Learning Semantic Perception for Cluttered Bin Picking

Sven Behnke and Max Schwarz

University of Bonn
Computer Science Institute VI
Autonomous Intelligent Systems

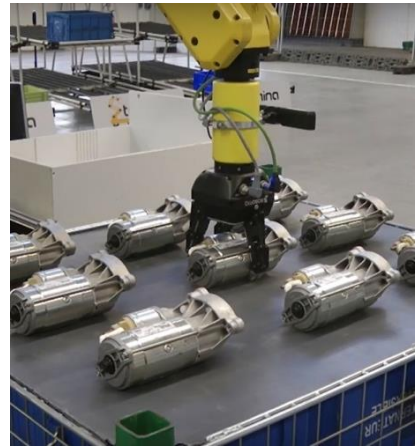
Bin Picking

- 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 Past Experience

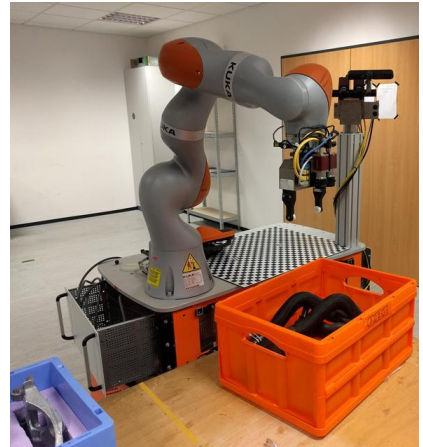
ActReMa



STAMINA

Amazon Picking

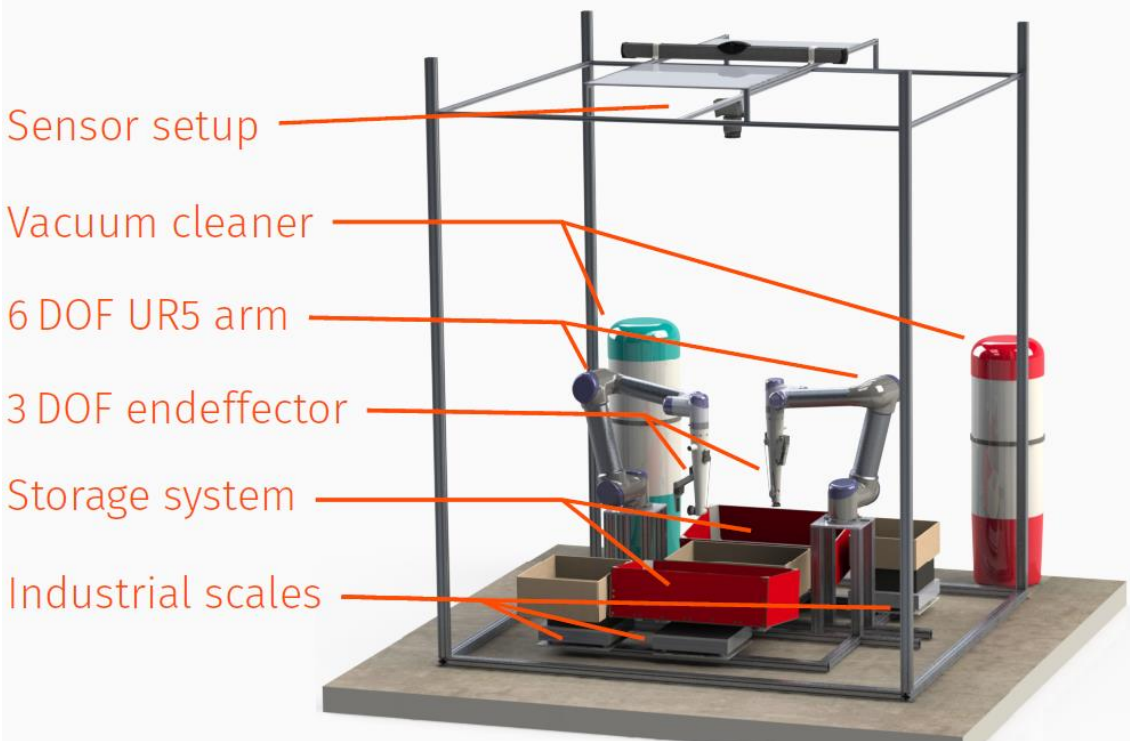
EuRoC C1

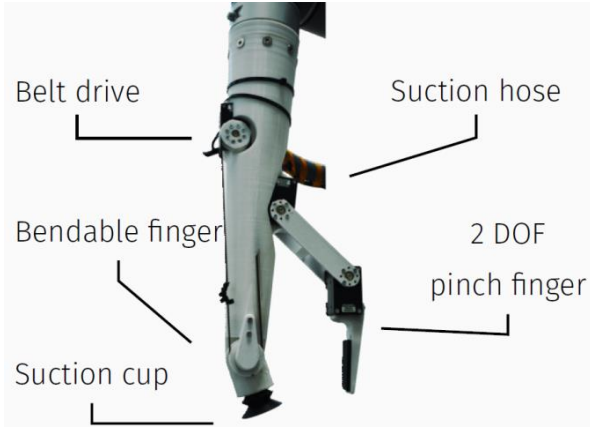


EuRoC C2

Amazon Robotics Challenge 2017

- Quickly learn novel objects
- Design own storage system

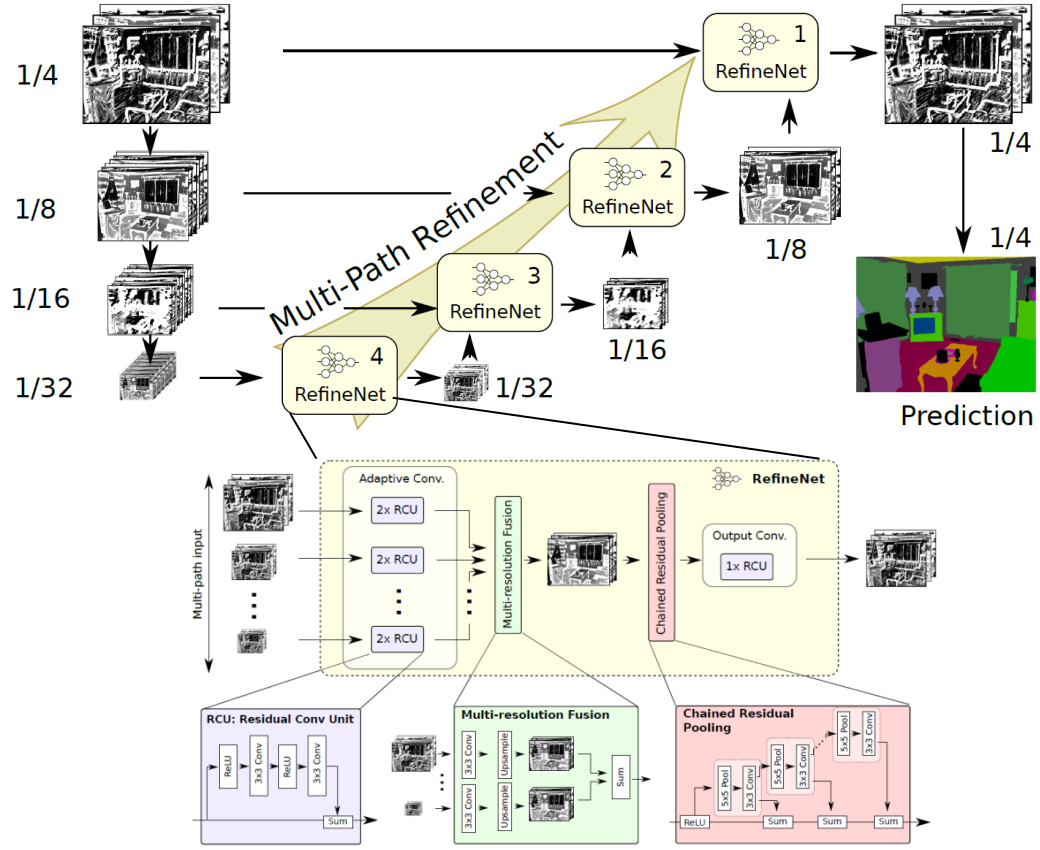




[Schwarz et al. ICRA 2018]

RefineNet for Semantic Segmentation

- Scene represented as feature hierarchy
- Corse-to-fine semantic segmentation
- Combine higher-level features with missing details



[Lin et al. CVPR 2017]

The Data Problem

- Deep Learning in robotics (still) suffers from shortage of available examples
- We address this problem in two ways:

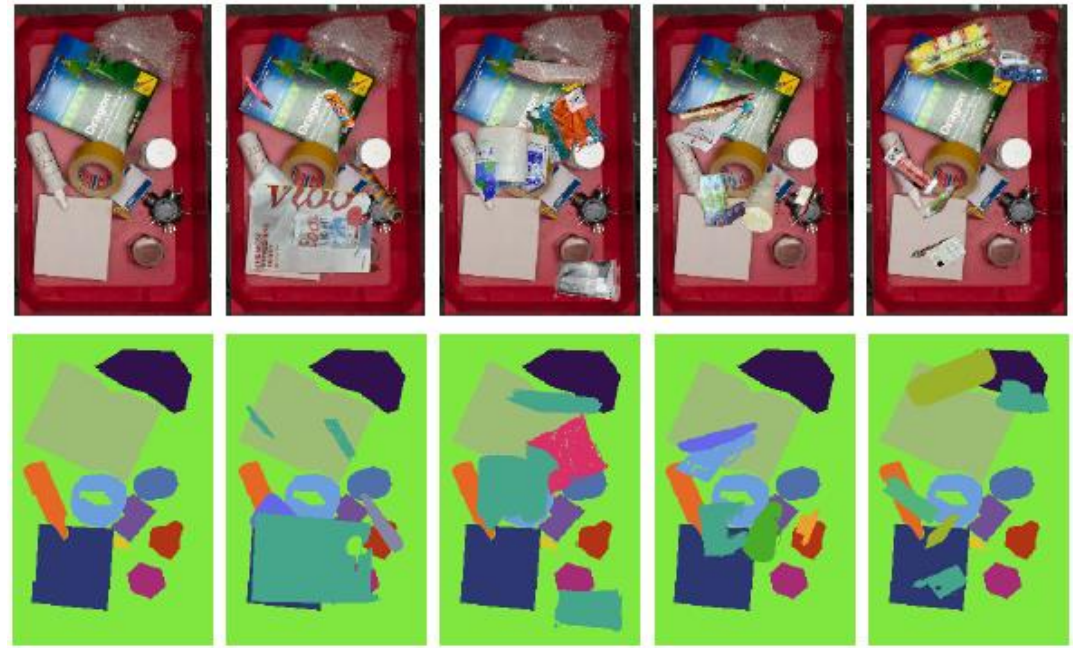
1. **Generating data**:
   Automatic data capture,
   online mesh databases,
   scene synthesis

2. **Improving generalization**:
   Object-centered models,
   deformable registration,
   transfer learning,
   semi-supervised learning

Object Capture and Scene Rendering

- Turntable + DLSR camera

- Rendered scenes



[Schwarz et al. ICRA 2018]

ARC 2017 Perception Example

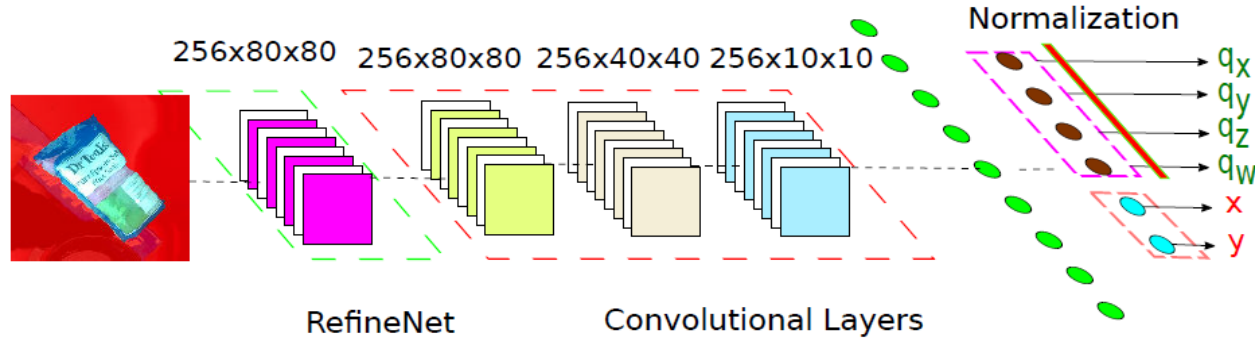
[Schwarz et al. ICRA 2018]

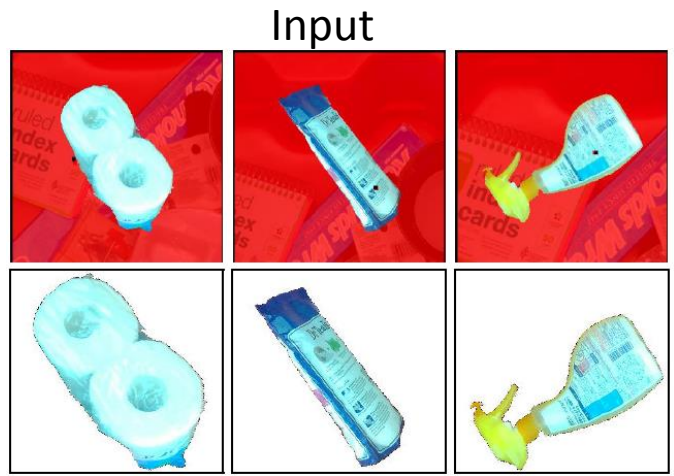
Amazon Robotics Challenge 2017 Final

[Schwarz et al. ICRA 2018]

Object Pose Estimation

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





Object	Translation [pixel[1]]		Orientation [°]	
	train	val	train	val
Browns brush	10.3	11.4	7.7	10.3
Epsom salts	11.2	12.5	7.4	10.5
Hand weight	9.6	10.4	2.1	2.6
Reynolds wrap	11.6	11.8	6.3	9.8
Utility brush	12.5	13.6	6.9	10.9

[1] Relative to the 320×320 crop centered on the object.

[Schwarz et al. ICRA 2018, Periyasamy et al. IROS 2018]

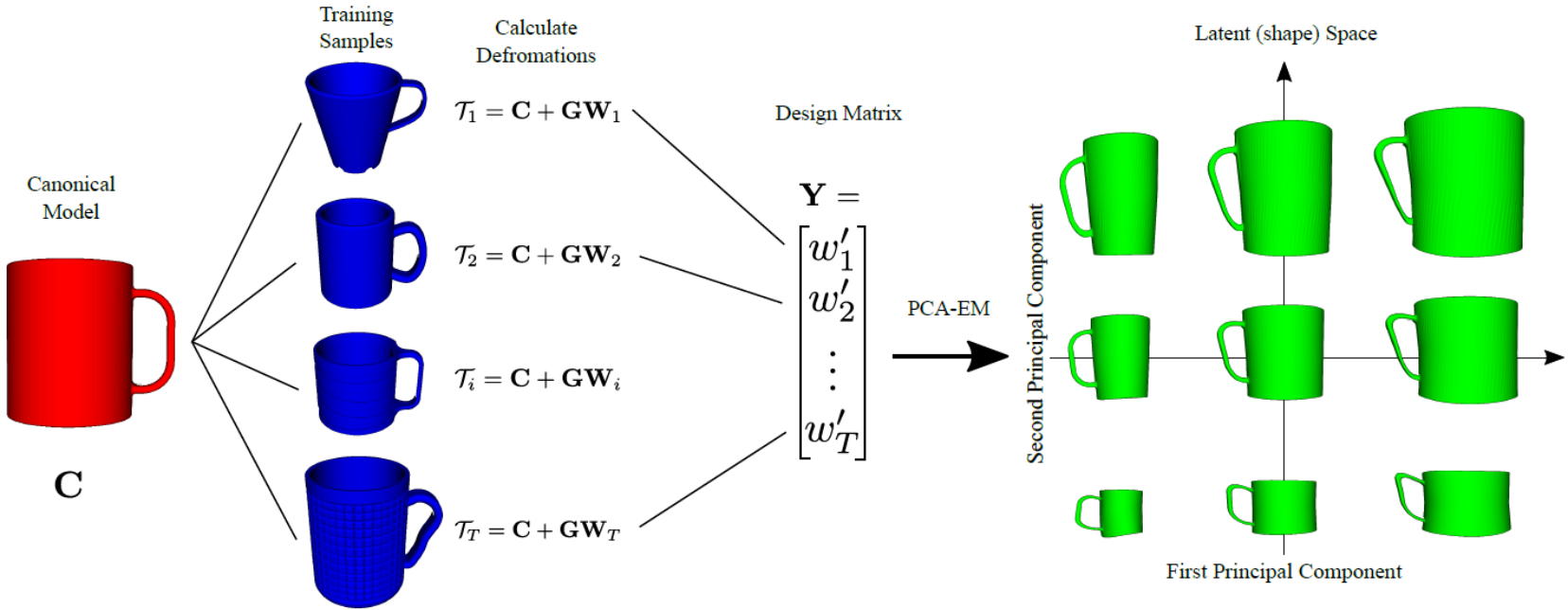
Transfer of Manipulation Skills

Knowledge Transfer

[Rodriguez and Behnke ICRA 2018]

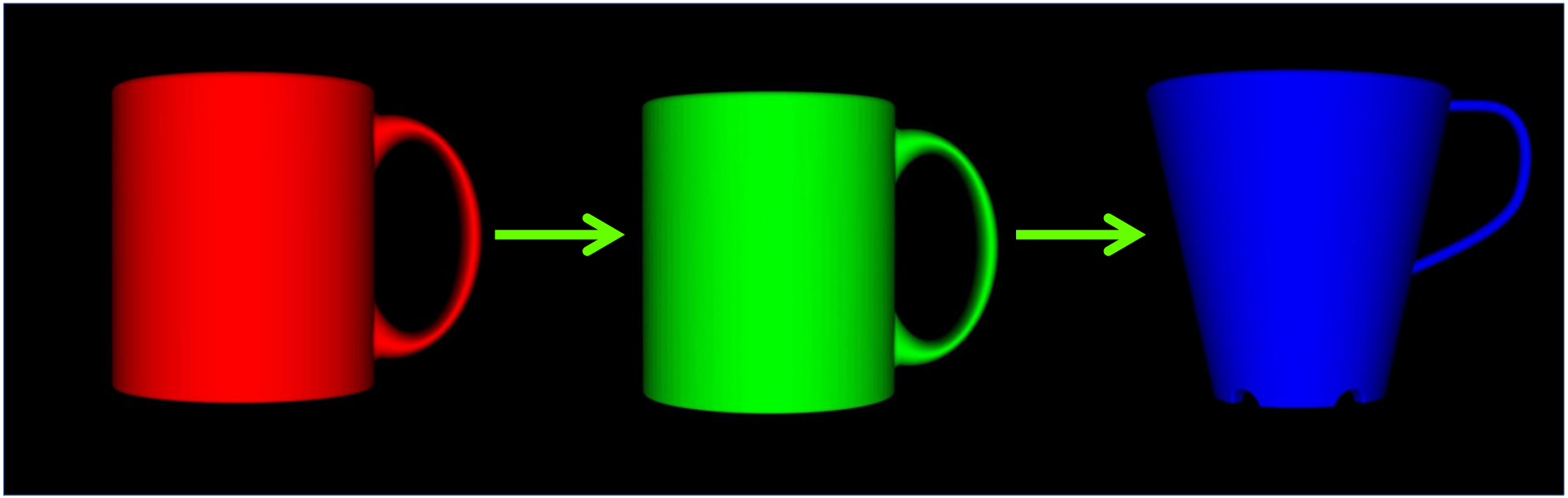
Learning a Latent Shape Space

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



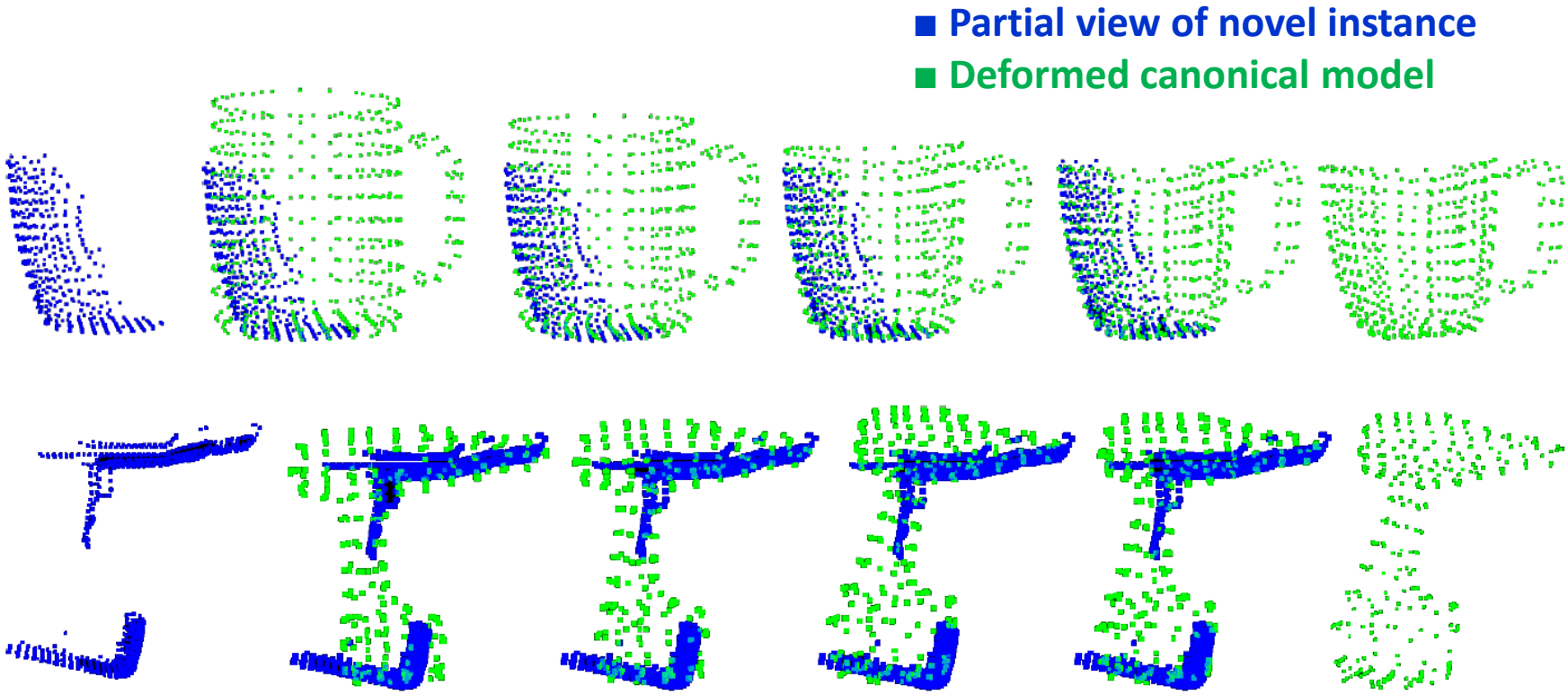
[Rodriguez and Behnke ICRA 2018]

Interpolation in Shape Space



[Rodriguez and Behnke ICRA 2018]

Shape-aware Non-rigid Registration



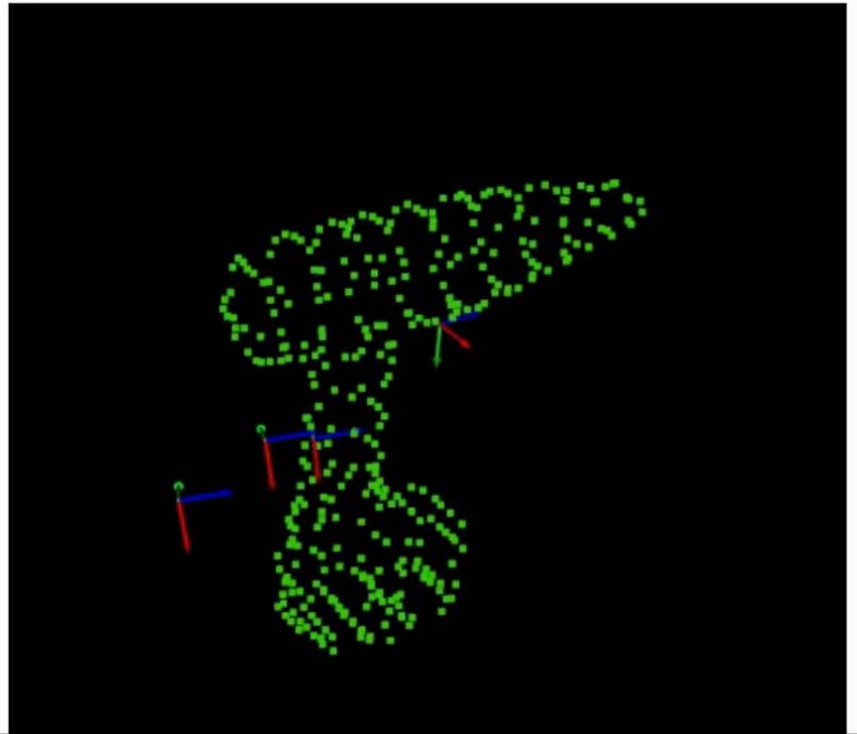
[Rodriguez and Behnke ICRA 2018]

[Rodriguez and Behnke ICRA 2018]

UNIVERSITÄT BONN

Transference of Grasping Skills

Warp grasping information

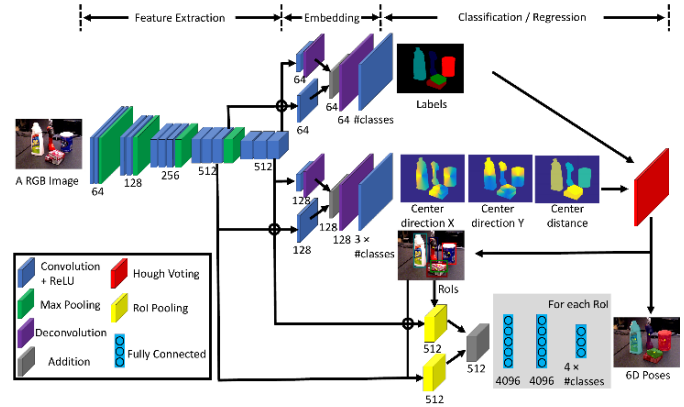


Grasping an Unknown Power Drill

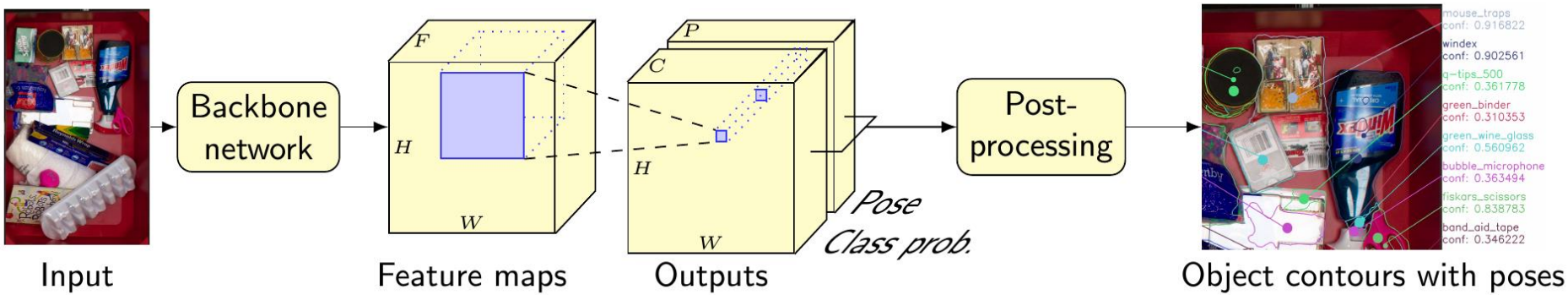
[Rodriguez and Behnke ICRA 2018]

Fully Convolutional 6D Pose Estimation

- Extending PoseCNN
  [Xiang, Schmidt, Narayanan, Fox: PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. RSS 2018]

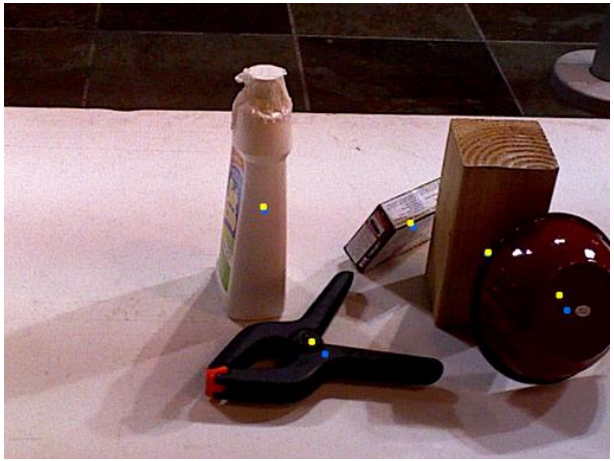


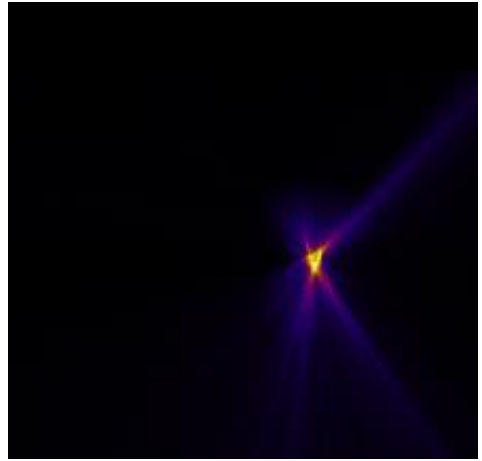
- Fully convolutional (per-pixel) prediction of pose parameters: 2D center offset, depth, orientation as quaternion

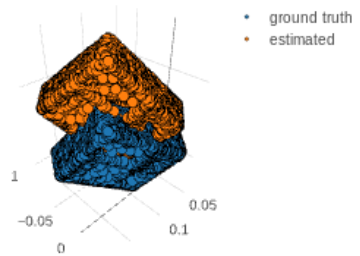


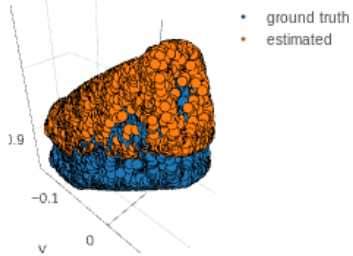
Fully Convolutional 6D Pose Estimation

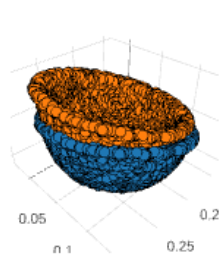
- Hough voting to find object centers in 2D
- Quaternions aggregated using Hough inliers and semantic segmentation

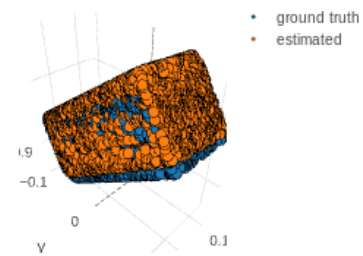












[Example from Pose-CNN data set Xiang et al. RSS 2018]

Capturing More Data: ARC Bin Dataset

Sensors: Nikon DSLR + Photoneo PhoXi XL scanner. 230 bin frames, 30 different objects

From Turntable Captures to Meshes

- Turntable setup:
  - DSLR (Nikon D3400)
  - Depth sensor (PrimeSense Carmine)
  - Dynamixel actuator
- Fast calibration:
  - Automatic capture of checkerboard images
  - Ceres-based optimization of camera extrinsics
- Meshing:
  - Masking via Background Subtraction
  - Extract isosurface of visual hull + TSDF from depth sensor
  - Texturing with Color Map Optimization

Range Sensor and Silhouette Fusion for High-Quality 3D Scanning
Karthik S. Narayan, James Sha, Arjun Singh, and Pieter Abbeel
ICRA 2015

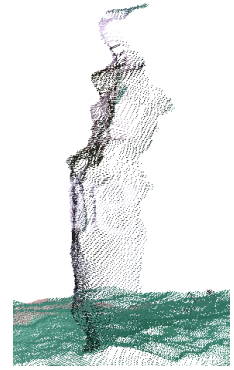
Color Map Optimization for 3D Reconstruction with Consumer Depth Cameras
Qian-Yi Zhou, Vladlen Koltun
ACM TOG 2014

From Turntable Captures to Meshes

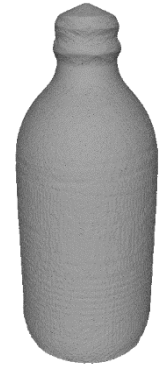
Fused & textured result

Reflections & Unreliable Depth

DSLR image



Depth camera



Visual hull

Concavities

Visual hull

Fused result

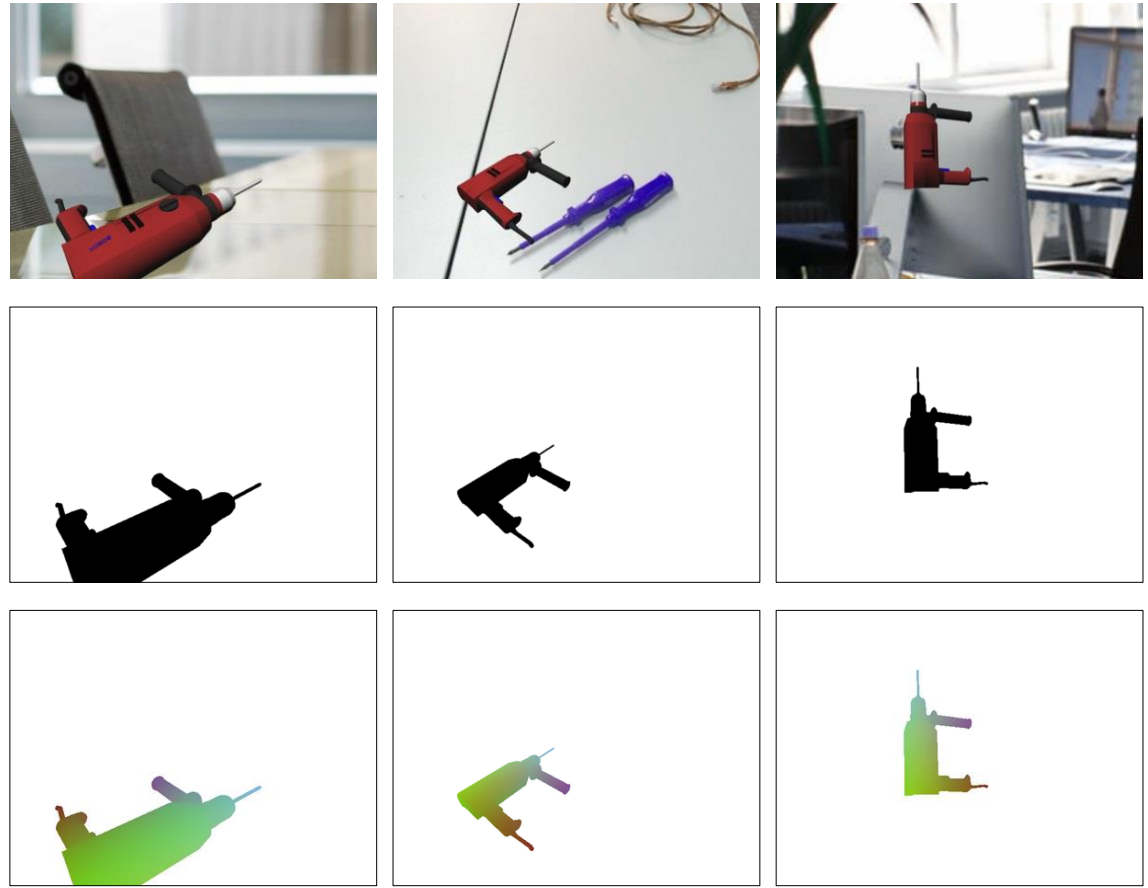
An Alternative: CAD Meshes

- Meshes found by search term "drill" on https://sketchfab.com

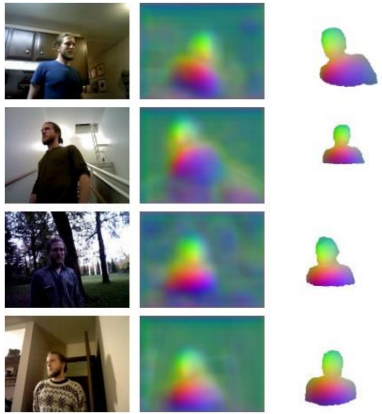
Rendering 3D Scenes

Advantages of mesh-based scene synthesis:

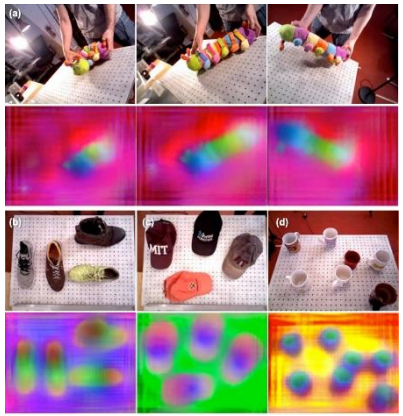
- Generate new scenes on-the-fly during training
- OpenGL/CUDA interop
- Simulate variations in hue, lighting, scale, rotation, camera intrinsics, …
- Cheap ground truth:
  - segmentation labels
  - object-centric coordinates
  - occlusion information
  - …



Self-Supervised Feature Learning



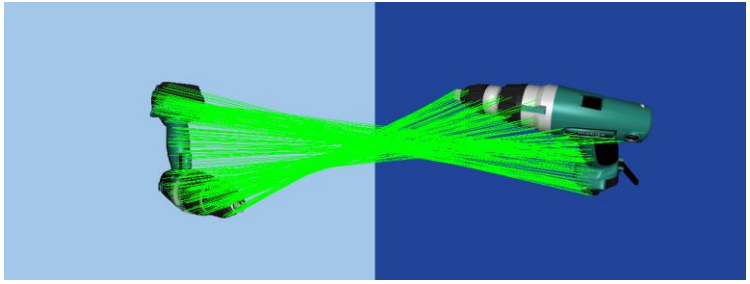
Self-supervised Visual Descriptor Learning for Dense Correspondence
T. Schmidt, R. A. Newcombe, D. Fox
Robotics and Automation Letters, 2017



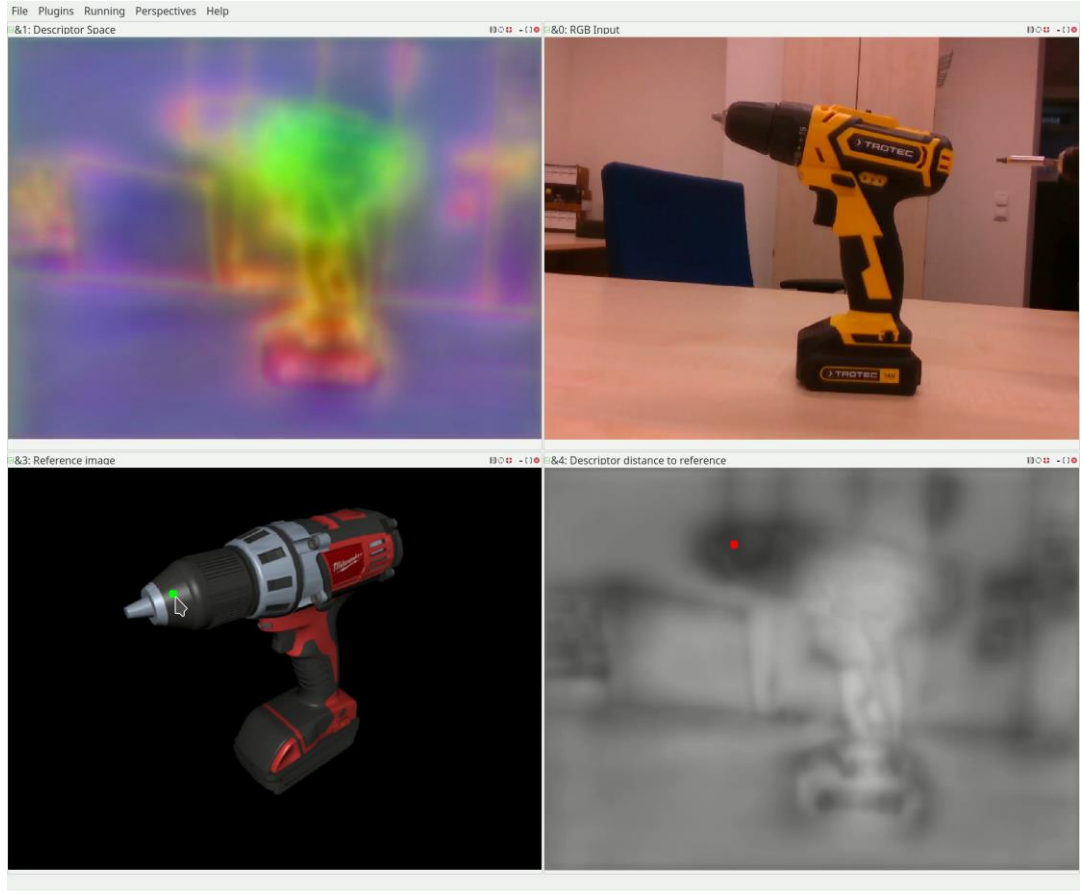
Dense Object Nets: Learning Dense Visual Object Descriptors
by and for Robotic Manipulation
P. Florence, L. Manuelli, R. Tedrake
CoRL 2018

Visual Descriptor Learning

- Trained on 1100 frames rendered from 22 CAD meshes

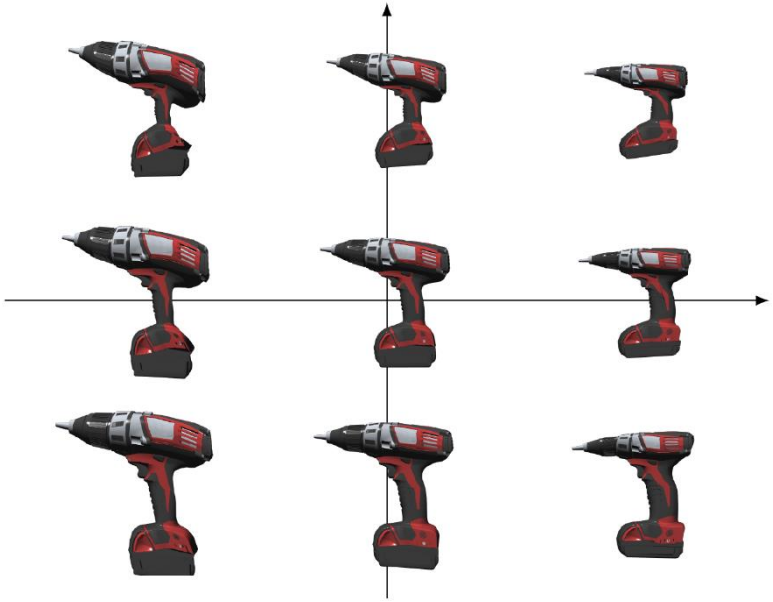


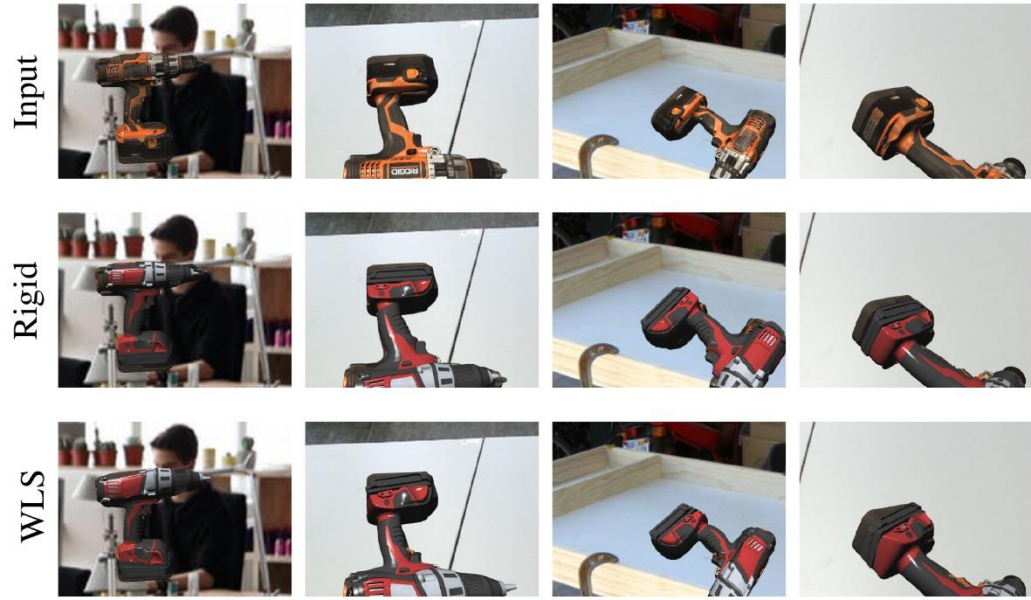
- Pixel-wise contrastive loss
- No training signal between different instances!



Combination with Non-Rigid Registration

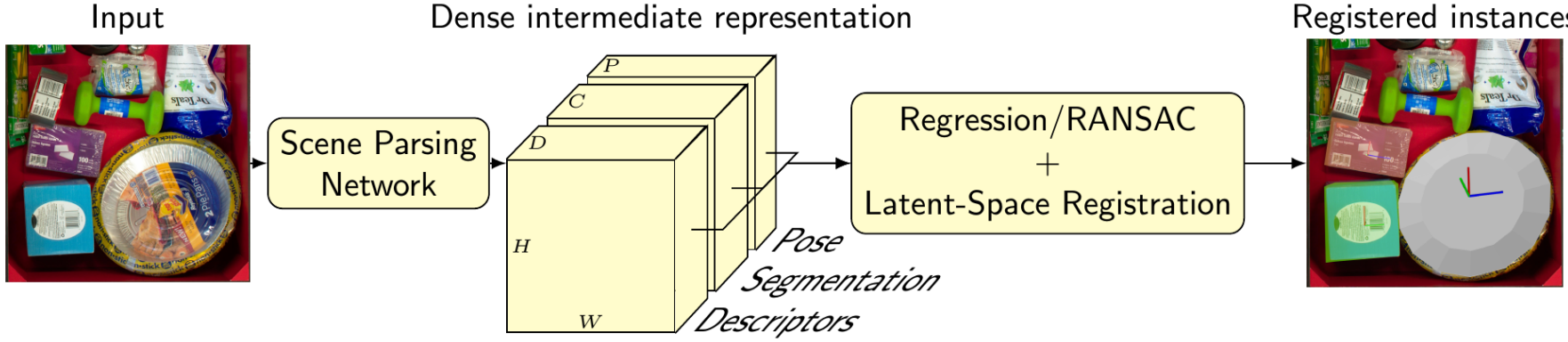
- Shape space creation using correspondence from visual descriptors
- Inference: Semantic segmentation, RANSAC, shape-aware fine registration



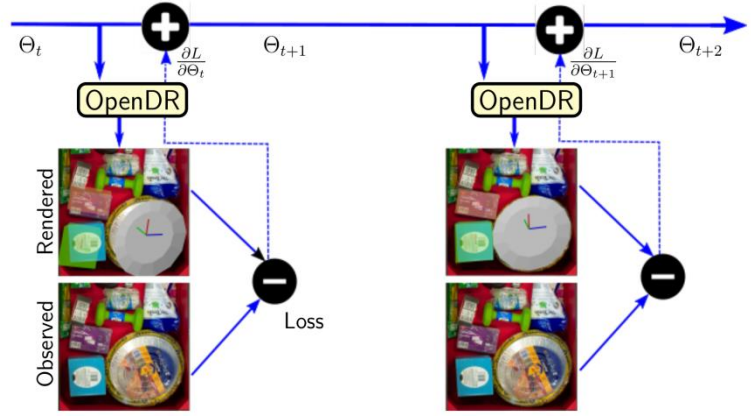


Outlook

- Complete scene parsing pipeline utilizing learned descriptors and shape models

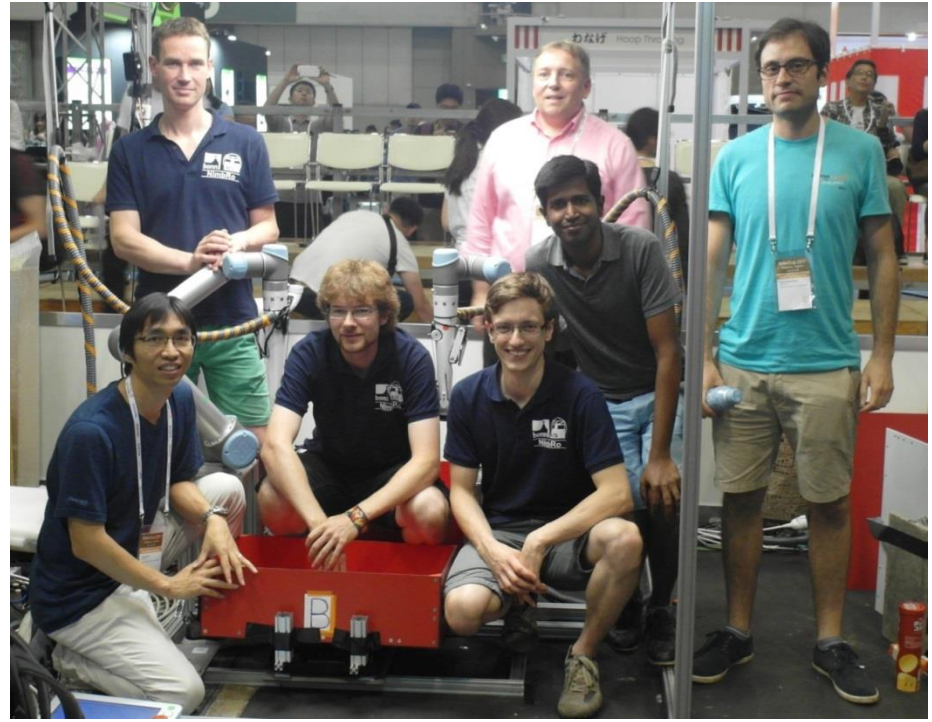


- Iterative refinement by render and compare



Conclusions

- Developed methods for learning semantic perception of cluttered bin scenes
- Integration to APC/ARC systems
- Addressed data problem by
  - Data capture and annotation
  - Synthesizing scenes
  - Deformable models
- Much further research needed for complete scene understanding



ARC 2017 team NimbRo Picking