



# Getting Robots to Touch Things Appropriately

Jen Jen Chung

7 February 2024





THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

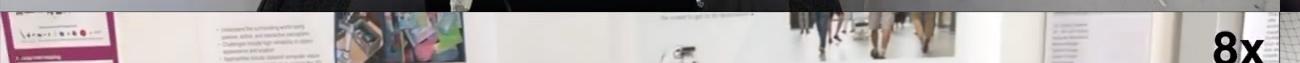


**ETH** zürich



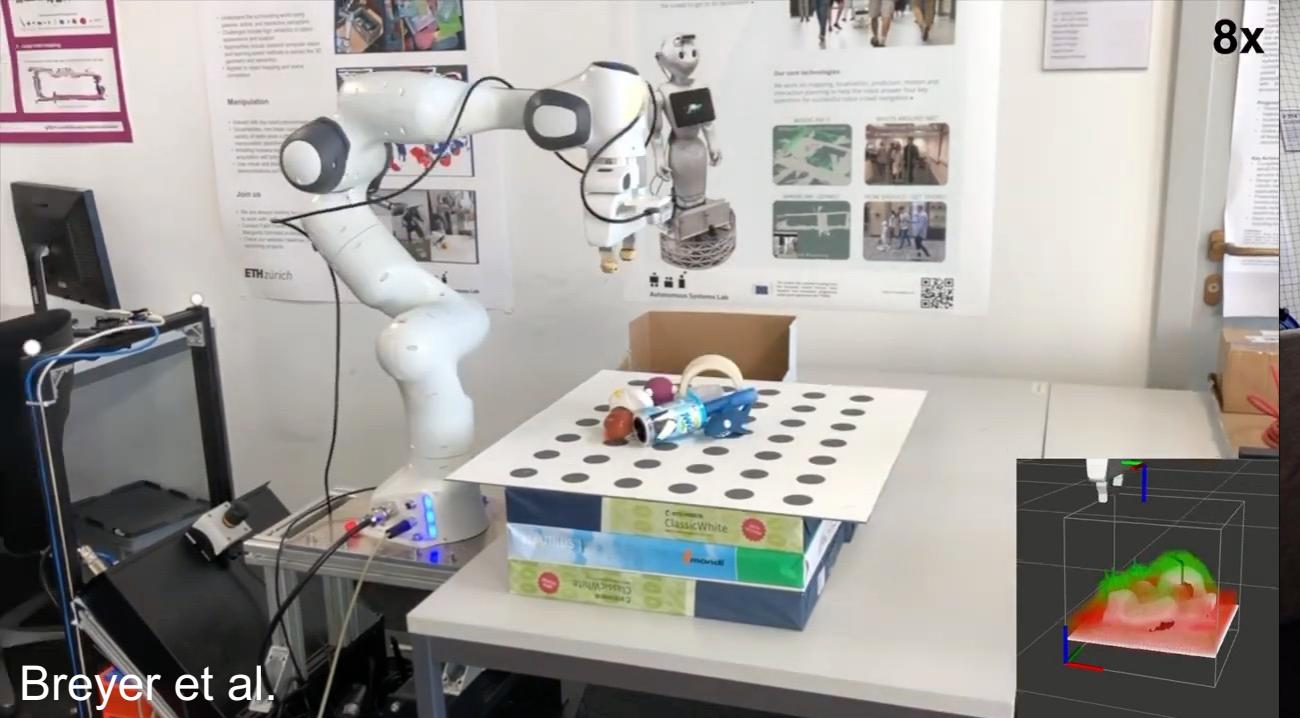
Autonomous Systems Lab

Grinvald et al.



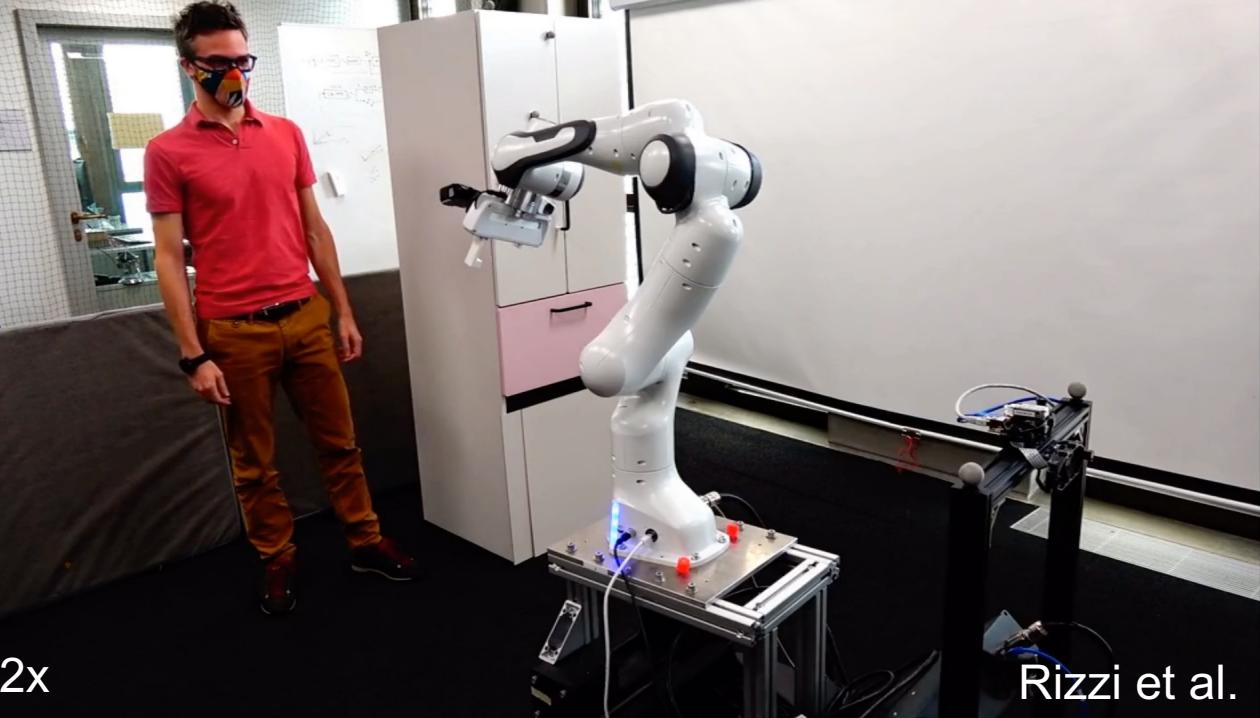
Primesense RGB-D camera facing forwards

Breyer et al.



8x

2x



Dugas et al.

Rizzi et al.



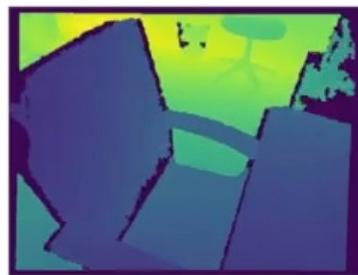
# Voxblox++: Combining geometry and semantics



RGB



Mask R-CNN



Depth



Depth segmentation

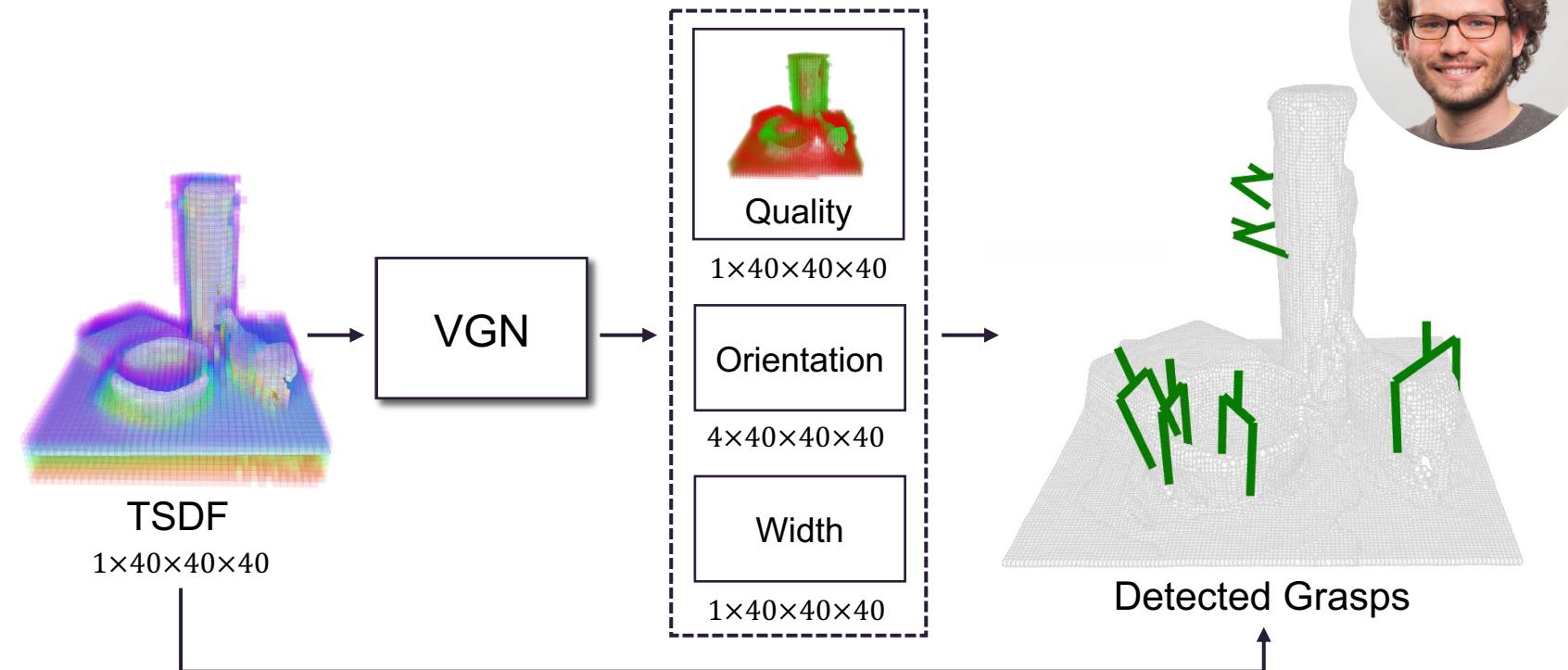
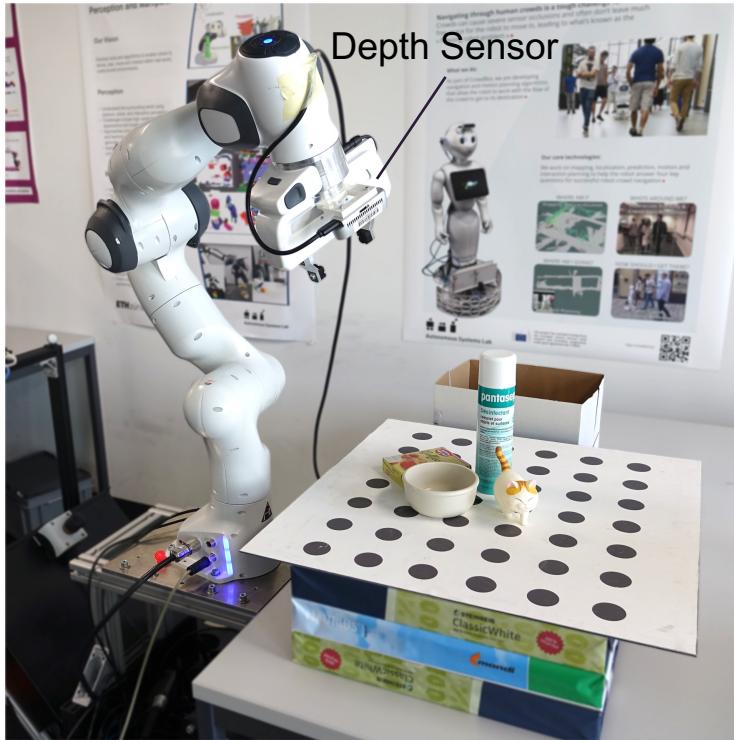


\*not actual speed

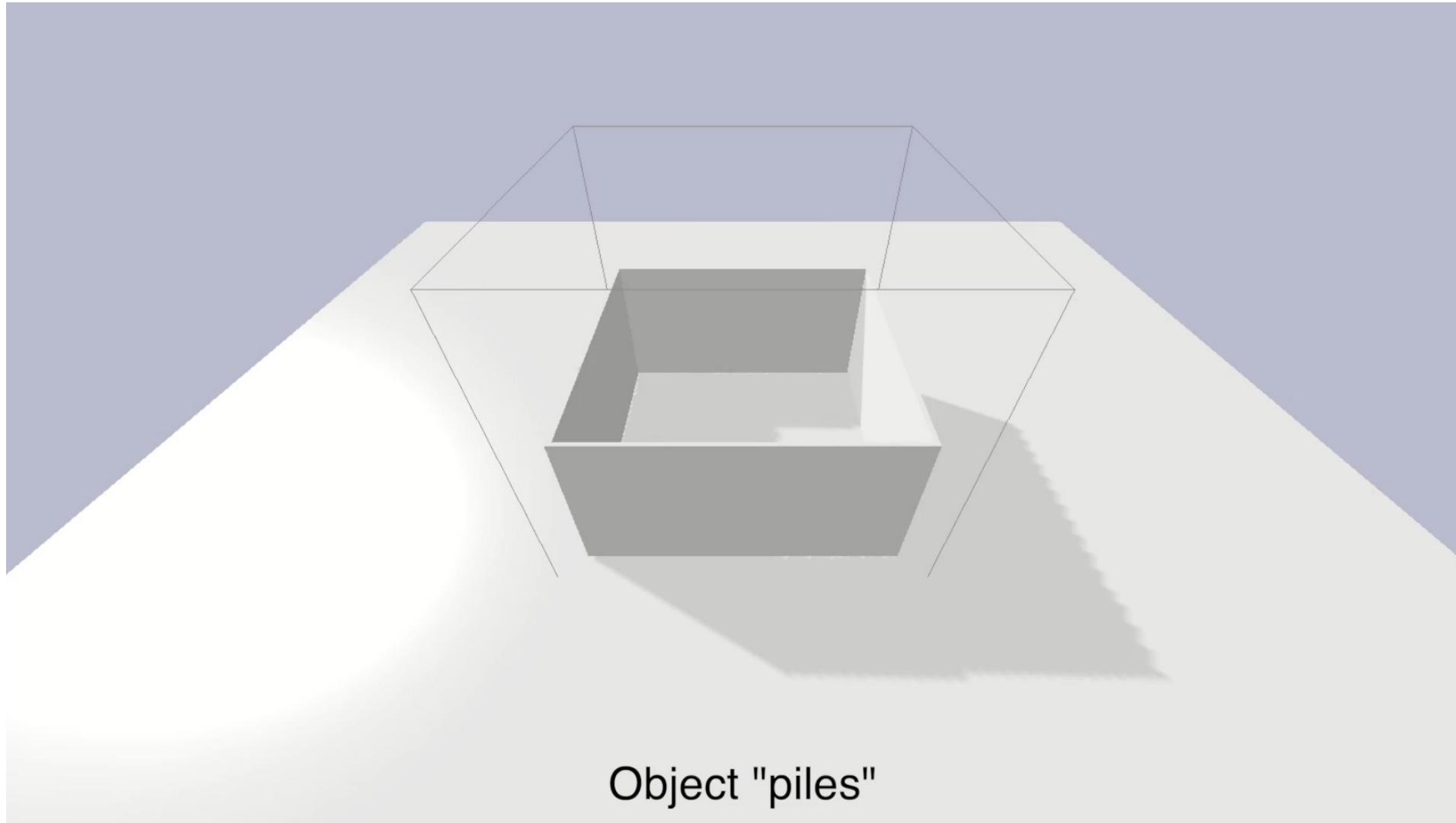
Segments are incrementally integrated into a volumetric map, where a fusion strategy keeps track of the 3D geometry and location of the individual object instances detected.



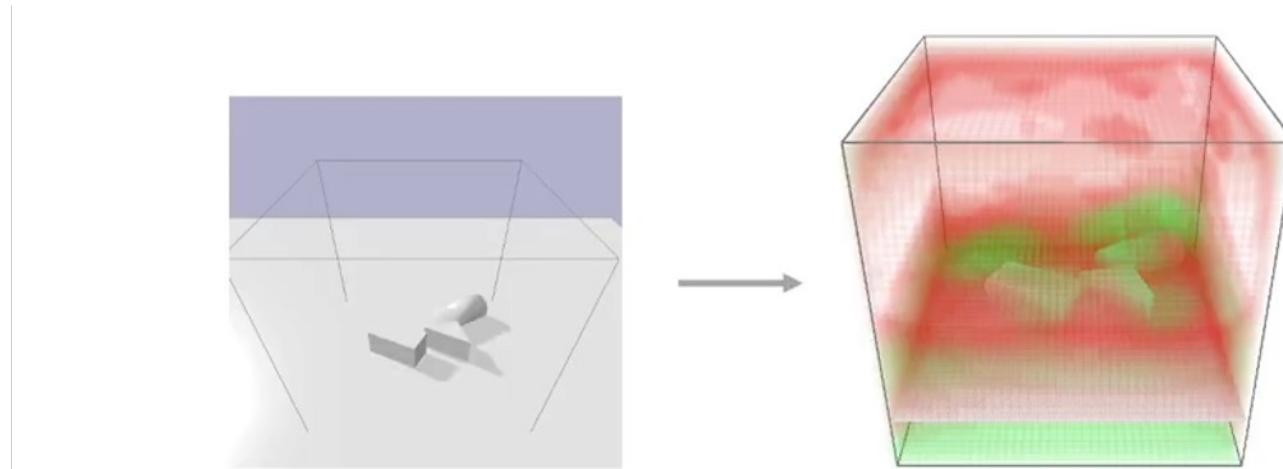
# Scene-aware grasping in 6DoF



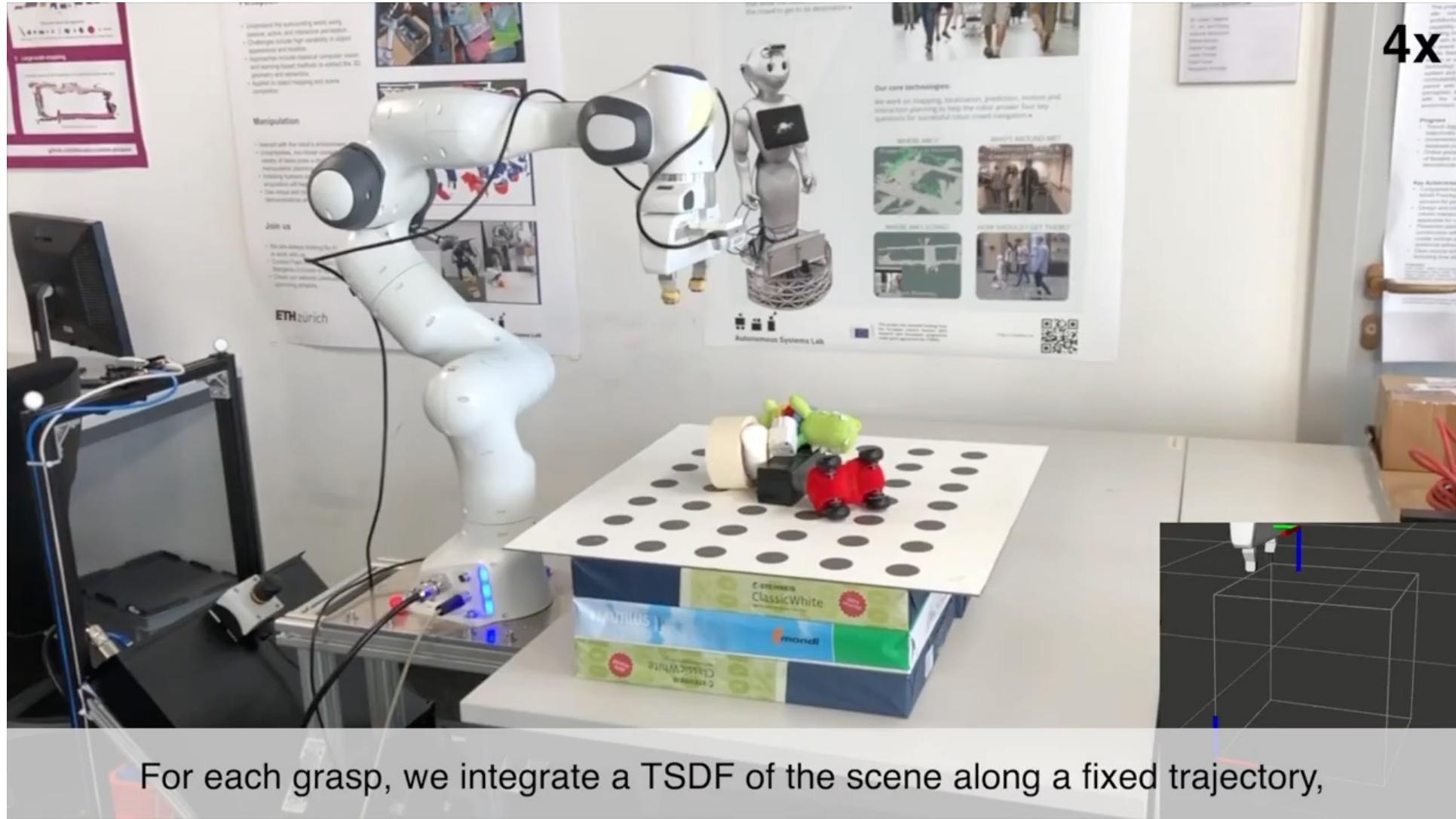
# Training data collection in simulation



# Filtering the network output



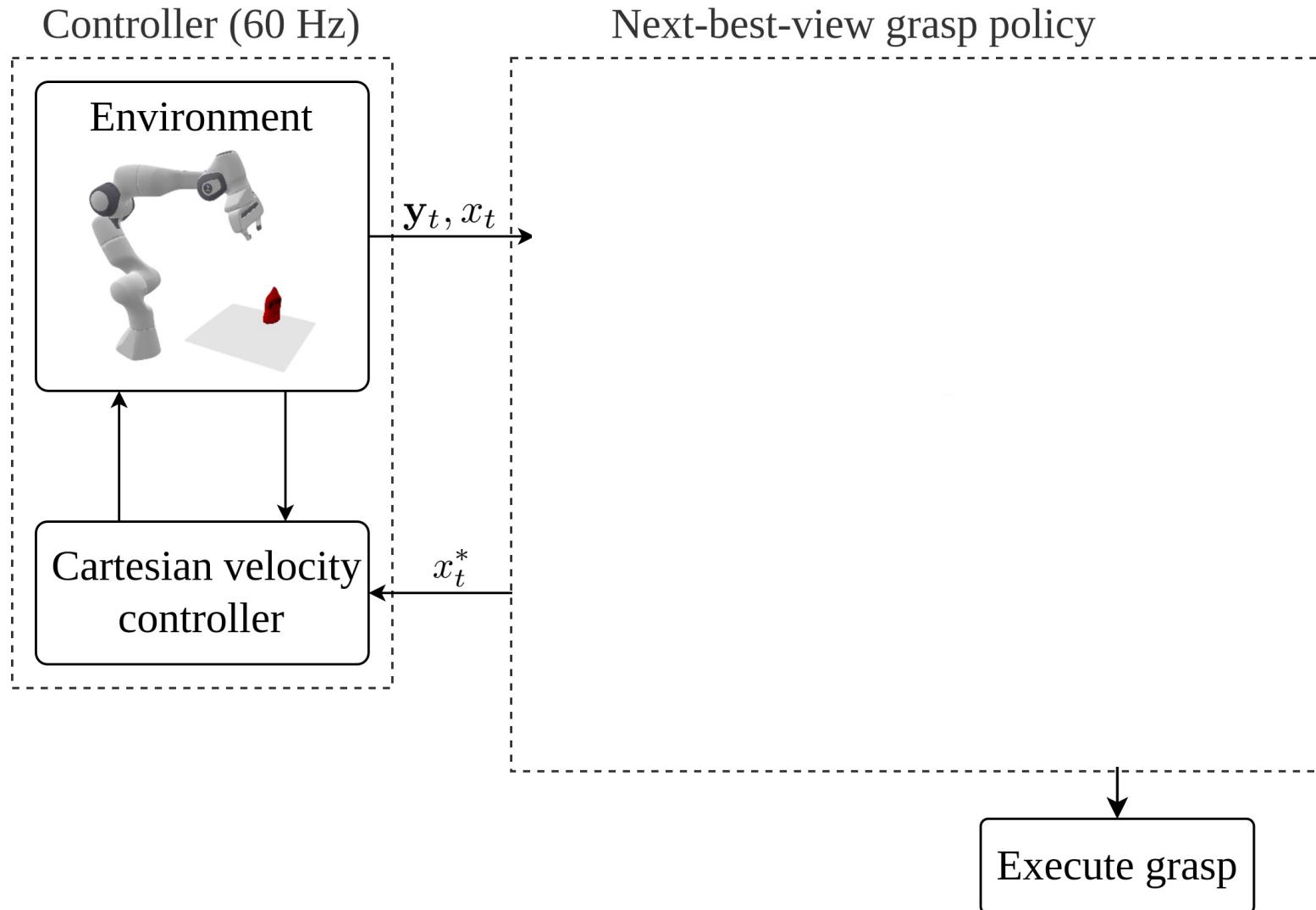
# VGN: Volumetric Grasp Network



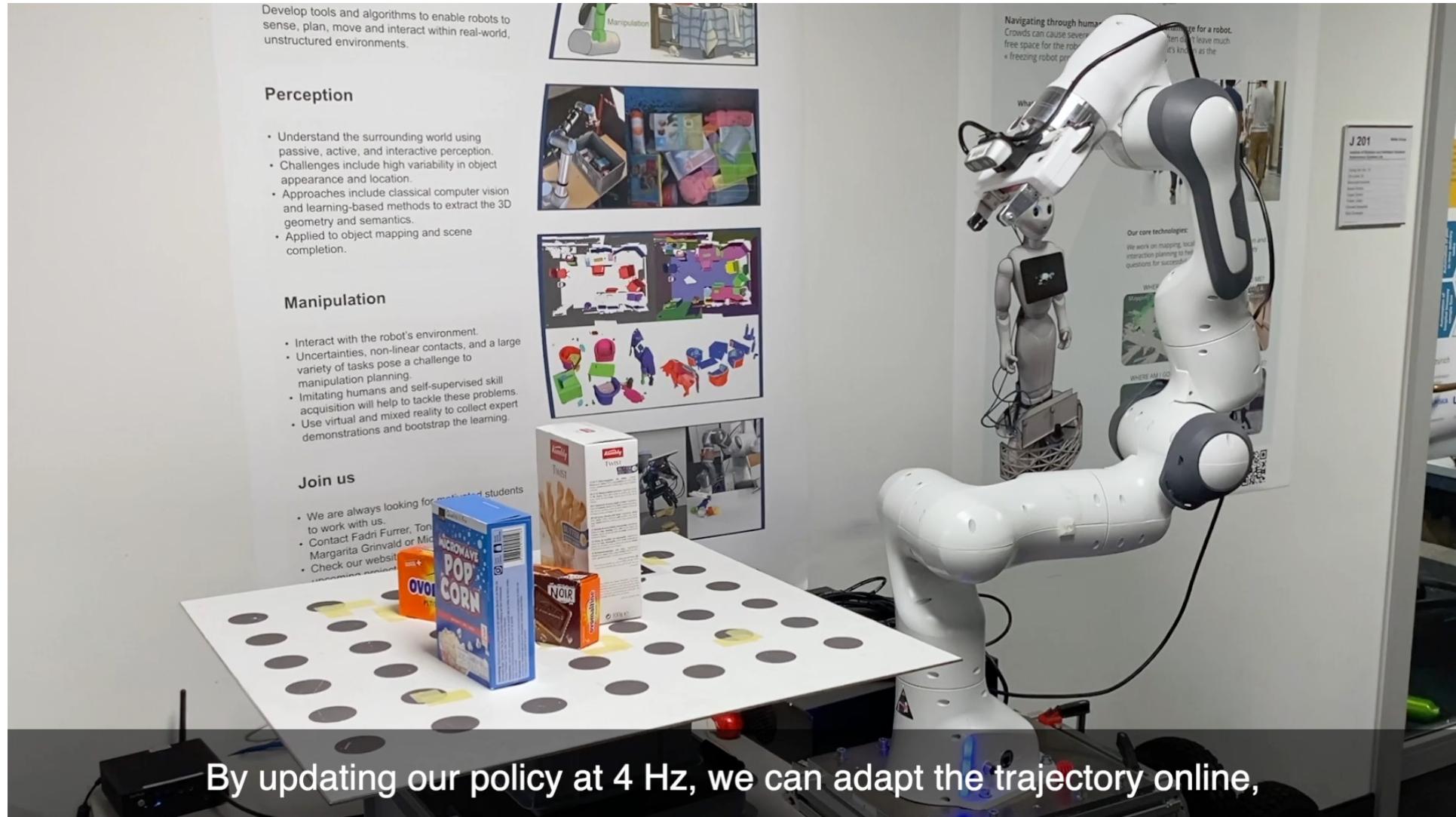
# Next-best-view for grasp detection



# Closing the loop

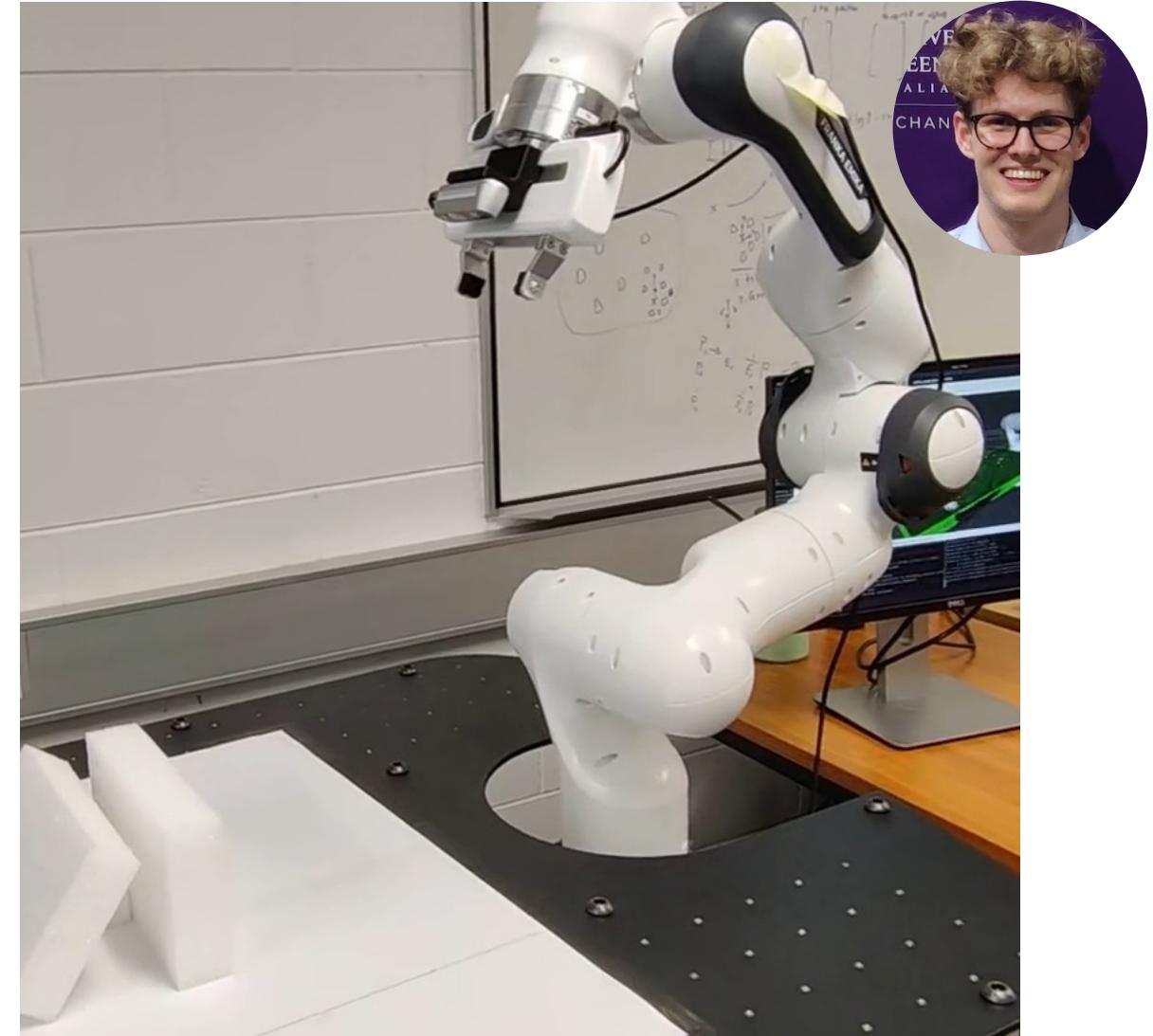
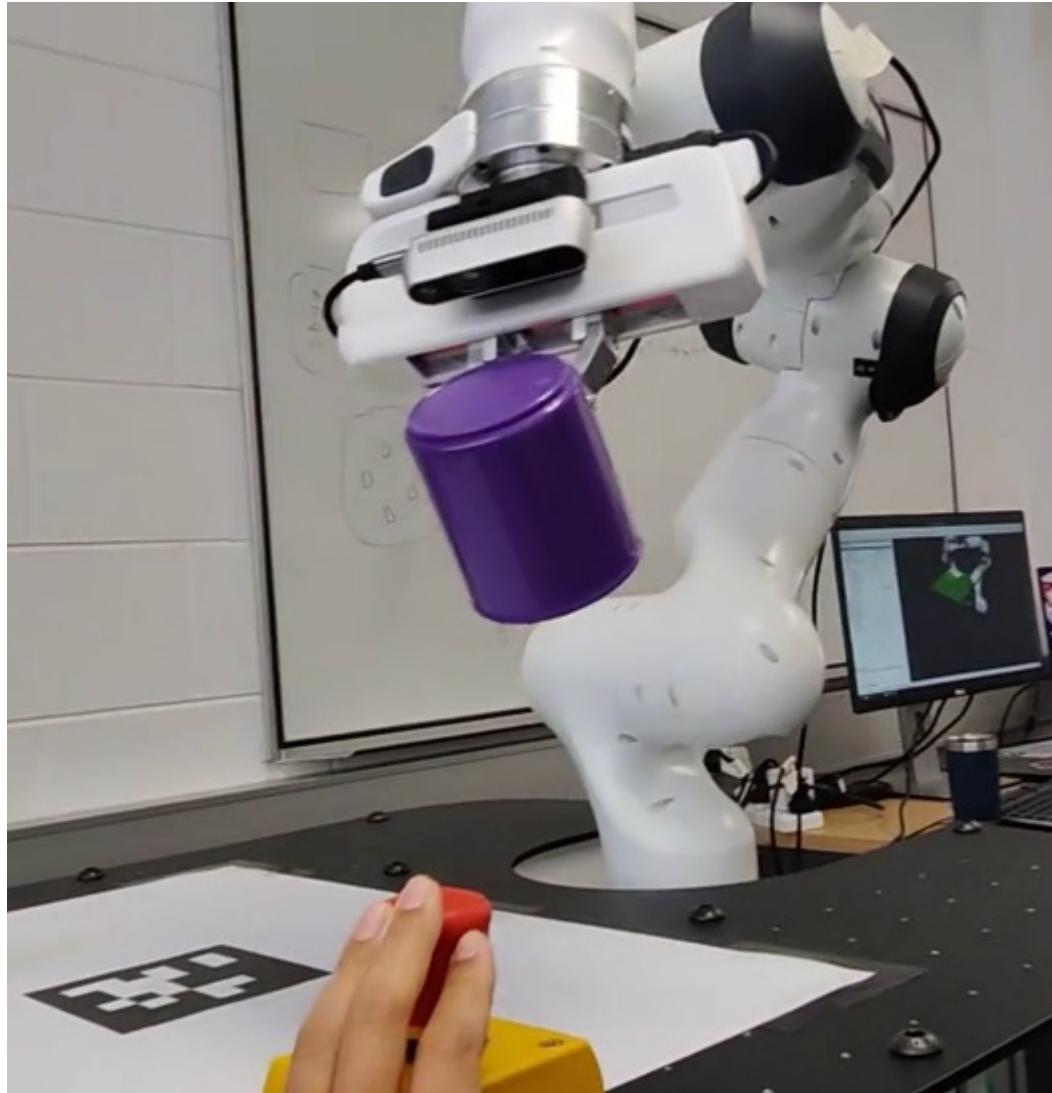


# Active perception for grasping

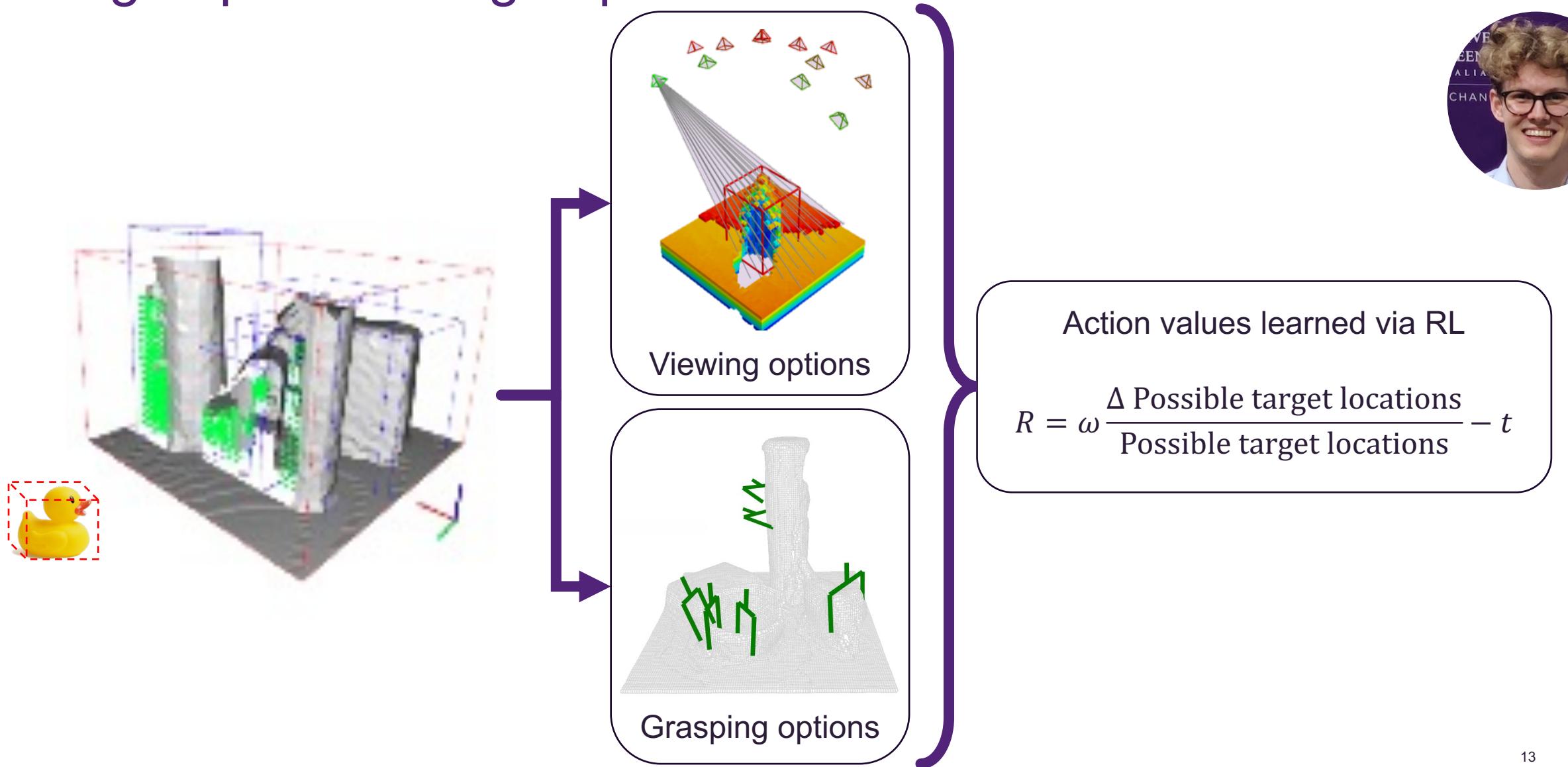


Breyer et al., "Closed-loop next-best-view planning for target-driven grasping", IROS 2022

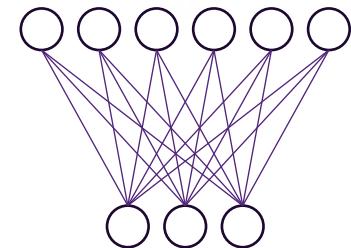
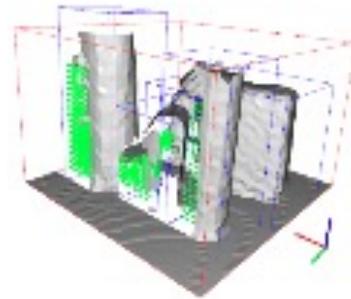
# What to do when you can't see the target object?



# To grasp or not to grasp?



# Dual-headed deep Q-network



View agent  
input

encoded  
TSDF

end-effector  
pose

Grasp agent  
input

encoded  
TSDF

end-effector  
pose

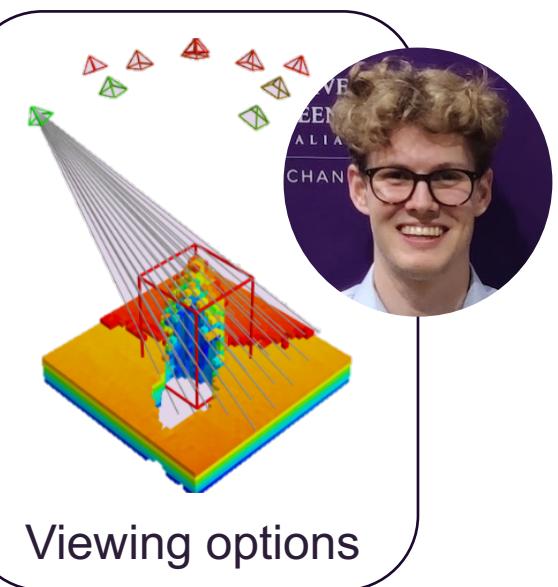
view\_pose0

view\_pose1

view\_pose2

:

view\_poseM



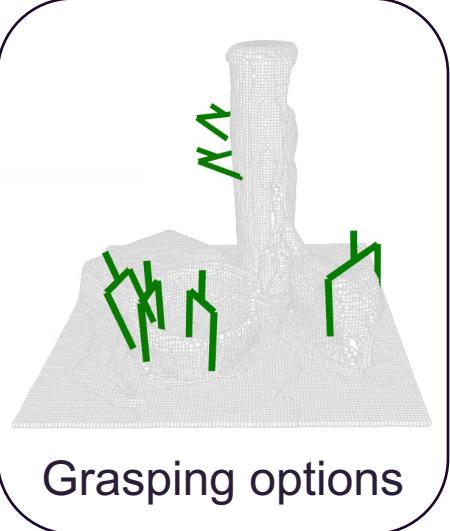
grasp\_pose0

grasp\_pose1

grasp\_pose2

:

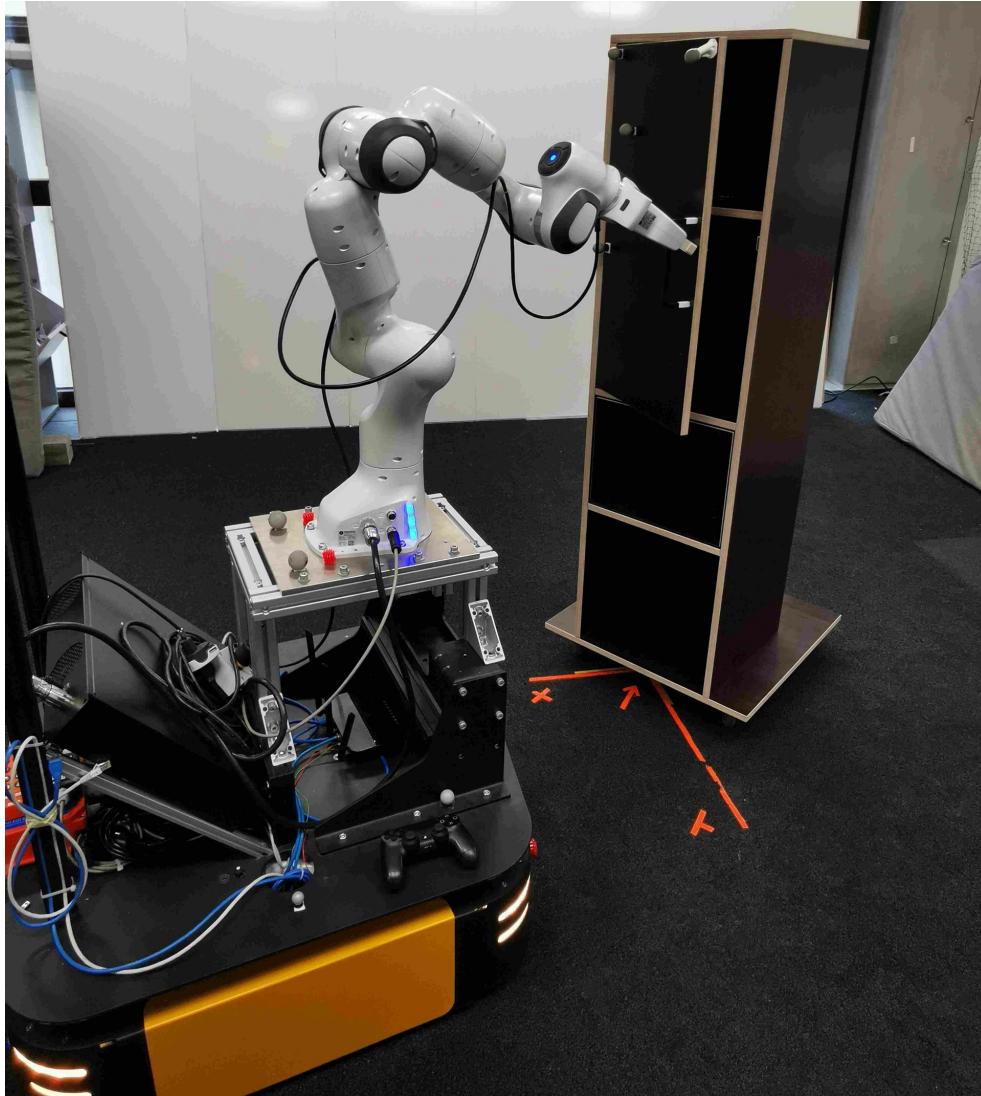
grasp\_poseN



# Interactive perception for grasping



# Interaction with articulated objects



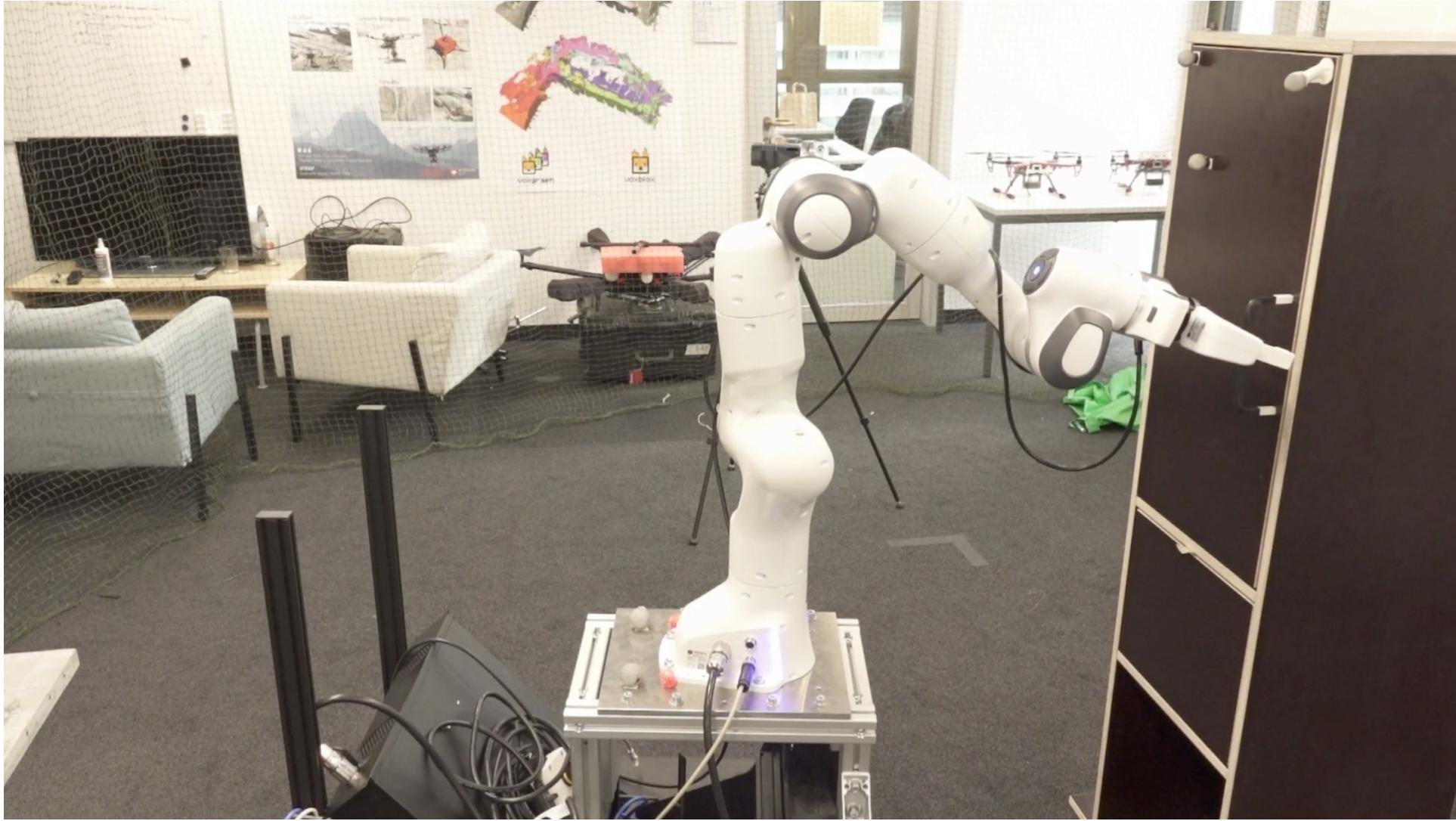
- Grasping → force-based interactions
  - ✗ Synthesise grasp and plan
  - ✓ Continuous control
- Switching contact dynamics
  - ✗ Gradient-based optimisation
  - ✓ Sampling-based optimisation
- Unknown object properties
  - ✓ Need built-in safety and robustness

Model predictive path integral control (MPPI)

Control barrier function  
+  
Energy tank



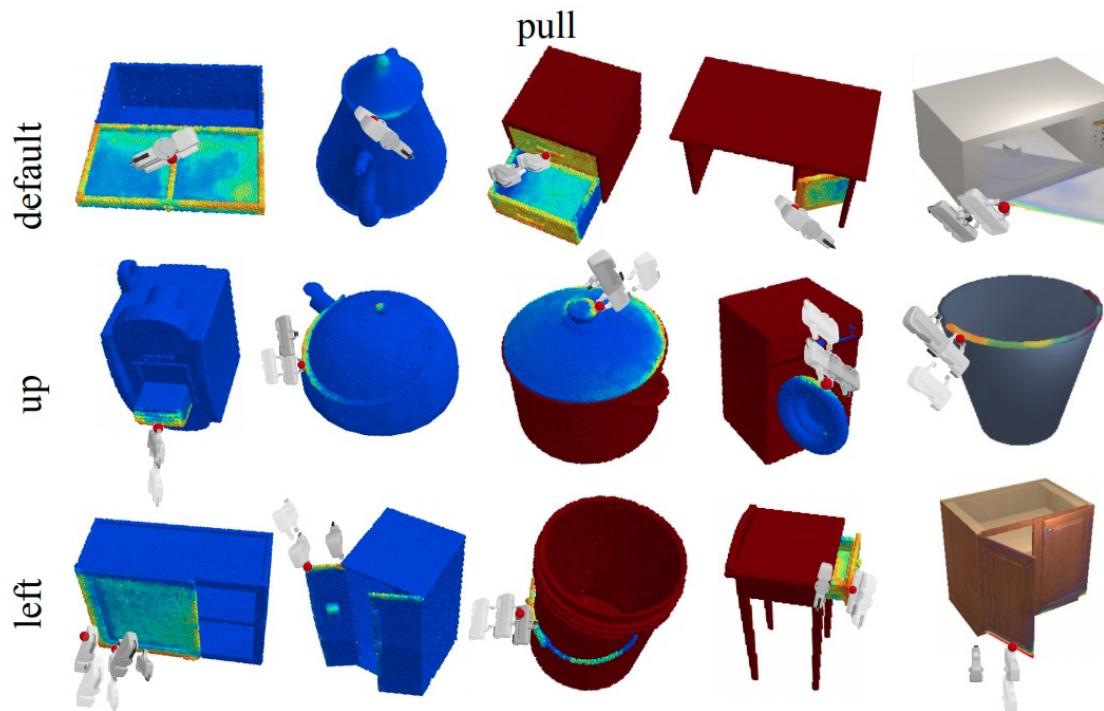
# Safe whole-body interaction



Rizzi et al., "Robust sampling-based control of mobile manipulators for interaction with articulated objects", TRO 2023

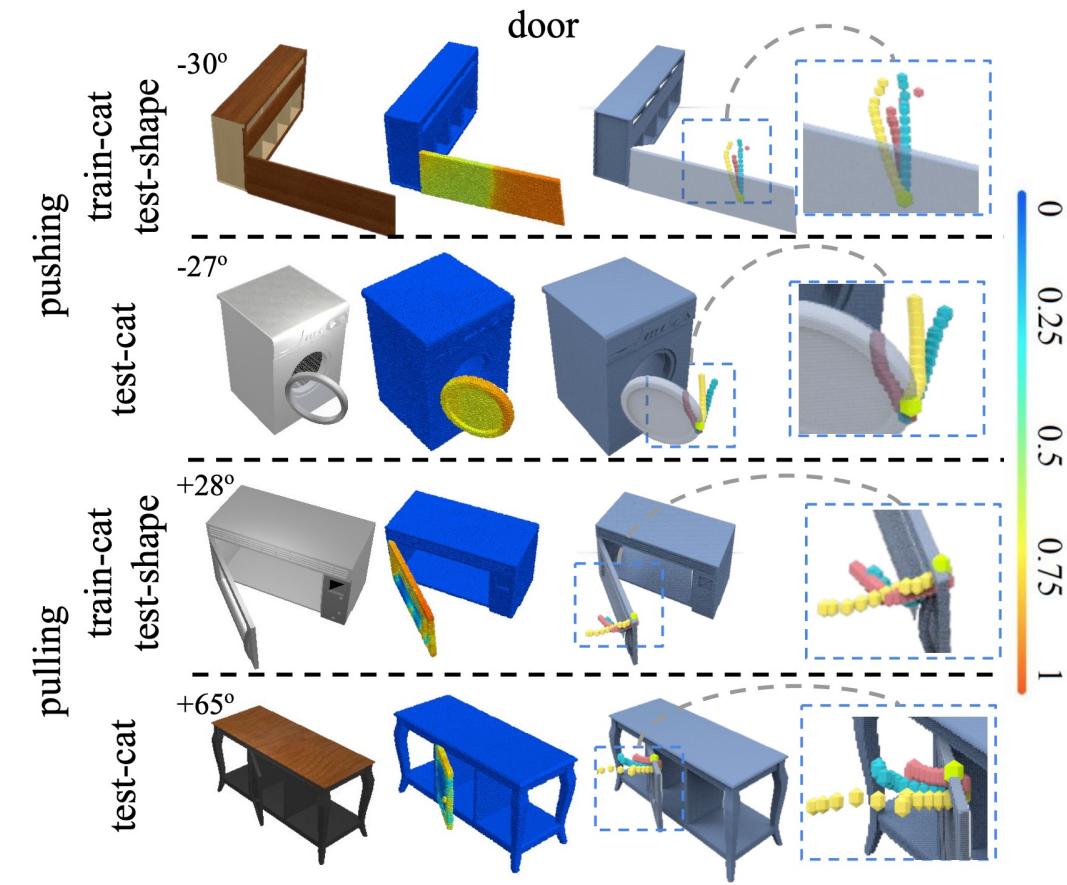
# Learning *where* to interact with articulated objects

Where2Act



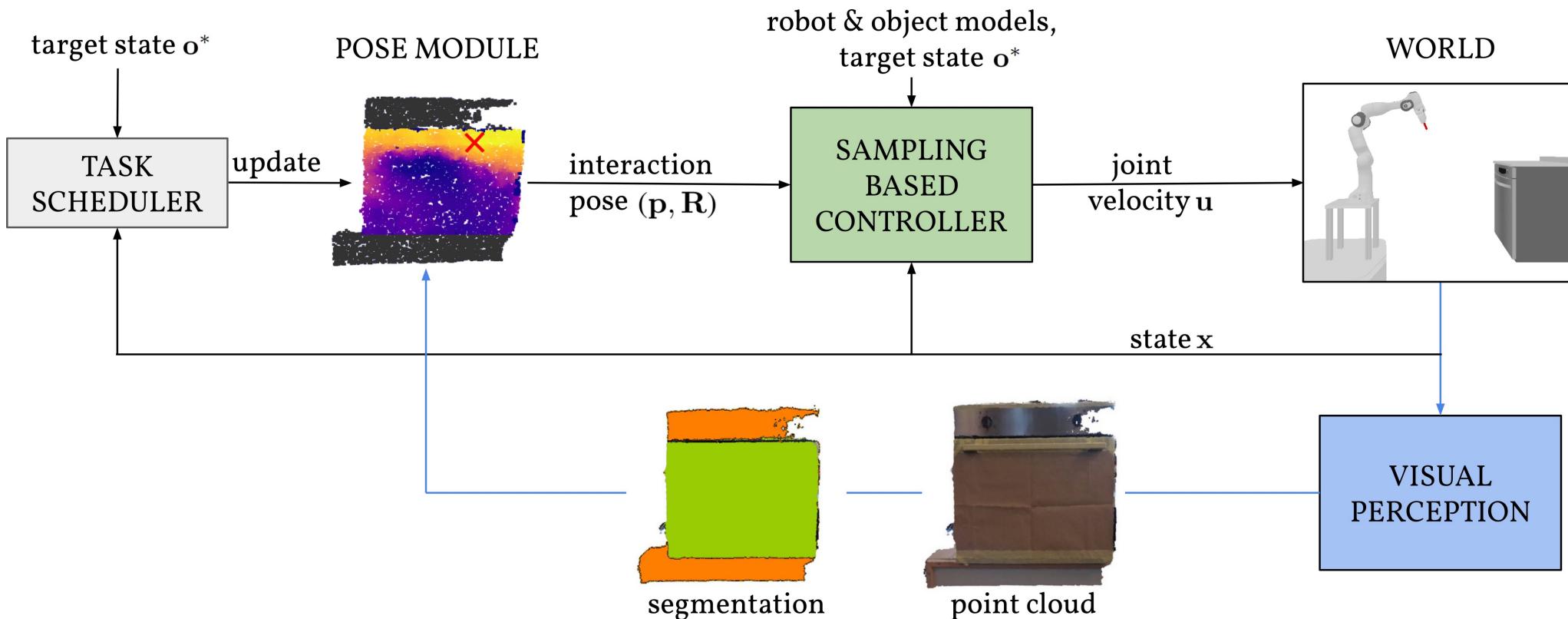
Mo et al., 2021

VAT-Mart



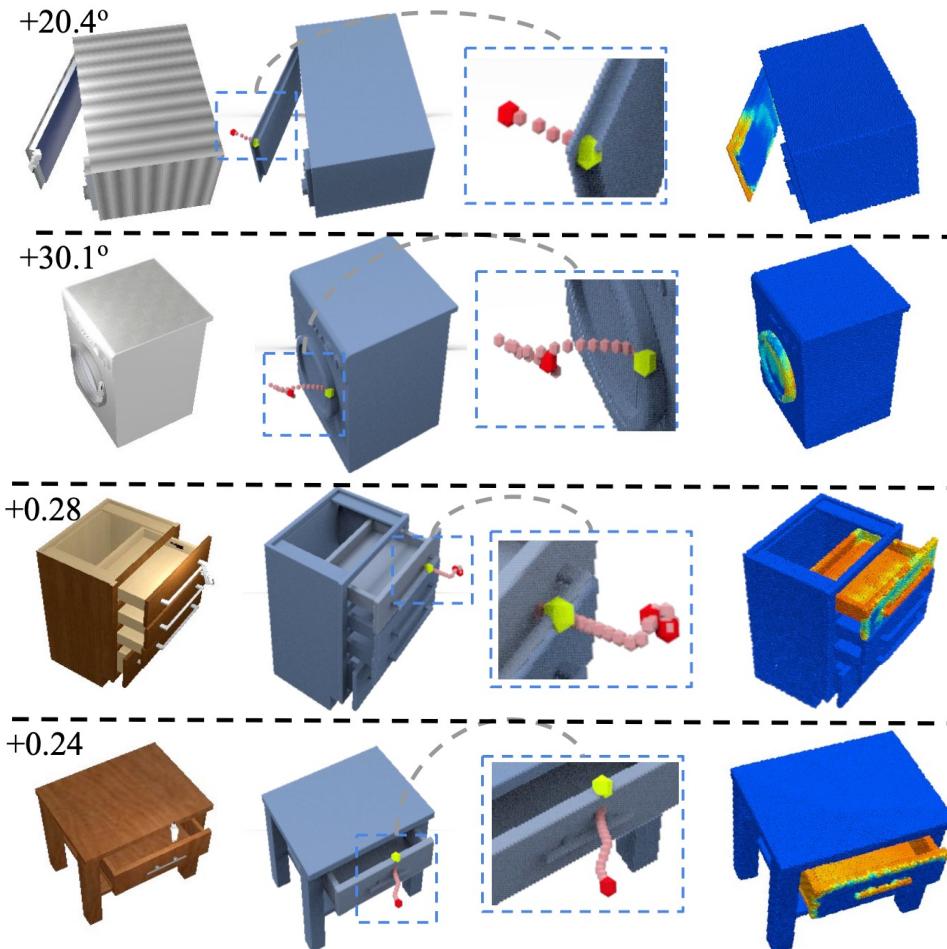
Wu et al., 2022

# Agent-aware affordance learning

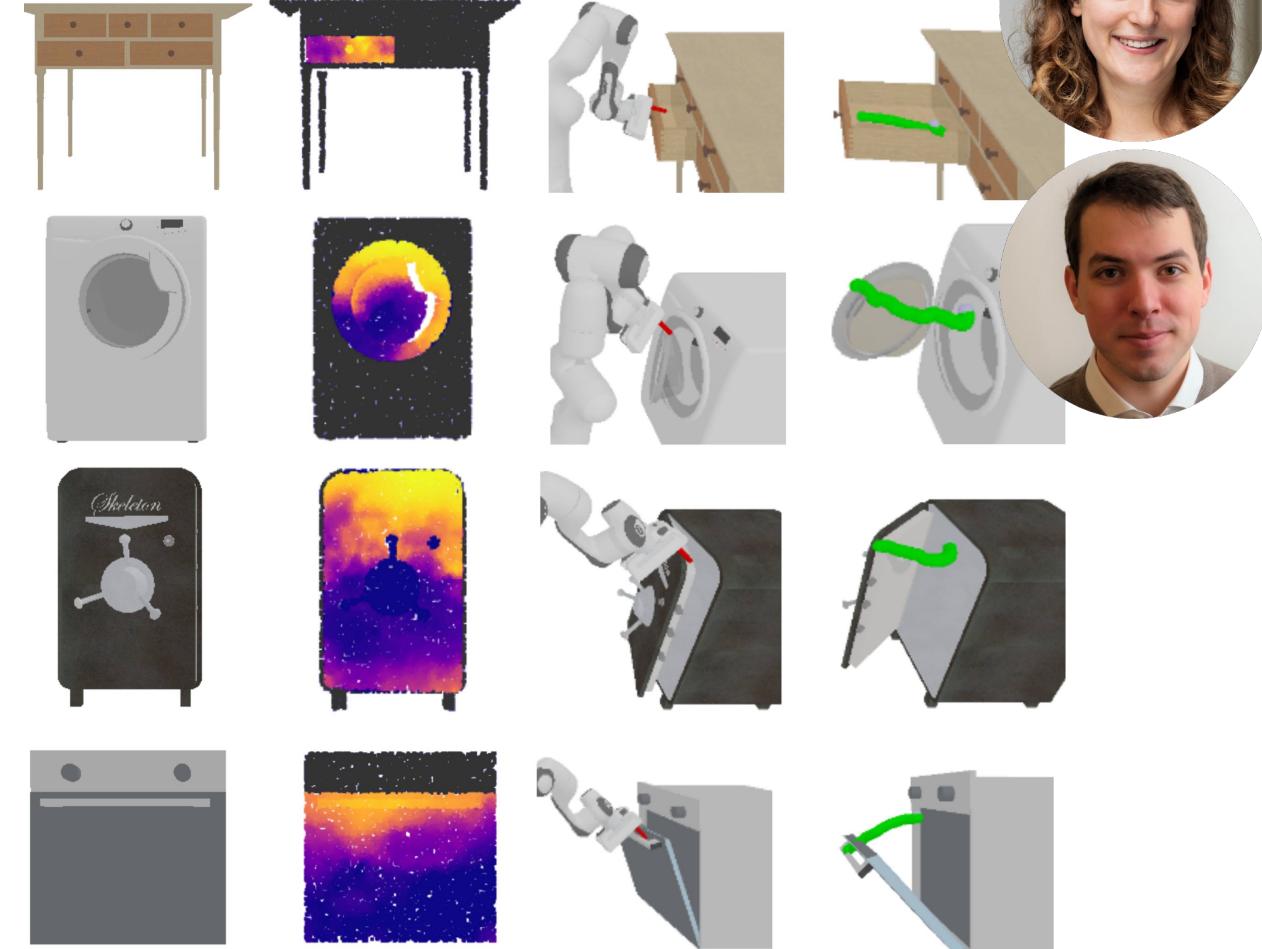


# Opening task

Agent agnostic (VAT-Mart)

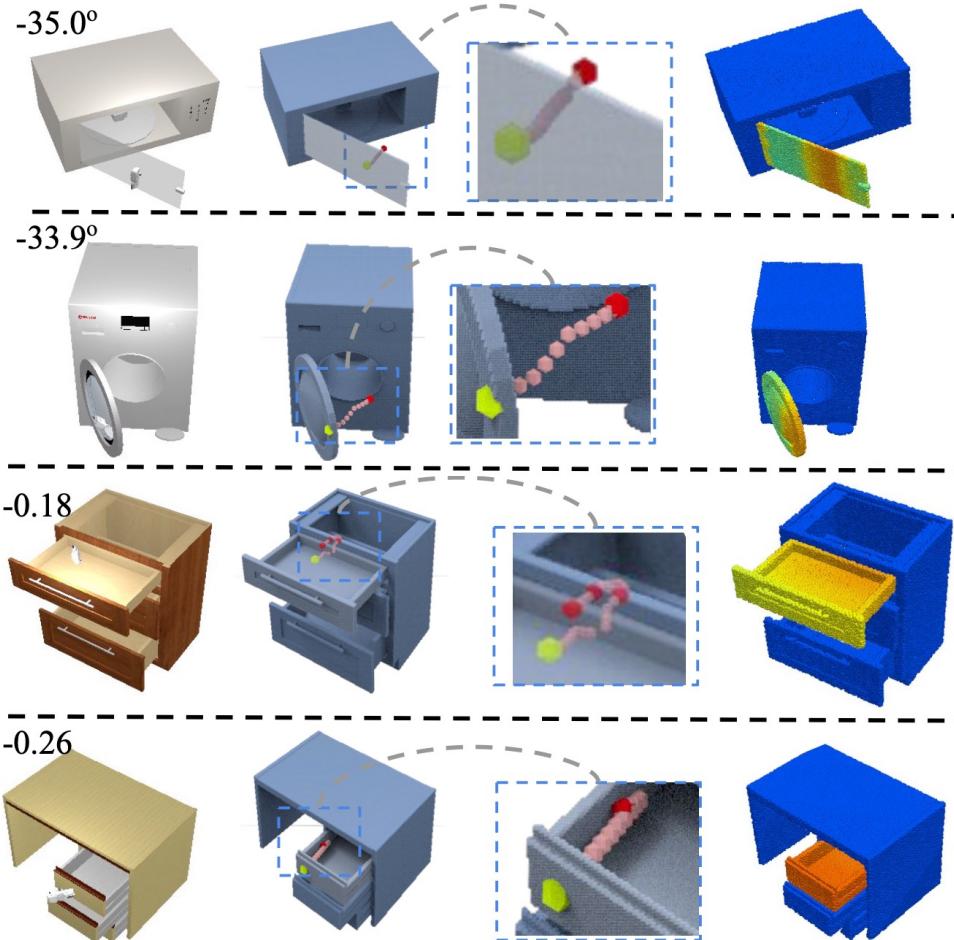


Agent-aware (Ours)

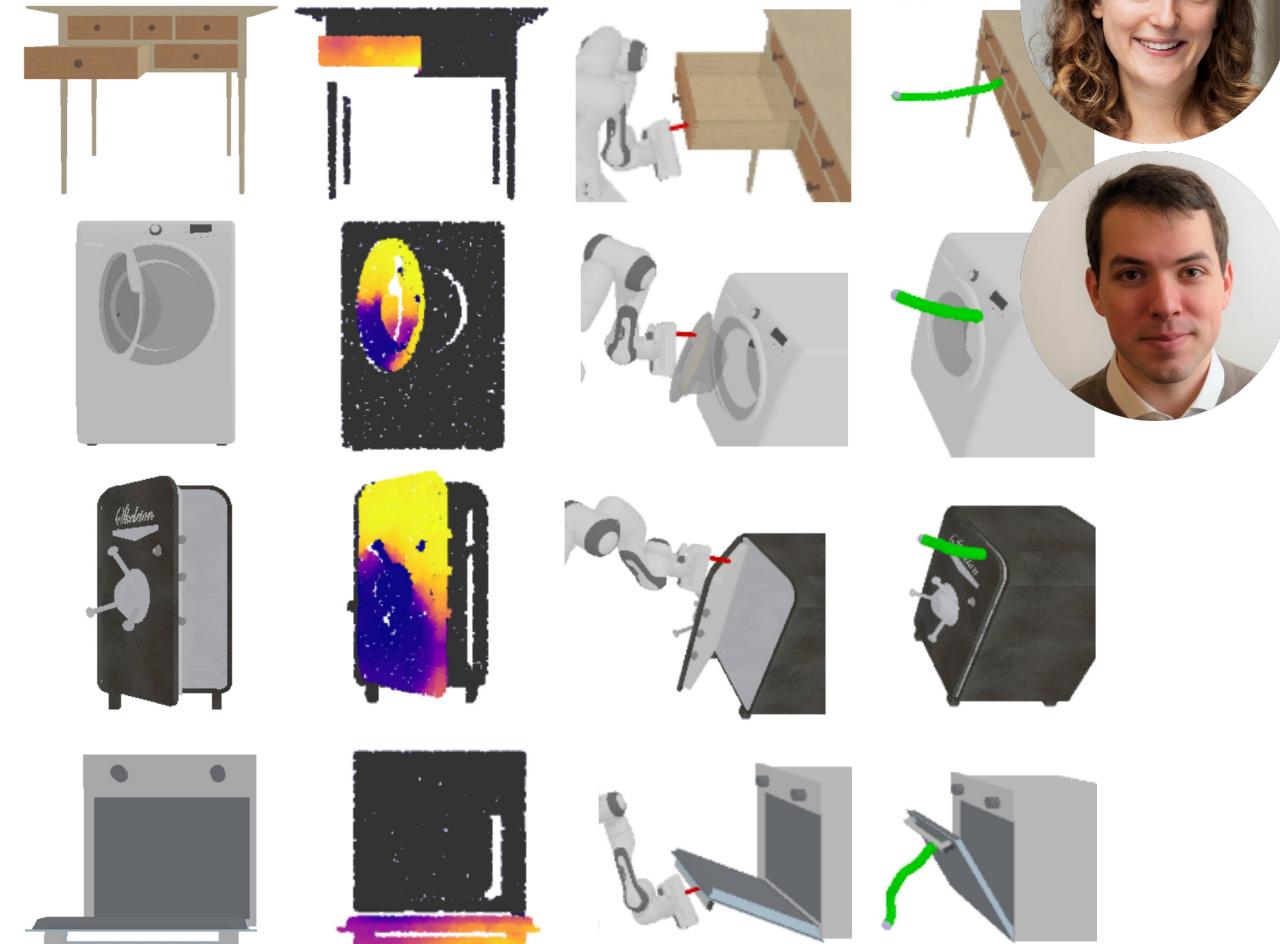


# Closing task

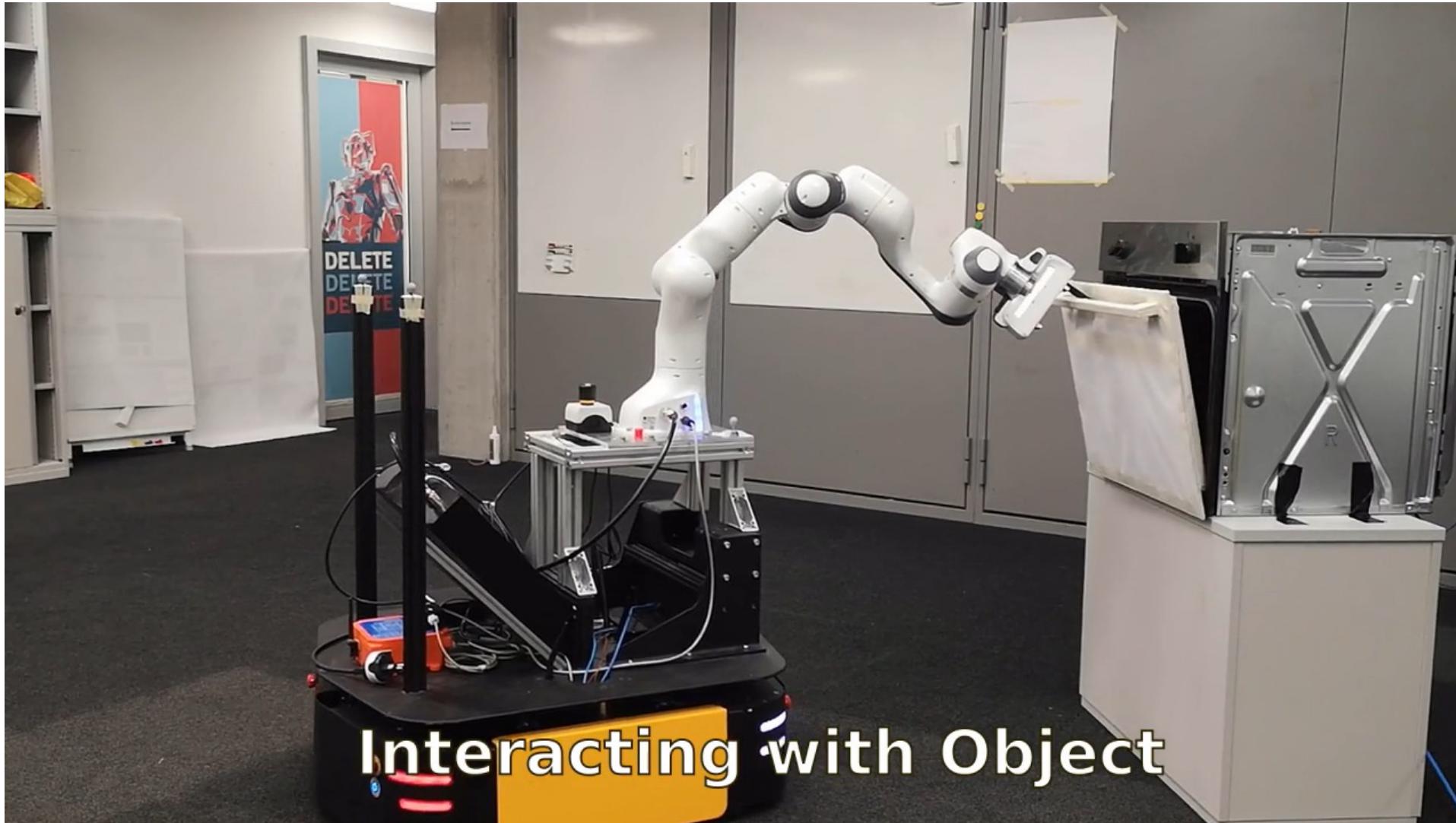
Agent agnostic (VAT-Mart)



Agent-aware (Ours)

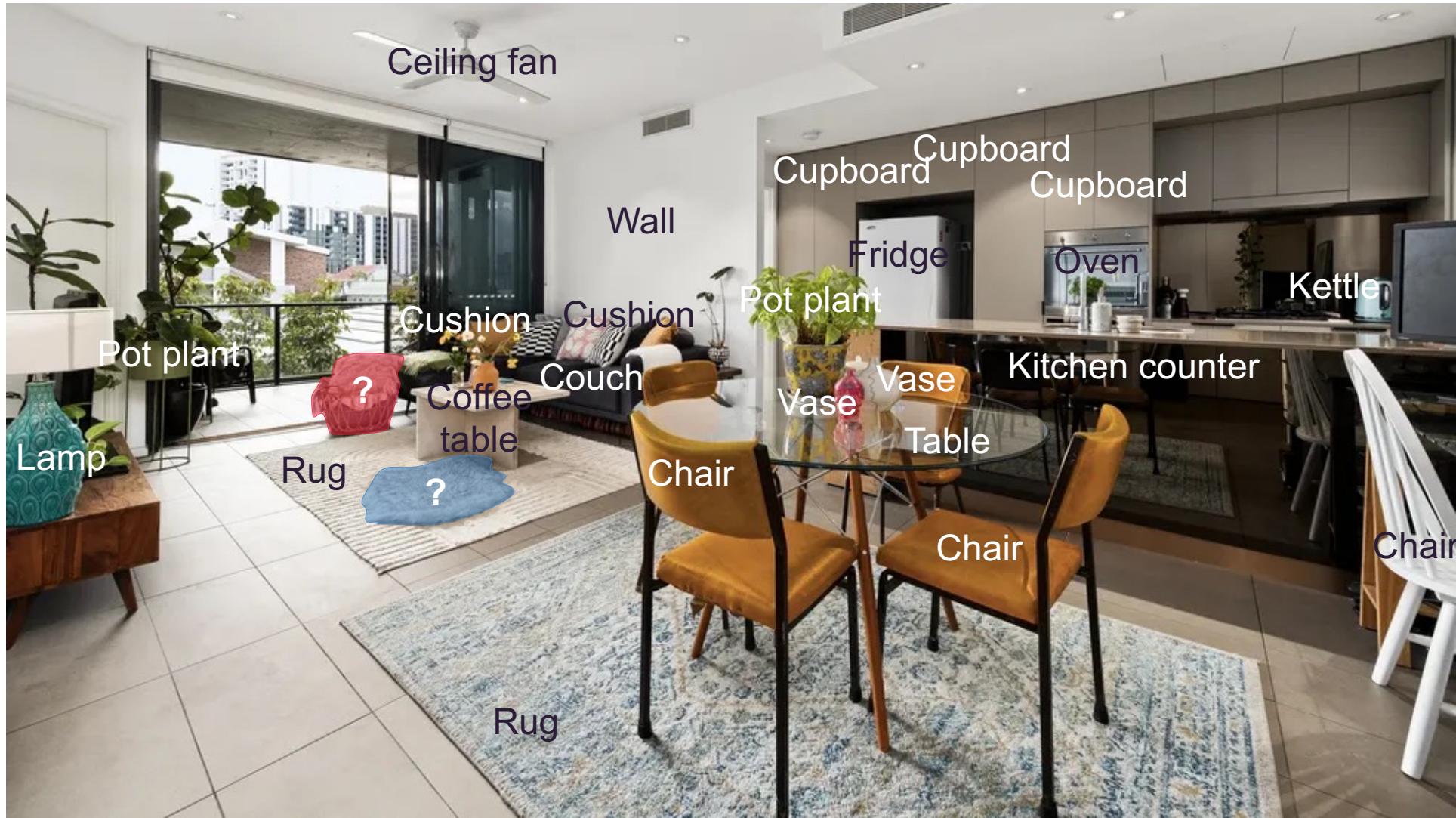


# Agent-aware affordance learning



Schiavi et al., "Learning agent-aware affordances for closed-loop interaction with articulated objects", ICRA 2023

# Learning *what* to interact with

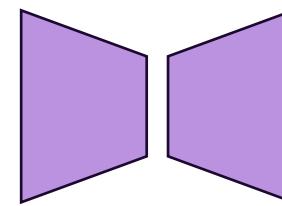
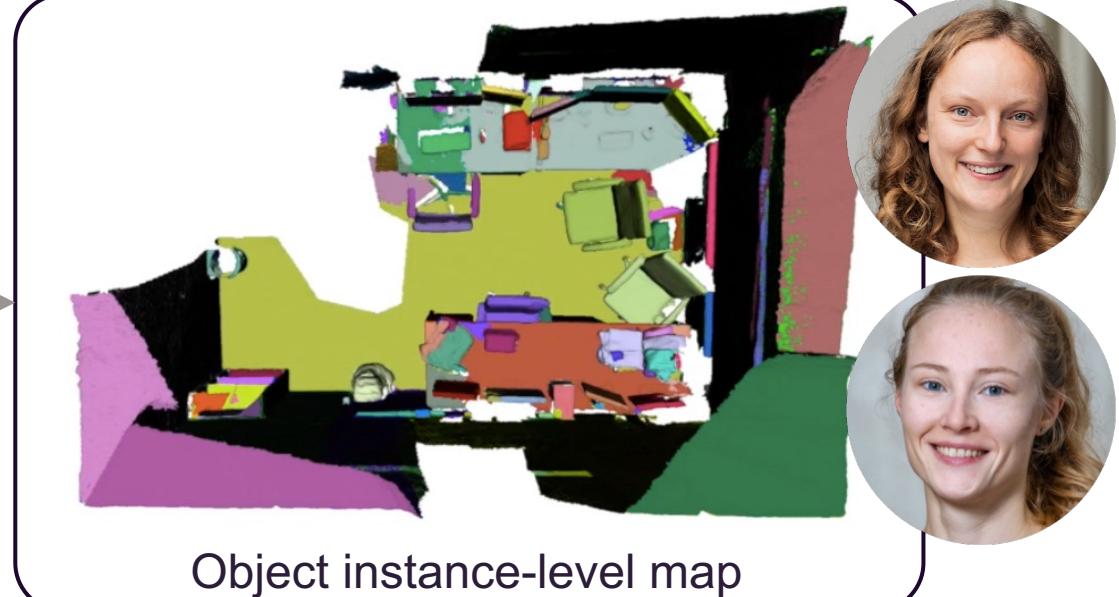


# Exploring interactions in an object-level map



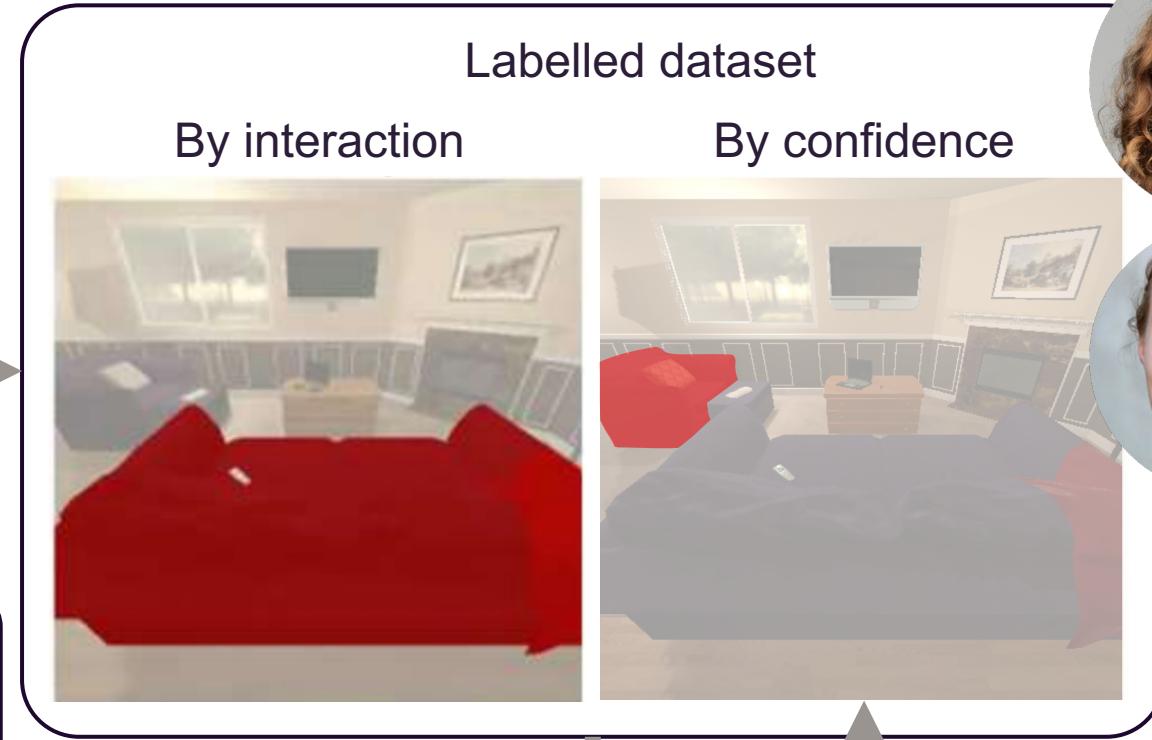
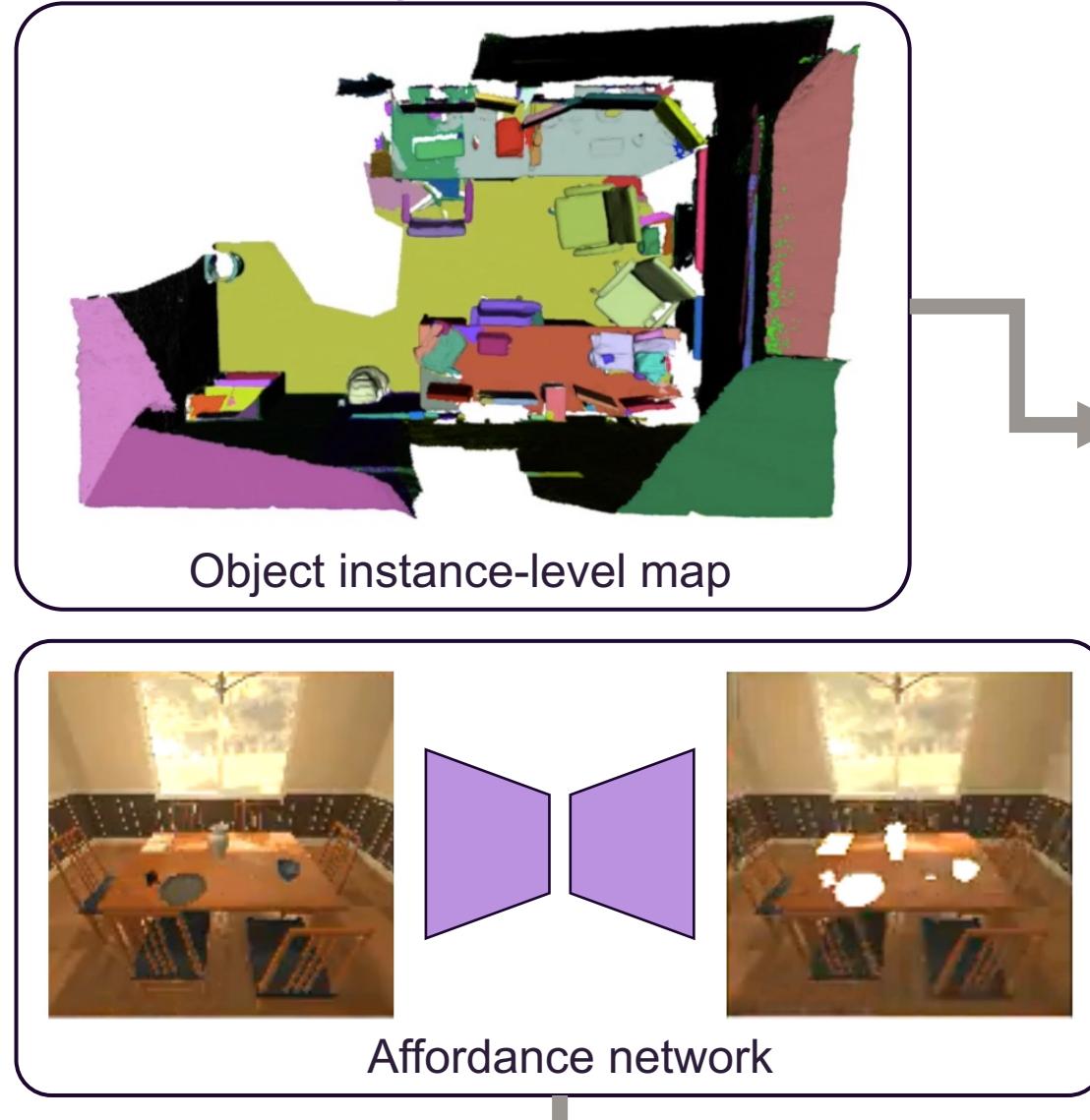
iTHOR simulation environment

- ✓ Robot pose
- ✓ Instance mask
- ✓ RGB-D
- ✓ Interaction result

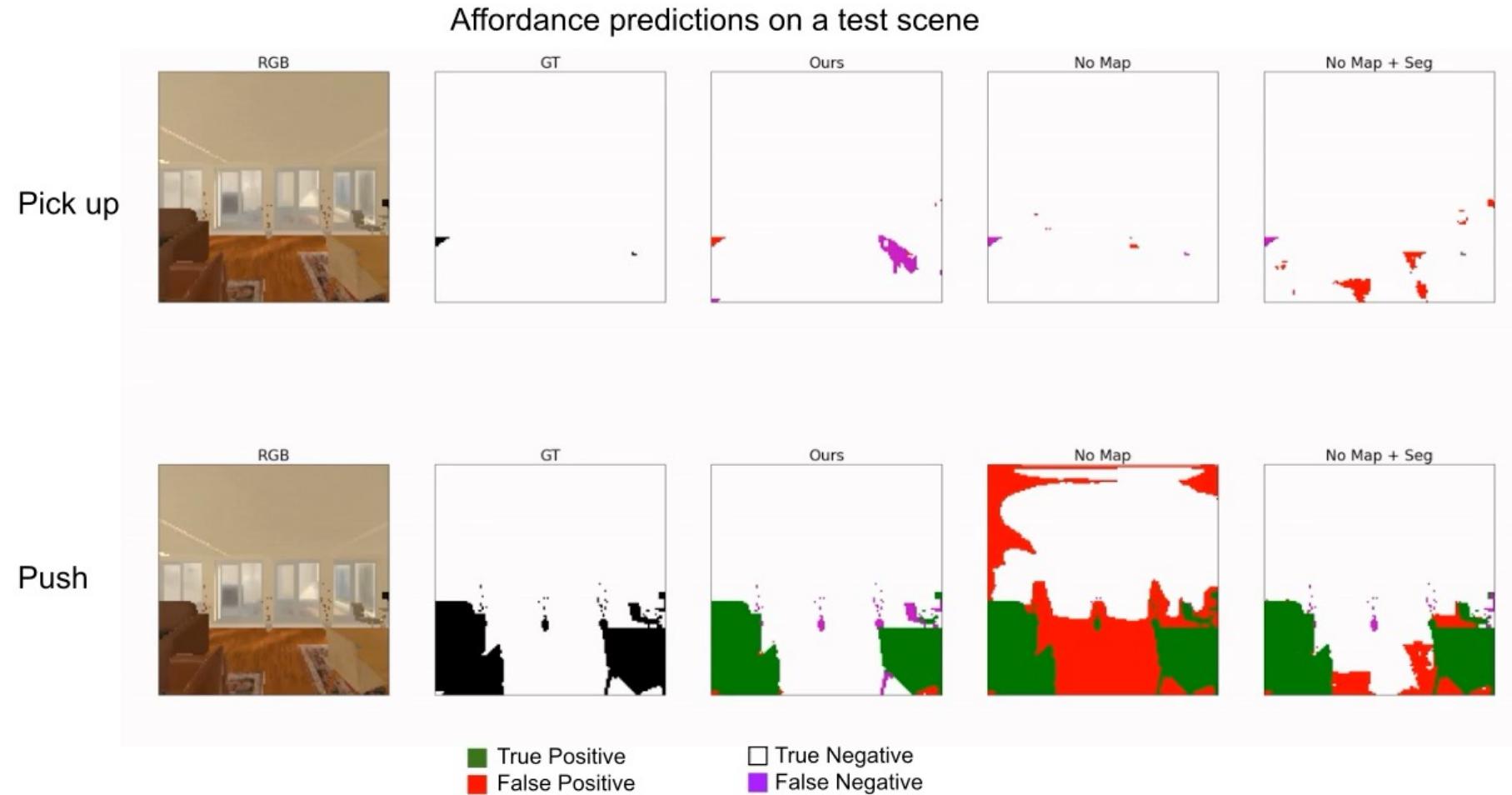


Affordance network

# Exploring interactions in an object-level map



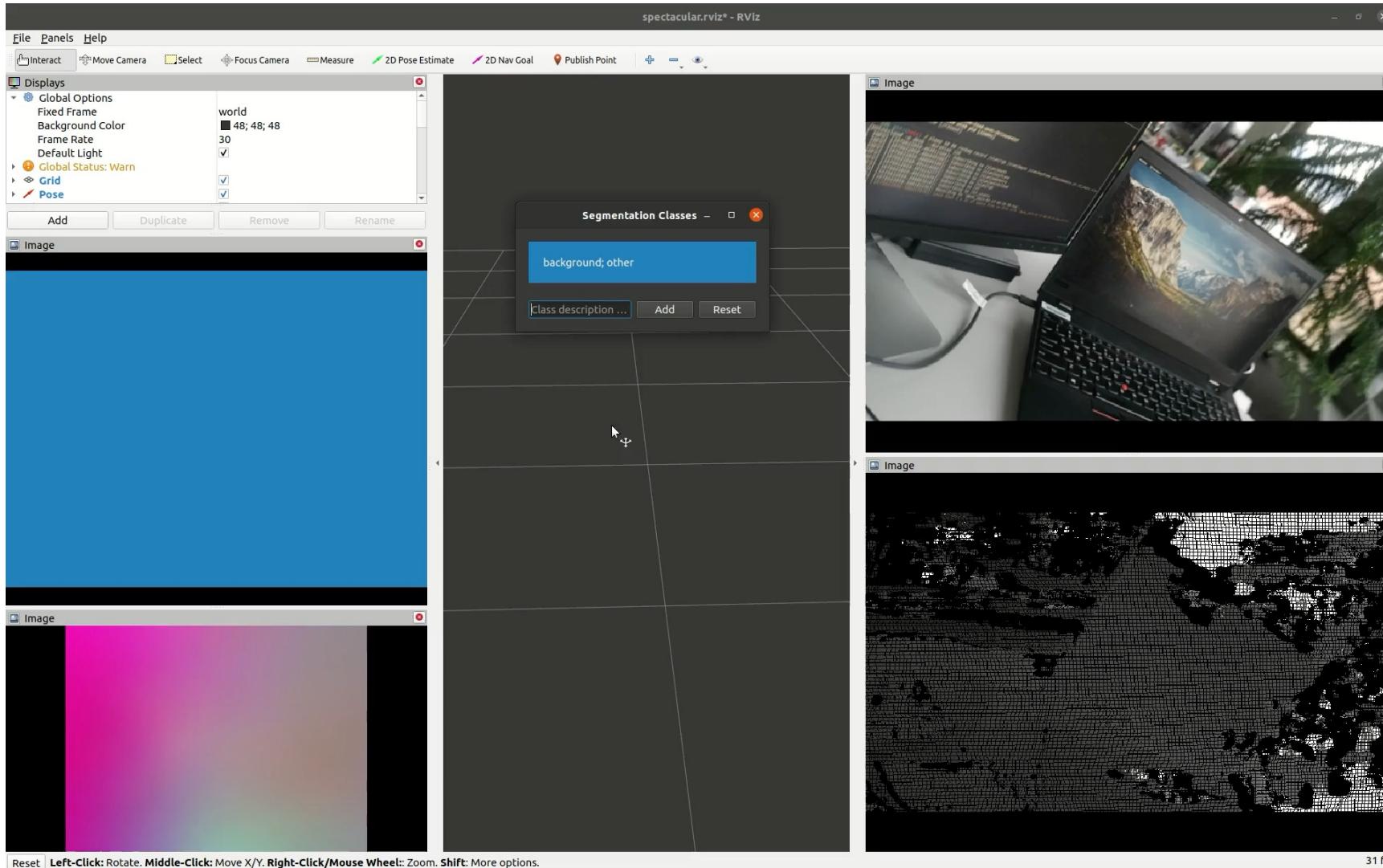
# Learned affordances from interactive exploration



# One feature field to rule them all?



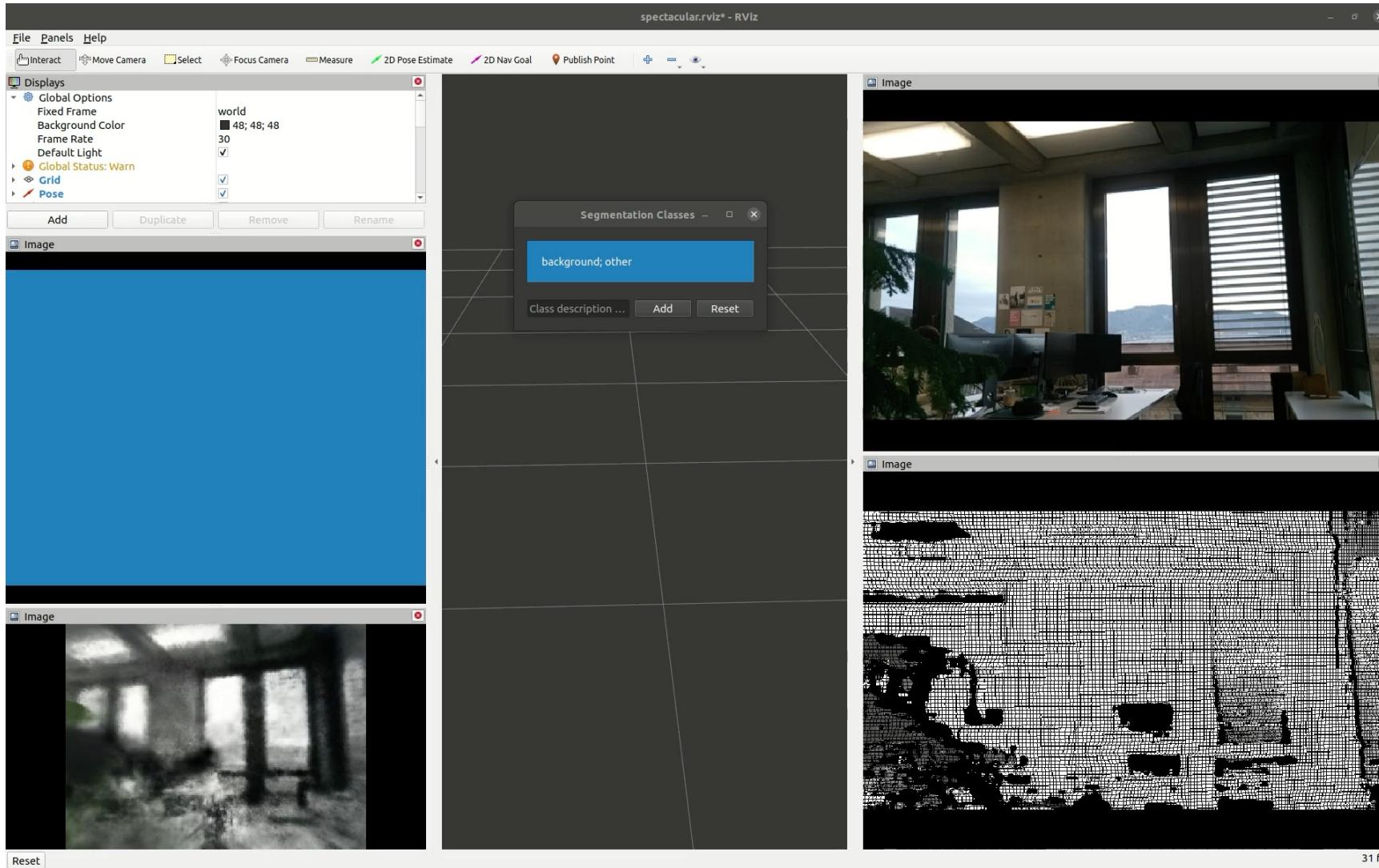
# Neural implicit vision-language feature fields



Blomqvist et al., “Neural implicit vision-language feature fields”, IROS 2023



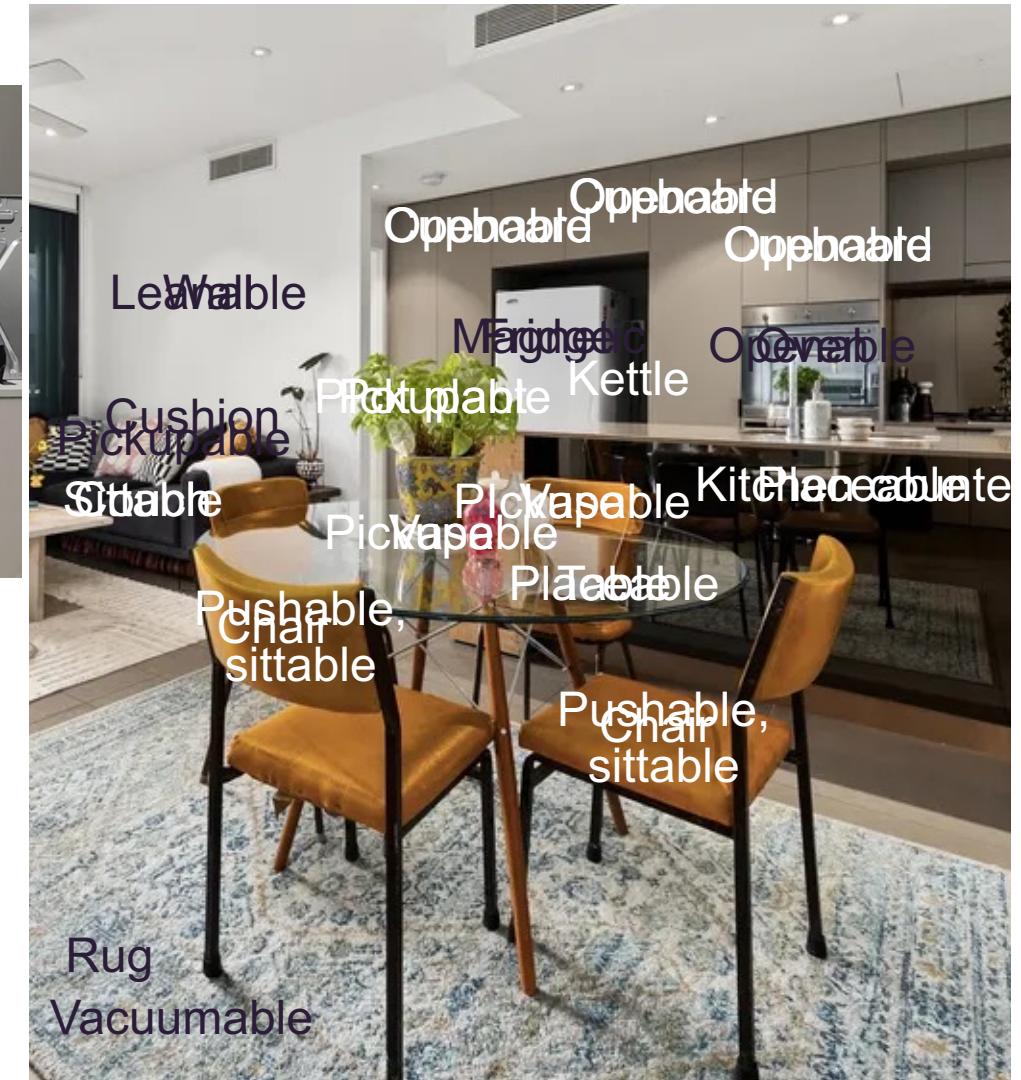
# Neural implicit vision-language feature fields



Blomqvist et al., "Neural implicit vision-language feature fields", IROS 2023



# Can we move beyond semantics and reach affordances?



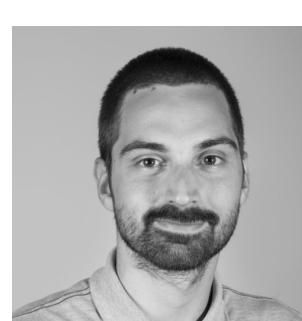
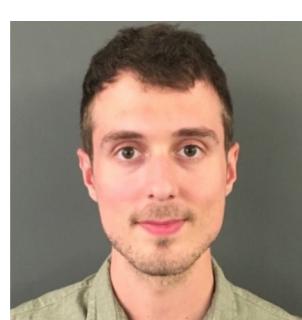
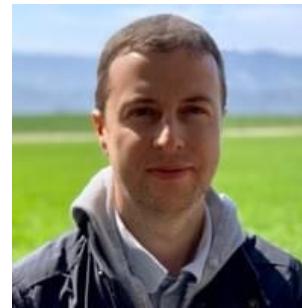


**Harmony**

Assistive robots for healthcare



**CROWDBOT**



# Getting Robots to Touch Things Appropriately

Jen Jen Chung

Robotic Perception, Planning and Learning Lab

[jenjen.chung@uq.edu.au](mailto:jenjen.chung@uq.edu.au)

