INDEPENDENT STUDY REPORT

Students --

Aditya Agarwal (2021701011) and Bipasha Sen (2021701003)

Advisor --

Dr. Madhava Krishna

Objectives met --

- OCRTOC competition – 3rd position globally
- 6D pose estimation – ongoing collaboration with Dr. Srinath Sridhar from Brown University
- Exploring synergies in push and grasp manipulation – currently in the problem formulation stage

Link to the code repository -- https://github.com/skymanaditya1/ocrtoc_iiith

Final standings in the competition -- (IIIT Hyderabad team name is Lumos)

Rank	Team	error (cm)	1st best task			2nd best task			3rd best task			4th best task		
			task index	error (cm)	pick-up	task index	error (cm)	pick-up	task index	error (cm)	pick-up	task index	error (cm)	pick-up
1	BUAA-GR	18.84	2-2-1	2.66	0	2-2-8	2.84	1	1-1-5	3.31	4	2-1-9	4.25	4
2	CASIA-Robot Team	19.44	2-2-1	1.34	0	2-2-9	1.56	0	2-2-7	3.31	3	2-2-6	2.63	3
3	Lumos	23.33	1-1-5	3.3	4	1-1-8	4.76	4	5-2-9	7.9	5	1-5-2	5.92	5
4	Wuhan & Chal mers	25.69	3-2-3	5.55	4	1-1-3	5.31	3	1-5-2	6.33	4	3-3-8	5.91	4
5	NTXZ_robotic s	26.89	1-1-5	5.9	4	1-1-9	6.08	4	1-1-3	7.56	4	1-1-4	7.75	4
6	THU-VCLab	28.03	1-1-7	4.82	4	1-1-6	5	4	1-5-2	6.21	5	1-1-5	6.35	4
7	EDDURI-S	31.99	2-1-4	3.16	5	1-5-2	5.02	5	1-5-5	5.82	4	1-1-3	6.74	4
8	SIA Smart Gro up	33.56	3-2-4	20.59	2	3-2-2	22.6	2	3-2-3	18.08	2	3-1-3	24.09	1
9	THU-SUN-LA B	34.84	1-1-7	7.81	4	1-1-5	9.65	3	3-3-3	8.47	0	1-5-2	13.53	4

OCRTOC (Open Cloud Robot Table Organization Challenge) is a cloud-based benchmark for robotic grasping and manipulation. It is a global ICRA competition that focuses on the object rearrangement problem and specifically on the task of table organization. This year, the competition witnessed the participation of 24 robotics teams and institutes from all over the world, including some top Chinese universities such as Tsinghua, Peking, the Chinese Academy of Sciences, and Beihang University.

OCRTOC competition website -- http://ocrtoc.org/#/

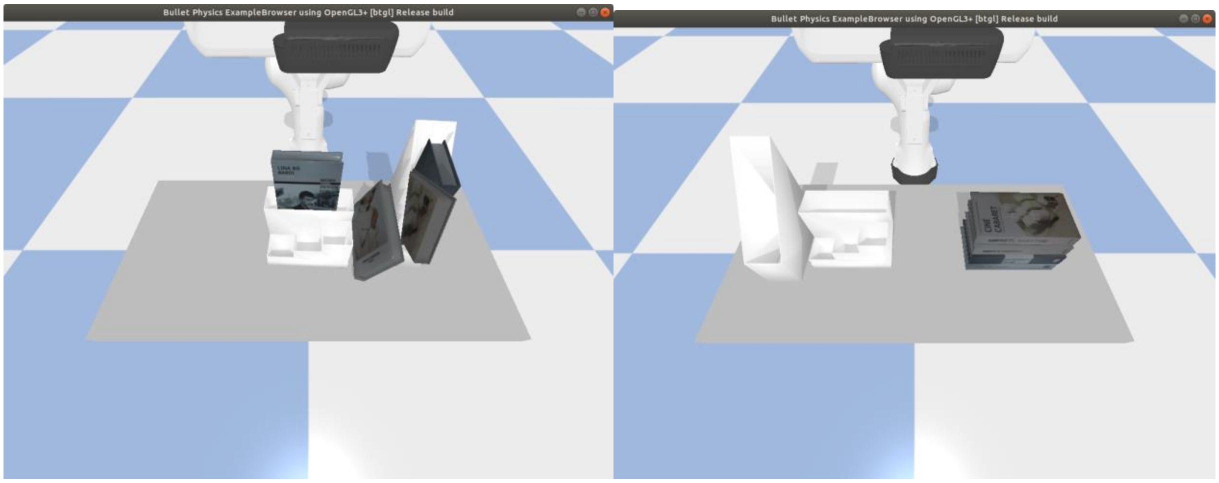
ICRA 2022 Competition Website -- https://rpal.cse.usf.edu/rgmc_icra2022/

The competition involved a 6DOF Franka Panda robot arm (attached below) with a 3D real-sense camera attached to its wrist, a stationary RGB-D Kinect camera, and several objects placed on a tabletop. The competition's goal was to maneuver the robot arm in complex scenes and achieve a goal state configuration given an initial scene. The task was to solve multiple tasks of varying complexity, such as stacking objects, picking and placing objects in clutter, swapping objects, etc., with as little error as possible. Building an end-to-end pipeline that generalizes to multiple scenes was challenging since the pipeline involved multiple moving parts, including deep learning, computer vision, and robot planning stacks.

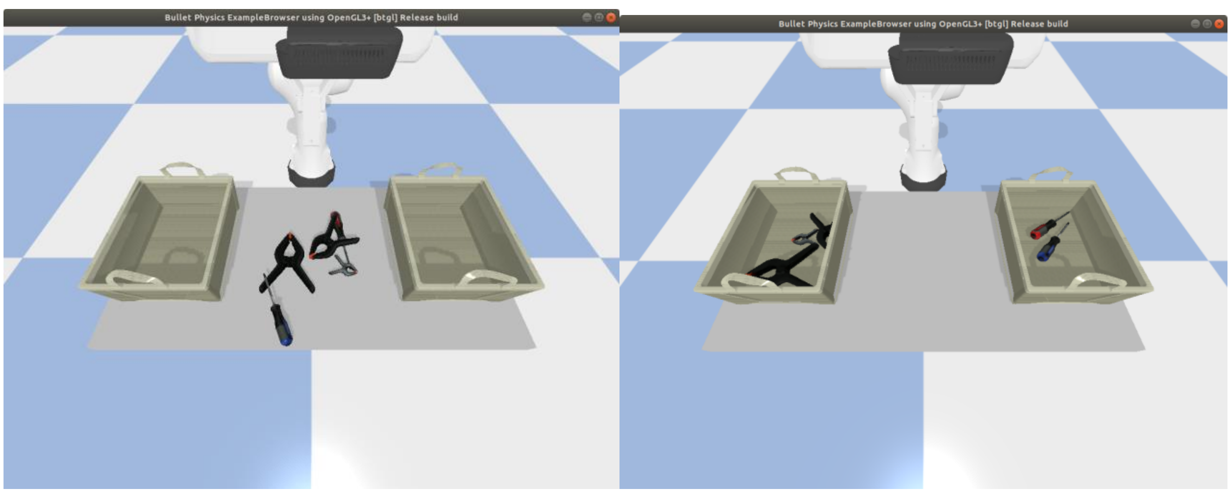
The evaluations were conducted on 30% of scenes seen (made available to the participants) and 70% unseen scenes (evaluated online and hidden from the participants). The competition included multiple scenes of varying diversity and complexity.

<u>**Different scenes of varying complexity --**</u>

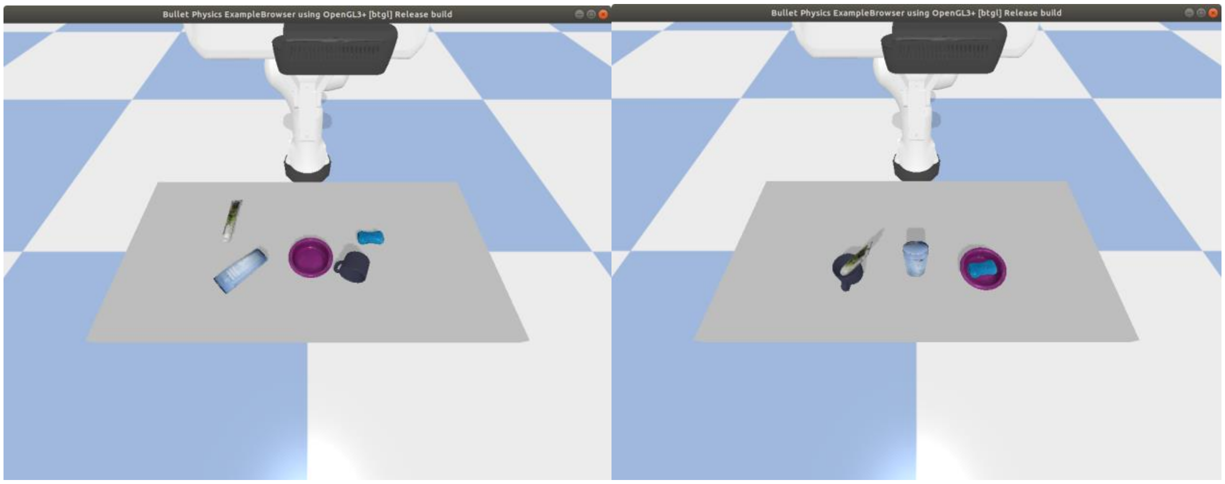
I. Stacking



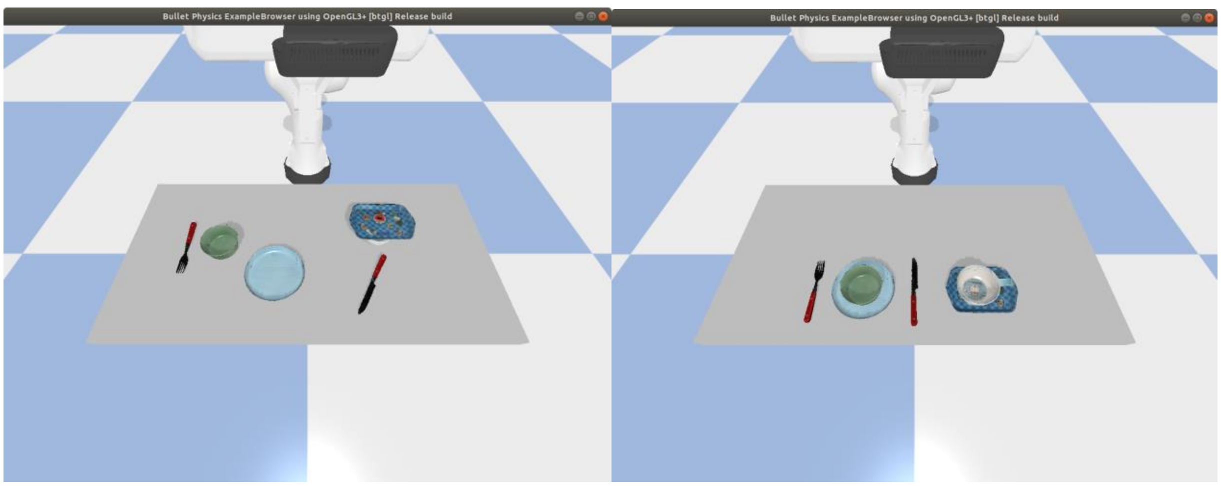
II. Picking objects in clutter



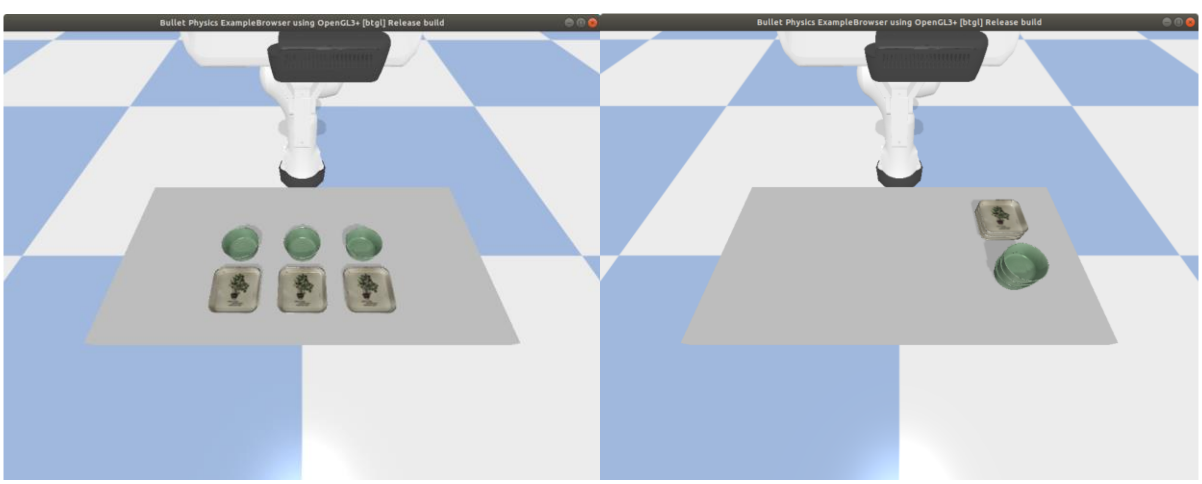
III. Buffer placement and object swapping



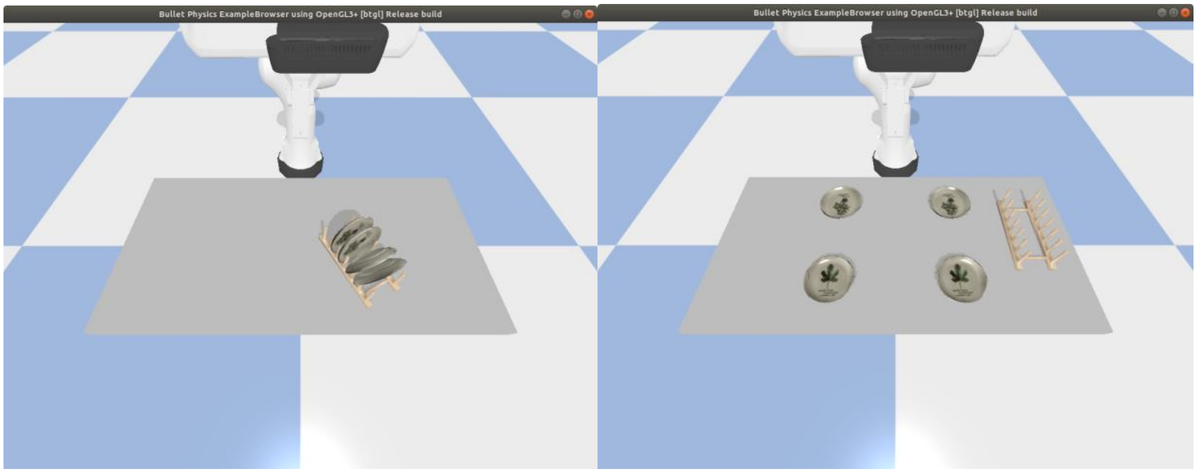
IV. Picking difficult to pick objects (flat-plates, knives, forks etc.)



V. Duplicate objects



VI. De-stacking



We had to improve upon a baseline solution provided by the OCRTOC team. The aim was to minimize the error between the desired final configuration of the table and the configuration achieved by our solution. We added many changes to the baseline and were able to solve many of the challenges. Some of the changes were:

(1) Improving the 3D reconstruction of the tabletop scene,

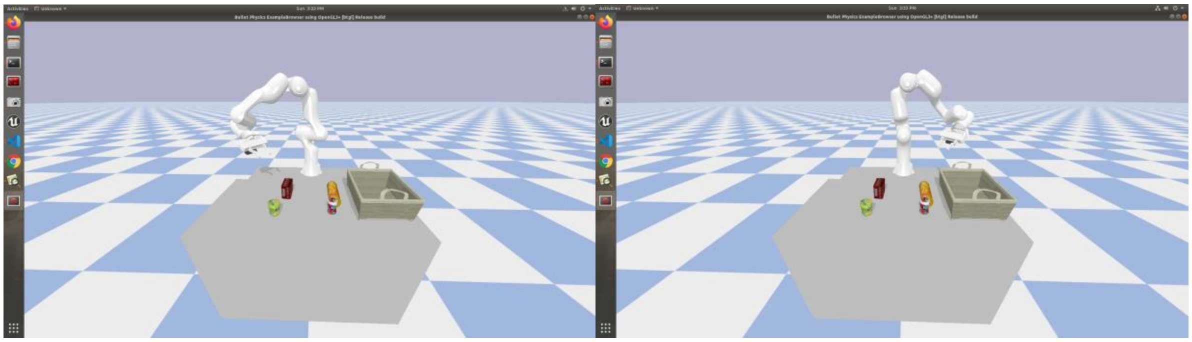
(2) Improving the way the robot arm grabbed the objects,

(3) Smoothening the overall movement of the robot arm,

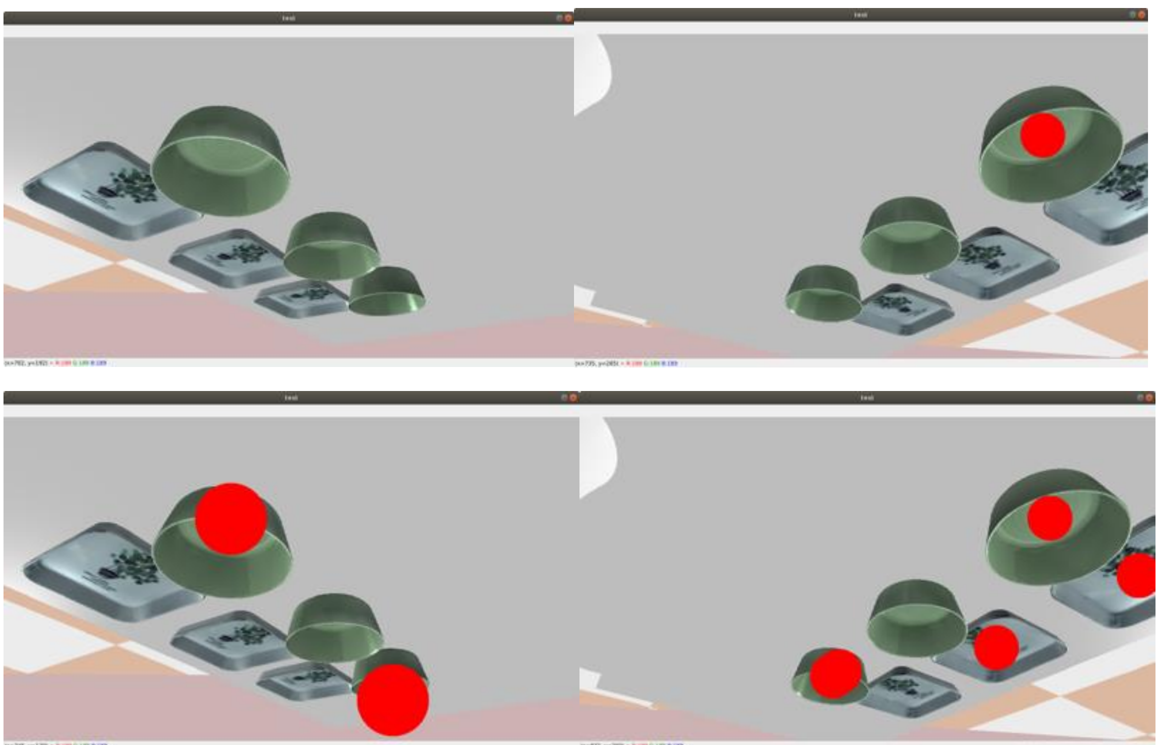
(4) Building a full-blown planning module from scratch that could handle cases that the existing SOTA planners could not, such as stacking and swapping objects.

Challenges solved -- (All the approaches and video demonstrations are available in the PPT and Video directory)

1) Integrating camera poses for reducing holes during scene registration (more clearly demonstrated in the video) -- The scene registration is performed by capturing the scene from different camera angles/poses. Finding the optimal camera poses that would lead to dense scene registration in the least amount of time was a key objective. We found the 4 camera poses empirically which would lead to robust 3D point cloud scene registrations.

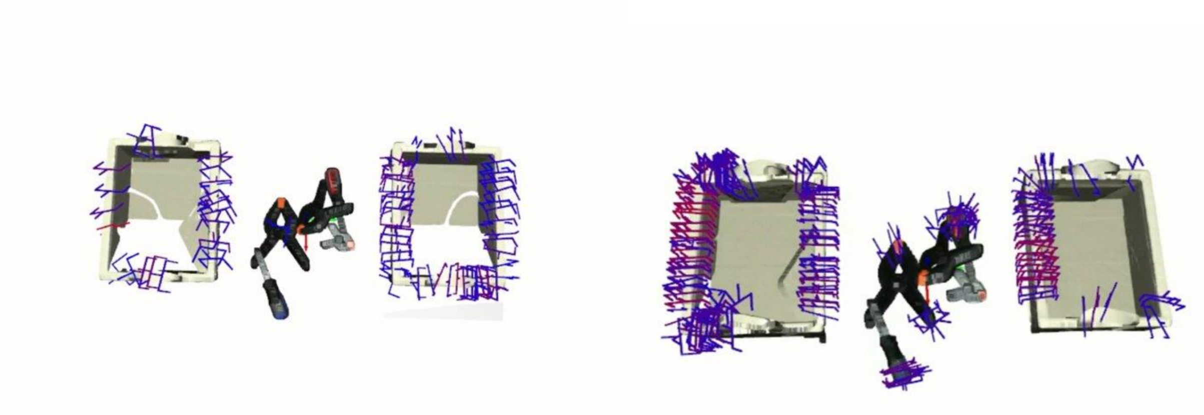


2) Handling duplicate objects by marking the objects already seen – The feature matching code provided in the baseline would fail in cases containing duplicate objects as it would assign a target pose to an already detected object. We added markers on already detected objects to prevent assigning multiple target poses to the same object.



3) Integrating and modifying the grasp proposal logic (using a bounding box approach) -- grasp proposals are the possible poses that the robot arm can use to grab the object. Dense grasp proposal generation is a key component of ensuring that all objects are picked up. We integrated a grasp proposal approach based on Contact Graspnet and handled the corresponding coordinate transformations that followed leading to denser grasp proposals. We

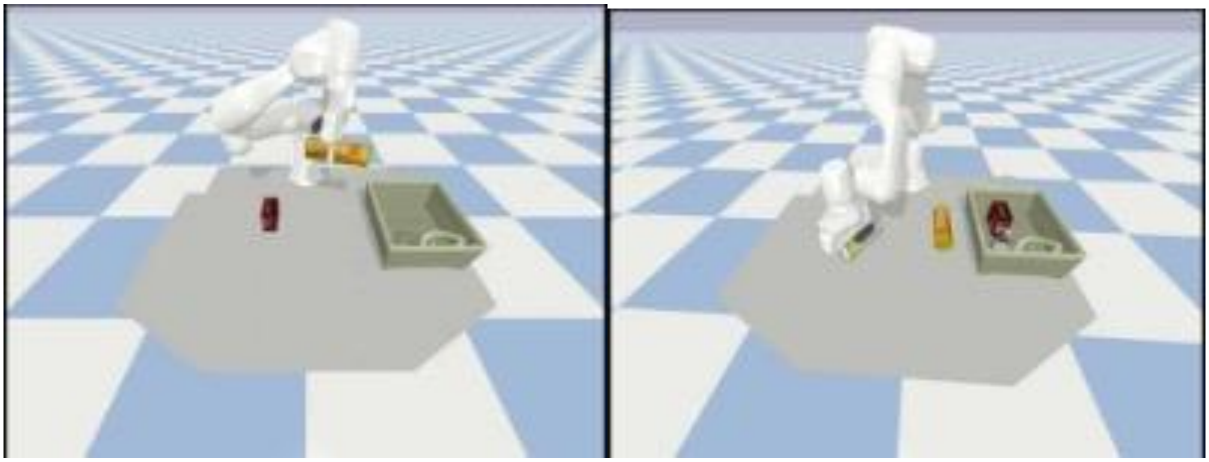
also changed the grasp proposal logic using a 3D bounding box detection-based approach to assigning grasp poses to objects accurately, especially in scenes with clutter (as shown below).



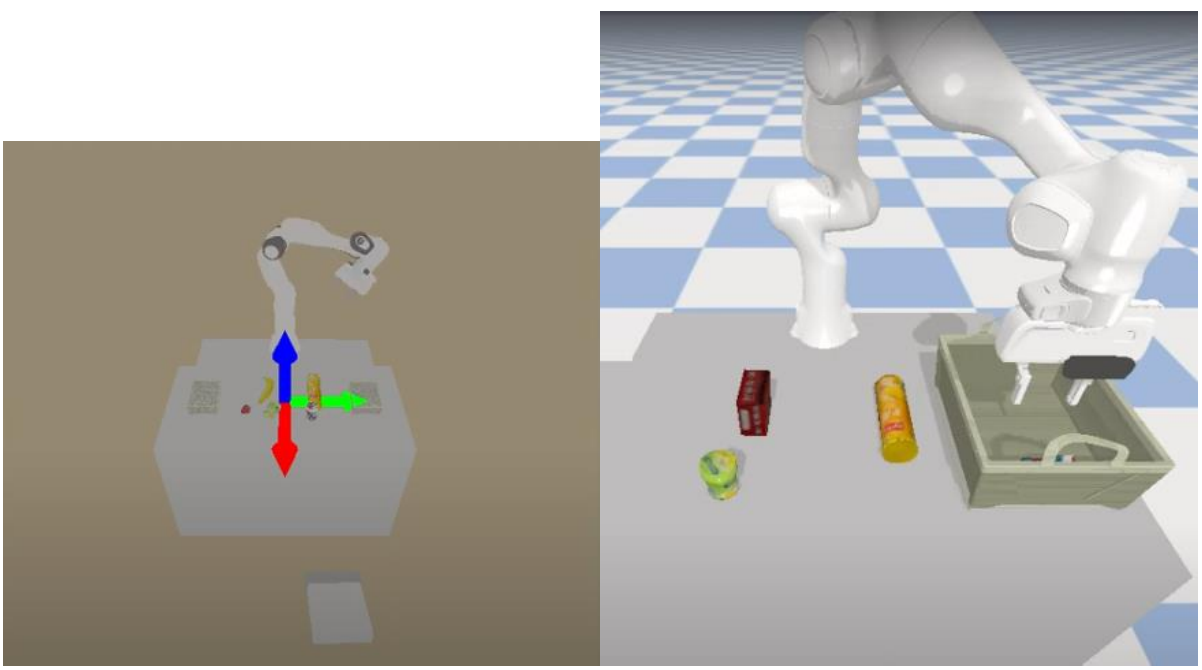
4) Custom task planner – intermediate position before the object is placed at its target position. The baseline task planner could not handle scenes where object stacking or swapping were required. We implemented a custom task planner that performed better than SOTA task planners in handling scenes where object stacking, and swapping were required. Attached below is the pseudo-code for the custom planner involved.

```
Initial object pose list: I_list
Final object pose list: F_list

object_stacks = get final scene object stacks

while all objects not done:
  2D_occ_grid = get occupancy grid using the current point cloud

  current_object = get the object that has free goal spot and is supposed to be
                   placed at the bottom of its stack
  if current_object not None:
    place the current object
    continue
  else:
    for object in I_list:
      p = check if the objects that are supposed to be placed under it are placed
      if p is true:
        current_object = object

    if current_object not None:
      place the current object
      continue
    else:
      current_object = choose some random unfinished object
      buffer_pose = sample a collision-free buffer spot closest to goal pose
      place the current_object in the buffer spot
      continue
```

5) Removing erratic movements while moving in Cartesian space – adding rest poses to prevent erratic movements. We observed that the robot arm would make weird movements while moving in the Cartesian space leading to object collisions. Like other SOTA approaches, we added rest poses after successful object picks, which led to minimal object collisions and overall smoother robot motion.



6) Experimenting between different simulators (PyBullet vs Sapien) -- Different simulators play a different role for different scenes as the underlying physics and object-arm interactions are different. We evaluated the performance of two simulators PyBullet and Sapien and all the tasks and optimized for the one with the higher accuracy (Sapien).



Video demonstrations are provided which show how we handle scenes of varying complexities.

We observed that integrating these small yet incremental changes helped us identify issues early, and our position changed several times in the top 3 standings during the competition. Participating in the competition helped us identify critical challenges in this area and formulate several research directions such as robust 6D object pose estimation, inpainting an incomplete 3D pointcloud registration, and exploring synergies between pushing and grasping objects in scenes involving clutter.

The tabletop arrangement and scene arrangement are critical and futuristic problems. They have tremendous potential in real-life practical settings, such as using service robots in warehouse automation, in the dining industry, and assisting humans by performing dirty, dull, distant, dangerous, or repetitive jobs. We believe that working on these research areas will help us build a fast, robust, and generalized pipeline, and we wish to explore our work on real robot setups.

Team Lumos from IIIT Hyderabad has been invited to ICRA 2022 on 24th May to be held in Philadelphia USA, to give a presentation about their solution and the different challenges they faced. They will also participate in running their solution on real-robot setups at the conference.

Our solution was also demoed to the head of TCS Research India Dr. Gautam Shroff (Vice President and Chief Scientist) and the CTO of TCS K Ananth Krishnan (TCS EVP & CTO).

List of resources included --

1. Video presentation to the CTO of TCS - https://drive.google.com/file/d/1FYvp3YkjV63wag53viGkVxhJ1HrxgMja/view?usp=sharing
2. Weekly presentation presented to Dr. Madhava Krishna from November 2021 – May 2022 --

https://drive.google.com/file/d/15EvQ8uwDaLokQhyhhCXriVkTWpXrMzAG/view?usp=sharing

3. ICRA 2022 workshop video submission - https://drive.google.com/file/d/1-sc21YAEKk1Ynb2rET6OSTuQCAtcXbE6/view?usp=sharing
4. ICRA 2022 competition demonstration video -- https://drive.google.com/file/d/113Vsqr82ar8hpPwL4lW9-h-8zKTgo8rP/view?usp=sharing
5. ICRA workshop paper -- https://drive.google.com/file/d/11YlzlNR7WOrvfbEp5r4Wtl1l8SbgdUbb/view?usp=sharing also available at -- https://arxiv.org/abs/2205.04090