Data-efficient Deep Learning for RGB-D Object Perception in Cluttered Bin Picking

Max Schwarz and Sven Behnke

Warehouse Picking Automation Workshop 2017
May 29, 2017

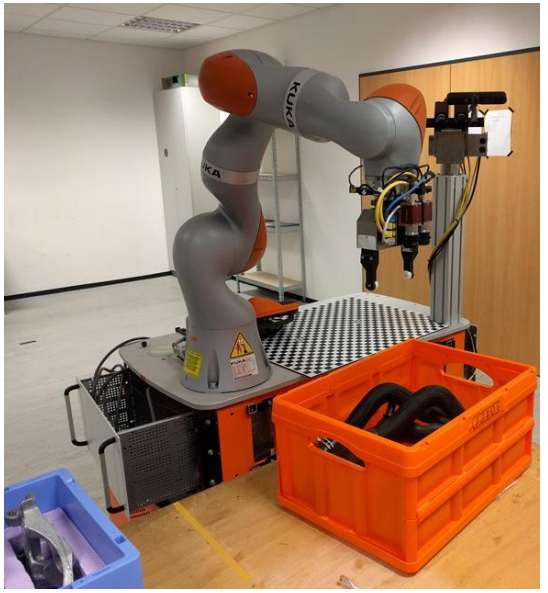
Rheinische Friedrich-Wilhelms-Universität Bonn
Computer Science Institute VI: Autonomous Intelligent Systems

Picking Activities in our Group

ActReMa

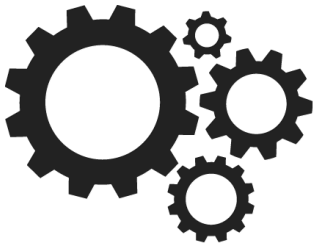
STAMINA

EuRoC C1



EuRoC C2

Outline



System

Control

Perception

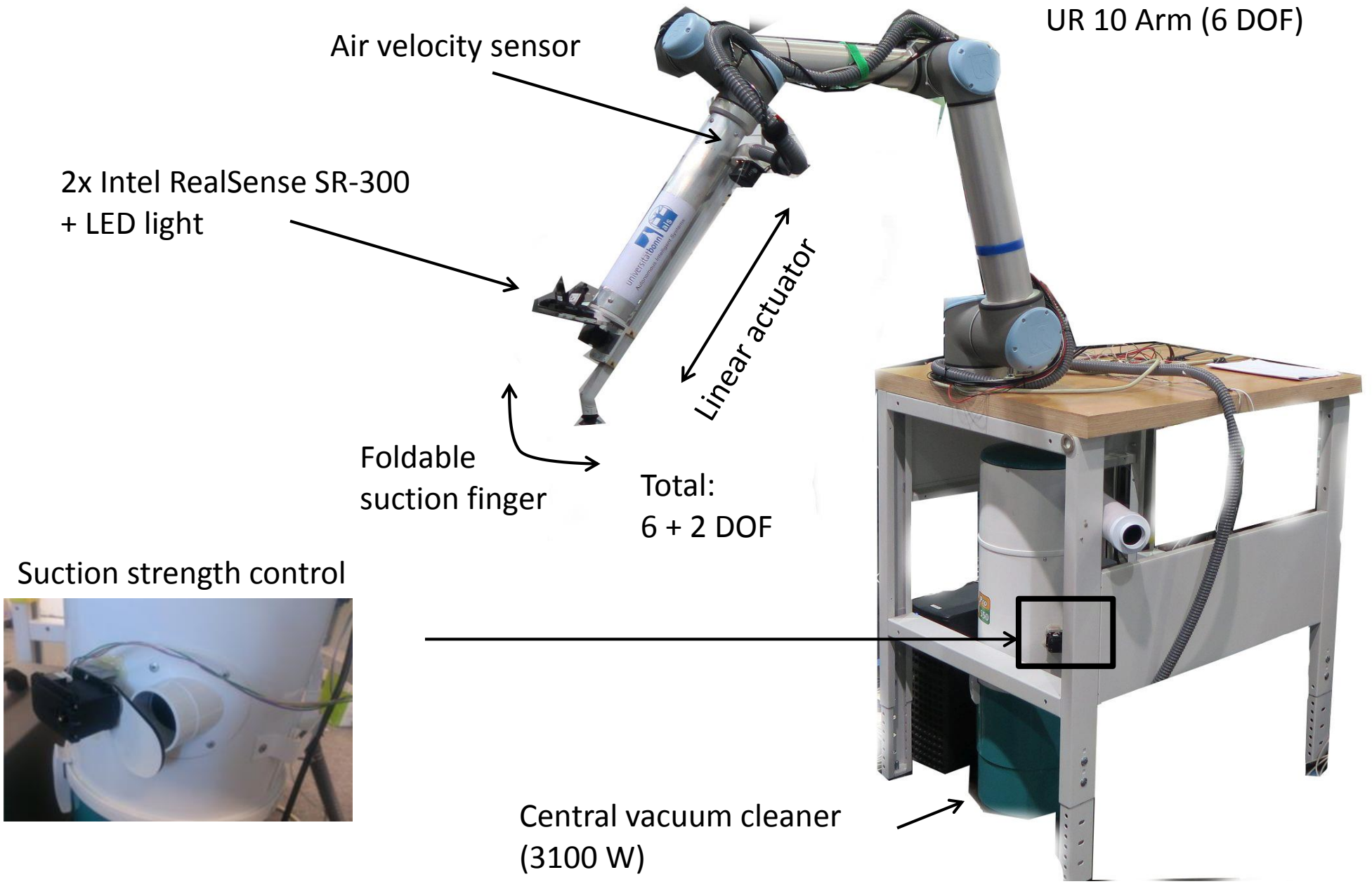
Special Features

Concept and Design

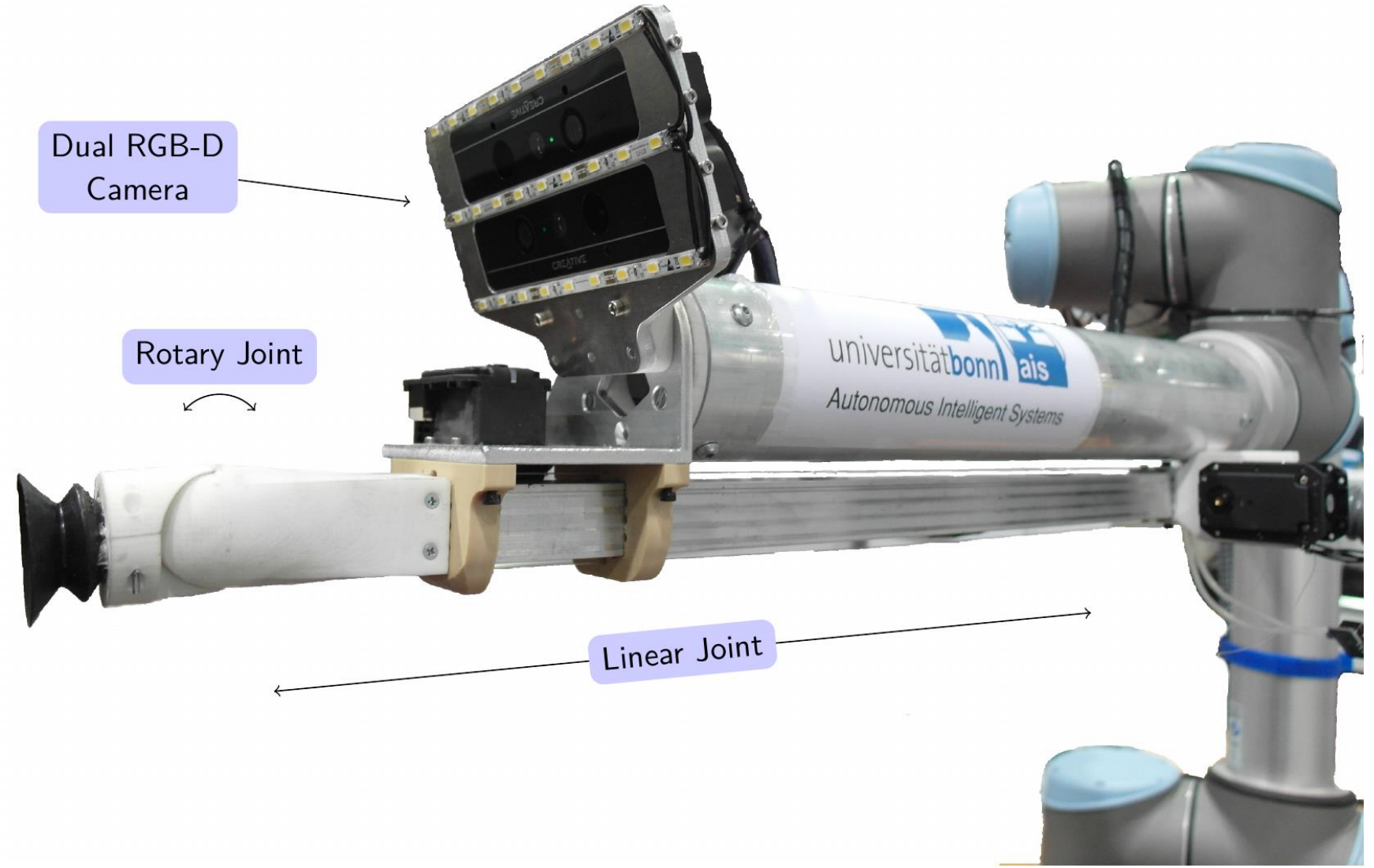
- UR 10: Workspace, Payload, Cost, Safety
- Single suction gripper: Avoid design complexity
  - Second supporting finger planned
- Folding finger:
  Front, top, and side grasps

Aim for highest performance at lowest complexity!

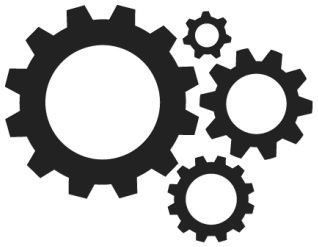
System Overview



Endeffector Design



Outline



System

Control

Perception

Special Features

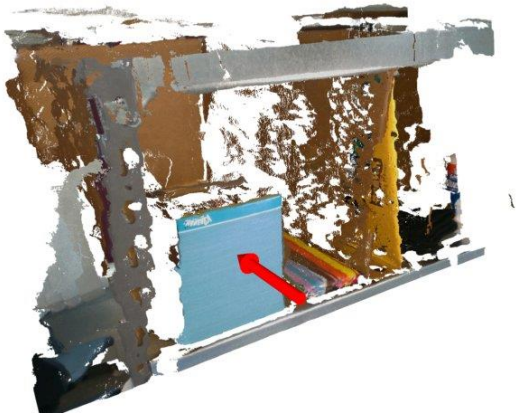
Motion Generation

- Heuristic-based grasp planning
- Replace complex motion planning with:
  - Keyframe-based motion generation
  - Collision detection at runtime (triggered in picking run)
- Assumption: If we can see a point, we can retrieve it using suction
- Self-collision detection using MoveIt!
- Avoid collisions with shelf in IK solver!

Generic Grasp Pose Selection

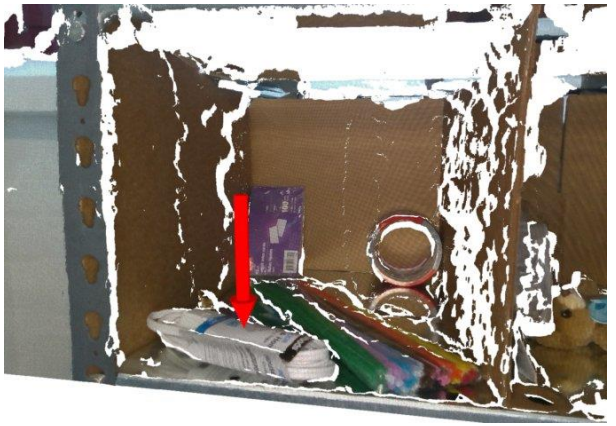
Center grasp for “standing” objects:

- Find support area for suction close to bounding box center



Top grasp for “lying” objects:

- Find support area for suction close to horizontal bounding box center

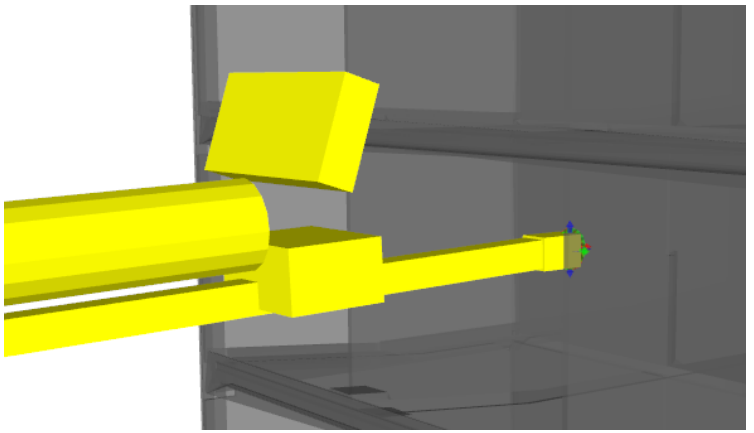


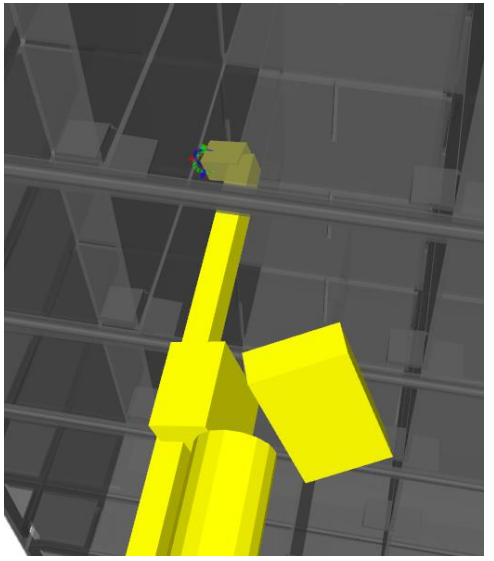
Custom rules for specific objects

(9 rules in total)

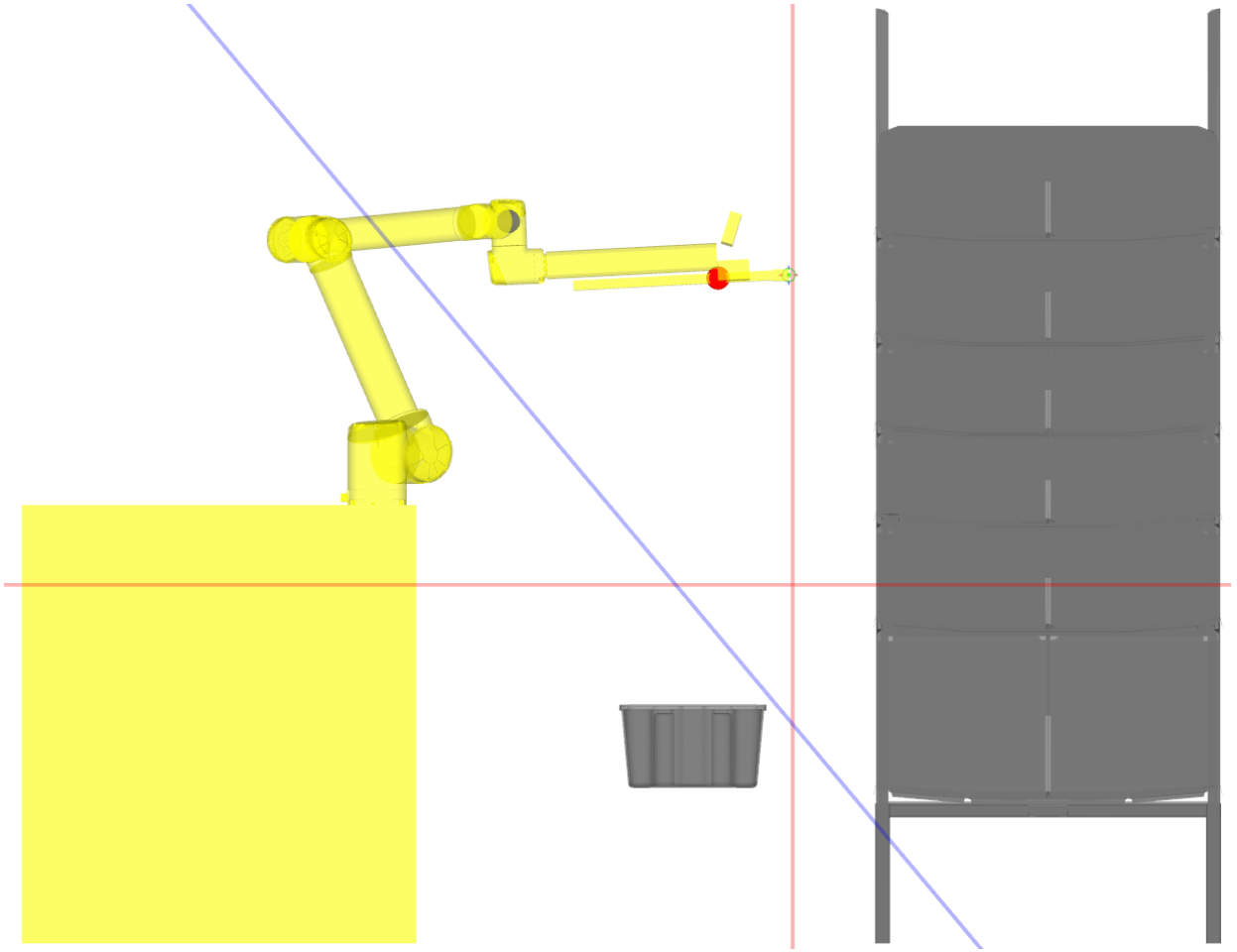
Inverse Kinematics

- Redundancy resolution by null-space cost optimization:
  - Joint limit avoidance
  - Cartesian plane avoidance (keep wrist out of shelf/tote)
  - Keep linear extension short
- Robust solution using damped least squares
- For in-shelf manipulation, only position + suction direction (5D IK)

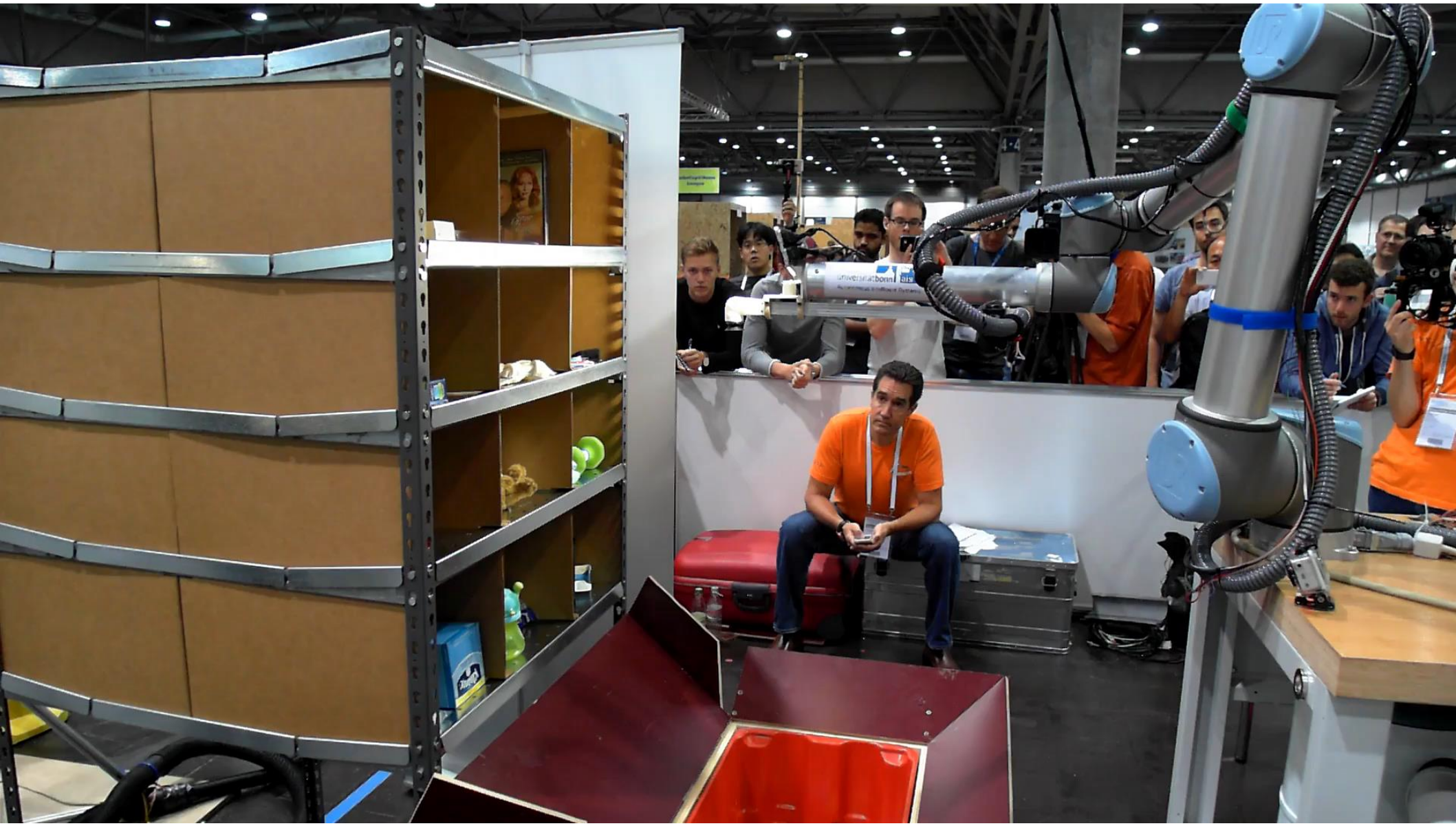




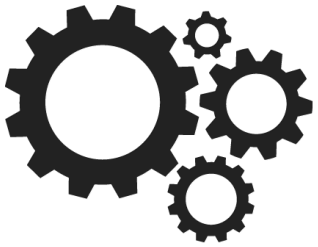
Limiting Planes for IK



Motion Execution



Outline



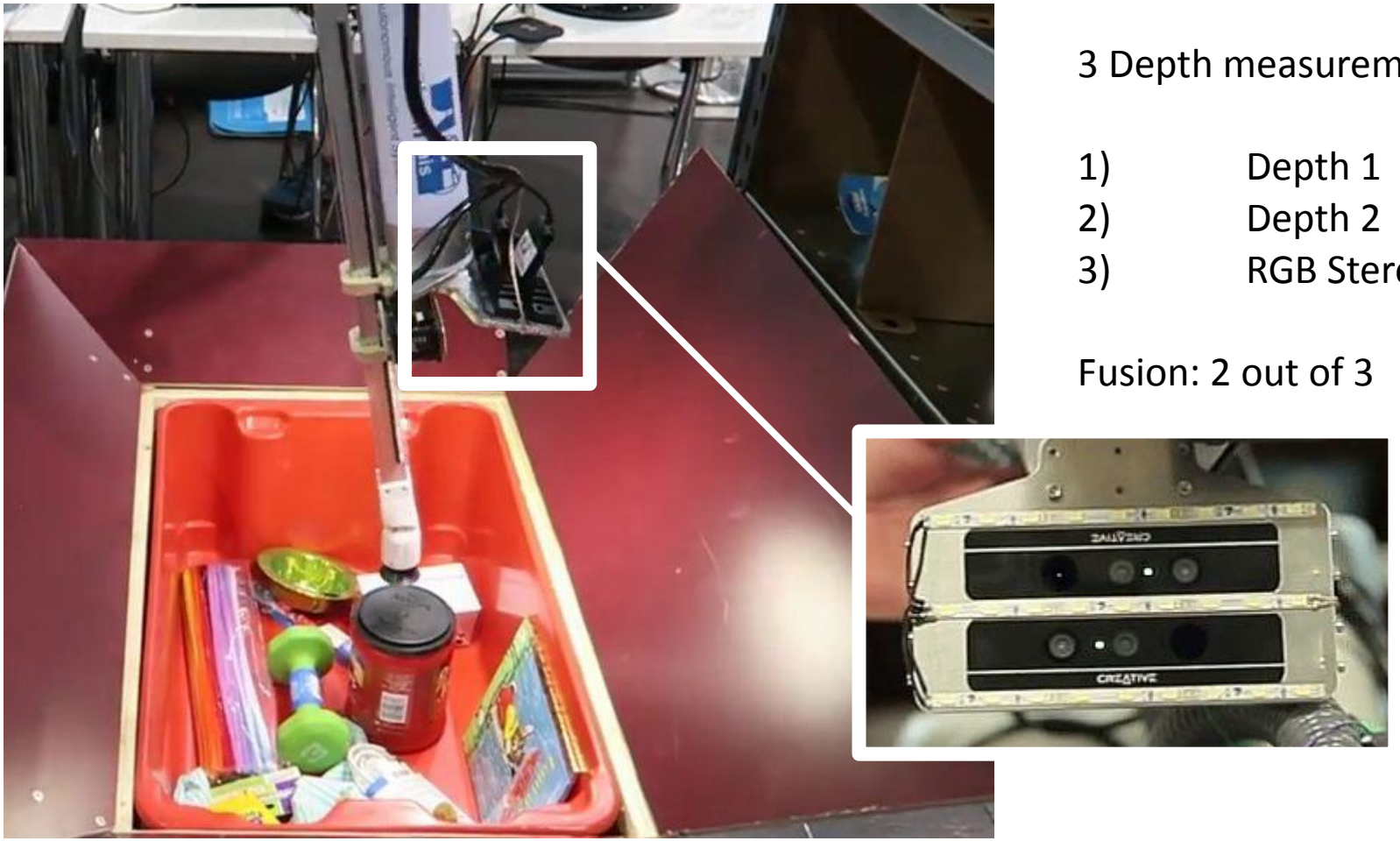
System

Control

Perception

Special Features

Sensors



2x Intel RealSense SR-300

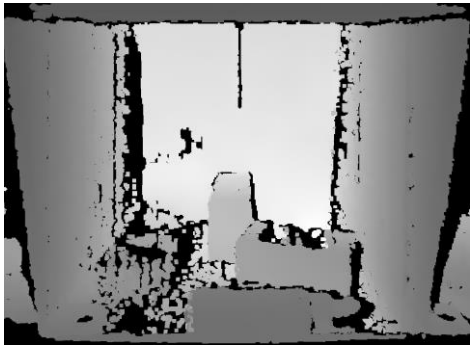
3 Depth measurements per pixel

1) Depth 1
2) Depth 2
3) RGB Stereo

Fusion: 2 out of 3

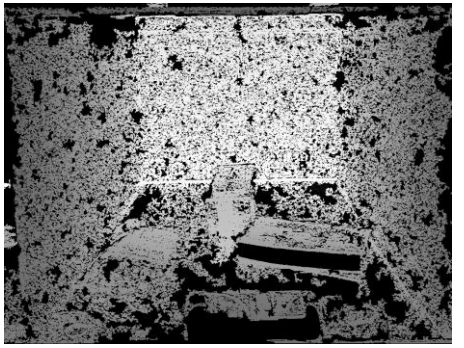
Depth Fusion

RGB frame



Upper depth

Lower depth



Stereo depth

$$\min_{u,v} \left[\alpha_1 \int_{\Omega_H} |T^{\frac{1}{2}}(\Delta u - v)| \, \mathrm{d}x + \alpha_0 \int_{\Omega_H} |\Delta v| \, \mathrm{d}x + \int_{\Omega_H} \mathbf{w} |(u - D_s)|^2 \, \mathrm{d}x \right]$$

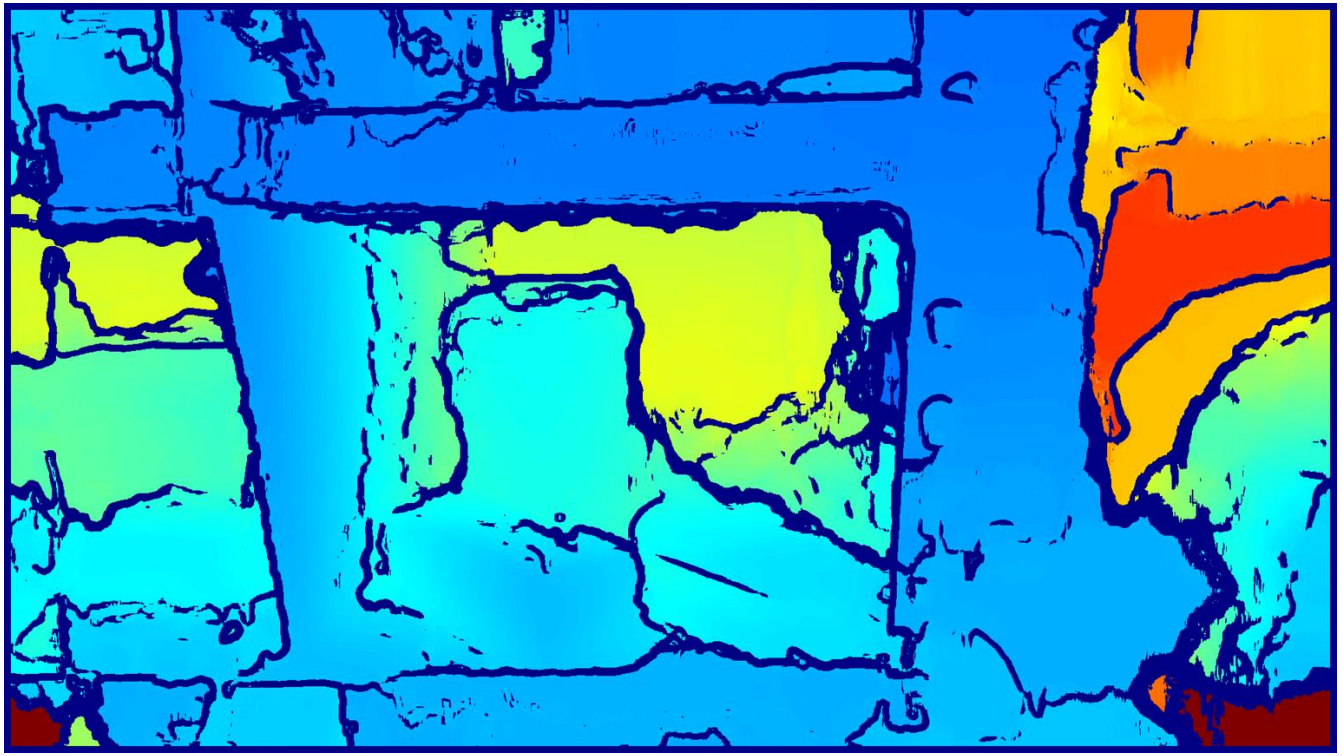
Image Guided Upsampling
(Ferstl at al. 2013)

Fused result

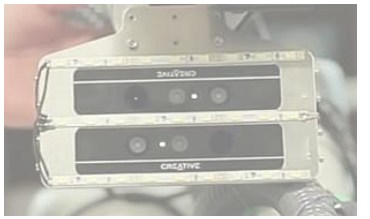
RGB


Glue-All
RICH. PURE TASTE
Folgers
CLASSIC ROAST
EXPO
WHITE BOARD CARE

Depth

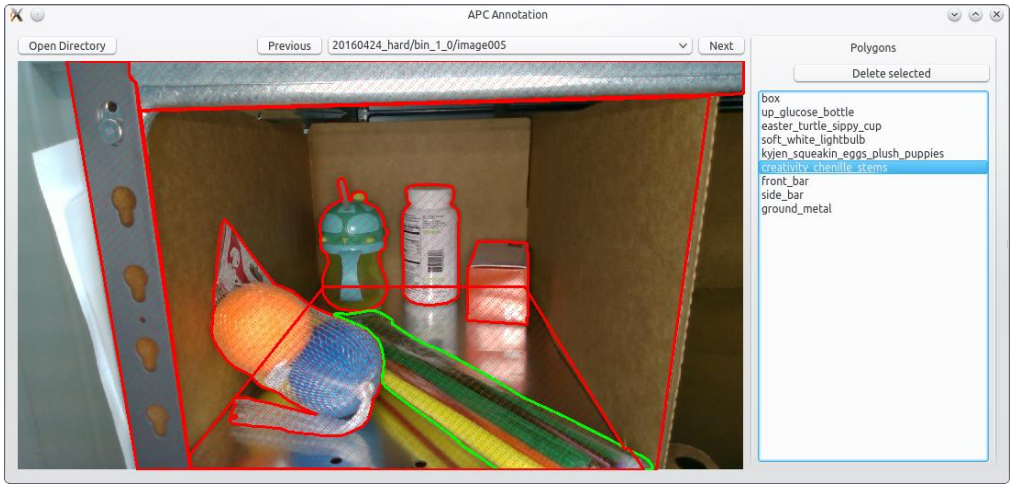


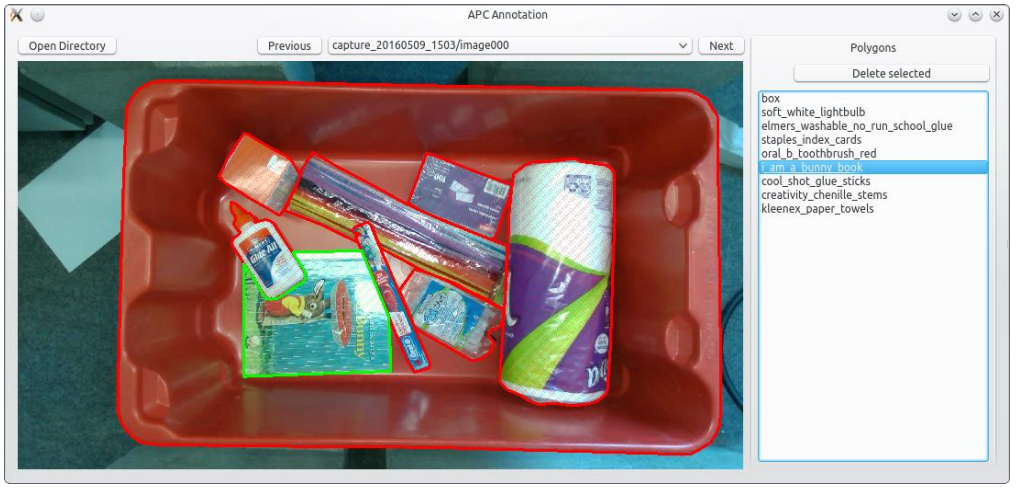
Data Acquisition & Annotation

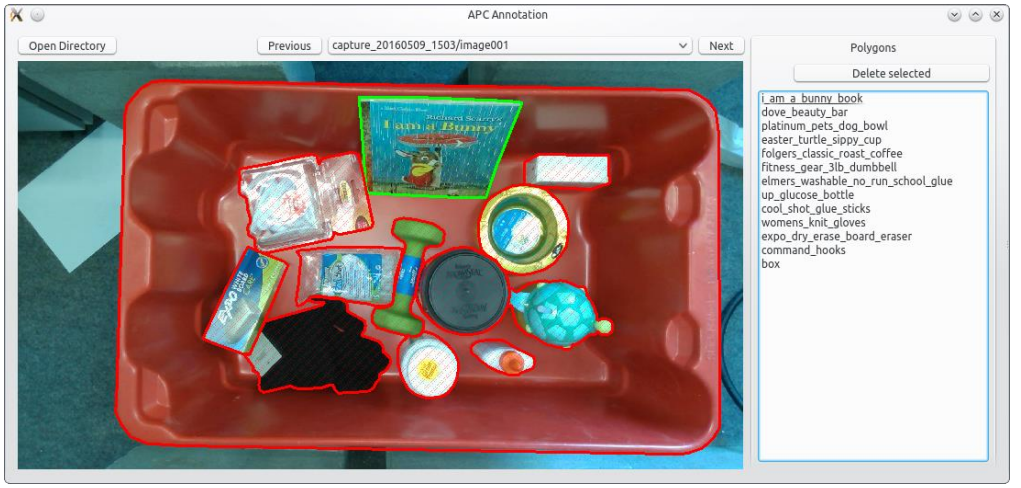


APC: ca. 100 images per setting (shelf/tote)

- → 10 images per object

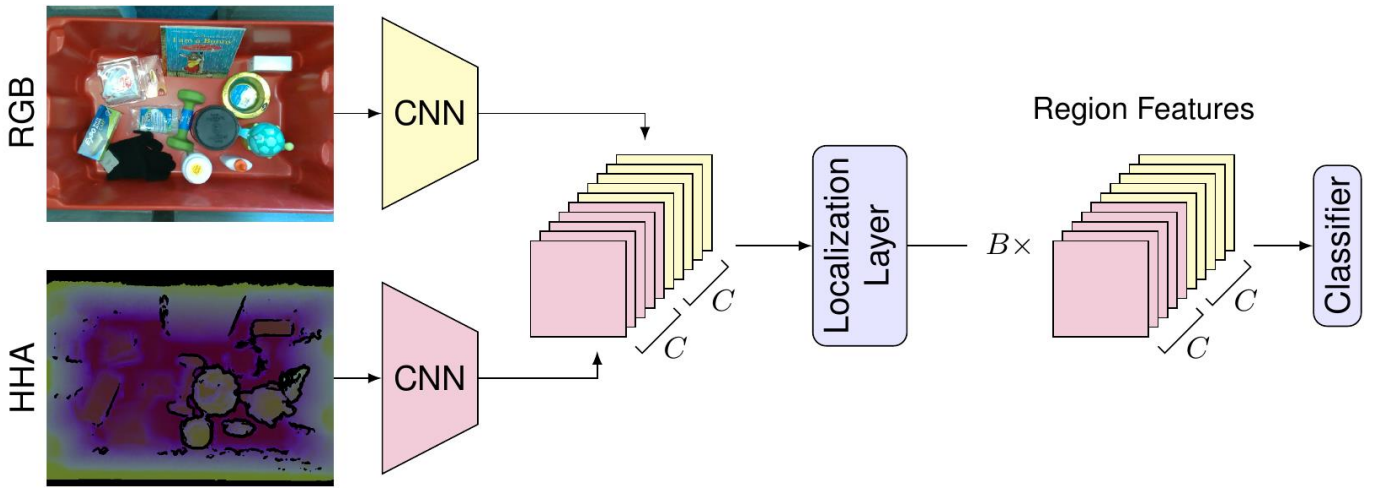




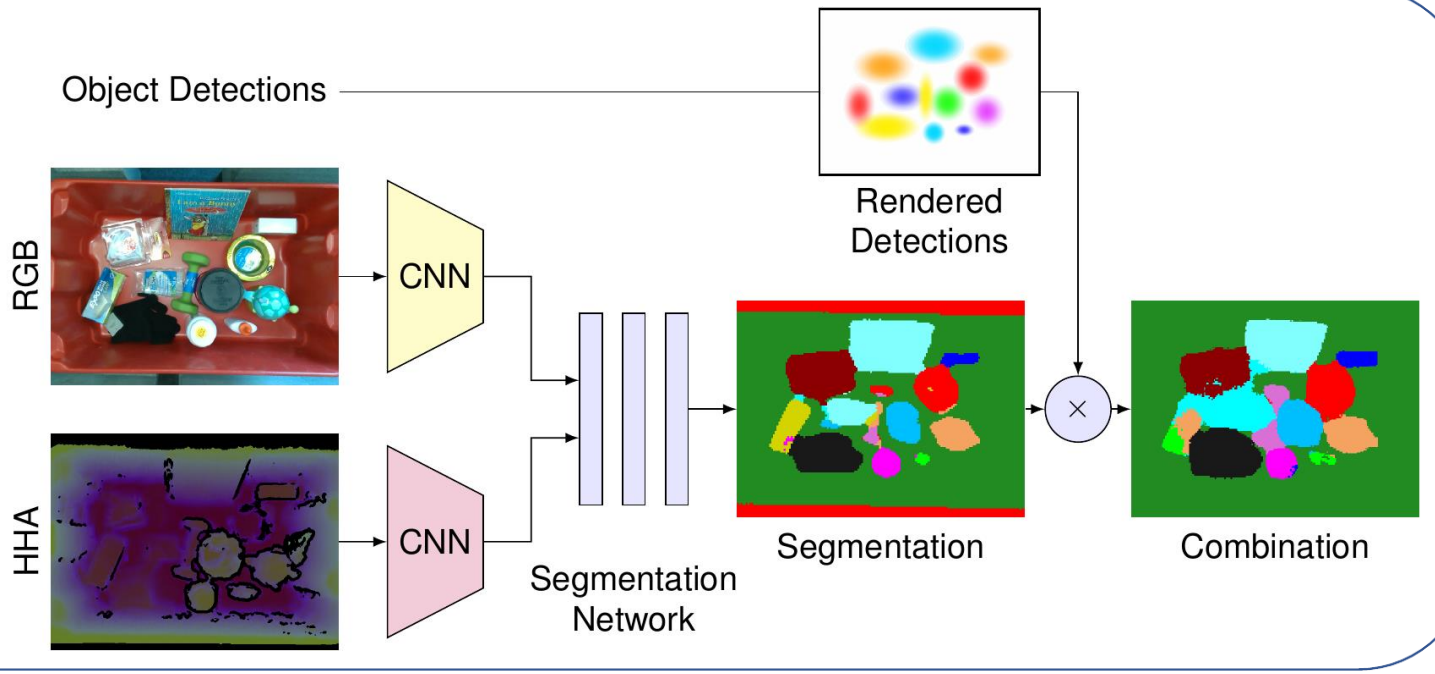


Deep Learning Architecture

1) Object Detection

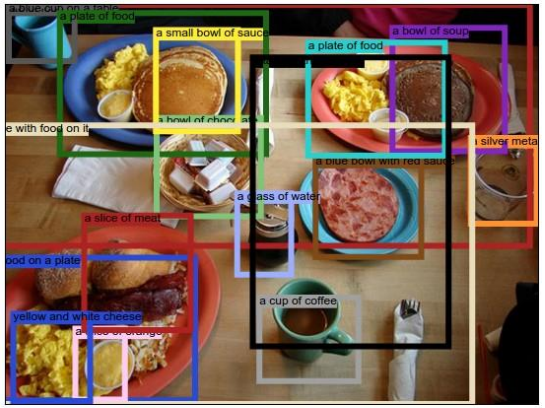


2) Semantic Segmentation



Object Detection

[Johnson et al., CVPR 2016]

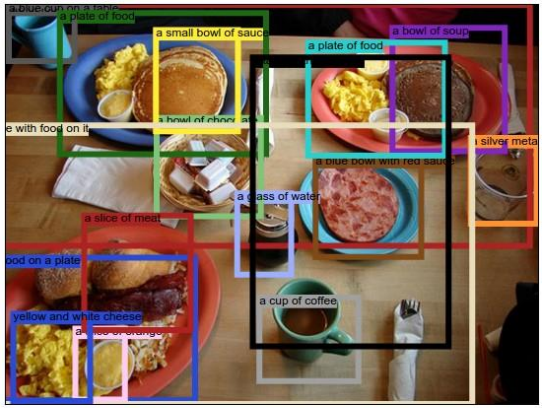


a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on it. a slice of meat. yellow and white cheese.

DenseCap

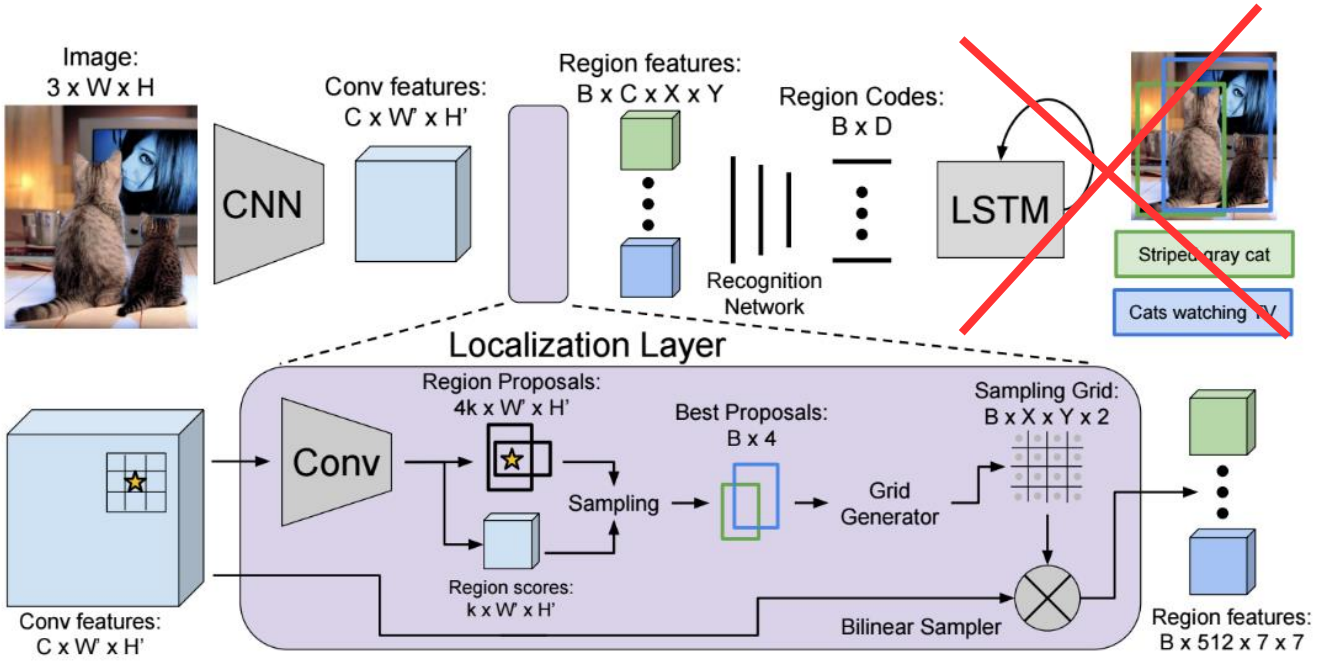
Object Detection

[Johnson et al., CVPR 2016]



a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on it. a slice of meat. yellow and white cheese.

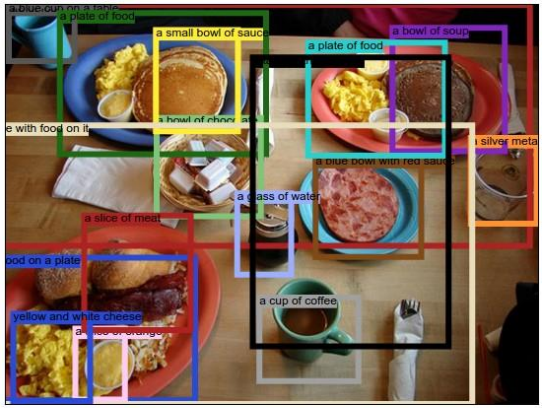
DenseCap



- End-to-end training
- External proposals
- Popular deep learning framework (Torch)
- Well documented
- Pretrained!

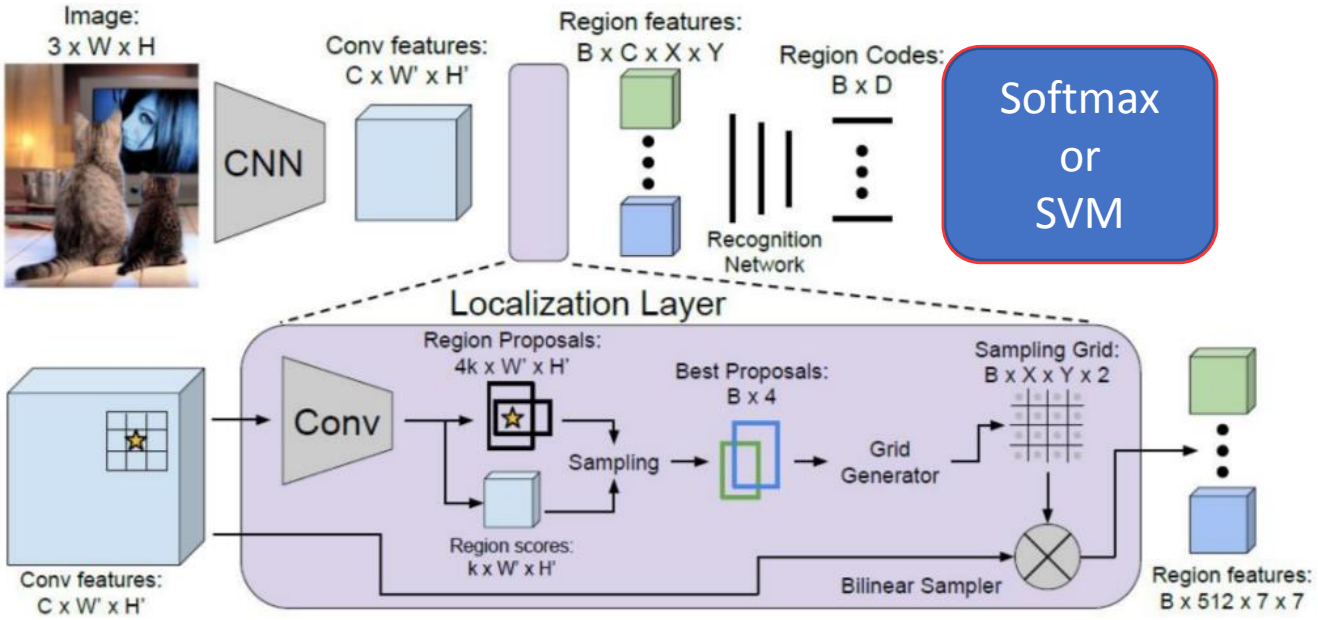
Object Detection

[Johnson et al., CVPR 2016]



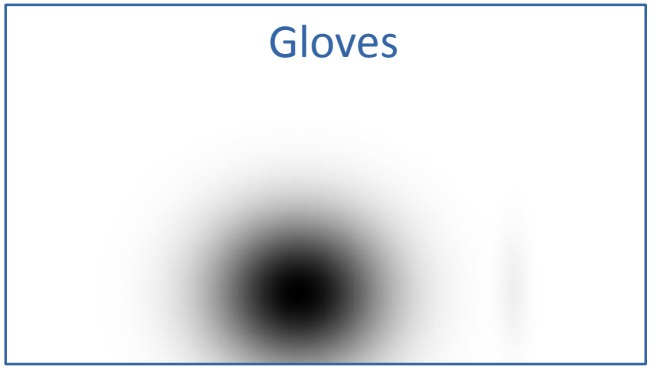
a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on it. a slice of meat. yellow and white cheese.

DenseCap



Replace Text Generation with Softmax Classifier or Online SVM Training

Deep Features and Online Learning



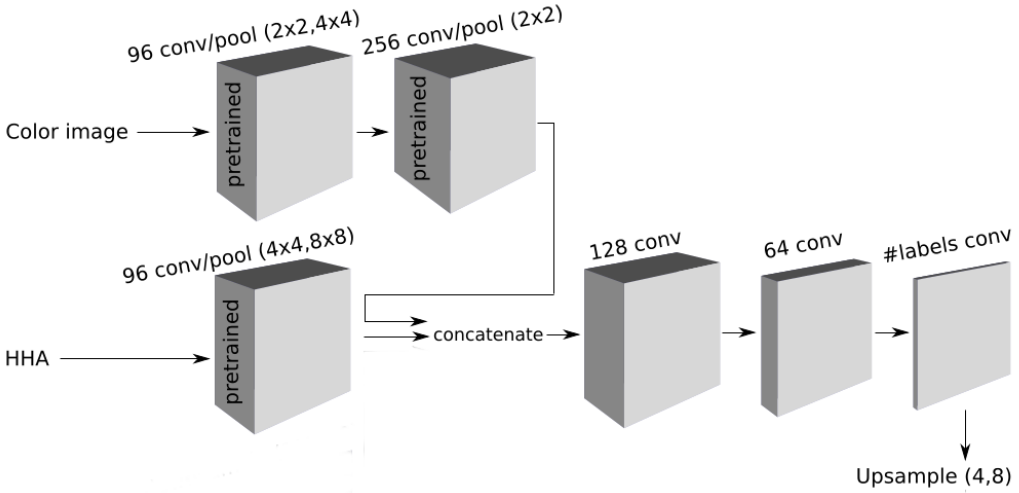
Evaluation on APC dataset

Table 6.1: Evaluation of object detection architectures on the shelf dataset. The mAP score is reported for the uninformed and informed case.

		mAP		
Input	Variant	Uninf.	Inf.	F1
RGB	SVM (plain)	–	28.83	68.50
RGB	SVM (tailor)	–	28.87	68.35
RGB	Softmax (no augmentation)	86.04	88.97	76.88
RGB	Softmax (with augmentation)	86.49	89.56	77.10
RGB-D (TGV)	HHA Features (Sec. 5.4.2)	86.53	89.81	77.58
RGB-D (TGV)	Ext. Proposals (Sec. 5.4.1)	87.01	89.84	77.46
RGB-D (TGV)	HHA CNN (Sec. 5.4.3)	86.47	90.12	78.98
RGB-D (TGV)	Distillation (Sec. 5.4.4)	**87.87**	**91.19**	**79.84**
RGB-D (single)[1]	Distillation (Sec. 5.4.4)	86.50	90.13	78.71
RGB-D (DT)[2]	Distillation (Sec. 5.4.4)	87.48	90.32	78.85

Semantic Segmentation

[Husain et al., RA-L 2016]



Fully Convolutional Neural Network

- Pre-trained OverFeat on ImageNet
- Fine-tuned (last 3 layers) on APC Data

Training:

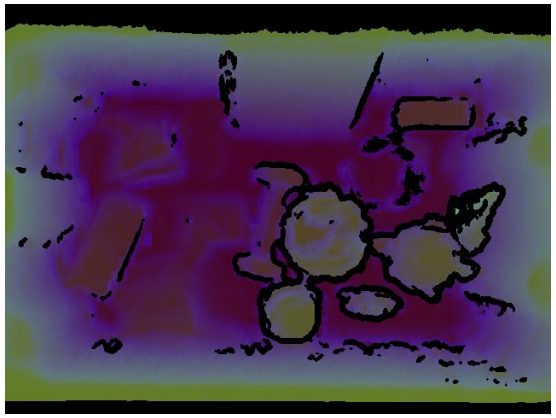
~ 3 hours on multiple GPUs

Testing:

~ 200 ms per image

Semantic Segmentation

RGB



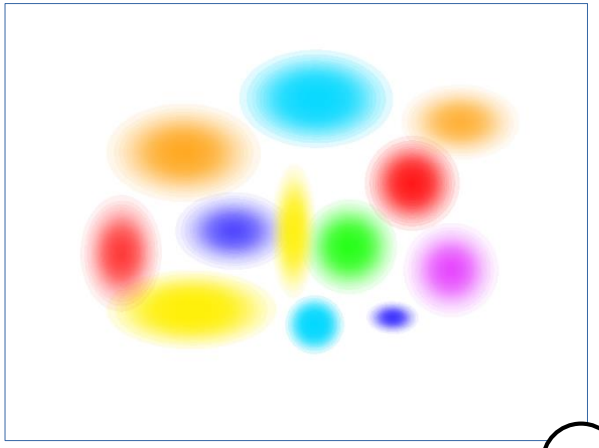
HHA

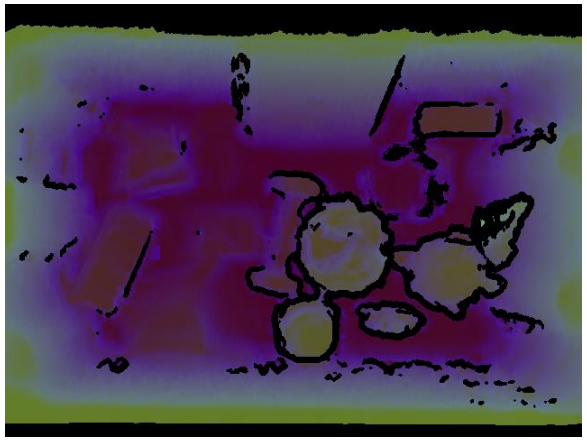
Table 6.2: Final results on the APC dataset.

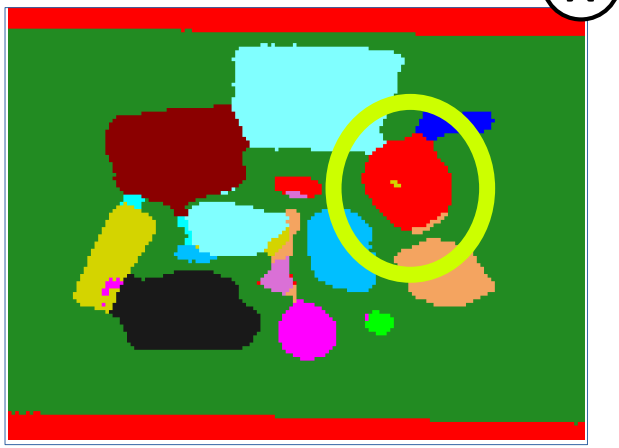
	mAP		F1
Dataset	Uninformed	Informed	
Shelf	87.87	91.19	79.84
Tote	87.00	88.65	77.90

Result

Combined Detection and Segmentation







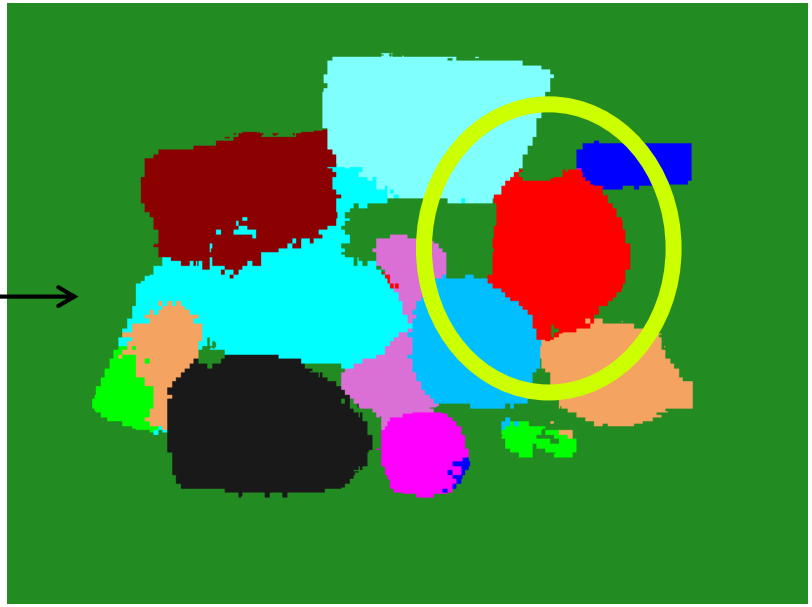
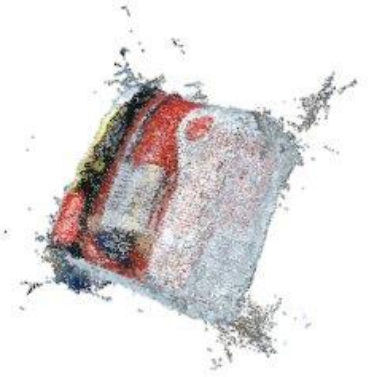


Table 6.5: F1 scores for semantic segmentation.

	Shelf		Tote	
Method	Uninf.	Inf.	Uninf.	Inf.
Segmentation	0.757	0.787	0.789	0.816
Combination[1]	0.782	0.813	0.823	0.833

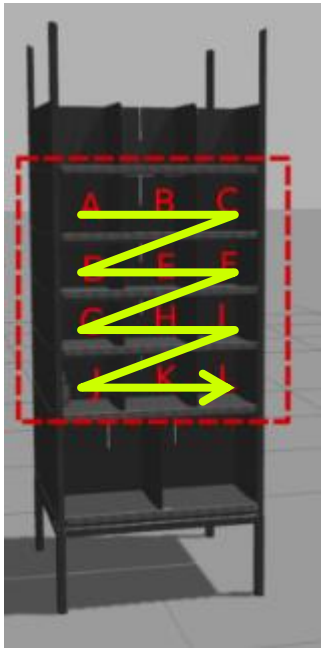
6D Pose Estimation

- Capture item on turn table
- Build 3D model
- Generate proposals
- Register to test image



V4R Library [Aldoma et al., ICRA 2013]

Pick / Stow Strategy



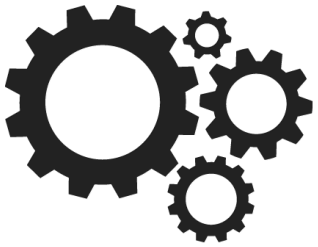
Pick:

- order A … L
- On failure, retry at end
- Drop at 3 predefined positions in tote

Stow:

- Try to put all items into one 20 points bin
  - (select the one with most free space)
- Stow “large” items into own bin
  - (coffee, socks, paper towels, tissue box, curtain, pencil cup, mailer)
- If leftover object at end, retry segmentation with all classes

Outline



System

Control

Perception

Special Features

Foldable Funnel

Tricky Items to Grasp

Heavy / cylindrical

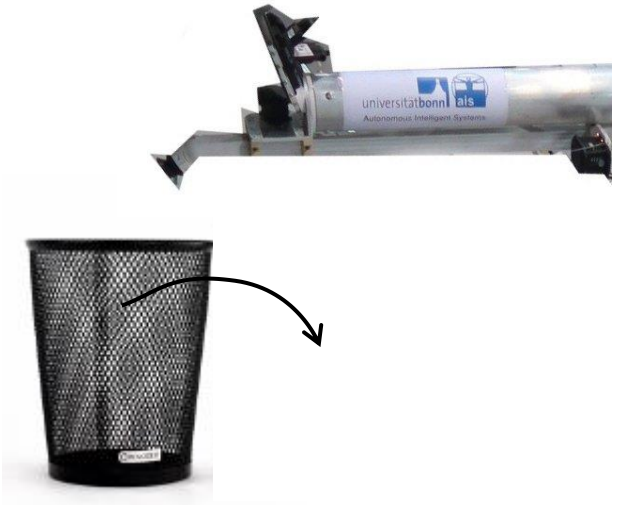
→ Ensure that grasp is on **center of mass**!

?

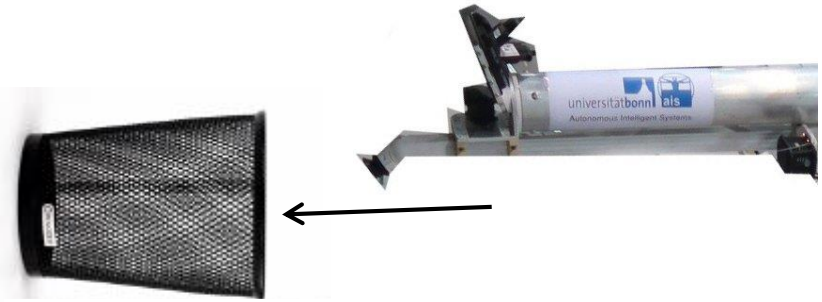
Hard to suck

→ Grasp on one ball

Sucking the Pencil Cup



1. Knock over



2. Suck on bottom

Sucking the Pencil Cup

1.5x

What happened at APC 2016?

Table 7.2: Picking Run at APC 2016

Bin	Item	Pick	Drop	Report
A	duct tape	×	×	×
B	bunny book	✓	✓	×[2]
C	squeaky eggs	✓	×	✓
D	crayons[1]	✓	×	✓
E	coffee	✓	✓	×[2]
F	hooks	✓	×	✓
G	scissors	×	×	×
H	plush bear	✓	×	✓
I	curtain	✓	×	✓
J	tissue box	✓	×	✓
K	sippy cup	✓	×	✓
L	pencil cup	✓	✓	×[2]
	Sum	10	3	7

[1] Misrecognized, corrected on second attempt.
[2] Incorrect report, resulting in penalty.

Stowing:

- 11/12 items stowed ☺
- 1 misrecognition ☹

Picking:

- 10 correct items picked ☺
- 3 items dropped ☹
- 3 wrong reports ☹

Summary

Stow: 2nd place

DELFT	214
NimbRo	**186**
MIT	164

Pick: 3rd place

DELFT	105
PFN	105
NimbRo	**97**

Do it as simple as possible,
but not simpler!

More Information

Max Schwarz, Anton Milan, Christian Lenz, Aura Munoz, Arul Selvam Periyasamy, Michael Schreiber, Sebastian Schüller, and Sven Behnke:
NimbRo Picking: Versatile Part Handling for Warehouse Automation
IEEE International Conference on Robotics and Automation (ICRA) 2017
Talk: Wednesday 9:30, room 4813/4913.

Max Schwarz, Anton Milan, Arul Selvam Periyasamy, and Sven Behnke:
RGB-D Object Detection and Semantic Segmentation
for Autonomous Manipulation in Clutter
International Journal of Robotics Research (IJRR), Sage Publications, to appear 2017.

Source code release: https://github.com/amazon-picking-challenge/nimbro_picking

Thank you

All team members (left to right):

Anton Milan, Michael Schreiber, Sebastian Schüller, Max Schwarz, Arul Selvam Periyasamy, Christian Lenz, Sven Behnke, Aura Muñoz.