

基于RDK的目标识别算法研究

李俊, 兰红

江西理工大学信息工程学院, 江西 赣州

收稿日期: 2022年10月8日; 录用日期: 2022年11月7日; 发布日期: 2022年11月14日

摘要

目前, 目标检测算法的应用场景不易灵活变动, 机器人在执行复杂问题时, 面临较高的计算和存储资源的问题, 以及机器人的可扩展性较低的问题。针对机器人进行目标检测方面的扩展, 提出用基于RDK云服务机器人实现多场景式的目标检测, 基于YOLOv5优化的目标检测算法对环境中的多种物体进行识别。云机器人服务的分布式部署可提高资源的利用率。通过编辑蓝图调用算法模型和基于Blender设计的动作, Smart Voice语音系统可以调用整个蓝图事件, 完成数据采集、目标导航、物体识别、机器人运动等多种工作, 达到实现基础服务功能的目标。

关键词

RDK蓝图, 服务机器人, YOLOv5算法, Blender动作设计, Smart Voice语音系统

Research on Target Recognition Algorithm Based on RDK

Jun Li, Hong Lan

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: Oct. 8th, 2022; accepted: Nov. 7th, 2022; published: Nov. 14th, 2022

Abstract

At present, the application scenarios of target detection algorithms are not flexible, and robots face the problems of high computing and storage resources and low scalability when executing complex problems. Aiming at the expansion of robot target detection, this paper proposes to use RDK-based cloud service robots to achieve multi-scene target detection, and recognize multiple objects in the environment based on YOLOv5 optimized target detection algorithm. Distributed deployment of cloud robot services can improve resource utilization. By editing the blueprint calls the algorithm model and actions based on Blender design, Smart Voice voice system can call the whole blueprint event to complete data collection, target navigation, object recognition, robot movement and other work, and achieve the goal of realizing basic service functions.

Keywords

RDK Blueprint, Service Robot, YOLOv5 Algorithm, Blender Action Design, Smart Voice Voice System

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着技术的进步, 以及人类对智能服务机器人的要求越来越高, 兼具高机动性能、智能化交互的机器人正在成为全球的研究焦点。并且现阶段新冠肺炎疫情防控压力犹存, 减少公共场所人员接触是防止疫情扩散的有效手段, 在这样的背景下, 用机器人代替人工的需求被激发。

就目前而言, 在机器人方面的研究如雨后春笋般涌现。但许多的研究方向都是基于具有极佳运动控制系统的仿生机器人, 刘京运[1]所阐述的四足机器人进化史介绍了仿生机器人在运动控制方面的快速发展。在基于深度强化学习方面的机器人路径规划[2]和机器人运动控制[3]方面都有长足的发展。在大多数的服务机器人工作的环境下, 极佳的运动控制系统和仿生机器人的优势都不能够得到十足的体现, 在服务机器人领域, 最为关注的问题还是资源的利用和人机交互的优化方面。而伴随着 5G 技术的发展与突破, 通过云端建立云服务机器人系统是提升机器人计算和存储资源的重要途径。基于 SOA 的云服务机器人框架设计[4]主要针对于资源共享负载均衡进行了阐述, 具体针对与机器人的运转并未过多介绍。也有部分针对于机器人操作和目标检测相结合的研究, 例如结合目标检测和机器人抓取实现番茄的采摘[5]、水面垃圾的清理[6]以及针对识别算法进行优化, 提升机器人的抓取成功率[7], 但大都是在针对于固定任务场景下的对单项任务的完成或是抓取的运动策略方面的优化。鲜有涉及到云机器人交互设计、多场景任务模式等方面的研究。虽然目标检测已应用在各个广泛领域, 但是云服务机器人并未得到广泛的应用, 目前, 在基于云端机器人系统向目标检测应用方向扩展的研究较少。

因此, 为解决多适应场景下的基于云服务机器人的目标检测算法的实现问题。本文以智能服务机器人在仿真咖啡厅场景下的多种交互操作任务为基础, 将 YOLOv5 目标检测算法结合机器人自然语音理解、导航行走、物品交互等多模态技术。通过达闼科技[8]的 HARIX 海睿云端大脑完成对云服务机器人的控制, 让目标检测算法能够上传至云端, 并在机器人上得到调用实现, 达到实现多应用场景下目标检测的实现, 能够完成识别并触发空调和灯的控制开关、冲泡咖啡、倒水等多种应用场景下的任务实现。

2. 总体设计

结合功能需求, 本作品需要以下 6 个方面的支持, 分别是 RDK 蓝图设计、数据采集和处理、YOLOv5 算法训练、算法镜像部署、Blender 动作设计、Smart Voice 智慧语音。技术方案主要是采用 YOLOv5 算法训练识别模型, 使用 docker 将模型封装为镜像上传至云端, 进行 RDK 蓝图编程调用云端部署的算法, 使用语音驱动意图进行识别, 使用 Blender 动作进行位姿的调整, 完成各项任务。在整个实现的过程中, 研究方法主要包括理论研究、仿真实验、对比分析、验证优化等方法, 在对机器人本身进行操作的过程中, 需要通过不断地仿真实验和对比分析进行调整参数的设置, 例如在蓝图的设计中要所有的蓝图节点能够按顺序准确运行, 每个节点函数的调用都要合理合规, 在动作设计中, 要求对位姿能有精准的调整, 对抓握等动作要有准确的设计, 以求达到最佳的使用效果。在对算法的训练中, 要不断地进行验证优化,

让算法在准确度和泛化能力上都有较优的效果。研究方法主要包含作品的技术路线, 如图 1 所示。

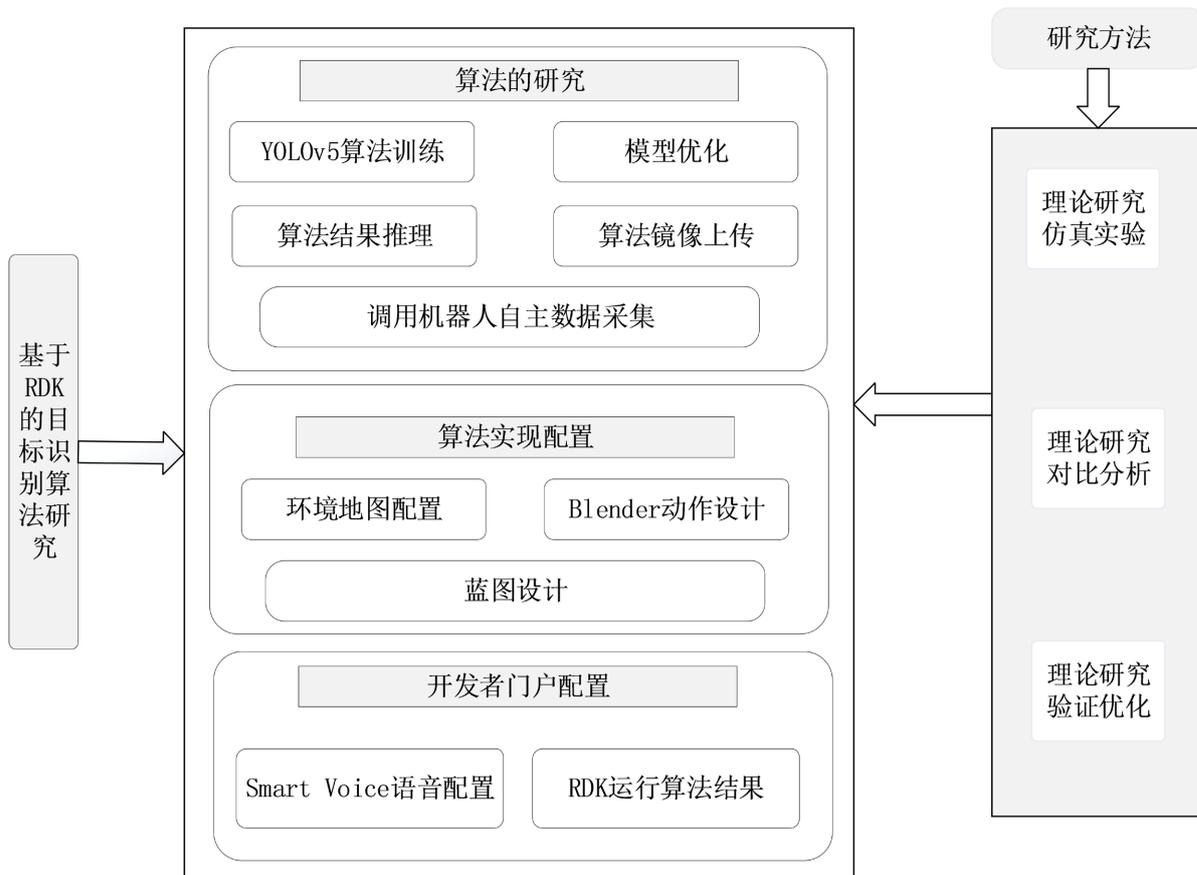


Figure 1. Overall system architecture diagram
图 1. 系统总体架构图

3. RDK 蓝图设计

蓝图也被称为可视化编程, 将 C++语言的代码可视化, 通过逻辑性编入到引擎中。在此主要包含数据采集和算法实现两个蓝图的设计。

3.1. 数据采集蓝图的设计

在虚拟环境中进行数据采集, 需先通过 Smart Voice 语音系统调用数据采集的蓝图事件, 再通过导航

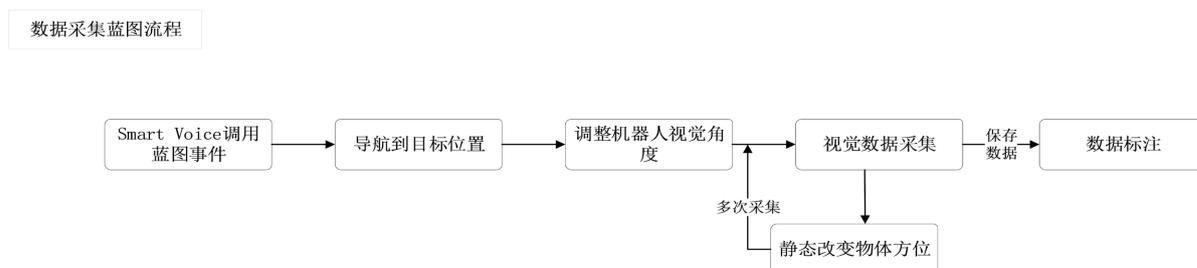


Figure 2. Design drawing of data collection blueprint
图 2. 数据采集蓝图设计图

让机器人到达指定位置, 经过对机器人位姿的调整后使目标物体在视觉范围内, 便可开始调用数据采集蓝图开始采集数据, 每采集一张图片都对目标物体进行静态调整, 调整后便可继续采集, 以提高数据的泛化能力。数据采集的蓝图设计如图 2 所示。

数据采集可以通过 for 循环进行多次采集, 避免了单次采集需要大量调用相同蓝图节点, 减少了蓝图节点的冗余, 循环采集部分的蓝图如图 3 所示。

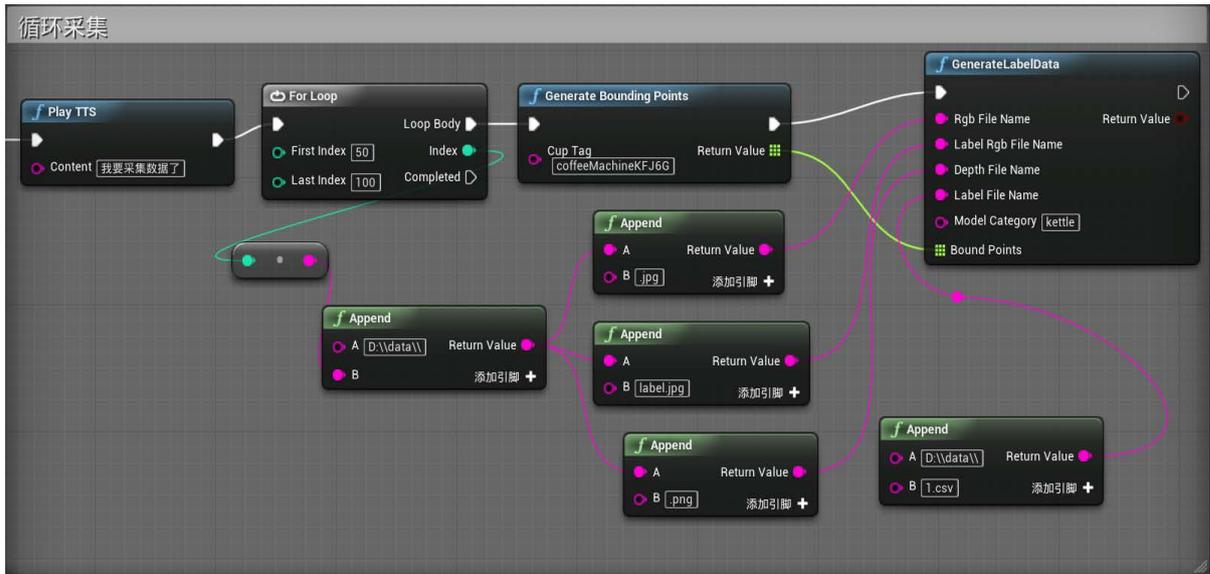


Figure 3. Blueprint of data collection

图 3. 数据采集蓝图

以虚拟环境中的开关为例, 在图片中包含选择采集的开关并且相应开关有红色方框进行标识, 在此需要进行准确的位姿调节, 摄像头的角度向下倾斜, 机器人的机身向右轻微旋转, 便可以采集到开关在图片中央的数据, 为提高数据的多样性, 可以在机器人的位姿上进行多种调整, 以获得不同效果的大量数据, 如图 4 所示。

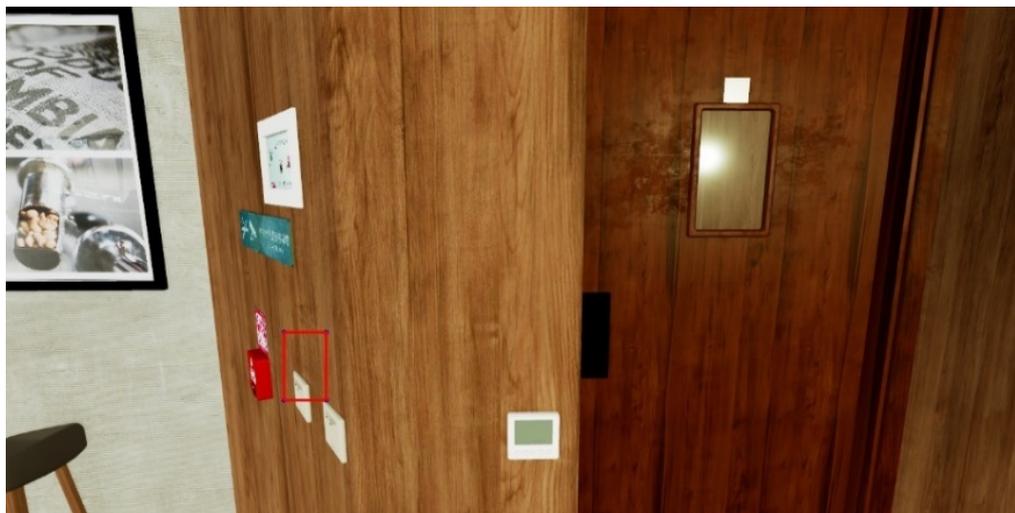


Figure 4. Data acquisition results at the switch

图 4. 开关处数据采集结果

3.2. 算法调用蓝图的设计

实现目标检测算法的运行, 首先, 要通过 Smart Voice 语音调用蓝图事件, 驱动导航到目标位置, 在目标位置进行简单的位姿调节后, 调用算法识别蓝图节点, 在识别结束后, 可通过语音反馈结果, 并结合检测结果判断是否要进行后续的动作运行。算法调用的蓝图设计如图 5 所示。

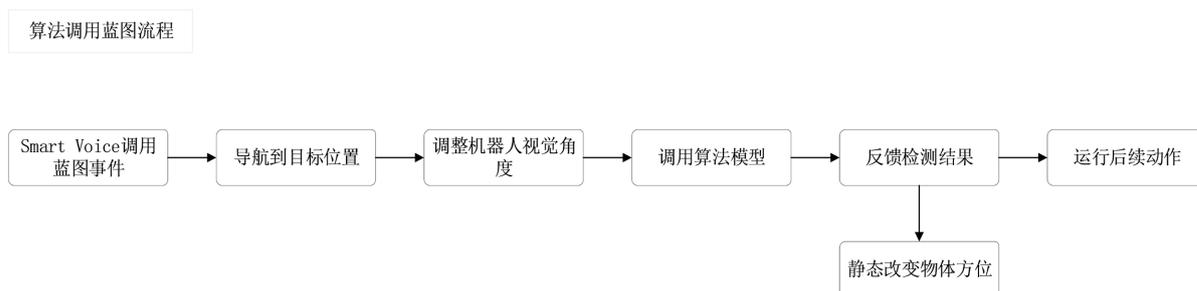


Figure 5. Design drawing of algorithm call blueprint

图 5. 算法调用蓝图设计图

蓝图设计是整个机器人运行的基础, 机器人通过蓝图的流程顺序逐个完成相应节点的任务, 节点的内容环环相扣, 当上一节点的内容没有完成时, 很大概率会影响到后续节点的运行, 所以蓝图的设计要经过多次的模拟实验, 以求各个环节都能达到理想的效果, 最终完成整个动作流程。

4. YOLOv5 目标检测算法

4.1. 算法架构

在识别算法上采用 YOLOv5 [9] [10]进行算法训练。YOLO 是端到端目标检测算法的代表, 与 R-CNN 系列最大区别是不需求取候选区域, 只需通过一个单独的 CNN 网络就能够实现目标的定位和识别。YOLO 的显著特点是检测速度快, 但其缺点是精度相比于 Fast-RCNN 等有所下降, 是一种典型的精度换速度的做法, 网络结构图如图 6 所示。

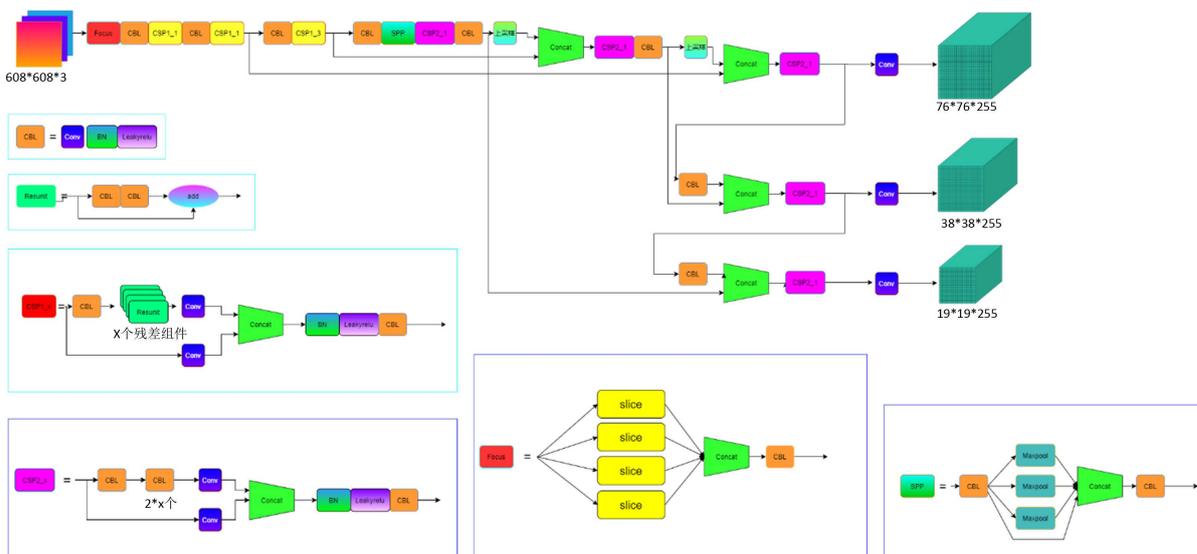


Figure 6. YOLOv5 network structure diagram

图 6. YOLOv5 网络结构图

YOLOv5 的网络结构主要包括以下四个部分, 各个模块功能如图 7 所示。

- 输入端: Mosaic 数据增强、自适应锚框计算、自适应图片缩放;
- Backbone: Focus 结构, CSP 结构, SPP (Spatial Pyramid Pooling [空间金字塔池化]);
- Neck: FPN + PAN 结构;
- Prediction: GIoU_Loss。

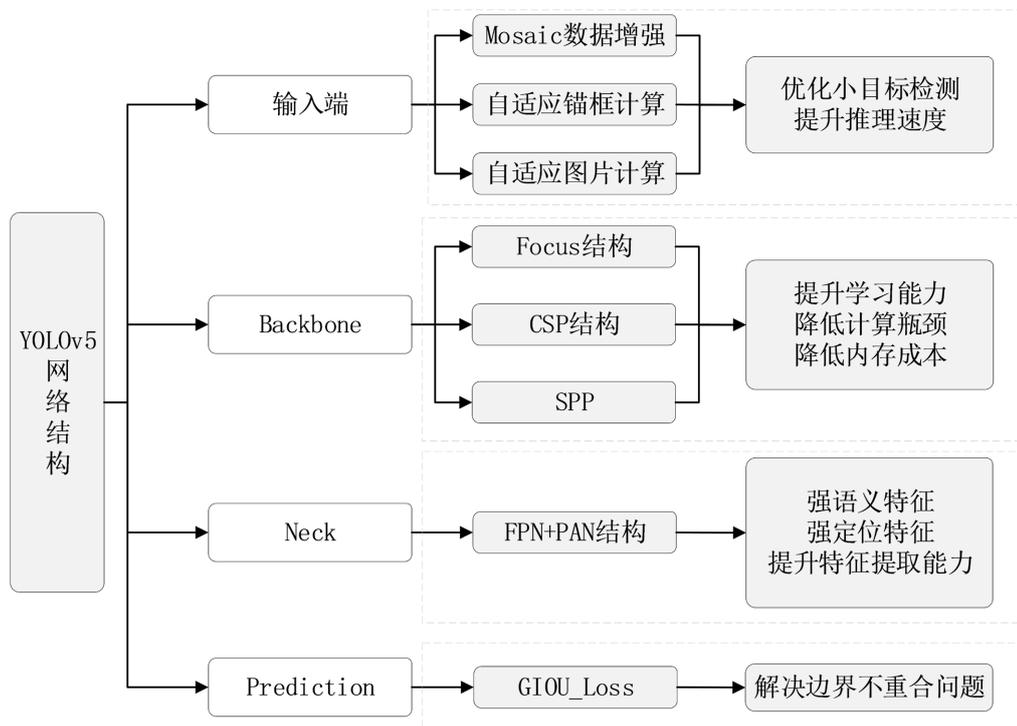


Figure 7. YOLOv5 structure classification diagram

图 7. YOLOv5 结构分类图

4.2. 算法优化

算法训练时先将数据集图片导入输入端经 Mosaic 数据增强、自适应锚框计算、自适应图片缩放后进入 Focus 切片, 通过卷积核得到特征图, 然后进入 Neck, 经过 FPN + PAN 操作, 使用 CSP2 加强网络特征融合能力, 输出端计算 Bounding box 使用 GIoU_Loss 做损失函数, 输出结果。

YOLOv5 小目标检测效果不佳的原因之一便是小目标样本的尺寸较小, 而采样倍数较大、较深的特征图很难学习到小目标的特征信息, 因此, 增加小目标检测层对较浅特征图于较深特征图拼接后进行检测, 添加[5,6,8,14,15,11]的小检测层。

在 YOLOv5 中插入卷积注意力模块 CBAM, CBAM 可以无缝衔接到 CNN 架构中, 并且对算力和内存的开销不会产生较大影响, 可以与 CNN 网络一起进行端到端的训练。若有一个中间特征作为卷积层输出, 则根据层的深度每个这样的特征层都会捕获边缘、形状等信息, 以获取更复杂的输入语义表示。使得网络在训练过程中更多地关注特征图的重要部分。

4.3. 模型训练

将在仿真环境中收集的图片进行标注作为模型训练的数据集。在 YOLOv5/data 文件夹下创建自己数据集的路径、种类数量、种类名称的.yaml 文件, 将自己的数据集以 data/image/train、data/image/val、

data/labels/train、data/labels/val 的结构来组成。然后使用 `python train.py—img 640—batch 8—epochs 500—data data/dataset.yaml—cfg yolov5m.yaml—weights yolov5m.pt` 的命令运行, 结果会在 YOLOv5 文件夹下的 runs/train 中。

5. 机器人运行实现

5.1. 机器人运行规划

在算法训练优化完成后得到算法模型, 还需将算法模型进行镜像制作并上传, 在结合机器人语音匹

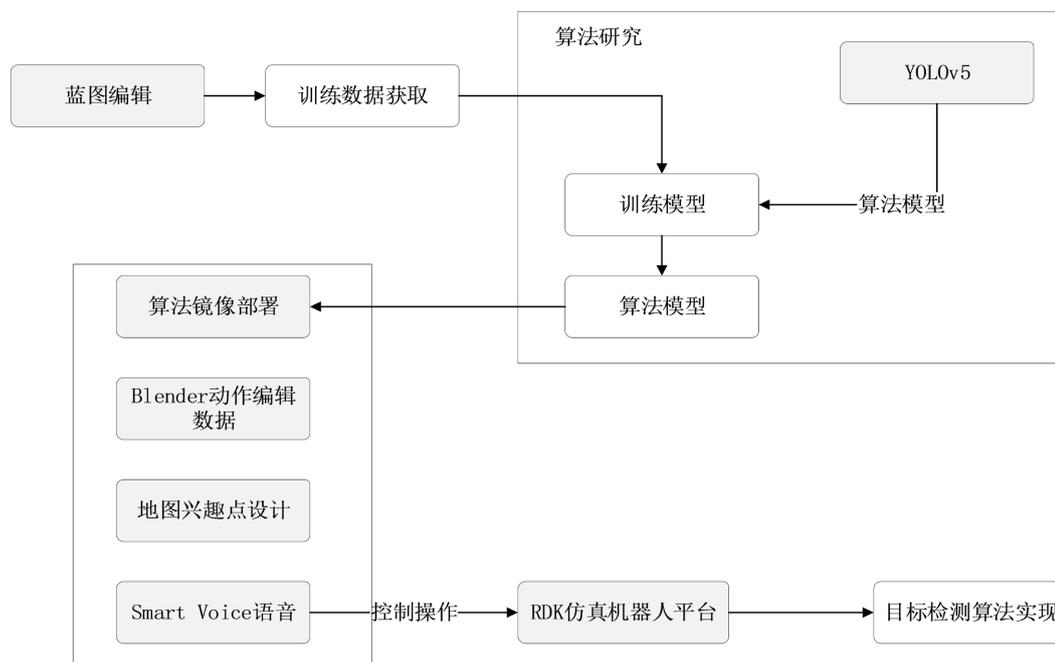


Figure 8. Design drawing of algorithm implementation

图 8. 算法实现设计图

配及动作设计的基础上, 才能驱动机器人进行目标算法的识别。算法的实现设计包含算法镜像部署、Blender 动作设计、地图兴趣点设计、Smart Voice 等 4 个部分, 如图 8 所示。

5.2. 算法镜像部署

将已经训练好的模型通过 docker 制作镜像, 再将镜像上传到云端以获得唯一域名地址, 通过调用地址直接运行识别模型。在蓝图中通过 ObjectDetect 类可以调用模型。将已配置好使用对应镜像的蓝图进行打包上传, 再将已经上传的技能与机器人选择的角色进行配对绑定, 则配对成功的机器人角色便可再虚拟环境中使用训练好的目标识别算法。基于 RDK 平台通过 Smart Voice 语音系统对已经设置好的机器人角色进行任务调用, 即可在仿真环境中运用算法进行操作。

5.3. Blender 动作设计

机器人的动作规划通过 Blender 平台和蓝图中关节组数的节点进行编辑规划, 经过实际的测试, setJointStates 节点设置关节数组来产生的动作更适应于瞬间动作, 通过 Blender 进行帧动作编辑, 对关键帧进行动作设定, 在数据导出时会自动平滑地补充间隙帧的数据, 减少了对冗余信息的重复操作, 从而大量缩短了动作设计时间。在蓝图中拖拽 PlayMotionAsync 节点来播放指定路径动作, 因此, 在设计动

作的时候需要与物体的位置和机器人所处的位置相匹配, 否则, 动作正确也难以达到预期效果。

首先, 通过 RDK 环境获取物体的位置信息, 利用 Blender 编辑器按时间轴的编辑方式编辑每个关节的状态, 并添加蓝图给出到达物体的路径策略, 控制仿真机器人实现位姿的调整和目标物品的识别和抓取。根据以上任务描述, 项目需要进行的步骤包括: 获得仿真环境下机器人与物品之间的高度和角度关系、Blender 编辑器控制不同关节运动的实现、语音下兴趣点导航和动作之间的蓝图配合等。动作规划如下图 9 所示。

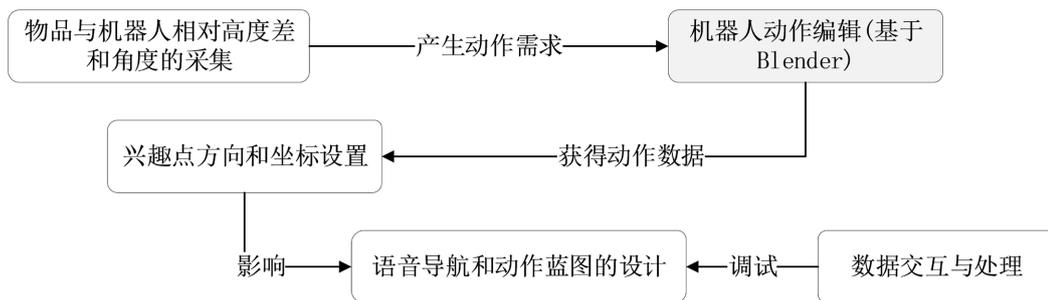


Figure 9. Design drawing of Blender action planning
图 9. Blender 动作规划设计图

如图 10 所示, 采用不同视角获取的方案, 进行 2D 基本几何规划方法, 可以有效提高获取物体信息的效率, 采用多次信息对比, 减小误差, 以提高精准度。由于导航下会将位姿进行更改, 一定程度上需要多次实验不同坐标和角度的兴趣点, 提高动作的精准度。此方法适用于粗略估计物品与本体之间的相对位置, 实际调整数据, 需要对关节进行多次调整, 才能获得较为精准地动作数据。将最终获得的合格的数据导出到本地, 再通过蓝图节点 PlayMotionAsync 调用已导出的动作数据, 便可以通过蓝图运行动作数据完成相应的任务需求。

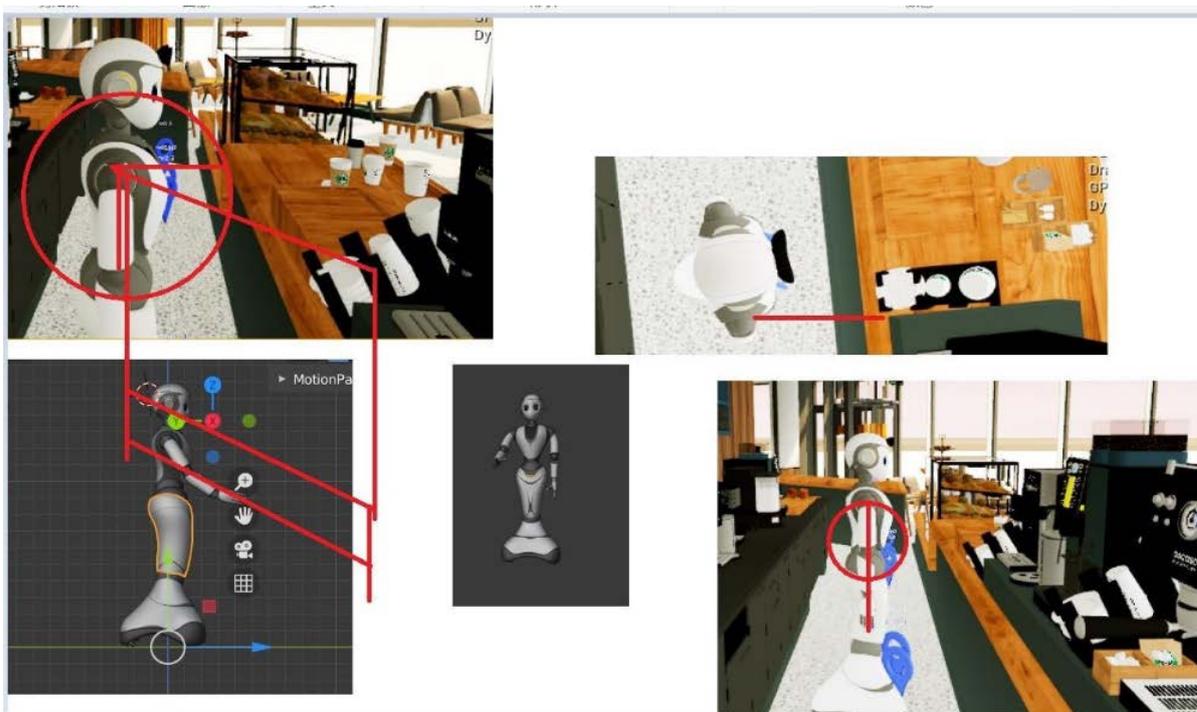


Figure 10. Coffee making plan
图 10. 咖啡制作规划图

5.4. 地图设计

在开关处包含控制灯、空调、窗帘、门的多个开关，并且开关的碰撞体积极小，需要精准调节兴趣点的位置方向，并且还需精确设计动作规划，才能精确地执行对各个开关进行触发。如图 11 是各个不同开关之间的兴趣点设置。

5.5. Smart Voice 语音设计

在 Smart Voice 网站中设置服务项目，添加 GoToCurtain、GoToDesk、GoToCoffee 等意图。当语音识别结果命中该意图时，控制蓝图中对应的程序将开始执行，即完成该意图所对应的机器人动作。Smart Voice 相当于一个将语音转换为文字的过程，在转换为文本之后与已经设定的文本内容进行对比，如果有



Figure 11. Display diagram of navigation point setting at the switch
图 11. 开关处导航点设置展现图

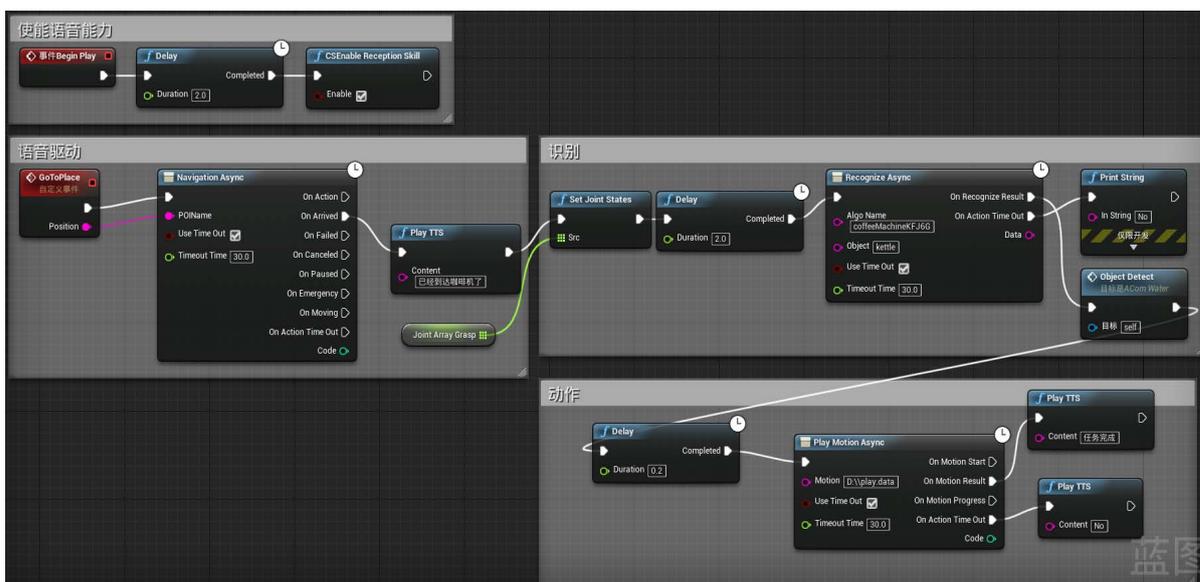


Figure 12. Blueprint of algorithm call
图 12. 算法调用蓝图

比对成功则进行任务执行, 如果没有比对成功则任务不在设定范围之内, 所以现需要设置大量的替换词等, 让一种任务的执行尽可能包含多种的语言表达形式。

6. 项目实施

在实际测试的过程中, 需要实现图 12 蓝图所示的调用, 通过语音导航到目标位置后再通过调用 URL 地址运行识别模型。可得如图 13 和图 14 所示结果, 识别的大致准确度分别为 90% 和 92.3%。完成目标识别。



Figure 13. Detection and identification results of kettle
图 13. 水壶检测识别结果



Figure 14. Detection and identification results of the switch
图 14. 开关检测识别结果

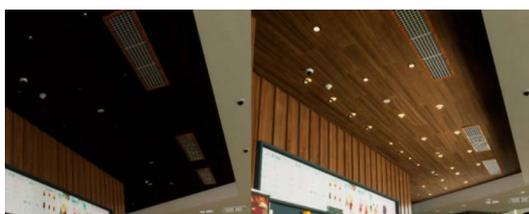


Figure 15. Comparison diagram of lights on and off
图 15. 开灯关灯对比图



Figure 16. Deliver coffee
图 16. 递送咖啡

在对物体进行目标识别之后, 对应的蓝图节点将识别结果返回, 后续的蓝图节点根据返回的结果判断执行何种方案。

在图 15 和图 16 中, 展示的是在进行目标检测识别之后实际任务的执行完成。

7. 总结

在实际运行的测试过程中, 实验测试效果与模型检测效果相差无异, 结果显示, 识别算法在实际运用中抗干扰能力较强, 并且基本可以达到实际需求的准确度。根据机器人运行的蓝图可以实现算法的调用和检测分析。在一定程度上扩展了机器人的可执行性。在仿真环境中的识别应用展现得较为完善, 例如可以识别到水壶, 并在识别完成之后, 实现倒水的动作, 基本达到了预期的结果。

参考文献

- [1] 刘京运. 从 Big Dog 到 Spot Mini: 波士顿动力四足机器人进化史[J]. 机器人产业, 2018(2): 109-116.
- [2] 王珂, 卜祥津, 李瑞峰, 赵立军. 景深约束下的深度强化学习机器人路径规划[J]. 华中科技大学学报(自然科学版), 2018, 46(12): 77-82.
- [3] 董豪, 杨静, 李少波, 王军, 段仲静. 基于深度强化学习的机器人运动控制研究进展[J]. 控制与决策, 2022, 37(2): 278-292.
- [4] 刘晓军, 类成玉, 吴皓, 杨勇. 基于 SOA 的云机器人服务框架设计[J]. 计算机工程与设计, 2021, 42(2): 567-573.
- [5] 李昱华. 番茄采摘机器人目标检测与抓取的关键技术研究[J]. 中国新技术新产品, 2018(23): 55-57.
- [6] 汤伟, 刘思洋, 高涵, 陶倩. 基于视觉的水面垃圾清理机器人目标检测算法[J]. 科学技术与工程, 2019, 19(3): 136-141.
- [7] Jain, S. and Argall, B. (2016) Grasp Detection for Assistive Robotic Manipulation. 2016 *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, 16-21 May 2016, 2015-2021. <https://doi.org/10.1109/ICRA.2016.7487348>
- [8] 孟月. 达闼科技: 向上生长的 2020 年[J]. 通信世界, 2021(3): 26-27.
- [9] Zhu, X.K., Lyu, S.C., Wang, X. and Zhao, Q. (2021) TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, 11-17 October 2021, 2778-2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>
- [10] Tutorials of Ultralytics YOLOv5. GitHub. <https://github.com/ultralytics/yolov5>