**睿尔曼新零售机器人落地方案**

**技术方案（v1.0）**

**目录**

[睿尔曼新零售机器人落地方案 1](#_Toc28712)

[技术方案（v1.0） 1](#_Toc21806)

1. **整体方案技术架构**

本项目旨在开发一款高度自动化的智能机器人系统，用于超市环境中的补货和出货任务。该系统将集成先进的感知传感器和机械臂技术，以实现高效、准确的物品处理和库存管理。

* 1. **项目目标**

1. 自动化补货：根据所给指令将补货篮中的物体补货到货架上，通过机器人自带的感知传感器，规划行动路线，精简补货流程形成自动化。
2. 智能出货：根据顾客所需购买产品，自动规划到达货架的路线，通过传感器感知环境并从货架上取得商品并放置于商品篮中。
   1. **产品形态**

产品由四个重要硬件机械部分组成，分别是：用于操作的7自由度升降双臂系统；用于夹取或者吸取的末端执行工具；用于感知环境的单相机或多相机系统；用于移动的地盘基座。



图1-1机器人产品形态概念图

其中：

1. 双臂升降系统用于使得6自由度机械臂能够处理不同高度的货架商品，扩大机器人的工作空间，更好的避免双臂系统之间的操作影响。可以选用力纳克的LC3 IC：



图1-2：力纳克DL6

LC3 IC 电动升降柱配有内置控制器，便于集成到工业应用中。LC3 IC 可以 100 mm/sec 的速度推动或拉动 1000 N 的负载。速度可调，不受负载影响。

以及SKF FRE系列升降柱



图1-3：SKF FRE系列升降柱

SKF FRE系列可以 10-15mm/sec 的速度推动或拉动 2000 N 的负载。动态偏载250Nm。

1. 吸取或夹取的末端执行工具，可以采用市面技术上较为成熟的二指夹爪和独立吸盘。二指夹爪可选用Robotiq 2F-85:



图1-4：Robotiq 2F-85

Robotiq 2F-85夹爪最大行程为85mm,夹力为20-235N，抓握荷载5KG，位置分辨精度可达0.4mm。

以及Onrobot RG6：



图1-5：Onrobot RG6

Onrobot RG2夹爪最大行程为160mm，夹力25-120N，抓握荷载6KG.位置分辨率0.1mm。

吸盘可选择Robotiq E Pick吸盘：



图1-6：Robotiq E Pick

吸盘采用电动式，最大负载10KG，真空度80%。

以及范德华力吸盘-VFC:



图1-7：范德华力吸盘-VFC

吸盘采用电动式，最大有效荷载1KG。

1. 机器人视觉系统：机器人需要相机组成的视觉系统来感知周围的环境，在经过手眼标定后，机械臂基座与相机之间的转换关系确定，才能让机器人具有感知周围环境的能力。在头部安装相机可以让机器人获取更大范围的视角，相机可以选用Gemini336L：



图1-7：Gemini336L

Rgb模式支持1280x800@60fps，深度模式支持1280x800@30fps，推荐工作区间0.25-6.0m。

以及realsenseD435i：



图1-8：RealsenseD435i

Rgb模式支持1920x1080@30fps，深度模式支持1280x720@30fps，推荐工作区间0.3-10.0m。

1. 移动底盘基座（黄泊凯）  
   描述一下底盘上放置货物篮

1

2

**3**

* 1. **产品工作形式**

1. 商品管理系统

本项目的商品管理系统是用于管理商店的商品详细信息、库存容量、商品定价等业务流程的软件系统，主要功能包括：

1. 商品信息管理：实时管理记录商品类目、所在货架、入库时间、保质期等信息，并能完成添加、修改和删除指令。
2. 库存管理：实时监控库存水平，自动计算补货需求，确保库存充足，在发起补货指令后，指引机器人前往补货平台，完成补货。
3. 订单管理：处理客户发出的订单指令，指引机器人前往货架处，夹取指定货物于货物篮中。
4. 促销等其他管理：创建和管理促销活动、打折等其他操作。
5. 补货

当机器人受到来自商品管理系统的补货指令后，将前往理货台。理货台上放置需要配送到货架上的商品，机器人根据头部相机的视觉系统以及理货台上的重量传感器联合识别当前货架上是否存在货物，当识别到物品分类后，机器人会使用合适的执行工具夹取或吸取商品，并移动到货架放置商品。

理货指令

机器人移动到理货台

视觉系统和重量传感器联合识别

分拣商品

移动并放置商品到对应货架

图1-9：机器人理货流程

1. 取货

当机器人收到取货指令时，根据管理系统内部给定的商品货架坐标移动到指定点位，通过视觉感知夹取商品，放置于货篮中并带回。

移动到指令位置

取货指令

识别商品

分拣商品

图1-10:机器人取货流程

1. 机器人调度相关（黄泊凯）
2. **国内外研究现状**
   1. **根据RGB图像的目标检测网络分类现状（周俊桦）**
   2. **6-DoF位姿估计现状（林伟鸿）**

六自由度（6-DoF）目标姿态估计在各种应用中具有重要意义，这些应用包括增强现实、机器人操作和自动驾驶等。由于光照和遮挡的变化，不同姿态的对象在不同视角下的外观可能会有很大差异，这给从 单一彩色图像中估计6-DoF对象姿态带来了巨大挑战。

基于学习的早期阶段方法通常一次性估计6-DoF姿态。Yu等人提出使用卷积神经网络直接回归对象中心和距离以及旋转分量。近期的工作则提出首先估计观测图像与对象模型之间的2D-3D对应关系，然后使用PnP算法求解对象姿态。这些一次性方法的性能通常受到网络容量的限制，并且容易受到不良条件（如光照变化和遮挡）的影响。PVNet基于关键点的预测，在点云质量良好的情况下在LINEMOD数据集上可以左右99%以上的精度。

随着深度学习的普及，为了避免昂贵的深度传感器，部分算法使用纯粹RGB预测位姿估计的方法，代表的方法有YOLOV5-6D，利用yolo网络的主题框架，引入空间注意力，相比于其他的RNN算法，在推理时常和预测准确度中做到了不错的平衡。 SAM-6D通过两个步骤来实现零样本 6D 物体姿态估计，包括实例分割和姿态估计，给定任意目标物体，SAM-6D 利用两个专用子网络，即实例分割模型（ISM）和姿态估计模型（PEM），来从 RGB-D 场景图像中实现目标；其中，ISM 将 SAM 作为一个优秀的起点，结合精心设计的物体匹配分数来实现对任意物体的实例分割，PEM 通过局部到局部的两阶段点集匹配过程来解决物体姿态问题。表现最出色的方法通常包含姿态细化过程，这一步骤极大提升了性能。FoundationPose 为减少大规模训练的人工工作，利用LLM和Diffusion开发了一种新的合成数据生成Pipeline。为了弥补无模型和基于模型的设置之间的差距，FoundationPose利用以对象为中心的神经场来进行随后的渲染和新视图RGBD渲染。通过两层网络，粗位姿估计和姿态细化，结果表明相比直接预测物体位姿大大提高了准确度，降低了模型参数量。

* 1. **抓取姿态估计现状（施翰煜）**
  2. **机器人根据目标避障导航移动现状（黄泊凯）**

1. **项目关键研究以及拟解决的关键问题**
   1. **关键研究内容（林伟鸿）**

本项目是基于视觉驱动的机器人超市场景，结合人工智能实现机器人在无人超市中完成取货补货流程。整体研究内容架构如下图所示：

实际作业场景验证

应用验证

* 精确目标检测识别
* 目标实例分割
* 基于RGB-D的位姿估计
* 基于RGB-D的抓取估计
* 空间路径规划

复杂商品目标检测与实例分割技术

实时高精度物体位姿估计技术

物体稳定抓取点估计技术

机器人移动避障以及运动规划技术

4项关键技术

自动取货送货机器人具身智能技术

数据采集困难，复杂商品准确分类和实例分割困难

不同商品的抓取点估计不同，且抓取后需要稳定姿态放置

机器人移动与路径碰撞检测困难

关键难点

场景提炼

图 3.1 项目总体研究框架

* 1. **拟解决关键问题：**

项目关键解决问题主要从实时场景下物体目标检测识别技术、高精度的物体实时位姿估计及位姿跟踪技术、常见物体的实时抓取估计技术、复杂环境中机器人作业运动轨迹规划技术等方面展开，具体如下：

* + 1. **实时场景下物体目标检测识别技术（周俊桦）**
    2. **高精度的物体实时位姿估计及位姿跟踪技术（林伟鸿）**

机器人抓取需要3D相机告诉机器人当前的物体位置，并由算法给出物体的抓取位置，目前常用的解决方法一方面是计算物体的6-Dof位姿，并根据抓取物体单独设定抓取点，完成一次单独的抓取动作。但 由于本项目涉及的商品数量多且商品种类未来可能增加，单纯依靠人力设定抓取点不具备很好的泛化性。另一方面是直接根据获取的点云针对单个点云单独生成抓取估计，在限定条件的情况下，算法可以快速给出末端执行器的抓取点位估计，具有良好的泛化性。但是在本项目中，抓取后需要考虑的一大问题是如何将商品稳定放置到指定地点。综合以上两种方法，结合两种网络是一个可预见的解决方案，在预测抓取点的同时，预测物体的当前位姿并进行位姿跟踪。根据项目需要，需要一个兼顾实时性和稳定性的方法，如何在复杂物体中，有效利用有效利用高精度的物体实时位姿估计及位姿跟踪技术对于机器人抓取来说至关重要。这不仅涉及到对物体的精确识别和定位，还需要实时更新物体的位姿信息。考虑到这些问题，采用相对轻量级的网络来保证位姿的估计方法的实时性，融合物体的先验mask、深度等多种信息，构造形状先验网络，可有效克服物体种类多的影响，使物体位姿跟踪更具有鲁棒性。

* + 1. **常见物体的实时抓取估计技术（施翰煜）**
    2. **复杂环境中机器人作业运动轨迹规划技术（黄泊凯）**

1. **拟采取的研究解决方案**
   1. **总体技术架构（林伟鸿）**

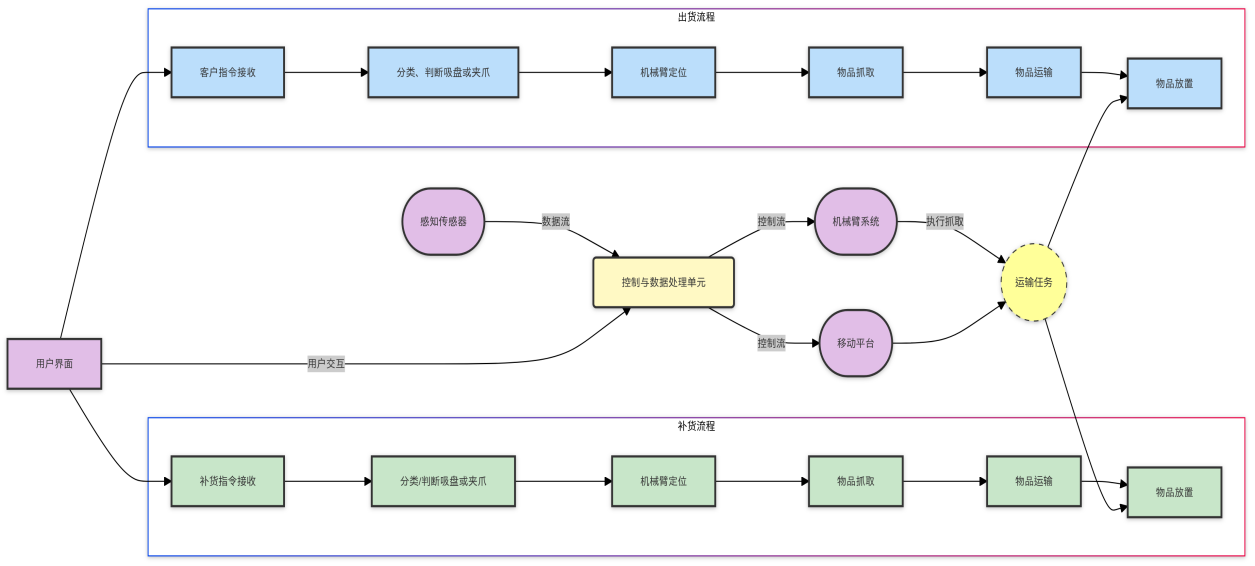
****

图4-1：整体架构图

* + 1. **实时场景下物体目标检测识别技术（周俊桦）**
    2. **高精度的物体实时位姿估计及位姿跟踪技术（林伟鸿）**

在分类物体后，需要选定合适的末端执行器工具，吸取或夹取商品，同时对商品进行姿态估计以及跟踪，以便感知商品实时姿态以及确定后续放置姿态。

在深度学习时代之前，许多基于手工制作特征的方法，比如SIFT、FPFH、VFH和点对特征（PPF），被设计用于对象姿态估计。然而，当面对复杂场景时，这些方法在准确性和鲁棒性方面存在缺陷。这些传统方法现在已经被利用深度神经网络学习高维特征表示的数据驱动深度学习方法所取代，从而提高了准确性和对复杂环境的鲁棒性。在本项目中，需要考虑到使用轻量级的网络做到兼顾实时性和精确性。故选用FoundationPose作为基础网络，FoundationPose借助于两层网络估计物体位姿，而不是单层网络直接回归物体的位姿，大大减少了网络层参数，Bop是专门用于物体位姿估计排序的榜单，为位姿估计领域做出了极大的贡献，FoundationPose在精度上截至24年3月一直处于榜首位置。

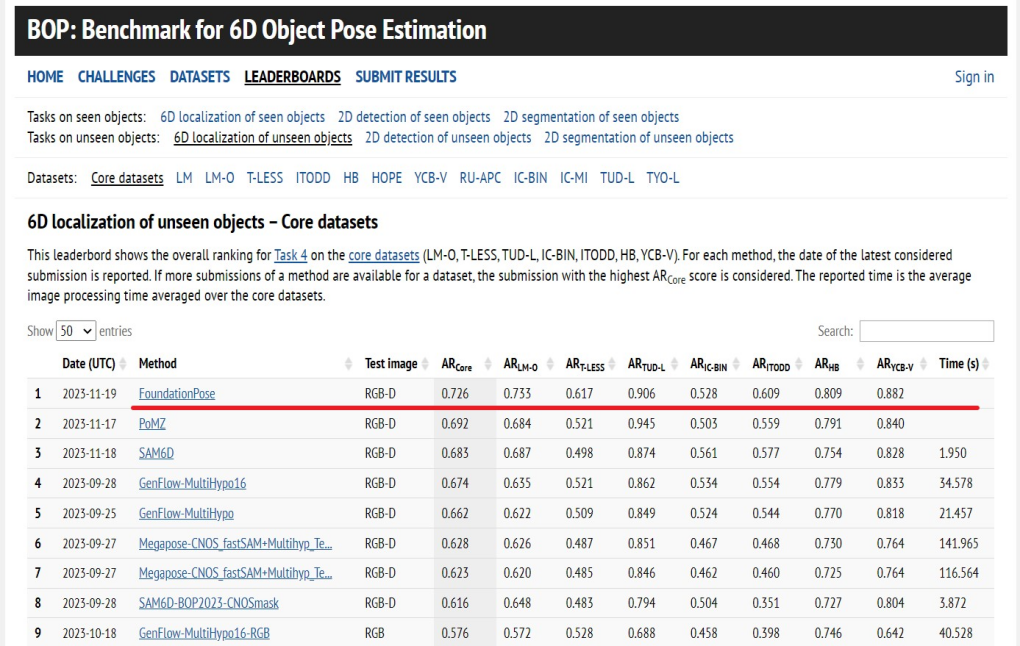


图4-2：Bop排名

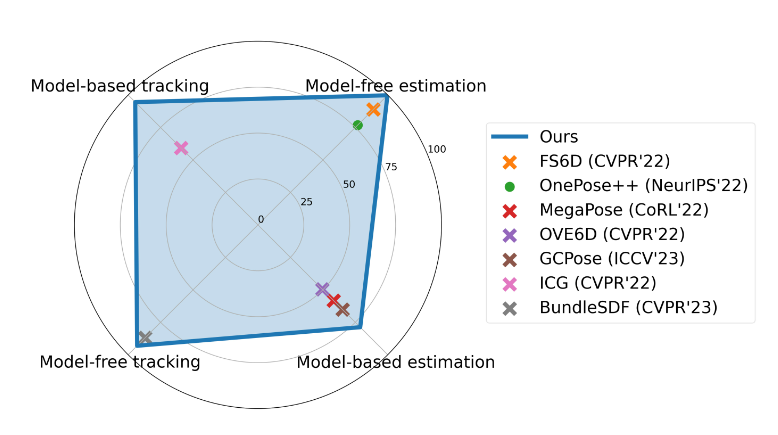


图4-3：Foundation与不同算法的对比

为了使模型具有良好的泛化性，使其对未见过的新模型快速进行姿态估计，需要大量的数据集，但真实数据集在现实世界中的采集十分耗费人力与物理。

1. 基于LLM与Diffusion结合的大规模数据集合成

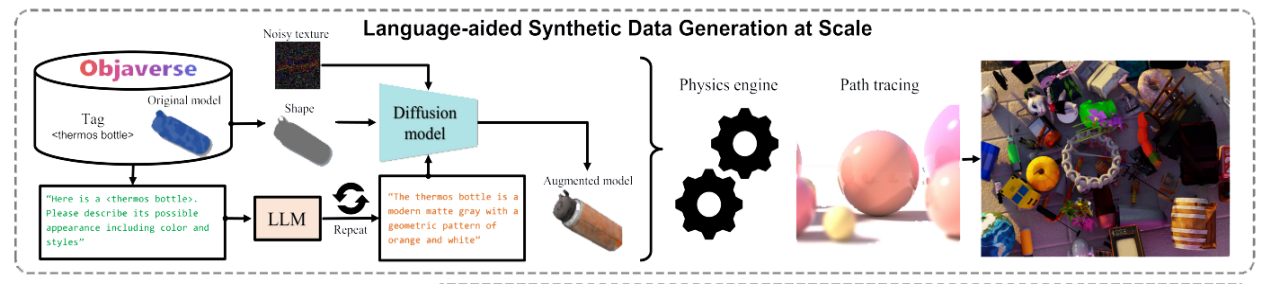


图4-4：数据集生成PipeLine

得益于LLM与Diffusion以及真实物理引擎的快速发展，仿真环境下可以采集到高质量的数据。如上图所示，首先在原有的3d模型上获取没有纹理的基础模型，并通过LLM固定句式的描述生成不同纹理描述，添加一定噪声纹理后交由Diffusion模型生成带有不同纹理的3d模型，通过这种数据增强的方式，可以迅速扩展原有的模型数量；新生成的带有纹理的模型在真实物理引擎的渲染下，可以随机生成在预设的场景环境中，并通过路径追踪取得物体的位姿。经过大量数据集训练后，模型的鲁棒性会进一步提升，并能让模型通过微调适应未来新的商品分类。

有别于直接回归生成的方法，FoundationPose采取两步位姿估计，先假定一个初始位姿，根据位姿 细化网络网络迭代更新这些初始位姿，最后根据位姿排名网络输出一个位姿分数，巧妙的设计绕过直接回归生成方法所带来的非全局最优解的问题，并且算法参数量和稳定性大大增加。

1. 粗位姿估计的生成和位姿微调

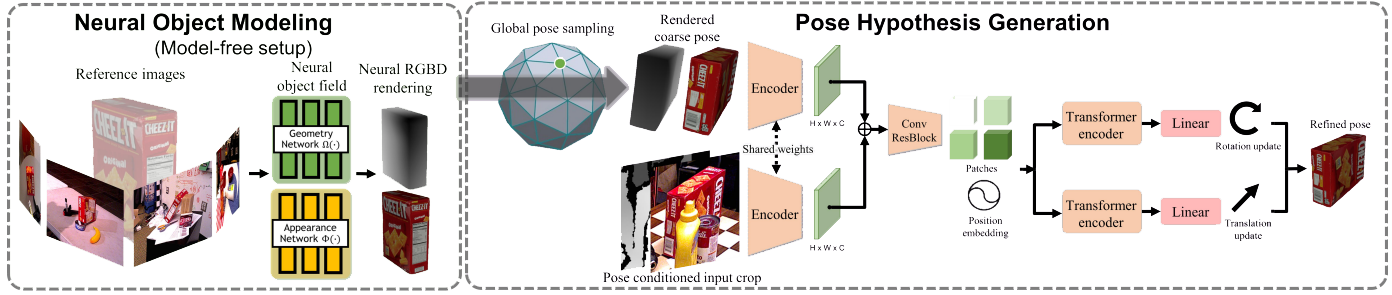


图4-5：物体粗位姿生成以及位姿细化

1. 姿态初始化

给定输入一张rgb图片，物体的二进制mask编码，物体模型，深度图，相机内参。如图3.2.4所示，预先定义Global pose sampling，以从不同视角获取渲染图片并作为初始化旋转变换R。同时根据给定mask和深度，计算场景中物体mask框内的平均深度作为初始化平移变化t。这样的粗姿态获取设计避免了类似其他算法繁杂的位姿初始化过程，预先定义好的视角可以很好适配不同情况且减少网络负担。

1. 位姿调整

姿态初始化后的姿态与实际姿态之间有很大的差距，通过创建一个姿态细化网络，将粗姿态对应的渲染图和来自给定的rgb图片的裁剪作为输入，该网络输出一个姿态更新，包括旋转和平移更新量。如图所示，先使用共享的CNN编码器从两个rgb输入分支提取特征，特征图连接之后并嵌入位置信息用以标记，通过transformer编码器模块处理后，再通过mpl层单独输出旋转更新IMG_256和平移更新IMG_257

IMG_256

IMG_257

网络使用L2监督损失 IMG_258，其中 IMG_259表示权重，IMG_260表示ground truth。

1. 姿态选择

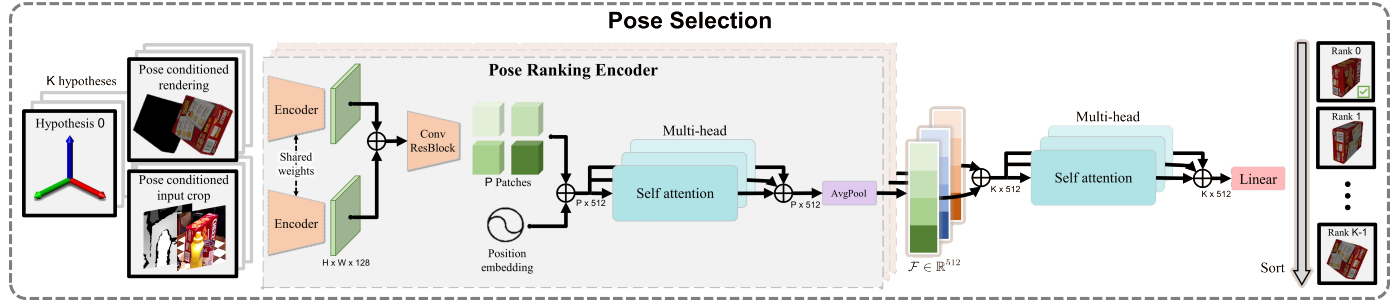


图5-6：位姿选择

完成上述姿态更新后，将获得一个细化的姿态假设列表，使用一个分层的姿态排名网络计算他们的分数，选择具有最高分数的姿态作为最终估计。这样的网络设计通过特定的迭代次数，可以稳定的从粗姿态估计到位姿调整获取一个最佳位姿。相比于单步直接回归的网络算法运行速度大大提升。

与位姿调整的网络相似，输入为与位姿调整后相对应的渲染图和给定rgb图片，使用与细化网络中相同的特征提取主干架构。所提取的特征被连接、位置标记并被转发到多头自注意模块，以便利用图像全局信息进行比较，姿态排序网络经过平均池化，输出描述渲染和给定rgb之间的对齐质量的特征向量，为了充分利用所有的位姿假设信息，再对所有k个姿态假设进行多头子注意力，最终得到有效排名得分，获取最佳位姿。

1. 姿态跟踪

相比于位姿估计，姿态跟踪不需要预先Global pose sampling，只需要计算两帧之间物体的变化，根据上述网络跟新旋转向量和平移向量。

6D位姿估计依赖于高精准的3D模型，但若暂时没有3D模型，也有一些商业或开源的方案，例如Qlone和luma ai大模型，只需要几秒的视频，就能生成逼真的三维模型。

Qlone 是一款基于增强现实 (AR) 技术的 3D 扫描应用程序，它允许用户使用智能手机或平板电脑轻松创建物体的 3D 模型。只需围绕物体移动设备，Qlone 就会自动捕捉物体的图像，并利用 AR 技术构建物体的 3D 模型。同时Qlone 提供了一些基本的模型编辑功能，例如删除不需要的部分、修复模型缺陷等。

luma ai同样可以从一系列照片/视频中自动生成 3D 模型，无需使用专业的 3D 扫描设备。通过照片采集，图像配准，深度估计，表面重建，纹理映射等步骤生成逼真的3D模型。

* + 1. **常见物体的实时抓取估计技术（施翰煜）**
    2. **复杂环境中机器人作业运动轨迹规划技术（黄泊凯）**

1. **预期成果与工作计划（林伟鸿）**
   1. **预期成果**

开展机器人超市具身智能技术研发，针对超市场景，在大量商品的补货取货流程下，实现鲁棒高精度的商品目标检测与实例分割、6-DoF位姿估计、基于RGB-D商品抓取点估计以及机器人作业实时动态规划控制，预期成果与指标具体如下：

1. 目标检测识别准确率：针对典型商品，在实验室模拟测试场景下达到93%以上；
2. 目标6-DoF位姿检测精度：实验室模拟测试场景下位姿估计的平均距离（ADD）不超过2cm；其中：平均距离（ADD）指计算真值姿势和估计姿势转换的两个3D模型中对应点之间距离的平均值；
3. 位姿识别的图像处理帧率：目标检测识别与位姿估计达到15FPS以上的图像处理速度，支撑机器人作业动作连贯性；
4. 机器人动态规划实时性：相对位姿变化时，机器人运动响应时间小于0.5s，实现对相对位姿变化的实时响应，并需要具备在障碍空间条件下的自主作业规划能力。
5. 知识产权：发明专利申请5项；软件著作权3项。其中：
   1. 在商品目标检测识别方面，完成发明专利申请1项，软件著作权1项；
   2. 在商品位姿估计与抓取点估计方面，完成发明专利申请2项，软件著作权1项；
   3. 在机器人作业实时动态规划控制方面，完成发明专利申请2项，软件著作权1项。
6. 联合人才培养：项目实施期内，联合培养2-3名从事机器人视觉智能、自主规划技术研发的硕士研究生。
   1. **项目研究计划**

本项目研发工作计划需1年时间，具体工作计划如下：

|  |  |
| --- | --- |
| 时间 | 实施阶段 |
| 2024.11~2025.01 | 1. 项目详细技术方案制定； 2. 试验环境搭建； 3. 机器人视觉检测技术研发； 4. 数据集收集制作； 5. 6-DoF位姿估计等算法研发与仿真模拟验证； 6. 实时运动规划算法研发与仿真模拟验证； |
| 2025.02~2025.05 | 1. 机器人视觉检测、位姿估计与实时运动规划等算法的模拟场景应用测试验证与算法集成测试 2. 完成发明专利与软件著作权申请。 |
| 2025.06~2025.10 | 1. 进行电网作业机器人视觉检测与运动规划技术的实际应用验证； 2. 进一步完善算法，完成项目技术成果交付。 3. 完成项目验收。 |