ROMBUS

Robotic Operation for Manipulation and Benchmarking Under SLM

Setting the Standard for AI-Driven Robotics

Benchmarking AI for Precision, Efficiency, and Real-World Impact

Problem

We conducted interviews with Subject Matter Experts (SMEs) from Microsoft, robotics engineers, and AI researchers, highlighted gaps in AI-driven robotic automation. Additionally, we performed a rigorous literature review on language models and robotic automation, identifying key limitations and research gaps. We discovered inefficiencies in task execution, adaptability, and real-time decision-making. Provided in more detail below:

Manual Robotic Programming — Traditional robotic systems require **explicit coding** for every task, making automation **slow, rigid, and difficult** to scale

Lack of SLM Benchmarks — There is **no structured evaluation framework** to measure Small Language Model (SLM) efficiency, accuracy, and feasibility in robotic automation

Computationally heavy LLMs — Large Language Models (LLMs) provide strong reasoning abilities but are **too slow and resource-intensive** for real-time robotic execution

Approach

Our system integrates **Small Language Models (SLMs) like Phi-3.5, Llama 3, Mistral, and Qwen** with robotic systems, enabling natural language task planning without manual programming. We enhance precision through Vision Language Models (Phi-3.5 vision, Qwen2 VL, Llava-hf) that provide scene descriptions for object detection, segmentation, and depth estimation. Our structured benchmarking framework measures **SLM efficiency, execution accuracy, and success rates**, optimizing the system for **edge computing** deployments.

Solution

AI-Powered Task Planning
Converts natural language commands into robotic actions

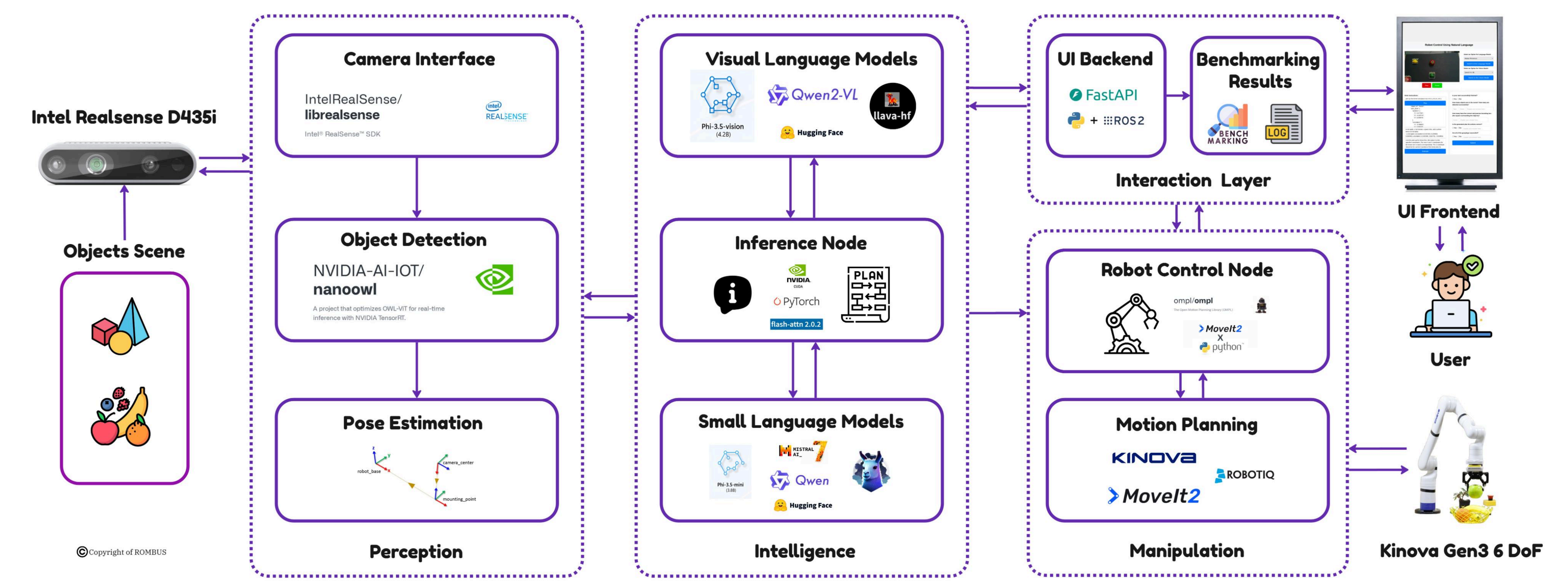
»»

Low-Latency Execution
Deploys SLM-based planning for real-time automation

»»

Benchmarking Framework
Measures SLM efficiency, accuracy, and success-fail rates

System Architecture



Process

Primary & Secondary Research

- Conducted stakeholder interviews and market research to identify gaps in AI-driven robotics.
- Performed literature review on small language models, robotics and benchmarking methods.

Prototyping & Development

- Set up robotic environments and integrated Phi-3.5, Llama 3, and NanoOWL, Phi-vision VLM.
- Developed SLM-driven task planning models and tested perception accuracy.

Iteration & User Evaluation

- Conducted 10+ usability tests to refine robotic execution and real-time inference.
- Optimized SLM response time, accuracy, and perception and control-based grasping.

Benchmarking & Evaluation

- Compared SLM-based robotics vs. traditional automation in execution speed and efficiency.
- Measured task success rates, adaptability, and computational performance of different SLMs.

Final Integration & Deployment

- Demonstrated functional live demo of AI-driven robotics.
- Provided a structured SLM benchmarking framework for researchers and industry partners for robotics pick & place tasks.

UNIVERSITY of WASHINGTON | GLOBAL INNOVATION EXCHANGE

Microsoft

Lakshita Singh : MSTI' 25
Aayush Kumar : MSTI' 25
Kendra Yang : MSTI' 25
Matthew Zhang : MSTI' 25