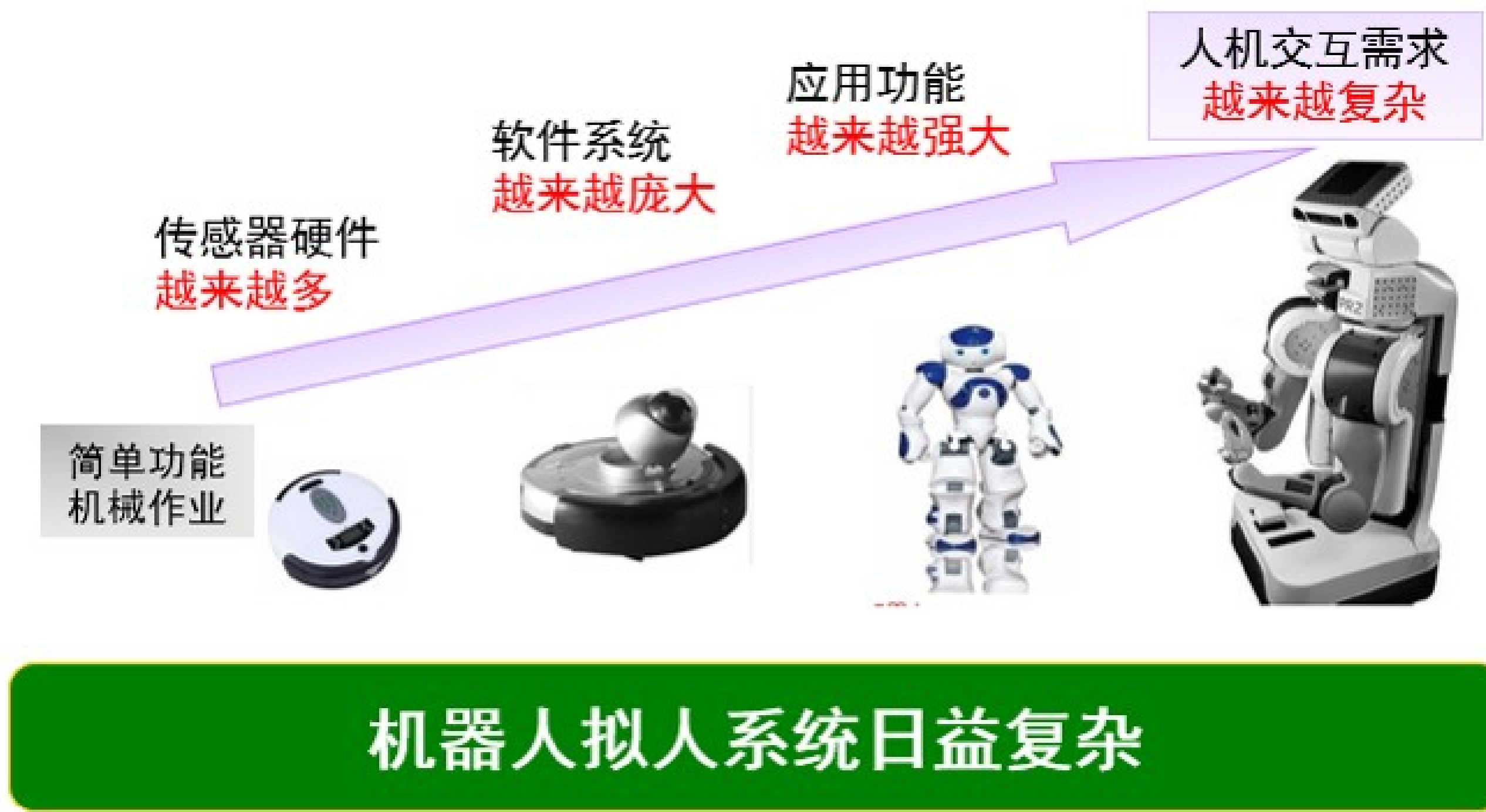




# 多模态智能机器人视觉感知与人机交互

## 背景

随着人工智能技术与产业的发展，智能机器人与人们生活具有越来越密切的联系。在儿童教育陪伴、医疗看护、导览讲解、商场导购等智能机器人应用中，视觉感知与人机交互是关键技术，即机器人通过获取场景中视觉与语音信息，理解用户需求从而更准确提供各种服务。



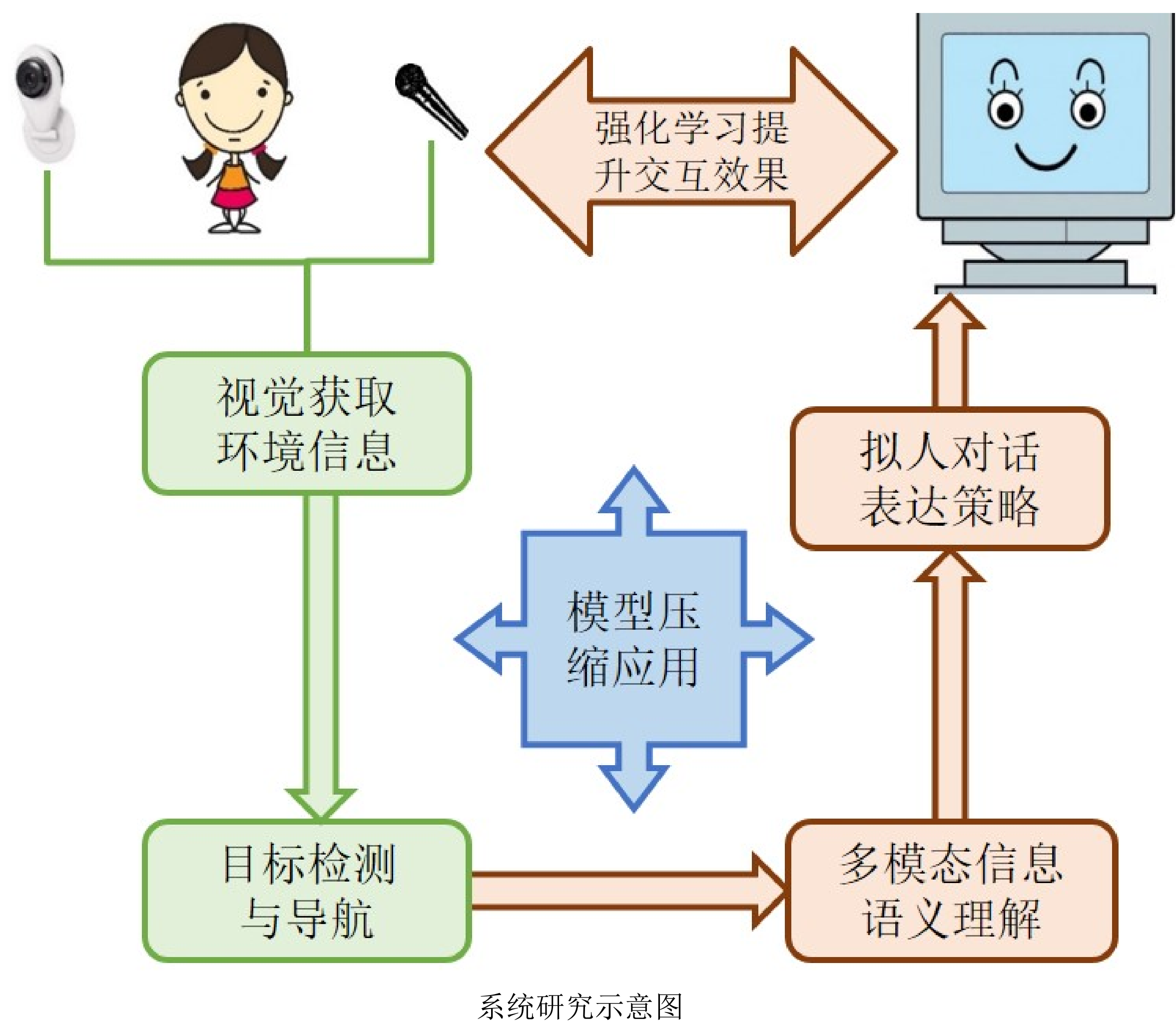
近年来，随着机器视觉感知、语音识别与合成技术的发展，人机交互已经获得很大的进步。智能机器人需求持续增长，未来发展潜力巨大。

## 关键研究问题与创新点

- 视觉感知：构建动态的视觉感知与检测模型，在真实环境中检测物体库同步动态更新，提高机器人对环境的信息获取能力与视觉导航精度，提升机器人对环境的适应性
- 人机交互：提升外部感知的多模态信息的智能理解；基于强化学习理论设计对话激励和评价策略，强化人机交互拟人表达能力
- 多模态信息理解：基于累积注意力机制提取多模态信息关键内容，提升机器人对外部感知的信息的智能理解能力
- 轻量级模型：研究通道剪枝的方法压缩深度学习模型，使得算法模型更适合部署至机器人系统上

## 研究概况

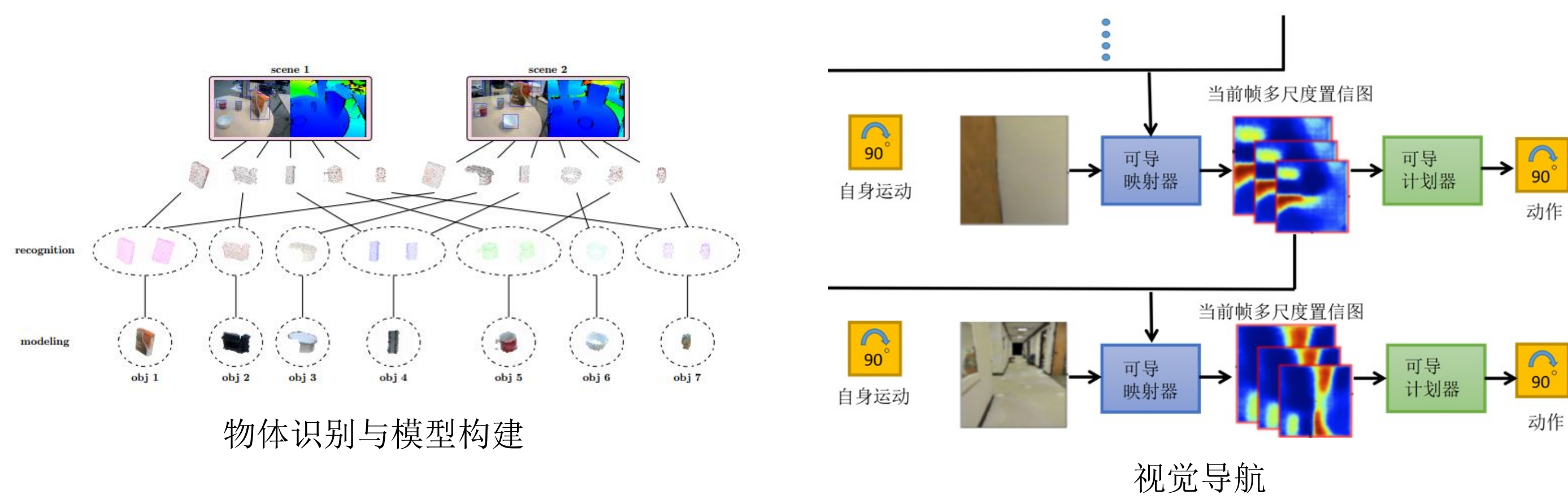
首先研究模型剪枝压缩技术以降低深度模型资源要求，围绕多模态信息展开基于累积注意力机制和最优传输理论的信息理解与融合，并基于此实现精准的机器人视觉引导，最后利用强化学习技术寻求更好的交互效果以达到智能机器人系统更为实用的目的。



## 视觉感知与导航

研究机器人场景目标视觉信息的自动获取方法；实现真实环境下的机器人自主视觉导航。

- 准确有效地获取场景视觉信息是机器人可靠运行的重要基础
- 基于视觉的机器人精准导航是机器人移动执行任务的前提

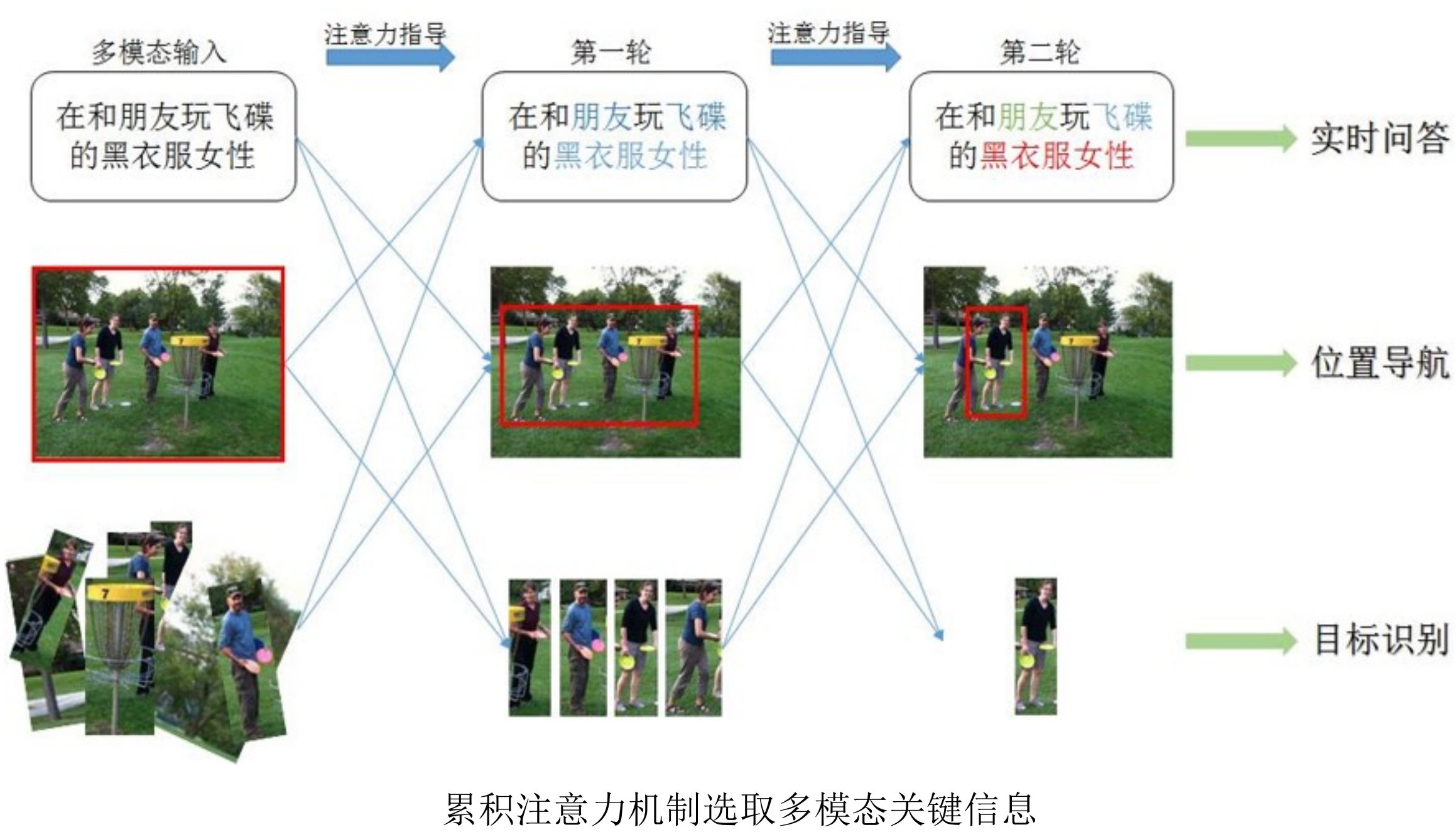


难点: 环境中信息复杂多变，识别物体时保证识别效率和准确率; 从像素点到动作构建端到端动态导航，保证导航实时性和精准度

## 人机交互

研究多模态智能人机交互技术，包括多模态信息智能理解与多模态拟人表达（语音、文本、动作、表情）实现人机自然交互。

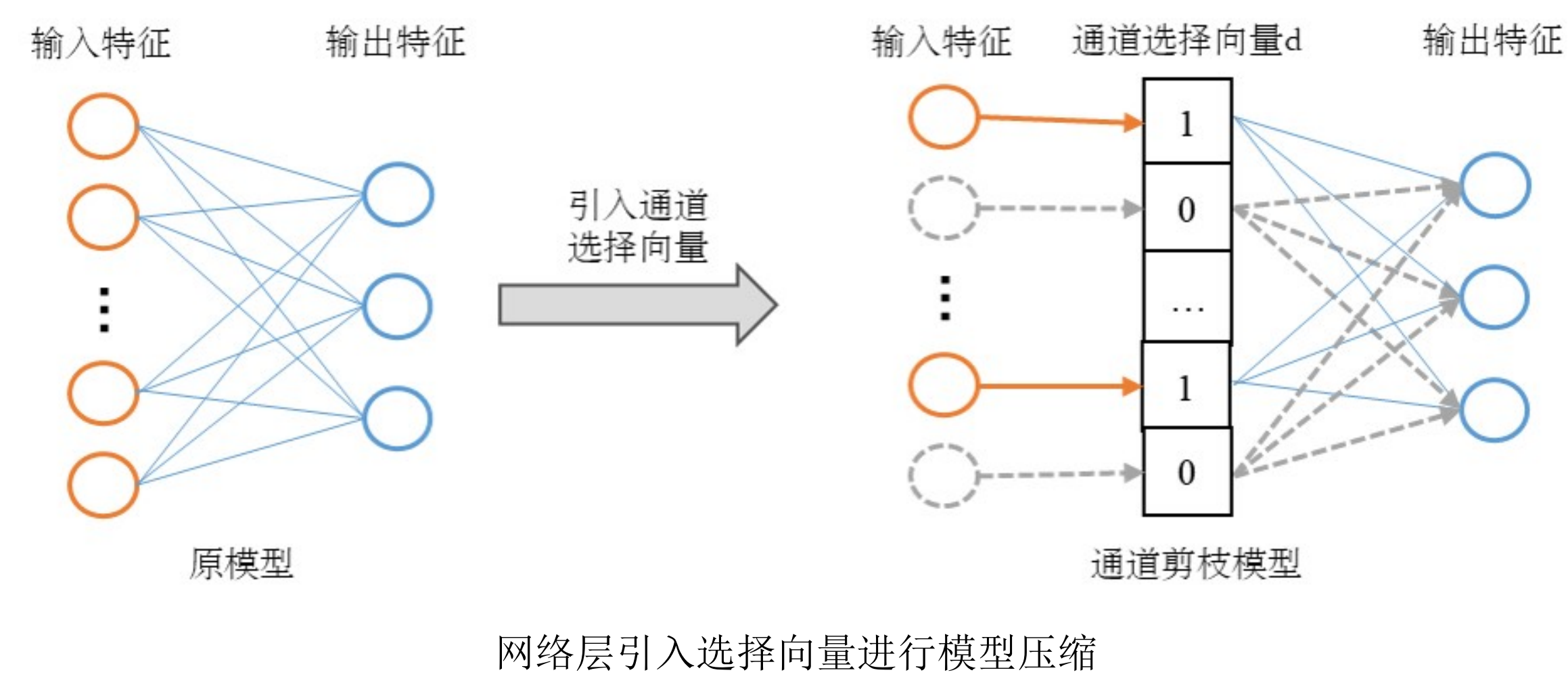
- 对外部感知的多模态信息进行智能理解与信息融合让机器人更深刻地理解真实世界
- 自然、高效、友好和智能地进行人机交互是拟人机器人重要目标



难点: 外部信息存在异构性、复杂性、冗余性，如何对信息进行智能理解获取关键信息; 人机交互对话中逻辑、任务、目的需要多样化处理

## 轻量级模型压缩

研究深度模型冗余去除方法，构建高性能的轻量级神经网络模型。



- 深度网络模型精度高与复杂度高并存，需要较高的计算硬件支撑，智能机器人设备往往计算资源有限，导致难以部署
- 将深度网络模型压缩为轻量级模型使其便于部署在机器人设备上

难点: 压缩模型以降低复杂度的同时，保障模型的精度效果