

# 实战营第六组-中期汇报

专项工作开展原则：

- 1.业务问题解决导向：**以解决实际业务问题为项目成功的最终标准，专注于深入洞察和有效解决问题，超越单纯完成项目任务。
- 2.质量与长期视角：**追求卓越，不因赶工期而牺牲质量，以长远眼光审视和加强组织的长期能力建设。
- 3.高效执行原则：**践行“三三三”、“24小时”原则，确保决策和行动的高效性和快速响应。
- 4.持续成长心态：**享受工作带来的个人成长，将工作视为一场提升自我、实现价值的“修行”。
- 5.团队协作与沟通：**建立开放、透明的沟通环境，鼓励团队成员之间的协作，共同面对挑战，实现项目目标。

**项目名称：**面向智能零售终端的服务机器人：融合图像识别与条码扫描的自助结算与语音交互式购物引导方案

**核心成果：**实现一款能应用于7FRESH七鲜等超市零售场景，为顾客提供商品结算等服务的机器人。基础功能为通过结合图像识别与扫描条形码两种方法，自动计算顾客选取的商品总价并结算，过程中通过对话来引导顾客或进行其他交互。进阶功能为帮助顾客拿取购物袋并协助商品装袋；提供购物小票，协助开发票或退换货凭证；告知促销信息、办理优惠活动；提供一次性勺、叉、筷、吸管等特定商品所需餐具，等等。

**组长：**王裕硕

**成员：**金叶晨，冼政霖，高玉正，徐文江，朱辰宇

**项目里程碑：**

序号	里程碑	时间	成员	项目进展
1	建立视觉识别与商品检测模块	第2周	徐文江	完成手部识别（Hamer模型）、商品检测（Grounding DINO）、图像编码（CLIP）、商品筛选逻辑初步实现，解决多手干扰和误识别问题，正在升级至DINO V2模型以提高准确性。
2	完成机械臂与手部跟踪系统集成	第2~3周	王裕硕	将方舟R5机械臂与RealSense相机物理集成，实现位姿控制与手部跟踪，解决SDK与IK适配问题，通过URDF重建与PID调参实现动作平滑，加入滤波机制以应对跟踪误差。
3	构建语义理解与多轮问答模块	第2~3周	冼政霖	搭建基于大语言模型的FAQ系统，支持商品寻址、购物建议等功能，解决模糊语义、多轮提问问题，引入上下文跟踪与PE机制，有效提升用户问题识别准确性与交互自然度。
4	语音识别与合成模块开发	第2~3周	金叶晨	完成初版语音识别与语音合成模块，适配超市噪音环境，使用SenseVoice和ChatTTS，处理识别误差和语音延迟问题，引入噪声抑制、情感语音合成和本地轻量部署策略。

5	设计商业化模式与运营推广路径	第2周	朱辰宇	构建“基础租赁+增值服务”商业模型，明确收益来源与成本控制措施；制定三阶段推广计划（试点期、扩张期、生态期），提出模块化维护和远程诊断机制，支撑产品大规模落地。
---	----------------	-----	-----	--

## 项目方案

### 一、概述：

#### 1.1 需求背景、目标及收益

提示：1、选题在行业内有什么特色和先进的地方？2、选题对业务产生什么价值？

项目选题为“智能机器人售货员”，计划在一个月时间内，实现一款能应用于7FRESH七鲜等超市零售场景，为顾客提供商品结算等服务的机器人。基础功能为通过结合图像识别与扫描条形码两种方法，自动计算顾客选取的商品总价并结算，过程中通过对话来引导顾客或进行其他交互。进阶功能为帮助顾客拿取购物袋并协助商品装袋；提供购物小票，协助开发票或退换货凭证；告知促销信息、办理优惠活动；提供一次性勺、叉、筷、吸管等特定商品所需餐具，等等。

### 问题现状

- 1. 现有自助结算的局限：**目前主流的自助结算机虽能完成扫码、支付等基础功能，但因位置固定、无法移动，无法为顾客提供额外的交互式服务（如帮助装袋、推送促销信息等），限制了使用场景和用户体验。

2. **相关机器人系统尚处于研发或试点阶段**: 如红旗连锁已在四川地区筹备机器人无人售货系统，通过与人形机器人企业合作，整合AI视觉识别与智能路径规划等核心技术，但仍处于技术迭代和场景验证阶段，尚未大规模商业落地 [news.qq.com](http://news.qq.com)。
3. **针对餐饮托盘的视觉识别方案难以直接复用**: 法国初创公司Retail Robotics Solutions推出的Robot-Cashier能在托盘上直接识别食物，省去条码，但该方案依赖相对规则的餐盘场景，不适用于货架、购物车中多样、密集商品的超市零售环境 [rrs-catering.eu](http://rrs-catering.eu)。

## 选题意义

1. **大幅降低人力与维护成本**: 移动机械臂+自主导航使机器人可24小时无人值守，减少现场人工及维护开销，提升运营效率与投资回报 [taobotics.com](http://taobotics.com)。
2. **满足零售行业技术创新需求**: 面对传统消费增长乏力，机器人售货员被视为行业破局之道，可更快速地处理大量商品信息与顾客需求，为消费者提供高效、便捷、个性化的购物体验 [news.qq.com](http://news.qq.com)。
3. **机械臂末端易于功能扩展**: 在实现基础扫码结算的同时，可拓展为协助装袋、递送购物袋、打印发票/小票、推送促销及提供一次性餐具等多种增值服务，极大丰富零售业态。

## 1.2 竞品分析（选填）

提示：分析市场上已有竞争产品，同时分析所开发产品的优势、劣势。

### 市场竞品

#### 1. Trigo (Tesco “GetGo” ) — 视结合视觉与传感器的无收银台系统

Trigo 使用计算机视觉构建“just-walk-out”购物体验，已在伦敦Tesco试点店部署 [WIRED+6维基百科+6卫报+6](#)。其优势在于无需扫码，购物通畅，但高成本且依赖固定布局，不具备移动和实境交互功能。

## 2. 法国Retail Robotics Solutions “Robot-Cashier”

实现在托盘上识别食物，跳过条码扫描，但依赖高规整托盘场景，难以适配超市环境  
arXiv+15StrongPoint+15Retail Systems+15 Retail Customer Experience。其方法专一，通用性和鲁棒性不足。

## 3. Sweet Robo 在 Food Bazaar 的互动售卖机

为特色体验型自动售卖市场提供视觉识别和互动服务，提升购物乐趣The Sun+3新闻频道内布拉斯加州东北部+3Retail Systems+3。但不支持移动，也无法处理多商品、结算或袋装服务的复杂场景，属轻体验型产品。

## 本项目优势

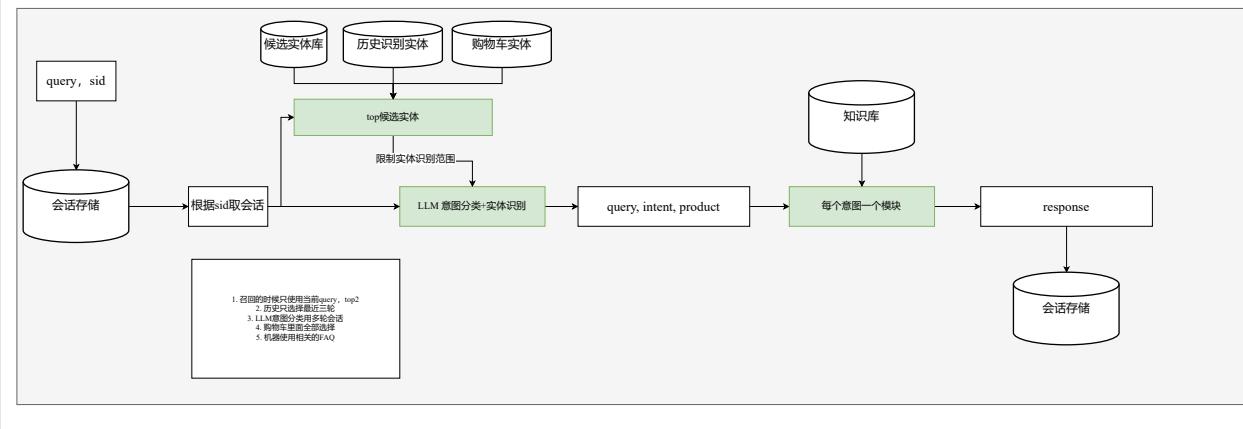
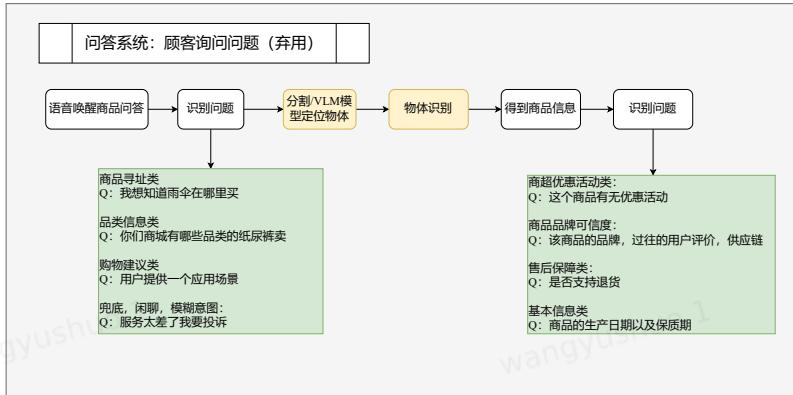
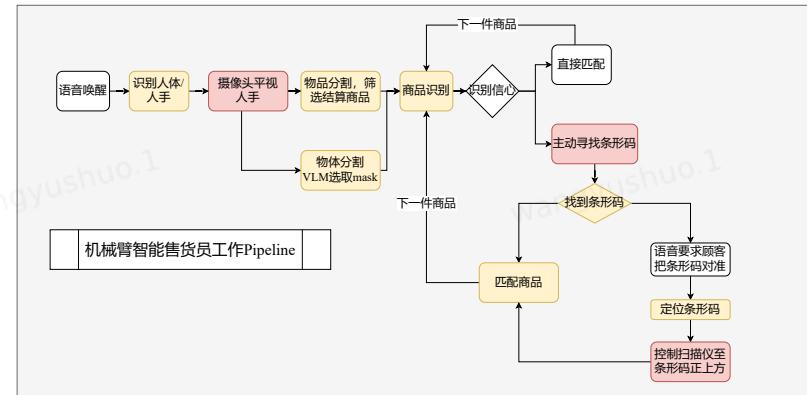
- 移动型机械臂 + 自主导航：**相较于固定式系统，机器人可主动走向顾客，提供结算引导、装袋、打印等服务，显著提升交互覆盖度与运营灵活性。
- 扫码 + 图像识别双模态为基础结算方式：**在超市环境中适配不规则商品布局，可切换识别方式，提升识别覆盖率与准确率。
- 功能可扩展性强：**内置机械臂末端接具，可集成装袋、递送发票、小票打印、餐具发放等服务，不只是单一结算。
- 行业成熟性与经济可行性：**目标锁定7FRESH等大型零售场景，在AI服务与成本优化前提下，具备真实可量产潜力；与当前试点案例相比，将更具落地性。

## 1.3 用户分析

用户类型	核心诉求	满意标准
日常购物者	快速、准确扫码及图像识别；人性化结算流程；简单交互体验	少于30秒完成结算；互动自然；精准识别99%以上商品
行动不便/老年者	机器人引导、装袋辅助与打印凭证等贴心服务	全程无需繁琐操作，解决物理隔阂；凭证实时可得
科技尝鲜用户	机器人主动接待、提供促销提醒与互动乐趣	有趣交互、推荐贴合个人消费标签；稳定体验
零售商	降本增效；覆盖峰值期客流；减少人工失误；品牌传播效应	投资回报率提升；运营平稳；顾客与媒体正面反馈显著
维护人员	稳定性、可维护性、安全性；运行异常可预测	故障率低；监控界面明确；异常风险可及时拆解与处理
财务/稽核部门	数据合法、可追溯；支付渠道多样；凭证（发票、小票等）符合合规性要求	数据安全存储；结算通道符合标准；凭证打印准确可用

## 二、项目方案

## 流程图



## 三、项目进展

### 机械臂控制模块 - 王裕硕

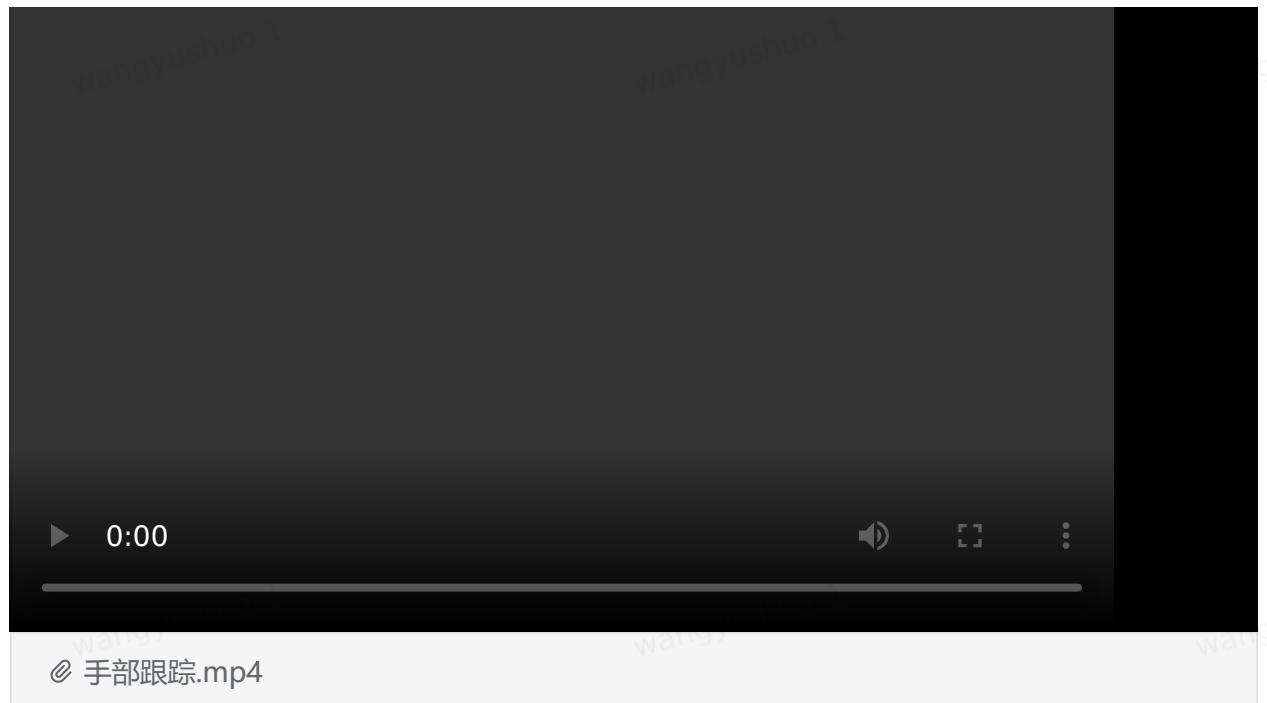
**已完成的任务：**方舟无限R5机械臂与RealSense相机物理集成，位姿或关节转角控制，手部识别与跟踪

**遇到问题：**

- 1，原始提供的控制SDK未针对R5型号进行充分适配，使用过程中发现逆运动学（IK）求解结果不稳定，导致机械臂部分关节出现异常振动，影响控制精度与安全性。
- 2，在实际测试中，Mediapipe在光照变化或遮挡情况下容易出现手部关键点丢失，进而导致系统对手部位置估计突然跳变，机械臂表现出剧烈移动或错误响应，影响整体稳定性与使用体验。

**解决办法：**

- 1，参考R5机型结构，重新编写URDF文件，确保模型关节定义、连杆长度等参数与实际硬件一致，以提升运动学解算的准确性。对控制器中的PID参数进行实验性调优，优化每个关节的响应速度与稳定性，抑制电机振动问题，提升轨迹执行的平滑度。
- 2，滤波以提高运行稳定性，引入时间序列滤波器对手部关键点数据进行处理，有效降低由短时识别误差引发的位姿突变。



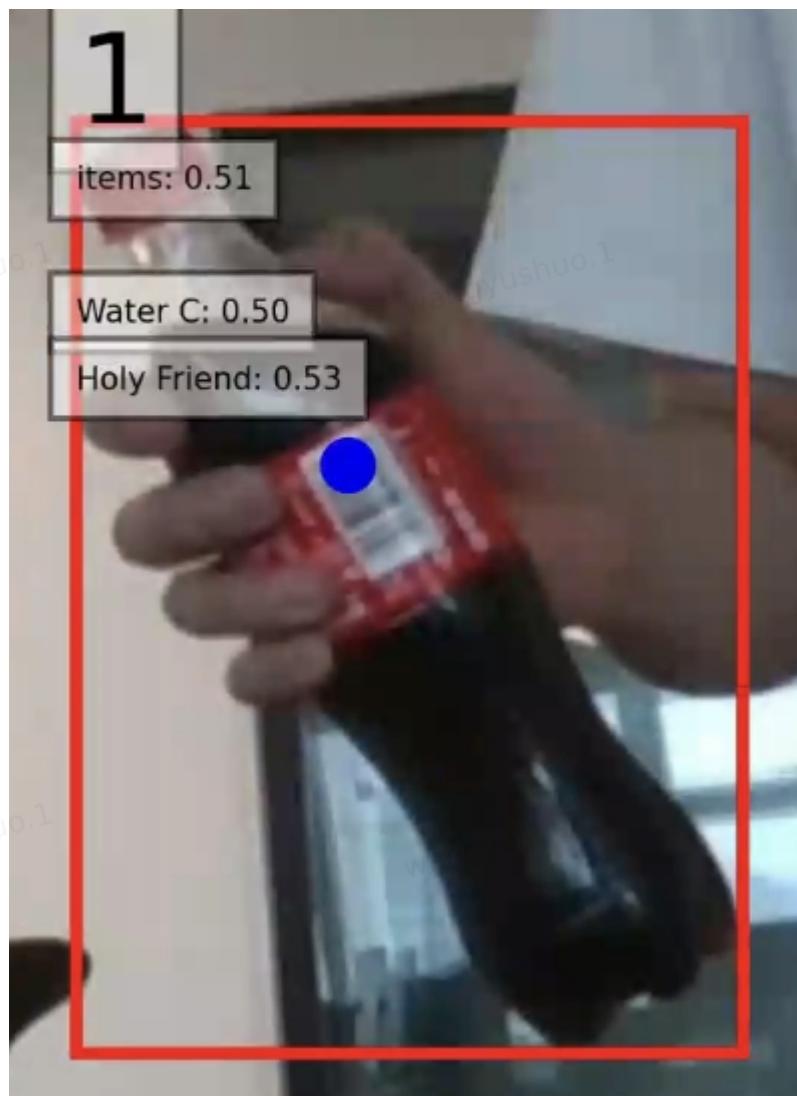
# 视觉识别与扫描模块 - 徐文江

已完成的任务：

1. 实现人体手部识别功能 (使用Hamer模型)



2. 商品检测功能 (使用Grounding dino模型)



3. 人手和商品筛选功能 (几何方法)

4. 商品品类识别功能 (使用Clip图像编码器)

```
[[ 0.13818905  0.21053804 -0.1792292 ...  0.75840706 -0.17232406  
-0.04856042]  
[-0.4172747   0.19332348  0.16669334 ...  0.37086672 -0.0251078  
-0.10604114]  
[-0.03615847 -0.12036143 -0.00796025 ...  0.34287548 -0.22884487  
-0.16591227]]
```

编码成512维嵌入向量作为商品特征

遇到问题：

1. 画面中不止一个人手



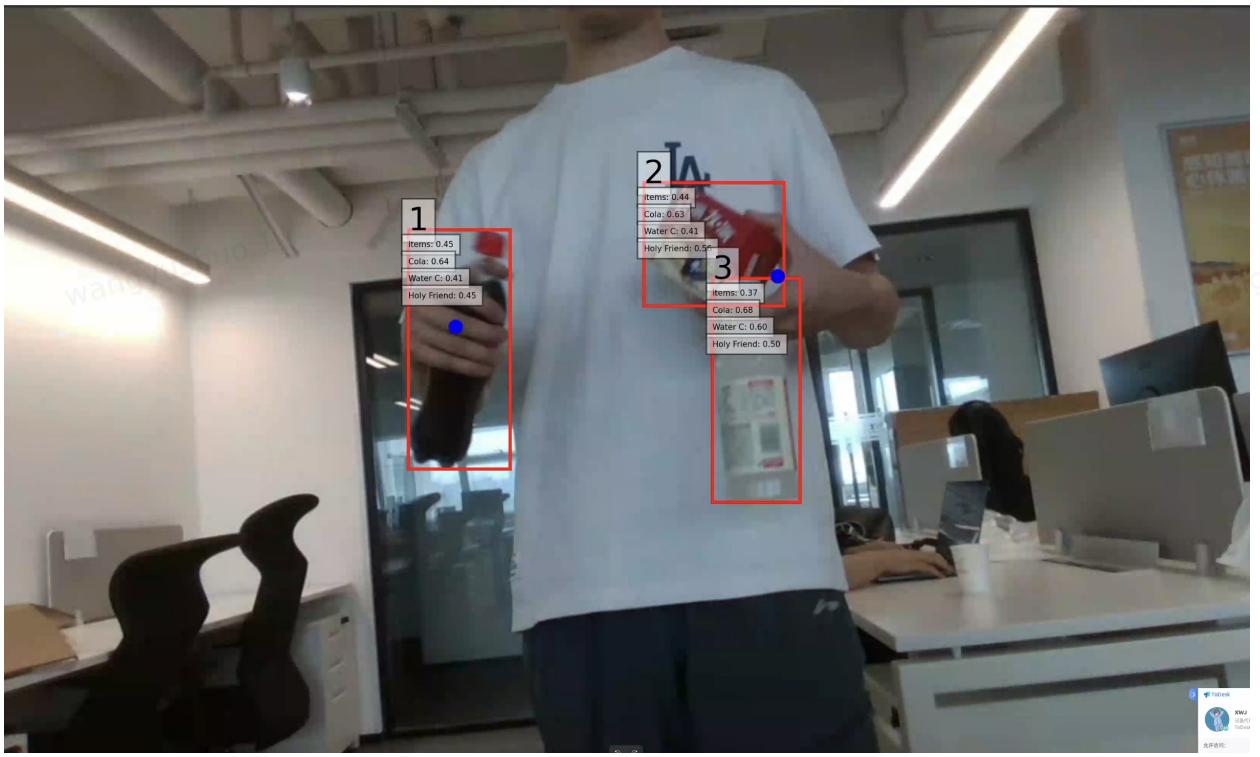
如示例图，模型检测出三只人手

## 2. 场景中非手持商品也被检测到



如示例图，模型检测出场景中无关物品

## 3. 商品特征匹配失败



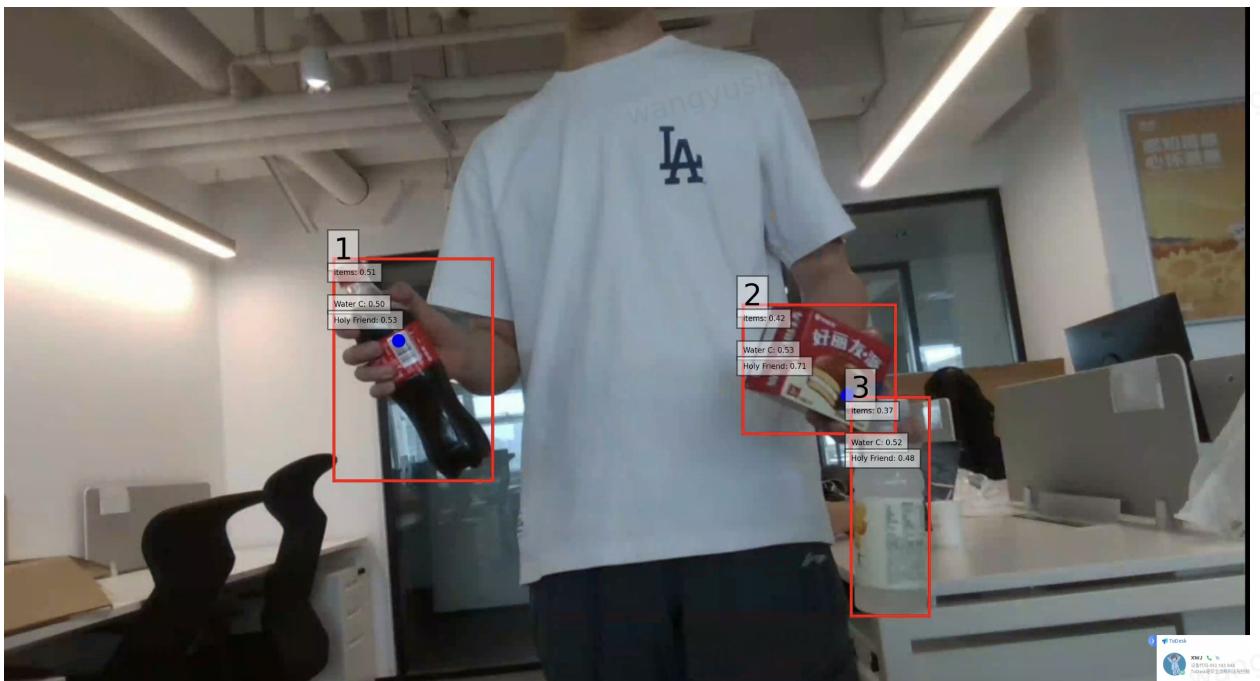
如图，水溶C被误判为可乐

解决办法：

1. 人手根据大小筛选box最大的两只



2. 筛选出人手box与商品box存在重合的商品作为手持商品



3. 选择其他更强的视觉编码器Dino V2进行特征提取 (进行中)

智能对话模块 - 洗政霖

已完成的任务：已完成商品寻址类，品类信息类，购物建议类的FAQ模块开发

遇到问题：我们面对的是一个线下零售FAQ场景。线下零售FAQ场景的挑战在于

1. 一个会话中询问的用户可能有多个，询问的问题可能有多个，询问的商品也可能有多少个
2. 用户表达他想询问的商品时可能比较模糊，比如把商品名表达成日常用语 黄元帅表达成黄色的苹果。

解决办法：

1. 对多轮会话做好存储管理输入给LLM，利用LLM强大的上下文理解能力区分不同问题之间的边界
2. 在候选实体中用BM25召回top5+历史识别的实体作为补充信息辅助大模型实现更准确的实体识别
3. 做好PE确保大模型避免幻觉识别到正确的实体和意图，如果识别模糊引导用户进一步明确
4. 丰富完善知识库，提供相应的知识让大模型应答

示例1：引入上下文帮助模型理解多轮对话中用户的不同问题

```
● (xz1) PS D:\xianzhenglin.1\Desktop\实战营FAQ> python .\FAQ_main.py "苹果在哪里我没有找到" JD1404
处理查询：苹果在哪里我没有找到
会话ID：JD1404
多轮会话读取：['苹果在哪里我没有找到']
大模型响应：{"intent": 1, "product": "苹果"}
大模型响应：{"Answer":"您好，苹果位于进门左手边的果蔬区，您可以在那里找到。"}
● (xz1) PS D:\xianzhenglin.1\Desktop\实战营FAQ> python .\FAQ_main.py "都有哪些品种？" JD1404
处理查询：都有哪些品种？
会话ID：JD1404
多轮会话读取：['苹果在哪里我没有找到', '都有哪些品种？']
大模型响应：{"intent": 2, "product": "苹果"}
大模型响应：{"Answer":"我们这里有多种苹果品种，包括红富士、黄元帅和青苹果。您可以根据自己的口味选择喜欢的品种。"}
● (xz1) PS D:\xianzhenglin.1\Desktop\实战营FAQ> python .\FAQ_main.py "我想给孩子做沙拉推荐哪一种" JD1404
处理查询：我想给孩子做沙拉推荐哪一种
会话ID：JD1404
多轮会话读取：['苹果在哪里我没有找到', '都有哪些品种？', '我想给孩子做沙拉推荐哪一种']
大模型响应：{"intent": 3, "product": "苹果"}
大模型响应：{"Answer":"如果您想给孩子做沙拉，推荐选择青苹果。青苹果酸脆清爽，硬度高，非常适合用于健康轻食和沙拉制作。"}
○ (xz1) PS D:\xianzhenglin.1\Desktop\实战营FAQ>
```

示例2：正确应答用户的模糊问题

```
(xz1) PS D:\xianzhenglin.1\Desktop\实战营FAQ> python .\FAQ_main.py "你们这里有卖黄色的苹果吗" JD1360
处理查询：你们这里有卖黄色的苹果吗
会话ID：JD1360
多轮会话读取：['你们这里有卖黄色的苹果吗']
大模型响应：{"intent": 2, "product": "苹果"}
大模型响应：{"Answer":"是的，我们这里有黄色的苹果，品种叫黄元帅苹果。如果您需要购买，可以到水果区的苹果专区找到。"}
```

## 语音识别与合成模块 - 金叶晨

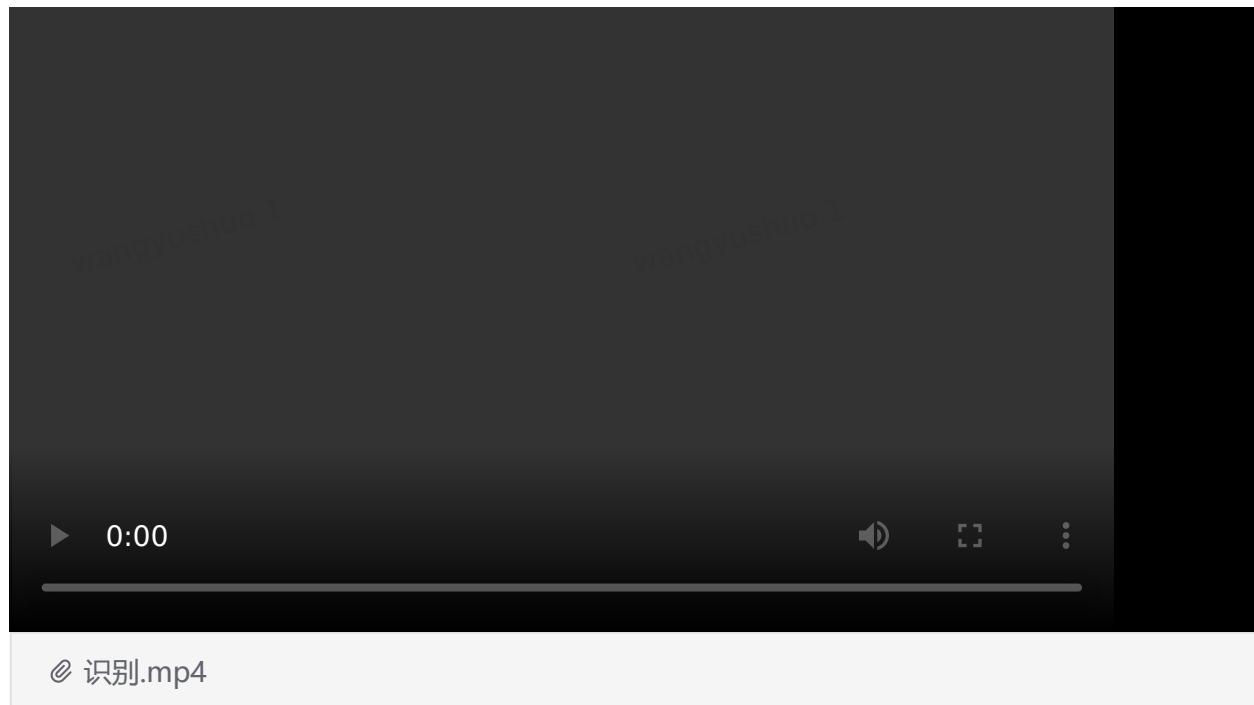
已完成的任务：

遇到问题：超市环境嘈杂（背景音乐、人声、货架搬运声等），导致语音识别准确率下降；顾客普通话不标准或使用方言，导致识别错误；合成语音机械感强，影响用户体验；语音生成延迟导致对话卡顿。

解决办法：预训练噪声抑制模型；收集方言数据，微调模型；使用情感化TTS引擎；本地部署轻量级TTS模型

语音识别：

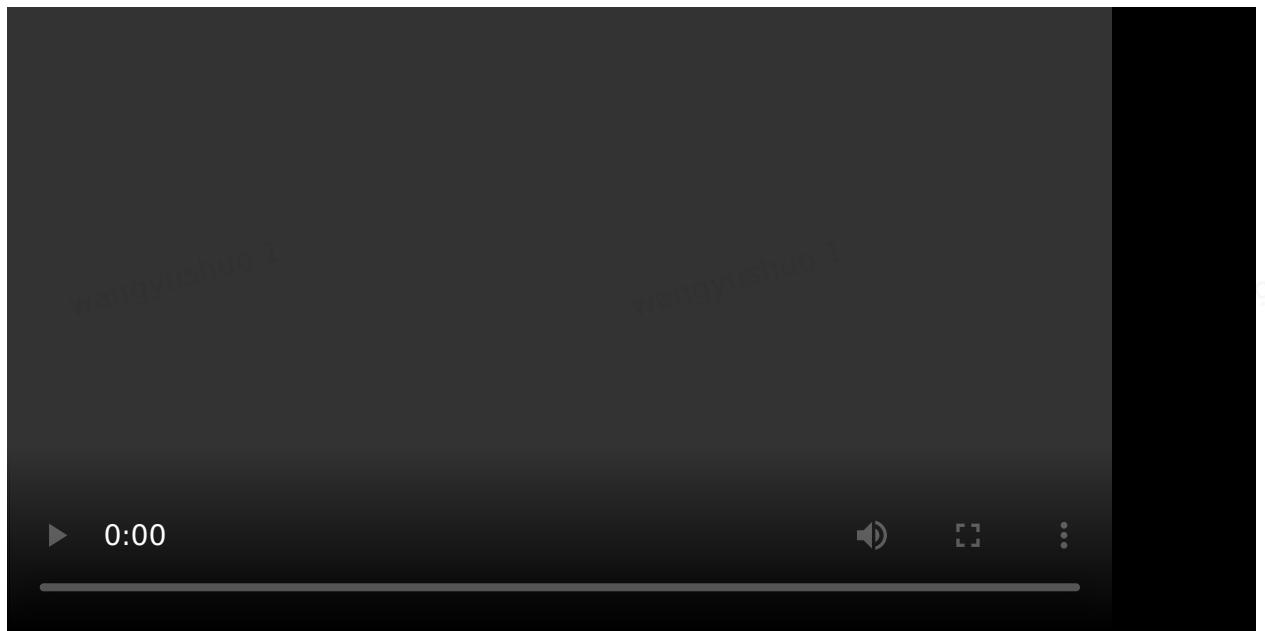
SenseVoice



<|zh|><|NEUTRAL|><|Speech|><|within|>他刚买了台新电脑，比之前的快多了。

语音合成：

ChatTTS



## 产品商业化运营方案 - 朱辰宇

已完成的任务：

收益：

模式	说明	收益来源
<b>硬件租赁</b>	向商超按月出租机器人（含维护）	租金（单台¥8,000-12,000/月）
<b>增值服务</b>	按功能模块收费： - 促销信息推送（品牌方付费） - 电子发票系统接入	广告分成/技术服务费

成本控制：

成本控制策略

## 1. 硬件成本优化:

- 机械臂成本控制
- 复用商超现有监控摄像头辅助定位，减少激光雷达数量。

## 2. 运维降本:

- 云端远程诊断系统：故障通过OTA更新/参数调整解决。
- 模块化设计：扫描仪故障时可快速插拔更换 (<5分钟)。

落地推广：

阶段	行动方案	KPI
试点期 (1-3个月)	在10家7FRESH门店部署3台： - 提供免费试用期 - 收集顾客动线数据优化路径规划	单台日均服务≥50单，故障率<3%
扩张期 (4-6个月)	在全部7FRESH门店部署	单台日均服务≥100单，故障率<1%
生态期 (1年+)	签约其他连锁商超，采用“基础租金+流水分成”模式	市占率覆盖目标区域30%

## 四、下一步计划

### 1. 模块调试与稳定性优化

- 在模拟真实超市场景中对各个模块（机械臂控制、图像识别、语音交互等）进行全面测试和联调，重点提升系统在嘈杂环境、多用户交互等情况下的响应稳定性与错误恢复能力。

- 引入错误注入机制，测试系统在关键部件（如图像识别失败、语音无法识别）异常下的鲁棒性。

## 2. 扩大商品图像数据集，提升识别精度

- 通过实地采集超市中常见商品图像，并针对遮挡、角度变化、光照变化等因素进行多样化增强，训练DINO v2等更强的视觉编码器模型。
- 构建“商品知识图谱”，辅助模型识别模糊品类（如“黄色的苹果”=黄元帅），提升语义理解能力。

## 3. 完成系统整合，打通工作流 pipeline

- 以用户动作为主线，联通“检测手部 → 识别商品 → 语音交互 → 控制机械臂”的完整流程，实现端到端自动结算体验。
- 搭建统一状态管理与调度模块，协调各子系统资源分配与并发处理，提升系统整体效率与响应速度。

## 4. 挑战拓展功能，丰富服务能力

- 尝试实现机器人辅助装袋功能，初步设计装袋路径规划与抓取策略，验证机器人能否在商品密集、袋口变形情况下成功完成动作。
- 打通打印发票、发放一次性餐具等后续功能的软硬件接口，初步测试商业化所需的增值服务链路。

# 五、总结和反思

## 1. 跨学科协作的挑战与成长

项目涉及计算机视觉、语音交互、机械控制、商业策略等多个技术与业务模块，组员们在合作过程中需要持续学习、相互协调，提升了对跨学科项目的理解与执行能力。虽然起初接口调试与任务分工略显混乱，但通过定期同步与明确负责人，逐步构建起有效的协作机制。

## 2. 从“功能完成”到“体验导向”的思维转变

我们认识到，项目成功不仅是模块能“跑起来”，更重要的是“用户能用好”。因此团队在后期更注重交互流畅度、容错机制与边缘案例处理能力，逐步向场景真实可用迈进。

### **3. 技术选型需平衡“前沿性”与“工程落地”**

初期曾尝试多个模型与算法，如Mediapipe、Clip、GroundingDINO等，发现部分算法虽具学术领先性，但在边缘设备上部署难度大。未来需进一步评估“性能-资源-稳定性”三者间的权衡，找到最适合商业化的技术路径。

### **4. 团队执行节奏和反馈机制仍有提升空间**

虽已贯彻“三三三”与“24小时原则”，但仍存在部分子任务进度滞后、问题反馈不及时的情况。接下来将进一步加强过程追踪与任务透明化，建立更精细化的推进机制。