自动化学报
ACTA AUTOMATICA SINICA

基于改进YOLOX的移动机器人目标跟随方法

万琴 李智 李伊康 葛柱 王耀南 吴迪

Target Following Method of Mobile Robot Based on Improved YOLOX

WAN Qin, LI Zhi, LI Yi-Kang, GE Zhu, WANG Yao-Nan, WU Di

在线阅读 View online: https://doi.org/10.16383/j.aas.c220344

基于改进 YOLOX 的移动机器人目标跟随方法

万 琴[1,2] 李 智[1] 李伊康[1] 葛 柱[1] 王耀南[2,3] 吴 迪[1]

摘 要 针对移动机器人在复杂场景中难以稳定跟随目标的问题, 提出基于改进 YOLOX 的移动机器人目标跟随方法, 主要包括目标检测、目标跟踪以及目标跟随三个部分. 首先, 以 YOLOX 网络为基础, 在其框架下将主干网络采用轻量化网络 MobileNetV2X, 提高复杂场景中目标检测的实时性. 然后, 通过改进的卡尔曼滤波器获取目标跟踪状态并采用数据关联进行目标匹配, 同时通过深度直方图判定目标发生遮挡后, 采用深度概率信息约束及最大后验概率 (Maximum a posteriori, MAP) 进行匹配跟踪, 确保机器人在遮挡情况下稳定跟踪目标. 再采用基于视觉伺服控制的目标跟随算法, 当跟踪目标丢失时, 引入重识别特征主动搜寻目标实现目标跟随. 最后, 在公开数据集上与具有代表性的目标跟随方法进行了定性和定量实验, 同时在真实场景中完成了移动机器人目标跟随实验, 实验结果均验证了所提方法具有较好的鲁棒性和实时性.

关键词 移动机器人, YOLOX, 重识别, 目标跟随

引用格式 万琴, 李智, 李伊康, 葛柱, 王耀南, 吴迪. 基于改进 YOLOX 的移动机器人目标跟随方法. 自动化学报, 2023, **49**(7): 1558−1572

DOI 10.16383/j.aas.c220344

Target Following Method of Mobile Robot Based on Improved YOLOX

WAN Qin[1,2] LI Zhi[1] LI Yi-Kang[1] GE Zhu[1] WANG Yao-Nan[2,3] WU Di[1]

Abstract A target following method of mobile robot based on improved YOLOX is proposed to solve the problem that mobile robots are difficult to follow the target stably in complex scene. This method mainly includes three parts: Target detection, target tracking and target following. Firstly, the lightweight MobileNetV2X network is adopted under the YOLOX framework to improve the real-time performance of target detection in complex scene. Then, the improved Kalman filter is proposed to obtain the tracking state and data association is used for target matching. When the target is judged by depth-histogram, the depth probability constraint and maximum a posteriori (MAP) probability are utilized for matching, which ensure that the robot tracks the target stably under occlusion. Moreover, target-following algorithm based on servo control is proposed, and re-id feature is introduced to actively search for disappeared targets. Finally, qualitative and quantitative experiments on public data set and in real-world environments demonstrate the efficiency of the proposed method.

Key words Mobile robot, YOLOX, re-id, target following

Citation Wan Qin, Li Zhi, Li Yi-Kang, Ge Zhu, Wang Yao-Nan, Wu Di. Target following method of mobile robot based on improved YOLOX. *Acta Automatica Sinica*, 2023, **49**(7): 1558−1572

移动机器人在安防、物流和医疗等领域应用广泛[1−2], 其中机器人目标跟随算法引起了广泛关注, 但移动机器人目标跟随算法的鲁棒性和实时性仍是亟待解决的关键问题[3−4].

机器人目标跟随算法分为生成式模型方法和检测跟踪方法两大类[5−6]. 生成式模型主要通过构建目标模型实现跟随, 如 Yoshimi 等[7] 利用视觉传感器获取行人颜色和纹理特征, 机器人在视野范围内寻找与之相匹配的区域, 融合行人与位置速度信息构建模型, 采用基于生成式的目标跟踪算法跟随行人. 然而, 此类算法关注目标本身, 忽略背景信息, 经常出现跟踪丢失的情况.

为同时考虑目标与背景信息, 检测跟踪方法得

收稿日期 2022-04-27 录用日期 2022-09-26
Manuscript received April 27, 2022; accepted September 26, 2022
国家自然科学基金 (62006075), 湖南省自然科学杰出青年基金 (2021JJ10002), 湖南省重点研发计划 (2021GK2024), 湖南省教育厅重点项目 (21A0460), 湖南省自然科学基金面上项目 (2020JJ4246, 2022JJ30198) 资助
Supported by National Natural Science Foundation of China (62006075), Foundation Project for Distinguished Young Scholars of Hunan Province (2021JJ10002), Key Research and Development Projects of Hunan Province (2021GK2024), Key Projects of Hunan Provincial Department of Education (21A0460), and General Project of Hunan Natural Science Foundation (2020JJ4246, 2022JJ30198)
本文责任编委 程龙
Recommended by Associate Editor CHENG Long
1. 湖南工程学院电气与信息工程学院 湘潭 411104 2. 湖南大学机器人视觉感知与控制技术国家工程研究中心 长沙 410082 3. 湖南大学电气与信息工程学院 长沙 410082
1. College of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan 411104 2. National Engineering Research Center for Robot Visual Perception and Control Technology, Hunan University, Changsha 410082 3. College of Electrical and Information Engineering, Hunan University, Changsha 410082

到了越来越多的关注, 此方法通过构建分类器区分目标及背景, 其跟踪效果普遍优于生成式模型方法. 余铎等[3] 通过快速判别尺度空间切换相关滤波算法与卡尔曼滤波算法实现稳定跟踪. 另外, 移动机器人在跟随控制过程中常受到背景杂斑、光照变化、目标遮挡、尺度变化等干扰, 导致跟随目标丢失. 因此传统的检测跟踪方法不适用于移动机器人在复杂多变场景中的目标跟随[2].

基于深度学习的移动机器人目标跟随算法具有鲁棒性强等优势[8]. Zhang 等[9] 通过基于目标轮廓带采样策略来提高移动机器人跟踪性能, 但未对遮挡、行人消失等情况进行处理. Pang 等[10] 提出一种基于深度学习的目标检测器, 引入卡尔曼滤波来预测目标位置, 加入重识别模块处理遮挡问题, 但此类算法需先获取精度较高的目标检测结果. 鉴于上述问题, JDE (Jointly learns the detector and embedding model) 检测模型可用来融合重识别与检测分支[11], 提高目标检测精度. YOLO (You only look once) 系列算法则是一类基于 JDE 检测模型的一阶段框下的目标检测算法, 具有高效、灵活和泛化性能好的优点.

YOLO 算法包括了 YOLOV1 ~ YOLOV7 系列算法以及一系列基于改进 YOLO 的目标检测算法. Redmon 等[12] 提出 YOLO 算法进行目标检测, 直接采用回归的方法进行坐标框的检测以及分类, 使用一个端到端的简单网络实现坐标回归与分类, 能够极大地提升目标的检测速度. 此后, YOLO 的网络结构不断优化, 已经成为目标检测领域主流的算法. Hsu 等[13] 引入比率感知机制, 动态调整 YOLOV3 的输入层长度和宽度超参数, 从而解决了长宽比差异较大的问题, 能够有效地提高平均跟踪精度. Huang 等[14] 引入改进的 YOLOV3 模型, 此模型将预测尺度从 3 个增加到 4 个, 并使用额外的特征图来提取更多的细节. YOLOV3 的目标位置识别精度较差, 在目标分布密集、尺寸差异较大的复杂场景中, 检测效果较差. YOLOV4[15] 开发了 Darknet53 目标检测模型, 此模型具有更高的网络输入分辨率, 网络层参数多, 计算复杂度高, 对小目标检测效果较差. 对此, YOLO-Z[16] 提出了一系列不同尺度的模型, 提高 YOLOV5 检测小目标的性能. Cheng 等[17] 提出一种单阶段 SSD (Single shot multibox detector) 微小目标检测方法, 此方法可提高微小目标检测的实时性, 但其使用的两阶段式目标检测器使目标定位精度有所下降. YOLOV6[18] 设计了更高效的主干网络和网络层. YOLOV7[19] 扩展了高效长程注意力网络, 加入了基于级联的模型缩放方法, 均可一定程度提高检测精度和推理效率, 但由于未引入重识别分支, 无法提取浅层特征用于后续跟踪. YOLOX[20] 在 YOLO 系列的基础上做出了一系列改进, 相比于 YOLO 系列目标检测算法, 其最大的不同是采用了无锚框检测器. 而 YOLOV1 ~ YOLOV5 采用有锚框的检测器, 由于可能会被多个锚框同时检测且与检测框中心存在误差, 并不适用于 JDE 检测模型. 因此, 采用无锚框的 YOLOX 目标检测算法更加适合于 JDE 检测模型.

移动机器人检测与跟踪跟随目标的核心问题是其在运动过程中, 复杂场景干扰影响其检测精度以及跟随性能. YOLOX 以 Darknet53 网络结构为主干, 有较高的检测精度, 但模型较大、推理速度较慢, 不适用于移动机器人实时跟随. 在 YOLOV5 的网络模型中, 虽然网络的特征提取能力随着深度的增加而增强, 但下采样次数的增加会导致梯度的消失, 这极大影响了移动机器人的检测精度[21]. 为了提升移动机器人的检测精度, DeepSORT 目标跟踪算法[22] 采用卡尔曼滤波更新目标位置, 并与当前检测目标关联匹配, 但未解决因遮挡跟踪造成的目标丢失问题. Han 等[23] 提出 PSR (Peak side-lobe rate) 目标跟踪算法, 引入深度信息来评估跟踪可信度, 并可主动检测跟踪丢失目标. 但其采用相关滤波法实现目标跟踪, 在复杂场景下的跟踪鲁棒性低. 可见, 改进网络结构的同时引入深度信息, 是提升移动机器人检测跟随性能的一种亟待探索的方法.

综上所述, 基于 YOLO 系列的移动机器人目标跟随算法的鲁棒性强且精度高, 但对于变化环境迁移和泛化能力弱, 且运行速率低. 传统移动机器人目标跟随算法速度快, 但是当目标发生形变、尺度变化和严重遮挡等情况时, 跟踪过程容易出现目标跟踪丢失. 因此, 为实现复杂场景下移动机器人稳定跟随目标, 本文提出改进 YOLOX 的移动机器人目标跟随方法 (Improved YOLOX target-following algorithm, IYTFA). 主要工作如下:

1) 为提高目标检测精度和速度, 提出基于 YOLOX-MobileNetV2X 网络 (YOLOX-M2X) 的目标检测算法, 使用交叉熵损失、回归损失以及重识别损失函数, 共同训练检测与重识别分支.

2) 为提高目标预测与更新速率, 采用改进的卡尔曼滤波器获取目标跟踪状态. 同时加入基于深度直方图的遮挡检测机制, 并通过深度概率约束帧间目标匹配, 提高遮挡跟踪准确率.

3) 在目标跟随过程中, 提出基于视觉伺服控制的主动搜寻策略, 并在目标消失时引入重识别特征进行跟踪跟随, 保证移动机器人稳定跟随目标.

本文内容安排如下: 第 1 节介绍 IYTFA 算法,

包括目标检测部分、目标跟踪部分和目标跟随控制部分; 第 2 节为实验验证, 简要说明移动机器人和深度学习平台, 定性、定量分析目标跟踪算法, 并进行移动机器人目标跟随实验; 第 3 节对本文工作进行总结与展望.

1 IYTFA 算法

IYTFA 移动机器人目标跟随方法的结构框图如图 1 所示, 主要由目标检测、目标跟踪及目标跟随控制三部分组成. 首先, 将 YOLOX 的主干网络 Darknet53 替换为 MobileNetV2X, 通过获取的 RGB 视频序列输入训练完成的 MobileNetV2X 网络得到特征图, 再将重识别损失函数和检测损失函数分别训练重识别分支及检测分支, 从而得到目标检测结果. 然后采用改进的卡尔曼滤波器获取跟踪状态, 通过轨迹关联实现目标匹配, 同时引入遮挡判别机制, 如判断目标被遮挡则加入深度概率约束进行遮挡目标跟踪匹配. 最后采用基于视觉伺服控制的主动搜寻策略完成移动机器人目标跟随.

1.1 改进 YOLOX 的目标检测算法

目标检测是移动机器人目标跟随的关键问题, 目标检测精度很大程度上决定了移动机器人跟随的稳定性. 本文以 YOLOX 体系架构为基础进行改进, 优化网络结构与损失函数, 提高检测实时性. 主干网络使用 MobileNetV2X 网络, 再通过检测分支与重识别分支得到检测结果.

1.1.1 YOLOX-MobileNetV2X 网络

YOLOX 算法[20] 将解耦头、数据增强、无锚框以及标签分类等算法与传统的 YOLO 算法进行融合, 算法泛化能力强, 检测小目标精度高.

YOLOX 算法网络主要分为三个部分, 分别为主干网络、网络层和预测层. 其主干网络采用 Darknet53 特征提取网络, 网络层采用特征金字塔网络, 预测层使用了 3 个解耦头. 输入图片在主干网络部分进行浅层特征提取, 输出 3 个特征层传入网络层进行深层特征提取, 输出分别传入 3 个解耦头进行目标检测. 但是 YOLOX 主干网络通常使用 Darknet53 网络, 存在模型尺寸大、推理速度慢等问题. 因此为实现移动机器人实时目标检测, 本文提出 YOLOX-M2X 网络, 将 YOLOX 主干网络采用轻量级的特征提取网络 MobileNetV2X, 该网络的卷积核心层是深度可分离卷积层, 可将输出的特征图的通道数缩减至一半, 并再与原卷积层提取的特征图合并, 与仅使用一组深度可分离卷积的 MobileNetV2[24] 相比, 该网络可获得更多特征图的语义信息.

在 YOLOX-M2X 网络上, 先采用 COCO2017 训练集训练得到网络参数, 再移植至移动机器人平台进行实时检测. COCO2017 数据集是一个可用于图像检测的大规模数据集, 包含超过 330×10^3 幅图像 (其中 220×10^3 幅是有标注的图像), 涵盖 150 万个目标及 80 个目标类别 (行人、汽车、大象等)、91 种材料类别 (草、墙、天空等), 每幅图像包含 5 句语句描述, 且有 250×10^3 个带关键点标注的行人.

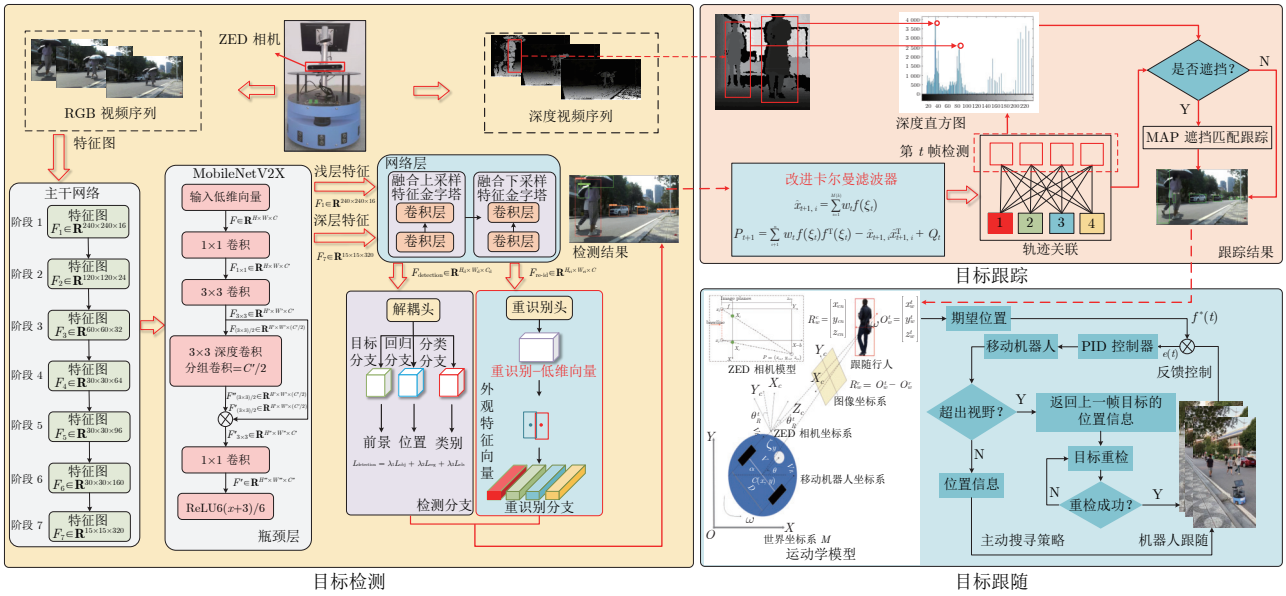


图 1 本文方法结构框图

Fig. 1 Structure block diagram of our method

MobileNetV2X 网络将目标检测时的分类分为 7 个阶段, 输入图片分辨率为 $H \times W$ (H 为图片高度, W 为图片宽度). 假设输入特征图表示为 $F \in \mathbf{R}^{H\times W\times C}$, 其中 H 为高度、W 为宽度、C 为通道数. 每个阶段的核心层为瓶颈层, 每个阶段的瓶颈层中包括 4 个步骤.

步骤 1. 使用 1×1 卷积核扩展特征图为 $F_{1\times1} \in \mathbf{R}^{H\times W\times C'}$, 大幅减少计算量.

步骤 2. 特征图 $F_{1\times1} \in \mathbf{R}^{H\times W\times C'}$ 进行逐点卷积, 再采用 3×3 深度可分离卷积得到特征图 $F_{3\times3} \in \mathbf{R}^{H'\times W'\times C'}$.

步骤 3. 为进一步获得更多的语义信息, 将特征图 $F_{3\times3} \in \mathbf{R}^{H'\times W'\times C'}$ 一分为二, 首先通过普通卷积将特征映射减少到原始通道数的一半, 得到特征图 $F_{(3\times3)/2} \in \mathbf{R}^{H'\times W'\times(C'/2)}$ 和 $F'_{(3\times3)/2} \in \mathbf{R}^{H'\times W'\times(C'/2)}$, 再将特征图 $F'_{(3\times3)/2} \in \mathbf{R}^{H'\times W'\times(C'/2)}$ 进行深度可分离卷积得到 $F''_{(3\times3)/2} \in \mathbf{R}^{H''\times W''\times(C'/2)}$, 然后将 $F''_{(3\times3)/2} \in \mathbf{R}^{H''\times W''\times(C'/2)}$ 与 $F'_{(3\times3)/2} \in \mathbf{R}^{H'\times W'\times(C'/2)}$ 两者结合在一起得到新的特征图 $F'_{3\times3} \in \mathbf{R}^{H'''\times W'''\times C'}$.

步骤 4. 将新特征图 $F'_{3\times3} \in \mathbf{R}^{H'''\times W'''\times C'}$ 使用卷积核为 1×1 的投影卷积层再次卷积, 得到特征图为 $F' \in \mathbf{R}^{H'''\times W'''\times C''}$, 即得到每个瓶颈层的输出特征图.

在 MobileNetV2X 网络中的第 7 个阶段得到深层特征图 $F_7 \in \mathbf{R}^{15\times15\times320}$, 在第 1 个阶段得到浅层特征图 $F_1 \in \mathbf{R}^{240\times240\times15}$, 经网络层后得到检测分支和重识别分支的输入特征图.

1.1.2　目标检测分支及损失函数

MobileNetV2X 网络输出的检测分支特征图 $F_7 \in \mathbf{R}^{15\times15\times320}$, 经网络层后得到特征图 $F_{\text{detction}} \in \mathbf{R}^{H_d\times W_d\times C_d}$, 再经过解耦头后得到检测分支, 包括目标、分类和回归三个分支. 输入目标分支的特征图为 $F_{\text{obj}} \in \mathbf{R}^{H'_o\times W'_o\times1}$, 分支中每个特征点代表对应预测框内被检测目标属于前景的概率, 据此判断目标是前景还是背景. 为稳定训练过程并加快收敛速度、精确定位目标, 该分支估计每个像素相对于目标中心的连续漂移, 以减少下采样的影响; 输入回归分支的特征图为 $F_{\text{reg}} \in \mathbf{R}^{H'_r\times W'_r\times4}$, 该分支对目标框的中心坐标点及高度宽度 $(x,\ y,\ w,\ h)$ 进行预测; 输入分类分支的特征图为 $F_{\text{cls}} \in \mathbf{R}^{H'_c\times W'_c\times1}$, 该分支得到对目标所属类别的预测评分, 如目标属于行人、车辆、动物等不同类别的评分, 其代表目标属于各个类别的概率值. 最后将三个分支的输出结果合并相加得到特征图 $F_{\text{detection}} \in \mathbf{R}^{H'_d\times W'_d\times6}$, 即为目标检测分支的信息.

为度量目标检测信息和真实目标信息之间的差值, 进一步定义损失函数, 损失函数值越小则差值越小, 训练模型准确度越高. 由于 MobileNetV2X 网络中的目标检测分支包括目标分支、回归分支和分类分支, 其对应损失函数由目标损失函数 L_{obj}、回归损失函数 L_{reg} 和分类损失函数 L_{cls} 三部分组成, 总的训练损失函数 $L_{\text{detection}}$ 表示为

$$L_{\text{detection}} = \lambda_1 L_{\text{obj}} + \lambda_2 L_{\text{reg}} + \lambda_3 L_{\text{cls}} \tag{1}$$

其中, λ_1, λ_2 和 λ_3 是损失平衡系数. L_{obj} 和 L_{cls} 采用二值交叉熵损失函数 (Binary cross entropy, BCE), L_{reg} 采用交并比 (Intersection over union, IoU) 损失函数.

在目标检测中, 需首先判定预测的目标属于前景或者背景, 目标损失函数 L_{obj} 采用 Focal 交叉熵损失函数度量其与真实值的差值, 即

$$L_{\text{obj}} = -\frac{1}{N_{\text{obj}}}\sum_{s}[y_s \times \lg(p_s) + (1-y_s)\times \lg(1-p_s)] \tag{2}$$

其中, N_{obj} 代表用于计算 L_{obj} 损失函数的视频帧目标总个数; y_s 表示测试样本 s 的标签, 前景标为 1, 背景标为 0; p_s 表示测试样本 s 预测为前景的概率.

L_{reg} 回归损失函数使用 IoU 损失函数来度量预测检测框与真实目标框的交并比 (面积重叠度). IoU 指标范围为 [0, 1], 当面积重叠率越大时, IoU 指标数值越大, 即

$$L_{\text{reg}} = 1 - IoU \tag{3}$$

其中, IoU 表示当前帧目标预测框和目标真实框的面积重叠率, 即交并比.

为评判当前视频帧目标所属的类别与真实值的差值, 分类损失函数采用多分类交叉熵损失函数对目标所属类别的预测进行评分, 即

$$L_{\text{cls}} = -\sum_{d=1}^{N_{\text{cls}}}\sum_{c=1}^{M}(y_{dc}\lg(p_{dc})) \tag{4}$$

其中, N_{cls} 代表用于计算 L_{cls} 损失函数的视频帧目标总个数; M 表示类别的数量; y_{dc} 为符号函数, 如果当前视频帧目标 d 的真实类别等于 c, y_{dc} 为 1, 否则取 0; p_{dc} 为当前帧目标 d 属于类别 c 的预测概率.

1.1.3　重识别分支及损失函数

为在目标消失再出现时完成视频连续帧间的目标匹配识别 (即目标重识别), 在 YOLOX-M2X 网络中加入重识别分支提取目标的颜色、纹理等浅层外观特征作为重识别特征.

MobileNetV2X 网络输出的重识别分支特征图

$F_1 \in \mathbf{R}^{240\times240\times16}$ 经网络层得到 $F_{\text{re-id}} \in \mathbf{R}^{H_{ri}\times W_{ri}\times C_{ri}}$, 首先使用 3×3 卷积核依次与输入特征图卷积, 得到特征图 $F'_{\text{re-id}} \in \mathbf{R}^{H'_{ri}\times W'_{ri}\times C_{ri}}$, 再通过 128 组 1×1 的卷积, 得到具有 128 个通道的特征图 $F''_{\text{re-id}} \in \mathbf{R}^{H''_{ri}\times W''_{ri}\times 128}$, 则在特征图中提取对应的目标框中心点 $(x,\ y)$ 处的浅层外观特征作为该目标重识别特征. 同时使用全连接层和归一化操作将其映射到特征分布向量 $\boldsymbol{C} = \{c(b), b \in [1, B]\}$.

为评判重识别特征图准确度, 定义重识别损失函数, 其值越小, 表示重识别特征图越准确. 并将重识别损失函数 L_{id} 定义为

$$L_{\text{id}} = -\sum_{la=1}^{N_{\text{re-id}}}\sum_{b=1}^{B} L_{(b)}^{la}\lg(c(b)) \tag{5}$$

其中, 目标真实框的标签编码为 $L_{(b)}^{la}$, B 是训练数据中所有身份 (ID) 的编号, 使用特征图对应的目标中心点重识别特征在 YOLOX-M2X 网络进行训练, $N_{\text{re-id}}$ 表示当前帧中目标所属类别的总数.

最后, 将检测和重识别损失函数相加, 同时使用不确定性损失函数[11] 来自动平衡检测和重识别损失函数. 与单独使用 L_{id} 和 $L_{\text{detection}}$ 训练模型相比, 训练效果得到提升, 同时减少了计算复杂度, 可达到实时性要求.

1.2 基于改进卡尔曼滤波的目标跟踪

首先使用第 1 帧检测的目标框初始化目标轨迹及跟踪状态, 然后通过改进的卡尔曼滤波器预测下一帧目标位置, 最后采用连续帧间的数据关联确定目标跟踪状态.

在当前视频帧下, 设 t 时刻检测到 M 个目标, $i = 1, \cdots, M$, t 时刻跟踪 N 个目标, $j = 1, \cdots, N$, 每一帧检测及跟踪结果实时更新, 则当前 t 时刻的第 i 个检测目标状态为 $x_{t,\,i}$, 第 j 个跟踪目标状态为 $z_{t,\,j}$.

本文使用改进的卡尔曼滤波器对行人轨迹的状态进行预测和更新, 设目标状态为 $x_{t,\,i} = (\beta, \dot{\beta})^{\text{T}}$, $\beta = \{u, v, \gamma, h\}$. 其中 β 表示目标的观测值, (u, v) 表示边界框中心位置, 高宽比为 γ, 高度为 h. 目标中心点变化速率为 $(\dot{u}, \dot{v})$, 高宽比变化速率为 $\dot{\gamma}$, 高度变化速率为 $\dot{h}$. 考虑带非加性噪声的一般非线性系统模型

$$\begin{cases} x_{t,\,i} = f\left(x_{t-1,\,i}\right) + w_{t-1} \\ z_{t,\,j} = h\left(x_{t,\,i}\right) + v_t \end{cases} \tag{6}$$

其中, w_{t-1} 和 v_t 分别是过程噪声序列和量测噪声序列, 并假设 w_{t-1} 和 v_t 是均值为 0 的高斯白噪声, 其方差分别为 Q_t 和 R_t, 即 w_{t-1}~$(0,\ Q_t)$, V_t~$(0,\ R_t)$.

为详细说明改进卡尔曼滤波器预测与更新的过程, 算法 1 给出此部分的伪代码.

算法 1. 改进卡尔曼滤波算法

输入. $x_{0,\,i}$, P, Q_t, R_t.

输出. x_{t+1}.

初始化. $x_{0,\,i}$ 和协方差矩阵 P, 确定噪声协方差矩阵 Q_t 和 R_t, 并且设置时间 $t = 0$.

1) **for** $[1, t+1]$.

2) $\psi = V\sqrt{D}V$, 其中, $[V, D] = eig(P)$, eig 表示矩阵特征值分解, ψ 为正半定平方根矩阵, P 为协方差矩阵.

3) 计算均值 φ 和协方差矩阵 P_i 之间余弦相似度 $\rho_i = |\langle\varphi, P_i\rangle| / (|\varphi| \times |P_i|)$.

4) 计算容积点 $\xi_{t,\,i} = \varphi + \sqrt{\sum_{i=1}^{M}\rho_i/(\lambda\rho_i)}\psi_t$.

5) 计算权重 $\omega_{t,\,i}(\varepsilon) = \lambda\rho_i/(4\sum_{i=1}^{M}\rho_i)$.

6) 生成容积点 $\xi_{t,\,i}$ 和 $\omega_{t,\,i}$.

7) 通过以下公式估计一步状态预测 $\hat{x}_{t+1,\,i}$ 和 P_{t+1} 方差.

$$\hat{x}_{t+1,\,i} = \sum_{i=1}^{M} w_{t,\,i} f\left(\xi_{t,\,i}\right)$$

$$P_{t+1} = \sum_{i=1}^{n} w_{t,\,i} f\left(\xi_{t,\,i}\right) f^{\text{T}}\left(\xi_{t,\,i}\right) - \hat{x}_{t+1,\,i}\hat{x}_{t+1,\,i}^{\text{T}} + Q_t$$

8) 与步骤 3) 类似, 生成容积点 $\xi_{t+1,\,i}$ 和 $\omega_{t+1,\,i}$.

9) 估计输出预测 $\hat{z}_{t+1,\,j}$, 协方差 P_{t+1}^{z}.

$$\hat{x}_{t+1,\,i} = \hat{x}_{t+1,\,i} + K_{t+1}\left(z_{t+1,\,j} - \hat{z}_{t+1,\,j}\right)$$

$$K_{t+1} = P_{t+1}^{z}$$

$$P_{t+1} = P_{t+1} - K_{t+1}P_{t+1}^{z}K_{t+1}^{\text{T}}.$$

10) **end for**.

采用改进卡尔曼滤波器获取上一帧目标 i 的中心点在当前帧的预测位置 $z_{t,\,j}$, 同时通过重识别特征图 $F''_{t,\,\text{re-id}} \in \mathbf{R}^{H''_{ri}\times W''_{ri}\times 128}$ 中对应该预测中心位置, 得到上一帧目标在当前帧的预测外观特征 $F''_{t-1,\,\text{re-id}} \in \mathbf{R}^{H'_{ri}\times W'_{ri}\times 128}$. 移动机器人在跟随过程中, 会出现遮挡、快速移动等情况, 余弦距离具有快速度量的优点. 采用余弦距离 $q(i, j)$ [19] 判别当前帧中心点对应的外观特征向量 $F''_{t,\,\text{re-id}} \in \mathbf{R}^{H''_{ri}\times W''_{ri}\times 128}$ 与上一帧在当前帧的预测外观特征向量 $F''_{t-1,\,\text{re-id}} \in \mathbf{R}^{H'_{ri}\times W'_{ri}\times 128}$ 是否关联.

$$b_{i,\,j} = \boldsymbol{C}[q(i, j) \le \lambda] \tag{7}$$

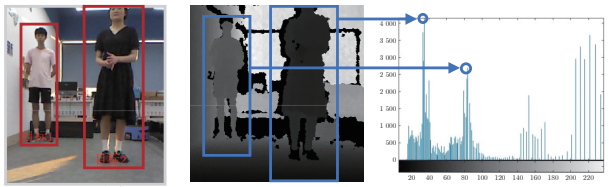
其中, $b_{i,\,j}$ 为正确关联轨迹集合. 在训练数据集上训练网络参数得到余弦距离, 并与训练集基准之间的余弦距离进行比较, 得到阈值 λ. 式 (7) 中, 当 $b_{i,\,j}$ 小于阈值 λ, 表示当前帧检测目标 i 与上一帧跟踪目标 j 关联, 则跟踪正常; 当 $b_{i,\,j}$ 大于阈值 λ, 表示未成功关联, 则继续判断目标是否遮挡或消失.

1.3　基于深度概率约束的遮挡目标跟踪

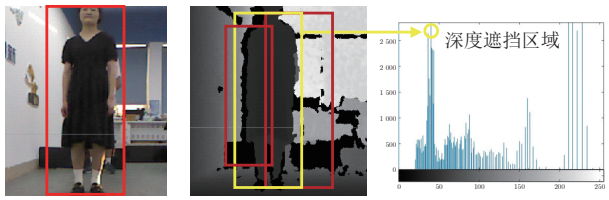
跟踪目标由于被遮挡, 目标外观会发生显著变化, 导致目标特征减少, 移动机器人跟踪目标丢失. 本文提出一种有效的遮挡处理机制, 当判断遮挡发生时, 采用深度概率对目标周围区域进行空间约束, 并通过最大后验概率 (Maximum a posteriori, MAP) 关联匹配实现遮挡跟踪.

1) 遮挡判断

由于多个目标相互遮挡时, RGB 外观被遮挡, 只可从深度信息区分不同遮挡目标, 而 ZED 相机获取的深度信息为多个遮挡目标中离该相机最近的目标深度信息. 因此将目标框在 RGB 图中的位置区域映射到深度图中并设定为深度遮挡区域, 若判定其他目标进入此区域表示发生遮挡, 具体判定如图 2 所示.



(a) 遮挡前
(a) Before occlusion



(b) 遮挡后
(b) After occlusion

图 2　遮挡前后深度直方图

Fig. 2　Depth histogram before and after occlusion

目标 1 遮挡前深度直方图最大峰值为 4000, 目标 2 遮挡前深度直方图的最大峰值为 2500, 发生遮挡后深度遮挡区域深度直方图最大峰值为 2500, 深度直方图的峰值从 4000 下降到 2500. 显然, 此时目标 1 的深度遮挡区域的深度直方图出现新的上升峰值 2500, 且小于遮挡前的峰值 4000, 则可见目标被遮挡后的深度直方图峰值明显减少.

因此, 可根据深度变化大小判定是否发生遮挡. 设跟踪目标 j 在 $t-1$ 帧和 t 帧之间的深度变化均值可以近似为高斯分布 $(d_t^j - d_{t-1}^j) \sim \mathrm{N}\left(w_t, \xi_t^2\right)$, 在此基础上判定是否发生遮挡.

$$D_t^{\mathrm{T}} = \mathrm{e}^{(1-T_t)} \times \max\left(\frac{\sum_j^{w_t-\xi_t} \left|\left(d_t^j - d_{t-1}^j\right)\right|}{\sum_j^{M} \left|\left(d_t^j - d_{t-1}^j\right)\right|} \times \left|\frac{w_t - \xi_t}{w_{t-1} - \xi_{t-1}}\right|, \frac{\sum_j^{w_t-\xi_t} \left|\left(d_t^j - d_{t-1}^j\right)\right|}{\sum_j^{M} \left|\left(d_t^j - d_{t-1}^j\right)\right|} \times \left|\frac{w_{t-1} - \xi_{t-1}}{w_t - \xi_t}\right| \right) \tag{8}$$

式中, d_t^j 是第 t 帧中跟踪目标 j 的深度值; $\sum_j^M |(d_t^j - d_{t-1}^j)|$ 表示第 $t-1$ 与第 t 帧之间所有目标深度差之和; $|(\omega_t - \xi_t)/(\omega_{t-1} - \xi_{t-1})|$ 表示第 t 帧与第 $t-1$ 帧之间深度值变化率, 其值足够大表示发生遮挡. $\sum_j^{w_t-\xi_t} |(d_t^j - d_{t-1}^j)|$ 代表小于 $w_t - \xi_t$ 的所有跟踪目标深度值差之和. 其遮挡判断准则为

$$T_j = \frac{|d_t^j|^2}{D_t^{\mathrm{T}}} \tag{9}$$

当目标未被遮挡时, d_t^j 接近 D_t^j, T_j 接近于 1; 当目标被遮挡时, 则 T_j 接近于 0.

2) 遮挡匹配跟踪

当目标发生遮挡时, 通过最大后验概率关联当前帧检测目标与上一帧跟踪目标, 可有效解决遮挡跟踪问题. 假设所有运动目标之间相互独立, 设单个目标轨迹组成为 S, 似然概率具有条件独立性, 则关联遮挡目标的目标函数为

$$S^* = \arg\max_S \prod_i P\left(x_{t,i} | z_{t-1,j}\right) \prod_{z_{t-1,j}} P\left(z_{t-1,j}\right) \tag{10}$$

式中, $P\left(z_{t-1,j}\right)$ 是所有跟踪目标的先验概率; $P\left(x_{t,i} \mid z_{t-1,j}\right)$ 表示当前检测目标属于跟踪目标的条件概率, 该条件概率通过检测目标与上一帧跟踪目标框的重叠率计算得到.

设当前帧检测目标 $x_{t,i}$ 的深度图对应的边界框为 $b(x_{t,i})$, 跟踪目标 $z_{t-1,j}$ 的深度图对应的边界框为 $b(z_{t-1,j})$, 通过判断 $b(x_{t,i})$ 与 $b(z_{t-1,j})$ 的重叠率来表示跟踪置信度, 式 (11) 用于求目标框的重叠率, 即

$$C = b(x_{t,i}) \cap b(z_{t-1,j}) > \sigma \tag{11}$$

式中, σ 为重叠区域, 若 C 大于 σ, 表示 $x_{t,j}$ 与 $z_{t-1,j}$ 关联匹配.

1.4　基于视觉伺服控制的目标跟随

在获取目标跟踪结果后, 选定感兴趣的一个目标作为移动机器人的跟随目标. 为使移动机器人实

现目标跟随, 本文采用基于视觉伺服控制的目标跟随算法, 使跟随目标框的中心点保持为视野范围中心点. 如目标消失, 则移动机器人按照目标运动轨迹进行主动搜索, 重新识别目标并使移动机器人继续跟随目标.

1.4.1 基于 ZED 相机的移动机器人运动学模型

由于 ZED 相机具有成像分辨率高、可获取远距离深度图像等优点, 本文采用 ZED 相机作为移动机器人视觉传感器, 其内参已标定.

假设 ZED 相机的镜头畸变小到可以忽略, 相机固有参数用针孔模型表示, ZED 相机成像原理图如图 3 所示. 在图像坐标系 Y 的跟踪目标坐标为 $P=(x_{cn}, y_{cn}, z_{cn})$, z_n 是从图像坐标和 ZED 相机的固有参数中获得

$$z_n = f \times \frac{b}{x_l - x_r} = f \times \frac{b}{d} \tag{12}$$

其中, f 为相机焦距, b 为左右相机基线, f 和 b 是通过先验信息或相机标定得到. 其中由极线约束关系, 视差 d 可由左相机中的像素点 x_l 与右相机中的像素点 x_r 对应关系计算得到.

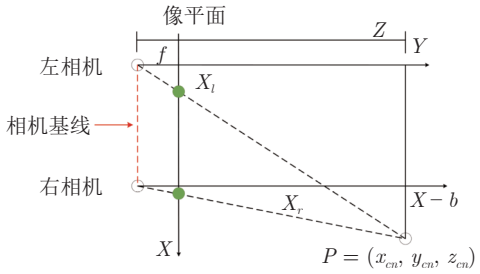


图 3 ZED 相机成像图

Fig. 3 ZED camera imagery

本文算法将移动机器人平台简化为基于 ZED 相机的两轮差速模型, 如图 4 所示. 图 4 中包括世界坐标系 G、机器人坐标系 PR、ZED 相机坐标系 Z 和图像坐标系 Y. 图中, $C(x, y)$ 为移动机器人运动中心点, D 为两轮之间的距离, θ 为方向角.

在世界坐标系 G 中, 跟踪目标位置 O_T 与机器人位置 O_M 之间的距离可表示为 $R_w^r = O_T - O_M$. 在 ZED 相机坐标系中, 目标到机器人的距离 Z_M^{T} 是由跟踪目标位置 O_T 和机器人位置 O_M 通过式 (13) 得到

$$Z_M^{\mathrm{T}} = [x_{cn}, y_{cn}, z_{cn}]^{\mathrm{T}} = R(\theta_Q, \theta_C) R_w^r - \delta d \tag{13}$$

其中, $R(\theta_Q, \theta_C)$ 表示从世界坐标系 Q 到 ZED 相机坐标系 Z 旋转矩阵. δd 表示在世界坐标系 M 中, 移动机器人与摄像机的距离.

1.4.2 机器人主动控制策略

前述跟踪算法完成目标跟踪, 并获取目标跟踪

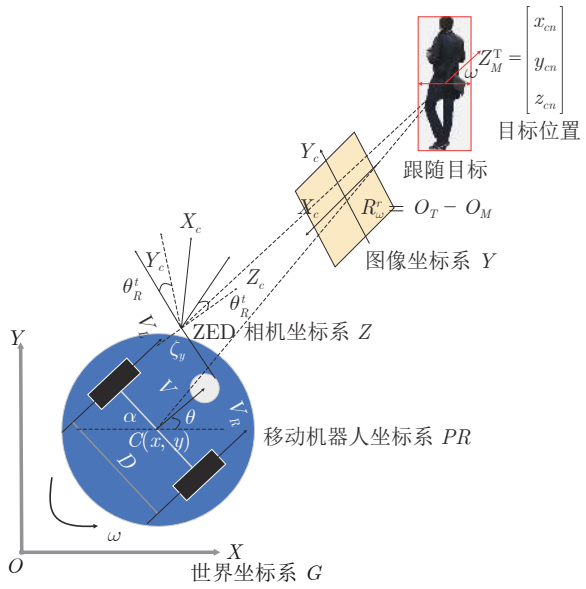


图 4 基于 ZED 相机的两轮差速驱动模型

Fig. 4 Two-wheel differential drive model based on ZED

框的深度信息, 但直接使用目标跟踪框的深度信息来计算机器人与跟随目标的距离, 会引入大量的背景信息. 因此需要对目标中心进行重定位, 为跟踪区域找到合适的位置, 提高机器人跟随精度.

目标跟踪框中心设为 $(x_l, y_l) \in \hat{\beta}$, $\hat{\beta}$ 表示目标跟踪框区域内所有像素坐标. 区域的精确位置在 $\hat{\beta}$ 内重定位, 利用循环矩阵 $\hat{k}_r$ 与 $\hat{\beta}$ 进行同或 $\odot$ 运算得到精确位置 $\hat{y}^*$, 计算式为

$$\hat{y}^* = \hat{k}_r \odot \hat{\beta} \tag{14}$$

在 $\hat{y}^*$ 中最大值 $\hat{y}_j^*$ 的位置 $(\Delta x, \Delta y)$ 即为精确目标跟踪中心 (x^*, y^*) 与 (x_l, y_l) 之间的位置偏差. 跟踪区域的精确位置 (x^*, y^*) 计算为

$$(x^*, y^*) = (x_l, y_l) + (\Delta x, \Delta y) \tag{15}$$

得到精确位置 (x^*, y^*) 后, 获取以 (x^*, y^*) 为中心区域框的 4 个顶点坐标, 计算中心点和顶点对应深度信息平均值 $f(t)$, 其值表示为移动机器人与目标的距离. 设移动机器人期望到达位置为 $f^*(t)$, 误差 $e(t)$ 可定义为

$$e(t) = f(t) - f^*(t) \tag{16}$$

机器人控制变量为 $X_{\mathrm{control}} = [U(t) = v_t, W(t) = w_t]$, v_t 代表移动机器人的线速度, w_t 代表移动机器人的角速度, PID 控制器设计为

$$\begin{bmatrix} U(t) \\ W(t) \end{bmatrix} = -\lambda \Big[k_D e(t) + k_I \sum e(t) + k_D (e(t) - e(t-1)) \Big] \tag{17}$$

其中, k_P, k_I 和 k_D 为 PID 系数, λ 是调整因子.

目标跟随控制部分的结构框图如图 5 所示, 误差 $e(t)$ 为 PID 控制器的输入端, 实时控制移动机器人角速度和线速度, 移动机器人与目标保持一定距离并稳定跟随目标. 在跟随过程中如果目标超出移动机器人视野范围, 移动机器人会保留消失前最后一帧跟踪目标的重识别特征和位置信息, 并朝着目标消失的方向继续运动. 若跟踪目标再次出现在视野内, 进行目标检测并提取该目标的重识别特征与消失前最后一帧跟踪目标的重识别特征关联匹配, 实现跟踪目标重识别即为重检成功, 则可实现目标消失后再重识别并稳定跟随目标.

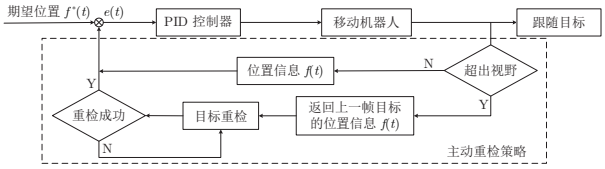


图 5　目标跟随器移动控制部分

Fig. 5　Target follower movement control section

2　实验验证

移动机器人目标跟随系统主要包括两个平台: 深度学习平台及移动机器人跟随控制平台. 深度学习平台 GPU 为 NVIDIA Geforce RTX2080SUPER, 内置软件环境为 Ubuntu16.04LTS 系统, 通过深度学习平台训练数据集获取 YOLOX-M2X 网络参数. 移动机器人跟随控制平台重 20 kg, 可承载 50 kg 的额外重量, 主要组件包括: ZED 双目相机、车体、NUC 机载微型计算机、STM32F103RCT6 底层驱动控制板等, 如图 6 所示. 机器人操作系统为 ROS (Robot operating system), ROS 具有良好的开源性和扩展性. 移动机器人跟随控制平台主要包括视觉检测模块、目标跟踪及跟随控制模块. 视觉检测模块采用 ZED 相机获取实时视频序列, 并通过深度学习平台训练数据集获取网络参数进行目标检测; 移动机器人目标跟踪及跟随控制模块采用两轮差速驱动, 移动机器人内部中心位置安装有数字姿态传感器, 可检测机器人的加速度、角速度. 移动机器人底盘前后端装有两个防跌落传感器, 其检测范围为 15 cm, 通过该模块实现目标稳定跟随.

为评估本文算法的性能, 本文在测试集和移动机器人平台进行了大量的实验. 将 DeepSORT 算法[22]、CTrack 算法[25]、FairMOT 算法[11] 与 Real-time MOT 算法[26] 等与本文算法在深度学习平台上进行验证, 并在 MOTchallenge 中的测试集上完成对比实验. 1) DeepSORT 算法的最大特点是加入外观信息, 借助重识别模型来提取特征, 减少目标 ID 切换的次数. 2) CTrack 算法是基于输入链式

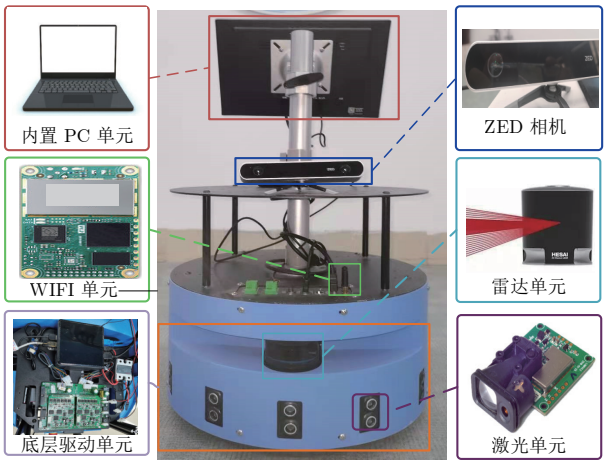


图 6　移动机器人平台

Fig. 6　Mobile robot platform

跟踪框架, 实现端到端联合检测, 同时设计了联合注意力模块来突出检测框在回归中的有效信息区域. 3) FairMOT 算法是将检测算法和重识别特征进行融合, 同时输出检测目标和重识别特征. 4) Real-time MOT 算法的创新在于将目标检测环节和外观特征融合在同一网络中, 大幅度提升多目标跟踪算法的速率. 5) 本文跟踪算法的目标检测部分采用基于无锚框的改进 YOLOX 检测算法, 同时将目标检测和重识别特征融合到轻量化网络中, 跟踪部分则采用改进卡尔曼滤波器, 可有效提高跟踪实时性.

为直观评估本文算法的性能, 对跟踪算法进行定性对比分析实验; 为定量分析本文算法性能, 在测试集上进行了网络消融实验、不同损失函数对比实验以及不同算法之间的遮挡跟踪对比实验; 最后, 在移动机器人平台上进行了室内、室外目标跟随实验.

如图 7 所示, 测试集为公开 MOTchallenge 中的测试集中的部分视频序列. 表 1 给出该测试集的帧率、分辨率、时长等属性. 测试集包括大量动态背景, 行人姿势变化多, 存在无遮挡、遮挡、交叉遮挡等情况.

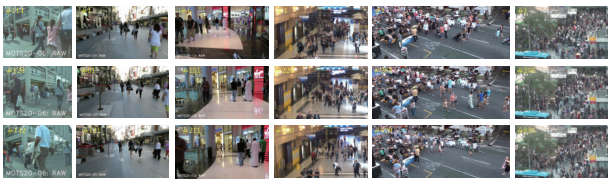


图 7　测试集

Fig. 7　Test set

2.1　定性分析实验结果

为验证本文算法的有效性, 分别在 MOT2008、MOT2002 和 HT2114 视频序列上, 将 DeepSORT

表 1 测试集视频序列
Table 1 Test set video sequences

名称	视频帧率 (帧/s)	分辨率 (像素)	视频时间 (s)	目标数量	目标框数	密集度	场景
MOT2008	25	1 920×734	806 (00:32)	279	145 301	180.3	步行街
MOT2002	25	1 920×1 080	2 782 (01:51)	296	202 215	72.7	室内火车站
HT2114	25	1 920×1 080	1 050 (00:42)	1 040	258 227	245.9	室内火车站
MOTS2007	30	1 920×1 080	500 (00:17)	58	12 878	25.8	步行街
MOTS2012	30	1 920×1 080	900 (00:30)	68	6 471	7.2	购物中心
MOTS2006	14	640×480	1 194 (01:25)	190	9 814	8.2	街道

算法、CTrack 算法、FairMOT 算法和 Real-time MOT 算法与本文算法进行目标遮挡跟踪对比实验. 图 8 是不同算法在 MOT2002 视频序列上的跟踪结果, DeepSORT 算法跟踪成功率最低, 出现了大量漏检目标. 图 8 中, Object-1 在第 26 帧时处于未遮挡状态, 除了 DeepSORT 算法外, 其余 3 种算法均能跟踪目标, 但是 FairMOT 算法中 Object-1 身份 ID 由 40 变为了 60, 出现身份 ID 信息切换. Object-2 在第 326 帧时处于未遮挡状态, Object-2 目标小、模糊且受光照影响, 仅有 Real-time MOT 和本文算法仍能稳定跟踪目标. Object-3 在第 2 701 帧时处于新出现的目标和周围物体交叉遮挡状态, DeepSORT算法、CTrack 算法和 FairMOT 算法出现跟踪丢失, 本文算法较其他 4 种算法跟踪效果更好, 区分目标与新出现的目标能力较强, 具有较好的跟踪鲁棒性. DeepSORT、CTrack、FairMOT 和 Real-time MOT 均出现一定程度上的漂移. 本文采用 YOLOX 框架下高性能的 MobileNetV2X 特征提取网络及改进的卡尔曼滤波器改善了目标在遮挡、交叉遮挡和完全遮挡的跟踪漂移问题.

接下来, 在实际场景中将本文算法与在测试集中跟踪性能较好的 FairMOT 跟踪算法在移动机器人平台上进行对比实验, 选用了实验室室内场景、学校食堂场景. 图 9 为在实验室室内场景的实验结果, 图中标注 Person ID-x 框为采用本文算法的目标跟踪框, 标注 Preson_x 框为采用 FairMOT 算法的目标跟踪框. 由图 9 可见, 遮挡前以上跟踪算法均能准确跟踪行人. 当第 122 帧行人 Person_2 被完全遮挡后, FairMOT 目标跟踪算法身份切换为 Person_3, 而本文算法仍然能稳定跟踪目标行人 ID-3. 通过与 FairMOT 算法在移动机器人平台上进行对比实验可知, 本文算法能有效地解决跟踪目标遮挡问题, 能够稳定跟随目标行人. 其原因在于本文算法相比于 FairMOT 算法加入了深度概率约束, 能够有效解决遮挡后身份 ID 错误切换问题.

如图 10 所示, 在学校食堂采集视频序列中进行了目标跟踪实验, 同时展示了 FairMOT 和本文算法的部分跟踪结果. 在第 10 帧行人 ID-1 遮挡了其他目标, 如 ID-10 被遮挡后本文算法可有效跟踪, 而 FairMOT 算法将该目标错误跟踪为 ID-29, 被遮挡的行人 ID 发生了变换, 可见本文算法相对于 FairMOT 算法 ID 错误切换较少. 分析原因得出, 本文算法引入重识别分支, 有效解决了遮挡后目标跟踪问题. 同时, 本文算法能很好地检测 FairMOT 未检测到的行人, 例如第 13 帧行人 ID-21, 得益于改进的 YOLOX 目标检测算法良好的检测性能. 通过在实际复杂场景中的跟踪结果对比, 证明了本文算法在移动机器人平台上具有良好的实时性和准确性, 同时也具有很好的鲁棒性.

2.2 定量分析实验结果

定量分析分为目标检测算法的两组消融实验和在测试集上的跟踪算法对比实验.

2.2.1 目标检测算法消融实验

为验证加入网络和重识别损失函数的有效性, 将本文算法与 3 种算法进行对比: 1) 传统 YOLOX 网络[20]; 2) 在 YOLOX 框架下用 MobileNetV2X 代替主干网络; 3) 在 YOLOX 框架下用 MobileNetV2X 代替主干网络, 再加入重识别损失函数. 为定量分析目标检测性能, 使用几种常用的评估指标[27], 如准确率 (Presion)、召回率 (Recall)、F1 分数和平均准确率 (Mean average precision, mAP)、每秒浮点运算次数 (Flops, 单位为 GHz), 结果见表 2.

首先, 比较算法 1) 与算法 2) 的性能指标, 在准确率、召回率、F1 分数、平均准确率相差不大的情况下, 每秒浮点运算次数从 21.79 GHz 大幅下降到 8.65 GHz, 可见运算复杂度大幅下降. 由于触发器导致的精度损失可以忽略不计, 因此在移动机器人平台 CPU 计算能力有限的情况下, 受计算能力限制的轻量级模型 MobileNetV2X 更加适合应用于移动机器人平台.

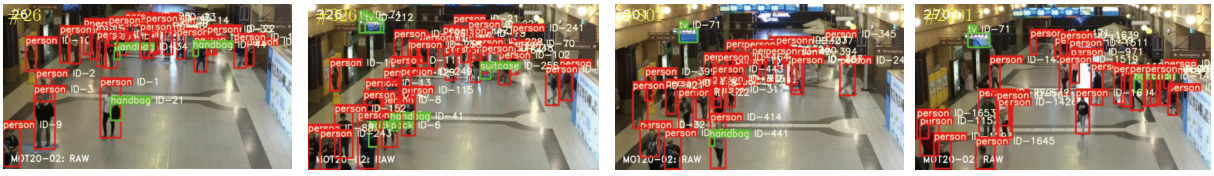
接着, 比较算法 2) 与算法 3) 的性能指标: 采用平均绝对误差 (Mean absolute error, MAE) 和均方误差 (Mean square error, MSE), 在 NWPU-

(a) DeepSORT 算法
(a) DeepSORT algorithm

(b) CTrack 算法
(b) CTrack algorithm

(c) FairMOT 算法
(c) FairMOT algorithm

(d) Real-time MOT 算法
(d) Real-time MOT algorithm



(e) 本文算法
(e) Our algorithm

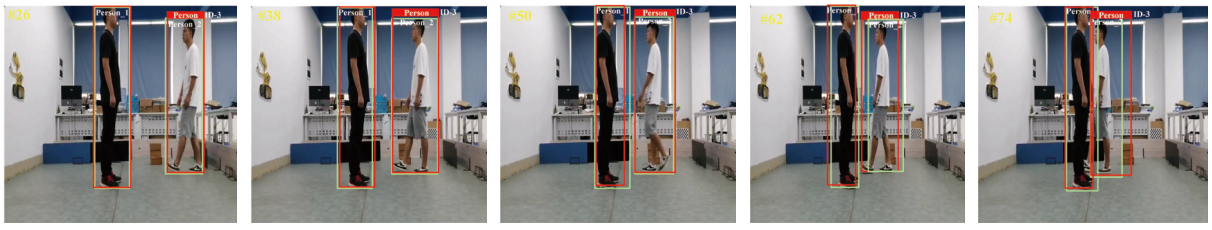
图 8　本文算法与 DeepSORT、CTrack、FairMOT、Real-time MOT 多目标跟踪算法对比分析

Fig. 8　Comparison and analysis of our algorithm with DeepSORT, CTrack, FairMOT, and Real-time MOT multi-target tracking algorithm

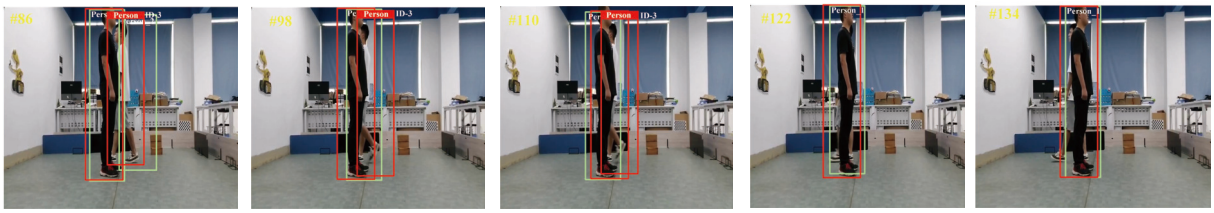
Crowd 和 UCF-QNRF 数据集上进行测试[28]. 在图 11 中, 在 2 个主干网络 (Darknet53、MobileNetV2X) 上分别进行测试. 比较不同的损失函数性能, 包括 L2 损失函数、贝叶斯损失 (Bayesian loss, BL) 函数[29]、NoiseCC 损失函数和 DM-count 损失函数. 从图 11 可见, 本文的损失函数与其他损失函数相比, Darknet53_MAE 与 MobileNetV2X_MAE 指标最低, 即本文使用的损失函数训练模型准确度高, 其重识别损失函数最后一层权重的梯度不再与激活函数相关, 只与输出值和真实值的差值成正比, 收敛快. 但 Darknet53_MSE 与 MobileNetV2X_MSE 指标略低于 DM-count 损失函数, DM-count 损失函数 MAE 指标偏高, 将 MAE 与 MSE 指标综合考虑, 本文提出的损失函数更加稳定. 从以上两组实验对比分析可见, 采用 MobileNetV2X 网络计算复杂度大大降低, 且重识别损失函数收敛快, 实时性增强.

2.2.2　目标跟踪算法实验

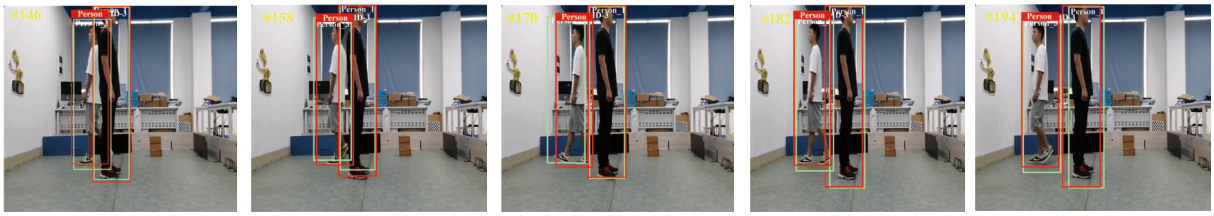
为定量分析目标跟踪算法在不同序列上目标遮挡跟踪效果, 选用 MOT2008、MOT2002 等 6 个视



(a) 遮挡前
(a) Before occlusion



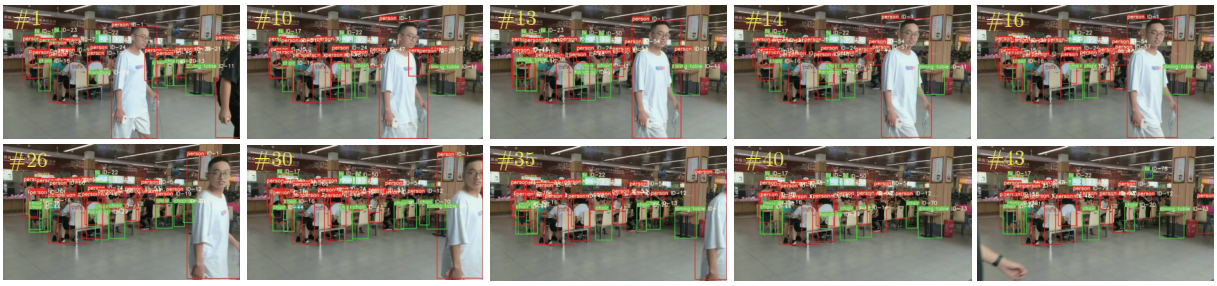
(b) 发生遮挡
(b) Occlusing



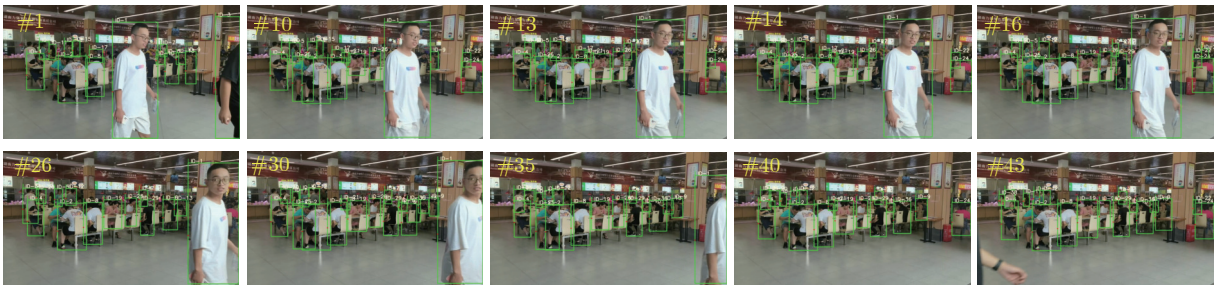
(c) 遮挡后
(c) After occlusion

图 9 移动机器人平台上 FairMOT 算法与本文算法对比实验

Fig. 9 Comparative experiment of FairMOT algorithm and our algorithm on mobile robot platform



(a) 本文算法
(a) Our algorithm



(b) FairMOT 算法
(b) FairMOT algorithm

图 10 学校食堂场景中本文算法与 FairMOT 算法对比实验

Fig. 10 Comparative experiment between our algorithm and FairMOT algorithm in school canteen scene

表 2　网络消融实验

Table 2　The ablation studies of the proposed network

ID	主干网络		Presion	Recall	F1	mAP	Flops (GHz)
	Darknet53	MobileNetV2X					
1	√		0.929	0.980	0.954	0.969	**21.79**
2		√	0.910	0.972	0.941	0.970	**8.65**
3		√	0.935	0.974	0.952	0.971	9.46
4		√	0.940	0.980	0.960	0.980	9.85

频序列进行对比分析和验证. 将本文算法与 4 种目标跟踪算法在深度学习平台上测试, 并进行对比分析. 对比算法包括 DeepSORT 算法[20]、FairMOT 算法[11]、CTrack 算法[23] 和 Real-time MOT 算法[24]. 本文在这些视频序列上测试多目标跟踪准确度 (Multiple object tracking accuracy, MOTA)、身份准确率与召回率的调和均值 (Identification F-score, IDF1) 以及跟踪目标轨迹数量 (Mostly tracked, MT) 等指标, 各项指标数据如表 3 所示. MOTA 指标评估算法性能, 基准测试范围为 MOTA $\in (-\infty, 100]$, 当跟踪器的错误数量超过场景中所有物体的数量时, 多目标跟踪精度 MOTA 为负. IDF1 为正确识别检测真实数量和平均数量之比; MT 为跟踪成功的目标数量在所有跟踪目标中所占的比例; ML 为跟踪失败的目标数量在所有跟踪目标中所占的比例; IDs 为单条跟踪轨迹改变目标标号的次数; FPS (Frames per second) 表示帧速率, 单位为帧/s. 表中的符号“↑”表示数值越大对应该符号指标性能越好; 符号“↓” 表示数值越小对应该符号的指标性能越好.

本文方法分别在 6 个数据集上测试, 如表 3 所示, 除本文算法, FairMOT 算法的 MOTA、IDF1 与 MT 指标最好. 在 MOT2008 数据集上, 本文算法除了 IDs 指标为 591, 略差于 FairMOT 算法指标 543, 其余指标均比其他算法好, 得益于本网络加入了重识别分支, 提高了遮挡跟踪身份匹配率, 同时采用改进的卡尔曼滤波器的预测与更新速率. 这几种方法的推理速度接近实时视频速率, 而本文算法更快, 在以上视频序列中, 本文算法在 MOTS2007 数据集中达到了 39 帧/s, 充分证明 YOLOX-M2X 网络实时性强, 适用于移动机器人目标跟随. 综上所述, 本文算法跟踪帧速率比其他目标跟踪算法要高, 在 MOTA 指标、IDF1 指标、MT 指标和 ML 指标上均为最优.

2.3　移动机器人目标跟随实验

为测试本文算法的鲁棒性和实时性, 在不同场景下进行了移动机器人目标跟随实验, 包括室内场景、室外场景. 首先在室内场景中进行目标跟随实

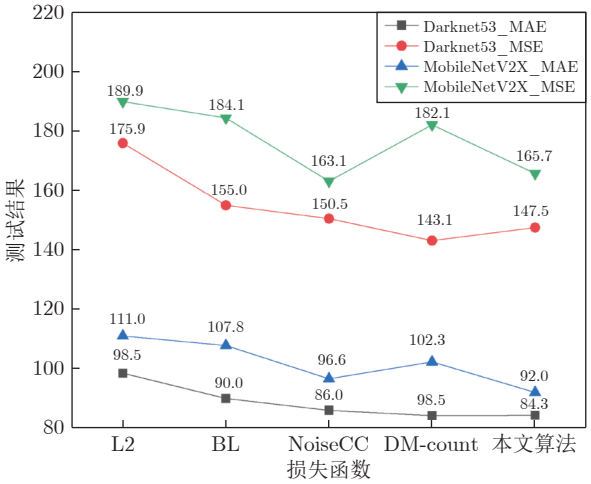


图 11　不同主干网络的不同损失函数的测试结果

Fig. 11　Test results of different loss functions for different backbone networks

验, 该室内场景存在旋转门, 柜台等障碍物, 宣传栏反光, 光照变化等影响. 如图 12 所示, 第 20 帧之后跟踪目标左转、第 23 帧避开柜台、第 35 帧跟踪行人穿过旋转门均能实时跟随目标. 实验证明, 本文算法目标检测及跟随精度高, 具有良好的实时性, 能够实现实时避障及稳定跟随目标.

然后在室外场景进行目标跟随实验, 分别选取走廊及学校道路两种场景进行移动机器人目标跟随实验. 在走廊场景中的实验如图 13 所示, 通过 ROS 可视化工具 rviz 在地图上记录移动机器人跟随路线, 中心区域为移动机器人建立的离线地图, 地图中的细线代表移动机器人运动轨迹. 跟踪行人识别身份为 ID-3. 移动机器人锁定跟踪行人 ID-3, 在锁定目标消失再出现及其他新目标出现时, 均能正确跟踪跟随锁定目标. 如第 35 帧时跟踪目标 ID-3 在转弯过程中, 快速移动消失在机器人视野范围内, 出现目标丢失情况, 移动机器人则通过主动搜寻策略, 在 rviz 界面显示了主动搜寻的轨迹, 能在目标消失再出现后继续稳定跟随目标. 如第 58 帧识别跟踪到新目标行人 ID-4 和第 73 帧识别跟踪到新目标行人 ID-5, 移动机器人仍能实时锁定跟随目标 ID-3.

在校园内进行大范围室外目标跟随的实验如图 14 所示, 左边地图显示了校园卫星地图, 图中曲线表示移动机器人运行轨迹, 包括了路面不平、光照变化、行人密集、下坡路段等复杂场景. 在该实验过程中, 机器人平稳跟随目标约 770 m, 移动机器人平均速度约为 0.42 m/s, 总跟随时间长度约为 30 min. 机器人轨迹上右边的 16 幅图像是这些位置对应的移动机器人实时跟随图. 第 1 段为机器人经过不平路面处, 机器人跟踪视野发生抖动, 影响了跟踪算

表 3 各项性能指标

Table 3 Each performance index

测试集	目标跟踪算法	MOTA ↑	IDF1 ↑	MT (%) ↑	ML (%) ↓	IDs ↓	FPS (帧/s) ↑
MOT2008	DeepSORT	47.3	55.6	30.10	18.70	625	22
	FairMOT	52.3	54.2	36.20	22.30	**543**	27
	CTrack	53.1	54.1	36.00	19.70	736	31
	Real-time MOT	52.9	52.3	29.20	20.30	709	34
	本文算法	**58.6**	**58.7**	**40.60**	**11.00**	591	**38**
MOT2002	DeepSORT	52.6	53.4	19.80	34.70	912	24
	FairMOT	59.7	53.6	25.30	22.80	1 420	28
	CTrack	61.4	62.2	32.80	18.20	781	32
	Real-time MOT	63.0	63.8	39.90	22.10	**482**	29
	本文算法	**67.9**	**68.8**	**44.70**	**15.90**	1 074	**35**
HT2114	DeepSORT	52.4	49.5	21.40	30.70	8 431	24
	FairMOT	63.0	58.6	31.20	19.90	4 137	25
	CTrack	66.6	57.4	32.20	24.20	5 529	22
	Real-time MOT	67.8	64.7	34.60	24.60	**2 583**	21
	本文算法	**73.7**	**68.3**	**38.20**	**17.30**	3 303	**27**
MOTS2007	DeepSORT	53.0	48.0	22.70	28.90	89	23
	FairMOT	60.1	49.9	28.40	25.00	135	27
	CTrack	61.2	54.0	30.60	21.60	**68**	30
	Real-time MOT	64.0	56.4	33.70	20.30	104	33
	本文算法	**67.9**	**59.3**	**35.90**	**18.40**	98	**39**
MOTS2006	DeepSORT	55.7	46.3	30.00	27.90	**67**	19
	FairMOT	63.8	49.7	31.80	25.50	79	23
	CTrack	65.3	52.6	34.10	24.00	84	25
	Real-time MOT	66.9	54.9	36.90	21.20	91	29
	本文算法	**69.1**	**57.0**	**38.00**	**18.00**	71	**31**
MOTS2012	DeepSORT	49.6	47.9	24.60	26.40	85	17
	FairMOT	52.8	49.3	27.50	24.90	68	20
	CTrack	55.7	52.1	29.90	21.70	78	23
	Real-time MOT	58.1	55.2	34.00	18.50	64	25
	本文算法	**61.3**	**56.4**	**37.50**	**15.90**	**58**	**27**

图 12 室内环境下移动机器人目标跟随实验

Fig. 12 Experiment of mobile robot target following in indoor environment

法特征网络提取特征, 经过不平的路面后, 利用主动搜寻和重识别功能, 重新捕获到目标并主动跟随. 第 2 段为机器人受到光照影响, 光照对目标的颜色、纹理特征提取影响大, 本文 YOLOX-MobileNet-V2X 网络引入数据增强, 提升了模型的鲁棒性, 保证了在光照变化下的移动机器人稳定跟随. 在第 3 段食堂附近出现了众多行人目标, 本文算法实时返回跟随行人的位置信息, 实现在复杂场景下稳定跟随. 第 4 段为下坡路面, 机器人能够实时控制角速度和线速度, 保持与目标相距一段稳定距离跟随. 可见, 在室外复杂情况下, 移动机器人仍能实现稳定地跟随目标.

3 结束语

本文提出了一种基于改进 YOLOX 的移动机

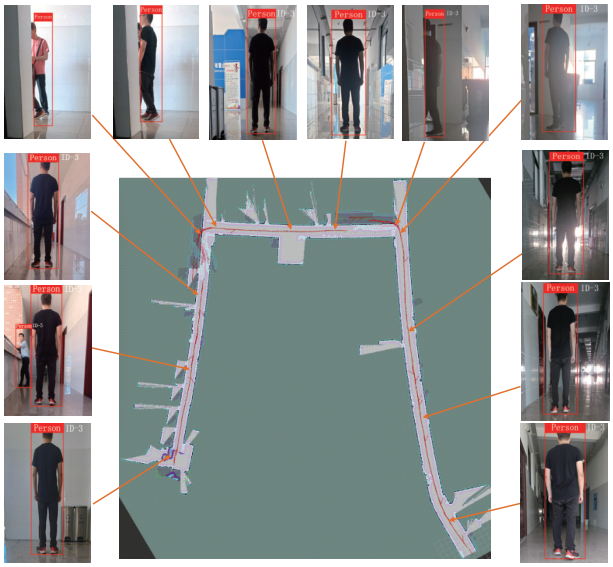


图 13　移动机器人跟随路线图

Fig. 13　Mobile robot following road map

图 14　室外环境下移动机器人目标跟随实验

Fig. 14　Experiment of mobile robot target following in outdoor environment

器人目标跟随方法, 以解决移动机器人在复杂场景中难以稳定跟随目标的问题. 针对目标检测实时性低的问题, 通过改进 YOLOX-M2X 网络实现目标检测; 针对复杂情况下目标跟踪问题, 提出改进的卡尔曼滤波器结合深度概率信息的方法, 确保机器人在遮挡情况下稳定跟踪目标; 针对机器人跟踪目标丢失情况, 设计了基于视觉伺服控制的目标跟随算法, 引入重识别特征主动搜寻目标, 实现目标跟随. 最后通过在多个测试集上的对比实验以及在移动机器人平台上的室内外实验验证, 证明了本文方法的有效性. 下一步工作将引入 Transformer[30] 作为目标检测的整体框架, 首先将 RGB-D 特征图输入目标检测网络, 采用编码器与解码器的架构, 然后在网络中提取浅层深度特征图, 用此时的深度特征图作为遮挡判断的特征, 相比直接使用深度特征, 能够提高遮挡判断的准确度, 从而进一步提高目标跟随的稳定性.

References

1　Wang Li-Jia, Jia Song-Min, Li Xiu-Zhi, Wang Shuang. Person following for mobile robot using improved multiple instance learning. *Acta Automatica Sinica*, 2014, **40**(12): 2916−2925
(王丽佳, 贾松敏, 李秀智, 王爽. 基于改进在线多示例学习算法的机器人目标跟踪. 自动化学报, 2014, **40**(12): 2916−2925)

2　Cao Feng-Kui, Zhuang Yan, Yan Fei, Yang Qi-Feng, Wang Wei. Long-term autonomous environment adaptation of mobile robots: State-of-the-art methods and prospects. *Acta Automatica Sinica*, 2020, **46**(2): 205−221
(曹风魁, 庄严, 闫飞, 杨奇峰, 王伟. 移动机器人长期自主环境适应研究进展和展望. 自动化学报, 2020, **46**(2): 205−221)

3　Yu Duo, Wang Yao-Nan, Mao Jian-Xu, Zheng Hai-Hua, Zhou Xian-En. Vision-based object tracking method of mobile robot. *Chinese Journal of Scientific Instrument*, 2019, **40**(1): 227−235
(余铎, 王耀南, 毛建旭, 郑海华, 周显恩. 基于视觉的移动机器人目标跟踪方法. 仪器仪表学报, 2019, **40**(1): 227−235)

4　Huang Yan, Li Yan, Yu Jian-Cheng, Feng Xi-Sheng. State-of-theart and development trends of AUV intelligence. *Robot*, 2020, **42**(2): 215−231
(黄琰, 李岩, 俞建成, 封锡盛. AUV 智能化现状与发展趋势. 机器人, 2020, **42**(2): 215−231)

5　Marvasti-Zadeh S M, Cheng L, Ghanei-Yakhdan H, Kasaei S. Deep learning for visual tracking: A comprehensive survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021, **23**(5): 3943−3968

6　Ciaparrone G, Sánchez F L, Tabik S, Troiano L, Tagliaferri R, Herrera F. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 2020, **381**(14): 61−88

7　Yoshimi T, Nishiyama M, Sonoura T, Nakamoto H, Tokura S, Sato H, et al. Development of a person following robot with vision based target detection. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China: IEEE, 2006. 5286−5291

8　Jiang Hong-Yi, Wang Yong-Juan, Kang Jin-Yu. A survey of object detection models and its optimization methods. *Acta Automatica Sinica*, 2021, **47**(6): 1232−1255
(蒋弘毅, 王永娟, 康锦煜. 目标检测模型及其优化方法综述. 自动化学报, 2021, **47**(6): 1232−1255)

9　Zhang M Y, Liu X L, Xu D, Cao Z Q, Yu J Z. Vision-based target-following guider for mobile robot. *IEEE Transactions on Industrial Electronics*, 2019, **66**(12): 9360−9371

10　Pang L, Cao Z Q, Yu J Z, Guan P Y, Chen X C, Zhang W M. A robust visual person-following approach for mobile robots in disturbing environments. *IEEE Systems Journal*, 2019, **14**(2): 2965−2968

11　Zhang Y F, Wang C Y, Wang X G, Zeng W J, Liu W Y. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 2021, **129**(11): 3069−3087

12　Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 779−788

13　Hsu W Y, Lin W Y. Ratio-and-scale-aware YOLO for pedestrian detection. *IEEE Transactions on Image Processing*, 2020, **3029**: 934−947

14　Huang J H, Zhang H Y, Wang L, Zhang Z L, Zhao C M. Improved YOLOv3 model for miniature camera detection. *Optics and Laser Technology*, 2021, **142**: Aricle No. 107133

15　Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection [Online], available: https://arxiv.org/abs/2004.10934, April 23, 2020

16　Benjumea A, Teeti I, Cuzzolin F, Bradley A. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles [Online], available: https://arxiv.org/abs/2112.11798, December 22, 2021

17　Cheng L, Liu W Z. An effective microscopic detection method for automated silicon-substrate ultra-microtome (ASUM). *Neur-*

al Processing Letters, 2021, **53**(3): 1723−1740

18 Conley G, Zinn S C, Hanson T, McDonald K, Beck N, Wen H. Using a deep learning model to quantify trash accumulation for cleaner urban stormwater. *Computers, Environment and Urban Systems*, 2022, **93**: Aricle No. 101752

19 Hussain M, Al-Aqrabi H, Munawar M, Hill R, Alsboui T. Domain feature mapping with YOLOv7 for automated edge-based pallet racking inspections. *Sensors*, 2022, **22**(18): Aricle No. 6927

20 Ge Z, Liu S T, Wang F, Li Z M, Sun J. YOLOX: Exceeding YOLO series in 2021 [Online], available: https://arxiv.org/abs/2107.08430, July 18, 2021

21 Yan F X, Xu Y X. Improved target detection algorithm based on YOLO. In: Proceedings of the 4th International Conference on Robotics, Control and Automation Engineering (RCAE). Wuhan, China: IEEE, 2021. 21−25

22 Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE, 2017. 3645−3649

23 Han D Y, Peng Y G. Human-following of mobile robots based on object tracking and depth vision. In: Proceedings of the 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA). Shanghai, China: IEEE, 2020. 105−109

24 Addabbo T, Fort A, Mugnaini M, Vignoli V, Intravaia M, Tani M, et al. Smart gravimetric system for enhanced security of accesses to public places embedding a mobilenet neural network classifier. *IEEE Transactions on Instrumentation and Measurement*, 2022, **71**: 1−10

25 Peng J L, Wang C G, Wan F B, Wu Y, Wang Y B, Tai Y, et al. Chained-tracker: Chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking. In: Proceedings of the European Conference on Computer Vision Springer. Glasgow, UK: Cham, 2020. 145−161

26 Wang Z D, Zheng L, Liu Y X, Li Y L, Wang S J. Towards real-time multi-object tracking. In: Proceedings of the European Conference on Computer Vision Springer. Glasgow, UK: Cham, 2020. 107−122

27 Milan A, Leal-Taixé L, Reid L, Roth S, Schindler K. MOT16: A benchmark for multi-object tracking [Online], available: https://arxiv.org/abs/1603.00831, March 2, 2016

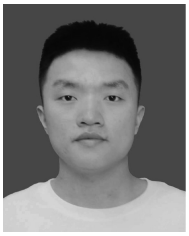
28 Song Y, Zhang Y Y, Liu L. Path following control of tracked mobile robot based on dual heuristic programming. In: Proceedings of the 5th International Conference on Control, Automation and Robotics (ICCAR). Beijing, China: IEEE, 2019. 79−84

29 Aetesam H, Maji S K, Yahia H. Bayesian approach in a learning-based hyperspectral image denoising framework. *IEEE Access*, 2021, **9**: 169335−169347

30 Koay H V, Chuah J H, Chow C O. Shifted-window hierarchical vision transformer for distracted driver detection. In: Proceedings of the IEEE Region 10 Symposium (TENSYMP). Jeju, South Korea: IEEE, 2021. 1−7

万 琴 湖南工程学院电气与信息工程学院教授. 2010 年获得湖南大学博士学位. 主要研究方向为机器视觉, 模式识别. 本文通信作者.
E-mail: wanqin_10@126.com
(**WAN Qin** Professor at the College of Electrical and Information Engineering, Hunan Institute of Engineering. She received her Ph.D. degree from Hunan University in 2010. Her research interest covers machine vision and pattern recognition. Corresponding author of this paper.)

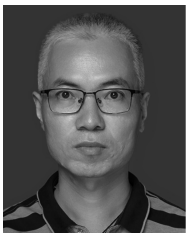
李 智 湖南工程学院电气与信息工程学院硕士研究生. 主要研究方向为目标跟踪, 目标跟随机器人.
E-mail: lizhi_09@126.com
(**LI Zhi** Master student at the College of Electrical and Information Engineering, Hunan Institute of Engineering. His research interest covers target tracking and target tracking method of mobile robot.)



李伊康 湖南工程学院电气与信息工程学院硕士研究生. 主要研究方向为微电网多目标成本优化模型构建.
E-mail: liyikang0906@163.com
(**LI Yi-Kang** Master student at the College of Electrical and Information Engineering, Hunan Institute of Engineering. His research interest covers multi-objective cost optimization modeling on microgrid.)

葛 柱 湖南工程学院电气与信息工程学院硕士研究生. 主要研究方向为目标检测, 机器人多目标跟踪.
E-mail: gezhu_06@163.com
(**GE Zhu** Master student at the College of Electrical and Information Engineering, Hunan Institute of Engineering. His research interest covers target detection and robot multi-target tracking.)

王耀南 中国工程院院士, 湖南大学电气与信息工程学院教授. 1995 年获得湖南大学博士学位. 主要研究方向为机器人学, 智能控制和图像处理.
E-mail: yaonan@hnu.edu.cn
(**WANG Yao-Nan** Academician at Chinese Academy of Engineering, professor at the College of Electrical and Information Engineering, Hunan University. He received his Ph.D. degree from Hunan University in 1995. His research interest covers robotics, intelligent control, and image processing.)



吴 迪 湖南工程学院电气与信息工程学院副教授. 2014 年获得兰州理工大学博士学位. 主要研究方向为多模态融合行人再识别, 目标检测.
E-mail: wudi6152007@163.com
(**WU Di** Associate professor at the College of Electrical and Information Engineering, Hunan Institute of Engineering. He received his Ph.D. degree from Lanzhou University of Technology in 2014. His research interest covers spatial-temporal person re-identification and target detection.)