔离墙的边缘这种点云表示类似人腿的物体,仅仅使用点云段检测往往会发生误分类,而结合视觉之后就可以更加鲁棒地完成跟踪。

由此可以看出,所提多模态融合室内人体系统针对室内环境都取得了可观的效果,证明了该方法的鲁棒性和有效性。

## 4 结束语

针对移动机器人室内人体跟踪的鲁棒性问题进行了相关研究,实现了一个多模态融合的室内人体跟踪系统,以较少的计算资源完成更加鲁棒的跟踪任务。将该系统应用于多种室内环境中,包括弱光照的走廊、

环境嘈杂的实验室区域,均完成了正确的人体检测,验证了该系统的有效性和通用性。

针对移动机器人室内跟踪的问题,如何以更少的计算资源完成实时的多模态融合的跟踪任务仍然需要进一步研究,可以使用 RGB-D 相机获得像素中每个点的深度后得到相应结构化点云,以减少传感器的数量;也可以探究更加轻量级的人体检测方法以减少系统的计算资源消耗。

## 参考文献:

- [1] HUANG W, ZHOU B, QIAN K, et al. Real-time multi-modal people detection and tracking of mobile robots with a RGB-D sensor[C]//2019 IEEE 4th international conference on advanced robotics and mechatronics (ICARM). Toyona-

- ka;IEEE,2019;325-330.
- [2] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago:IEEE, 2015;1440-1448.
- [3] 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 & Machine Intelligence, 2017, 39(6):1137-1149.
- [4] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Venice:IEEE, 2017;2961-2969.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE, 2016;779-788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European conference on computer vision. [s. l.]:Springer, 2016;21-37.
- [7] 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. Venice:IEEE, 2017;2980-2988.
- [8] ARRAS K O, LAU B, GRZONKA S, et al. Range-based people detection and tracking for socially enabled service robots[M]//Towards service robots for everyday environments. Berlin:Springer, 2012;235-280.
- [9] AGUIRRE E, GARCIA-SILVENTE M, PLATA J. Leg detection and tracking for a mobile robot and based on a laser device, supervised learning and particle filtering[C]//ROBOT2013: first iberian robotics conference. Berlin:Springer, 2014;433-440.
- [10] GUERRERO-HIGUERAS Á M, ÁLVAREZ-APARICIO C, CALVO OLIVERA M C, et al. Tracking people in a mobile robot from 2d lidar scans using full convolutional neural networks for security in cluttered environments[J]. Frontiers in Neurorobotics, 2019, 12(85):1-13.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. [s. l.]:Springer, 2015;234-241.
- [12] 徐思源, 储开斌, 张继, 等. 面向小目标检测的改进YOLOv3算法[J/OL]. 电光与控制;1-8[2022-04-26]. <http://kns.cnki.net/kcms/detail/41.1829.005.html>.
- [13] 温彬彬, 温鹏智, 胡雪花. EKF在机器人目标跟踪中的应用[J]. 电脑知识与技术, 2021, 17(36):116-118.
- [14] 顾文华. 基于扫描激光的室内移动机器人动态环境建模研究[D]. 南京:东南大学, 2014.
- [15] 林海, 庞妙珍, 刘天成. 基于改进DarkNet网络的轻量型车型识别方法[J]. 现代计算机, 2021, 27(24):100-104.
- (上接第37页)
- 通流量预测方法[J]. 计算机工程与设计, 2019, 40(9):2620-2624.
- [14] 曹堉, 王成, 王鑫, 等. 基于时空节点选择和深度学习的城市道路短时交通流预测[J]. 计算机应用, 2020, 40(5):1488-1493.
- [15] 李佩钰. 一种基于小波和神经网络的短时交通流量预测[J]. 计算机技术与发展, 2020, 30(1):135-139.
- [16] 赵刚, 王梦灵. 基于模糊分析的LSTM交通流量预测[J]. 计算机工程与设计, 2021, 42(4):1103-1108.
- [17] 蒋杰, 张江鑫. 改进ACO优化的BP神经网络短时交通流量预测[J]. 计算机仿真, 2021, 38(07):97-101+180.
- [18] 王伟, 过秀成, 陈学武, 等. 交通工程学[M]. 第2版. 南京:东南大学出版社, 2011.
- [19] KIRANYAZ S, AVCI O, ABDELJABER O, et al. 1D convolutional neural networks and applications: a survey[J]. Mechanical Systems and Signal Processing, 2021, 151:107398.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [21] DU Y, CUI N, LI H, et al. The vehicle's velocity prediction methods based on RNN and LSTM neural network[C]//2020 Chinese control and decision conference (CCDC). Hefei:IEEE, 2020;99-102.
- [22] SIAMI-NAMINI S, TAVAKOLI N, NAMIN A S. The performance of LSTM and BiLSTM in forecasting time series[C]//2019 IEEE international conference on big data (big data). Los Angeles:IEEE, 2019;3285-3292.
- [23] TREISMAN A M, GELADE G. A feature-integration theory of attention[J]. Cognitive Psychology, 1980, 12(1):97-136.
- [24] 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展[J]. 计算机工程与科学, 2021, 43(2):370-380.
- [25] 刘建伟, 刘俊文, 罗雄麟. 深度学习中注意力机制研究进展[J]. 工程科学学报, 2021, 43(11):1499-1511.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30:6000-6010.