The coupling of perception and interaction

For object discovery and understanding

Jen Jen Chung, Francesco Milano

14 October 2024

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ETH zürich

Why dense object instance-aware scene reconstruction?

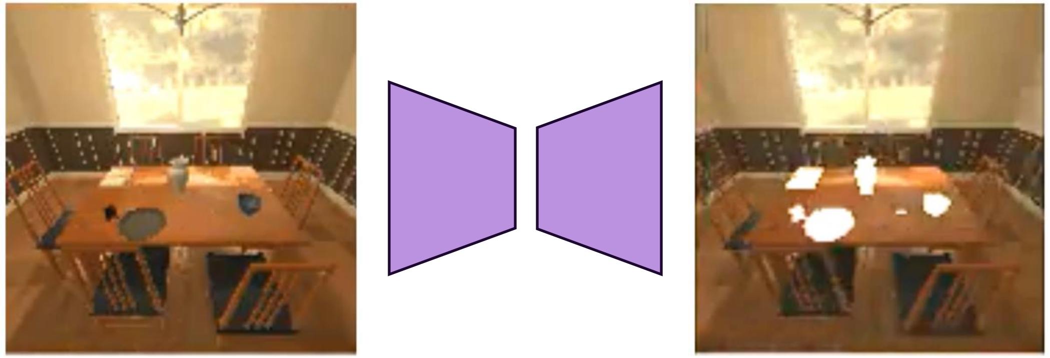
Grinvald et al., “Volumetric instance-aware semantic mapping and 3D object discovery”, RAL 2019

Exploring interactions in an object-level map

iTHOR simulation environment

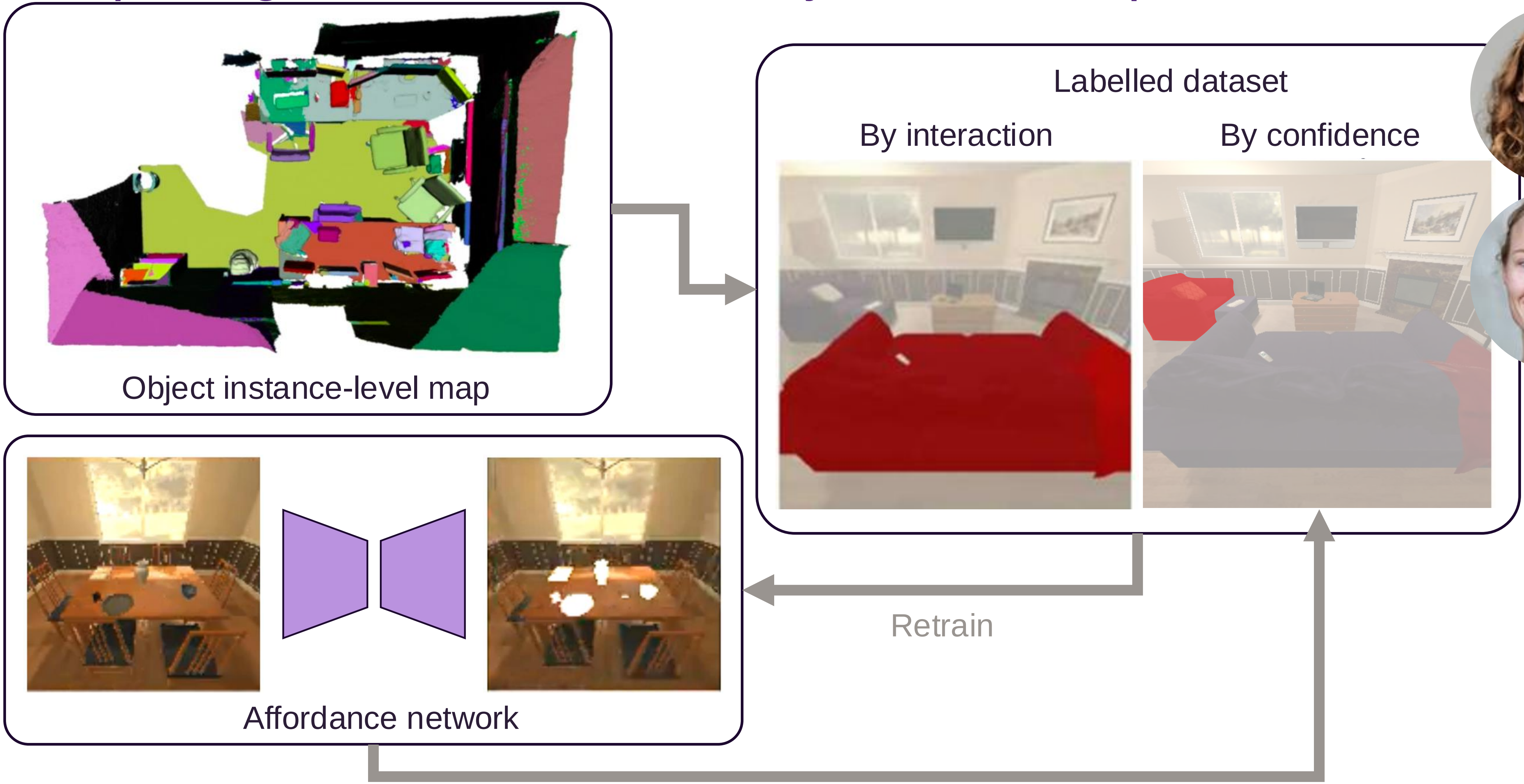
- ✓ Robot pose
- ✓ RGB-D
- ✓ Instance mask
- ✓ Interaction result

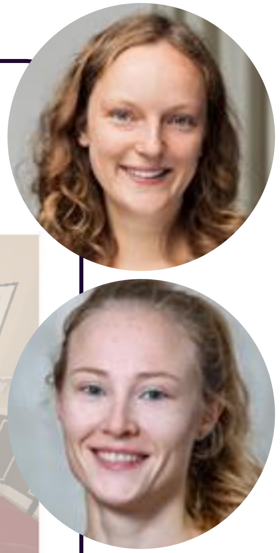
Object instance-level map



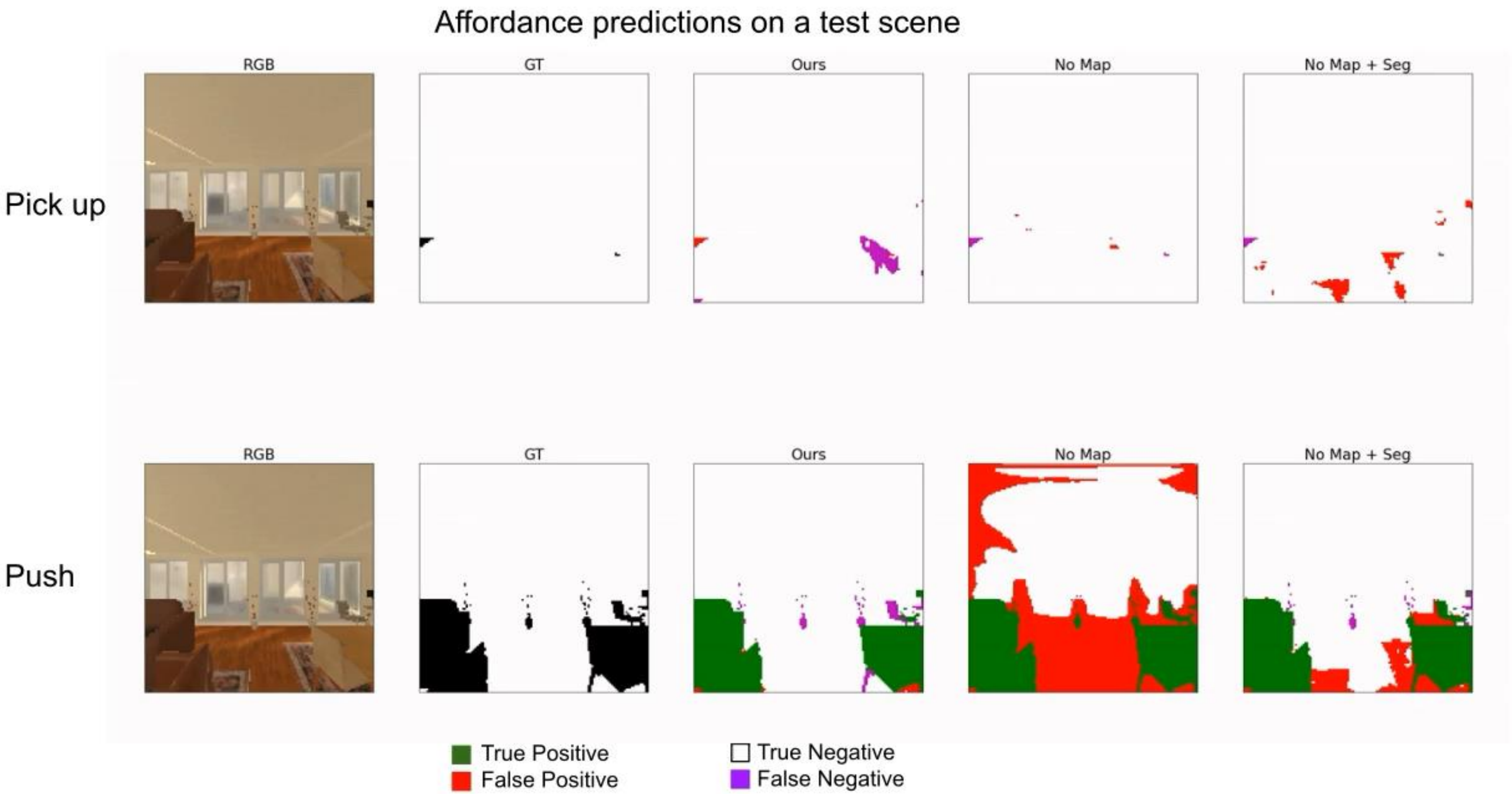
Affordance network

Exploring interactions in an object-level map



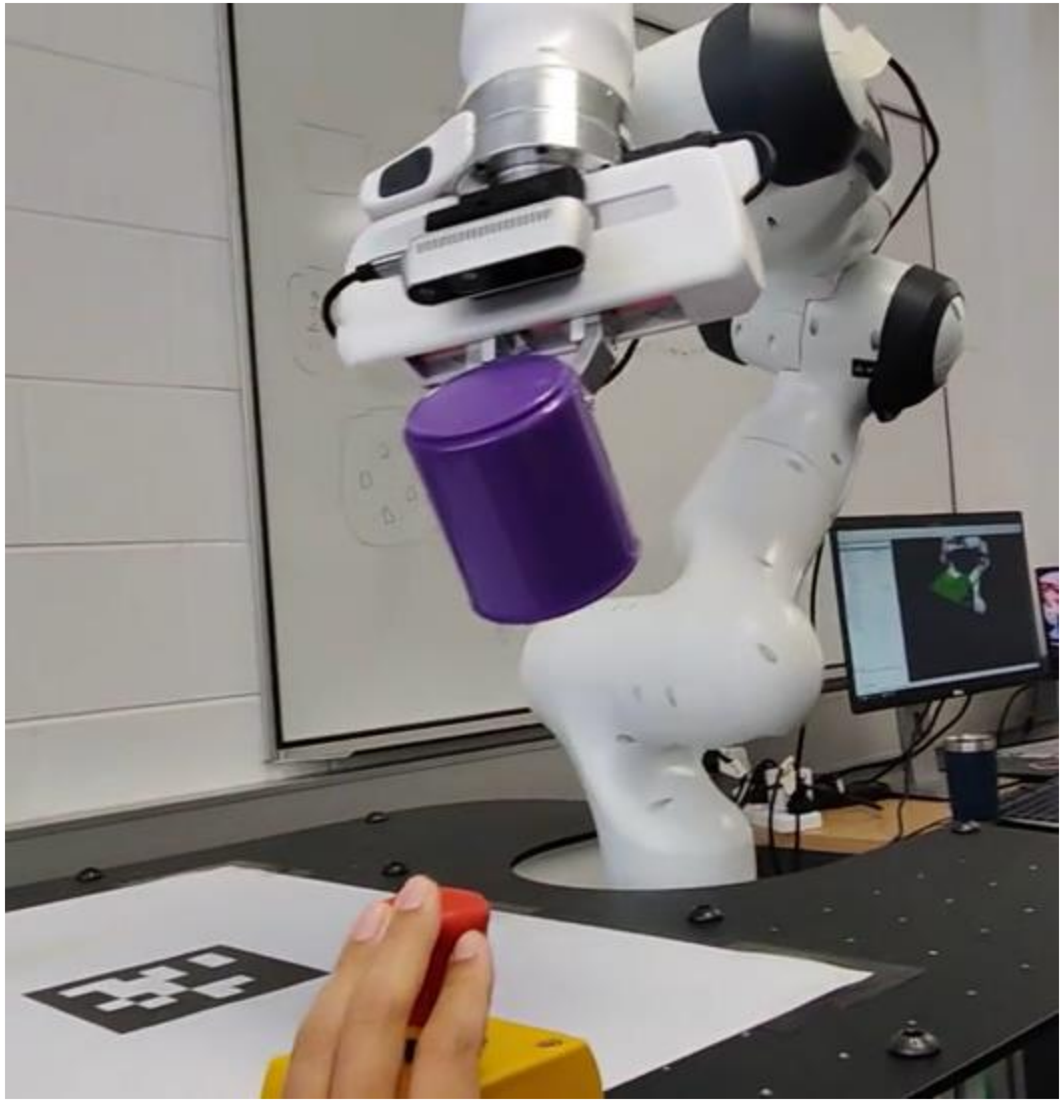


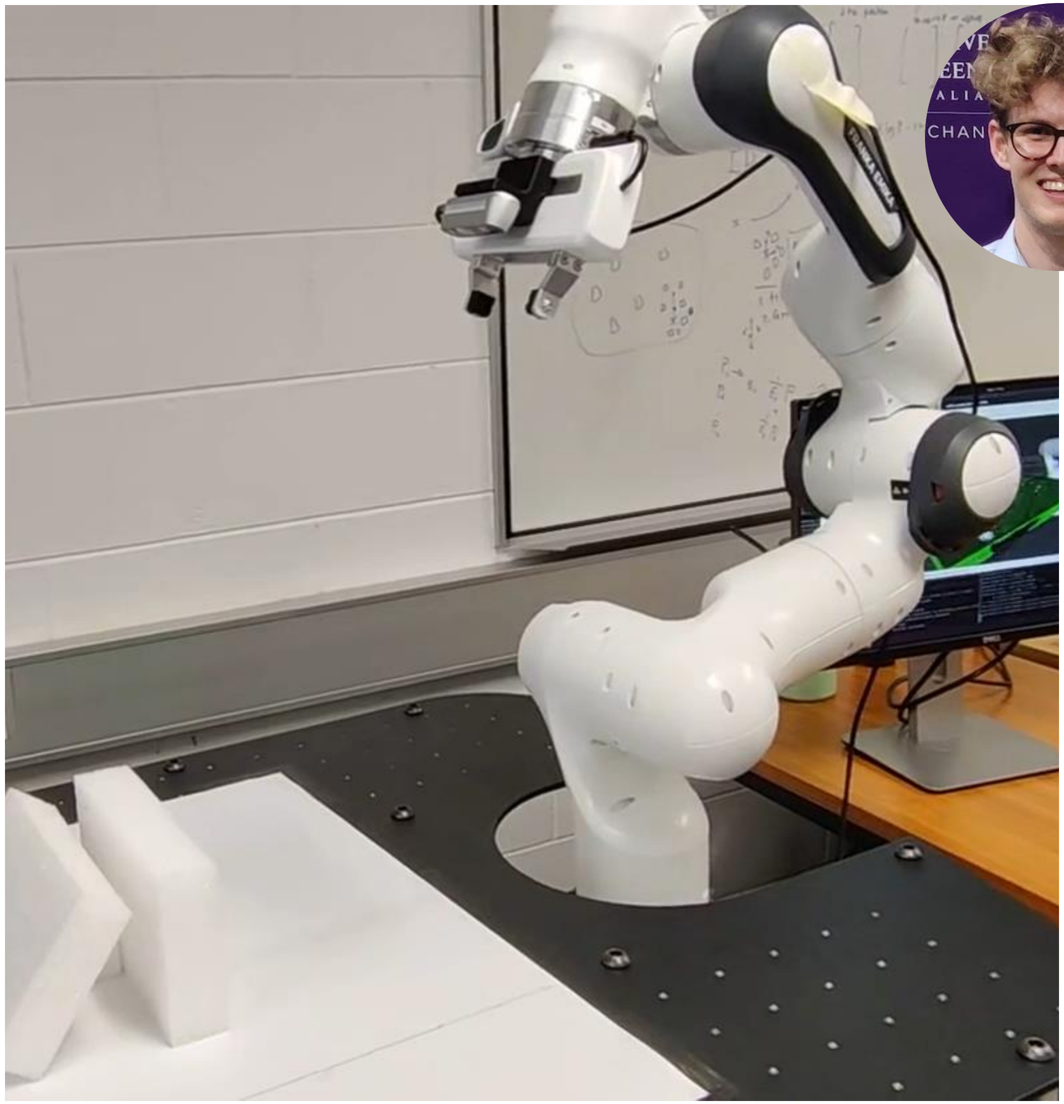
Learned affordances from interactive exploration



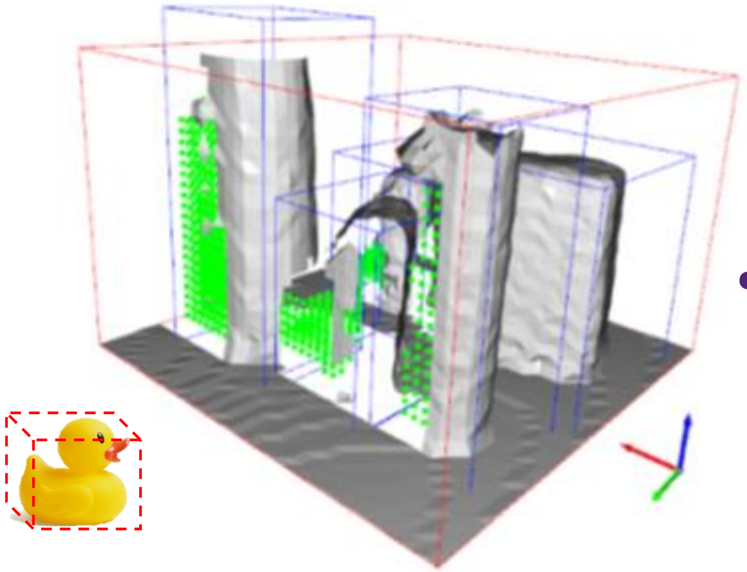
Wulkop et al., “Learning affordances from interactive exploration using an object-level map”, ISRR 2024

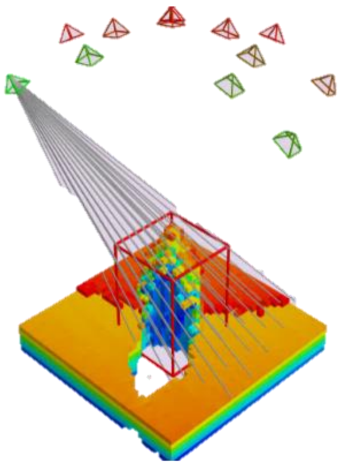
Finding and retrieving hidden objects



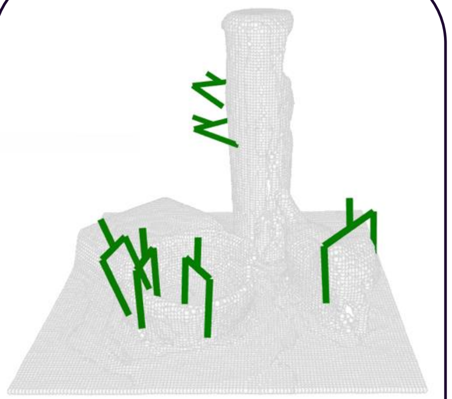


To grasp or not to grasp?





Viewing options

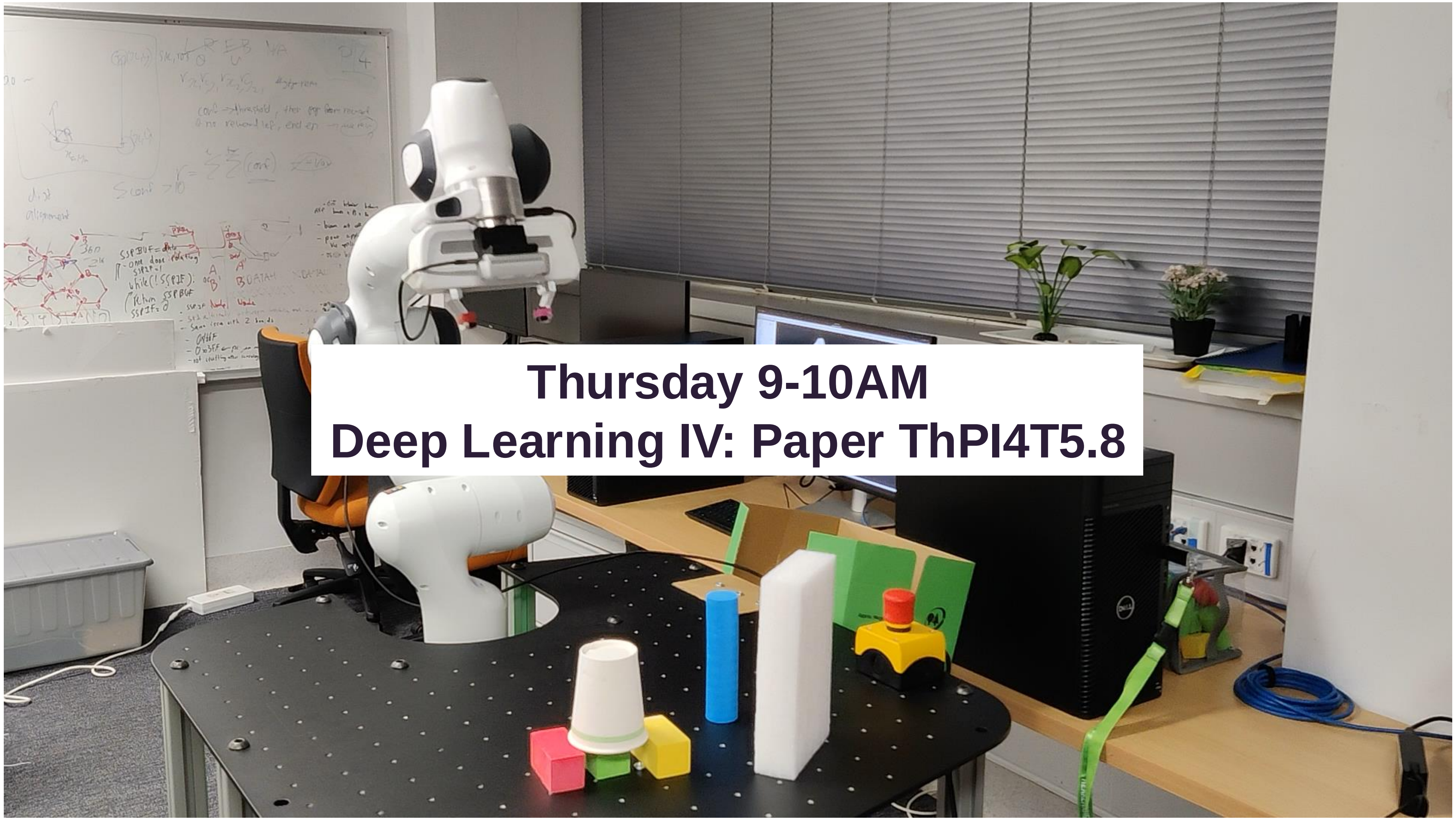


Grasping options

Action values learned via RL

$$R = \omega \frac{\Delta \text{ Possible target locations}}{\text{Possible target locations}} - t$$

Active search and grasp in clutter



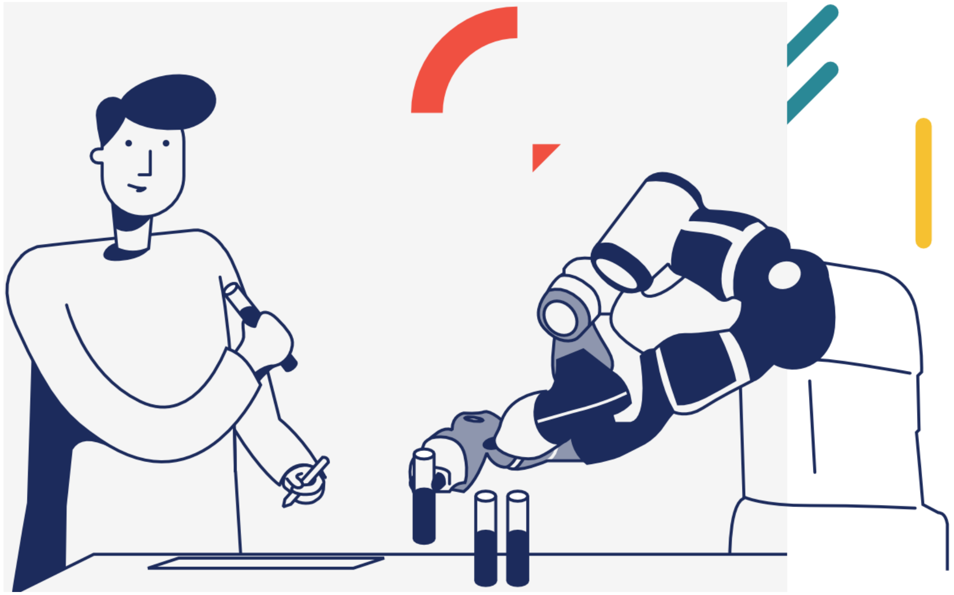
Thursday 9-10AM

Deep Learning IV: Paper ThPI4T5.8

Pitcher et al., "Reinforcement learning for active search and grasp in clutter", IROS 2024

Object-level representations for robotic interaction

Harmony: Assistive robots for healthcare



European Union’s Horizon 2020 research and innovation programme, grant agreement No. 101017008 - **https://harmony-eu.org/**

Harmony: Assistive robots for healthcare

European Union’s Horizon 2020 research and innovation programme, grant agreement No. 101017008 - **https://harmony-eu.org/**

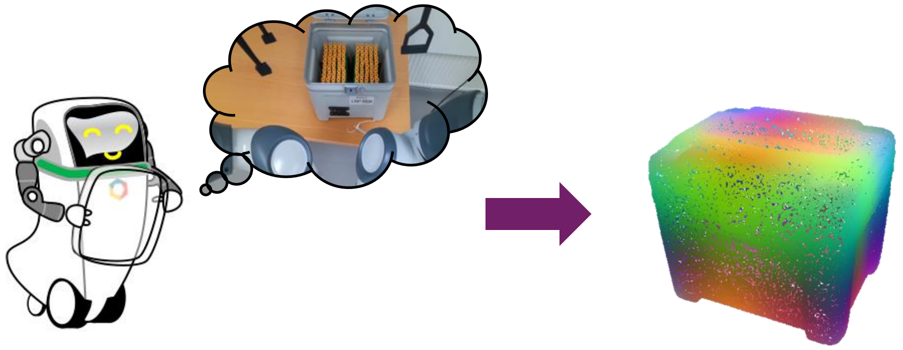
Harmony: Assistive robots for healthcare

European Union's Horizon 2020 research and innovation programme, grant agreement No. 101017008 - **https://harmony-eu.org/**

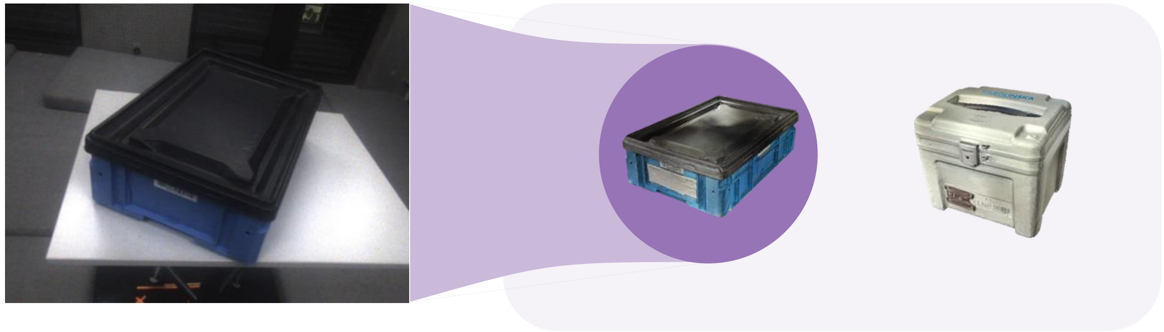
Perception in support of robotic interaction

Desiderata:

Accurate reconstruction (geometry, appearance)



Flexibility: Encode **task-specific properties**

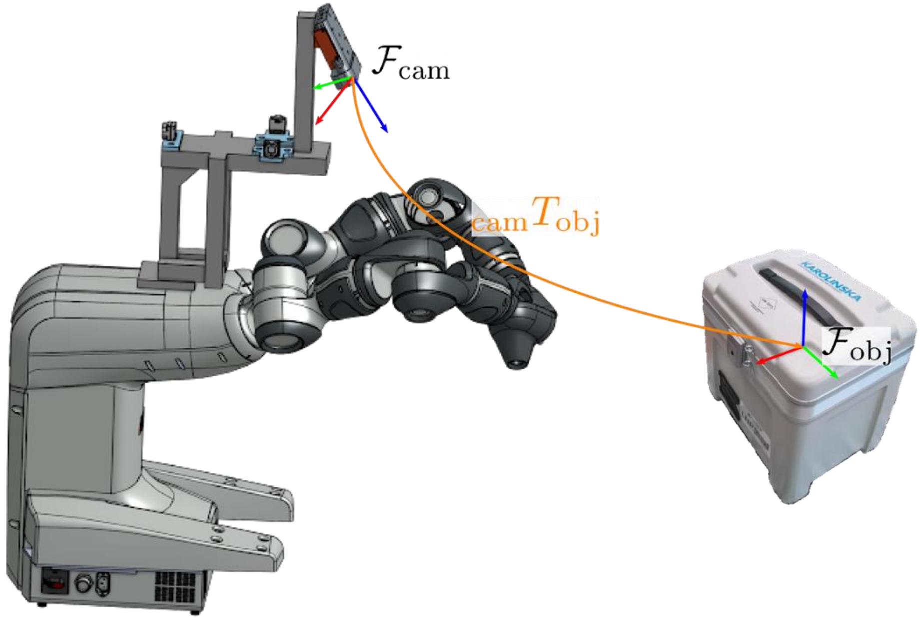


Ability to **easily incorporate new representations**

How? Our hypothesis: **Neural Fields** + **Neural Rendering**

Perception in support of robotic interaction

One example task: 6-DoF Object Pose Estimation

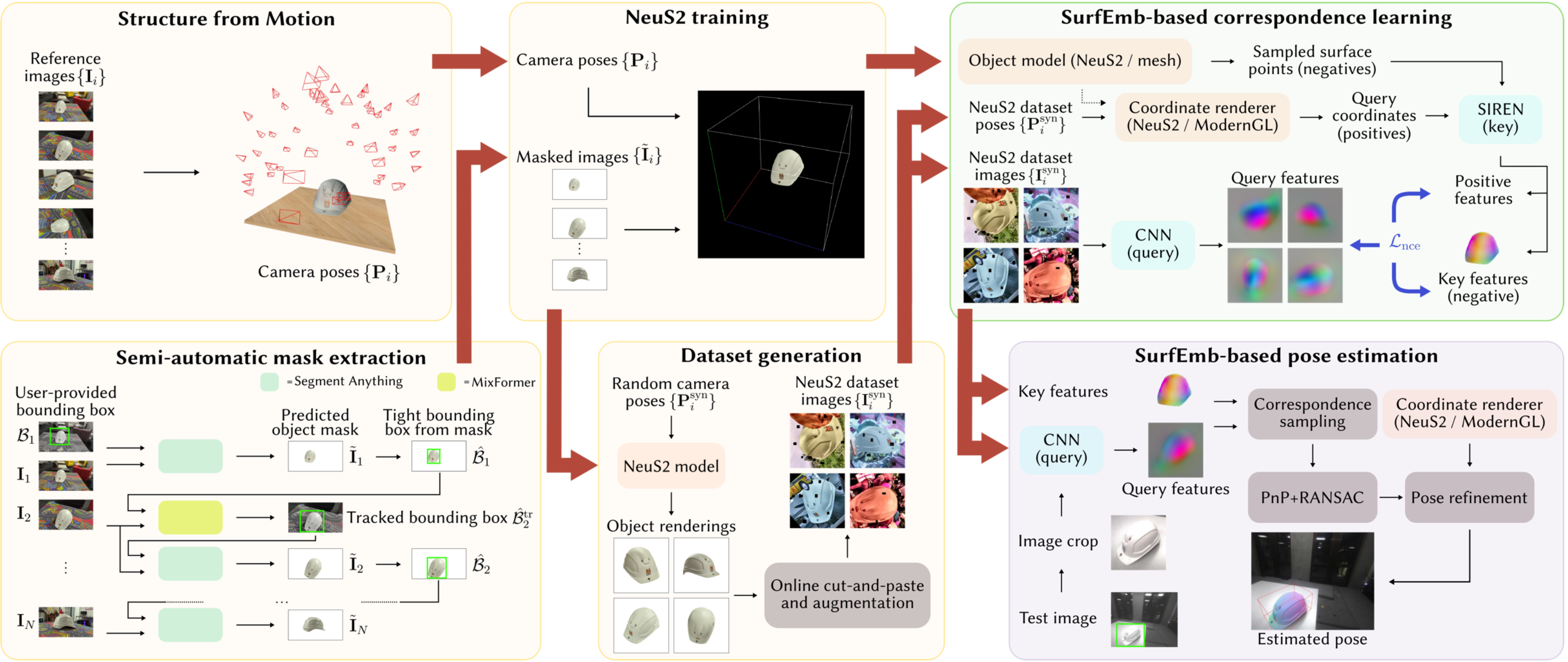


- State-of-the-art approaches rely on **textured CAD models** and **photorealistic synthetic datasets** (PBR)

How can neural fields and neural rendering help?

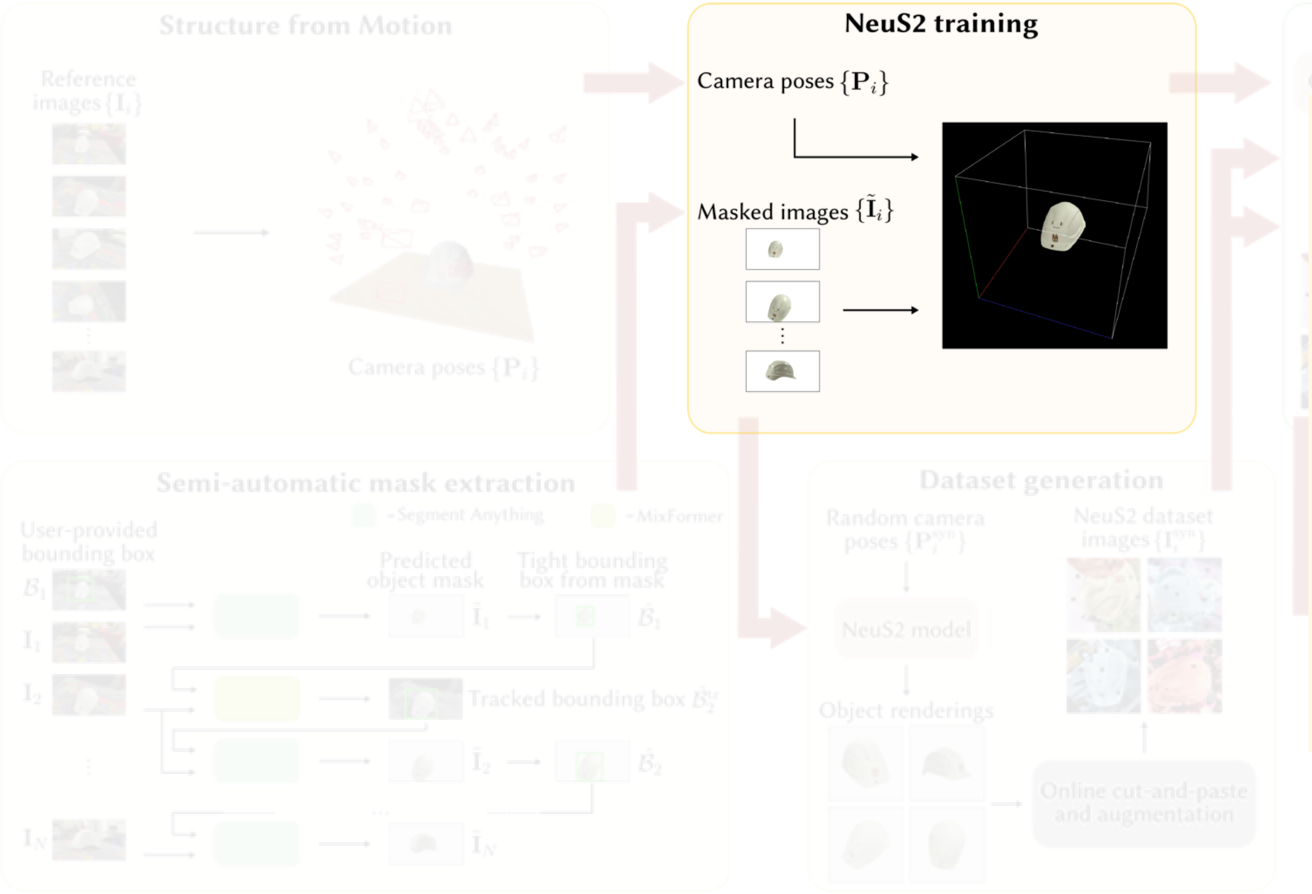
ETH zürich

Neural fields for object pose estimation – NeuSurfEmb



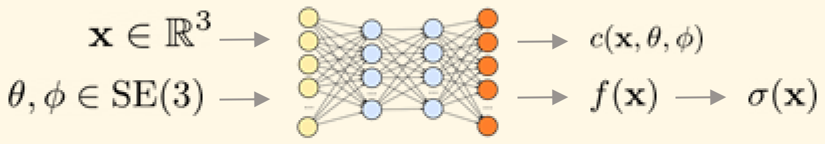
Milano et al., "NeuSurfEmb: A Complete Pipeline for Dense Correspondence-based 6D Object Pose Estimation without CAD Models", IROS 2024

Neural fields for object pose estimation – NeuSurfEmb



TL;DR:

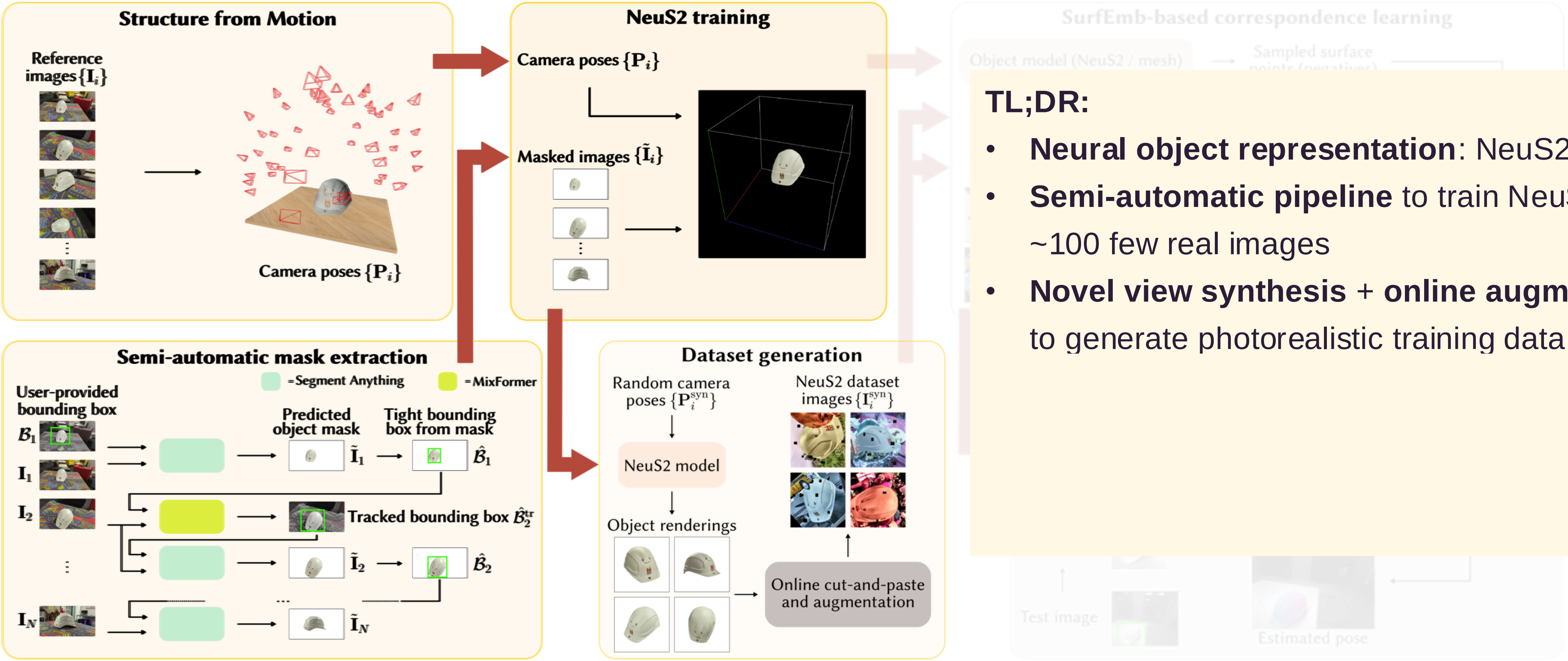
- **Neural object representation**: NeuS2



Milano et al., “NeuSurfEmb: A Complete Pipeline for Dense Correspondence-based 6D Object Pose Estimation without CAD Models”, IROS 2024

ETH zürich

Neural fields for object pose estimation – NeuSurfEmb



TL;DR:

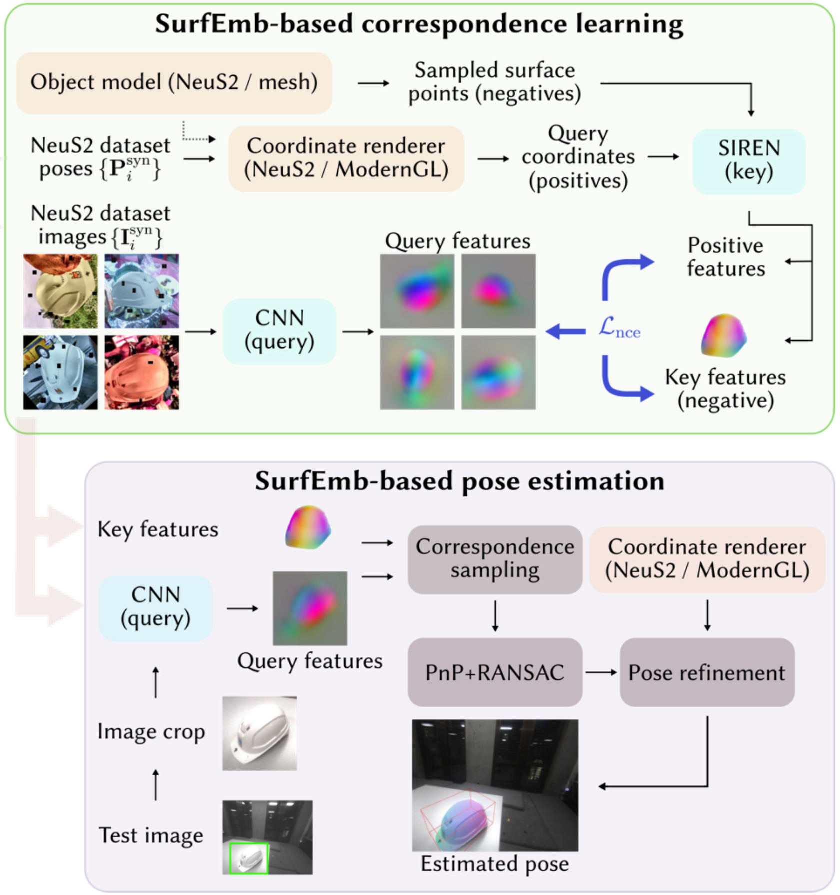
- **Neural object representation**: NeuS2
- **Semi-automatic pipeline** to train NeuS2 from ~100 few real images
- **Novel view synthesis** + **online augmentation** to generate photorealistic training data

Milano et al., "NeuSurfEmb: A Complete Pipeline for Dense Correspondence-based 6D Object Pose Estimation without CAD Models", IROS 2024

Neural fields for object pose estimation – NeuSurfEmb

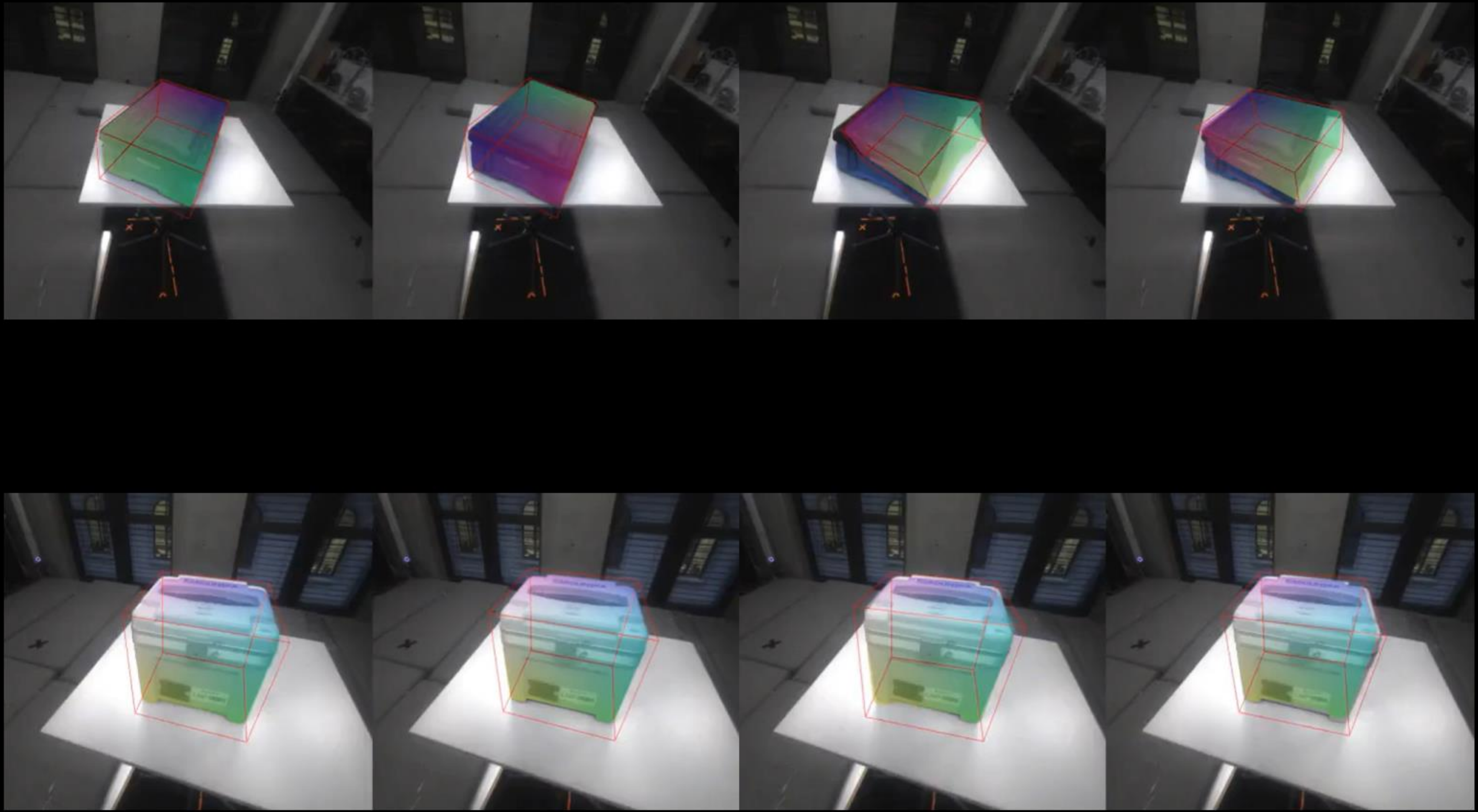
TL;DR:

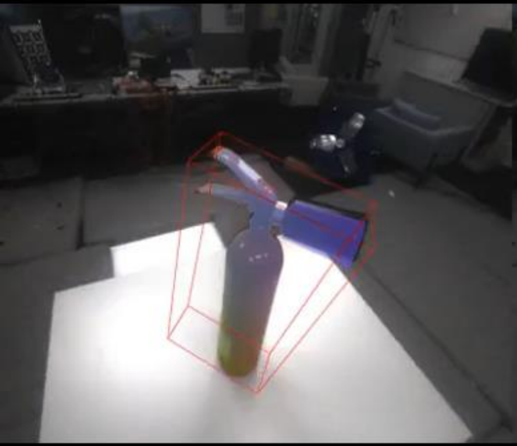
- **Neural object representation**: NeuS2
- **Semi-automatic pipeline** to train NeuS2 from ~100 few real images
- **Novel view synthesis** + **online augmentation** to generate photorealistic training data
- A **feature field** is contrastively trained with a CNN to **learn 2D-3D correspondences** (SurfEmb)
- **Correspondence-based pose estimation**



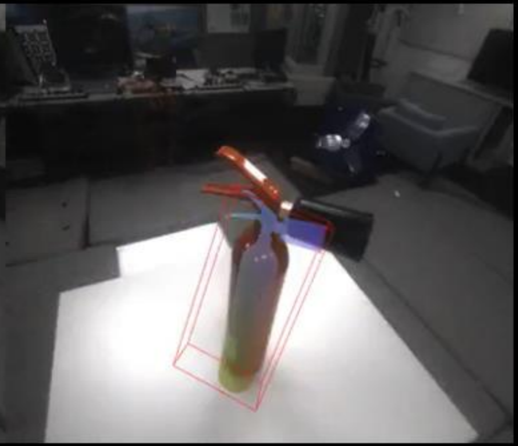
Milano et al., "NeuSurfEmb: A Complete Pipeline for Dense Correspondence-based 6D Object Pose Estimation without CAD Models", IROS 2024

NeuSurfEmb | OnePose++ | Gen6D (with tracking) | Gen6D (w/o tracking)

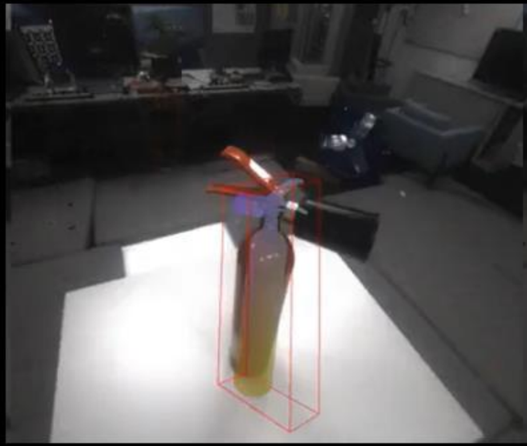




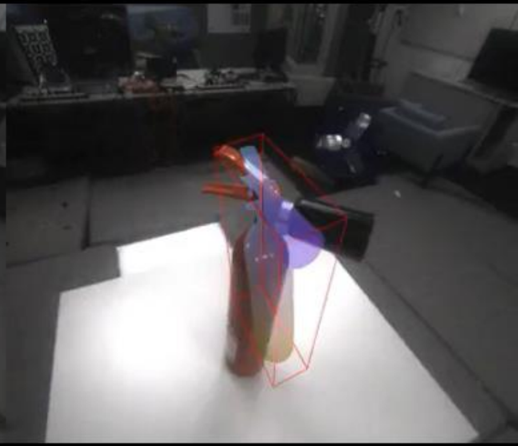
NeuSurfEmb



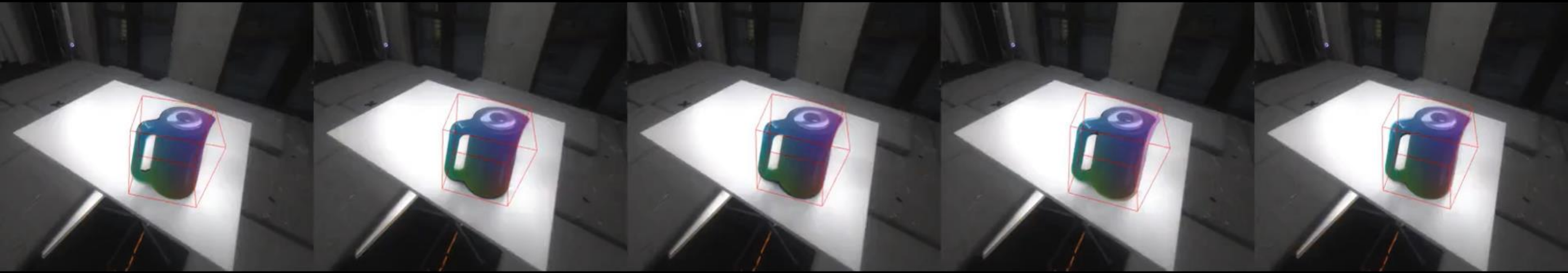
OnePose++ (w/o tracking, orig. recrop.)



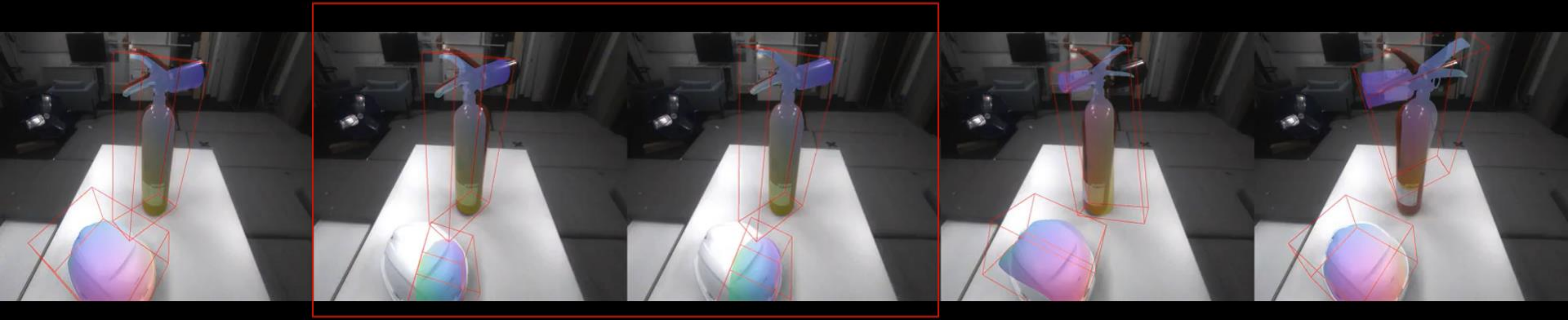
Gen6D (with tracking)



Gen6D (w/o tracking)

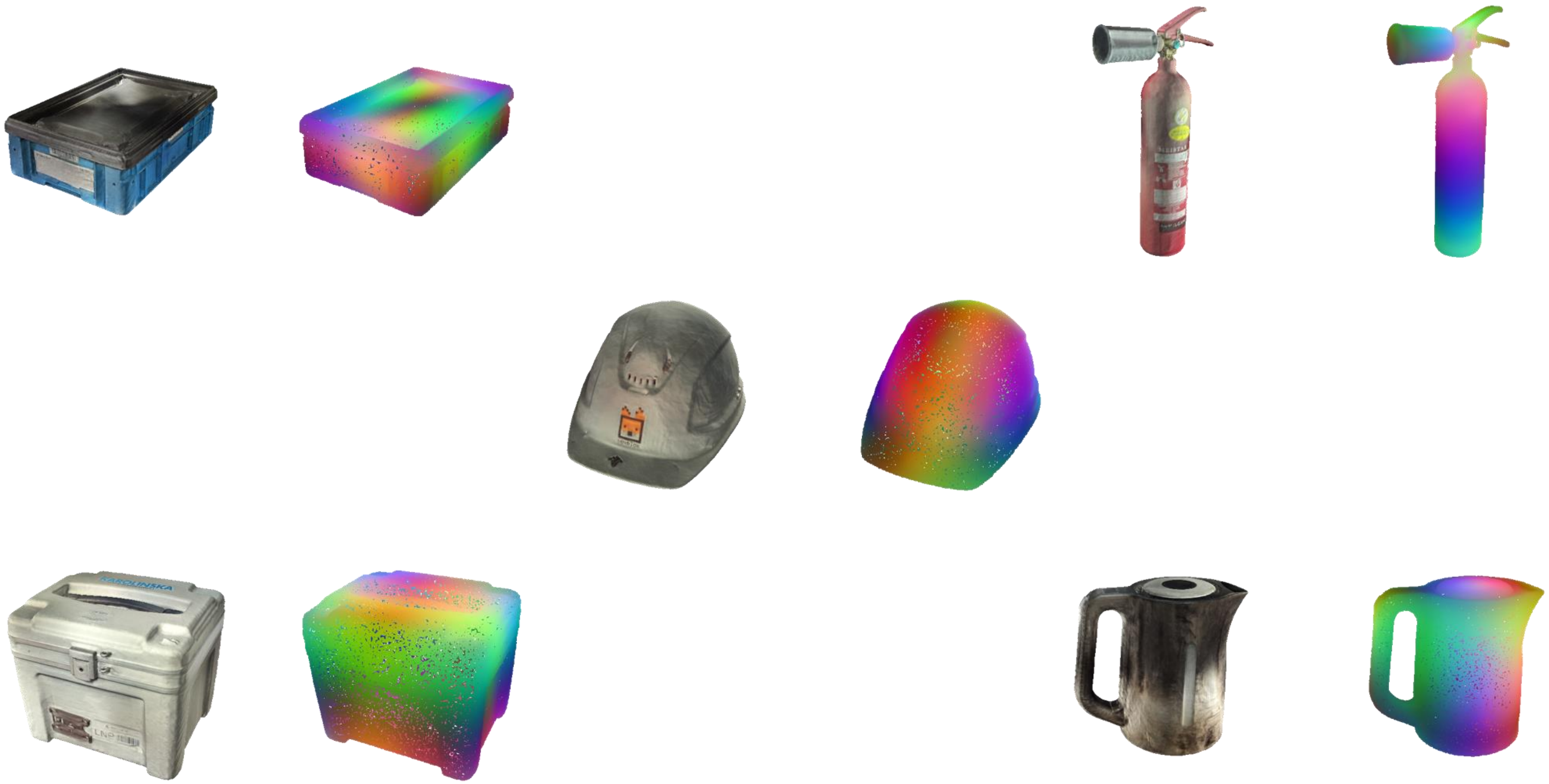


NeuSurfEmb | OnePose++ (w/o tracking, orig. recrop.) | OnePose++ (w/o tracking, prop. recrop.) | Gen6D (with tracking) | Gen6D (w/o tracking)



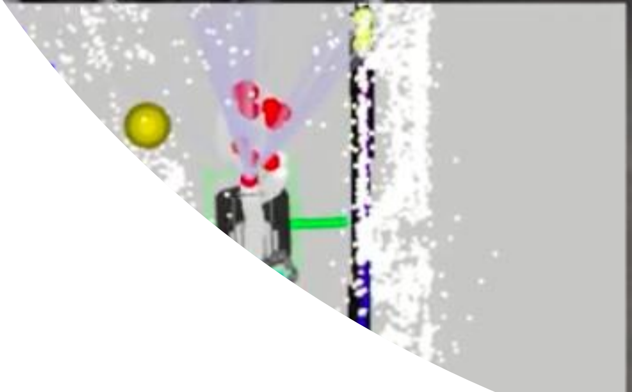
NeuSurfEmb | OnePose++ (w/o tracking, orig. recrop.) | OnePose++ (w/o tracking, prop. recrop.) | Gen6D (with tracking) | Gen6D (w/o tracking)

The low amount of texture generally causes less accurate predictions for OnePose++



What is the future of object representations for robotics?

- How can we form object representations even more efficiently?
- What type of properties should we additionally incorporate?
- Is an explicit database needed or will implicit, large-scale priors be the future?
- …



Jen Jen Chung

Robotic Perception, Planning and Learning Lab, UQ

jenjen.chung@uq.edu.au

Francesco Milano

Autonomous Systems Lab, ETHZ

francesco.milano@mavt.ethz.ch