POLITECNICO DI TORINO
Repository ISTITUZIONALE

The First WARA Robotics Mobile Manipulation Challenge - Lessons Learned

Original
The First WARA Robotics Mobile Manipulation Challenge - Lessons Learned / Domínguez, D.C., Iannotta, M., Kashyap, A., Sun, S., Yang, Y., Cella, C., Colombo, M., Pelosi, M., Preziosa, G.F., Tafuro, A., Zappa, I., Busch, F., Dong, Y., Longhini, A., Lu, H., Cabral Muchacho, R.I., Styrud, J., Fregnan, S., Guberina, M., Jia, Z., et al.. - (2025), pp. 1-8. (European Conference on Mobile Robots (ECMR) Padova (ITA) 02-05 September 2025) [10.1109/ecmr65884.2025.11163319].

Availability:
This version is available at: 11583/3003493 since: 2025-09-30T09:20:24Z

Publisher:
IEEE

Published
DOI:10.1109/ecmr65884.2025.11163319

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

10 October 2026

The First WARA Robotics Mobile Manipulation Challenge - Lessons Learned

David Cáceres Domínguez[a*], Marco Iannotta[a*], Abhishek Kashyap[a*], Shuo Sun[a*], Yuxuan Yang[a*], Christian Cella[b*], Matteo Colombo[b*], Martina Pelosi[b*], Giuseppe F. Preziosa[b*], Alessandra Tafuro[b*], Isacco Zappa[b*], Finn Busch[c*], Yifei Dong[c*], Alberta Longhini[c*], Haofei Lu[c*], Rafael I. Cabral Muchacho[c*], Jonathan Styrud[c*], Sebastiano Fregnan[d*], Marko Guberina[d*], Zheng Jia[d*], Graziano Carriero[d,e*], Sofia Lindqvist[f], Silvio Di Castro[f], Matteo Iovino[g]

***Abstract*— The first WARA Robotics Mobile Manipulation Challenge, held in December 2024 at ABB Corporate Research in Västerås, Sweden, addressed the automation of task-intensive and repetitive manual labor in laboratory environments — specifically the transport and cleaning of glassware. Designed in collaboration with AstraZeneca, the challenge invited academic teams to develop autonomous robotic systems capable of navigating human-populated lab spaces and performing complex manipulation tasks, such as loading items into industrial dishwashers. This paper presents an overview of the challenge setup, its industrial motivation, and the four distinct approaches proposed by the participating teams. We summarize lessons learned from this edition and propose improvements in design to enable a more effective second iteration to take place in 2025. The initiative bridges an important gap in effective academia-industry collaboration within the domain of autonomous mobile manipulation systems by promoting the development and deployment of applied robotic solutions in real-world laboratory contexts.**

***Index Terms*— Mobile Manipulation, Collaborative Robotics, Lab Automation**

I. Introduction

Modern biomedical research often relies on precise and repetitive lab procedures carried out by highly qualified personnel. Despite advancements in experimental techniques and research technologies, routine tasks — such as preparing solutions, cleaning instruments, or organizing equipment — continue to be performed manually in many laboratories. In particular, washing glassware remains a labor-intensive process that, while crucial for ensuring cleanliness and adherence to quality standards, does not require specialized expertise. Instead, it consumes the valuable time of researchers who could otherwise devote their skills to scientific duties.

The WASP Research Arena (WARA)-Robotics, hosted by ABB Corporate Research Center in Västerås, Sweden is financially supported by the Wallenberg AI, Autonomous Systems, and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

*These authors have contributed equally to this work.

[a]Center for Applied Autonomous Sensor Systems (AASS), Örebro University, Örebro, Sweden

[b]Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Milano, Italy

[c]Division of Robotics, Perception and Learning, Royal Institute of Technology (KTH), Stockholm, Sweden

[d]Lund Technical University (LTH), Lund, Sweden

[e]University of Basilicata, Italy

[f]Pharmaceutical Technology & Development, R&D, AstraZeneca, Gothenburg, Sweden

[g]ABB Corporate Research, Västerås, Sweden

Fig. 1: Lab Automation use-case in the AstraZeneca chemical laboratories: a human operator drives a cart around the lab to collect dirty glassware, organizes it in a dishwasher tray, and finally inserts the tray in the dishwasher.

The primary motivation behind the first WARA Robotics[1] Mobile Manipulation Challenge, hosted at ABB Corporate Research in Västerås, Sweden, was to address the problem of task-intensive lab work. The challenge aimed to push the boundaries of mobile manipulation research by tackling a highly relevant but under-explored real-world use case: automating routine laboratory workflows. Specifically, it aimed to develop systems capable of autonomously navigating human-populated spaces to transport carts of glassware from one location to another and manipulate items for loading into an industrial dishwasher. The challenge use-case was proposed by AstraZeneca (as depicted in Figure 1) and it encompassed multiple interconnected sub-tasks, including safe navigation, grasping and dexterous manipulation, as well as perception for object detection and pose estimation. By tackling these complex research areas, participating teams created innovative robotic solutions that have the potential to transform lab routines, thereby allowing scientific staff to reallocate their time to high-value-adding research activities.

Engaging in this type of challenge enables WASP industrial partners to guide the direction of academic groups, ensuring alignment with practical, real-world applications. This strategic collaboration accelerates progress in the field while providing mutual benefits to all parties involved,

[1]https://wara-robotics.se/

particularly in learning and knowledge development. While the specifics of these benefits cannot be precisely predicted, it is expected that they will emerge in the future, contributing to the ongoing growth and advancement of the robotics community.

This paper describes the four different approaches to solve the task as proposed by the participating teams and outlines the results and lessons learned. Importantly, the challenge also served as a way to evaluate the technology readiness of academic research, offering a realistic testbed for translating theoretical advances into deployable systems. These insights will help formulating the 2^{nd} iteration of the challenge.

II. Challenge Rules

The participating teams were challenged to develop a system that autonomously and safely navigates in a human-populated lab environment, localizes a cart loaded with glassware that needs washing, transports the cart to a designated dishwasher room, and manipulates the glassware for loading into an industrial dishwasher. There were no restrictions on the choice of the robotic platform nor the solution strategy.

A challenge mock-up kit (shown in Figure 2) was sent to the teams to allow them to develop and test their own solutions before coming to the WARA Robotics lab in Västerås. In the mock-up kit, the lab glassware was replaced by plastic items, comprising 3 bottles and 3 beakers.

The challenge was composed of the following sub-tasks.

A. Carting glassware that needs washing

The first sub-task concerned a mobile robot navigation and manipulation setup. In this task, the robot had to navigate an environment to a preset goal location where a cart had been positioned. A bin filled with randomly placed plasticware was positioned on the cart. Then, the robot had to pull/push the cart and navigate again to a second marked location in a room where the dishwasher tray was set up. It was assumed that there were no closed doors for the robot to manipulate along the way. For this task the robot had only onboard sensors available which included for instance lidars and cameras. WARA Robotics provided a 2D occupancy map of the environment as prior and allowed teams to run a (teleoperated) mapping session prior to addressing the challenge, if desired. The task was considered successfully completed if the robot was able to move the cart from the start to the goal location. It was also expected that the path the cart needs to take included at least one left and at least one right turn.

B. Manipulating glassware and dishwasher

The second sub-task assumed that the robot was already at the dishwasher tray location and was presented with a table-top scenario. A bin filled with randomly placed plasticware was provided at one end of the table and the dishwasher tray at the other end, with both placed roughly within predefined workspace zones. The task of the robot was then to select items from the bin, to pick them up, and to safely insert them onto the pins sticking out from the dishwasher tray. Different levels of difficulty were envisioned for this task, ranging from very easy (small workspace feasible for a fixed-base robot, homogeneous non-transparent items tagged with QR codes) to very hard (large workspace that requires a mobile base, heterogeneous transparent plastic objects). This task focused on random bin picking and automatic loading, while unloading the tray was still performed manually.

Fig. 2: All the teams received equipment to mock-up the lab automation use-case. The equipment included a cart with 4 castor wheels, a plastic bin, a plastic base hosting 2 different types of dishwasher pins (to mimic the dishwasher tray), 3 plastic bottles, and 3 plastic beakers.

III. Approaches

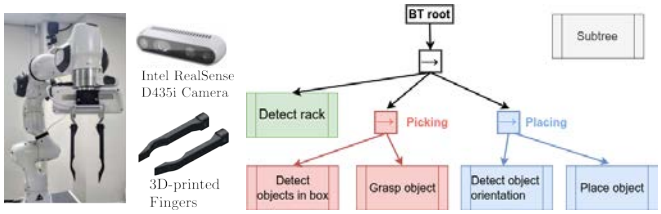
This section provides an overview of each team's approach to solve the proposed challenge. Four teams from different universities took part in the challenge: Örebro University (ÖrU), Politecnico di Milano (PoliMi), KTH Royal Institute of Technology (KTH), and Lund University (LTH). A short overview of the strategy used by each team and what task they focused on is reported in Table I, where Task A represents the mobile navigation sub-task (II-A), while Task B indicates the dishwasher loading sub-task (II-B).

TABLE I: Summary of the strategy adopted by each team.

Team	Platform	Gripper	Perception	Task
ÖrU	Franka Panda	parallel	RGB-D camera	B
PoliMi	GoFa 5	parallel + suction	RGB-D camera	B
KTH	Mobile YuMi	parallel	RGB-D camera	A&B
LTH	Mobile YuMi	parallel	RGB-D camera	A

A. Örebro University (ÖrU)

The ÖrU team focused on subtask B, dishwasher loading. The robotic manipulation system consisted of a Franka Emika Panda arm with custom 3D-printed fingers and an Intel RealSense D435i camera, as shown in Figure 3a. Together, the vision, grasping, and task execution systems enabled the robot to detect, grasp, and place transparent bottles and beakers onto designated dishwasher pins.



(a) (b)

Fig. 3: (a) Robot hardware setup. (b) Behavior Tree.

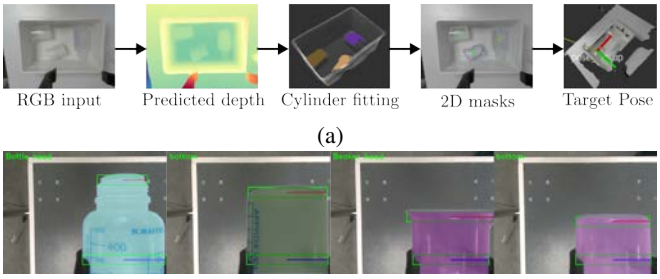
1) Vision System: The vision system relied on the RGB-D camera to detect and localize objects and pins. It needed to overcome challenges inherent to transparent surfaces, such as unreliable depth data and noisy point clouds.

a) Bottle and Beaker Detection and Localization: To overcome the problem of poor depth information, they employed MoGe [1] for monocular depth estimation from RGB images which produced more refined depth information compared to the camera depth image. The point cloud from estimated depth was used to segment the bin with the objects in it. They then applied cylinder fitting to the points inside the bin to detect the cylindrically shaped objects, mapping the inliers from the fitting process to pixels in the depth image. Using an empirically developed ranking system based on pixel count (indicating visibility) and aspect ratio (closeness to 1 indicating better geometry) that ensured robust object selection for manipulation, the best detection was selected for grasping. The process overview is shown in Figure 4a. The estimated depth is relative, not metric, but with a single object layer and near-vertical camera it was converted to Cartesian coordinates using the camera's intrinsic parameters.

b) Opening Direction Detection: Cylinder fitting locates bottles and beakers but doesn't reveal the openings' orientation. To ensure proper placement with the opening facing downward, an extra step was added to identify its direction. Once the object is held in the gripper, they segmented it with SAM2 [2] and determined its orientation through a custom feature detection module that analyzed structural differences between the upper and lower parts of the object, as shown in Figure 4b. The system then decided whether the object should be flipped to ensure correct orientation for placement.

c) Pin Localization: The pin localization process was based on the estimation module from FoundationPose [3]. This framework utilized an RGB-D image along with a mesh model of the plastic base and two pins to estimate the pose of the object. To improve accuracy, they incorporated a post-processing step that evaluated the quality of the estimated pose. This step relied on a simple registration error metric, which measured the sum of distances between mesh vertices and their nearest neighbors in the point cloud with a lower sum distance indicating a better alignment. To refine pose estimation, the system selected the estimate with the lower registration error between the proposal with initial orientation and a proposal rotated by 180° around the vertical axis.

2) Grasping System: The team used a custom-designed gripper and a Cartesian impedance controller for manipulation [4]. The gripper featured a curved front end that centered the object and aligned its axis with the tool tip, facilitating accurate placement (Figure 3a). The Cartesian impedance controller provided compliant behavior during motion execution in response to external forces, improving accuracy during contact-rich operations such as insertion reducing the impact of minor pose errors.



(a)

(b)

Fig. 4: (a) Bottle and beaker detection and localization. (b) Opening direction detection.

3) Task Execution: A Behavior Tree (BT) [5] was used to define and execute the robot's behavior, structured into three main sub-trees (Figure 3b): *Detect Rack*, *Picking*, and *Placing*, each handling a key phase of the manipulation task.

The Detect Rack sub-tree guided the arm to the placing area and performed the pin localization III-A.1.c, providing a reference for accurate object placement.

The Picking sub-tree began with *Detect objects in the box*, which moved the arm near the bin and ran the object detection routine III-A.1.a. This behavior included adaptive end-effector adjustments to improve detection in cluttered conditions. Upon successful detection, the system transitioned to *Grasp object*, using the gripper's position sensor to differentiate beakers from bottles based on diameter.

The Placing sub-tree included the *Detect object orientation* step III-A.1.b, followed by the placement action, aligning the object with the correct dishwasher pin.

To increase robustness, the BT incorporated two recovery mechanisms: one for failed grasps, using finger feedback to retry detection and grasping; and another during insertion, using force sensing to detect incorrect orientation, reorient the object, and retry placement. This modular structure enabled reliable task execution while handling common errors in a cluttered lab environment.

4) Limitations: The proposed approach made the following assumptions: (i) the bin contained only a single layer of objects, (ii) the environment was relatively uncluttered, and (iii) the system did not have a recovery mechanism if the placement failed. Potential solutions to overcome these limitations included aligning the estimated depth to metric depth, incorporating Vision-Language Models (VLM) to detect the plastic bin and dishwasher pins in a cluttered environment and to generate masks for subsequent processing, and designing a robust placement detection module using a VLM to verify the correct placement of objects.

B. Politecnico di Milano (PoliMi)

The PoliMi team focused on subtask B, dishwasher loading. Their setup was built around an ABB GoFa 5 equipped with a gripper featuring custom 3D-printed fingers, shown

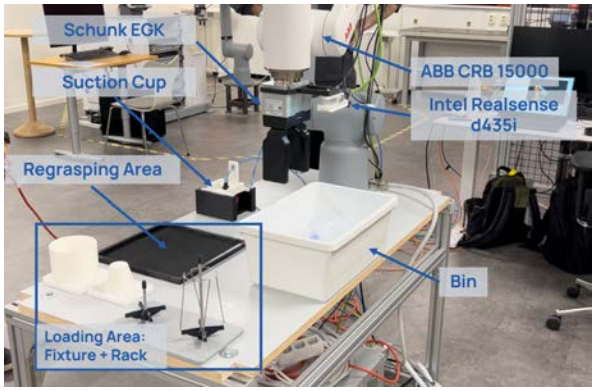


Fig. 5: The setup proposed by PoliMi was built around an ABB GoFa 5 collaborative robot, equipped with a Realsense D435i stereo camera and a custom parallel gripper optimized for the pin insertion task. A detachable suction cup was also employed for the bin-picking task.

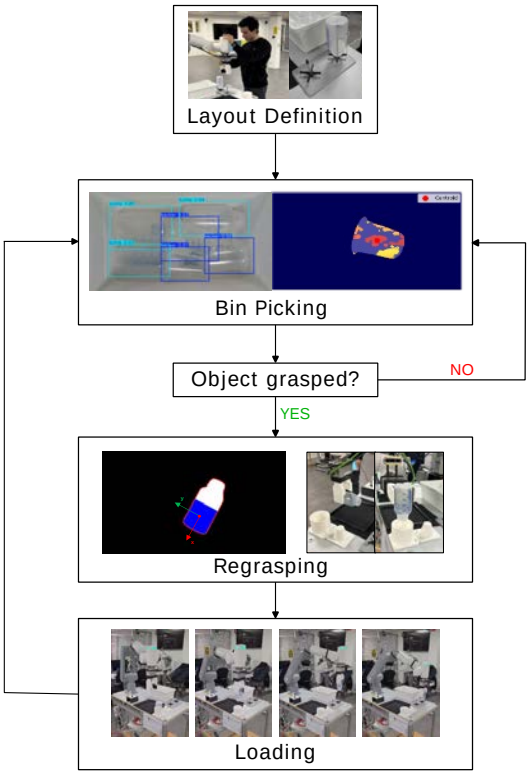


Fig. 6: Team PoliMi approach involved teaching the desired tray layout through single-shot kinesthetic demonstration. Bin-picking and regrasping actions were repeated iteratively until the dishwasher tray was fully loaded.

in Figure 5. The gripper was mounted on a flange holding an RGB-D Realsense d435i camera. The setup also comprised a suction cup which was used to extract bottles and beakers from the bin to overcome the limitations of unfeasible grasping poses due to the finger encumbrance. The team proposed to leverage the flexibility of a one-shot kinesthetic teaching approach, which they used to teach dishwasher loading to the robot. Moreover, they used visual perception to detect the desired target during bin picking and to estimate object poses for a regrasping action, which was used to position the parts on the fixtures before the loading phase. Their framework consisted of four principal components, as shown in Figure 6 and detailed below.

1) Layout definition: This initial step allowed the user to specify the number of bottles and beakers to be loaded into the dishwasher, along with their layout on the tray. This information was provided to the system by leveraging Programming by Demonstration. To define a new layout, the user first selected the type of item to be loaded. Then, they hand-guided the robot from the item's dedicated fixture to the dishwasher pins, storing movement waypoints and actuating the gripper when needed through the GoFa arm-side interface. Thus, the entire layout was defined as a set of loading skills [6]. A dedicated user interface allowed the user to easily teach and store a new layout or load a previously defined one.

2) Bin-picking: Given a predefined layout, the bin-picking phase was used to extract from the bin one of the items required by the layout. For this purpose, the robot grabbed the suction cup from its holder with the gripper and moved above the bin-picking area to allow a proper view of the entire bin with the in-hand camera. A fine-tuned YOLO-based Deep Neural Network (DNN) model [7] recognized objects in the bin as bottles or beakers. Then, a SAM2 [2] module segmented the items and estimated their picking pose. For each localized object, the depth was extracted by averaging the perceived depth over the mask. Among the objects required by the predefined layout, the closest one to the camera was selected for grasping, to avoid occlusion and facilitating extraction. Then, the picked item was released in a regrasping area in front of the robot.

3) Regrasping: The extracted object was prepared to be placed in the tray through a regrasping pipeline. A fine-tuned YOLO-based DNN model was used to detect the object in the grasping area that matched either the "bottle" or "beaker" label, selected during bin-picking. The region of interest obtained from a successful detection was fed to a SAM2 module that extracted the object mask, which was then refined through erosion and dilation steps. These morphological transformations mitigated inaccuracies arising from object transparency, which might have led to poor segmentation performances. Finally, a template-matching procedure determined the mask centroid. A candidate, top-down, grasping pose was computed by shifting this point along the principal axis of the item, identified by fitting a minimum area aligned bounding box to the mask, towards the bottom side of the object. The computed grasping pose was used to correctly grasp the object with the parallel gripper and place it in the corresponding fixture.

4) Loading: Once the item was regrasped and placed in the fixture, the system loaded the waypoint list corresponding to the action being taught by the user. Thus, through simple linear waypoint movements and gripper actuation, the robot was able to load the object onto the target pin. The system then updated the item list by removing the object that had just been loaded. After that, it iterated back to the bin-picking phase until every item in the layout was loaded.

The four modules were used to build the pipeline shown in Figure 6. The initial layout definition was used to de-

fine the loading task, stating which item to place on each dishwasher pin. The loading actions composing the layout were saved as list of waypoints, taking the item from the corresponding loading area (shown in Figure 5) to the desired pin. During the loading phase, these actions were executed as fixed linear waypoints motions to move the item from the initial fixture to the target pin. This approach enabled reuse of the layout across dishwasher cycles without relying on complex vision systems. It ensured precise and safe loading, allowing users to easily define robot movements in cluttered environments. In contrast, using vision to detect pins and loading areas on real dishwashers would have been challenging due to complex tray layouts, reflective surfaces, and moisture—factors that significantly increase perception errors. Regarding the intermediate regrasping step, its purpose was twofold: it was used to prepare the object, with its base facing upwards, for the loading phase, but also to check whether the grasp executed during the bin-picking phase was successful or not. If no graspable object was detected, the bin-picking module was repeated until an object of the desired type was extracted and placed on the table.

Overall, the proposed solution showed reliable performance on the challenge day, being able to load both pins on every run. The cycle time for the full pipeline took around two minutes per object, from bin-picking to loading. Despite the heavily reflective objects, the bin-picking pipeline also showcased good performance, as image post-processing through different color filters and depth averaging managed to return feasible target poses for the suction cup.

The main drawback of the method was its reliance on static fixtures in the regrasping area, required as an intermediate step to complete the loading procedure. Another element that penalized cycle times was the choice of a six-degrees-of-freedom (DoF) robot, as having no redundancy translated into longer unwinding motions to prevent reaching joint limits or cable entanglement The main failures stemmed from the suction cup getting stuck in bin corners or struggling with cluttered layers of mixed objects—though none occurred on the challenge day. Future setups should integrate the suction cup into the gripper and eliminate regrasping fixtures to enhance dexterity and flexibility.

C. Royal Institute of Technology (KTH)

The KTH team used the ABB Mobile YuMi research platform, customized multipurpose grippers, and a ROS2-based perception and navigation software stack. Their solution addressed the complete WARA Robotics Mobile Manipulation Challenge, thus tackling both subtasks.

The approach made several operational assumptions to limit the task complexity. It assumed that the approximate initial position of the cart, the positions for offloading, and the dishwasher locations were predefined. Regarding manipulation, it assumed that cups and bottles were initially placed upright on surfaces of known height that ideally had distinct, non-reflective, and non-white appearances to facilitate reliable visual detection. Additionally, tables and trays were assumed to be visually distinguishable.

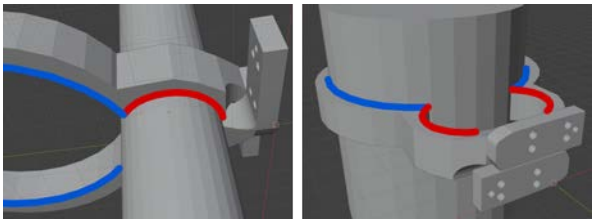


Fig. 7: The KTH multi-purpose finger design is shown by the active regions, in blue for plasticware manipulation and red for cart handling.

1) Hardware Setup: The ABB Mobile YuMi research platform was chosen to address the challenge. It comprises an ABB Dual-Arm YuMi robot mounted on a telescopic pillar fixed on a custom-made omnidirectional base. The robotic base uses four independently steerable and drivable wheels to achieve almost-instantaneous arbitrary change in planar velocity, i.e. directional and rotational. The mobile base is equipped with two 2D lidar sensors (front and back) for mapping and localization. The manipulator is a dual-arm ABB YuMi robot. Each arm has 7-DoF with a payload of 500g. An Azure Kinect RGB-D camera, mounted atop the robot, provides visual and depth information required for perception tasks.

The team mounted custom-designed grippers (Figure 7) that offered compliant and adaptable grasping capabilities suitable for grasping of both the cart handle and cylindrical plasticware. This multi-purpose gripper design has demonstrated its effectiveness in manipulation with multiple modalities [8], [9]. As the gripper closed, the handle or the plasticware naturally slid to tightly match the inner surface of the fingers despite the positional uncertainty, exemplifying a mechanical funnel [10]. It also showcased the usage of pregrasp cages as waypoints for object grasping [11], [12].

2) Software Stack: The software architecture utilized ROS 2 and integrated existing open-source packages with custom modules. Navigation functionalities were provided by SLAM Toolbox for environment mapping, Nav2 with Adaptive Monte Carlo Localization (AMCL) [13], and Smac Hybrid-A* planner [14] for path generation. The team implemented an adaptive robot footprint that dynamically accommodated the additional geometry of the cart during transport.

The object-detection module employed a YOLO-based [7] DNN to recognize and localize bottles and cups in the RGB images. Bounding boxes were projected into the depth space to compute the three-dimensional positions on a table or within a workspace. Plane segmentation was then applied to localize the height of objects. For dishwasher trays, a separate detection pipeline using a transformer-based model (Grounding DINO [15]) located black plates (the tray surface) and pins. Point-cloud processing through principal component analysis (PCA) and convex hull optimization identified the tray corners and pin positions.

The robot control was realized through high-level primitives interfacing with ABB hardware controllers through ABBs *RWS* API [16] with an adapted version of the interface from [17]. An impedance control strategy ensured robustness

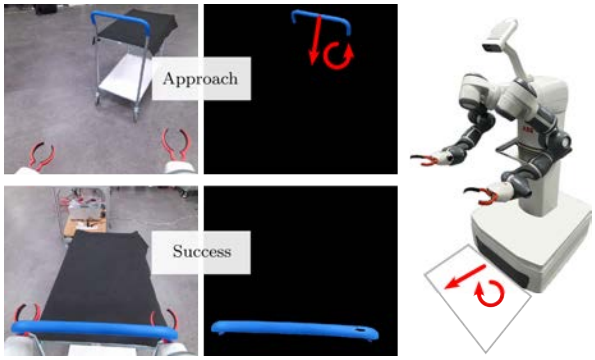


Fig. 8: In KTH's carting method, the mobile base was velocity controlled and followed the illustrated color-based visual servoing policies. The impedance control of the robot arms enabled robust grasping and transportation.

against positional uncertainties and alignment inaccuracies during grasping and manipulation.

3) Task-Specific Strategies: The cart handling procedure used visual servoing based on HSV color segmentation for alignment with cart handles, followed by impedance-controlled lateral grasping. After grasping, the navigation module dynamically updated the robot's collision footprint to safely transport the cart. The masking and visual servoing approach are shown through real examples in Figure 8.

The plasticware manipulation subtasks involved detecting upright cups and bottles using YOLO-based detection, computing suitable grasp poses, and executing impedance-controlled grasps. Placement into dishwasher trays leveraged transformer-based detection to identify tray structures and pins, with impedance-controlled insertion motions, ideally ensuring safe and precise object placement.

4) Results and Discussion: Subsystem evaluations demonstrated reliable performance, although full integration between navigation and manipulation subsystems was not completed within the available timeframe. Plasticware detection was qualitatively accurate, and multiple successful grasps and placements were observed during experimental trials. An example successful trial is shown as a step sequence in Figure 9. Navigation and cart pushing tasks achieved reliable performance, validating the effectiveness of visual servoing and adaptive footprint planning. An example successful cart approach and grasp is shown in Figure 8. Overall, the described approach showcased a high degree of modularity, where off-the-shelf algorithms were integrated with custom gripper hardware and task primitives.

While the color-based cart detection operated effectively in controlled lighting conditions, it may require more robust segmentation methods for environments with similarly colored backgrounds. Furthermore, the dishwasher-loading pipeline relied on known and consistent object orientation, limiting its applicability to objects lying on their sides or in cluttered bins. Further potential sources of failure were (i) the absence of explicit failure detection and recovery mechanisms, (ii) errors in object and goal pose estimation from the perception pipeline, and (iii) variability in wireless network communication quality with the mobile robot.

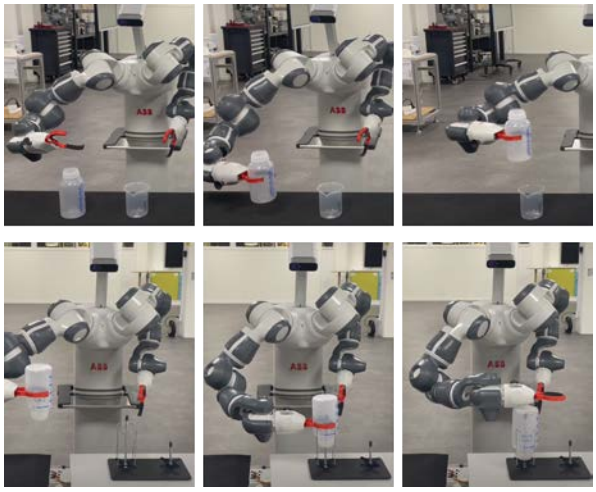


Fig. 9: The KTH plasticware manipulation is shown through an example sequence. The top row shows pick procedure through the bottle-detection, grasping, and moving steps. The bottom row shows the placing procedure through the pin-detection, insertion, and compliant pushing steps.

Future iterations of the manipulation challenge could be centered around the encountered limitations and sources of failure, leading to more reliable and widely applicable solutions. The original plan in this project was to incorporate the skills to solve each sub-task into a reactive Behavior Tree policy [18]. While that was not possible within this project due to time constraints, it would be an interesting extension that should notably improve robustness.

D. Lund Technical University (LTH)

The LTH team focused on the autonomous cart navigation subtask A, particularly cart pulling. Their approach was to use Model-Predictive Control (MPC) to achieve holistic control of the ABB Mobile YuMi research platform. To reduce the complexity of the problem, some initial assumptions were made, which then got slightly relaxed: the robot's, cart's, and goal positions were known and accurate and the cart's handlebar was easily recognizable in the environment.

1) Hardware Setup: The ABB Mobile YuMi research platform was chosen for the cart navigation task with standard off-the-shelf ABB fingers mounted on both grippers, as the gripping width was enough to hold the cart by its handlebar.

2) Software Stack: The actual control of the robot was performed using ROS 2, with hardware specific drivers for each sub-robot (i.e. the torso, the pillar, and the base) and a whole-body control interface for the entirety of the robot. The control signals were generated by a custom-made MPC using the Crocoddyl optimal control library [19]. A set of Cartesian trajectory tracking controllers were used as a fallback strategy in case of convergence issues.

3) Modeling: From a path-following perspective, both the dual-arm mobile manipulator and the cart were modeled as points in the xy-plane with a heading direction, expressed by two reference frames, Σ_B and Σ_C respectively, having the z-axis orthogonal to the world plane, expressed by the frame Σ_W. It was assumed that the robot was rigidly grasping the

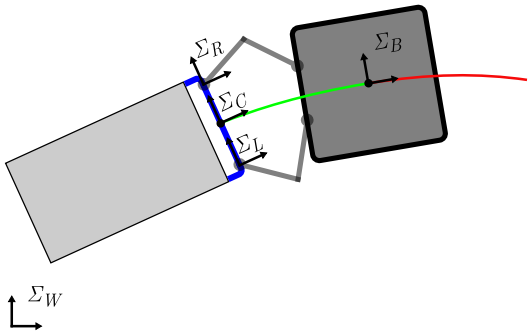


Fig. 10: Model schematic. The red line represents mobile base's obstacle-free path obtained from the path planner, while the green line represents the history of base's positions, which is converted into a reference trajectory for the handlebar and thus the cart.

cart at its handlebar with both arms, in a symmetric fashion. In this way, Σ_C was constrained to be between Σ_L and Σ_R, making the cart navigation problem an absolute-relative control task. An additional assumption was that the mobile base was constrained to non-holonomic movements, while the cart can move freely in space.

4) Method: The cart-pulling navigation problem was split into two separate subproblems. Firstly, a smooth, safe, online path was generated for the mobile base. This path was followed by a trajectory-tracking controller moving the base along the path. Secondly, the cart was steered so as to avoid obstacles while minimizing the arms' effort and maintaining manipulability. These goals were accomplished by prescribing the cart's path to be the path traversed by the base, i.e. the path comprised of previous base positions. This way the cart also followed a smooth obstacle-free path. The model with the constructed references can be seen in Figure 10.

a) Path generation: To navigate the cart-robot system in the environment, an obstacle-free path was generated. This was achieved in an online fashion, to account for dynamic obstacles. Before creating a path, obstacles were identified and converted to a useful representation for the planner. Lidar scanners were employed to collect distance measurements of the surrounding environment. The obtained distances were converted to a point cloud and morphologically inflated into circles in the xy-plane. A morphological union was then performed so to gather circles in bigger blobs, which were then simplified in the interest of computation time for the next step. The *starrification* algorithm in [20] was used to convert the obstacles in *starshaped* ones. The obstacles were further inflated with a clearance radius using [20] to create a *tube-path* that guaranteed convergence to the goal pose using the planner in [21]. The clearance radius was set to be the biggest radius that individually encircled the robot and cart. The path was a limited-horizon orbit of a dynamical system describing the flow towards the goal:

$$\dot{r} = M(r, \mathcal{O})(r^g - r)$$

with r the current robot position, r^g the desired goal, $M(\cdot, \cdot)$ a matrix that describes the attractive dynamics, and $\mathcal{O}$ is the set of world obstacles.

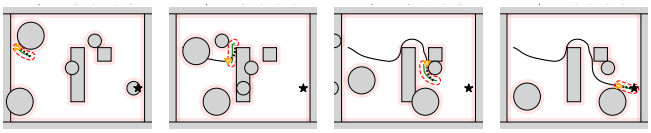


Fig. 11: Path generation for a unicycle robot with dynamic obstacles and safety guarantees over time (left to right). The robot (yellow arrow) follows the kinematically constrained trajectory (black dashed) while staying inside the safe area (red boundary) of the local desired trajectory (green solid), reaching the goal (star), never colliding with the dynamic obstacles (circles, translating vertically or horizontally). The overall motion is depicted in solid black, while the obstacle safety boundaries are depicted in light red. Generated with [20].

b) Base path following: Given a path, the robot's mobile base was tasked with moving backward while minimizing the time required to reach the goal. An MPC controller was employed to move the robot on track in the shortest time. The same controller was also tasked with arm control. The prescribed path consisted of both positions and orientations.

c) Cart trajectory following: The *SE(2)* path for the cart was constructed from the *SE(2)* path previously traversed by the base. Due to the cart being a rigid body and the rigid grasping of the cart's handlebar, the cart's path was transformed into *SE(3)* paths for the left and the right gripper. The (x, y) translations were provided by base's past path, as well as one degree of rotation. The height of the handlebar was fixed, and the rigid grasp covered the remaining 2 degrees of rotation. To avoid extending the arms and avoid collision between the cart and the base of the robot, the cart's path started at a predetermined fixed arc-length distance between the base's current position and a point among its previously traversed positions. Finally, the same timing law imposed on the base was used for the end-effector paths.

Upon grasping of the handlebar, the initial path between cart and robot was set to be a Bezier curve that smoothly connects the cart with the base. This path was subsequently updated with base's positions.

5) Results: The implementation was not finished in time for the challenge day, on which simulation results were shown within the presentation of the team's approach. However, the proposed solution was pursued further, with additional lab visits, for the purpose of a standalone publication which is in progress at the time of writing.

IV. Conclusions and Lessons Learned

The first WARA Robotics Mobile Manipulation Challenge proposed a highly unstructured lab automation use-case and gave participating teams the freedom to choose both the robotic platform and the solution strategy. A video of the challenge day is available online[2].

Teams from Politecnico di Milano (PoliMi) and Örebro University (ÖrU) addressed the manipulation sub-task (Section II-*B*), while Lund Technical University (LTH) focused on the carting sub-task (Section II-*A*). The Royal Institute of

[2]https://youtu.be/F9PRgmFeArM?si=dyfY8xrQoy5UWdw6

Technology (KTH) attempted the full challenge, integrating both sub-tasks.

A jury composed of employees from ABB and AstraZeneca evaluated the teams and named Örebro University the winner, awarding a 50,000 SEK participation bonus to the 2025 Automatica[3] Exhibition in Munich.

While the challenge format promoted creativity and flexibility, it also revealed several technical and methodological gaps in current academic approaches to mobile manipulation:

- Whole-Body Control for Redundant Robots: The need for torque-level interfaces was evident, especially in tasks requiring fine force control and coordination across multiple degrees of freedom. Lack of such access limited the ability to fully leverage the mechanical capabilities of redundant platforms like mobile manipulators.
- Perception and Grasping: Robustness in perception remains a key challenge. Cluttered environments, occlusions, and the presence of transparent objects exposed the fragility of current object detection and grasp planning pipelines. Additionally, the robotic platforms available in the WARA Robotics lab are equipped with parallel-jaw grippers, which impose significant limitations when performing general-purpose manipulation tasks.
- Task-Level Policies: Teams that employed Behavior Trees demonstrated better failure recovery and execution flexibility. This highlights the need for modular and reactive task-level frameworks capable of handling perception and actuation uncertainties without human intervention.
- Navigation with Passive Objects: The carting task underscored the limitations of traditional planners, which do not account for the maneuverability constraints introduced by pulling or pushing passive objects. Future planning systems must integrate these physical dynamics to ensure feasible and efficient motion strategies.

Additionally, the open-ended nature of the challenge made cross-team comparison difficult. Future editions will introduce standardized platforms, digital twin environments, and clearer evaluation metrics to promote reproducibility and fair assessment. For example, we will evaluate cart maneuverability (i.e., whether the robot can steer the cart through both left and right turns and dock it correctly), failure rates in perception, grasping, and insertion, the seamless transition between subtasks, and overall execution time. Moreover, to encourage collaboration, we will promote code sharing by awarding points to teams that upload their solutions to the challenge code repository during the development phase.

In conclusion, the challenge surfaced critical research directions while reinforcing the effectiveness of competition-based evaluation in bridging academic research with industrial relevance. This is particularly timely given recent trends in automating complex tasks in human-populated environments, such as chemical laboratories, where robots must operate safely and reliably alongside human workers.

[3]https://automatica-munich.com/en/trade-fair/

REFERENCES

[1] R. Wang, S. Xu, C. Dai, J. Xiang, Y. Deng, X. Tong, and J. Yang, "Moge: Unlocking accurate monocular geometry estimation for open-domain images with optimal training supervision," 2024. [Online]. Available: https://arxiv.org/abs/2410.19115

[2] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," *arXiv*, 2024. [Online]. Available: https://arxiv.org/abs/2408.00714

[3] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "Foundationpose: Unified 6d pose estimation and tracking of novel objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 868–17 879.

[4] N. Hogan, "Impedance control: An approach to manipulation," in *1984 American Control Conference*, 1984, pp. 304–313.

[5] M. Colledanchise and P. Ögren, "Behavior trees in robotics and AI: an introduction," *CoRR*, vol. abs/1709.00084, 2017.

[6] N. Lucci, E. Montini, I. Zappa, A. M. Zanchettin, and P. Rocco, "Intuitive cobot programming for small-medium enterprises," in *European Robotics Forum*. Springer, 2024, pp. 241–246.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[8] L. Kang, J.-T. Seo, S.-H. Kim, W.-J. Kim, and B.-J. Yi, "Design and implementation of a multi-function gripper for grasping general objects," *Applied Sciences*, vol. 9, no. 24, p. 5266, 2019.

[9] M. Aqib, A. Imran, K. Khan, M. Arsalan, S. Manzoor, K. Long, and B.-J. Yi, "Design and implementation of shape-adaptive and multifunctional robotic gripper," *Journal of Field Robotics*, vol. 41, no. 1, pp. 162–178, 2024.

[10] M. Mason, "The mechanics of manipulation," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2. IEEE, 1985, pp. 544–548.

[11] A. Rodriguez, M. T. Mason, and S. Ferry, "From caging to grasping," *The International Journal of Robotics Research*, vol. 31, no. 7, pp. 886–900, 2012.

[12] Y. Dong and F. T. Pokorny, "Quasi-static soft fixture analysis of rigid and deformable objects," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6513–6520.

[13] L. Zhang, R. Zapata, and P. Lépinay, "Self-adaptive monte carlo localization for mobile robots using range sensors," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 1541–1546.

[14] S. Macenski, M. Booker, and J. Wallace, "Open-source, cost-aware kinematically feasible planning for mobile and surface robotics," *arXiv preprint arXiv:2401.13078*, 2024.

[15] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su *et al.*, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," in *European Conference on Computer Vision*. Springer, 2024, pp. 38–55.

[16] "Rws." [Online]. Available: https://developercenter.robotstudio.com/api/RWS

[17] J. Styrud, M. Iovino, M. Norrlöf, M. Björkman, and C. Smith, "Automatic behavior tree expansion with llms for robotic manipulation," in *2025 International Conference on Robotics and Automation (ICRA)*, 2025.

[18] M. Iovino, E. Scukins, J. Styrud, P. Ögren, and C. Smith, "A survey of behavior trees in robotics and ai," *Robotics and Autonomous Systems*, vol. 154, p. 104096, 2022.

[19] C. Mastalli, R. Budhiraja, W. Merkt, G. Saurel, B. Hammoud, M. Naveau, J. Carpentier, L. Righetti, S. Vijayakumar, and N. Mansard, "Crocoddyl: An efficient and versatile framework for multi-contact optimal control," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2536–2542.

[20] A. Dahlin, "Reactive motion planning and control under constraints," Ph.D. dissertation, Chalmers Tekniska Hogskola (Sweden), 2023.

[21] L. Huber, A. Billard, and J.-J. Slotine, "Avoidance of convex and concave obstacles with convergence ensured through contraction," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1462–1469, 2019.