Toward Embedded LLM-Guided Navigation and Object Detection for Aerial Robots

Richie Suganda[2] (rrsugand@cougarnet.uh.edu), Bin Hu[1,2] (bhu12@uh.edu)

Dept. of Engineering Technology[1], Dept. of Electrical and Computer Engineering[2]
University of Houston, Houston, USA

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

NAIL Networked Autonomous Intelligent Learning

ABSTRACT

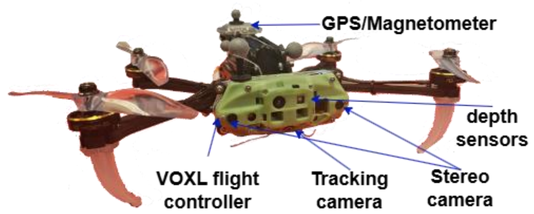
We present a novel framework for language-driven quadrotor navigation and object detection, targeting real-time deployment on edge platforms. This poster presents a proof-of-concept system and outlines a roadmap toward fully onboard, language-guided autonomy.

Ultimate Goal:

Develop fully integrated, language-driven autonomous systems for quadrotors, with real-time onboard LLM inference on embedded platforms.

Key Challenges:

- ✓ Bridging natural language understanding with low-level robotic control and perception.
- ✓ Meeting stringent compute, latency, and power constraints for real-time inference onboard small drones.
- ✓ Ensuring closed-loop performance in realistic, dynamic environments.



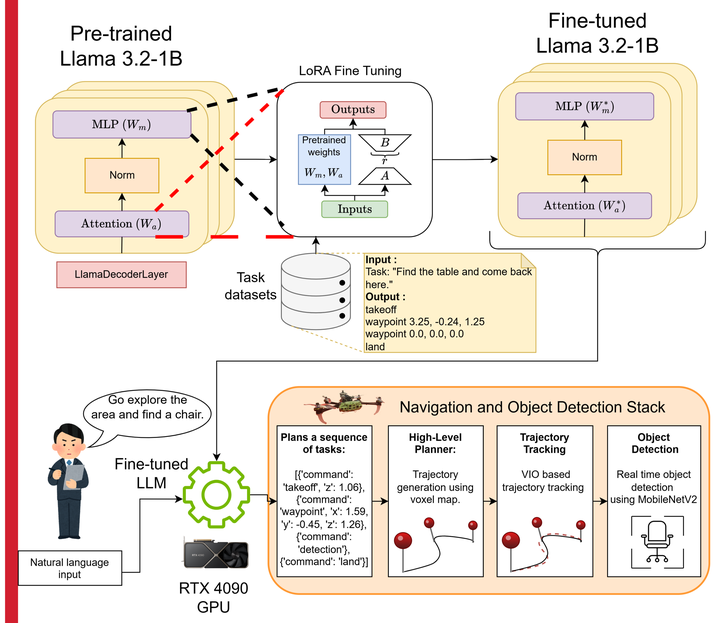
The ModalAI Seeker uses the VOXL CAM engine for VIO, onboard object detection, and path planning

- ✓ Tracking camera: VIO localization by capturing motion data
- ✓ Stereo cameras: depth mapping for obstacle detection and navigation
- ✓ Depth sensor: indoor depth perception
- ✓ VOXL flight controller: PX4 and ModalAI's flight core for agile maneuvers and robustness

KEY CONTRIBUTIONS

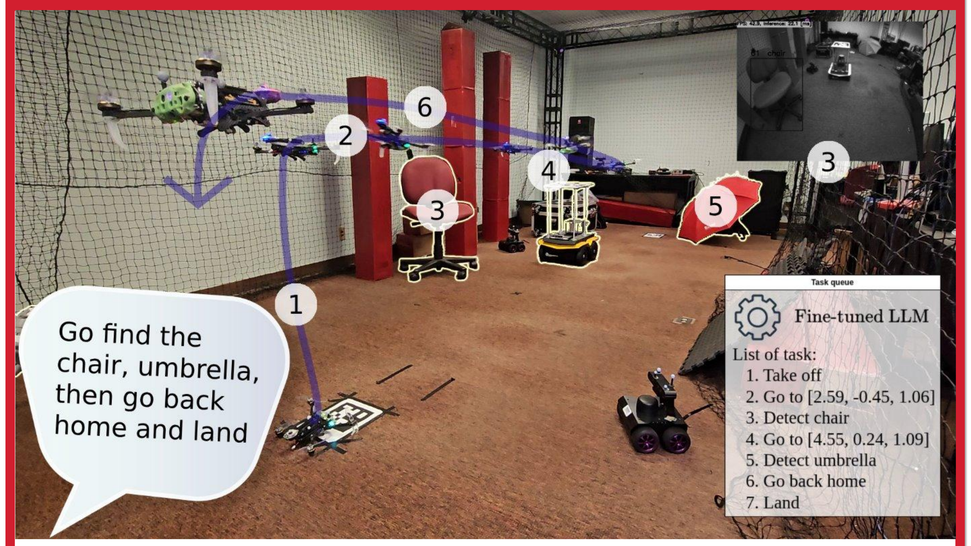
1. ***LoRA Fine-Tuning of Llama Model***: Fine-tuned Llama model with LoRA for quadrotor exploration and object localization/identification.
2. ***Hierarchical LLM Integration:*** Built a hierarchical LLM framework integrating human instructions with path planning, VIO control, and onboard object detection..
3. ***ModalAI Seeker Testbed:*** Created a proof-of-concept testbed using the ModalAI Seeker platform to validate the proposed approach

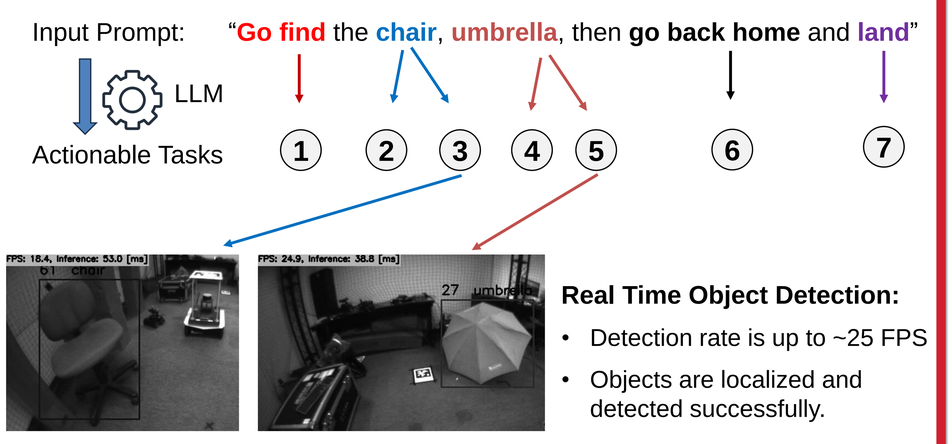
LLM based Navigation and Object Detection



- ✓ Fine tune a 1B-parameter LLaMA model using 4,500 prompt examples and ~3.5% of parameters via LoRA.
- ✓ LLM translates natural language into task-level goals for onboard execution.
- ✓ Onboard stack performs trajectory generation and tracking via VIO-based control, with real-time object detection using MobileNetV2.

PROOF OF CONCEPT DEMO





Real Time Object Detection:

- Detection rate is up to ~25 FPS
- Objects are localized and detected successfully.

FUTURE WORK

- ✓ **Onboard LLM Inference:** Apply model compression techniques such as activation-aware quantization and knowledge distillation to deploy LLMs on resource-limited embedded platforms.
- ✓ **Multimodal Input Integration**: Extend the interface to support spoken language commands, enabling natural voice-based control via audio-to-text pipelines or direct audio-conditioned LLMs.
- ✓ **Grounded Vision-Language Reasoning:** Fuse visual context from onboard cameras with language inputs to enable contextual understanding and dynamic task execution in complex environments.