

# Semantic Perception for Bin Picking

**Sven Behnke**

University of Bonn

Computer Science Institute VI

Autonomous Intelligent Systems



# Bin Picking

- Removing items from containers and shelves
- Still often performed by humans
- Difficulties include
  - Item variability
  - Problematic material properties
  - Articulation of objects
  - Lacking grasp affordances
  - Chaotic storage
  - Inaccessibility



[Amazon]

# Our Bin Picking Robots

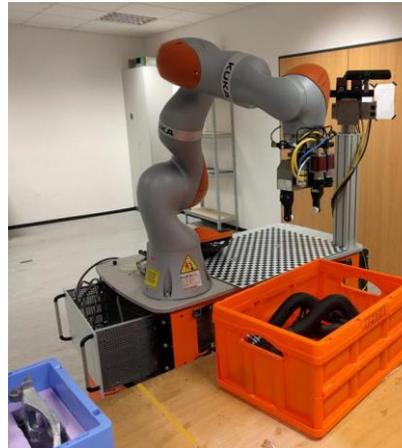
ActReMa



EuRoC C1



STAMINA



EuRoC C2

Amazon Picking

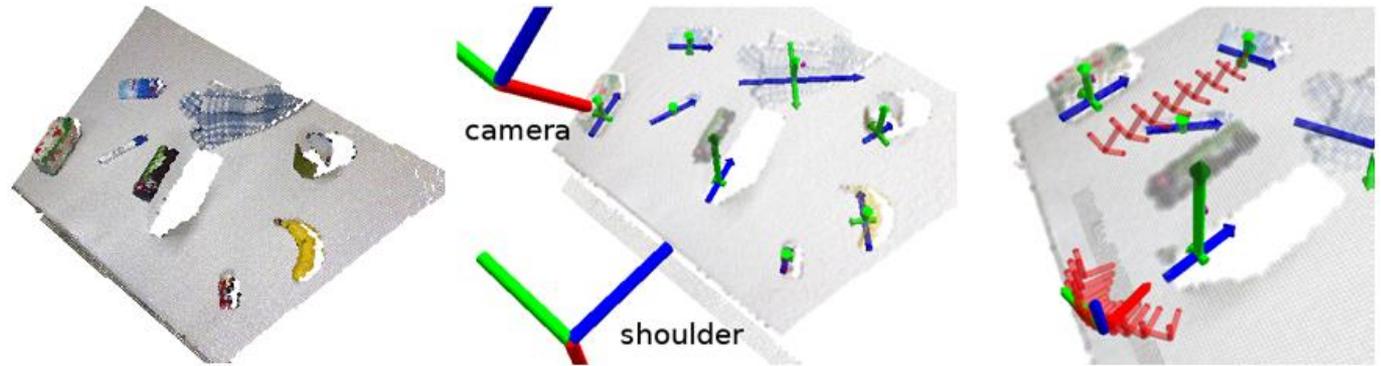


# Cognitive Service Robot Cosero

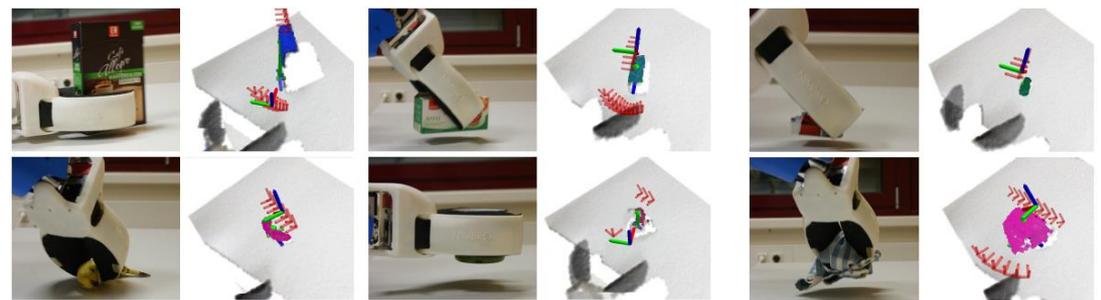


# Table-top Analysis and Grasp Planning

- Detection of clusters above horizontal plane
- Two grasps (top, side)



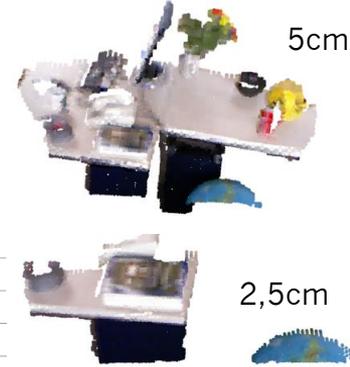
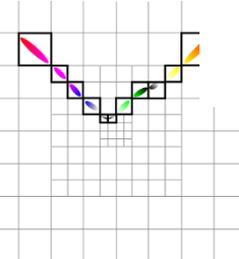
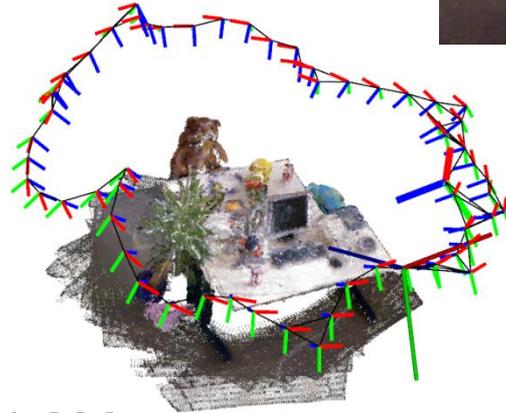
- Flexible grasping of many unknown objects



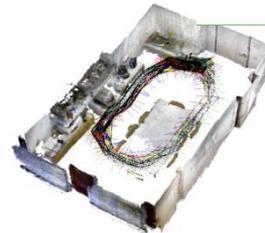
[Stückler et al, Robotics and Autonomous Systems, 2013]

# 3D Mapping by RGB-D SLAM

- Modelling of shape and color distributions in voxels
- Local multiresolution
- Efficient registration of views on CPU
- Global optimization

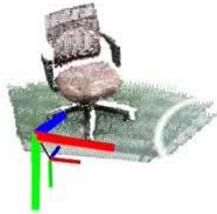


- Multi-camera SLAM

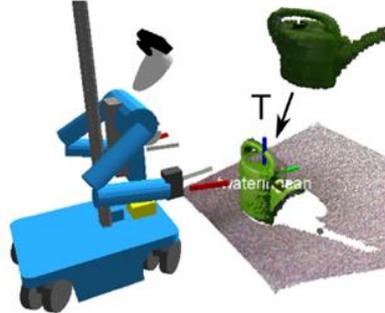


# Learning and Tracking Object Models

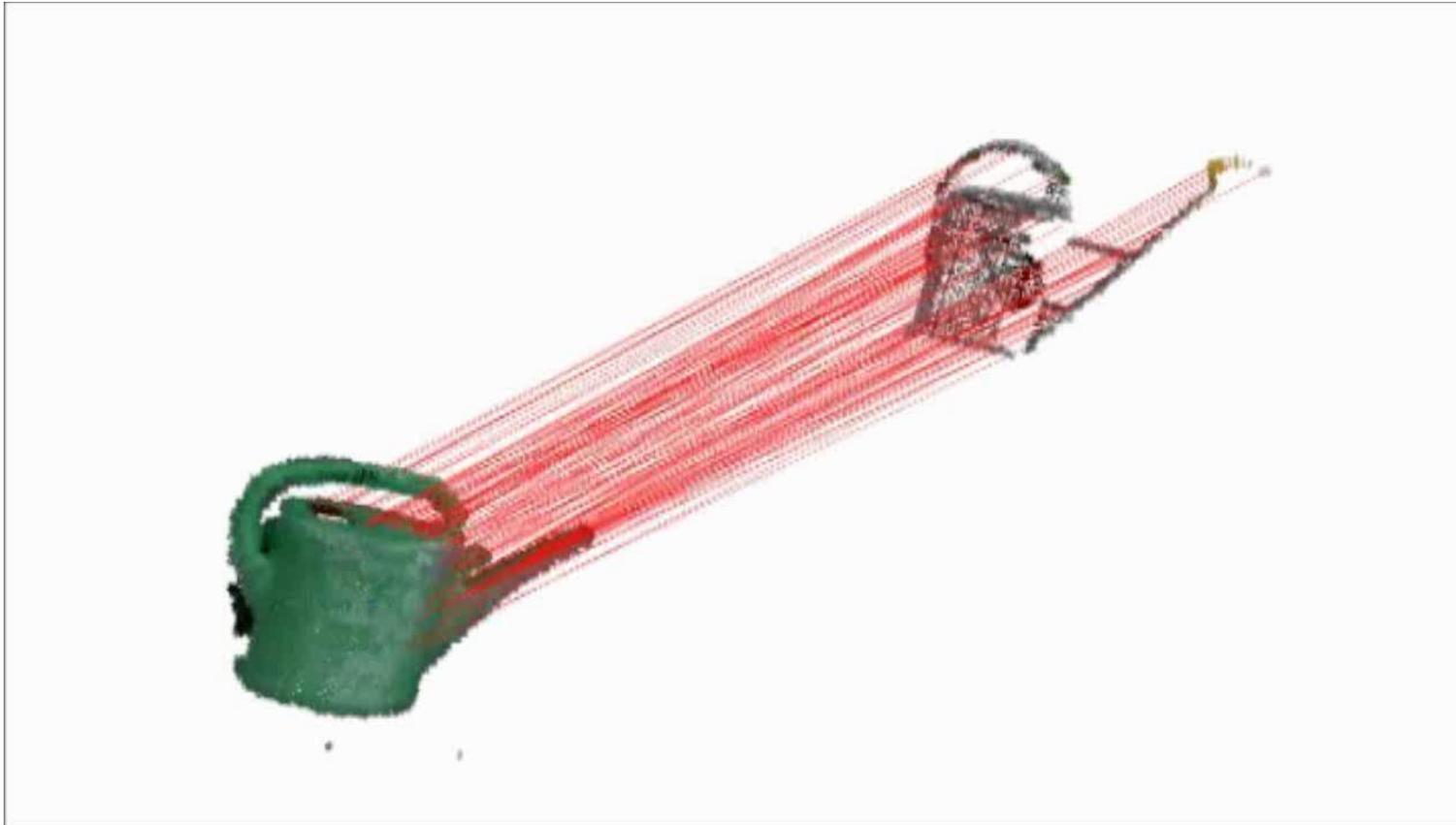
- Modeling of objects by RGB-D-SLAM



- Real-time registration with current RGB-D frame



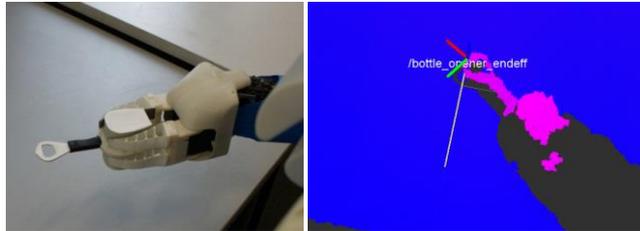
# Grasp & Motion Skill Transfer



[Stückler,  
Behnke,  
ICRA2014]

# Tool use: Bottle Opener

- Tool tip perception



- Extension of arm kinematics
- Perception of crown cap
- Motion adaptation



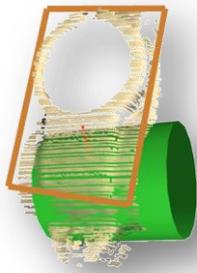
[Stückler, Behnke, Humanoids 2014]

# Shape-Primitive based Part Perception

- Known objects in transport box



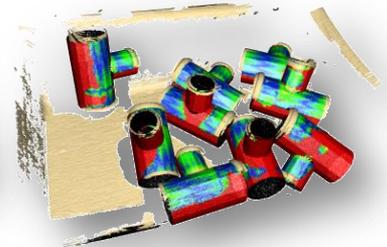
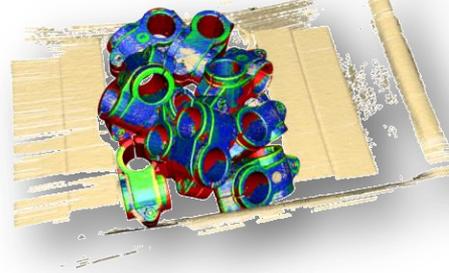
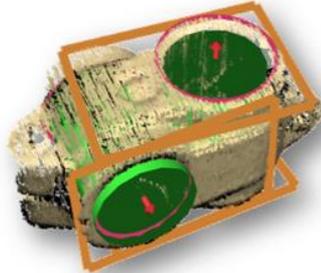
- Matching of graphs of 2D and 3D shape primitives



3D



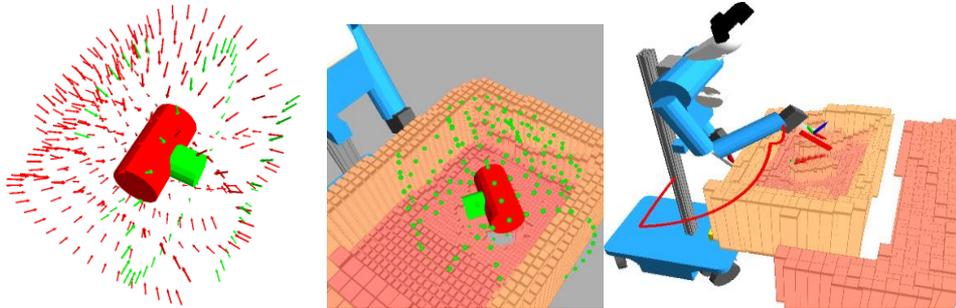
2D



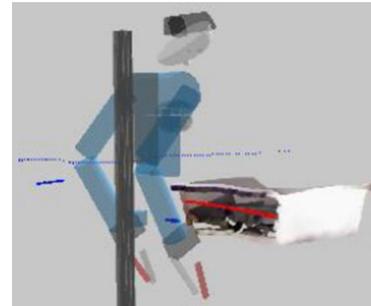
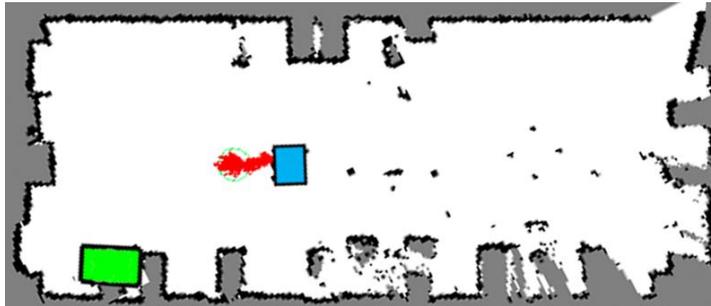
[Berner et al. ICIP 2013]

# Mobile Bin Picking

- Grasp and motion planning



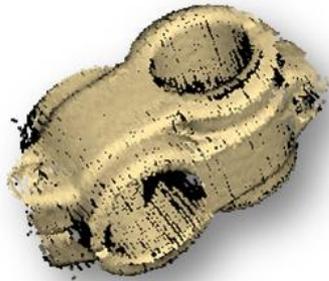
- Allocentric and relative navigation



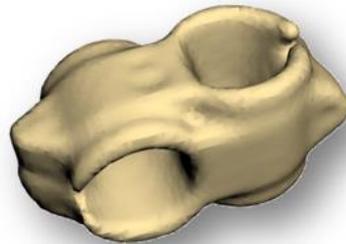
[Nieuwenhuisen et al.  
ICRA 2013]

# Learning of Object Models

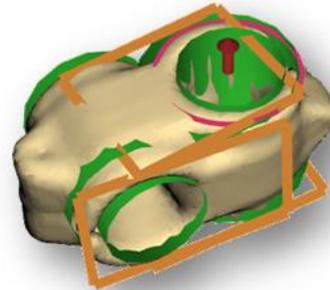
- Scan multiple objects
- Remove support plane
- Segment views
- Register views using ICP
- Recognize geometric primitives



Registered views



Surface reconstruction

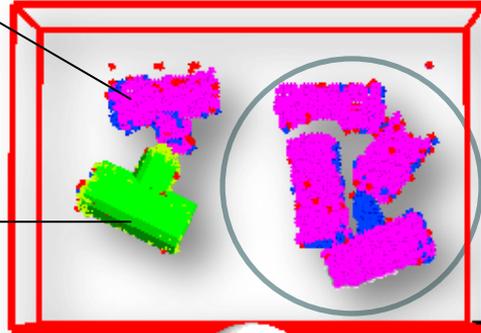


Detected primitives

# Active Object Perception

Detected cylinders

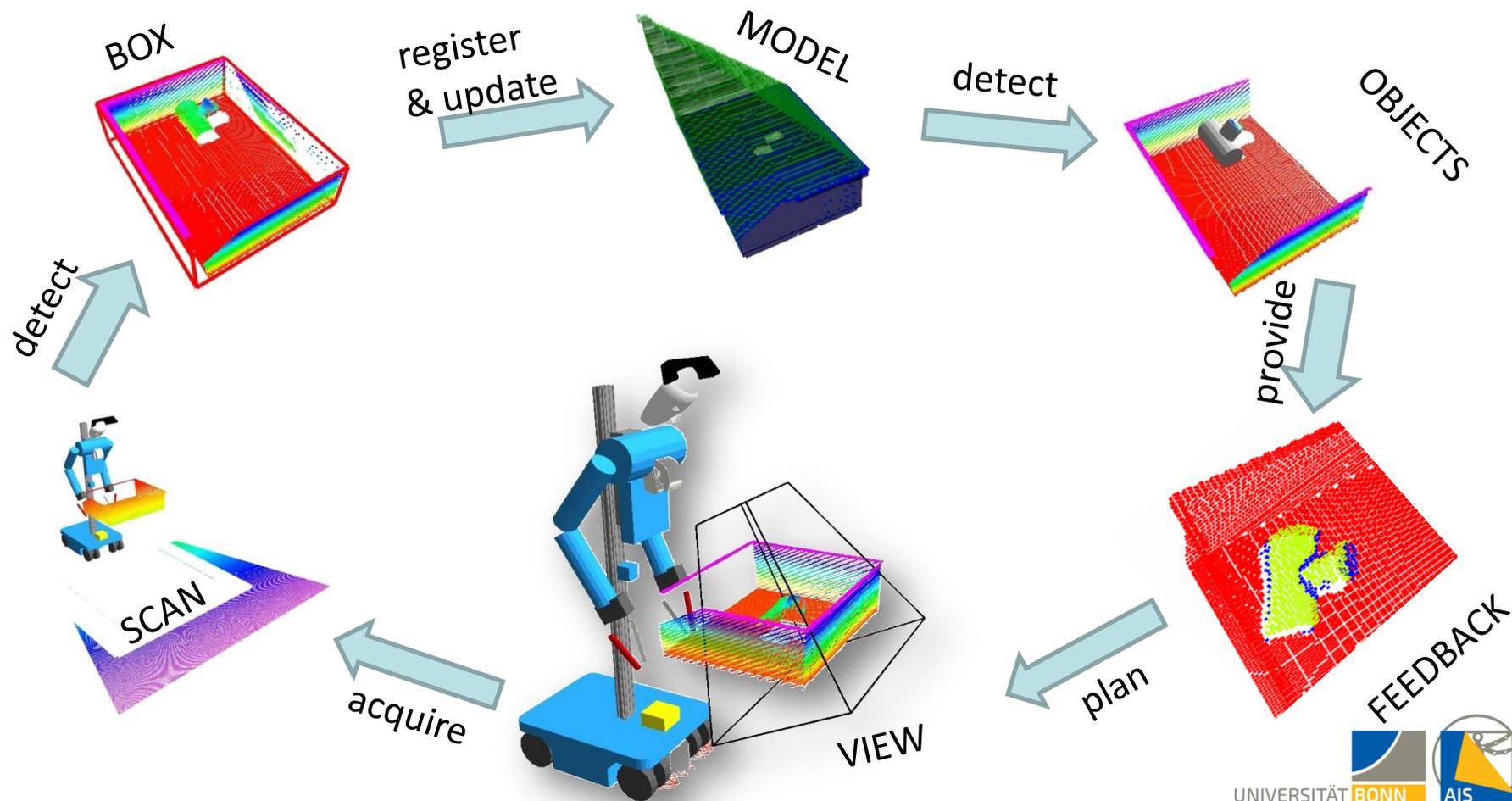
Detected object



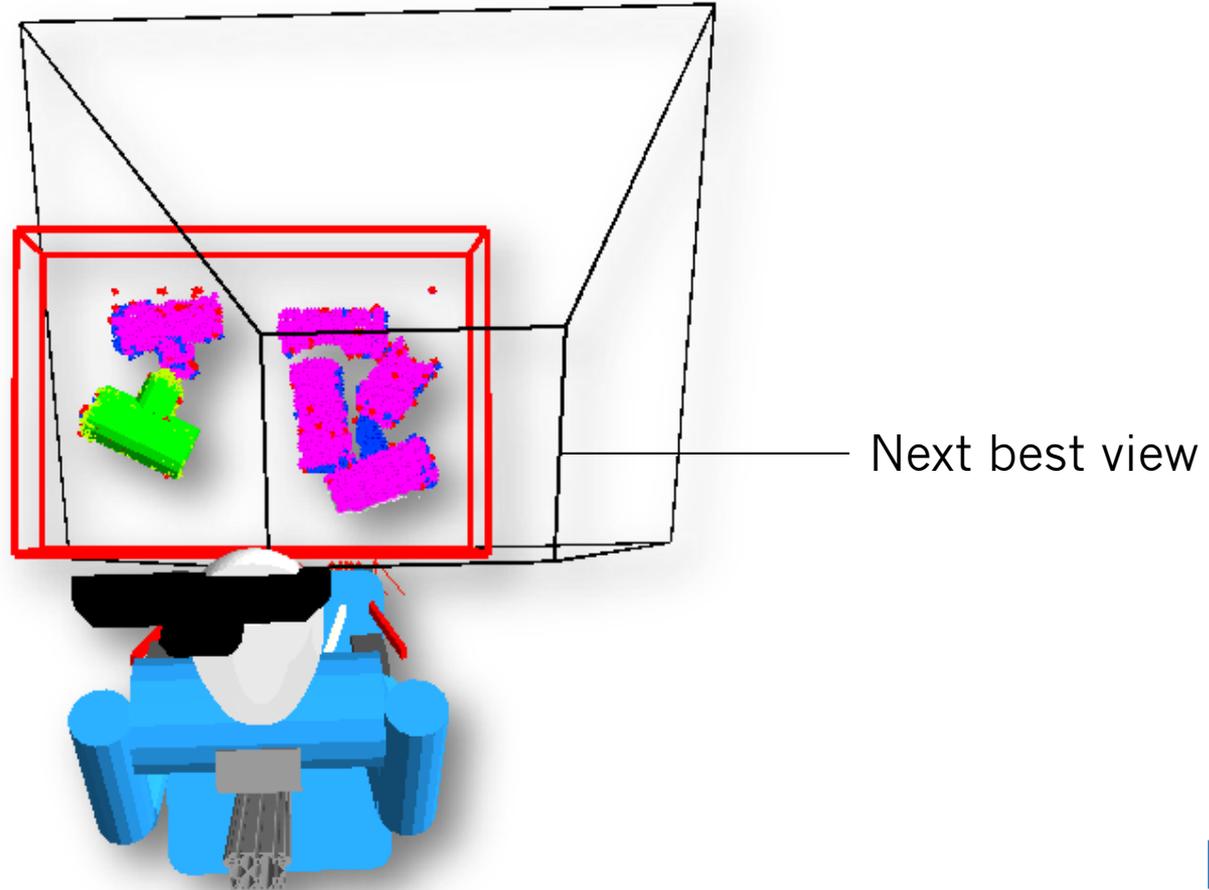
Partial occlusions



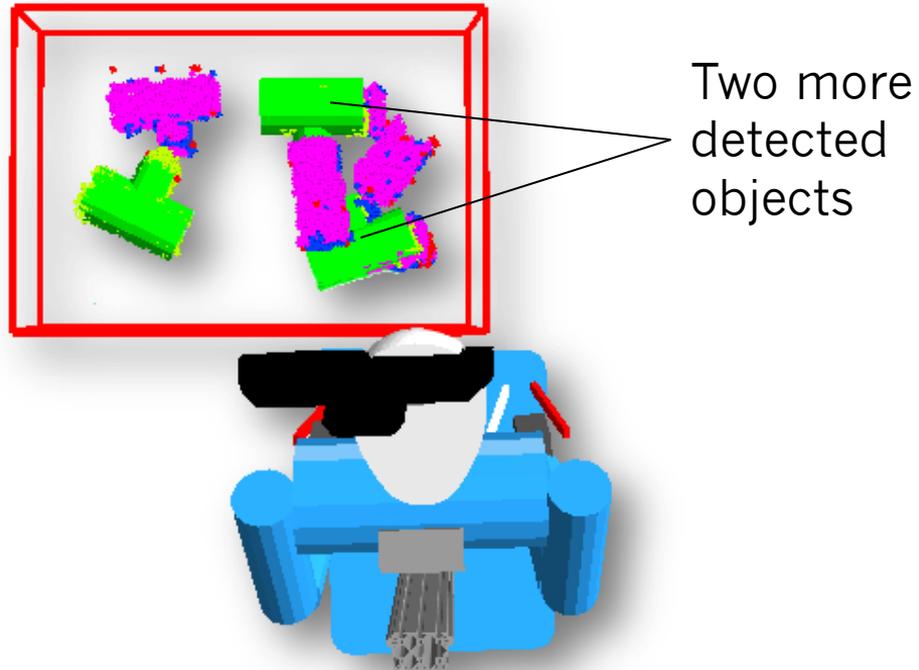
# Active Object Perception



# Active Object Perception

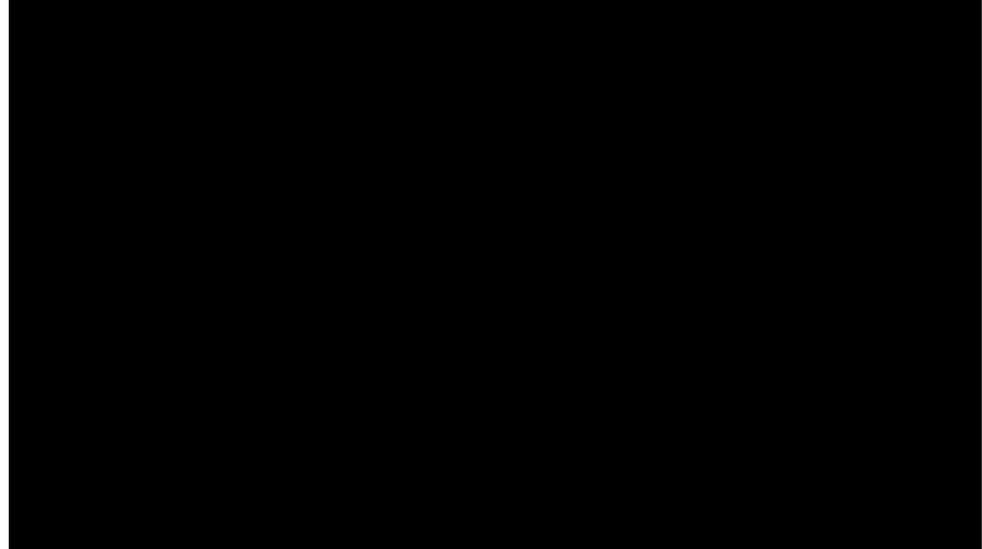


# Active Object Perception

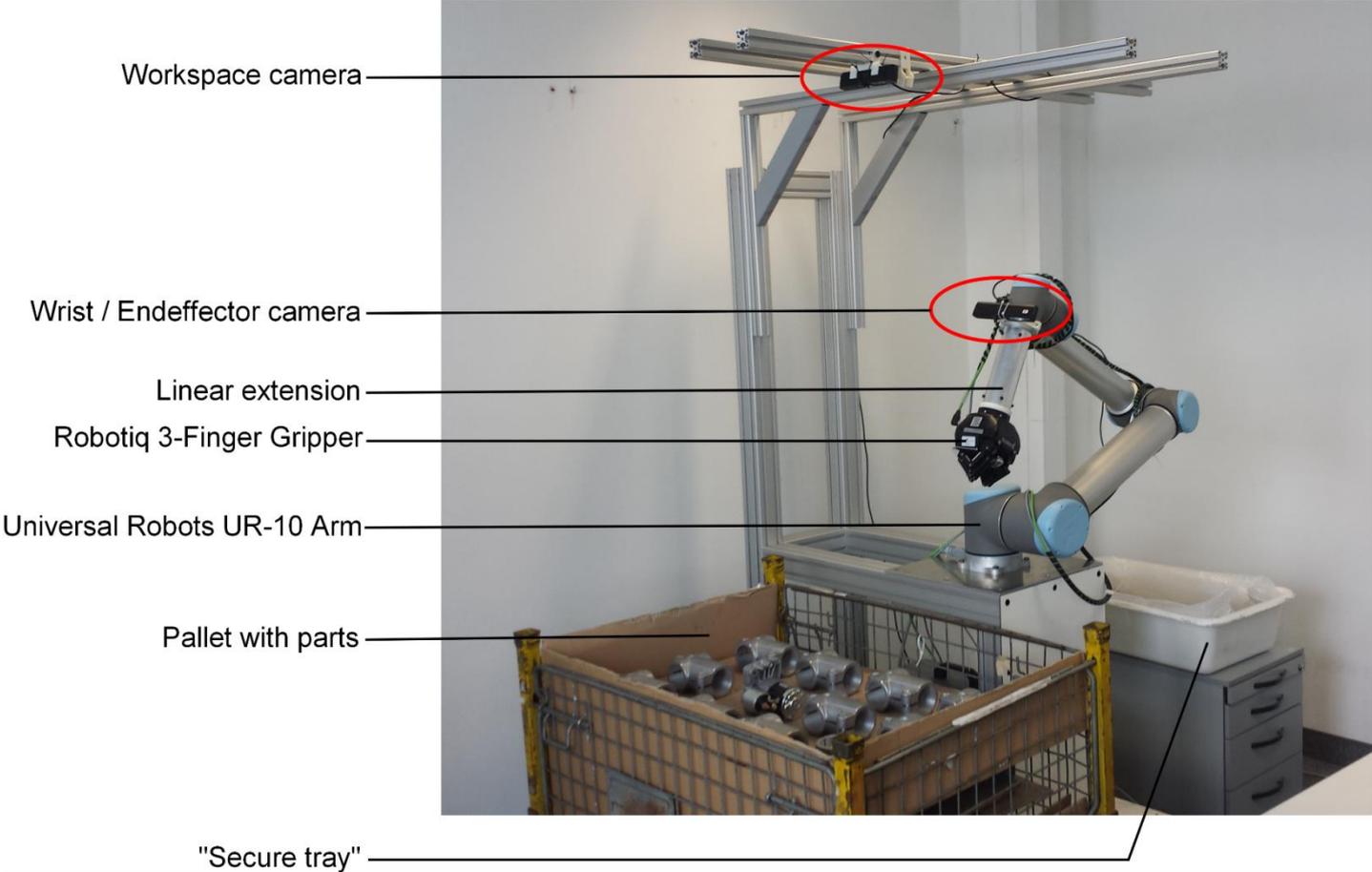


# Kitting of Automotive Parts

- Many car variants
- Collect the parts needed for the assembly of a particular car in a kit
- Parts in available in a supermarket
- Robot needs to
  - navigate to the transport boxes,
  - grasp the parts, and
  - place them in the kit

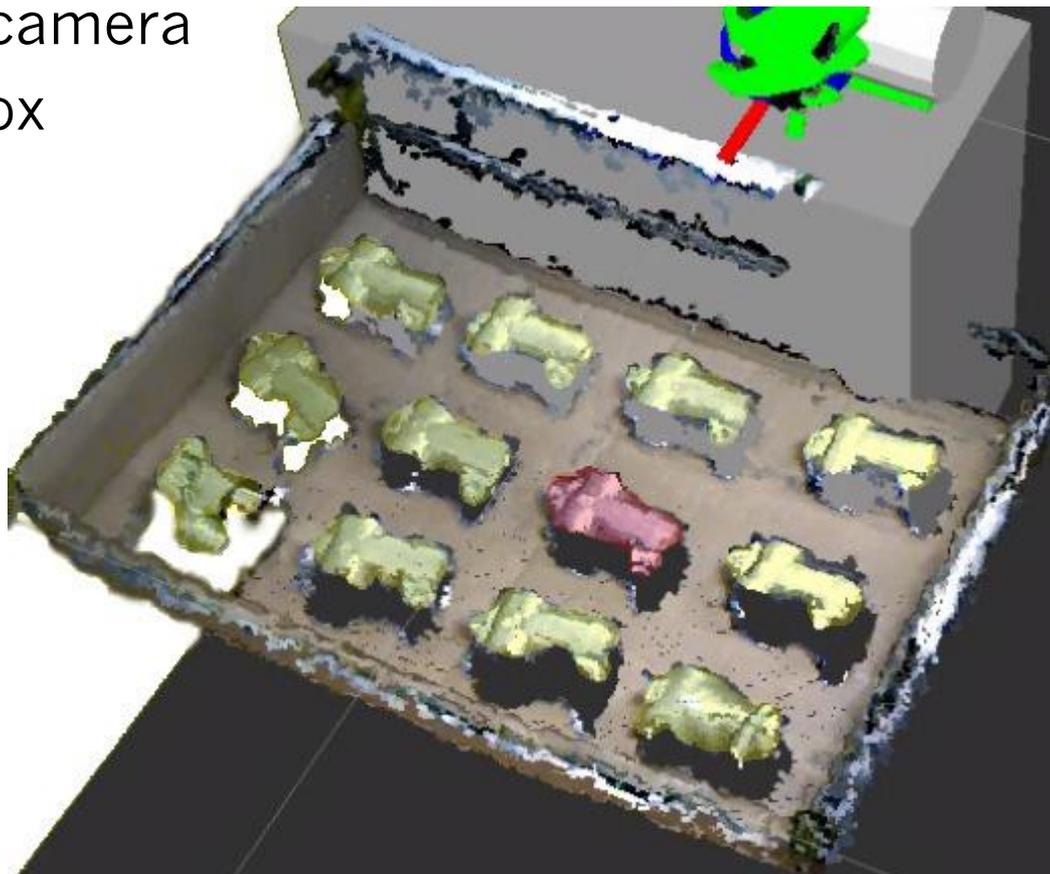


# Lab Demonstrator



# Part Detection

- Using work space RGB-D camera
- Initial pose of transport box roughly known
- Detect dominant horizontal plane above ground
- Cluster points above support plane
- Estimate main axes

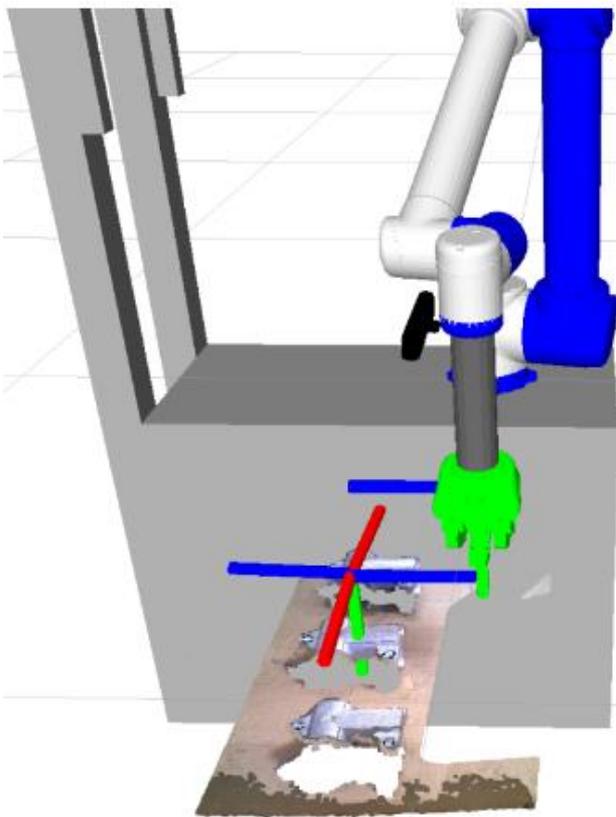


# Part Pose Estimation

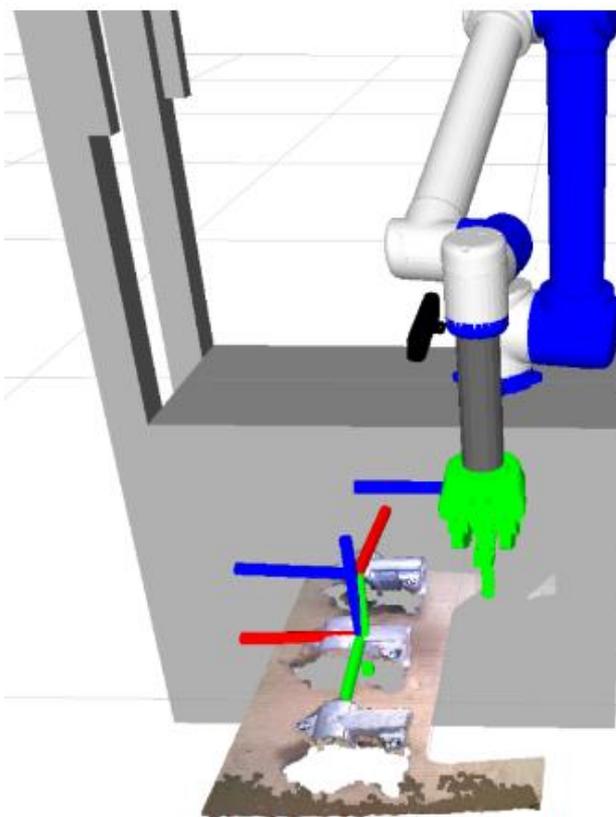
- Wrist RGB-D camera moved above innermost object candidate
- Object views are represented as Multiresolution Surfel Map
- Registration of object view with current measurements using soft assignments
- Verification based on registration quality



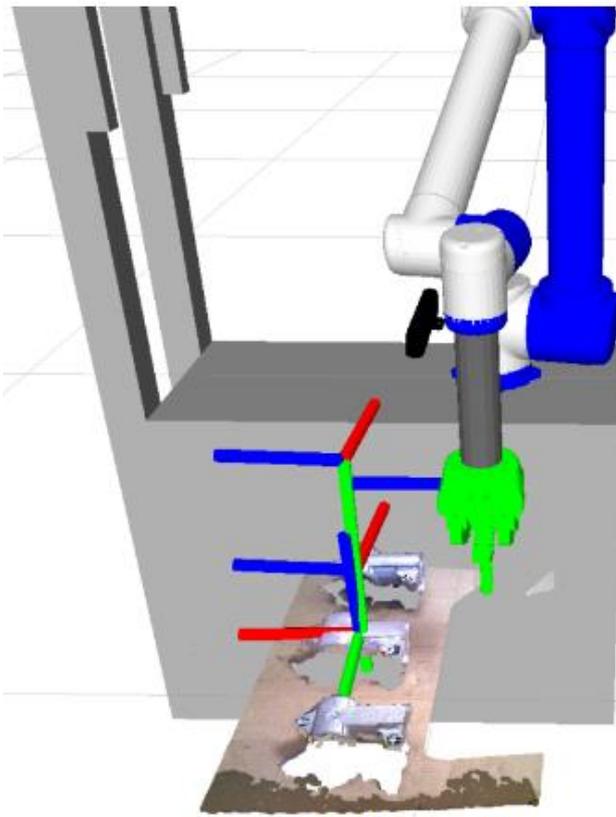
# Grasp Selection



Grasp Candidates



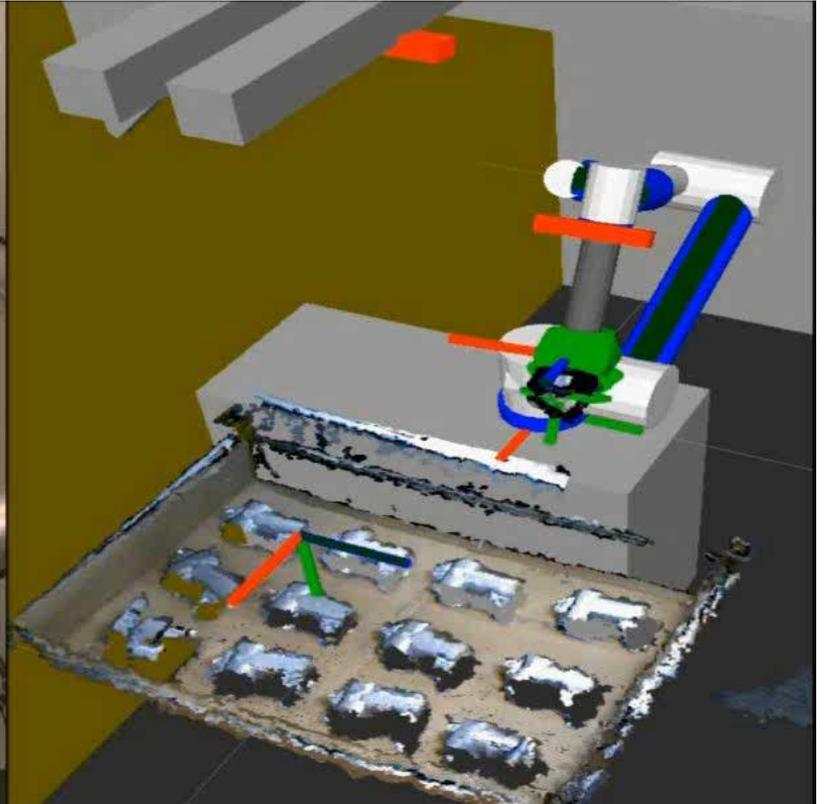
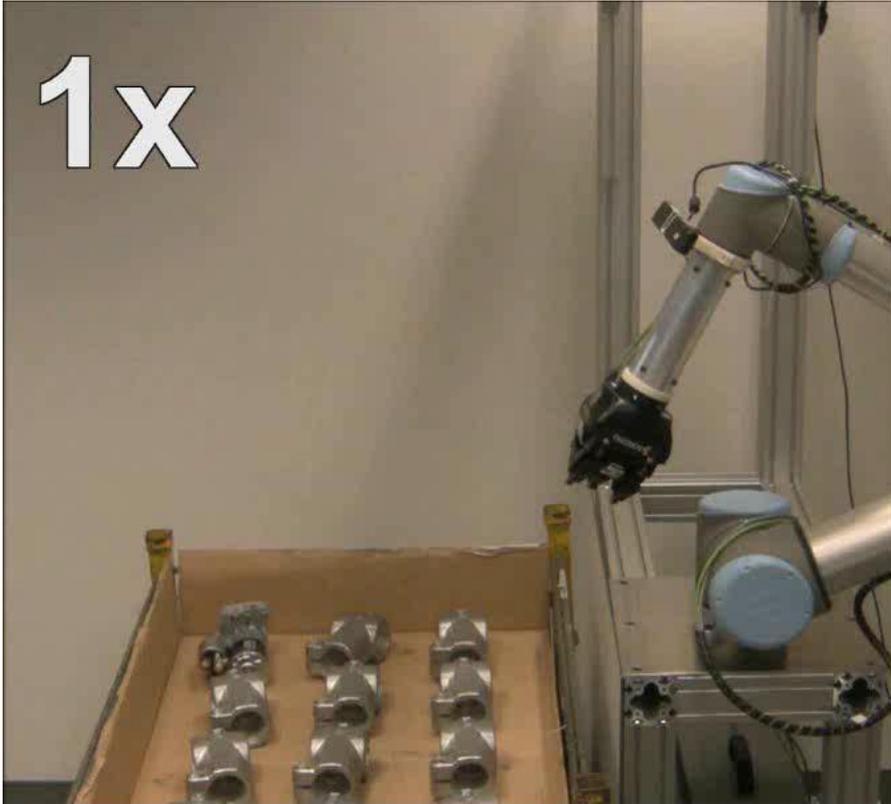
Selected Grasp and  
Object Pose



Pre-Grasp, Grasp  
and Object Pose

# Depalettizing of Parts

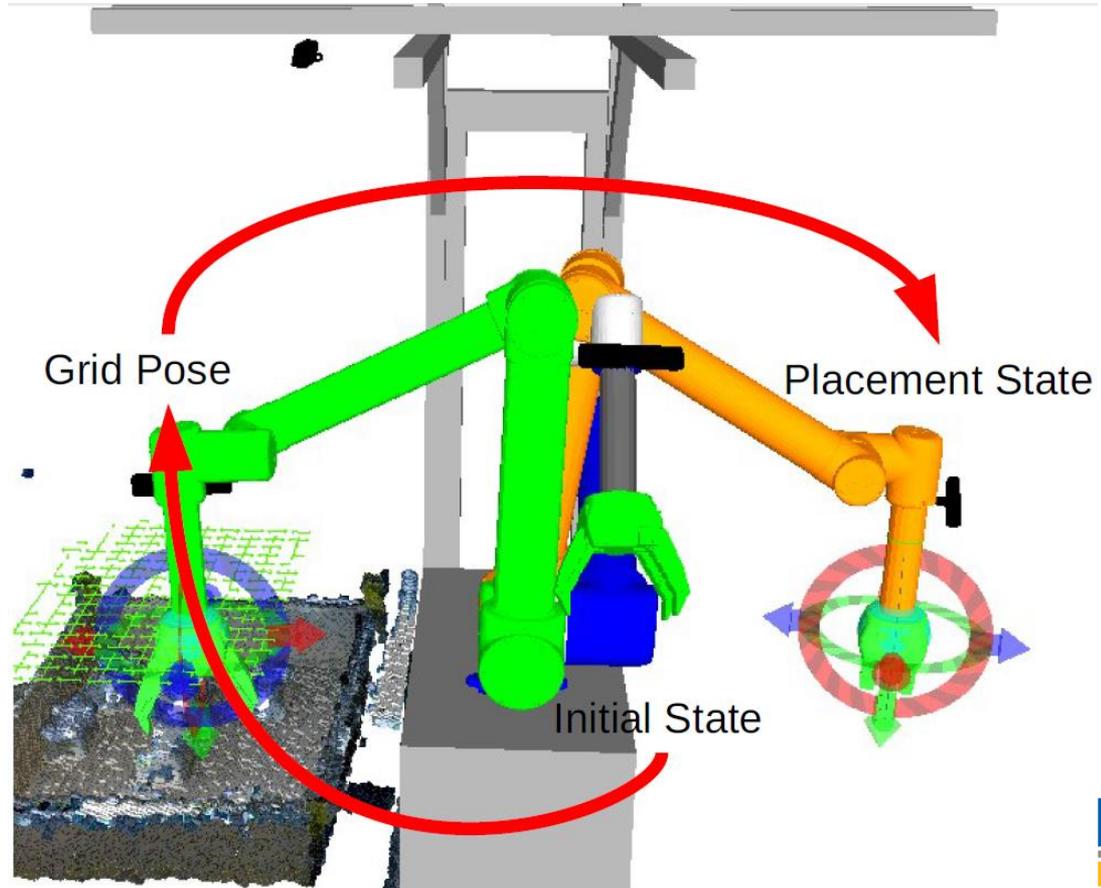
1x



We detect potential object candidates using the workspace camera.

# Motion Planning

- Predefined poses
- Preplanned paths
- Only short trajectories must be planned online



# Mobile Manipulation Robot

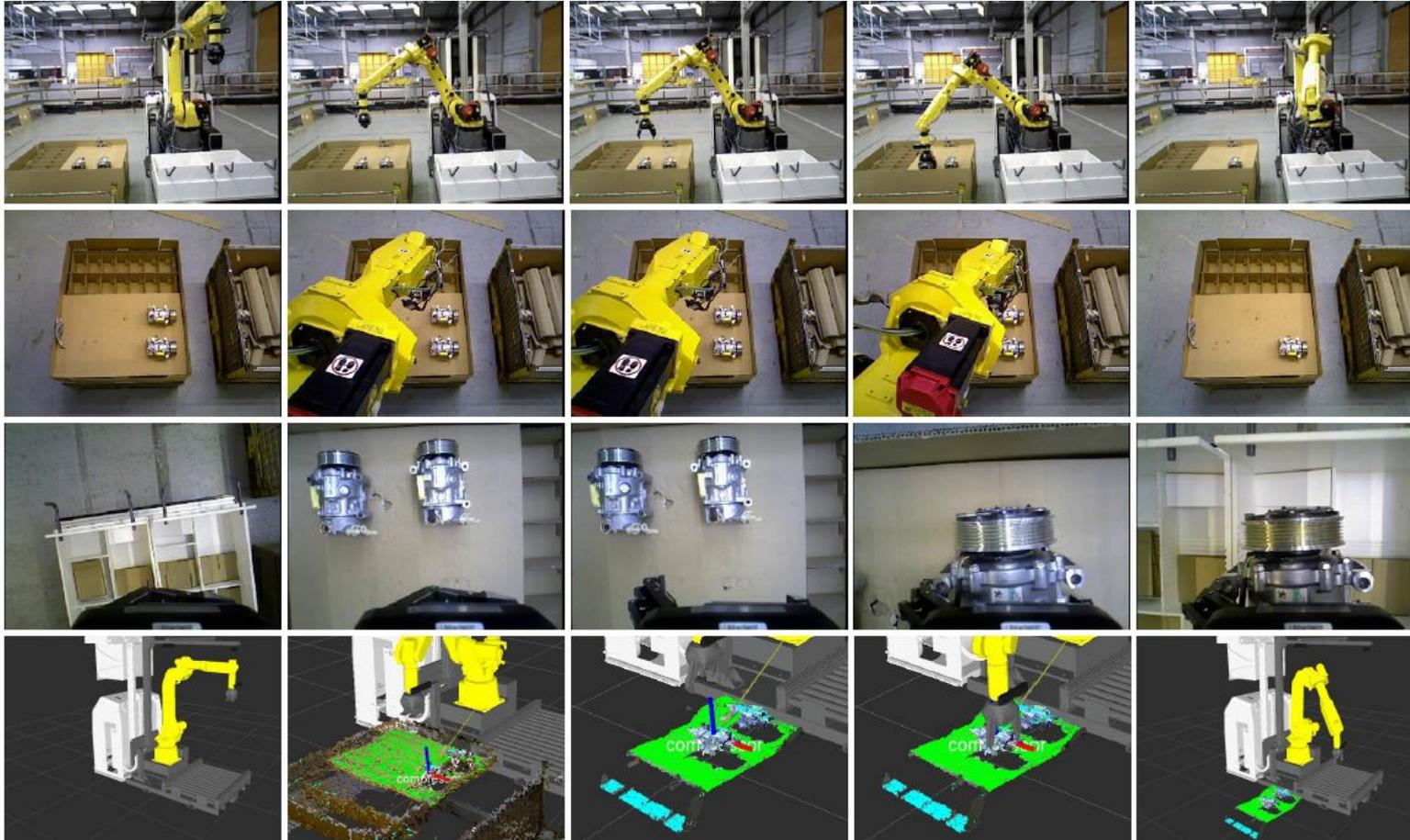
Initialization

Part detection

Approach

Grasping

Placing

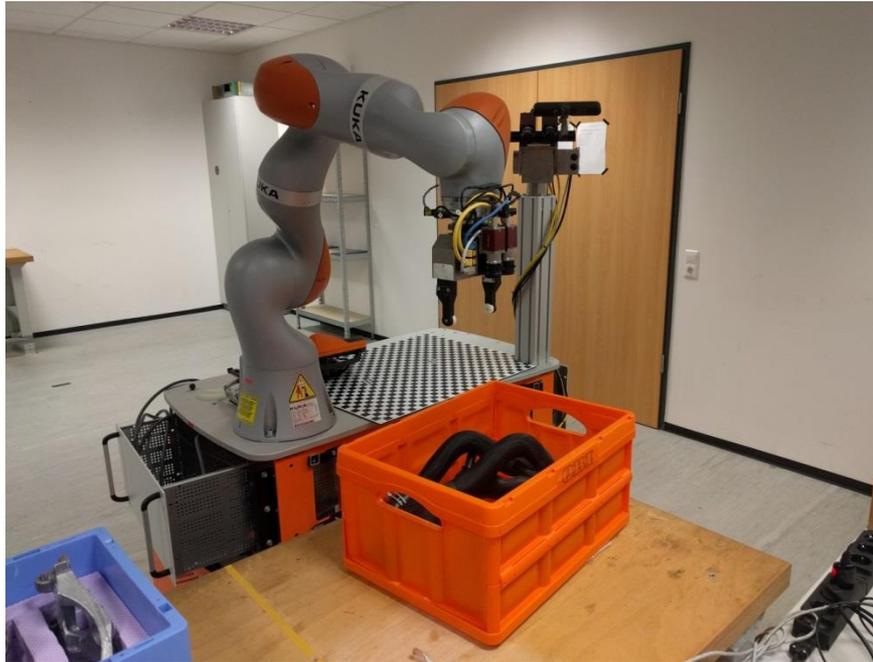


# Final Demonstration



# EuRoC Challenge 2: KittingBot

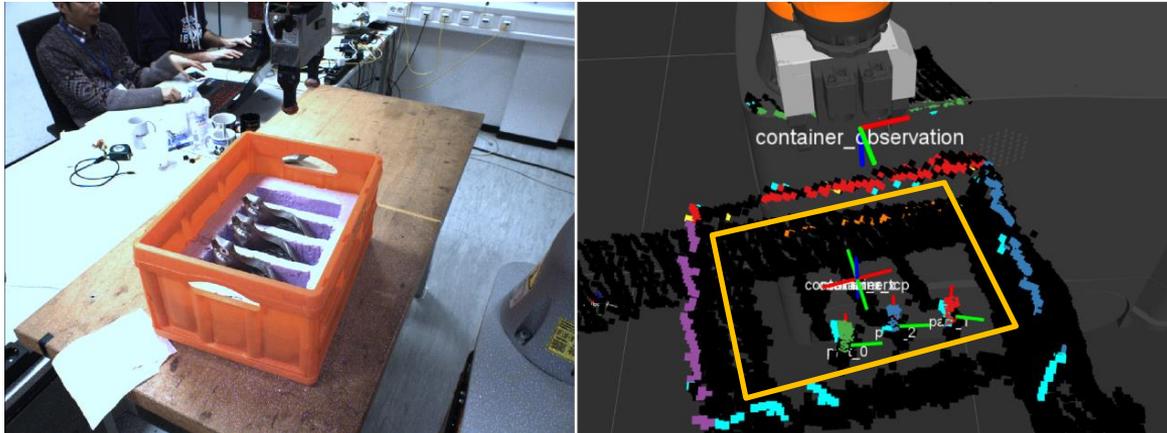
- Robot platform: Kuka miiwa
- Collaborative kitting



[Pavlichenko et al. IAS 2018]

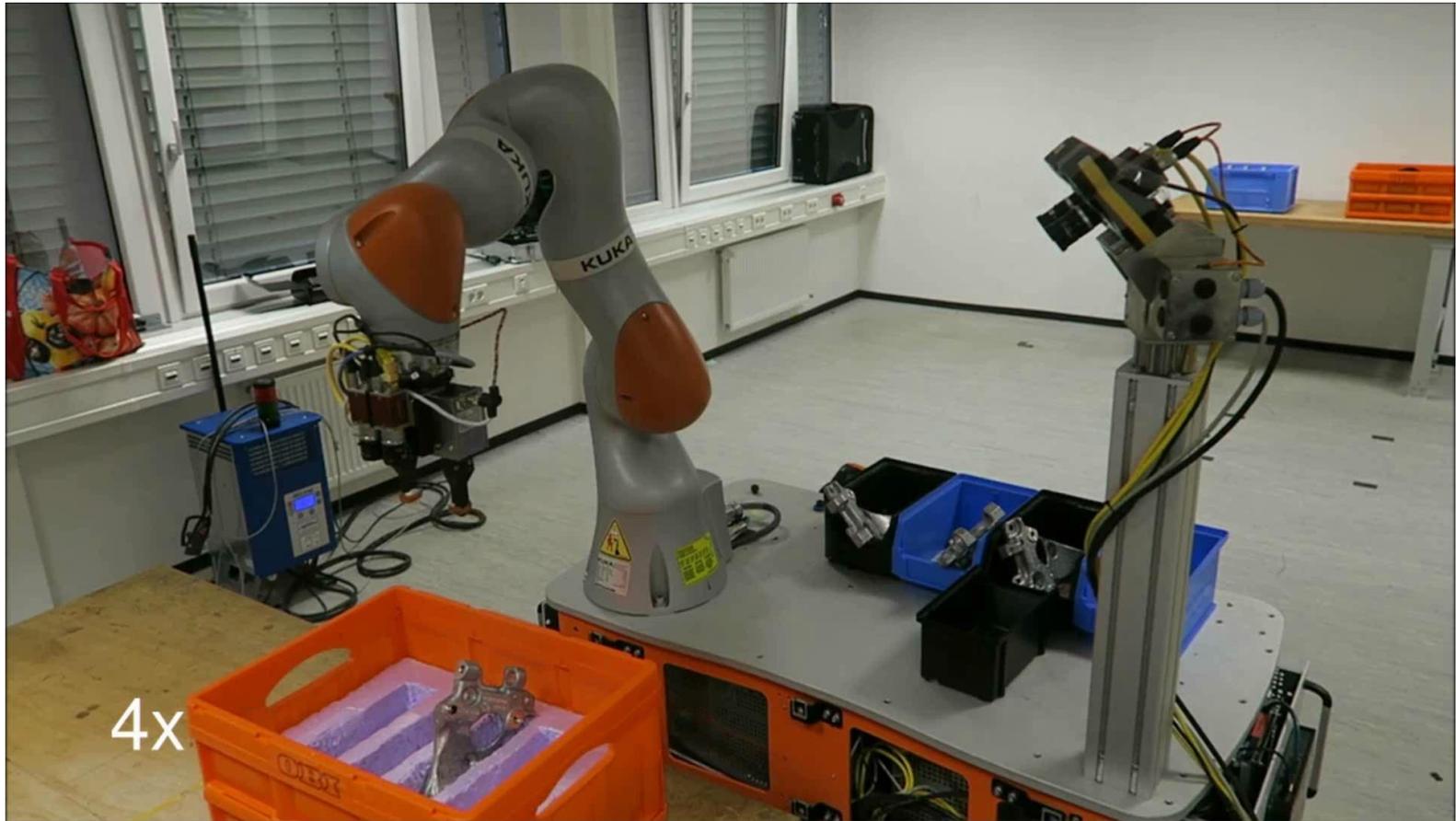
# Box Detection and Pose Estimation

- Detection of edges in depth image
- Grouping to lines
- Pose estimation for upper rim



[Holz und Behnke: ISR 2016]

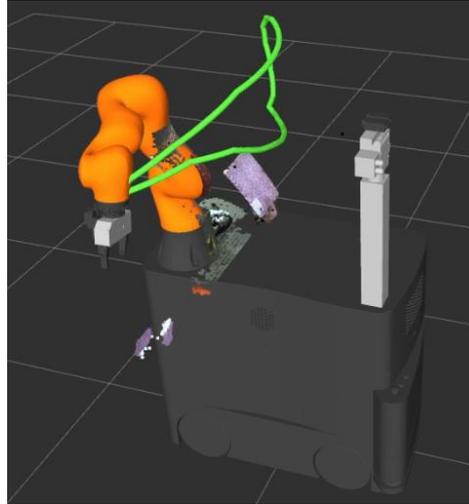
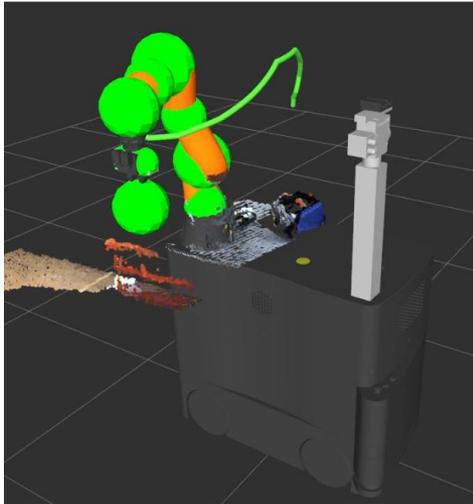
# KittingBot Showcase Demonstration



[Pavlichenko et al. IAS 2018]

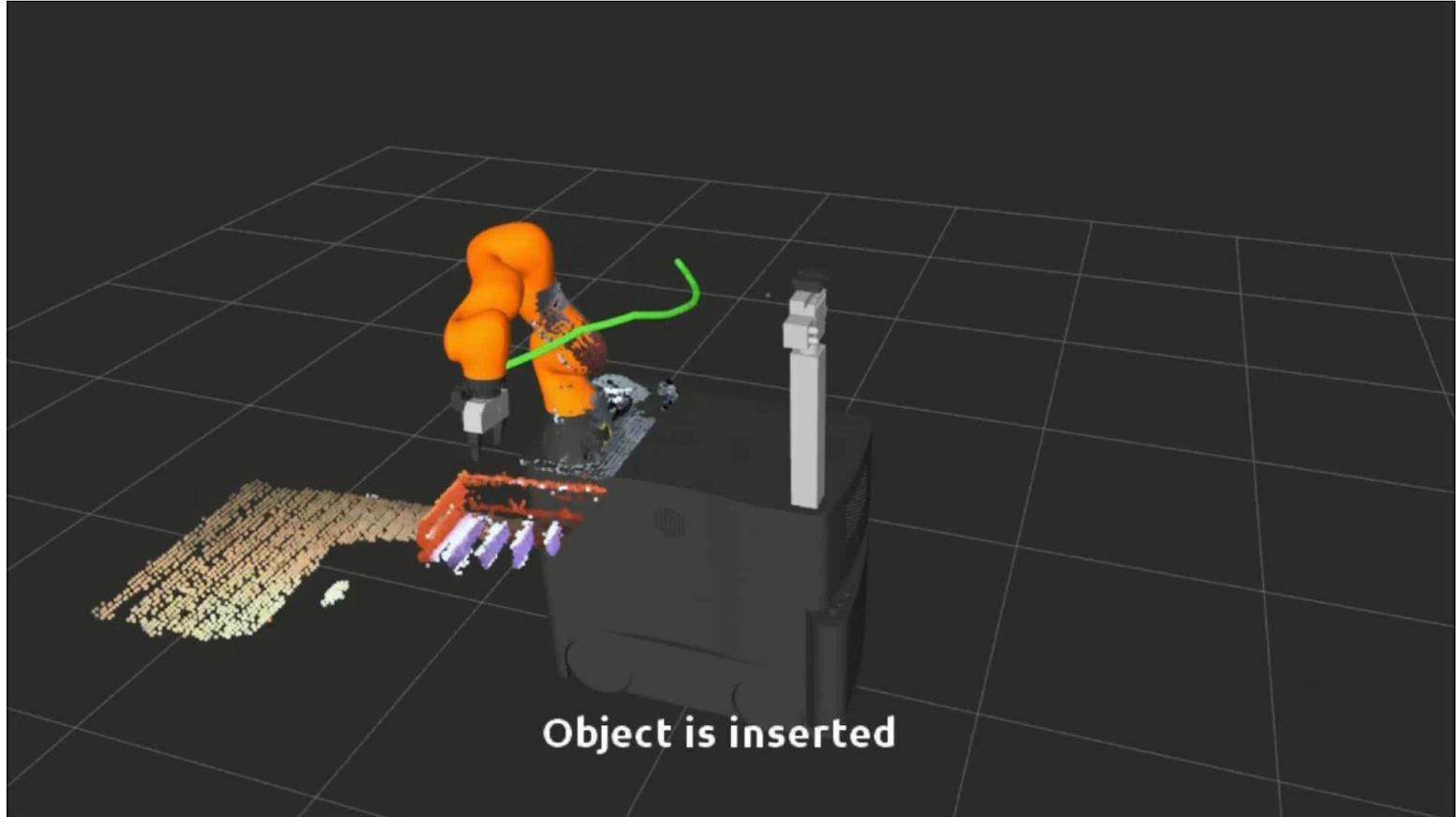
# Online Trajectory Optimization

- Joint optimization of multiple objectives
  - Obstacle avoidance
  - Speed
  - Torques



[Pavlichenko and Behnke: IROS 2017]

# KittingBot: Obstacle Avoidance



[Pavlichenko et al. IAS 2018]

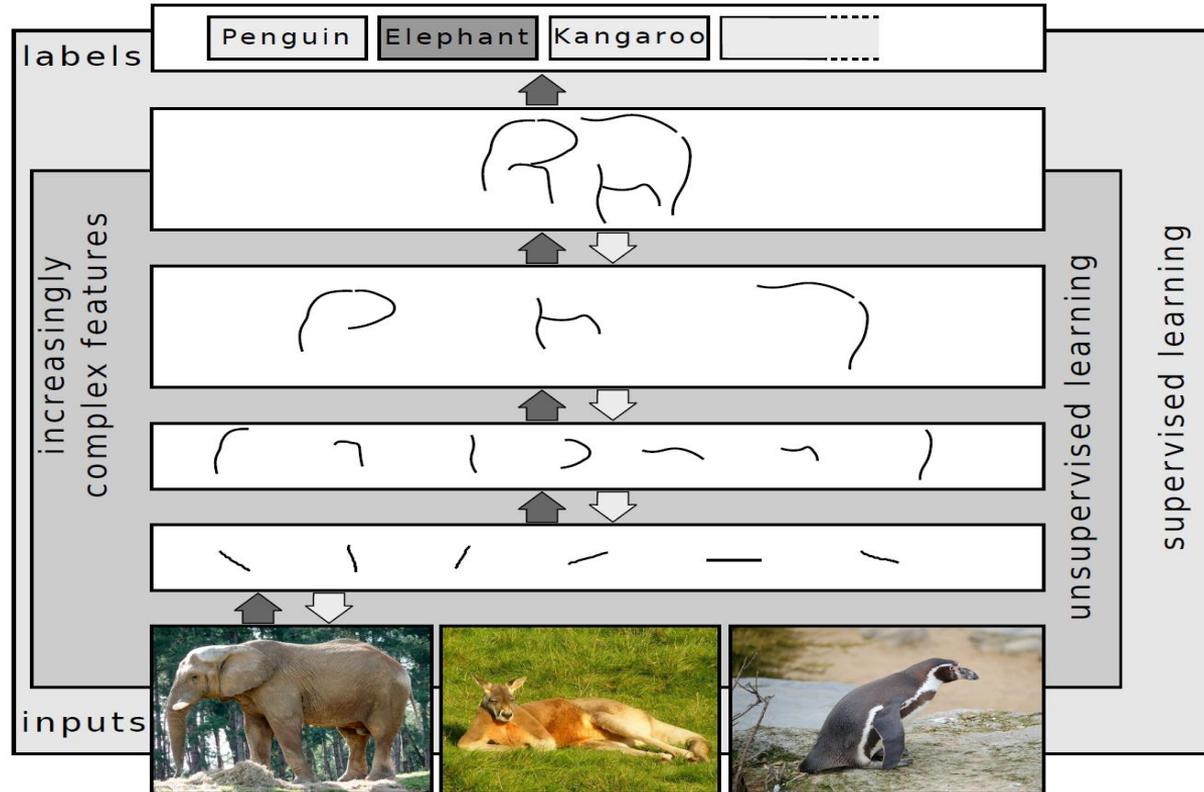
# KittingBot Showcase Demonstration



[Pavlichenko et al. IAS 2018]

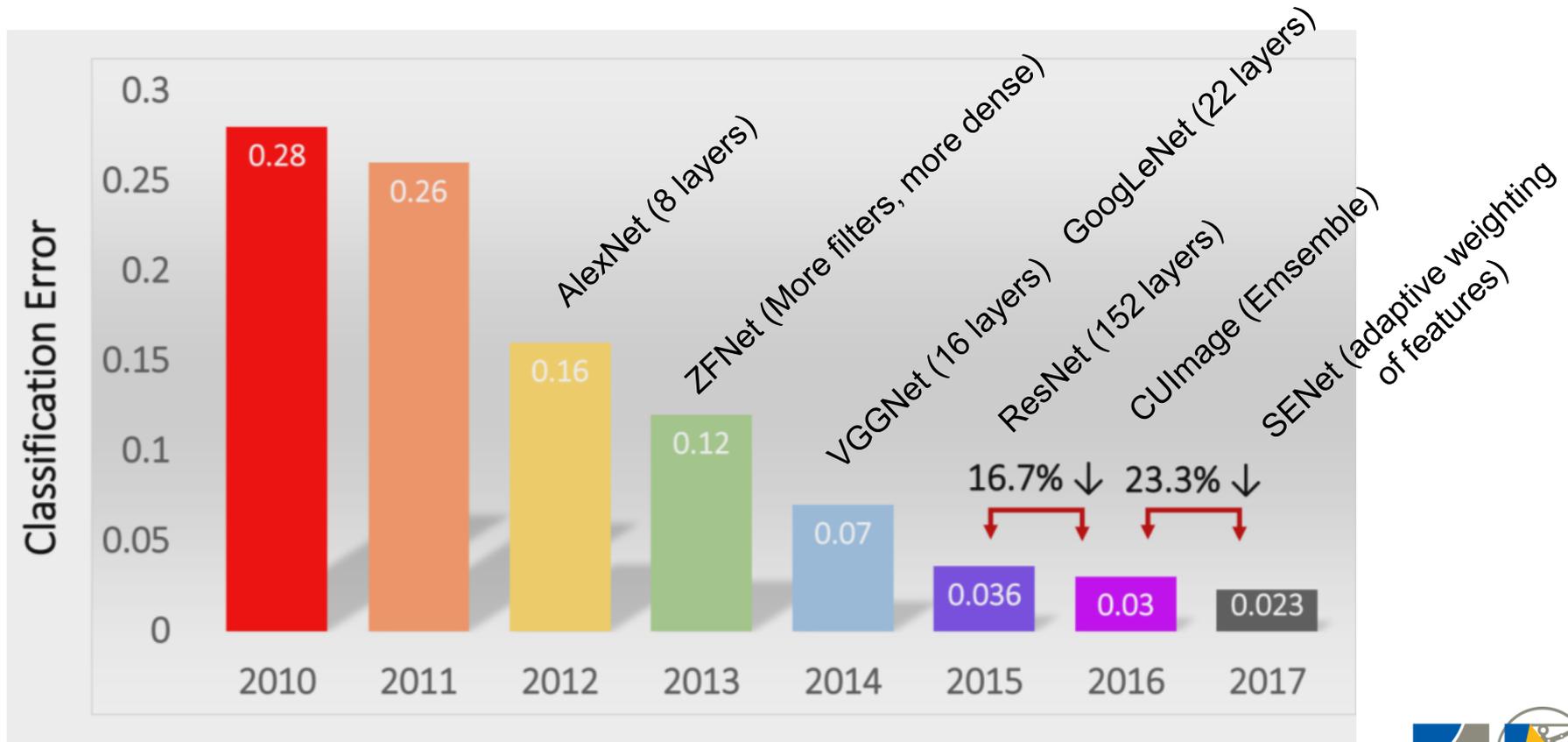
# Deep Learning

- Learning layered representations



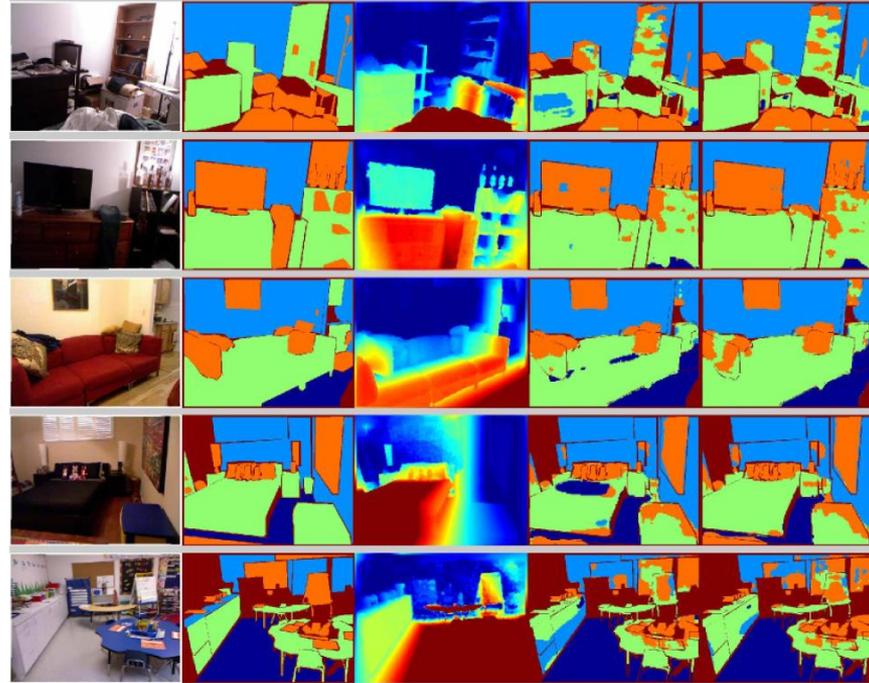
[Schulz;  
Behnke,  
KI 2012]

# ImageNet-Challenge (ILSVRC)



# Geometric and Semantic Features for RGB-D Object-class Segmentation

- New **geometric** feature: distance from wall
- **Semantic** features pretrained from ImageNet
- Both help significantly



[Husain et al. RA-L 2017]

RGB

Truth

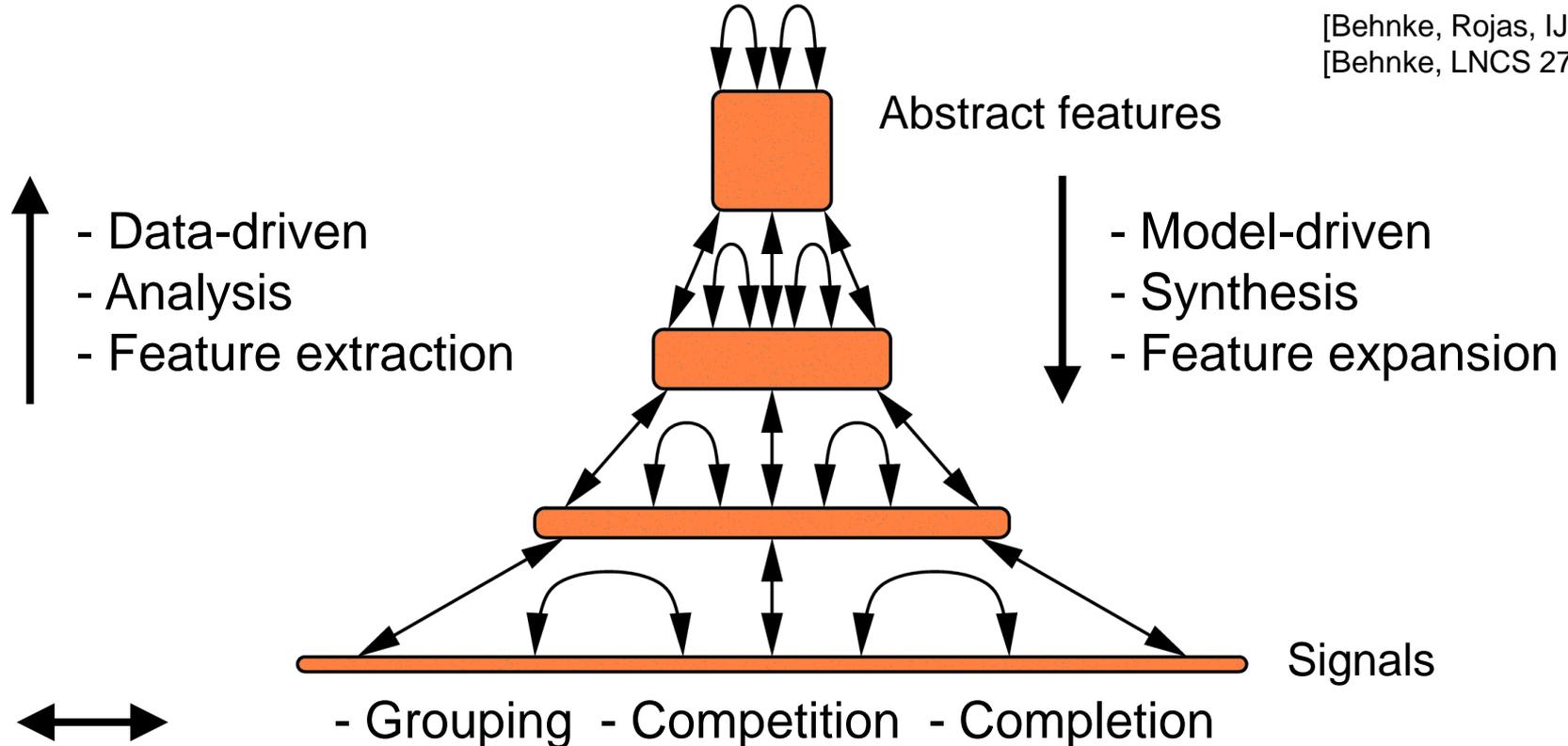
DistWall

OutWO

OutWithDistWall

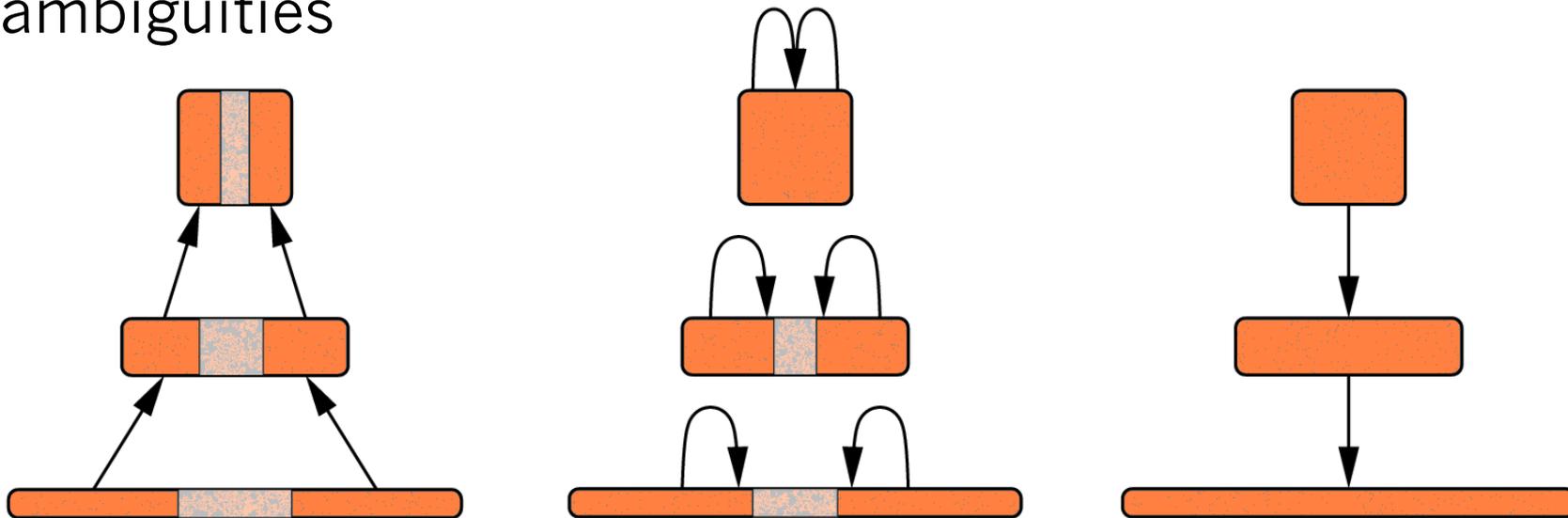
# Neural Abstraction Pyramid

[Behnke, Rojas, IJCNN 1998]  
[Behnke, LNCS 2766, 2003]



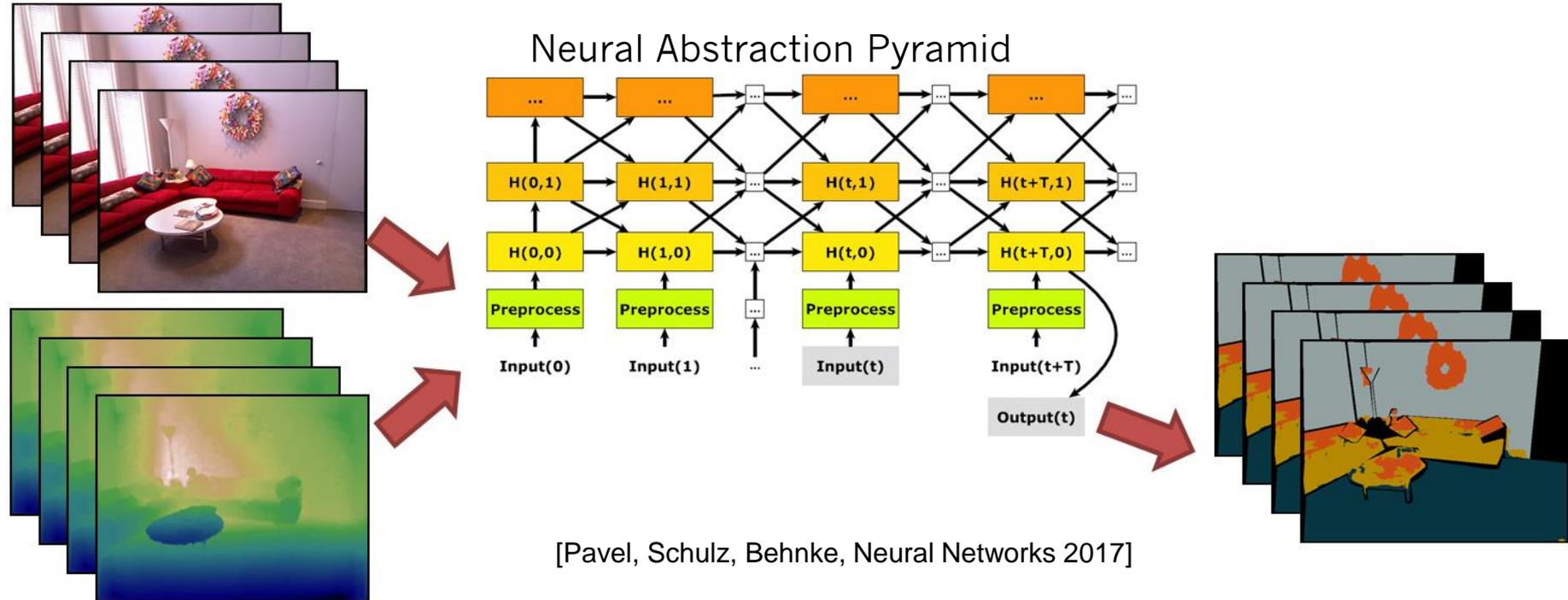
# Iterative Image Interpretation

- Interpret most obvious parts first
- Use partial interpretation as context to resolve local ambiguities



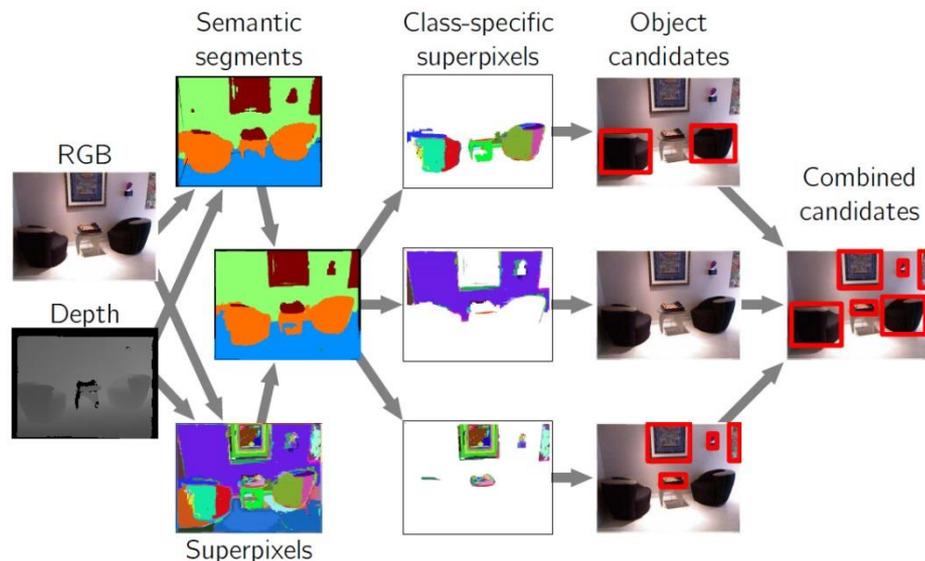
# Neural Abstraction Pyramid for Semantic Segmentation of RGB-D Video

- Recursive computation is efficient for temporal integration



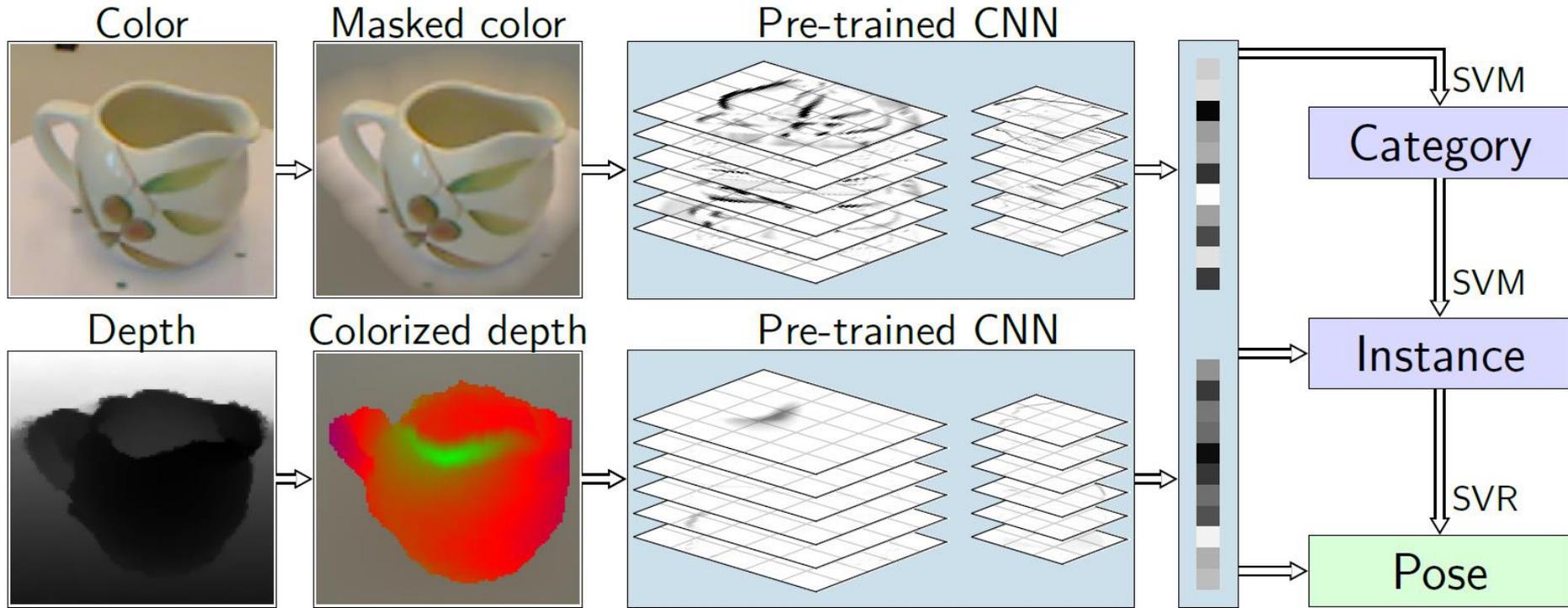
# Semantic Segmentation Priors for Object Discovery

- Combine bottom-up object discovery and semantic priors
- Semantic segmentation used to classify color and depth superpixels
- Higher recall, more precise object borders



[Garcia et al. ICPR 2016]

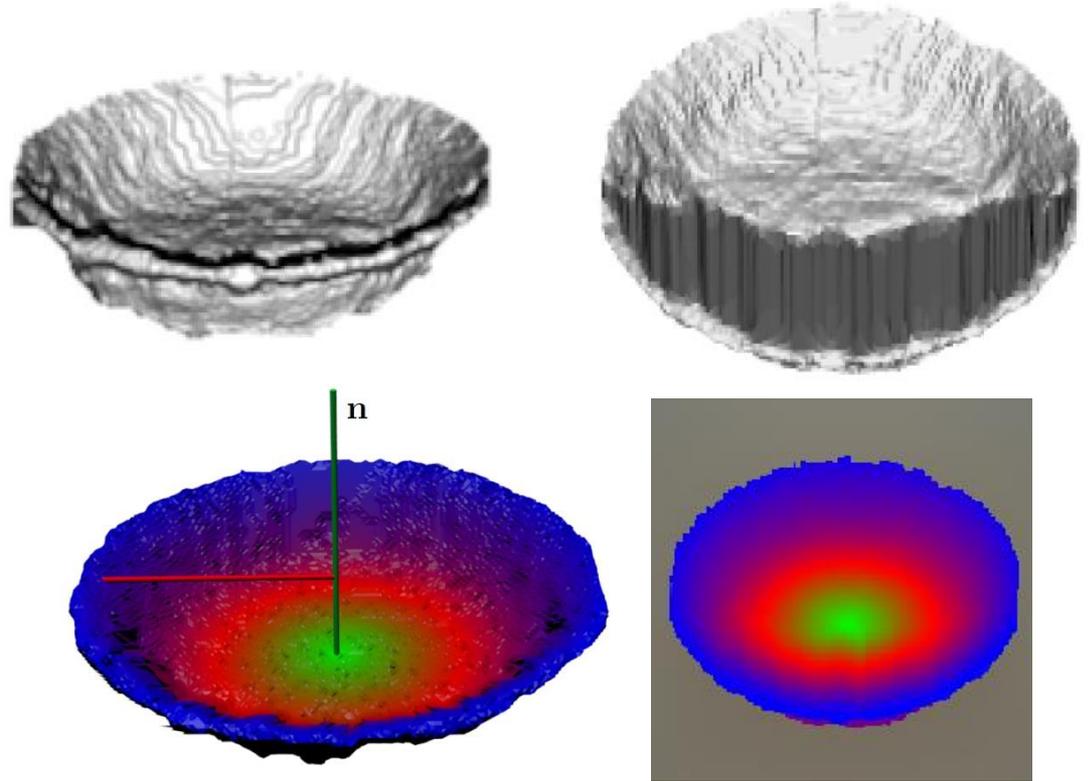
# RGB-D Object Recognition and Pose Estimation



[Schwarz, Schulz, Behnke, ICRA2015]

# Canonical View, Colorization

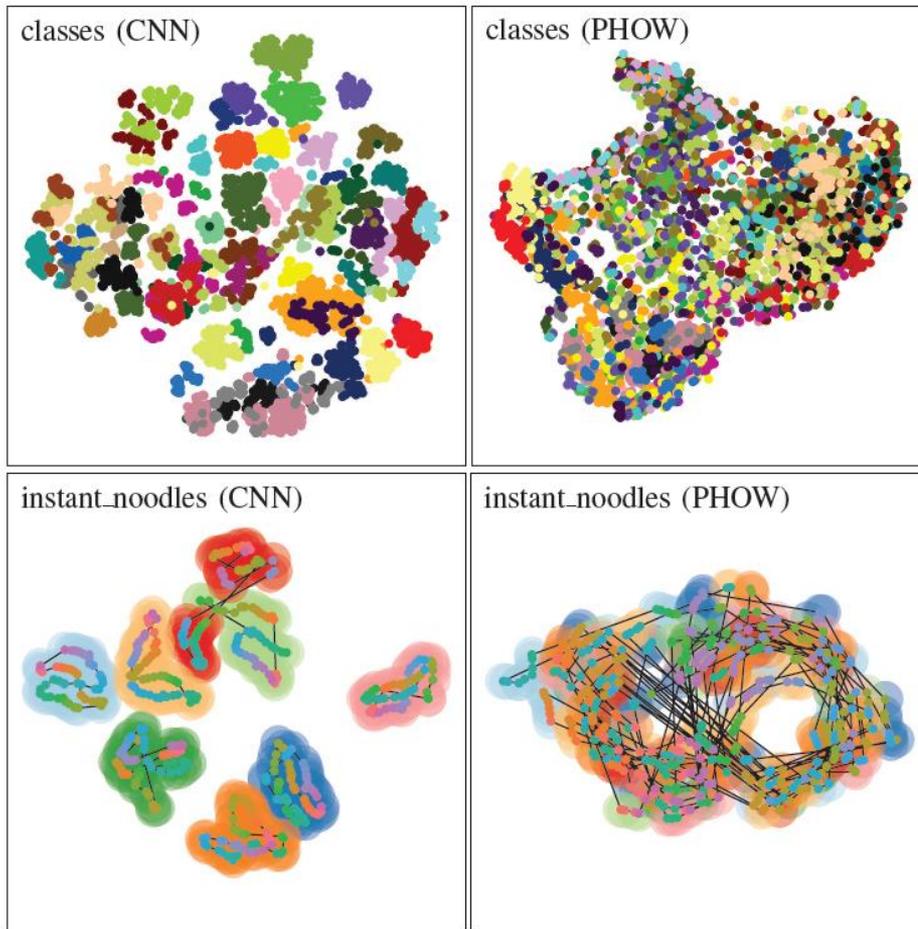
- Objects viewed from different elevation
- Render canonical view
- Colorization based on distance from center vertical



[Schwarz, Schulz, Behnke, ICRA2015]

# Pretrained Features Disentangle Data

- t-SNE embedding



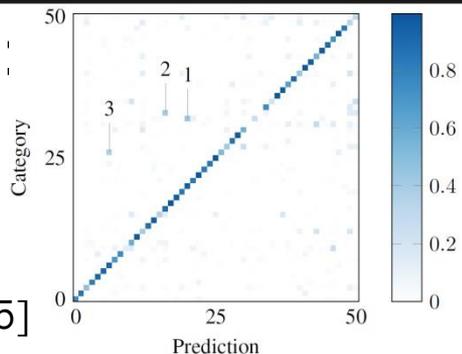
[Schwarz, Schulz,  
Behnke ICRA2015]

# Recognition Accuracy

- Improved both category and instance recognition

Method	Category Accuracy (%)		Instance Accuracy (%)	
	RGB	RGB-D	RGB	RGB-D
Lai <i>et al.</i> [1]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo <i>et al.</i> [2]	82.4 ± 3.1	87.5 ± 2.9	<b>92.1</b>	92.8
PHOW[3]	80.2 ± 1.8	—	62.8	—
<b>Ours</b>	<b>83.1 ± 2.0</b>	88.3 ± 1.5	92.0	<b>94.1</b>
<b>Ours</b>	<b>83.1 ± 2.0</b>	<b>89.4 ± 1.3</b>	92.0	<b>94.1</b>

- Confusion:



[Schwarz, Schulz, Behnke, ICRA2015]

1: pitcher / coffe mug



2: peach / sponge



# Amazon Picking Challenge

- Large variety of objects
- Unordered in shelf or tote
- Picking and stowing tasks



[Schwarz et al. ICRA 2017]

# Deep Learning Semantic Segmentation

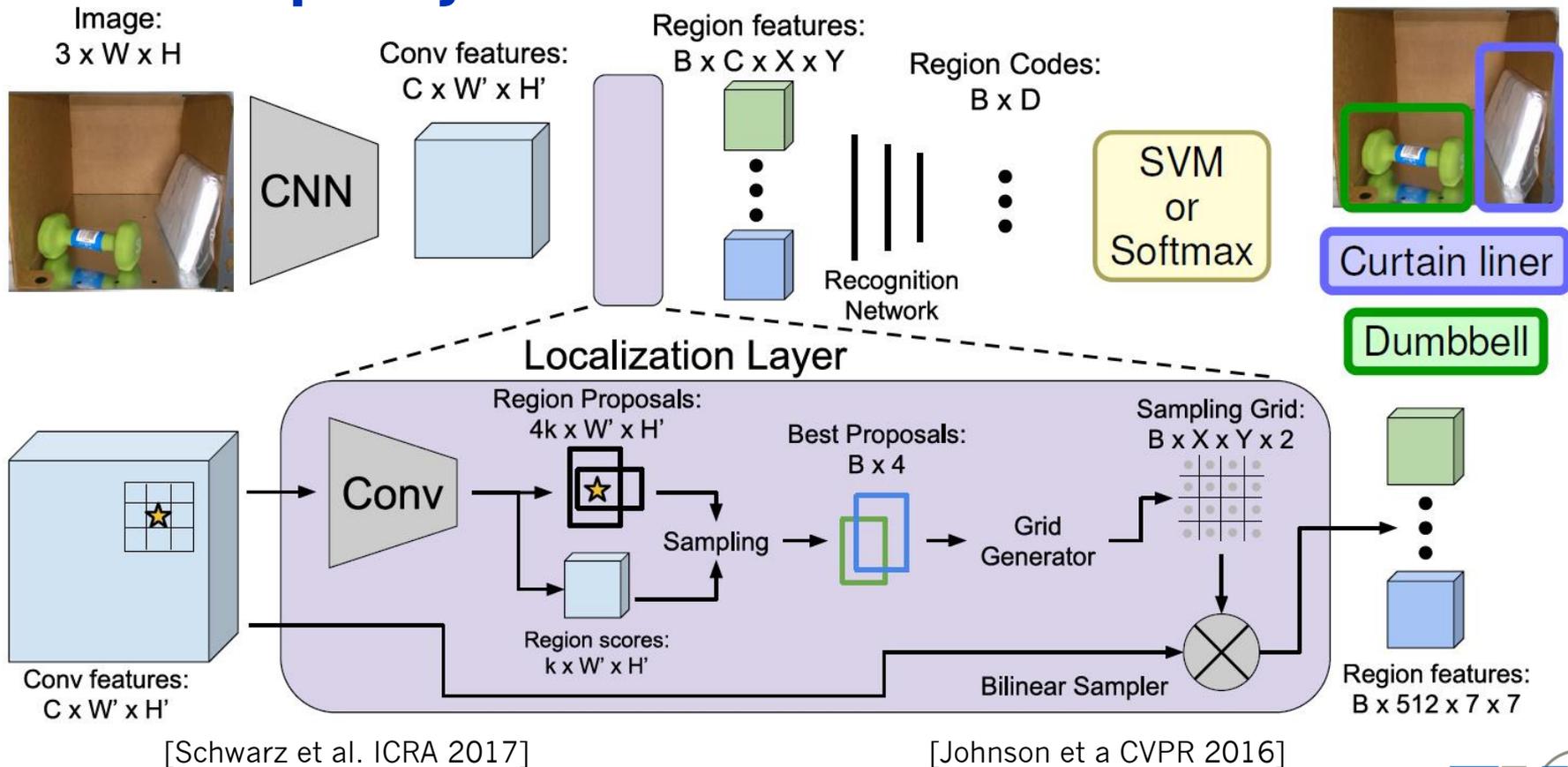
- Adapted from our segmentation of indoor scenes [Husain et al. RA-L 2016]



[Schwarz et al. ICRA 2017]

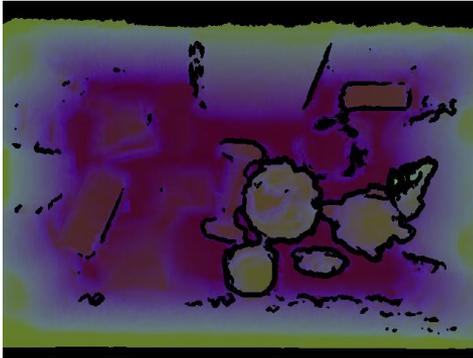
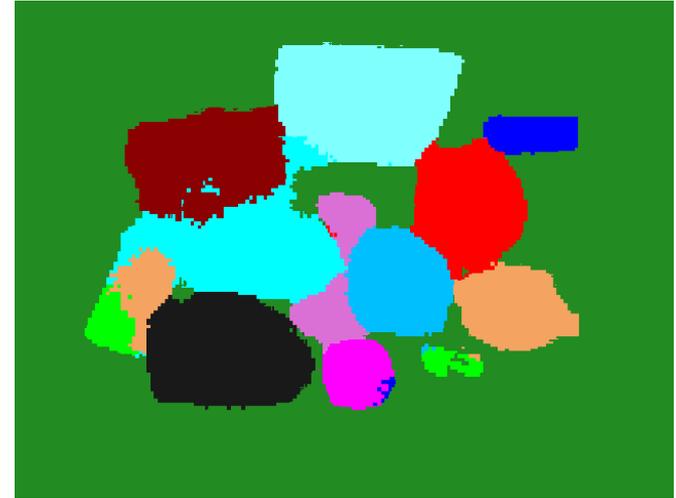


# DenseCap Object Detection



# Combined Detection and Segmentation

Detection



Segmentation

[Schwarz et al. IJRR 2017]

# Stowing



[Schwarz et al. ICRA 2017]

# Picking



# NimbRo Picking APC 2016 Results



- 2<sup>nd</sup> Place Stowing (186 points)
- 3<sup>rd</sup> Place Picking (97 points)



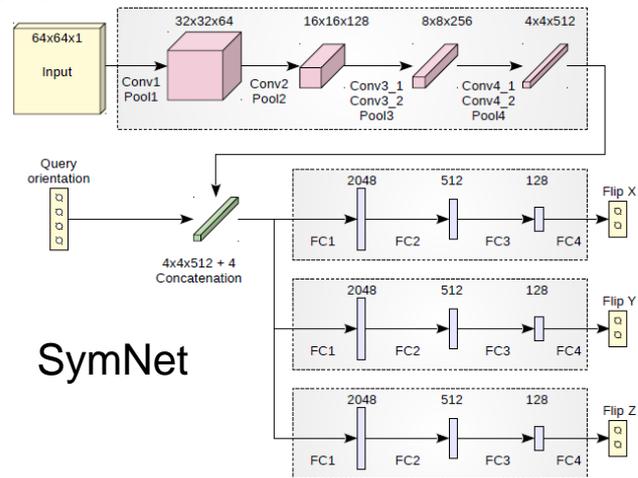
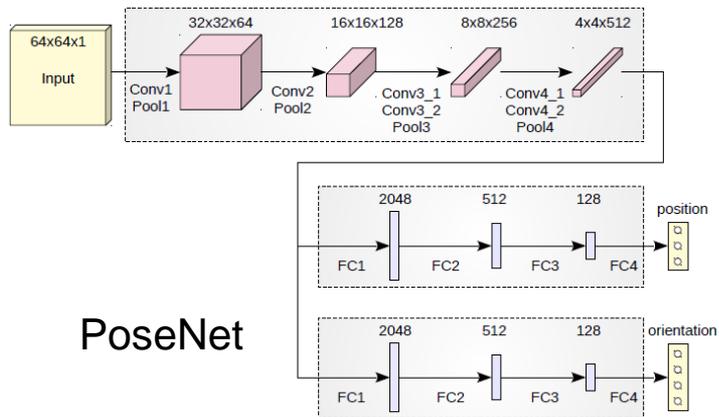
[Schwarz et al. ICRA 2017]

# EuRoC C1 Robolink Feeder: Bin Picking

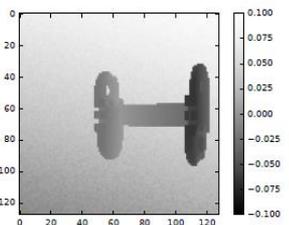
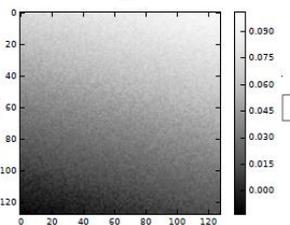
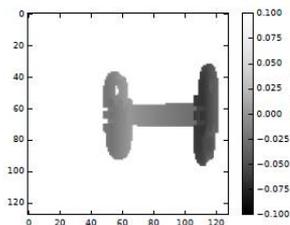


# Part Pose Estimation

- Two convolutional neural networks



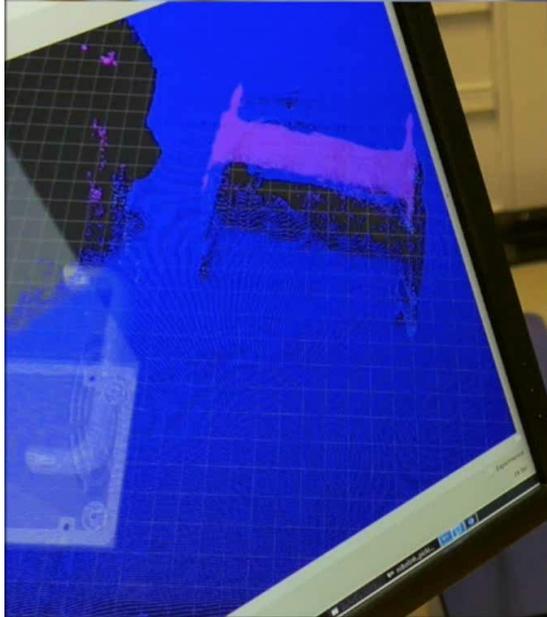
- Training with synthetic depth images



[Koo et al. CASE 2017]

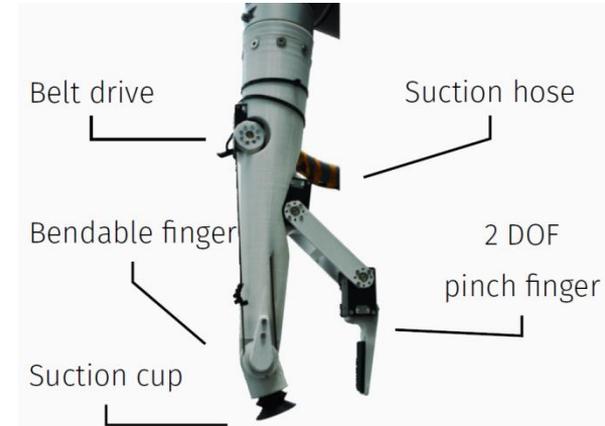
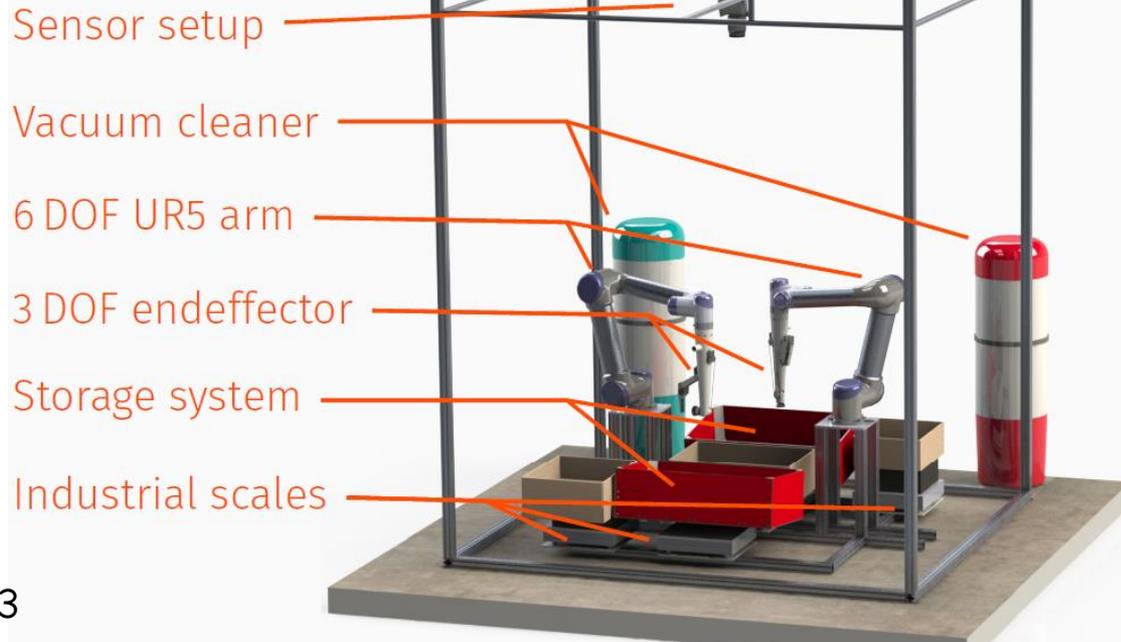
# Robolink Feeder: Regrasping and Placing

Pose estimation



# Amazon Robotics Challenge 2017

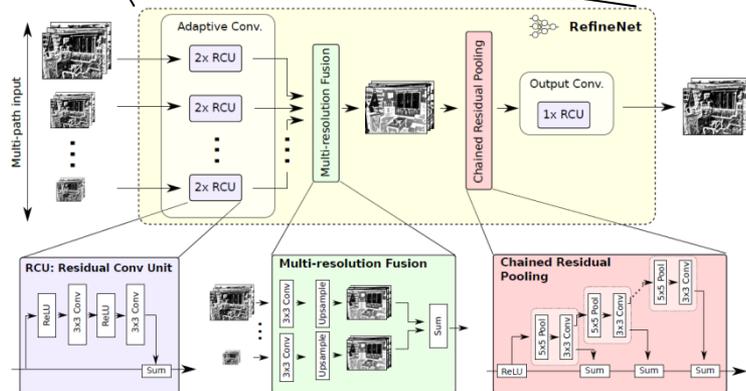
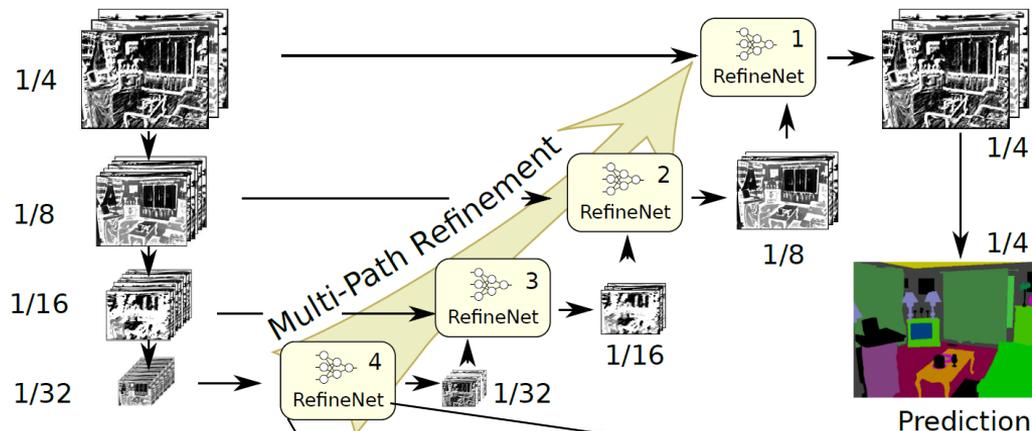
- Quickly learn novel objects
- Design own storage system



[Schwarz et al. ICRA 2018]

# RefineNet for Semantic Segmentation

- Increase resolution by using features from the higher resolution
- Coarse-to-fine semantic segmentation



[Lin et al. CVPR 2017]

# Object Capture and Scene Rendering

- Turn table + DLSR



Rendered scenes



# ARC 2017 Perception Example



- bronze\_wire\_cup  
conf: 0.749401
- irish\_spring\_soap  
conf: 0.811500
- playing\_cards  
conf: 0.813761
- w\_aquarium\_gravel  
conf: 0.891001
- crayons  
conf: 0.422604
- reynolds\_wrap  
conf: 0.836467
- paper\_towels  
conf: 0.903645
- white\_facecloth  
conf: 0.895212
- hand\_weight  
conf: 0.928119
- robots\_everywhere  
conf: 0.930464



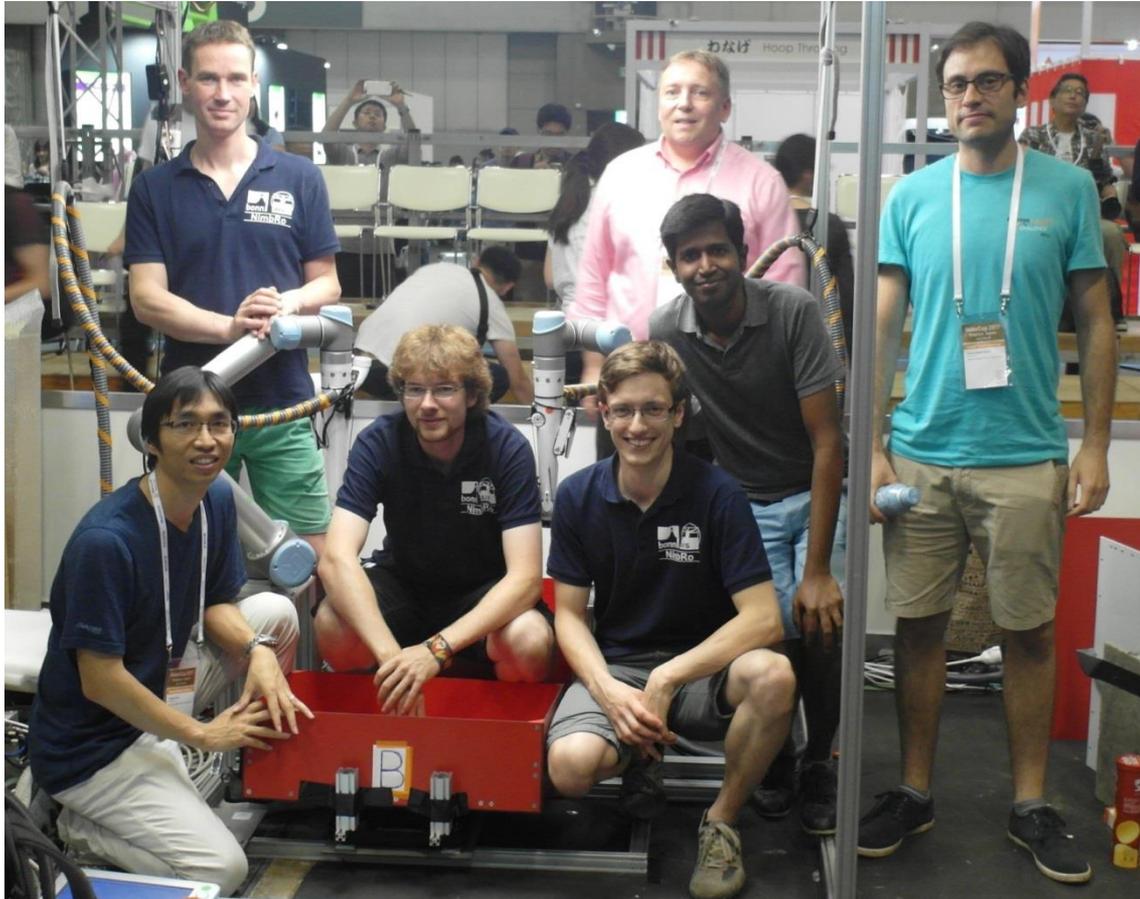
- mouse\_traps  
conf: 0.921731
- windex  
conf: 0.861246
- q-tips\_500  
conf: 0.475015
- fiskars\_scissors  
conf: 0.831069
- ice\_cube\_tray  
conf: 0.976856

# Amazon Robotics Challenge 2017 Final



# NimbRo Picking 2017 Team

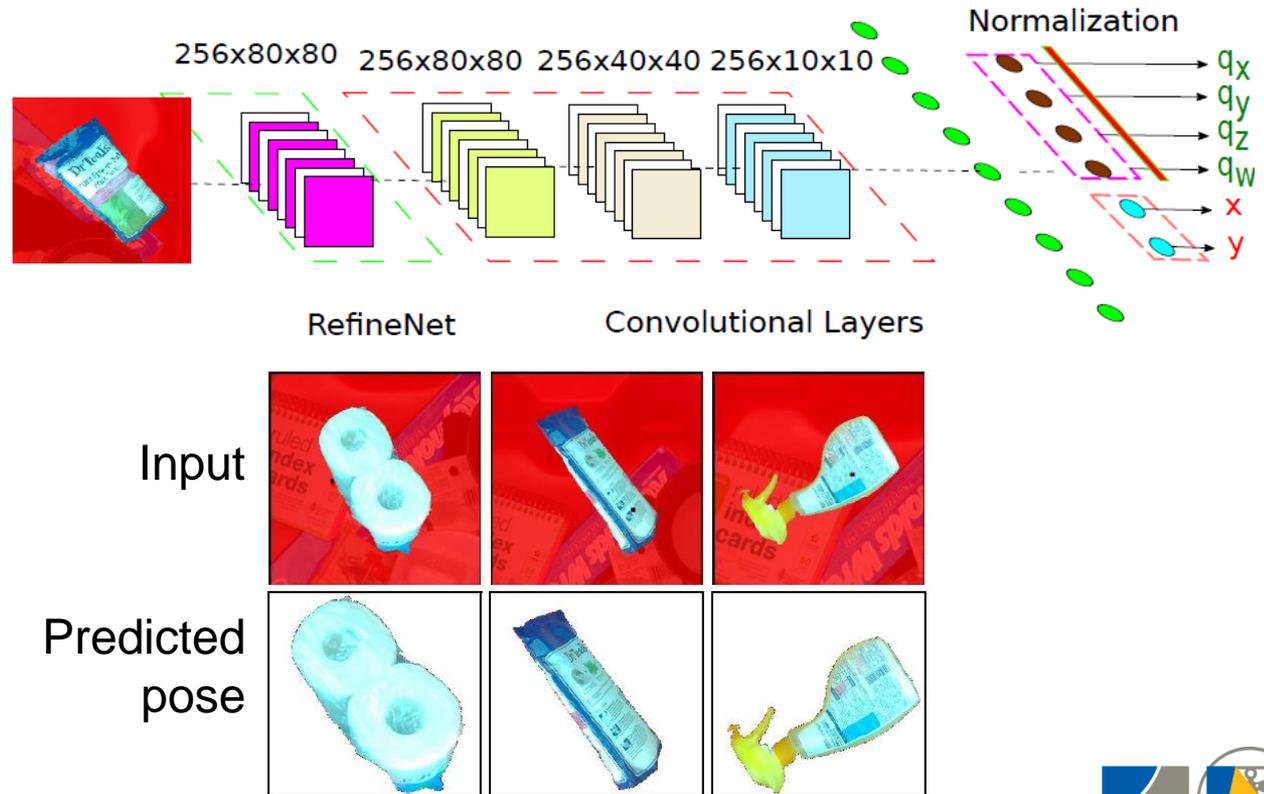
- 2<sup>nd</sup> place Pick
- 2<sup>nd</sup> place Stow-and-Pick Final



[Schwarz et al.  
ICRA 2018]

# Object Pose Estimation

- Cut out individual segments
- Use upper layer of RefineNet as input
- Predict pose coordinates



# Transfer of Manipulation Skills

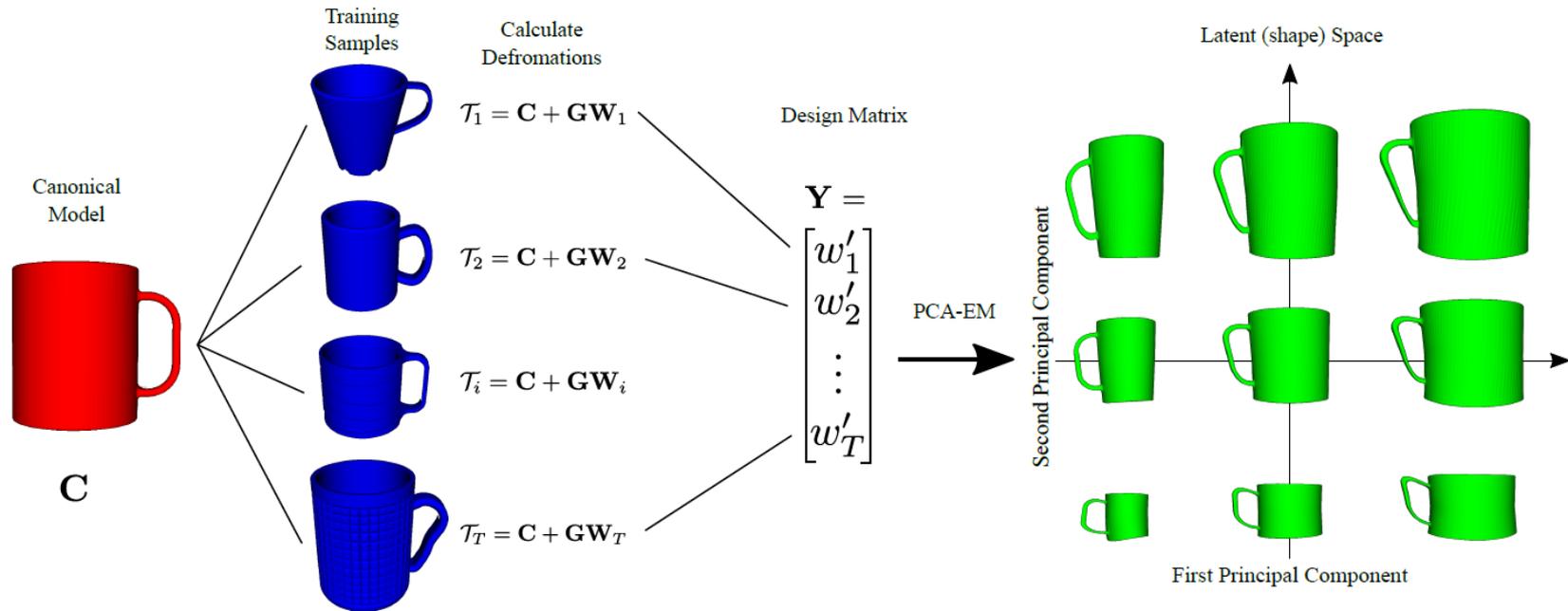


Knowledge  
Transfer

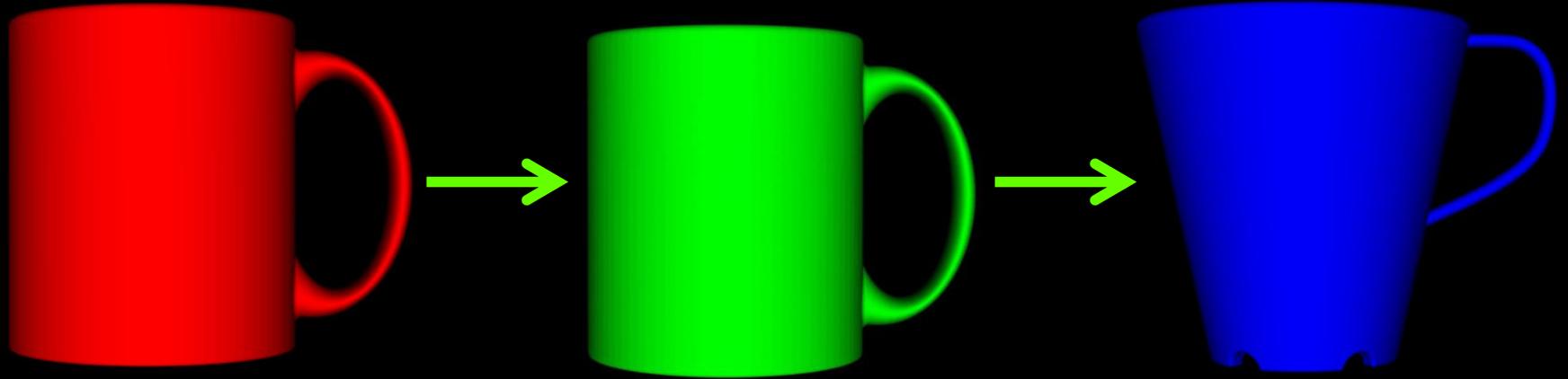


# Learning a Latent Shape Space

- Non-rigid registration of instances and canonical model
- Principal component analysis of deformations

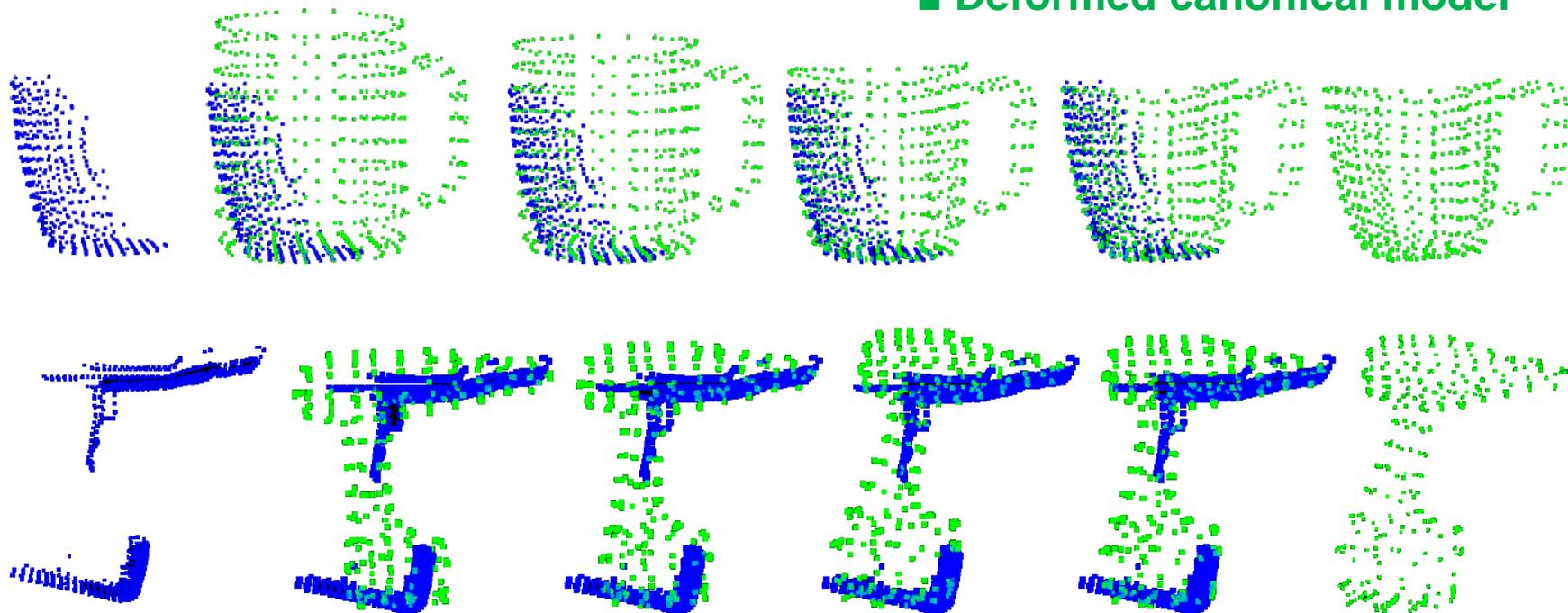


# Interpolation in Shape Space



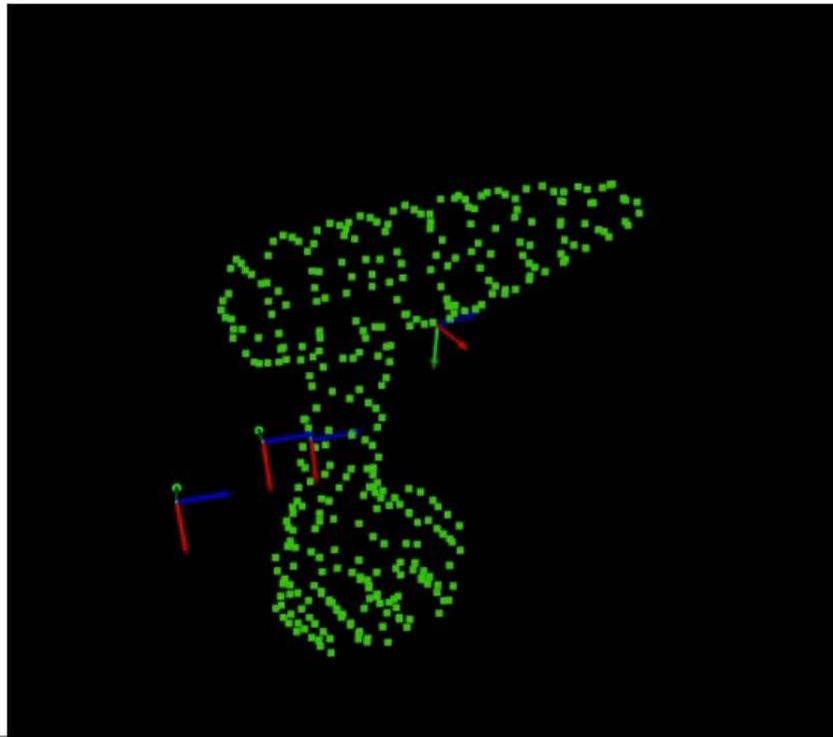
# Shape-aware Non-rigid Registration

- Partial view of novel instance
- Deformed canonical model



# Transference of Grasping Skills

Warp grasping information



# Grasping an Unknown Power Drill



# Conclusions

- Developed methods for item perception in bin picking
  - Shape primitive-based
  - View registration based
  - Deep learning based
- Challenges include
  - Item variability
  - Material properties
  - Articulation of objects
  - Occlusions
- Possible approaches
  - Deformable and articulated models
  - Active perception
  - Interactive perception



# Thank you very much for your attention!

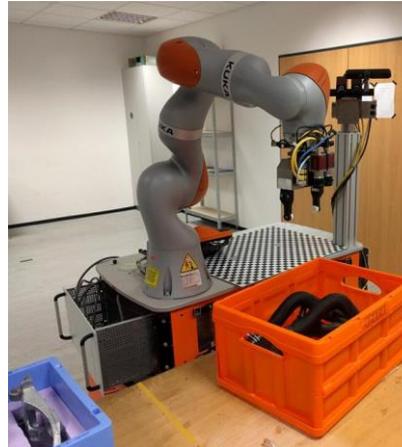
ActReMa



EuRoC C1



STAMINA



EuRoC C2

Amazon Picking

