

# Amazon Picking Challenge 2016: Team NimbRo of University of Bonn



Max Schwarz

Aura Muñoz

Sebastian Schüller

Michael Schreiber

Christian Lenz

Anton Milan

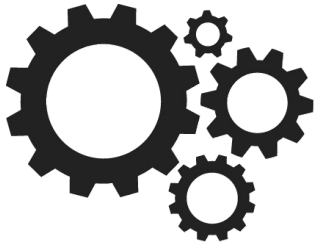
Arul Selvam Periyasamy

Sven Behnke



Computer Science Institute VI – Autonomous Intelligent Systems

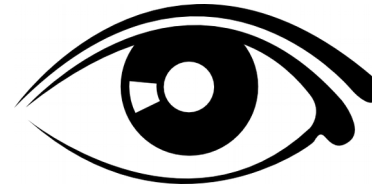
# Outline



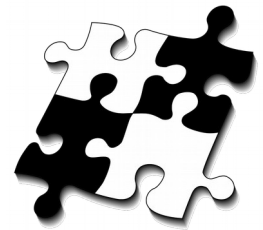
System



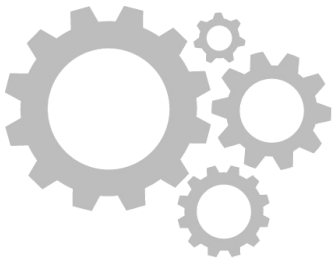
Control



Perception



Special  
Features



# System Overview

UR 10 Arm (6 DOF)

Air velocity sensor

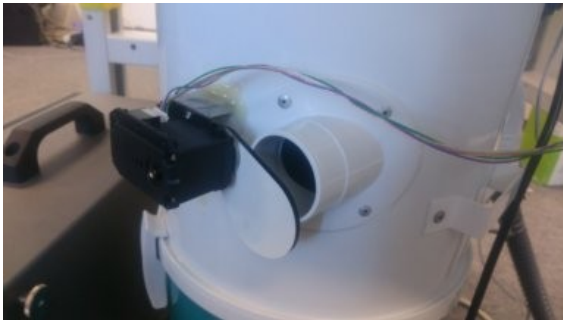
2x Intel RealSense SR-300  
+ LED light

Linear actuator

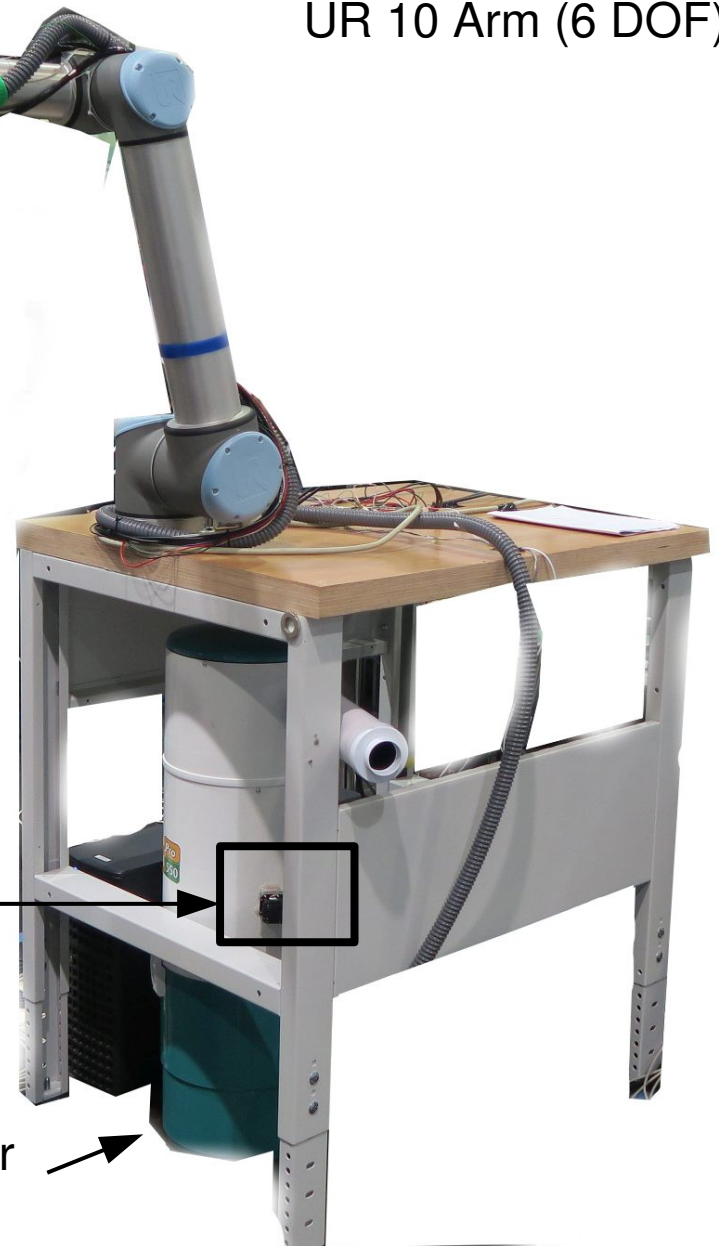
Foldable  
suction finger

Total:  
6 + 2 DOF

Suction strength control



Central vacuum cleaner  
(3100 W)





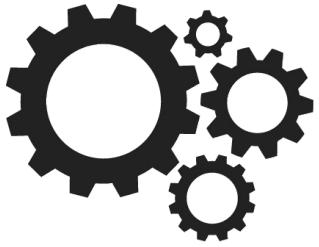
# Concept and Design

- 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!**



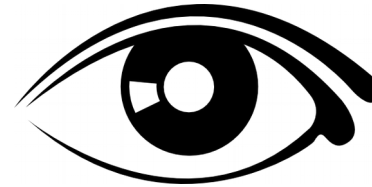
# Outline



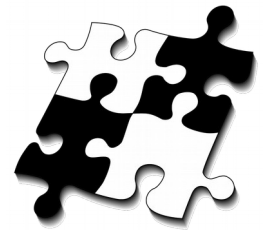
System



Control




Perception



Special  
Features



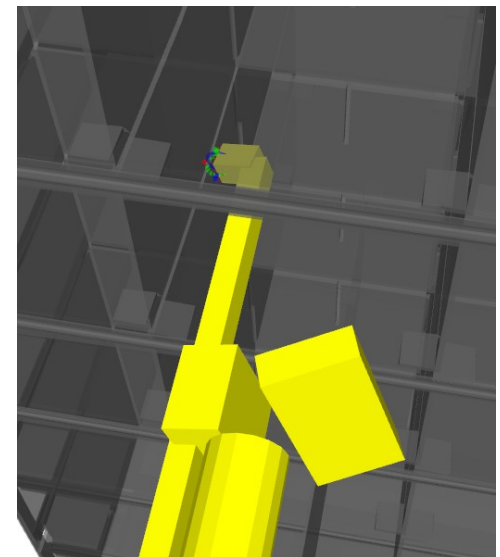
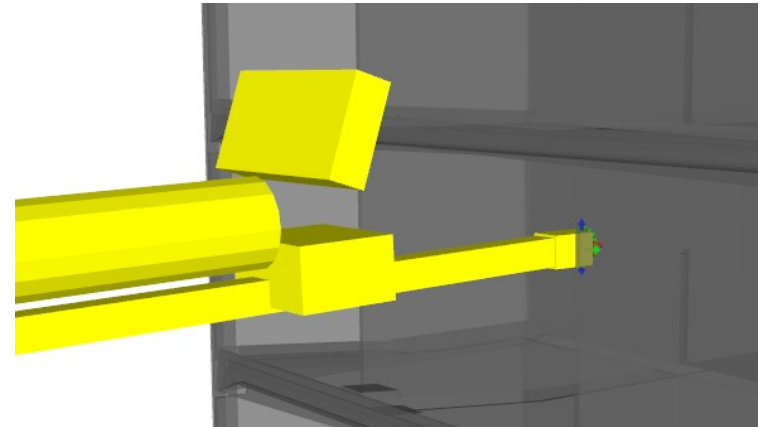
# Motion Generation

- 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 suction it
- Self-collision detection using  MoveIt!
- Avoid collisions with shelf in IK solver!

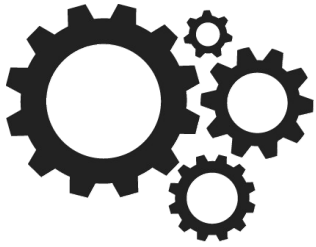


# Inverse Kinematics

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



# Outline



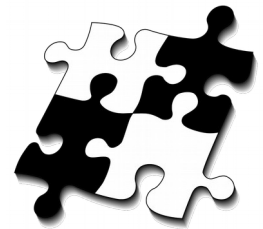
System



Control



Perception



Special  
Features





# Sensors

2x Intel RealSense SR-300

3 Depth measurements per pixel

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

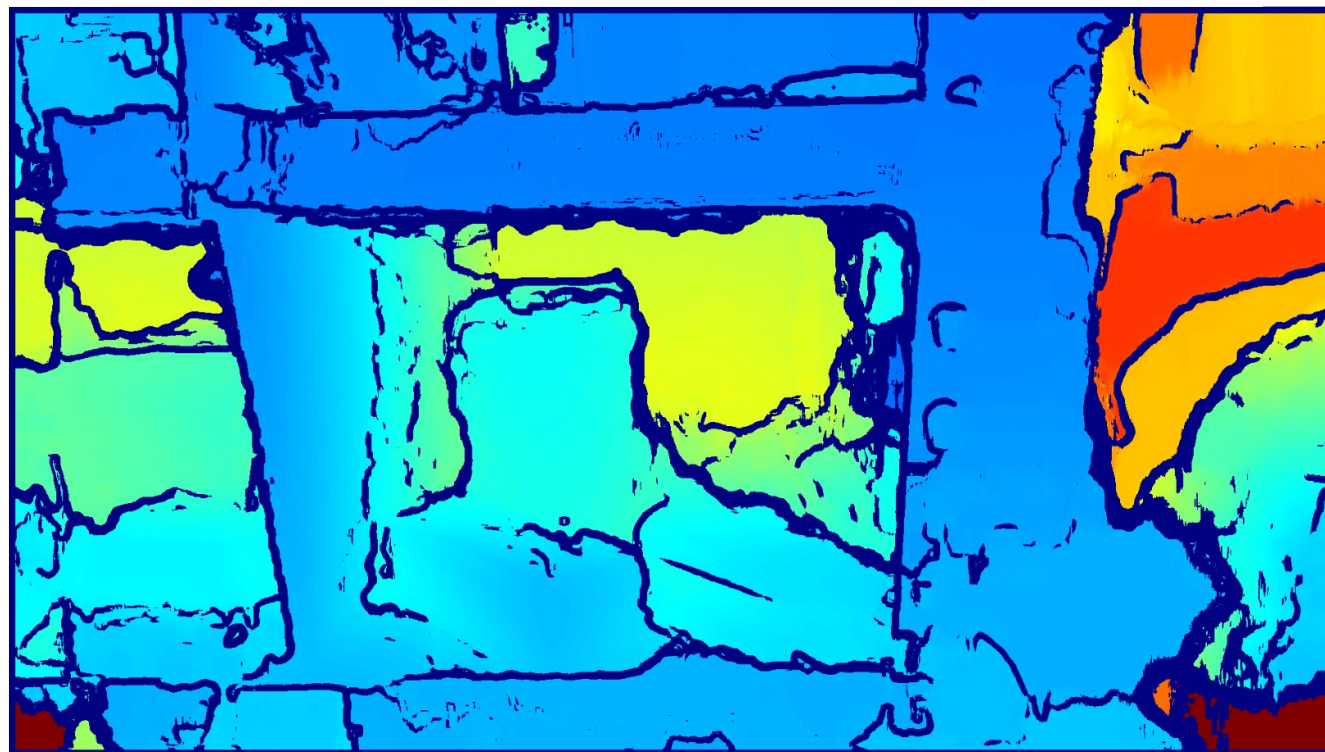
Fusion: 2 out of 3



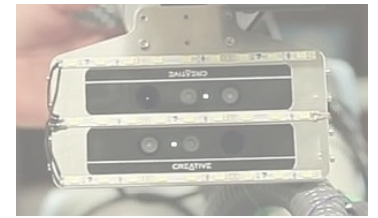
RGB



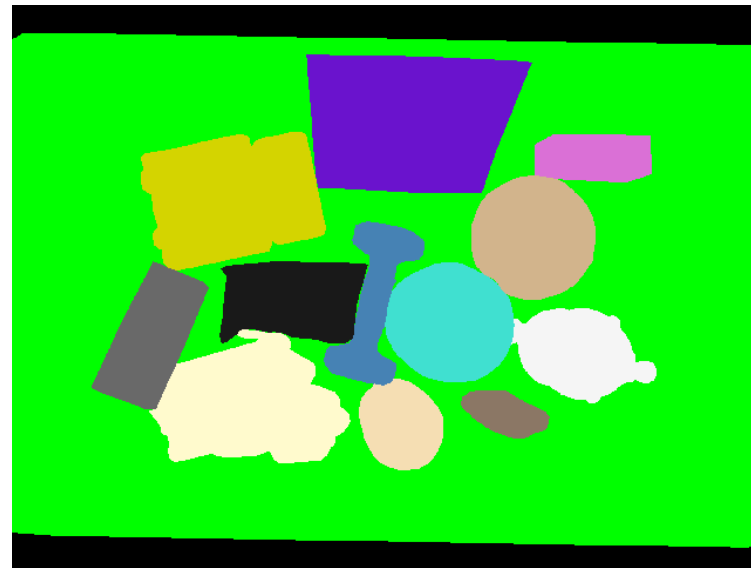
Depth



# Data Acquisition



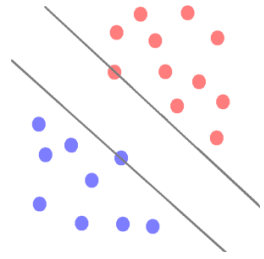
- ca. 100 images per setting (shelf/tote)
  - → 10 images per object
- Manual annotation



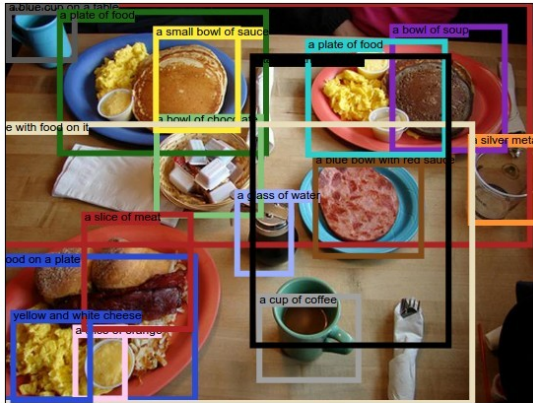




# Deep Features and Online Learning



[Johnson et al., CVPR 2016]

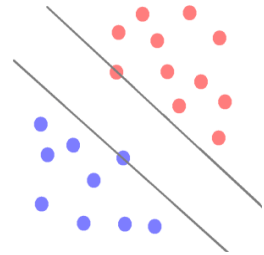


a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on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

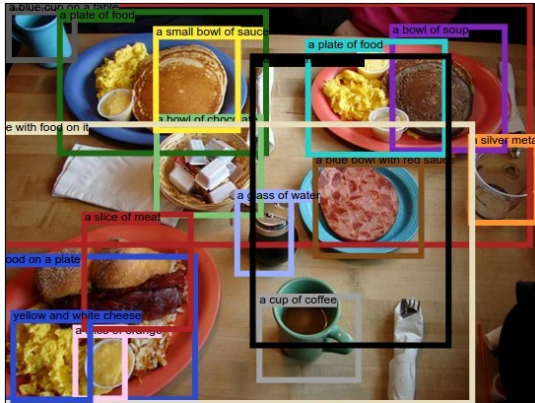
DenseCap



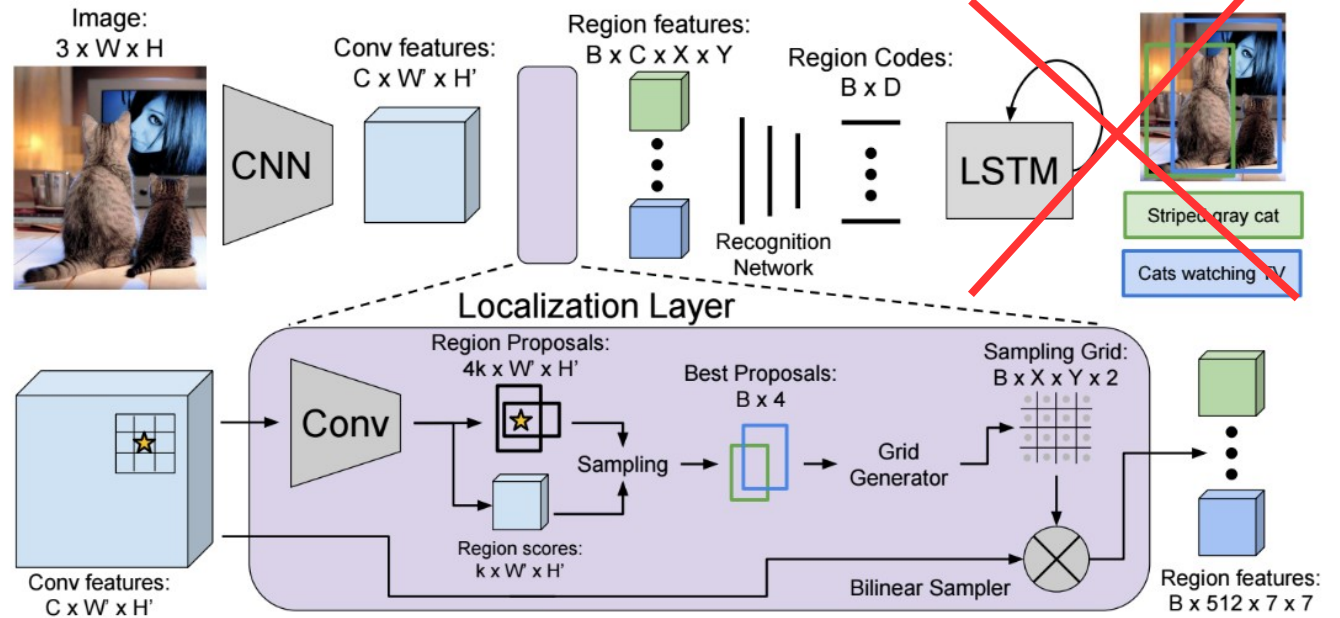
# Deep Features and Online Learning



[Johnson et al., CVPR 2016]



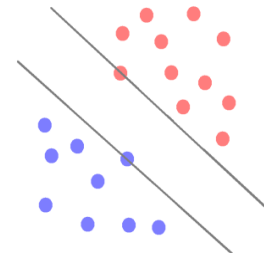
a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on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



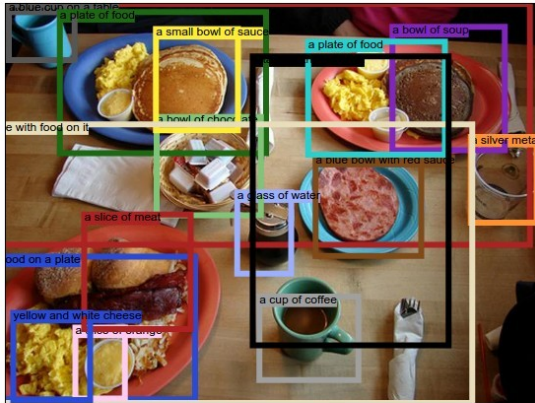
DenseCap



# Deep Features and Online Learning

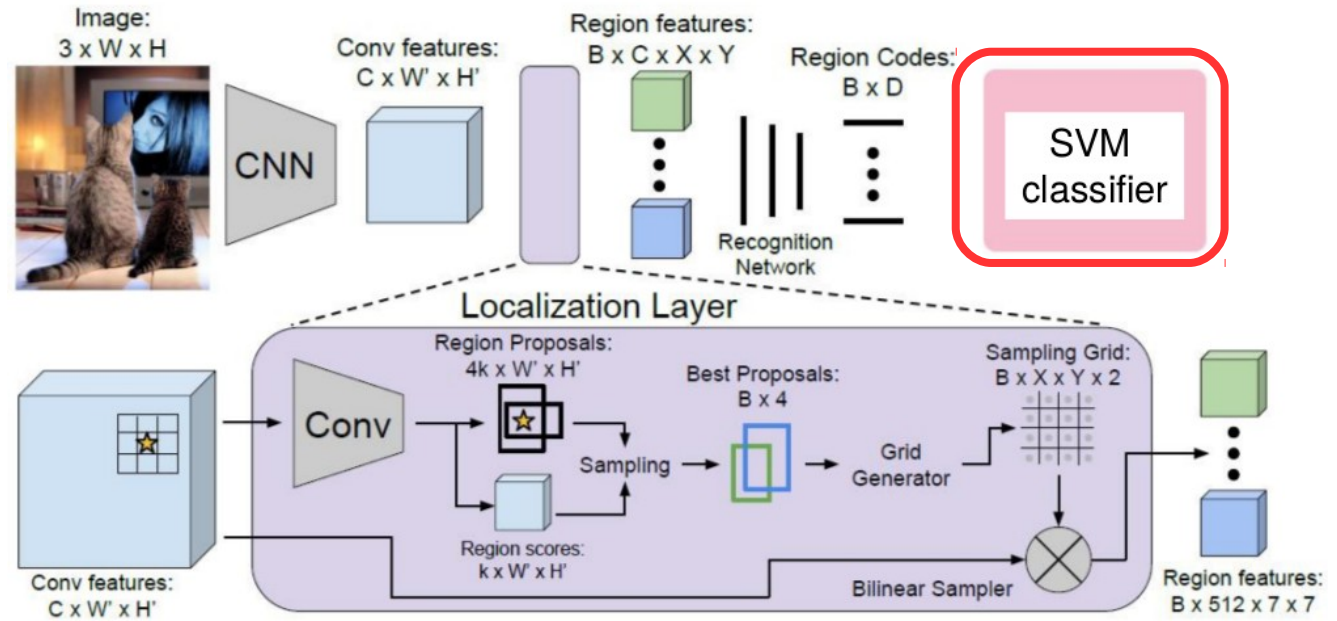


[Johnson et al., CVPR 2016]



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.

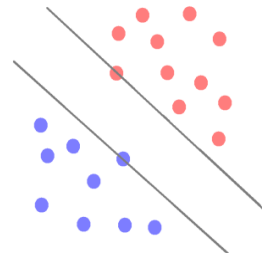
DenseCap



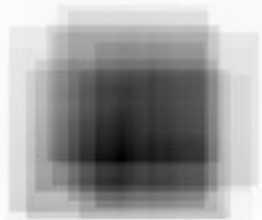
Replace Text Generation with Online SVM Training



# Deep Features and Online Learning



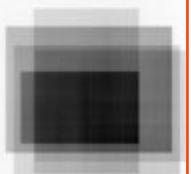
Gloves



Glue sticks



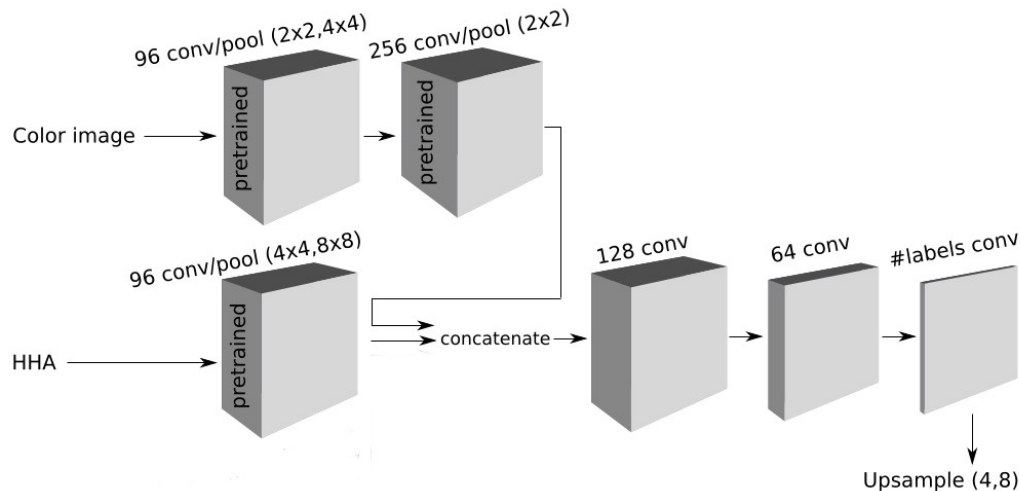
Sippy cup





# 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

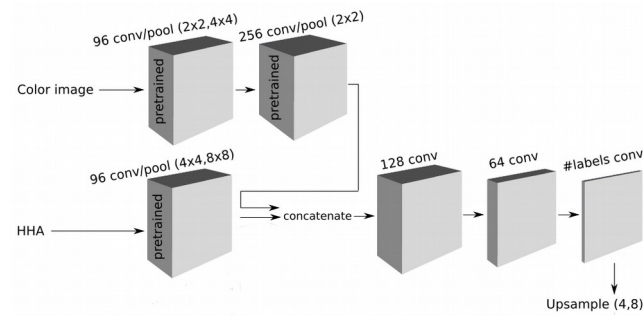
## Testing:

~ 200 ms per image

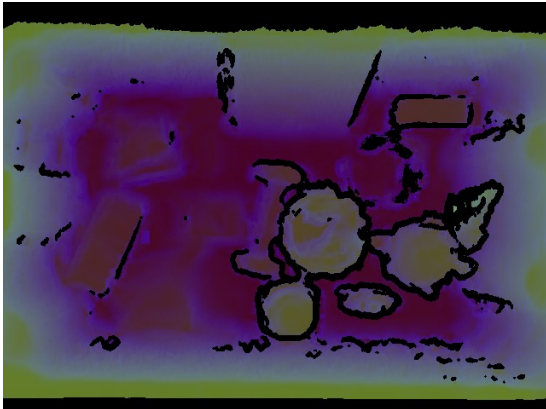


# Semantic Segmentation

RGB



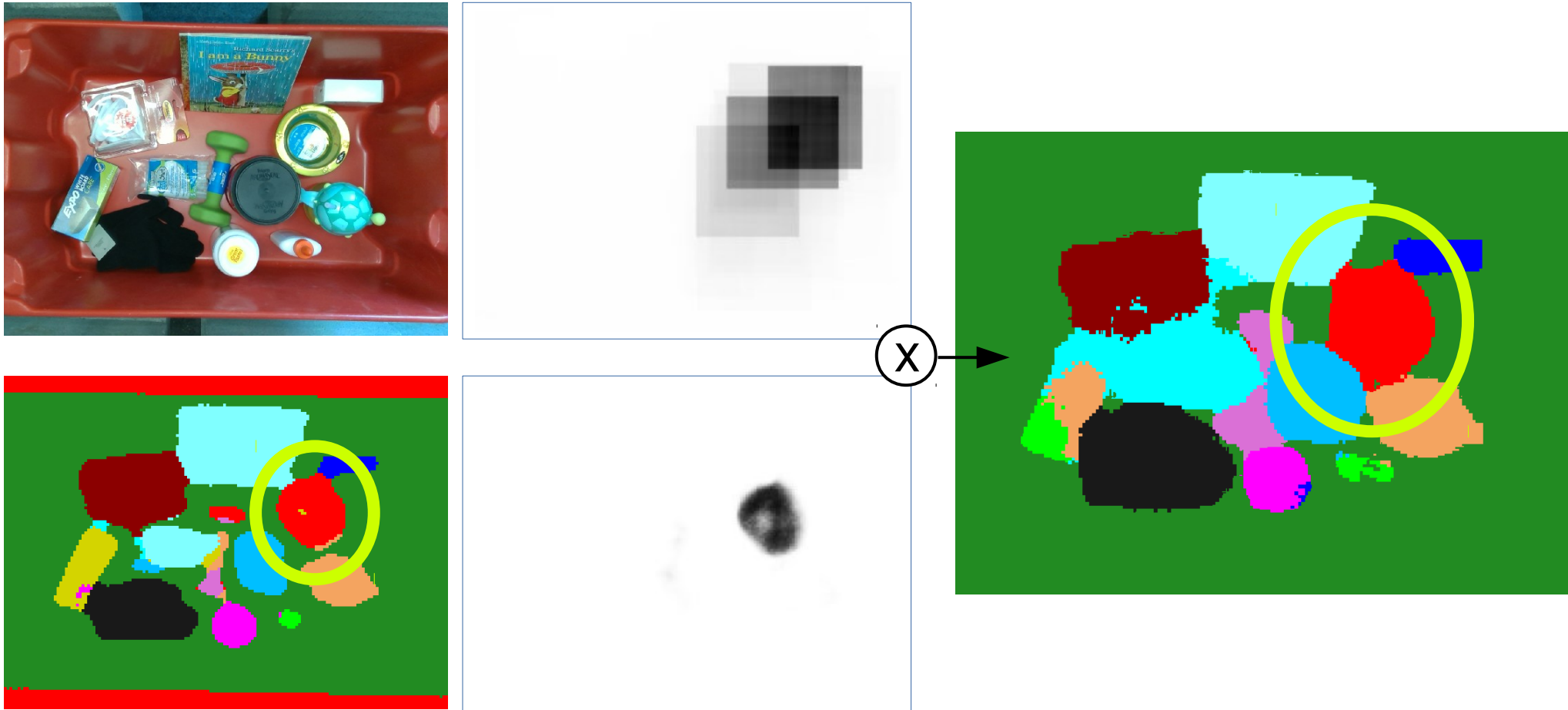
Result



HHA



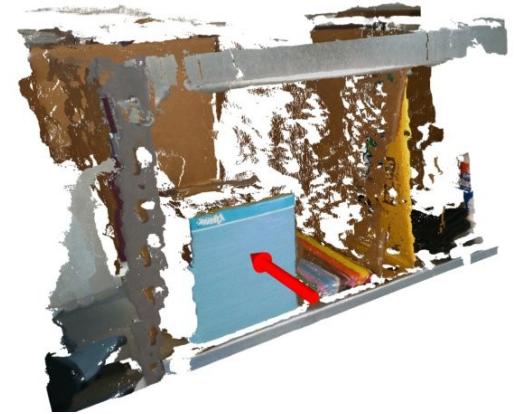
# Combined Detection and Segmentation





# 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)





# 6D Pose Estimation



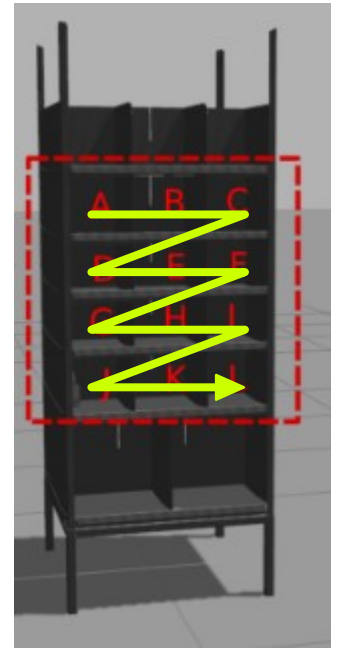
- Capture item on turn table
- Build 3D model
- Generate proposals
- Register to test image



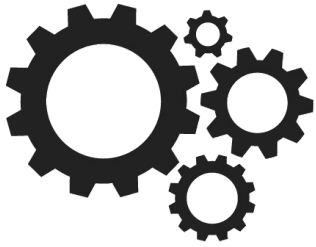


# Pick / Stow Strategy

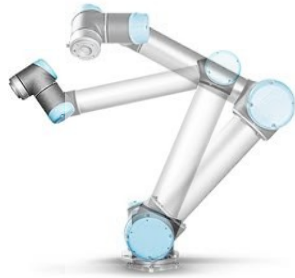
- Pick:
  - order A ... L
  - On failure, retry at end
  - Drop at 3 predefined positions in tote
- Stow:
  - Try to put all items into one 20 points bin
    - (select the one with most free space)
  - Stow “large” items into own bin
    - (coffee, socks, paper towels, tissue box, curtain, pencil cup, mailer)
  - If leftover object at end, retry segmentation with all classes



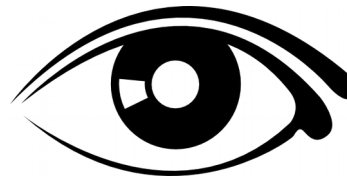
# Outline



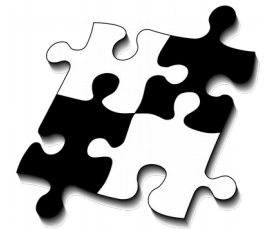
System



Control



Perception



Special  
Features



# Foldable Funnel







# Foldable Funnel







# Tricky Items to Grasp



Heavy / cylindrical

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

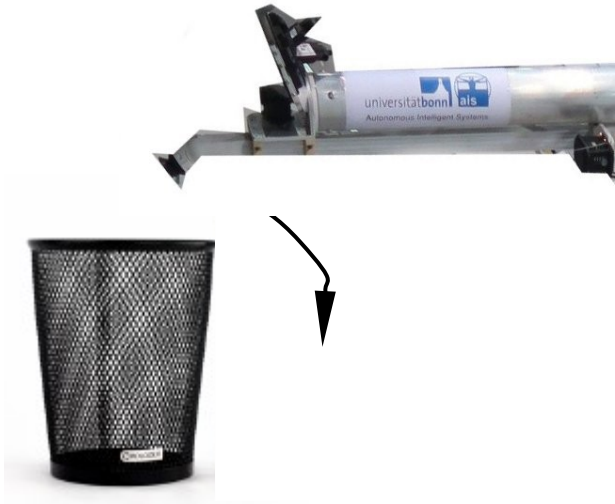


Hard to suck

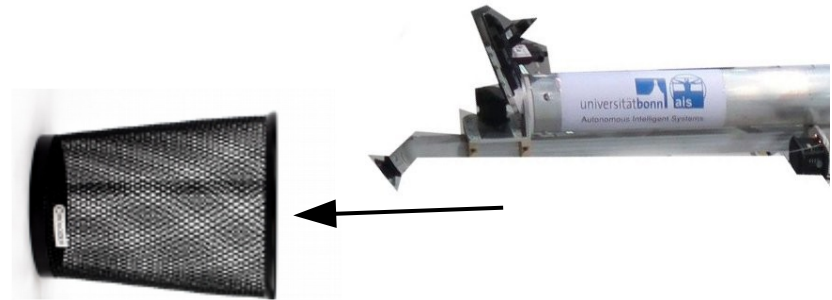
→ Grasp on one ball



# Sucking the Pencil Cup



1. Knock over



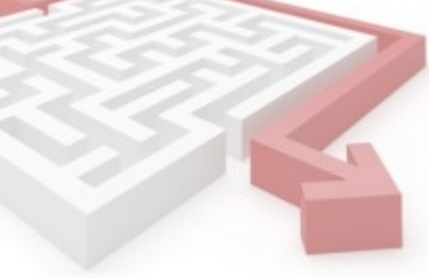
2. Suck on bottom



# Sucking the Pencil Cup

1.5x





# Summary

Stow: 2<sup>nd</sup> place

DELFT	214
<b>NimbRo</b>	<b>186</b>
MIT	164

Pick: 3<sup>rd</sup> place

DELFT	105
PFN	105
<b>NimbRo</b>	<b>97</b>

**Do it as simple as possible,  
but not simpler!**

# Thank you



Max Schwarz

Aura Muñoz

Sebastian Schüller

Michael Schreiber

Christian Lenz

Anton Milan

Arul Selvam Periyasamy

Sven Behnke