Robotic Grasping and Manipulation Competition at the 2024 IEEE/RAS International Conference on Robotics and Automation

By Yu Sun, Berk Calli, Kenny Kimble, Francis wyffels, Victor-Louis De Gusseme, Kaiyu Hang, Salvatore D'Avella, Alessio Xompero, Andrea Cavallaro, Maximo A. Roa, Jose Avendano, and Anastasia Mavrommati

OVERVIEW

The Ninth Robotic Grasping and Manipulation Competition (RGMC) took place in Yokohama, Japan, during the 2024 IEEE/RAS International Conference on Robotics and Automation (ICRA). The series of RGMC events started in 2016 at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) with strong support from the conference's organization committee, and since then they have been held each year at ICRA or IROS [1]. Across the editions, RGMC engaged the community in solving the open challenges associated with various robotic grasping and manipulation tasks for manufacturing, service robots, and logistics, and in advancing research and technology towards more realistic scenarios that can be encountered in daily activities at home or in warehouses. These tasks include assembling and disassembling boards, hand-in-hand grasping, picking and placing various objects, pouring liquids into a cup, bin picking, rearranging and setting formal tables, folding and unfolding cloths, and receiving objects handed over by a person. The goal of RGMC across these tasks is to assess the autonomous manipulation capabilities of a robotic arm when dealing with unknown or novel objects with varying physical properties and when handling scenarios with various degrees of uncertainty caused by a cluttered scene, random initial configurations, or human behaviors when interacting with the robot. For example, objects can vary in their shapes, appearances, transparency, deformability, and weight. RGMC challenges participants to design solutions that can perform outside of their labs and controlled conditions, within a limited time, and in some cases adapt to changing conditions within a task. For this year's event in Yokohama, RGMC organized five tracks grouped into two categories: the manufacturing track and four essential skill tracks. The essential skill tracks include cloth manipulation, in-hand manipulation, picking in clutter, and human-to-robot handovers. All tasks in the tracks are performed autonomously by the robot, without any human intervention. The tasks in all the tracks were co-developed based on the benchmarks published in the literature by the organizers and the benchmarks' authors [2], [3], [4]. Table 1 lists the organizers of this year's RGMC which attracted the interest of over 70 teams worldwide across the various competition tracks. Among the registered teams and through a reviewing process, 31 teams from 10 different countries were selected to participate in Yokohama. Only a fraction of these qualified teams could commit and complete the tracks on site, with 11 teams competing for the cloth manipulation track. Only 20 teams could commit and complete the tracks on-site, with three teams participating in three tracks and two teams participating each in two tracks. Table 2 provides the list of tracks, the reference benchmarking protocol, and the teams who competed on-site for each track.

Digital Object Identifier 10.1109/MRA.2024.3481609
Date of current version 12 December 2024

TRACKS AND SETUP

RGMC extended to five competition tracks, achieving the largest participation of the last nine editions. The cloth manipulation track achieved the first-ever large-scale out-of-the-lab benchmark for robotic cloth unfolding. For the first time, the competition included the human-to-robot handover track, which involves a human in the task, leading to higher uncertainty and safety concerns when controlling the robot. RGMC selected the competition tracks to favor the reusability and adaptation of previously defined benchmarking protocols to the on-site and time-limited competition beyond in-lab benchmarking. The picking in clutter track adapted the protocols of the cluttered environment picking benchmark (CEPB) [2]. The human-to-robot handover track adapted the protocol of the CORSMAL benchmark [3]. The in-hand manipulation track adapted the protocol of benchmarking in-hand manipulation [4].

The tracks ran for three days from Tuesday, 14 May, until Thursday, 16 May, attracting the audience of the conference to stop and observe the teams performing their tasks. Figure 1 shows the large audience engagement during the competition.

SETUP

To make the competition more inclusive, RGMC also provided the availability of two robotic arms thanks to the sponsorship of Universal Robots (UR), and travel expense support for shipping the equipment to selected teams. Most of the tracks had teams mounting their own robotic and perception systems, whereas

 1070-9932/24©2024IEEE

the cloth manipulation track required teams to interface their solutions with the already existing pipeline controlling the two available robotic arms to execute the task. The human-to-robot handover track had one team bringing and mounting their own setup, and the other two teams made use of the available robotic arms. The manufacturing track had teams using two robotics arms with different specialized end effectors. The picking in clutter track had two teams using multimodal grippers to maximize the grasping success rate in the competition and one team that used a reconfigurable end effector. The in-hand manipulation track had teams using their own robotic hands. To

TABLE 1. A list of organizers and their roles for the Ninth Robotics Grasping and Manipulation Competition.

NAME	AFFILIATION	ROLE
Yu Sun	University of South Florida (USA)	Contact and communication
Berk Calli	Worcester Polytechnic Institute (USA)	Essential skill tracks lead
Maximo A. Roa	German Aerospace Center (DLR) (Germany)	Logistics and competition management
Jose Avendano	MathWorks	Logistics and competition management
Anastasia Mavrommati	MathWorks	Logistics and competition management
Kenny Kimble	National Institute of Standards and Technology (USA)	Manufacturing track organizer
Francis wyffels	Universiteit Gent–imec (Belgium)	Cloth manipulation track organizer
Victor-Louis De Gusseme	Universiteit Gent–imec (Belgium)	Cloth manipulation track organizer
Kaiyu Hang	Rice University (USA)	In-hand manipulation track organizer
Podshara Chanrungmaneekul	Rice University (USA)	In-hand manipulation track organizer
Salvatore D'Avella	Sant'Anna School of Advanced Studies (Italy)	Picking in clutter track organizer
Andrea Cavallaro	Idiap Research Institute and École Polytechnique Fédérale de Lausanne (Switzerland)	Human-to-robot handovers track organizer
Alessio Xompero	Queen Mary University of London (U.K.)	Human-to-robot handovers track organizer

TABLE 2. Tracks of the 9th Robotic Grasping and Manipulation Competition with their reference benchmarking protocols and teams that competed and completed the track on-site in Yokohama, Japan.

RGMC TRACKS	REFERENCE BENCHMARKING PROTOCOL	TEAMS WHO COMPETED ON-SITE
Manufacturing	-	1. Tsinghua 3C United, Tsinghua University, China 2. AI & Robot Lab, Tsinghua University, China 3. JAKS, Kanazawa University, Japan
Cloth Manipulation	-	1. AIR-jnu, South Korea 2. Team Ljubljana 3. Ewha Glab 4. SCUT-ROBOT 5. Team Greater Bay 6. Samsung Research China - Beijing, China 7. Shibata Lab 8. AI & Robot Lab, Tsinghua University, China 9. UOS-Robotics 10. AIS Shinshu 11. 3C1S
In-hand Manipulation	Benchmarking In-Hand Manipulation [4]	1. XL team, Tsinghua University, China 2. AI & Robot Lab, Tsinghua University, China
Picking in Clutter	Cluttered Environment Picking Benchmark [2]	1. AIDIN-ROBOTORY from Sungkyunkwan University and AIDIN ROBOTICS Inc., China South Korea 2. THU-bot, Tsinghua University, China 3. THUDA, Tsinghua University, China 4. TCS Smart Machines, TCS Research Noida, India 5. Intelligent Control and Robotics, China
Human-to-Robot Handover	CORSMAL Benchmark [3]	1. SirsIIT, Italy 2. Air-jnu, South Korea 3. Smart Machines TCS Research, India

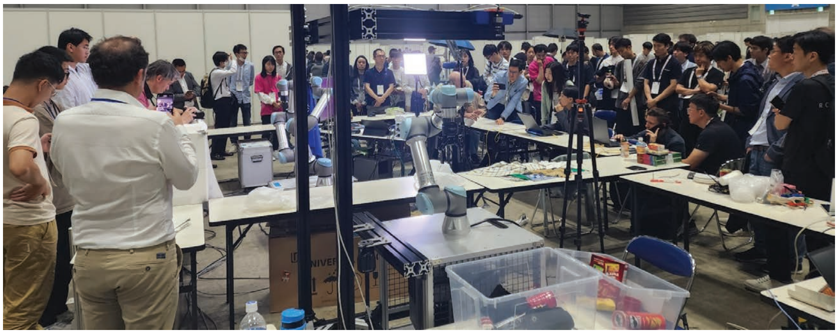


FIGURE 1. Large groups of ICRA 2024 attendees were attracted by the Robotic and Grasping Competition throughout the four-day period.

ers worldwide, and they most often produce in batches, with product variation from one batch to another. To support production in such an environment where robot expertise is often limited and cost is always a factor, robot systems must be easy to deploy and reconfigure with minimal retooling. This competition challenges teams to develop robot systems that are easy to deploy and program with the goal of handling the same small batch assembly operations present in the manufacturing industry (Figure 2).

encourage on-site participation and flexibility of the solutions in terms of both hardware and software across the tracks, teams were allowed to bring their own hardware and setup for the competition, including robotic arms, robot hands and various end effectors, and touch sensors.

PARTICIPATION

RGMC at ICRA 2024 attracted the interest of over 70 teams worldwide across the various competition tracks: nine teams for the manufacturing track, 11 teams for the cloth manipulation track, nine teams for the in-hand manipulation track, 15 teams for the picking in clutter track, and nine teams for the human-to-robot handovers track. The teams registered for the competition from all around the world including China, the United States, New Zealand, Japan, Ireland, Switzerland, Germany, Italy, the United Kingdom, India, and South Korea. Each team could register with a maximum of five members, including university professors, graduate and undergraduate students, and industry professionals. Among the registered teams and through a reviewing process, 31 teams from 10 different countries were selected to participate in Yokohama. Several of these qualified teams withdrew later on due to travel difficulties. Other teams that qualified for the competition dropped their on-site participation as they faced challenges such as insufficient performance of the implemented solution, lack or poor transfer of training manipulation models from simulation to the real robot, and difficulties in bringing their own hardware to the venue.

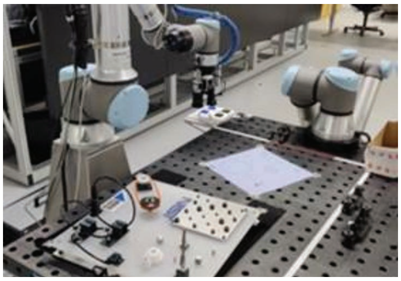


FIGURE 2. Robotic setup for the manufacturing track.

TRACK DETAILS AND RESULTS

MANUFACTURING TRACK

Small and medium-sized enterprises represent the majority of manufactur-

TASK

Given a board of manufacturing inspired tasks, such as pegs into holes, nuts threaded on screws, and belts routed through pulleys, the task is to assemble each part on one quadrant of the board at a time and then disassemble the parts into a bin within a time limit and having limited prior information about the board and parts. Before the execution of the task, parts are randomly placed with respect to the initial configuration, in terms of both position and orientation. Figure 3 shows an example of the task at three different states: initially disassembled, complete assembly, and disassembled.

RULES AND SCORING

The competition board had mostly similar parts to a practice board given to teams months in advance, but the part locations and orientations had been changed. Additionally, the competition board and kit mat, with

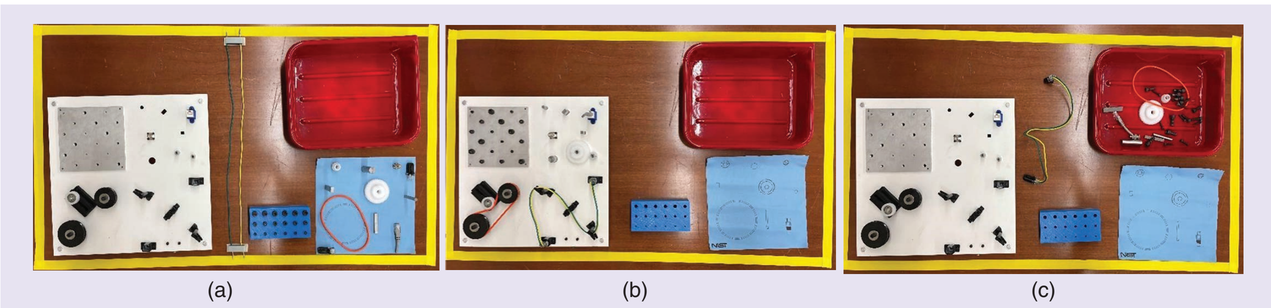


FIGURE 3. An example of the manufacturing track task where the robot needs to assemble each part on one quadrant at a time of the given board and then disassemble the parts into a bin. (a) Initial state, disassembled. (b) Complete assembly. (c) Final state, disassembled.

predetermined part locations, were placed at random offsets from the start location at time zero. Teams were tasked with rapidly reprogramming their systems to assemble each part on the board one quadrant at a time and then disassemble the parts into a bin. Teams had four hours to complete the task of the track, including 130 min for assembly and 90 min for disassembly. Points were earned on a per-part basis with partial points available for partial assemblies (Table 3).

TABLE 3. The performance and ranking of the teams for the manufacturing track.

	MANUFACTURING TRACK SCORES			
TEAM	ASSEMBLY	DISASSEMBLY	TOTAL	RANK
Tsinghua 3 C united team, China	31	48	79	1
JAKS, Japan	30	35	65	2
AI & Robot, China	22	39	61	3

HIGHLIGHTS

Teams were encouraged to solve the problem through whatever means necessary as there was no standard for hardware or software that must be used. To this end it was encouraging to see that every competing team this year utilized two robot arms with different specialized end effectors, resulting in higher scores than previous years. This shows that the technology is shifting to meet the needs of industry. This fact is further reinforced by the first year of funding for this track coming from the ASTM standards organization due to the inspiration from standards in development under the F45.05 subcommittee.

CLOTH MANIPULATION TRACK

Cloth manipulation skills are essential for diverse real-world applications, and cloth unfolding is a crucial initial step for various downstream tasks such as folding, hanging, ironing, and assisted dressing.

TASK

Given a cloth item randomly placed on a flat surface and then grasped and held in the air by one robotic arm, the robot needs to localize good grasping points in the color and depth images of the hanging cloth item to complete the unfolding procedure with the second robotic arm until the garment is fully stretched in the target state. Examples of cloth items are towels, shirts, and shorts. The cloth item used in the task can be known in a given image dataset or unknown. Figure 4 shows the cloth unfolding procedure designed for the track.

RULES AND SCORING

All participants used a shared dual-arm setup with two UR5e robots equipped with Robotiq 2F-85 grippers and a ZED 2i RGB stereo-depth camera. Participants focused on the critical step of grasp pose selection (step 4) once the cloth was lifted. Other steps, such as initialization and final stretching, were handled by the organizers. Grasp quality was evaluated by calculating the surface area of the stretched result. Coverage, a ratio between 0 and 1, was determined by dividing this surface area by the cloth's maximum possible surface area. Participants were scored based on their average coverage over 16 unfolding attempts. Evaluation items included towels and shirts, with half from the training dataset and half unseen.

RESULTS

Table 4 reports the performance scores of the teams and ranks the teams based on their average coverage over 16 unfolding attempts. Scores for the grasp success rate and coverage given successful grasps are reported in addition to the average coverage. The top-scoring team achieved an average coverage of 0.6. The results indicate that a high grasp success rate is necessary but not sufficient for achieving the best coverage performance. The two best-performing teams achieved a coverage given successful grasps of 0.73 and 0.69, highlighting the effectiveness of their grasp selection strategies for unfolding. These results are lower than the average coverage reported in previous works, e.g., 0.8 by FlingBot [5] and 0.85 by UnFoldIR [6], and show the importance of independent

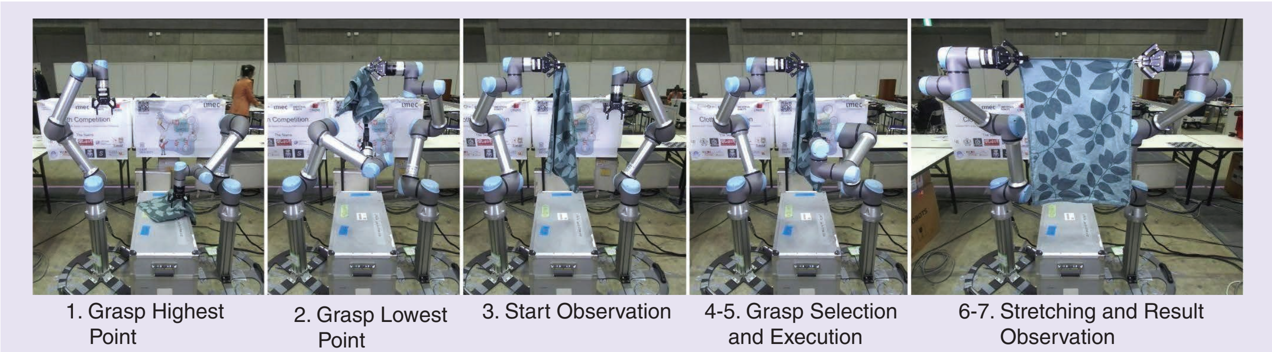


FIGURE 4. The seven-step unfolding procedure in the cloth manipulation track. The image depicts a successful attempt during the live evaluation, resulting in a perfect coverage score of 1.0. Steps 1 and 2 compose the initialization procedure, while step 3 shows the input RGB data provided to the participants. Steps 4 and 5 show the grasp selected by the participant and its execution. Steps 6 and 7 show the result after the final stretching motion.

evaluations outside controlled lab environments. Moreover, the strong performance of traditional methods used by the first- and third-best-performing teams show that learning-based methods, even if promising, may currently be hindered by the challenges and costs of collecting large datasets with real robot executions.

IN-HAND MANIPULATION TRACK

In-hand manipulation is an essential skill for robots to physically engage themselves into real-world tasks, especially in tasks where small-scale object reconfigurations within grasps are desired. Although various perspectives involved in this problem have been extensively investigated, including hand design, planning, control, state estimation, and learning, a robust solution still does not exist in any real systems. The goal of this competition is to provide a standardized task setup for researchers to share their solutions, while exchanging ideas on the challenges, opportunities, and potential next steps of in-hand manipulation to facilitate the advancement of relevant research.

TASK

Given an object already grasped by a robot hand, the robot hand needs to manipulate the object according to a sequence of predefined target positions (Task A) or rotate the object according to a sequence of specified orientations (Task B). For Task A, the target positions are defined in the camera's frame relative to the initial position of the object, as marked and tracked by an AprilTag. This way, the specification of target positions is system-agnostic in terms of the hand model, initial grasp, camera mounting pose, etc. For task B, the target orientations are marked and tracked by multiple AprilTags on each face of a cube object to avoid occlusions, while also being specified in the camera's frame to ensure they are system-agnostic. The task design was inspired by a published benchmark protocol for in-hand manipulation [4].

RULES AND RESULTS

Every team was required to bring their own robot hand for the competition. The hand needs to be installed on a stationary mount and the manipulated object is observed by a camera installed at a location chosen by the team. An example setup is shown in Figure 5. Object motions are required to be purely executed by the actions of the robot hand, and any motions by a "wrist" or "arm," in any form, to manipulate the object is strictly prohibited. When switching from task A to task B, teams are allowed to modify their hand setup as needed, such as remounting the hand in a different pose, or changing the location of the camera. It is worth noting that, although the object is manipulated in-hand, it is not required that the object is always stably grasped, as palm support can be leveraged to keep the object in-hand while it is manipulated by finger actions. Both tasks were evaluated by the accuracy, success rate, and total runtime. Table 5 shows the results of the teams for each task.

HIGHLIGHTS

Within the nine registered teams, six teams were qualified and worked for months to prepare for the competition. Among the qualified teams, only one team worked on both hand and algorithm designs, while all other teams used commercial hand models or open-source hand models that can be 3D printed. A few days before the competition, however, four teams decided to drop out at the very last minute as they were not able to secure a robust enough solution to compete on-site. Being one of the most challenging problems in robot manipulation, differently from many other tasks, in-hand manipulation often struggles to even finish a task, since many solutions can lose the object

TABLE 4. The cloth manipulation track scores and ranking of the teams.

TEAM NAME	METHOD NAME	GRASP SUCCESS	COVERAGE GIVEN GRASP SUCCESS	COVERAGE
Air-jnu	Intuitive grasping determination	69%	0.73	0.60
Team Ljubljana	CeDiRNet-6DoF	63%	0.69	0.57
Ewha Glab	Sharp edge detection	94%	0.57	0.55
SCUT-ROBOT	KeypointDetr	88%	0.58	0.53
Team Greater Bay	CopGNN	75%	0.6	0.53
Samsung Research China–Beijing	Grasp-Cloth	63%	0.59	0.48
Shibata Lab	Densenet Method	56%	0.57	0.46
AI & ROBOT LAB	CFAN	63%	0.57	0.45
UOS-Robotics	Affordance Edge Detection	19%	0.75	0.39
AIS Shinshu	Depth2Grasp-CNN	88%	0.37	0.37
3C1S	Grasp with PointNet-VAE	6%	0.91	0.35

Teams are ranked based on the coverage scored. The cloth track score table uses a color scale where darker green represents better performance. Outliers are displayed in gray to minimize their visual impact on the color distribution of the other scores.

in the middle of the task and cannot recover. We later learned from the dropped teams that all of them encountered the problem of "Sim-to-Real" gaps and failed to transfer their working solution trained in simulation to the real-world, indicating that end-to-end solutions are not yet able to prove feasibility for real-world in-hand manipulation. Nevertheless, it is interesting to note that the winning team's solution (XL team) also partially relied on simulation-based learning. However, their solution has wrapped the learned policy, which is effective for real-time manipulation planning, with a model-based control scheme to significantly improve the robustness of their solution. In the competition, their solution succeeded in all tasks and impressively achieved state-of-the-art performance.

PICKING IN CLUTTER TRACK

Pick and place is one of the most repetitive tasks for human workers and one of the golden standard tasks used to assess the capabilities of manipulator robots. Even today, in many warehouses, a lot of human pickers stand in front of a shelf all the time and repeatedly pick objects from that shelf to place them into bins. The automation of pick and place actions is important for industries to increase throughput while lowering expenses. On the other hand, it is a challenging problem for the robotics research community. Even if it is a simple task for a human, depending on the boundary conditions, it could be difficult for a robot to pick and place heterogeneous objects with a single end effector, especially when they are in a cluttered environment. The picking in clutter track is an adaptation of one of the tasks involved in the cluttered environment picking benchmark (CEPB) [2], which proposes several industrial-oriented task protocols spanning from tasks on individual objects for evaluating targeted components (i.e., perception, planning, and control) of the grasping system to heavily cluttered scenes that are meant to examine most of the recent needs of such an industrial revolution (Industry 4.0) and beyond (Industry 5.0), ranging from flexible automation and generalizable grasping to cluttered environments. The experimental setup, the procedures, and the evaluation metrics have been designed to aim at reproducibility without constraining the scenario and allowing the comparison among research groups through a novel evaluation metric that guarantees a fair comparison even with cluttered scenes.

TASK

Given a set of 20 known and unknown objects placed randomly within a clear box, the robotic arm needs to pick each object and place it into a second clear box within a maximum allowed time to fully clear the first box. The objects are heterogeneous, presenting different physical properties to pose difficulties to the vision component and the

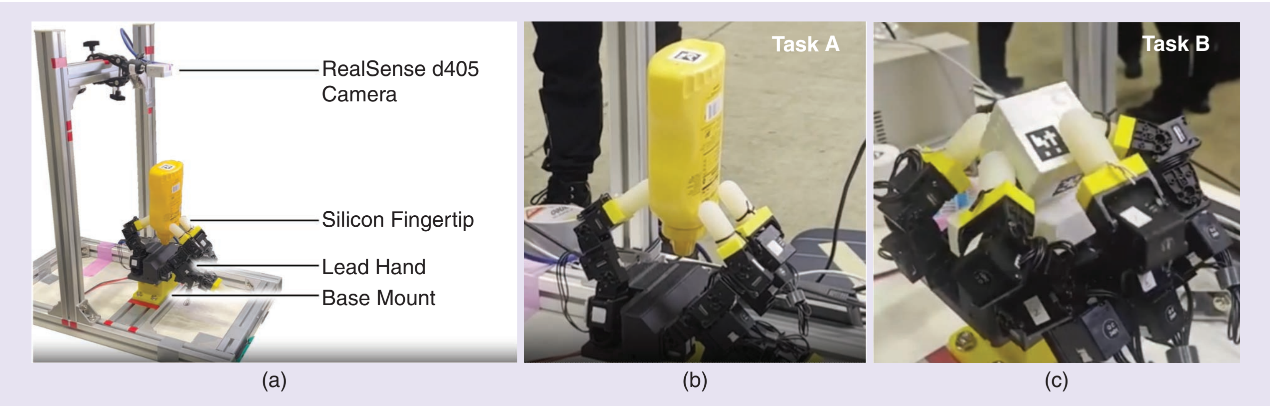


FIGURE 5. (a) An example setup of the in-hand manipulation competition. (b) Task A: the robot hand is required to move the grasped object to specified positions tracked by an AprilTag. (c) Task B: the robot hand is required to rotate a grasped cube object to specified orientations. The images were provided by the XL team.

TABLE 5. Results and ranking of the teams for the in-hand manipulation track.

	TASK A (CYLINDER OBJECT)			TASK A (NOVEL OBJECT)			TASK B	
TEAM	ERROR (CM)↓	WAYPOINTS REACHED↑	RUNTIME (S)↓	ERROR (CM)↓	WAYPOINTS REACHED↑	RUNTIME (S)↓	SUCCESS↑	RUNTIME (S)↓
XL Team, China	**0.54**	10	103.26	**0.62**	10	109.36	10	81.18
AI & ROBOT LAB, China	0.64	10	105.04	1.29	10	94.29	10	44.75

Bold values indicate the best performance per criterion.

grasping part of the manipulating system, and the problems are accentuated in the cluttered conditions of the bin-picking scenario.

PROTOCOL

The CEPB benchmark[1] has a modular design and is organized in *stages* with one or more *intermediate phases* and a *final phase test*. *Stages* are meant to represent an industrial-relevant task, which is identified by the *final phase test*, while the *intermediate phases* of each *stage* aim at evaluating a specific subproblem of the manipulation task. In particular, *stage 1* is for picking and placing of nonsequential objects in cluttered environments; *stage 2* concerns picking and placing of nonsequential unknown objects in cluttered environments; and *stage 3* focuses on picking and placing of sequential objects in cluttered environments. The set of 40 objects are selected from established datasets, such as YCB, ACRV-APC, and T-LESS, to challenge the end effector and the perception system. They present different levels of difficulty determined using a consensus protocol disseminating questionnaires among different colleagues working in the field. The total level of difficulty takes into account the gripper typology among parallel-jaw, suction, and soft grippers and the vision. Indeed, the objects can vary in size, shape, and weight and have diverse surface materials and texture properties. There are objects with reflective, perforated, or symmetric surfaces that are challenging for the vision; others have deformable surfaces or strong orientation constraints and shift their centers of mass when manipulated. All these problems are accentuated in the clutter because accurate segmentation and stable grasp are more difficult. The objects are subdivided into four subsets of ten elements each having an increasing overall difficulty. For every subset, the mean difficulty should be the same for all the gripper typologies (i.e., parallel-jaw, suction, and soft grippers) considered in the questionnaire, thus preventing the choice of a particular gripper in favor of the others, relying on the nature of the objects. The objective of the track was to perform a bin-picking task of known and novel objects in a cluttered scene similar to the *final phase* of *stage 2* of CEPB on *subset 1* plus *subset 4* (20 objects in total), where the four non-YCB objects in *subset 4* [i.e., 1) a plastic white cup, 2) a Rolodex jumbo pencil cup, 3) a glove, and 4) a joke book] were exchanged with novel objects belonging to the same object category [i.e., 1) a bigger plastic red cup, 2) a smaller squared pencil cup, 3) a wool hat, and 4) an issue of *IEEE Robotics & Automation Magazine*] and announced 1 h before the start of the competition to test the generalization capabilities of the system.

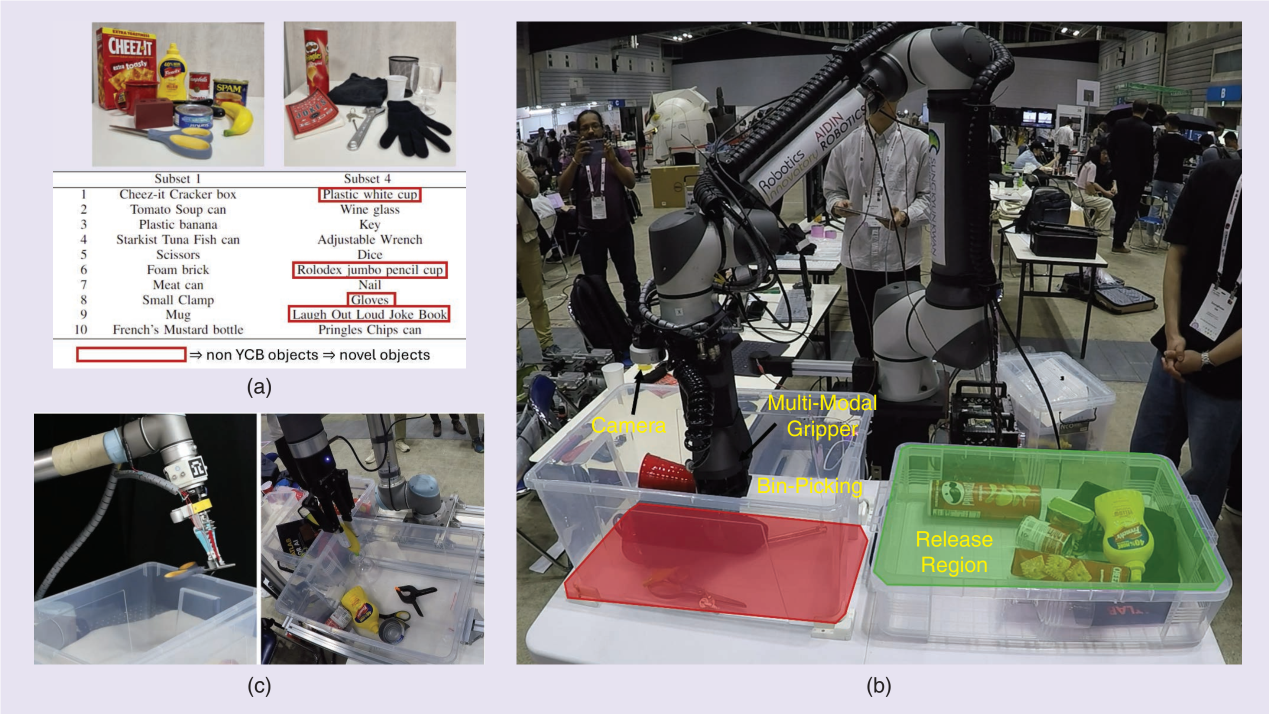
RULES

Each team had to bring and mount their own setup. Teams could decide their own setup according to the properties of their grasping system. They only had to declare in advance the picking and release regions, which could not be changed afterward on the fly. The task was repeated three consecutive times according to the following sequence of actions. A team member has to shake the objects in the smaller clear box (object 42 of CEPB) per subset (subset 1 and subset 4 separately) and throw them in the bigger clear box (object 43 of CEPB), starting from subset 1. Then the team member has to put a clear box in the designated release region and the judge activates the timer while another team member starts the system. After that, no human intervention or teleoperation is possible. The trial ends, and the judge stops the timer if one of the following four conditions holds: 1) the bigger clear box has been emptied after the last pick and place movement; 2) the system is stuck and unable to proceed, failing to grasp multiple times the exact same object, or in general, it does several consecutive fail grasps of different objects; 3) the maximum time (26 min) is reached; or 4) the system decides to quit earlier by going into the quit pose (e.g., the robot points at the ceiling) the team showed to the judge before the start of the task. Hardware faults (e.g., power drop or cable disconnections) that stop the system at the beginning of the trial or during the trial can be tolerated, resetting the timer and repeating the trial. If the system is blocked due to a collision with the environment, the team is allowed to reactivate it through the robot controller, but the timer must not be paused during this recovery operation. The maximum time is computed considering 40 s per double the number of objects, thus giving the chance to grasp a single object two times. An object is successfully grasped if it is completely placed in the release region. The evaluation metric is the average of the complexity of the object successfully grasped, taking into account the difficulty of the clutter normalized by the spent time [2]. To compute the clutter difficulty the team must collect a top-view picture of the starting configuration of the objects. Figure 6 shows the objects involved in the competition, an example of team setup, and some grasping attempts exploiting different grippers.

RESULTS AND HIGHLIGHTS

Bringing the system and setting it up on site was not easy for some teams. The illumination of the environment was different from laboratory conditions for almost all of the teams and affected the perception pipeline (Figure 7). However, the improvisation and adaptability skills of the teams made things work on site using umbrellas or external blockers. Teams who used a hybrid gripper, i.e., a gripper that was able to provide multiple grasping modalities, or a multimodal grasping solution on average performed better as they were able to handle the different properties of the objects and the object configurations in the clutter. Indeed, teams with a traditional gripper (parallel jaw) had difficulties grasping objects in a specific pose and location in the bin. Most of the teams had problems grasping very small objects like nails and keys or tricky objects like the magazine. In general, the introduction of novel objects caused some troubles not for the generalization of the vision part but rather for the final grasp (gripper–object

[1]http://cepbbenchmark.eu/.



	Subset 1	Subset 4
1	Cheez-it Cracker box	Plastic white cup
2	Tomato Soup can	Wine glass
3	Plastic banana	Key
4	Starkist Tuna Fish can	Adjustable Wrench
5	Scissors	Dice
6	Foam brick	Rolodex jumbo pencil cup
7	Meat can	Nail
8	Small Clamp	Gloves
9	Mug	Laugh Out Loud Joke Book
10	French's Mustard bottle	Pringles Chips can

FIGURE 6. (a) The objects involved in the competition (subset 1 and subset 4 of CEPB). (b) An example of a team setup during the competition on site. (c) Examples of grasping with different grippers during the competition.

interaction). Another challenge for many teams was grasping objects in the corner or near the wall of the bin. Indeed, a lot of hardware stops for collision detection were registered. Some grippers were designed for a power grasp that helped even with a nonoptimal grasp but damaged or broke some fragile objects like the plastic cup or the Cheez-It cracker box. Two teams implemented decluttered actions that helped to disentangle tricky configurations. Other teams developed a logic to understand if an object was correctly grasped to avoid wasting time while performing the release movement uselessly. Overall, teams had creative approaches to combine in a single pipeline the control of personalized hardware implementations, planning approaches, and visual detection. Table 6 summarizes the results achieved by the five teams during the three trials.

HUMAN-TO-ROBOT HANDOVERS TRACK

The real-time estimation through the vision perception of the physical properties of objects manipulated by humans is important to inform the control of robots and perform accurate and safe grasps of objects handed over by humans. However, estimating the physical properties of previously unseen objects using only inexpensive cameras is challenging due to illumination variations, transparencies, reflective surfaces, and occlusions caused by both the human and the robot. The essential skill track for dynamic human-to-robot handovers is based on an affordable experimental setup that does not use a motion capture system, markers, or prior knowledge of specific object models. Figure 8

FIGURE 7. Difficulties with illumination conditions on-site, different from lab conditions, were resolved with several strategies, e.g., using umbrellas to decrease the illumination intensity.

illustrates such a setup. The track focuses on food containers and drinking glasses that vary in shape and size and may be empty or filled with an unknown amount of unknown content. The goal is to assess the generalization capabilities of the robotic control when handing over previously unseen objects filled (or not) with unknown content, hence with a different and unknown mass and stiffness. No object properties are initially known to the robot (or the team), which must infer these properties on the fly during the execution of the dynamic handover, through the perception of the scene.

TASK

Given an unknown container with an unknown content and content amount and the container placed on a tabletop, a person grasps the object and approaches the robot placed on the other side of the table with the intention to hand over the container to the robot. By estimating the physical properties of the container and the intention of the person, the robot should simultaneously approach the person's hand to receive the container, completing a dynamic human-to-robot handover, followed by the delivery of the container to a predefined location on the table. The task is performed within a maximum allowed time to avoid assistive behaviors or making the handover a simpler static grasping task.

REVISED PROTOCOL AND SCORING

The track reuses the CORSMAL benchmark [3] and adapts its protocol for the in-person competition. In short, the CORSMAL benchmark[2] uses four rotationally symmetric cups that are either empty or full of rice as content for safety reasons when executing the handover tasks. The protocol defines 288 handover configurations that depend on the number of volunteers (four), the number of cups, the two content levels, three different grasp types, and three handover locations. The benchmark also defines 13 granular performance scores, split into five scores for the vision component, three scores for the robotic component, and five other scores for the overall task. A weighted average of these scores provides the overall score for the benchmark. To make the protocol more accessible for the competition while keeping the challenging goal of handing over

[2]https://corsmal.eecs.qmul.ac.uk/benchmark.html.

TABLE 6. The results and ranking of the teams for the picking in clutter track.

NUMBER	TEAM NAME	FINAL SCORE	TRIAL SCORE: FIRST–SECOND–THIRD	PICKED OBJECTS: FIRST–SECOND–THIRD	ELAPSED TIME: FIRST–SECOND–THIRD
1	THU-bot	**1.17**	1.03–1.11–1.37	16–19–**20**	10:00–11:03–8:35
2	AIDIN-ROBOTORY	0.94	**1.57**–0.38–0.85	**20**–8–**20**	**6:25**–18:31–15:58
3	THUDA	0.78	0.73–0.93–0.70	14–17–10	12:46–12:12–6:45
4	TCS Smart Machines	0.36	0.54–0.30–0.26	2–9–7	11:00–26:00–26:00
5	ICR USTC	0.053	0.14–0.02–0.02	11–6–6	14:14–11:02–14:01

Bold values indicate the best performance per criterion.

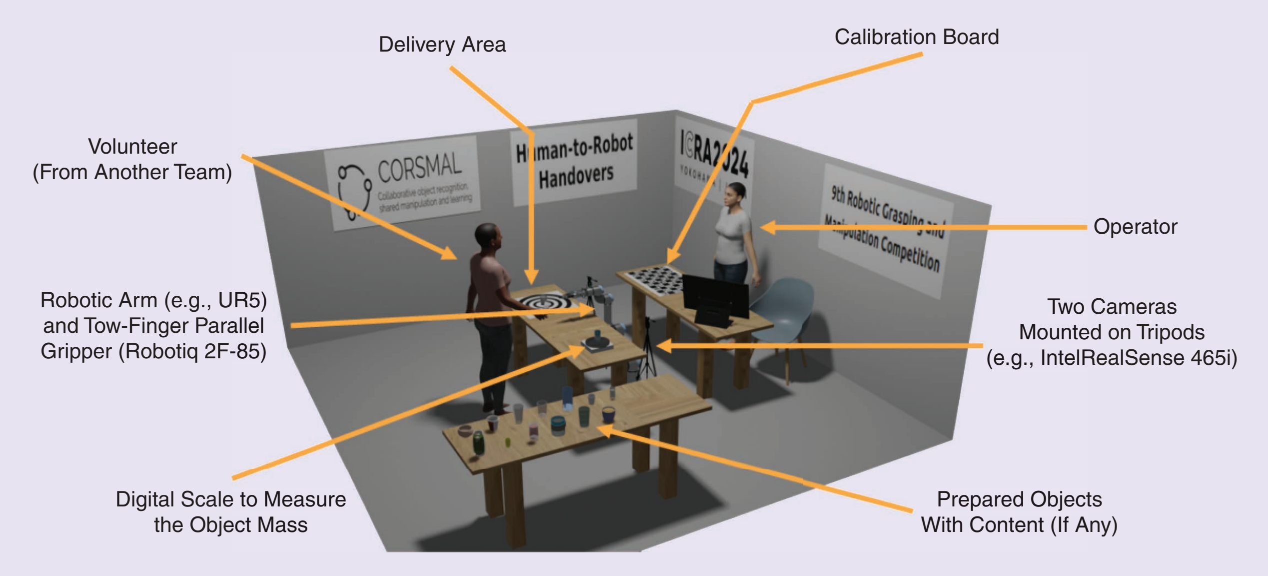


FIGURE 8. An illustration of the setup for human-to-robot handovers of unknown containers with unknown contents.

unknown containers with unknown content, we reduce and revise the number of handover configurations from 288 to 24 and the performance scores from 13 to 3. The configurations include a combination of a container, a content, and a content level. All configurations are executed by the same person with a natural grasp and approximately toward the same location in front of the robot. We select two cups from the CORSMAL benchmark set (known) and add four new containers (unknown to the teams) whose geometric shape is not necessarily rotationally symmetric. As the unknown content, we include pasta and beans in addition to rice, which is already known from the benchmark. Containers across the configurations can be either empty, half full, or full with any of the three contents. The 24 handover configurations are unknown to the teams until they are ready with their setup and software on the day of the competition. The configurations are split into groups of six configurations with increasing levels of difficulty and score: easy (5 points), medium (10 points), difficult (15 points), and hard (20 points). The level of difficulty is determined by the presence of an unknown container, unknown content, and unknown content level or by the nonvisibility of the content within the container with a camera. The points assigned to each configuration depend on the three selected performance measures from the benchmark:

1) the accuracy of the location where the container is delivered with respect to the initial target location with a maximum allowed distance of 500 mm
2) the total execution time of the handover from when the person enters in contact with the container until the robot releases the gripper at the delivery location, assuming a minimum execution time of 1 s
3) the final measured mass, in grams, of the delivered container and content, if any, with respect to the initially measured mass before the handover execution.

The total number of points (300) is divided by 3 to obtain the final 100-based point score for the competition track.

RULES

Each team should bring and mount their own setup, unless the team decided to use the available robotic arms. The setup includes a robotic arm with its end effector (e.g., a two-finger gripper), a workstation (a laptop or tower PC with a monitor), cameras with tripods and USB-C cables, a digital scale, a calibration pattern, and an ethernet cable connecting the PCs and the robotic arm. To prepare the sensing setup on site, teams need to synchronize and calibrate the cameras and calibrate the camera–robot system. Prior to the execution of the handover task, teams need to verify the behavior of the robotic arm, such as the end effector control, speed, and kinematics. For each configuration, before and after executing the handover to the robot, a member of the team needs to weigh the mass of the container and content, if any, using a digital weight scale. All configurations are executed only once by the same volunteer from another team, using the same hand and with a natural grasp of the container. The volunteer should aim to perform natural and dynamic handovers with no intention to help the robot (assistive behavior) or with no intention to make it difficult for the robot (adversarial behavior). At the beginning of a configuration, the container must be placed at a predefined initial location on the table that is not reachable by the robotic arm. Each configuration should be executed within 5 s to obtain points. Failed handover configurations include the robot not being able to grasp and/or hold the container during the delivery phase and the object falling after the robot places it on the table (0 points). To design their solutions, teams are allowed to perform calls to existing large language models and use other sensors (e.g., tactile or depth) in addition to the RGB camera(s). Any initial robot pose can be chosen with respect to the environment setup, but the volunteer is expected to stand on the opposite side of the table with respect to the robot. However, the solution cannot use prior knowledge of the objects (e.g., a prior 3D object model) or learn across executions of the configurations.

RESULTS

Team SirsIIT achieved the best-performing results with a score of 18 out of 100. Team Air-jnu achieved the second-best-performing results with a score of 15 out of 100. Team TCS followed with a score of 10 out of 100. Figure 9 shows and compares the scores obtained by the teams for each handover configuration. Most of the teams were able to complete the first 11 configurations, while two teams completed a few other configurations. Note that teams were free to choose which configurations to execute. Teams had only 30 min to execute the configurations on the competition day as the hardware and software took almost 4 h to be ready. Teams faced various challenges that included the calibration of the overall system, especially calibration of the cameras, and the illumination of the environment affecting the perception system, i.e., the recognition of the object. Two teams used the available robot on site, but they could not remotely try the setup before the competition to verify version compatibility and software and libraries to use. This aspect also impacted the preparation before the execution of the configurations.

CONCLUSION

The approach taken in this RGCM edition was to gather people from the community in a single hub that helps promote the use of existing benchmarking protocols already published in the literature. These benchmarking protocols also provide granular evaluations that help participants better understand their systems while encouraging research and assessments beyond the single overall scores limited to the competition. Based on the existing benchmarking protocols, RGMC set up five tracks, namely manufacturing, cloth manipulation, in-hand manipulation, picking in clutter, and human-to-robot handovers. Across the various tracks, the teams performed well overall, and many solutions were refreshing and inspiring. Some solutions by the participating teams, especially in the cloth manipulation track, still relied on traditional methods despite the popularity of learning-based approaches that can be hindered by the challenges and costs of collecting large datasets with real robot executions. In general, solutions based on machine learning techniques trained in simulation suffered from the reality gap during the competition on-site, and some teams even decided to not participate. Most of the teams in all the tracks required some effort to adapt to the on-site environment, different from the laboratory condition, demonstrating that the goal of a flexible and robust robotic grasping and manipulation system for the identified essential skills is not yet around the corner and showing that the competitions linked to benchmarking protocols are still an important tool to foster the research in the field. Figure 10 shows the winning teams for each track and the group picture involving participants and organizers.

What's next. The series of RGMC events received wide support from both academia and industry while also strengthening the ties between the two communities. The competitions demonstrated and assessed the state-of-the-art in the technical areas of robotic hands, grasping, and manipulation, and allowed us to identify new research problems and challenges to include within the tracks over time. The success of this competition, also demonstrated by the larger number of competition tracks, the larger number of organizers involved in the track, and the larger number of teams registering and participating in the competition, leads to the natural continuation of this event. The inclusion of novel tracks and tasks based on benchmarking protocols will enhance fair comparisons and evaluation of different scientific and technological developments, and foster research on under-investigated areas in the robotic grasping and manipulation domain. Next edition of RGMC will celebrate the 10th anniversary and reach an even broader audience by providing ten robots on-site (5 UR5 and 5 Franka Research 3/Pand robots) to

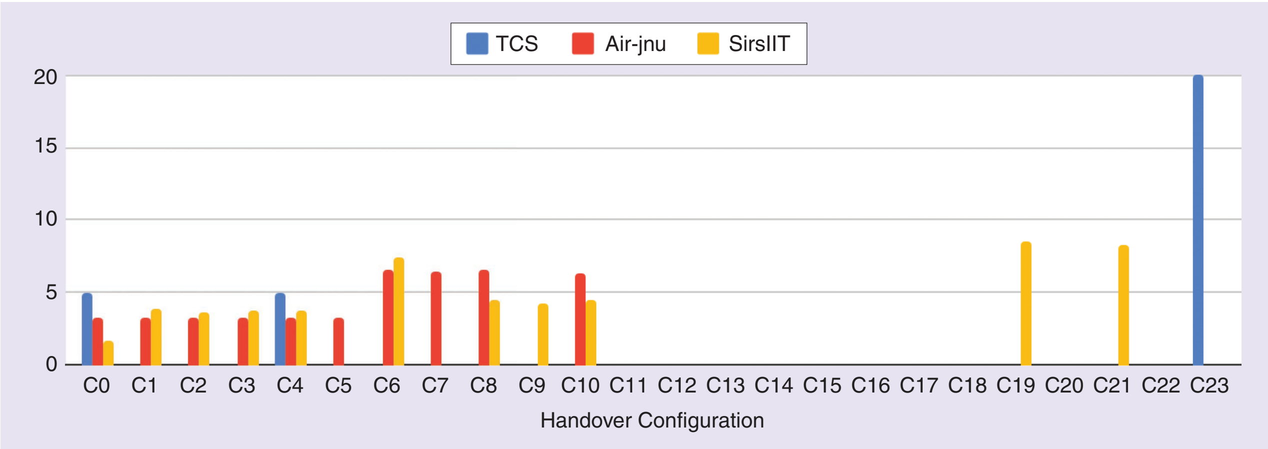


FIGURE 9. A comparison of the scores achieved by each team for each handover configuration. Note that the maximum target score for each configuration can vary depending on the difficulty level predefined for each configuration.

FIGURE 10. Pictures of the winning teams with the organizers of each track. (a) manufacturing track (Tsinghua 3C United team), (b) cloth manipulation track (team Air-jnu), (c) in-hand manipulation track (XL team), (d) picking in clutter track (Team THU-Bot), (e) human-to-robot handovers track (team SirsIIT), (f) team with the most elegant solution (XL Team), and (g) all winning teams and the RGMC organizers.

overcome the challenge for the teams to bring their own setup when joining the event. Furthermore, RGMC will promote the dissemination of competition results and publications of novel methods with the aim of building a community around benchmarking for robot grasping and manipulation by organizing conference workshops (e.g., the IROS 2024 workshop on Benchmarking via Competitions in Robotic Grasping and Manipulation) and collaborating with IEEE/RAS venues (e.g., IEEE Robotics and Automation Practice).

ACKNOWLEDGMENTS

RGMC was supported by ICRA 2024, IEEE-RAS, ASTM International, MathWorks, Amazon Robotics, Universal Robots, and the UMass Lowell NERVE Center. IEEE RAS sponsored travel reimbursements to help cover expenses, with the amount provided based on the distance traveled to the venue. ASTM International funded the cash prizes won by the teams competing in the Manufacturing Track: US$6,000 for first place, US$3,000 for second place, and US$1,000 for third place. IEEE RAS and Amazon Robotics sponsored the prizes for the winning team of each essential skill track. Each winning team received US$1,000. Universal Robots provided two UR5 robotic arms to be used during the competition by the cloth manipulation track and the human-to-robot Handover Track. The EU project euRobin sponsored the cloth manipulation track with US$5,000. UMass Lowell NERVE Center provided the YCB objects. Representatives from Amazon Robotics helped during the evaluation process by serving as judges for the picking in clutter track.

REFERENCES

[1] Y. Sun, J. Falco, M. A. Roa, and B. Calli, “Research challenges and progress in robotic grasping and manipulation competitions,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 874–881, Apr. 2022, doi: 10.1109/LRA.2021.3129134.

[2] S. D’Avella, M. Bianchi, A. M. Sundaram, C. A. Avizzano, M. A. Roa, and P. Tripicchio, “The cluttered environment picking benchmark (CEPB) for advanced warehouse automation: Evaluating the perception, planning, control, and grasping of manipulation systems,” *IEEE Robot. Autom. Mag.*, early access, Jan. 2023, doi: 10.1109/MRA.2023.3310861.

[3] R. Sanchez-Matilla et al., “Benchmark for human-to-robot handovers of unseen containers with unknown filling,” *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1642–1649, Apr. 2020, doi: 10.1109/LRA.2020.2969200.

[4] S. Cruciani, B. Sundaralingam, K. Hang, V. Kumar, T. Hermans, and D. Kragic, “Benchmarking in-hand manipulation,” *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 588–595, Apr. 2020, doi: 10.1109/LRA.2020.2964160.

[5] H. Ha and S. Song, “FlingBot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding,” in *Proc. 5th Conf. Robot Learn.*, 2022, vol. 164, pp. 24–33.

[6] R. Proesmans, A. Verleysen, and F. Wyffels, “UnfoldIR: Tactile robotic unfolding of cloth,” *IEEE Robot. Autom. Lett.*, vol. 8, no. 8, pp. 4426–4432, Aug. 2023, doi: 10.1109/LRA.2023.3284382.