

# Chief Scientist Office 2024 Team Description Paper

Takaaki Numai, Airi Yokochi, Joshua Supratman, Tatsuro Sakaguchi, Yushi Kaida, and Gakuto Okamoto\*

SoftBank Corp. Chief Scientist Office  
<https://www.robocup.ros-sier.com/>, <https://github.com/sbgisen>

**Abstract.** This paper presents SoftBank Corp. Japan team Chief Scientist Office’s new mobile manipulator “SOAR” for RoboCup 2024 @Home Open Platform League. Differing from our previous participation with “Cuboid-X,” a robot assembled from commercial products, SOAR integrates both custom-developed and commercial components, a significant step toward developing a customizable and affordable robotic platform. Our main achievement this year has been the successful integration of various hardware components and software development to tackle RoboCup challenges.

Our approach includes detailed hardware design and calibration, such as the arm and the sensor-rich head. The software focuses on developing systems, such as object and speech recognition, tailored to the dynamic environment of RoboCup.

SOAR embodies our vision of a robotic system that combines modularity with cost-effectiveness. As we continue to refine SOAR’s capabilities further, we are optimistic about its potential to create an adaptable, cost-effective robotic platform for a variety of applications and fields.

## 1 Introduction

This paper details the RoboCup @Home Open Platform League team Chief Scientist Office from SoftBank Corp. Japan. This is our second appearance at RoboCup @Home Open Platform League. Since our last participation in 2022, we have made significant improvements, which include the introduction of our new mobile manipulator “SOAR.”

At SoftBank Corp. Chief Scientist Office, we aim to pioneer a customizable and affordable commercial robot platform akin to the modular ease found in computer assembly or the historical T-Ford model. Our mission is to democratize robotics, making it accessible and adaptable for research and commercial use. We believe in a future where integrating robotics into business is as straightforward as selecting components for a custom PC build. With SOAR as a stepping stone, we aim to develop a versatile platform that celebrates modularity in both hardware and software, allowing seamless integration and upgrades through a suite of interchangeable modules.

---

\* The order of authors is based on age.

### 1.1 Focus of research

Our research is motivated by the ambition to develop a customizable and affordable robot platform. We recognize the constraints of current technologies and are committed to exploring their limits and potential. Key areas of our interest include but are not limited to:

- Modularity and Configurability: Investigate how to design a system that allows for easy integration and customization of hardware modules.
- Device Integration: Tackling the hardware challenges related to the seamless integration of diverse devices and components

We believe that these research areas are crucial to bridging the gap between the aspirational and the practical, laying the groundwork for a robot that can adapt to various needs.

## 2 Hardware Modules

SOAR is a mobile manipulator comprising four main components: the mobile base, the torso, the manipulator, and the head as shown in Fig. 1. The mobile base and the manipulator are commercially available products, while the torso and the head are custom-developed by our team. The head and the torso are separate components designed for specific functionalities; we designed the head to accommodate various interchangeable sensors and the torso to enhance the manipulator’s reach and versatility. All components are designed with minimal cable protrusion, with circuits centralized on the torso, thus streamlining the connection across components with minimal cabling. This modular approach not only facilitates customization but also organizes and simplifies maintenance and upgrades, allowing for the interchange of components as needed.

### 2.1 Mobile Base

SOAR’s mobile base (Fig. 1b) is a 2-wheel differential drive mobile base developed by F-Design<sup>1</sup>. It is supplemented with two casters, and each driving wheel and caster are equipped with suspensions for smoother movement and stability. We made additional modifications to the component, equipping it with a battery management system and a docking charging connector, enabling SOAR to recharge its battery autonomously once depleted, enhancing its operational autonomy and efficiency. We connect the motor drivers and the battery management system to a USB hub, resulting in only two cables extending out: the USB cable from the device hub and the power supply cable.

---

<sup>1</sup> <https://f-ds.jp/>



Fig. 1: SOAR and its four components.

## 2.2 Torso

The torso (Fig. 1c), developed by our team, is a linear actuator with one degree of freedom equipped with a constant-force spring to provide gravity compensation. The component elevates the head and the manipulator, providing a clearance of 410 mm, significantly enhancing the latter's reach and versatility. In addition, the component also stores the main PC and controller box as it serves as the central hub for SOAR's hardware components.

## 2.3 Manipulator

The manipulator (Fig. 1d) is xArm 7<sup>2</sup>, a commercially available product developed by uFactory. The end-effector is equipped with a two-finger gripper and a RGB-D hand camera.

## 2.4 Head

The head (Fig. 1e), designed as a sensory and interactive hub, supports a range of sensors and user interface devices for environmental understanding and social interaction. Equipped with a USB hub for efficient cable management and

<sup>2</sup> <https://www.ufactory.cc/xarm-collaborative-robot/>

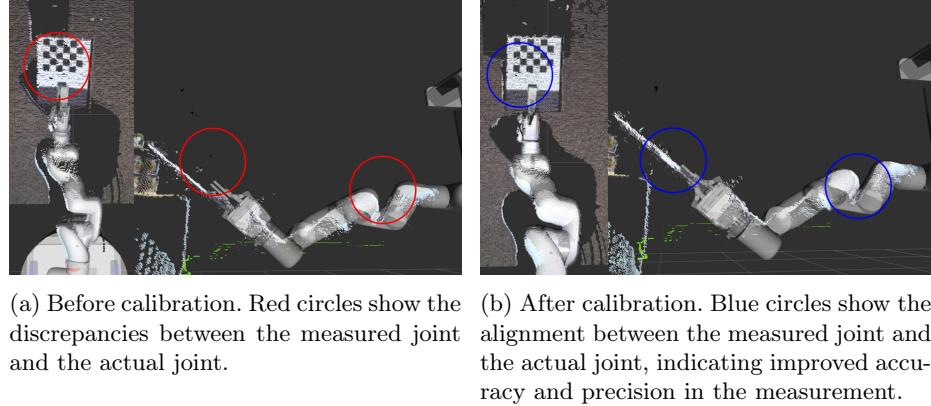


Fig. 2: Calibration results.

a separate USB for the high-bandwidth depth camera, it allows for easy swapping of sensors and devices, ensuring that the component can be customized for specific requirements. The head extends three specific cables: a USB cable from the sensor hub, a separate USB for the depth camera, and a power supply cable. In addition, the head has 2 degrees of freedom for pan and tilt movements, enhancing its SOAR's ability to survey diverse environments and engage in social interaction more effectively. Homing sensors have been integrated on both the pan and tilt for calibration and homing to enhance the head's functionality, eliminating the need for absolute encoders in motors, which reduces spatial requirements, eliminates the need for mechanical adjustments to achieve a homing position, and increases the flexibility in motor selection.

### 3 Alignment Between Modules

SOAR's modular design brings forth unique challenges in ensuring precise alignment between its components, mainly due to the frequent installation and removal of components during maintenance and upgrades, leading to performance variability. The interchangeability of modules requires a robust system to maintain consistent performances across different configurations.

To address these challenges, we used the robot\_calibration[1] ROS package. This package automatically fine-tuned the alignments of SOAR's components, ensuring each module aligns correctly with the system to maintain the robot's performance, significantly reducing the need to adjust the components to align with the system manually. Figure 2, shows an example with the manipulator and the head, where the measured joints align with their actual physical counterpart. While we achieve good calibration results with the manipulator, calibrating other components, like the mobile base, presents more challenges and is an ongoing process to achieve the desired level of alignment and performance.

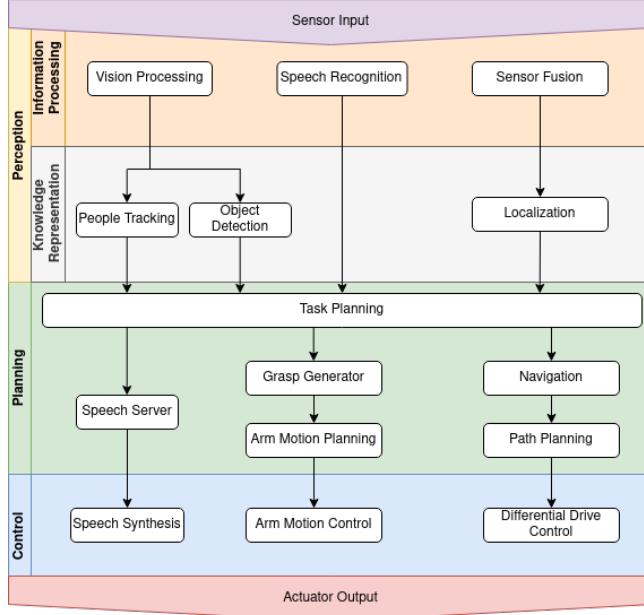


Fig. 3: SOAR software overview.

## 4 Software

The overview of our software system, shown in Fig. 3 is comprised of three main layers:

1. The perception layer, a collection of software modules that take the robot's sensor data and process them to create a symbolic representation of the environment,
2. the planning layer, a collection of software modules that take the state of the environment to evaluate and generate a sequence of appropriate robot behaviors, and
3. the control layer, the collection of software modules that translates the planned behavior and sends appropriate commands to the actuators for the robot to interact with the environment

The primary framework for software module interaction is Robot Operating System (ROS)<sup>3</sup>, with each software module representing one or several ROS nodes.

### 4.1 Object Recognition

Object detection and 3D localization are handled through the seamless integration of existing technologies. We employed ultralytics's YOLOv8[2] for its

<sup>3</sup> <https://www.ros.org/>



(a) Detecting cup noodles with YOLOv8. (b) Point cloud data and MoveIt planning scene. (c) 3D collision boxes generated by jsk\_pcl\_ros package.

Fig. 4: Object detection pipeline. YOLOv8’s bounding box (Fig. 4a) and the point cloud data (Fig. 4b) are given to jsk\_pcl\_ros package to generate 3D collision boxes (Fig. 4c), which are passed to MoveIt planning scene.

object recognition capabilities, producing bounding boxes with labels for each detected object in the robot’s field of view. To understand the object’s spatial orientation and structure, we used the jsk\_pcl\_ros<sup>4</sup> package for its robust point cloud processing. We used the EuclideanClustering and the ClusterPointIndices-Decomposer functionalities within this package to segment and cluster point cloud data within the detected bounding boxes from YOLOv8 and generate 3D collision boxes along with its pose. The pipeline is illustrated in Fig. 4.

#### 4.2 Creating dataset

To train YOLOv8 for object recognition in the dynamic nature of RoboCup competitions, where new objects are often introduced on the day of the event, our team has developed a robust method for dataset preparation. The process involves capturing images of objects from multiple angles and capturing background images on the RoboCup field. We then crop the object images using a GUI tool and overlay them onto the background images, creating composite images. We then generate a COCO-style dataset with the composite images. An example is shown in Fig. 5. This method significantly simplifies the annotation process, allowing us to create datasets with new objects within a day, thus enabling rapid adaptation and expansion of our object recognition system.

We are continuously working to enhance our model’s capability to recognize challenging objects, such as transparent objects or surfaces with reflective properties like silver. We are also refining our data collection techniques to consider factors such as varying camera angles and light intensities, which are critical not only for realistic and effective annotation but also for creating 3D simulation models. These improvements are aimed at increasing the robustness and accuracy of our object recognition system under varied and unpredictable competition conditions.

<sup>4</sup> [https://github.com/jsk-ros-pkg/jsk\\_recognition](https://github.com/jsk-ros-pkg/jsk_recognition)

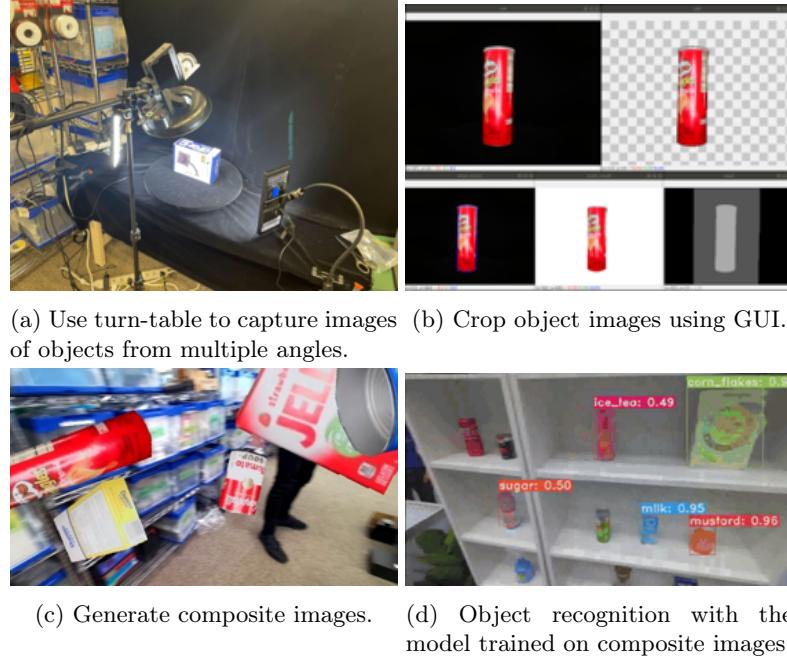


Fig. 5: Dataset creation process.

### 4.3 Speech Recognition And Interaction

We use EfficientWord-Net[3] to detect hotword to initiate the conversion. Once engaged, we employed Google Dialogflow for conversation, which can interpret spoken phrases and generate relevant responses based on our predefined conversational settings. For vocal output, we use Google Text-To-Speech.

In addition to the current pipeline, we are also exploring the integration of ChatGPT[4] as a potential replacement for Google Dialogflow to enrich SOAR's conversational capabilities. To facilitate this transition, we use Silero VAD[5] to segment the voice signals and Whisper[6] to convert speech to text.

## 5 Contribution

SoftBank Corp. has continued to sponsor various robotic competitions<sup>56</sup> and conferences<sup>78</sup> to advocate human-robot interaction research. Our sister company, SoftBank Robotics Corp. is a long-term global sponsor for RoboCup, which

<sup>5</sup> <https://wrs.nedo.go.jp/en/sponsor2020/>

<sup>6</sup> <https://tsukubachallenge.jp/2023/organization/sponsors>

<sup>7</sup> <https://ac.rsj-web.org/2023/>

<sup>8</sup> <https://roscon.ros.org/jp/2023/>

further exemplifies our commitment to advancing robotic technologies and education. Despite our low direct scientific contribution, SOAR, our current prototype, represents our venture into supporting the research community and industry by offering a platform for research and study. While it is still in development, SOAR is designed with the future in mind, a testament to our initiative in paving the way for affordable and customizable robots, which we believe will make significant contributions to both academic research and practical applications.

## 6 Conclusions and future work

In this paper, we presented our currently developing robot SOAR which will participate in the RoboCup@Home Open Platform League. We detailed the hardware specification and introduced several software modules, such as object recognition and speech recognition, which we will continue to refine to tackle the tasks in RoboCup @Home for our future work. By enhancing SOAR’s capabilities in real-world interaction, we aim to contribute not just a flexible and affordable platform but also a robust platform for research and commercial use.

## References

1. Michael Ferguson and Niharika Arora. Robust and Efficient Calibration of Mobile Manipulators. Technical report, Fetch Robotics, Tech. Rep, 2015.
2. Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8, 2023.
3. R. Chidhambararajan, Aman Rangapur, S. Sibi Chakkavarthy, Aswani Kumar Cherukuri, Meenalosini Vimal Cruz, and S. Sudhakar Ilango. EfficientWord-Net: An Open Source Hotword Detection Engine Based on Few-Shot Learning. *Journal of Information & Knowledge Management*, 21(04):2250059, 2022.
4. OpenAI. GPT-4 Technical Report, 2023.
5. Silero Team. Silero VAD: pre-trained enterprise-grade Voice Activity Detector (VAD), Number Detector and Language Classifier. <https://github.com/snakers4/silero-vad>, 2021.
6. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust Speech Recognition via Large-Scale Weak Supervision, 2022.

## SOAR Hardware Description

Specifications are as follows:

- Dimensions: (w) 550 mm (d) 650 mm (h) 1360 mm
  - 1770 mm
- Weight: 92 kg
- Max Speed: 0.83 m/s
- Mobile Base: differential drive
- Torso unit: 1 DOF linear joint
  - Sensors:
    - \* Hinson LE-50621 LiDAR
    - \* N100 9-axis IMU
- PC: Intel NUC (NUC11PHKi7C)
- Manipulator: 7 DOF, 3.5 kg payload
  - Sensors:
    - \* RealSense D435 RGB-D camera
- Head unit: 2 DOF pan-tilt
  - Sensors:
    - \* Azure Kinect RGB-D camera
    - \* NEEWER NW-410 condenser microphone
    - \* JM800S-30-360X zoom camera
    - \* Seek Thermal CompactPRO Fast Frame
  - Display: THANKO C-79D21B
  - Speaker: YAMAHA VXS1MLB



## SOAR Software Description

SOAR uses the following software components:

- Platform: Ubuntu 20.04 ROS Noetic
- Navigation: ROS Navigation Stack, EBand local planner
- Arm Motion Planner: MoveIt ROS package
- Object Recognition:
  - YOLOv8
  - jsk\_pcl\_ros ROS package
- Behavior Architecture: SMACH ROS package
- Hotword: EfficientWord-Net
- Speech Recognition: Google Dialogflow
- Speech Generation: Google Text-To-Speech
- Person Recognition: spencer\_people\_tracking ROS package