Towards Structured Model Learning, Perception and Planning for Cognitive Robots

Sven Behnke

University of Bonn
Computer Science Institute VI
Autonomous Intelligent Systems

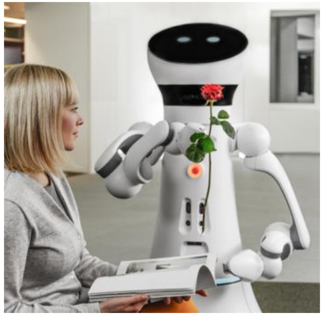
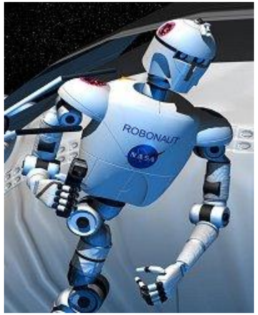
Many New Application Areas for Robots

- Self-driving cars
- Logistics
- Agriculture, mining
- Collaborative production
- Personal assistance
- Space, search & rescue
- Healthcare
- Toys

Need more cognitive abilities!






Some of our Cognitive Robots

- Equipped with numerous sensors and actuators
- Complex demonstration scenarios

Soccer

Domestic service

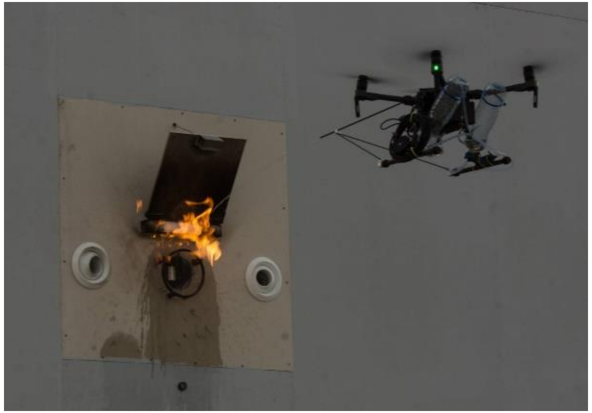
Mobile manipulation

Bin picking

Aerial inspection

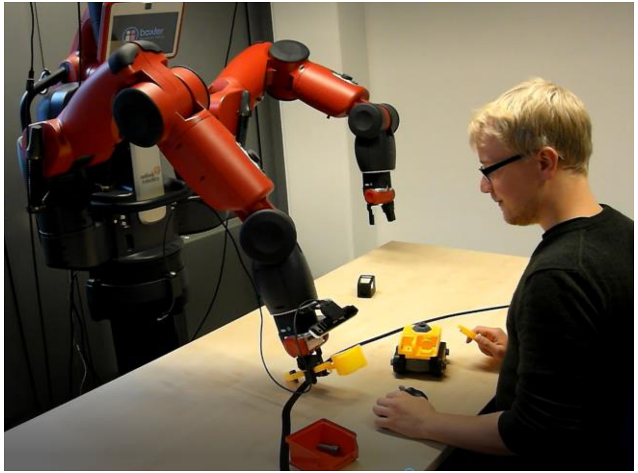
Some more of our Cognitive Robots

- Equipped with numerous sensors and actuators
- Complex demonstration scenarios

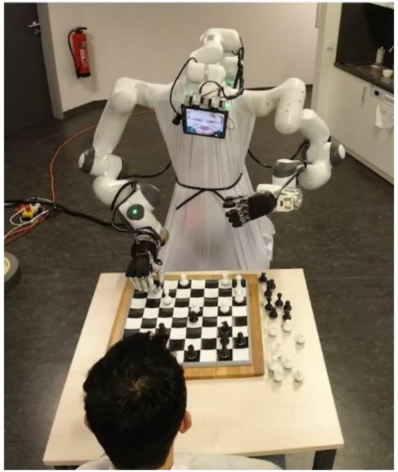


Rescue

Phenotyping



Human-robot collaboration



Telepresence

Computer Vision

3D Scene

2D Image

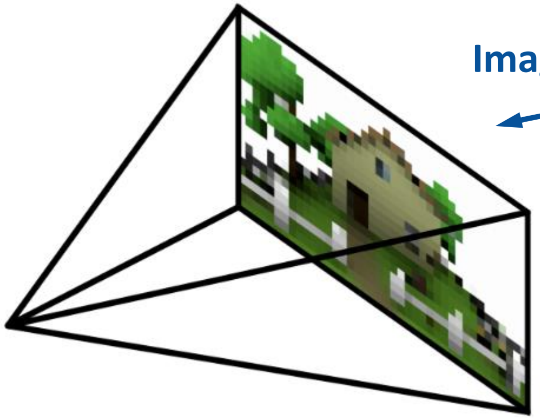


Image Capture / Rendering

Computer Vision

Pixel Matrix

217	191	252	255	239
102	80	200	146	138
159	94	91	121	138
179	106	136	85	41
115	129	83	112	67
94	114	105	111	89

- Objects, surfaces
- Geometry, 3D pose, shape
- Appearance, material properties
- Semantics

- Computer Vision is an **ill-posed inverse problem**:
  - Many 3D scenes yield the same 2D image

    => Additional constraints (knowledge about world) required

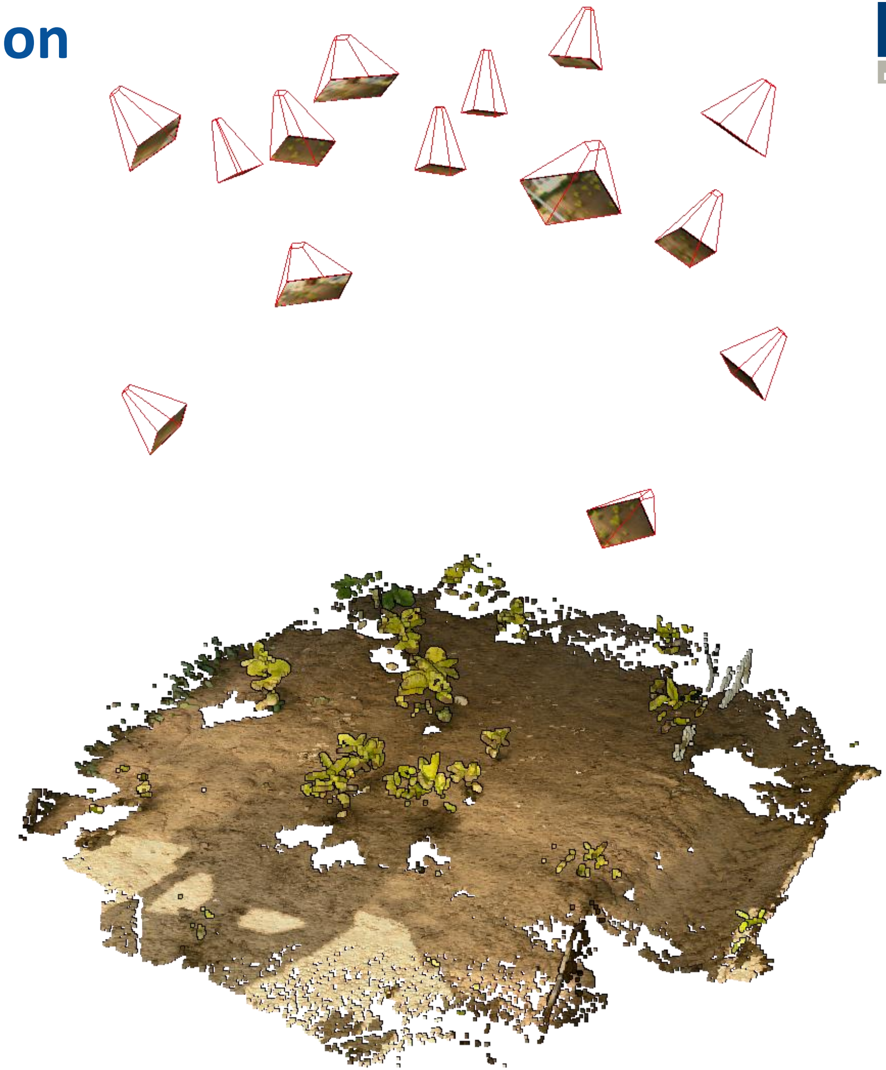
[Geiger]

Multi-view Plant Reconstruction

- 14× Nikon Z7 DSLR camera
  - 45 MP
  - 64–25600 ISO
  - 24-70 mm Lens

Multi-view Plant Reconstruction

- Recovered camera poses and semi-dense point cloud through Multi-View-Stereo



[Rosu 2022]

Multi-view Plant Reconstruction

- Geometry represented as Signed Distance Field (SDF)
- Color represented as a direction-dependent color field
- Transform SDF into radiance [1] and train similar to NeRF

Geometry

Color at the zero level-set of the SDF

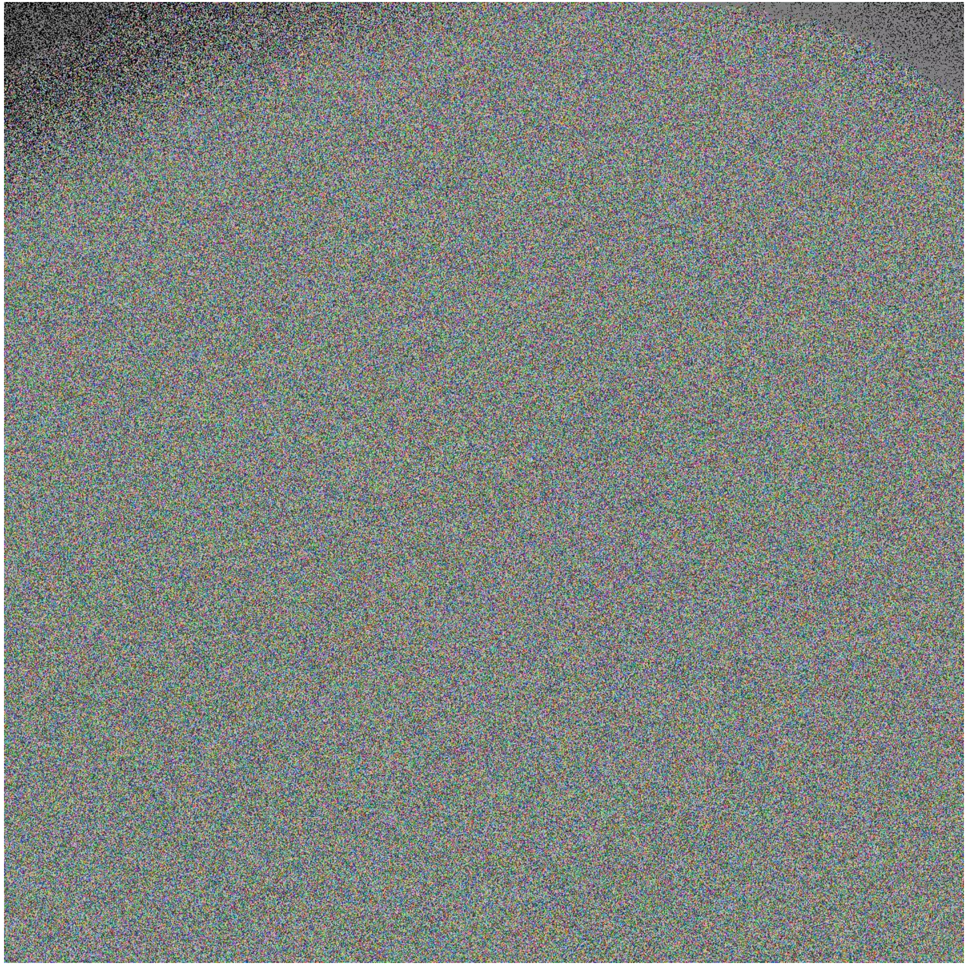
[Rosu 2022]

[1] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-View Reconstruction, NeurIPS 2021.

Multi-view Plant Reconstruction

- InstantNGP with a Multiresolution Hash Encoding [2]
- Small MLPs for SDF and color
- 25 M parameters
- 1 h training on Nvidia RTX 3090 GPU

[2] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding ACM Transactions on Graphics (SIGGRAPH 2022)



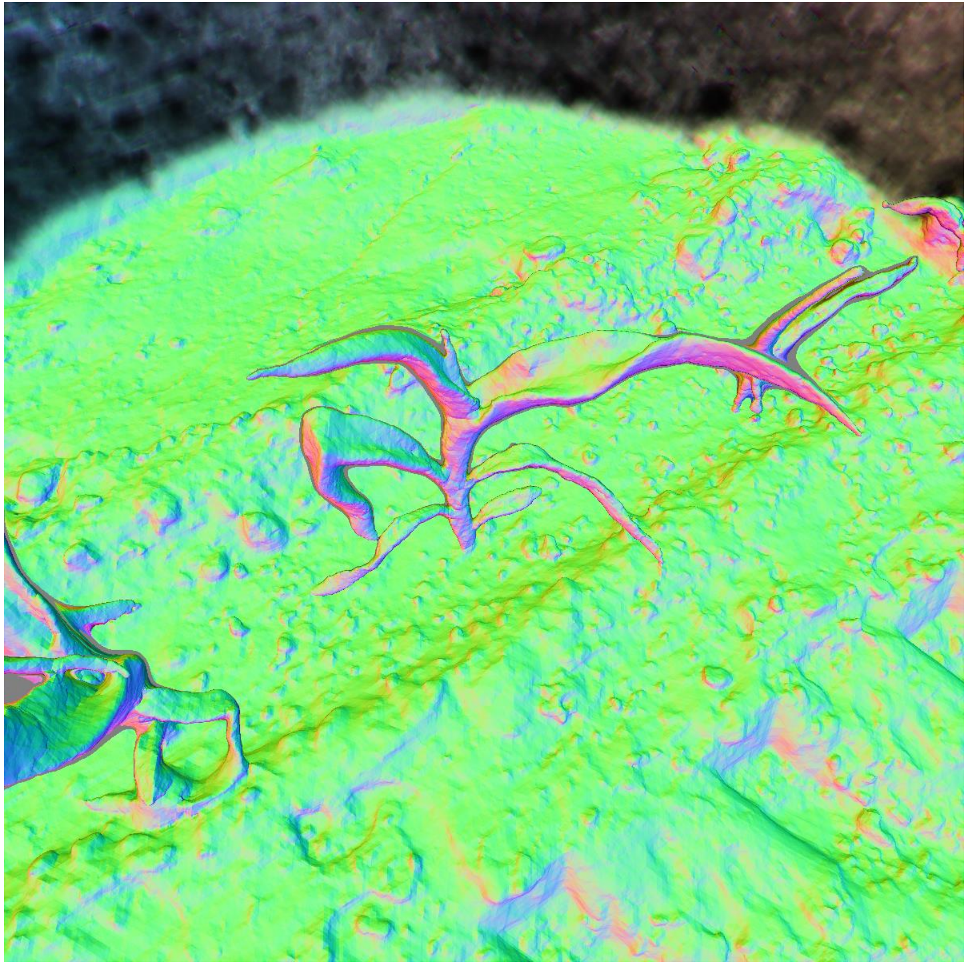
Surface normals

[Rosu 2022]

Multi-view Plant Reconstruction

- InstantNGP with a Multiresolution Hash Encoding [2]
- Small MLPs for SDF and color
- 25 M parameters
- 1 h training on Nvidia RTX 3090 GPU

[2] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding ACM Transactions on Graphics (SIGGRAPH 2022)



Surface normals

[Rosu 2022]

Multi-view Plant Reconstruction

- Rendered novel views

[Rosu 2022]

Plant Reconstruction over Multiple Days

Volumetric renders through
SDF + color

[Rosu 2022]

Plant Reconstruction over Multiple Days

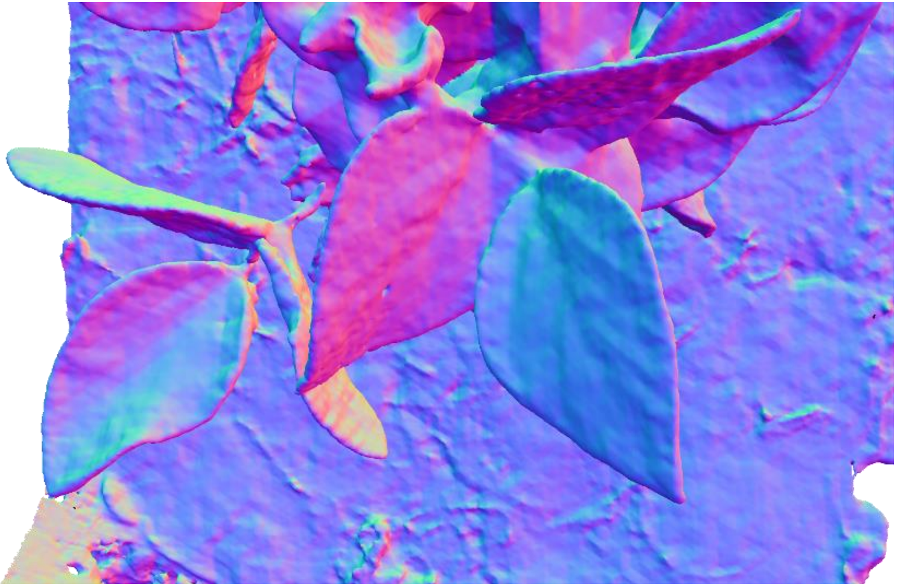
Predicted depth

[Rosu 2022]

High Geometric and Texture Detail

- Marching cubes on the SDF to recover mesh
- Learnable texture to match color images
- Rendering in real time

Textured mesh



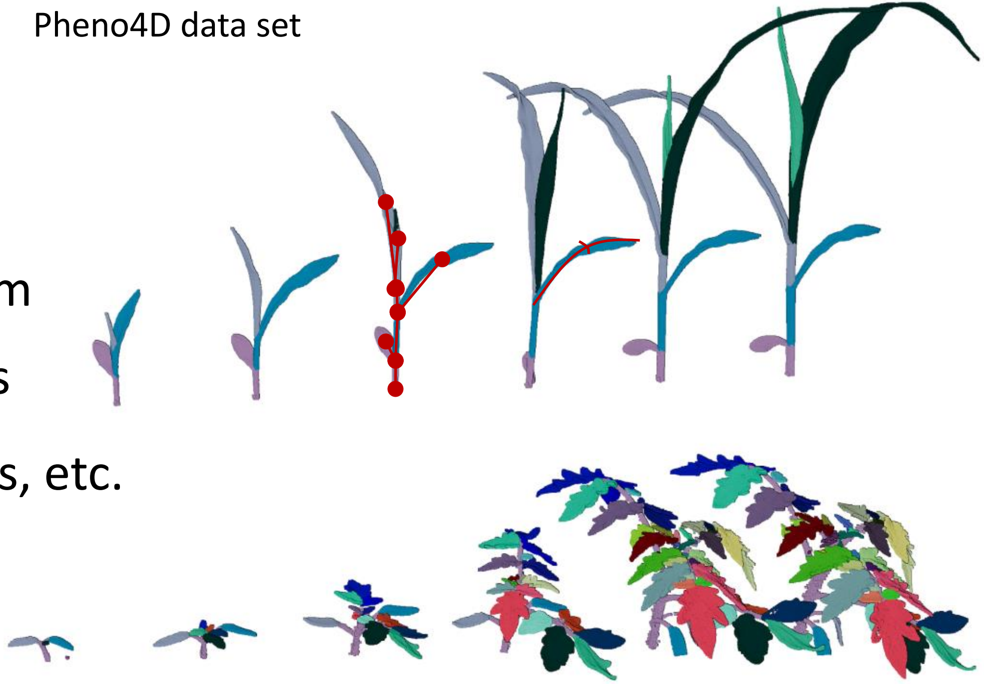
Mesh normal vector

[Rosu 2022]

Reconstruction of Plant Structure

- Identify individual plants
- Segment plant organ instances
- Model the plant as a **structural graph**
  - Establishes a plant coordinate system
- Associate instances over growth stages
- Model appearance, material properties, etc.
- This creates a **Digital Twin** of the plant
  - Basis for plant-science research
  - Basis for targeted interaction with the plant, e.g. contact measurements or harvesting => we must track it in real time while interacting with the plant

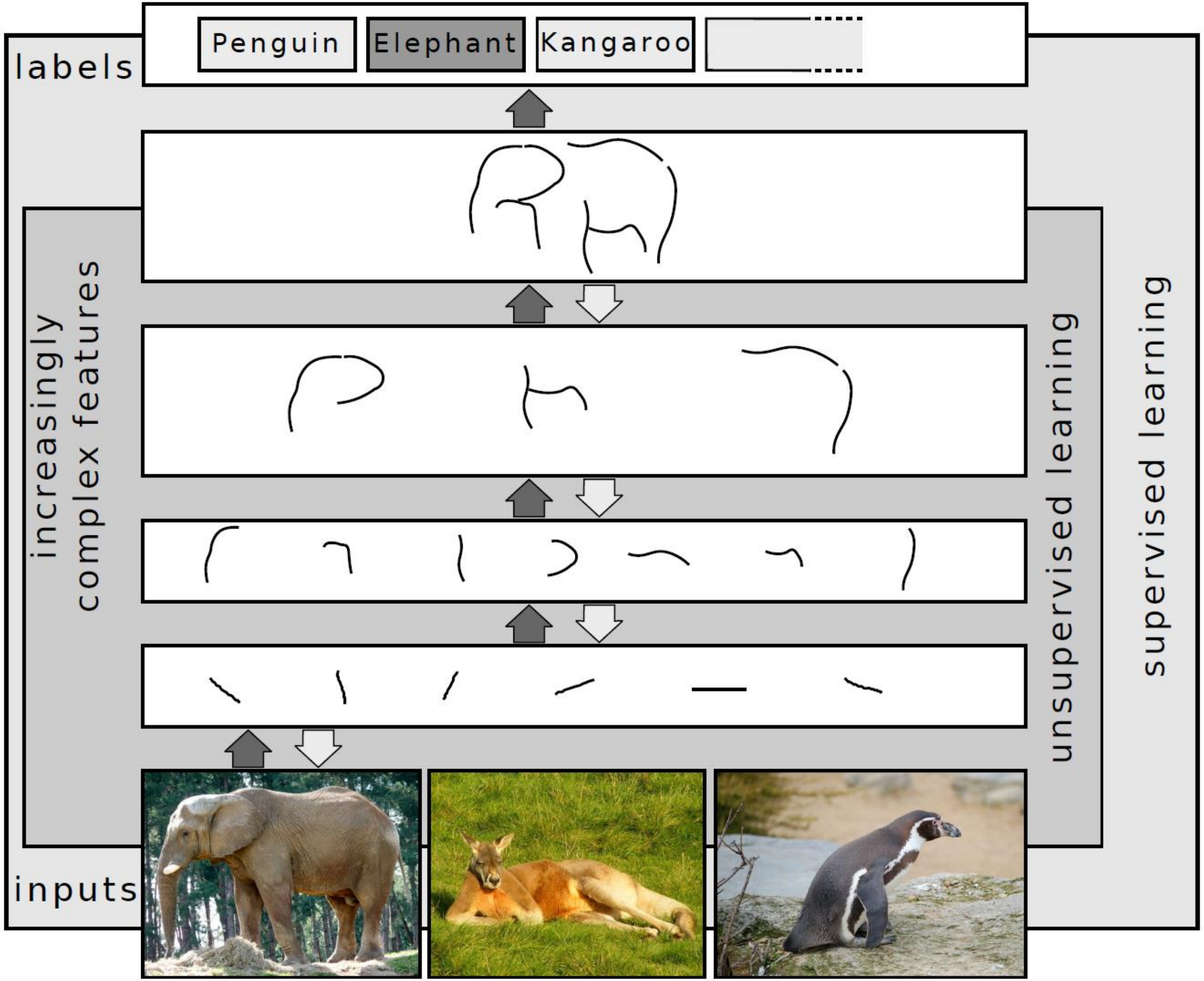
Pheno4D data set



[Schunck et al. PLoS ONE 16(8): e0256340, 2021]

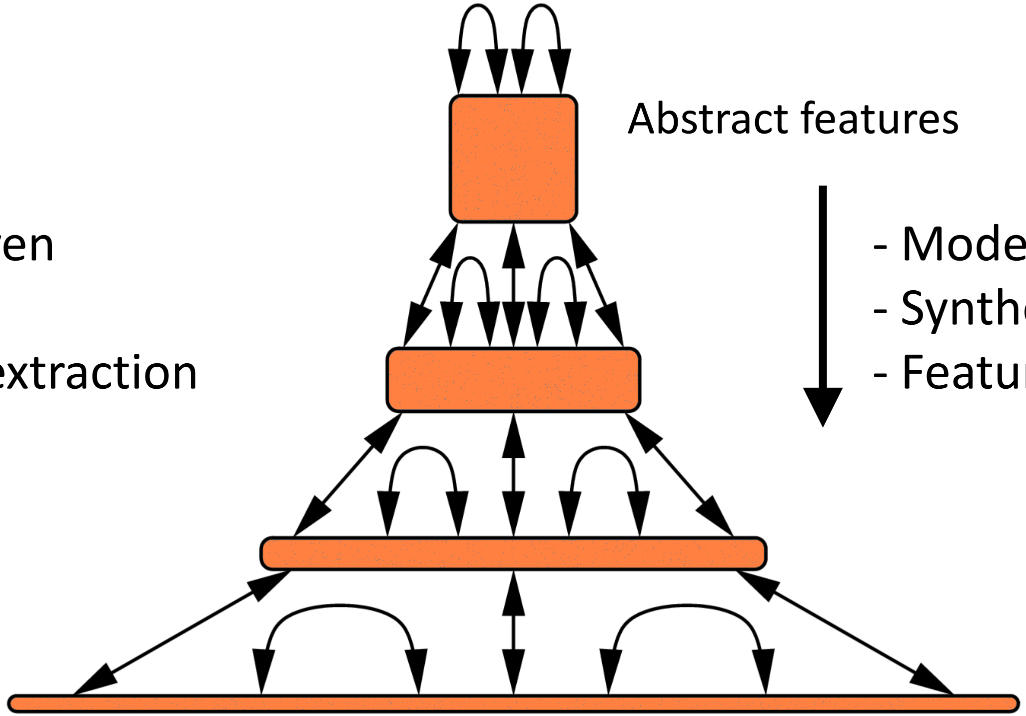
Deep Learning

- Learning layered representations
- Compositionality



[Schulz; Behnke, KI 2012]

Neural Abstraction Pyramid



Abstract features

Signals

- Data-driven
- Analysis
- Feature extraction

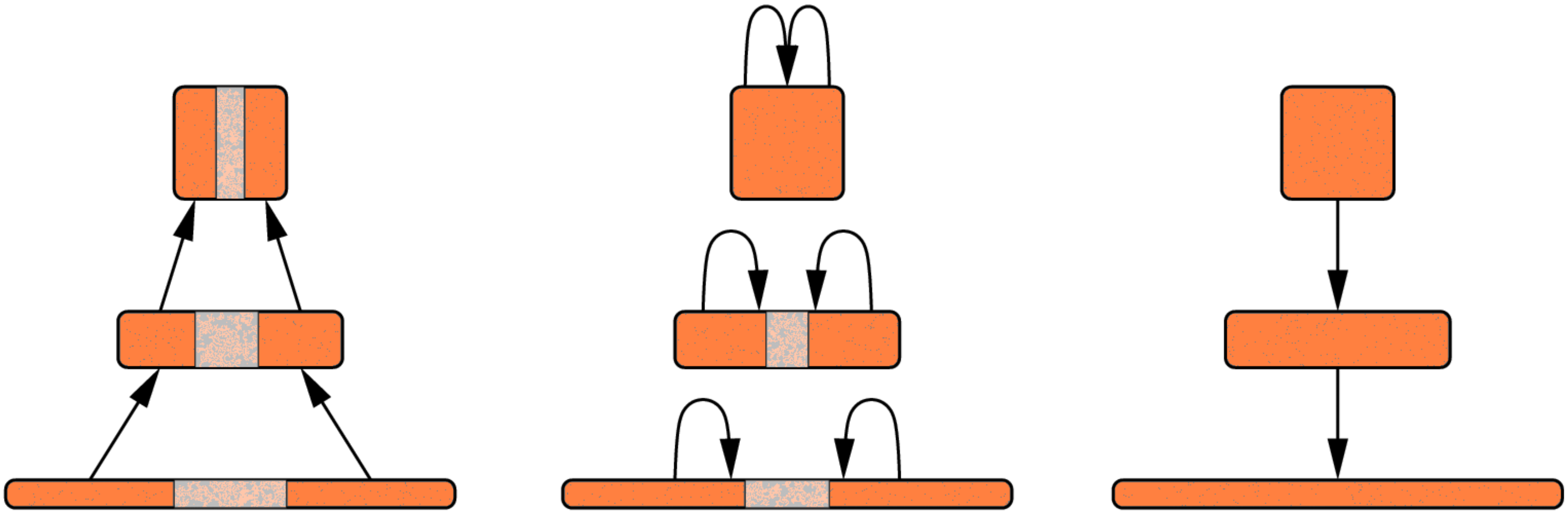
- Model-driven
- Synthesis
- Feature expansion

- Grouping - Competition - Completion

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

Iterative Image Interpretation

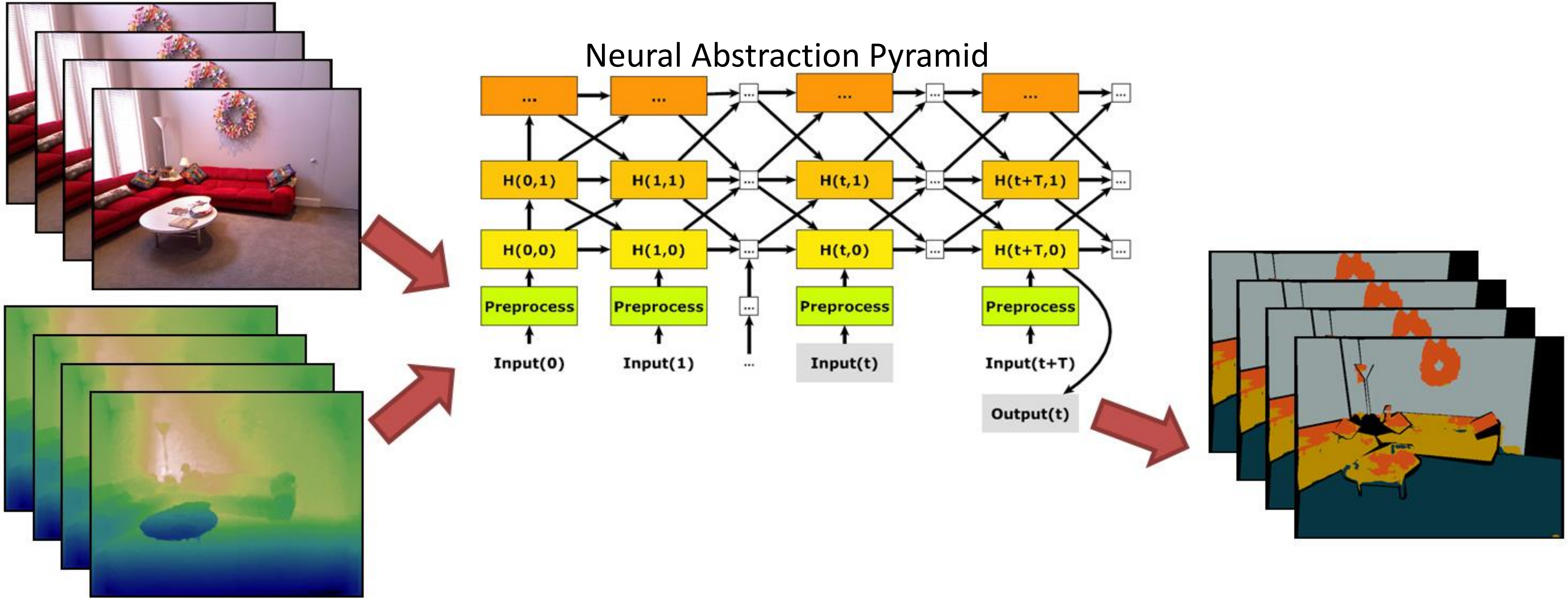
- Interpret most obvious parts first
- Use partial interpretation as context to iteratively resolve local ambiguities



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

Neural Abstraction Pyramid for Object-class Segmentation of RGB-D Video

- Recursive computation is efficient for temporal integration



[Pavel, Schulz, Behnke, Neural Networks 2017]

The Data Problem

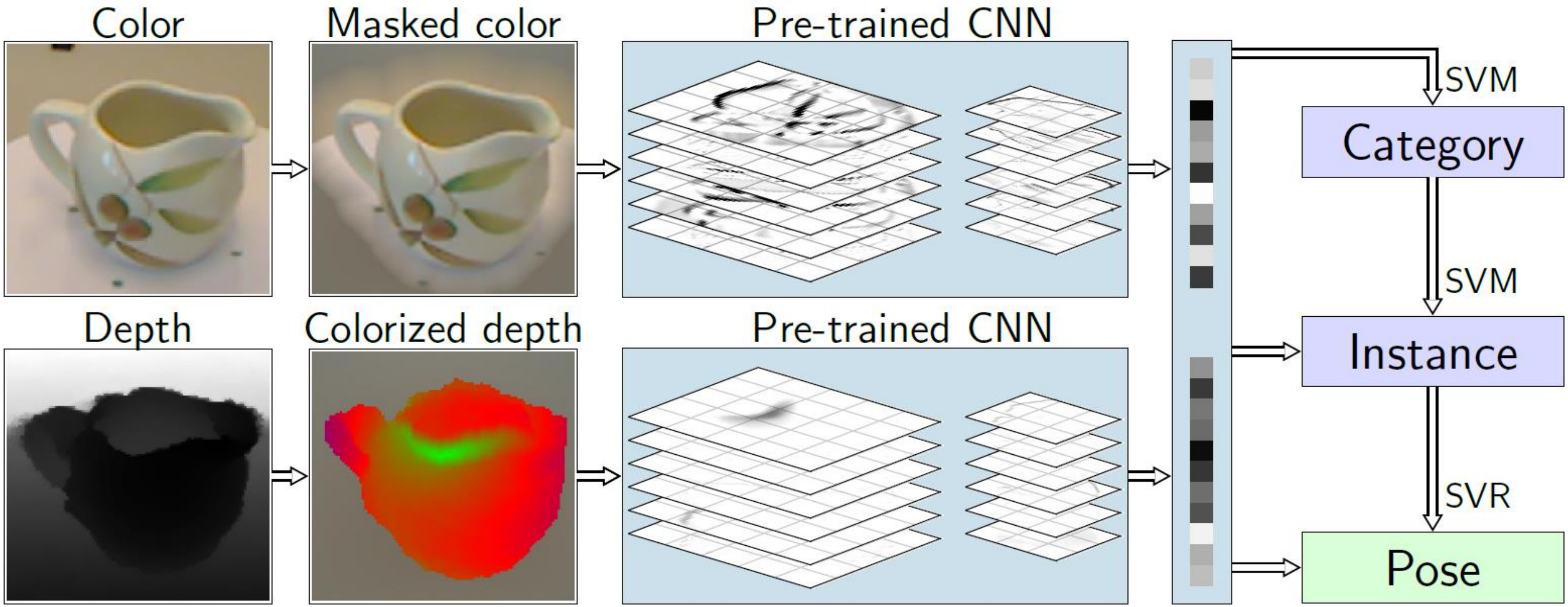
- Deep Learning in robotics (still) suffers from shortage of available examples
- We address this problem in two ways:

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

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

RGB-D Object Recognition and Pose Estimation

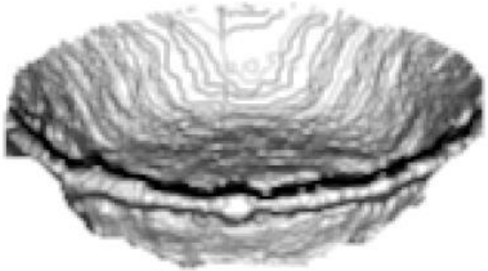
- Transfer learning from large-scale data sets

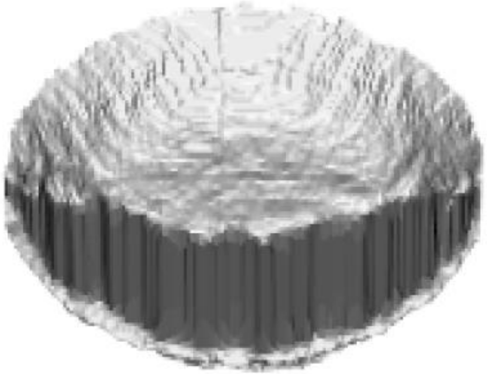


[Schwarz, Schulz, Behnke, ICRA2015]

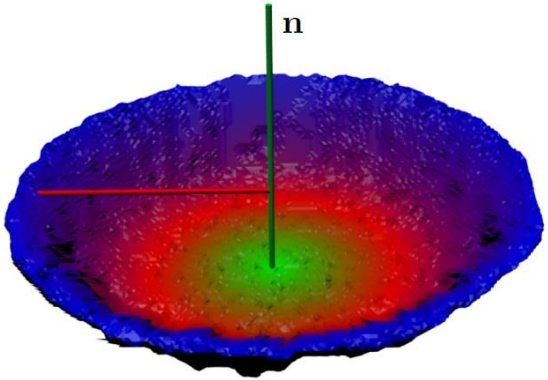
Canonical View, Colorization

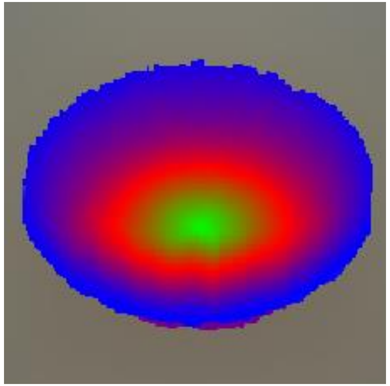
- Objects viewed from different elevation
- Render canonical view





- Colorization based on distance from center vertical

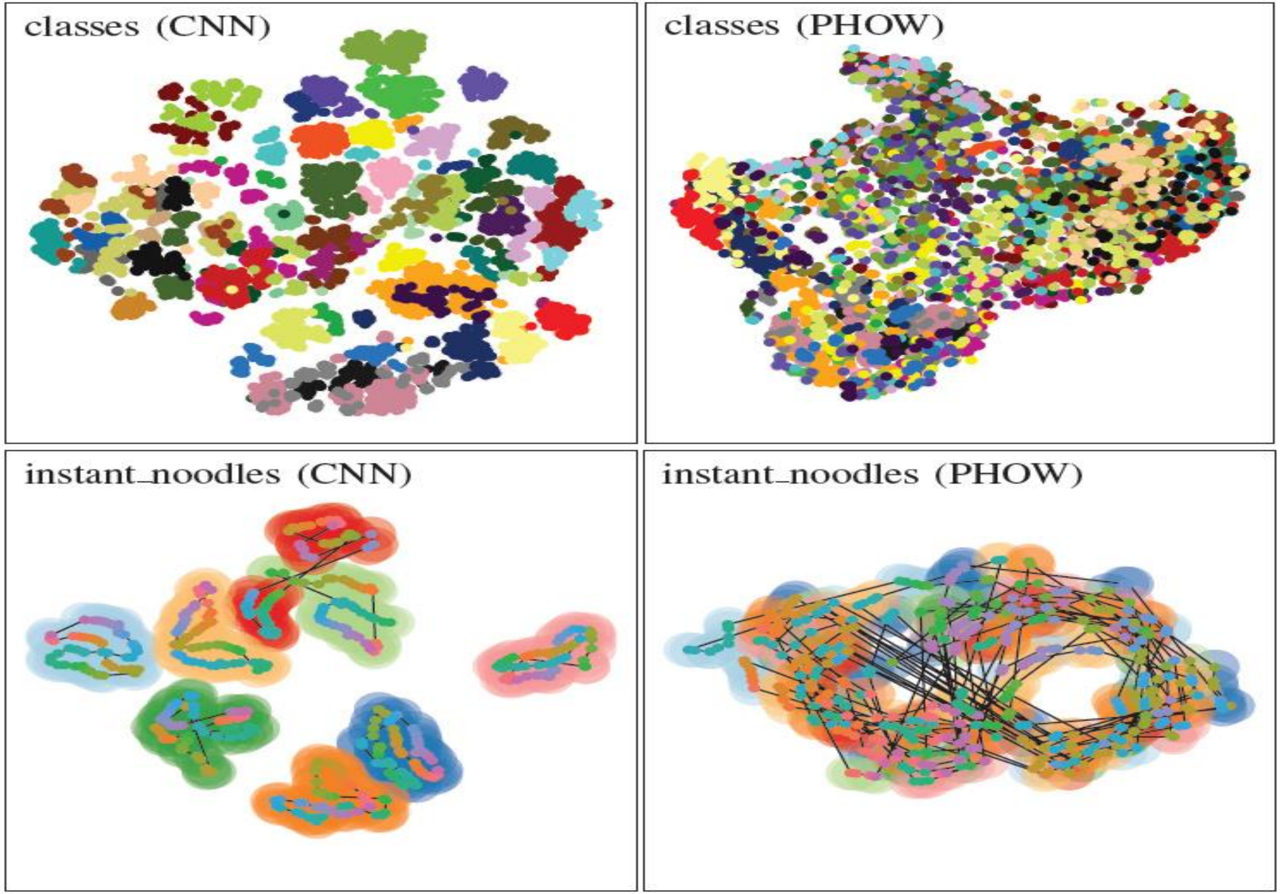




[Schwarz, Schulz, Behnke, ICRA2015]

Pretrained Features Disentangle Data

- t-SNE embedding



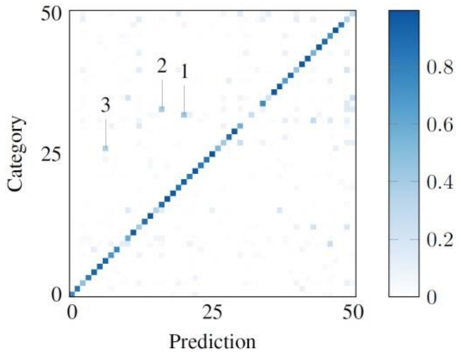
[Schwarz, Schulz, Behnke ICRA2015]

Recognition Accuracy

- Improved both category and instance recognition

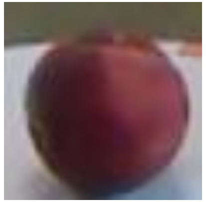
	Category Accuracy (%)		Instance Accuracy (%)	
Method	RGB	RGB-D	RGB	RGB-D
Lai *et al.* [1]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo *et al.* [2]	82.4 ± 3.1	87.5 ± 2.9	**92.1**	92.8
PHOW[3]	80.2 ± 1.8	—	62.8	—
Ours	$\mathbf{83.1 \pm 2.0}$	88.3 ± 1.5	92.0	**94.1**
Ours	$\mathbf{83.1 \pm 2.0}$	$\mathbf{89.4 \pm 1.3}$	92.0	**94.1**

- Confusion:



1: pitcher / coffe mug

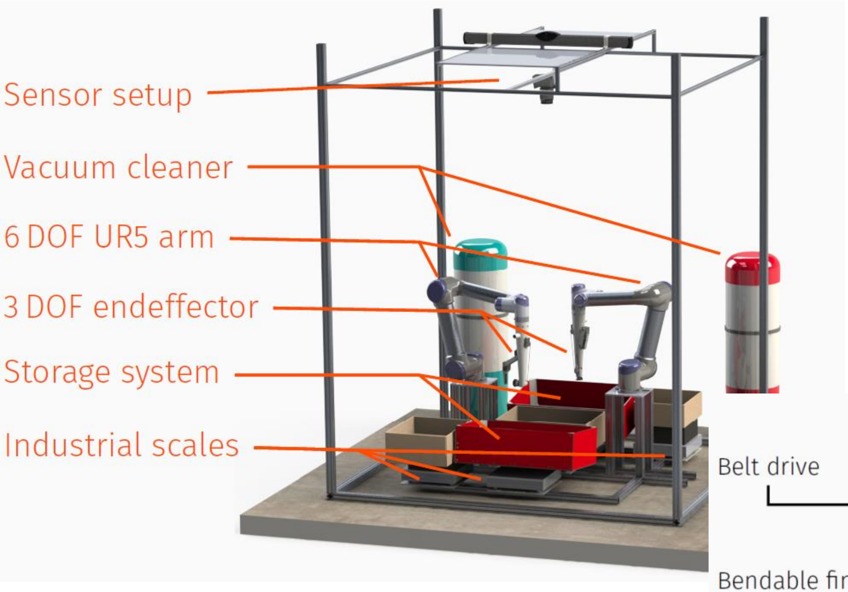

2: peach / sponge

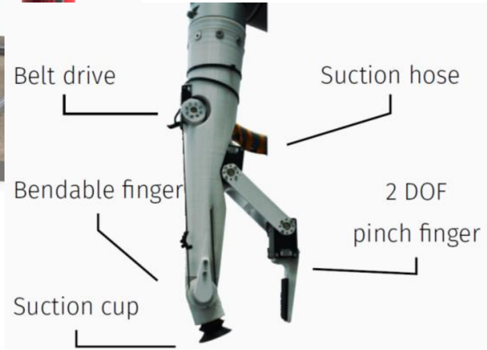



[Schwarz, Schulz, Behnke, ICRA2015]

Amazon Robotics Challenge

- Storing and picking of items
- Dual-arm robotic system





[Schwarz et al. ICRA 2018]

[Amazon]

Object Capture and Scene Rendering

- Turntable + DLSR camera

- Insertion in complex annotated scenes

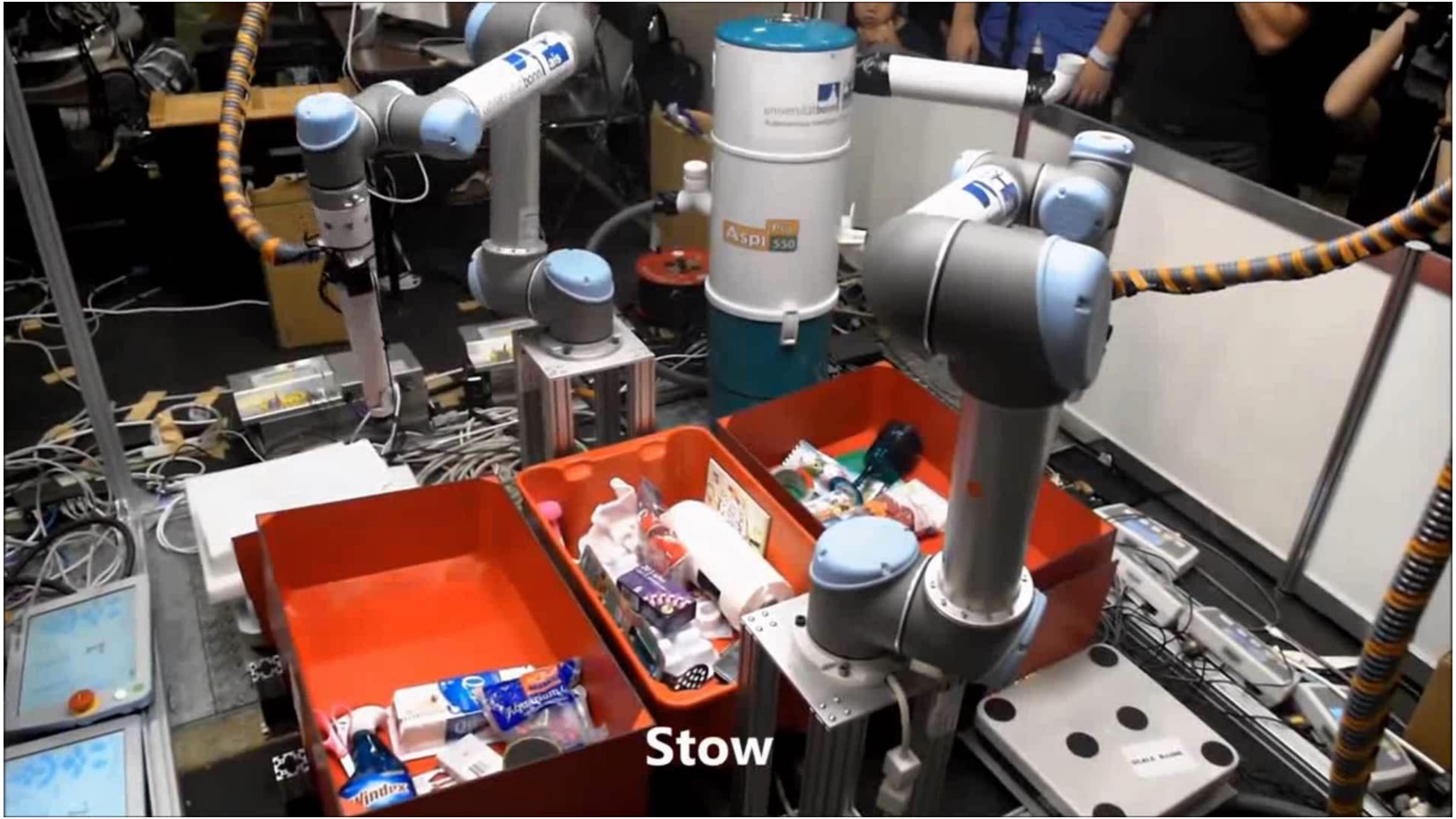
[Schwarz et al. ICRA 2018]

Semantic Segmentation and Grasp Pose Estimation

- Semantic segmentation using RefineNet [Lin et al. CVPR 2017]
- Grasp positions in segment centers

[Schwarz et al. ICRA 2018]

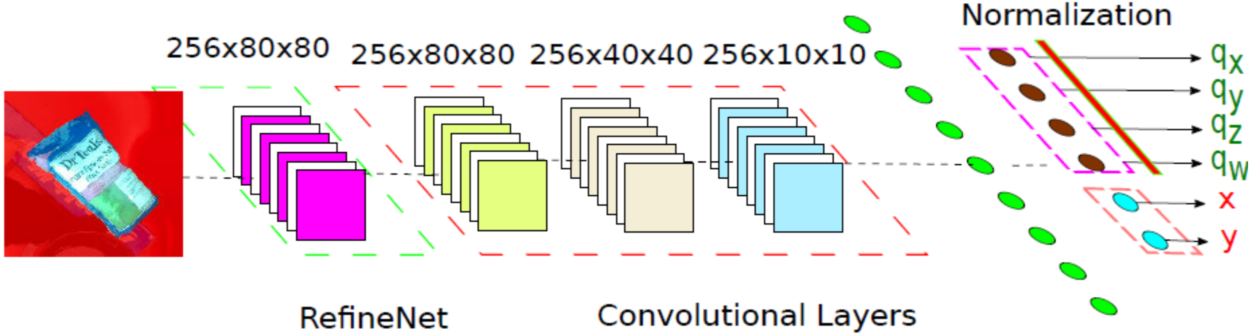
Amazon Robotics Challenge 2017



[Schwarz et al. ICRA 2018]

Object Pose Estimation

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



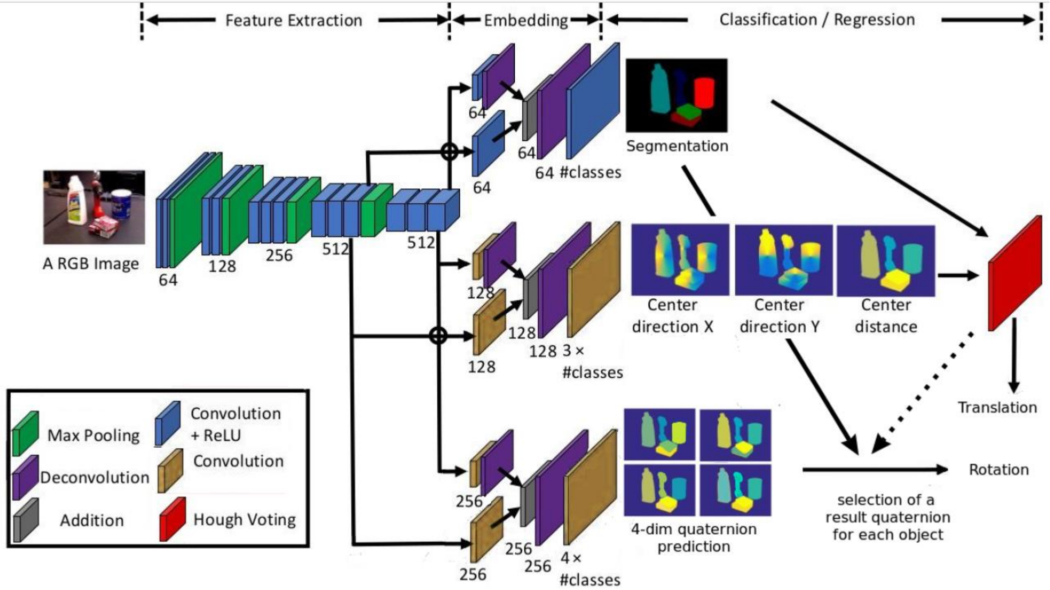
Input

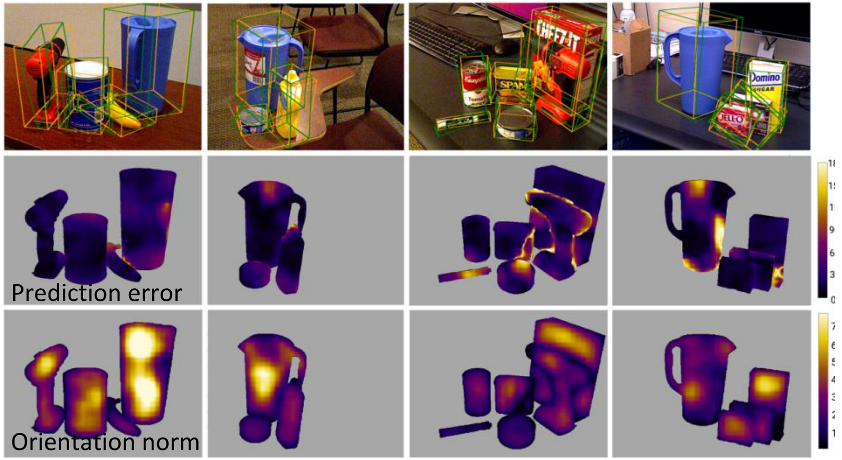
Predicted pose

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

Dense Convolutional 6D Object Pose Estimation

- Extension of PoseCNN [Xiang et al. RSS 2018]
- Dense prediction of object center and orientation, without cutting out

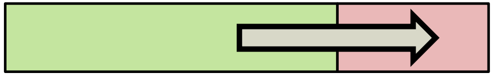




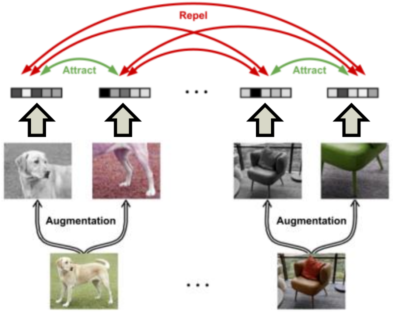
[Capellen et al., VISAPP 2020]

Self-supervised Learning

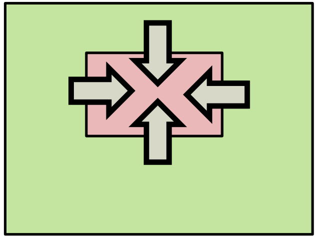
- Special case of unsupervised learning
  - Learning to represent the world in a non task-specific way
  - Learning predictive models for planning and control
- Define a **pretext task** without labels that needs some understanding of the data, e.g.



  - Predict the future from the past
  - Fill-in the gaps
  - Contrastive methods



Input: Unlabled images

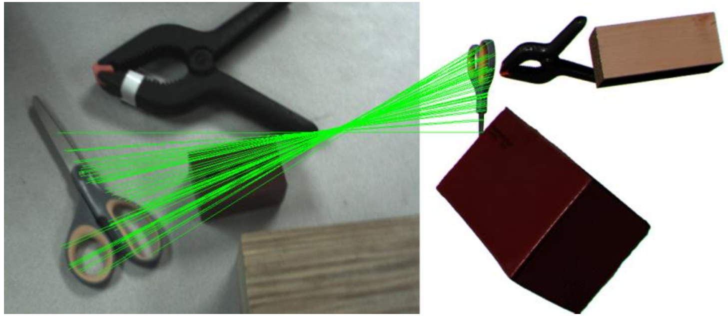


Downstream task

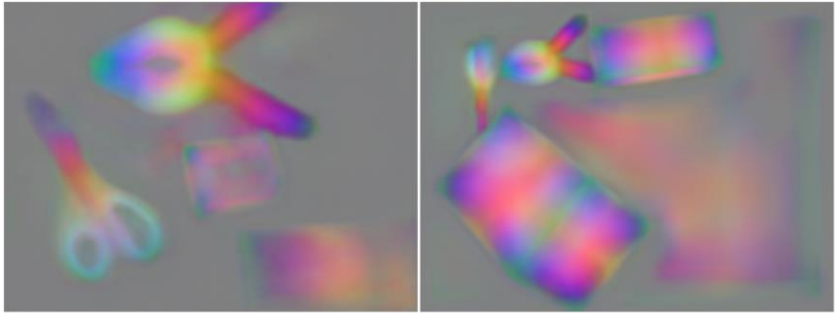
- Use learned representation to quickly learn **downstream task**
  - Supervised learning
  - Reinforcement learning

Self-Supervised Surface Descriptor Learning

- Feature descriptor should be constant under different transformations, viewing angles, and environmental effects such as lighting changes
- Descriptor should be unique to facilitate matching across different frames or representations
- Learn dense features using a contrastive loss



Known correspondences



Learned features

[Periyasamy, Schwarz, Behnke Humanoids 2019]

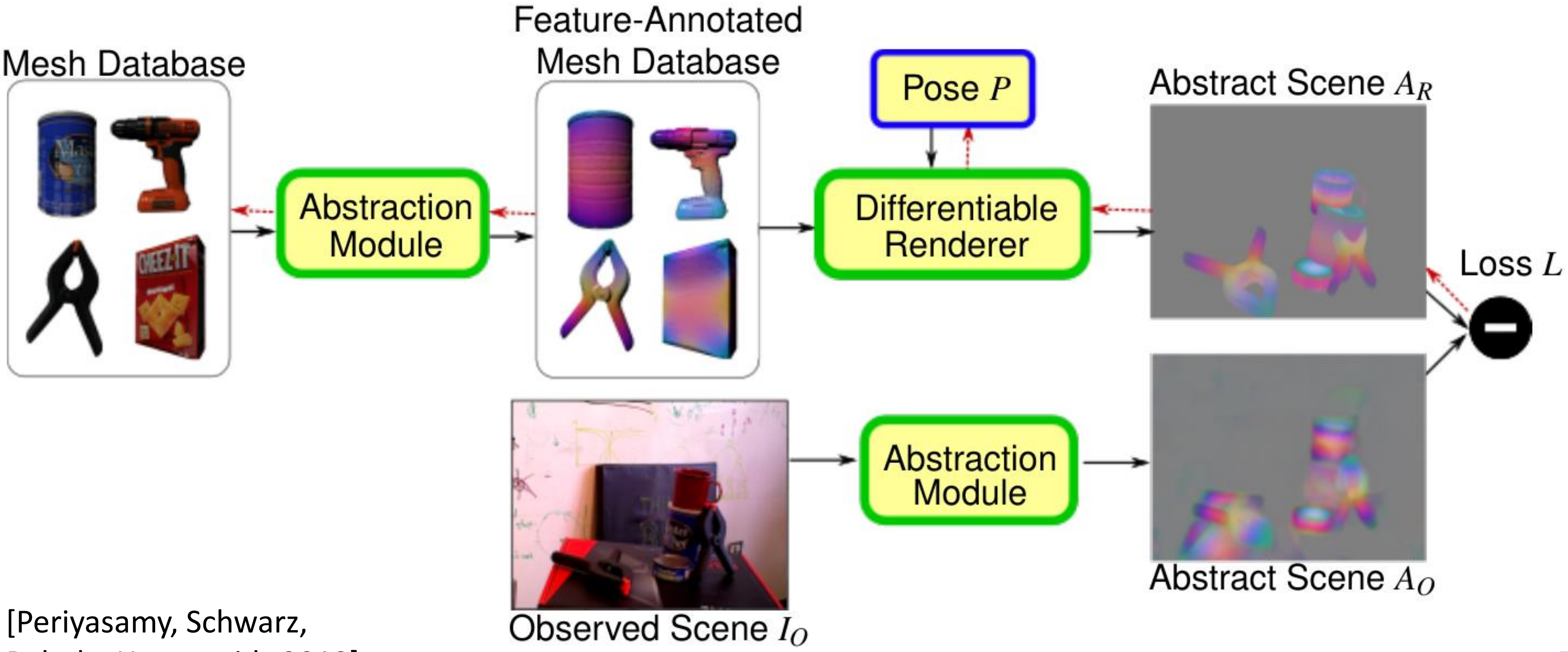
Descriptors as Texture on Object Surfaces

- Learned feature channels used as textures for 3D object models
- Used for 6D object pose estimation

[Periyasamy, Schwarz, Behnke Humanoids 2019]

Abstract Object Registration

- Compare rendered and actual scene in feature space
- Adapt model pose by gradient descent



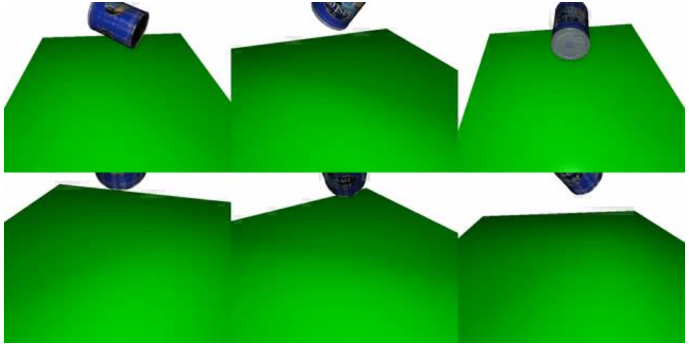
[Periyasamy, Schwarz, Behnke Humanoids 2019]

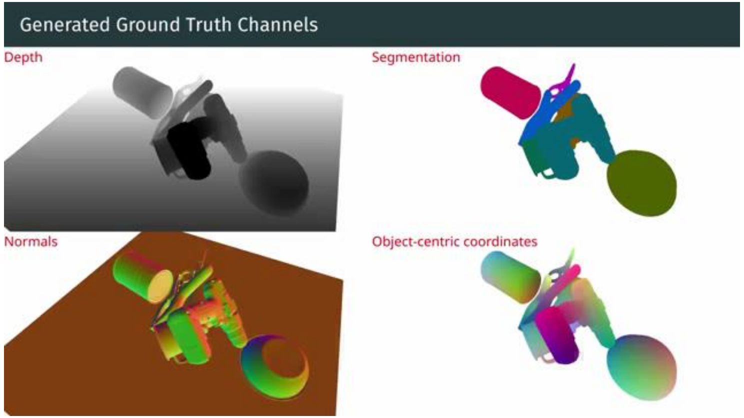
Registration Examples

[Periyasamy, Schwarz, Behnke Humanoids 2019]

Learning from Synthetic Scenes

- Cluttered arrangements from 3D meshes
- Photorealistic scenes with randomized material and lighting including ground truth
- For online learning & render-and-compare
- Semantic segmentation on YCB Video Dataset
  - Close to real-data accuracy
  - Improves segmentation of real data

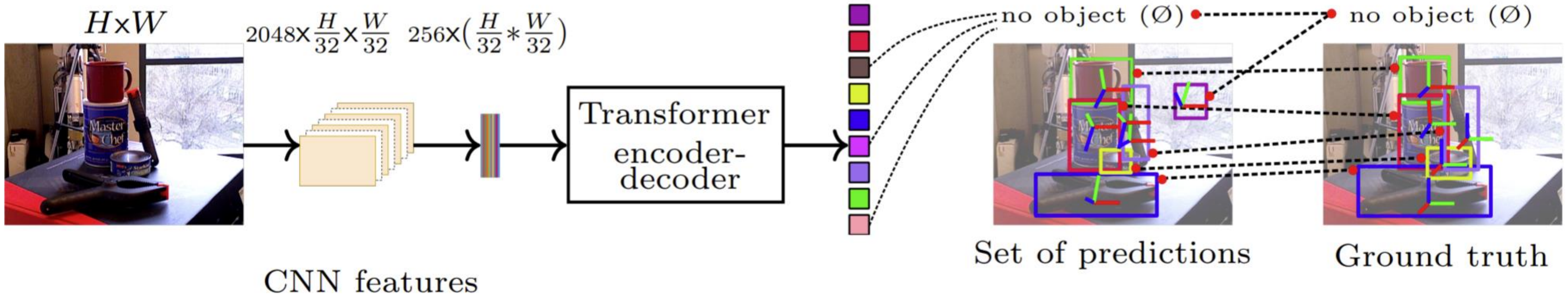




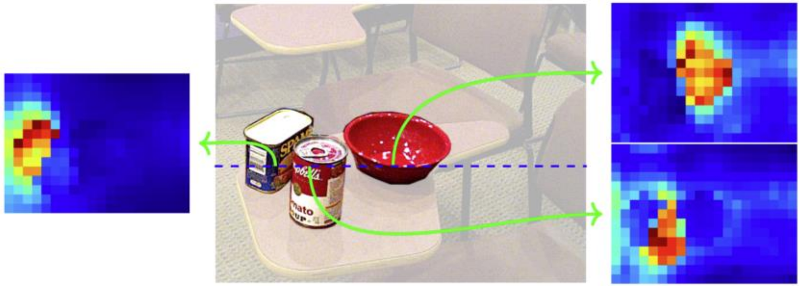
[Schwarz and Behnke, ICRA 2020]

T6D-Direct: Transformers for Multi-Object 6D Pose Direct Regression

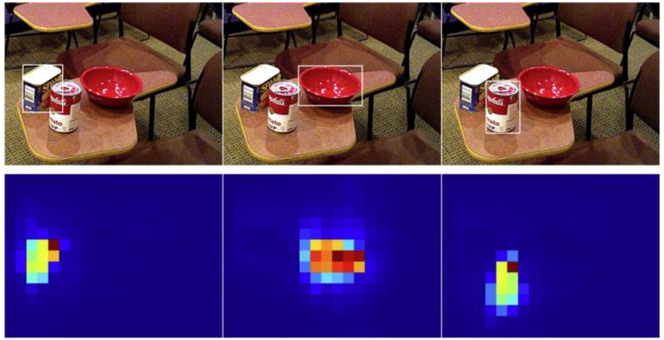
- Extends DETR: End-to-end object detection with transformers [Carion et al. ECCV 2020]
- End-to-end differentiable pipeline for 6D object pose estimation



Encoder self-attention

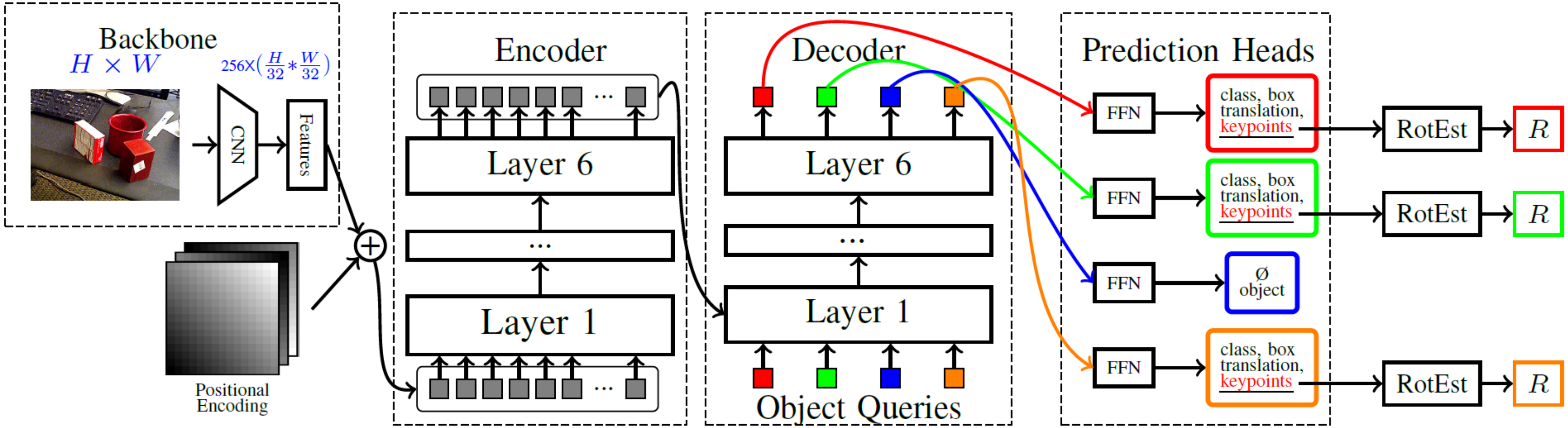


Object detections and decoder attention



[Amini et al. GCPR 2021]

Multi-Object 6D Pose Estimation using Keypoint Regression

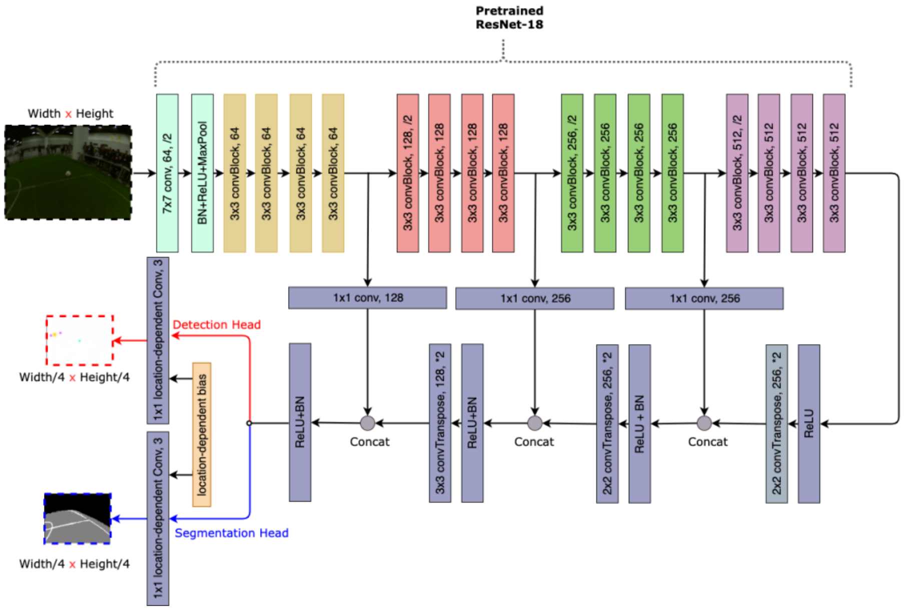


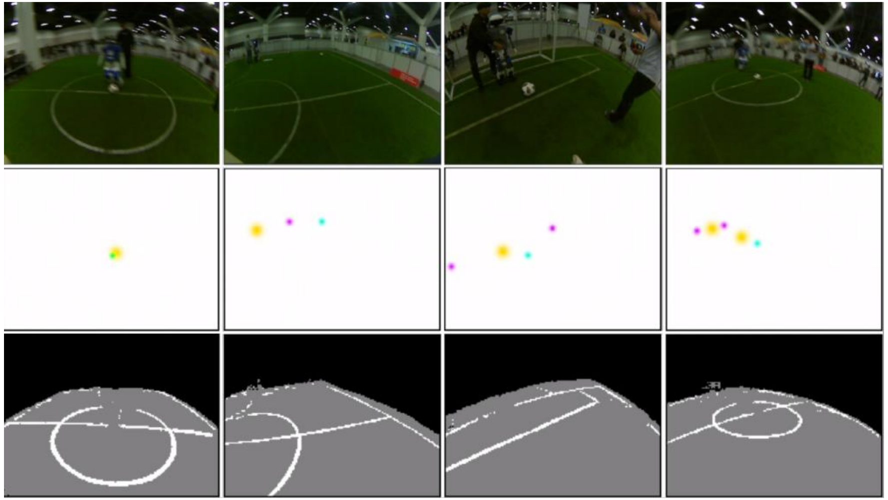
[Amini et al. IAS 2022, Best Paper Award]

RoboCup 2022 in Bangkok

Transfer Learning for Visual Perception

- Encoder-decoder network
- Two outputs
  - Object detection
  - Semantic segmentation
- Location-dependent bias



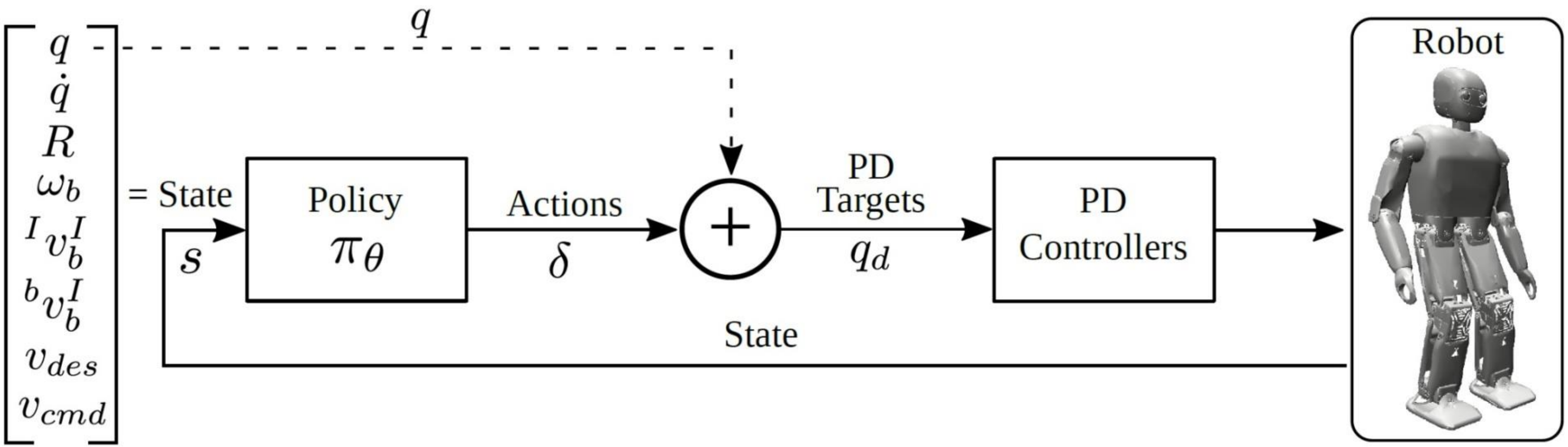


- Detects objects that are hard to recognize for humans
- Robust to lighting changes

[Rodriguez et al. RoboCup 2019]

Learning Omnidirectional Gait from Scratch

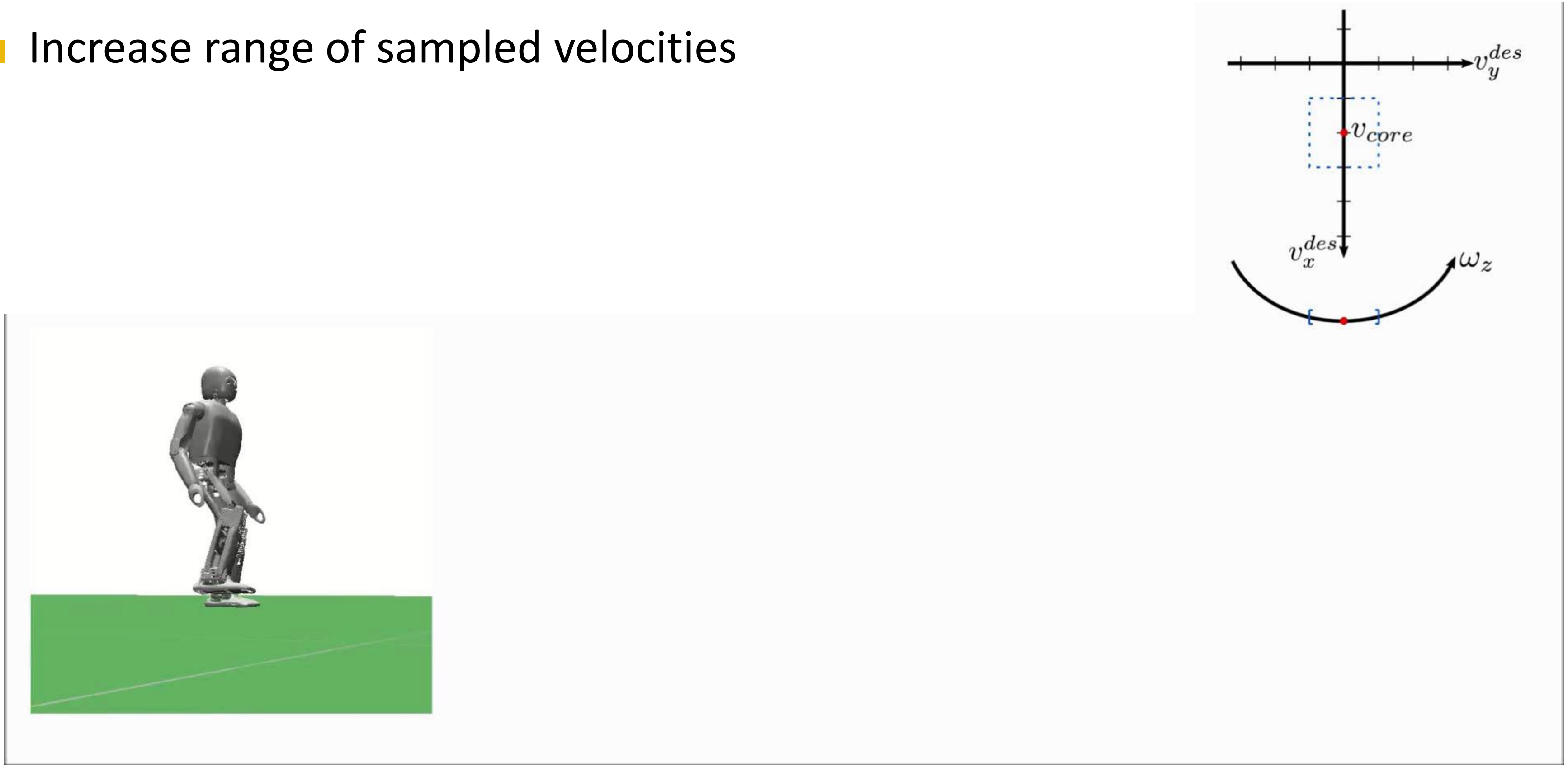
- State includes joint positions and velocities, robot orientation, robot speed
- Actions are increments of joint positions
- Simple reward structure
  - Velocity tracking
  - Pose regularization
  - Not falling



[Rodriguez and Behnke, ICRA 2021]

Learning Curriculum

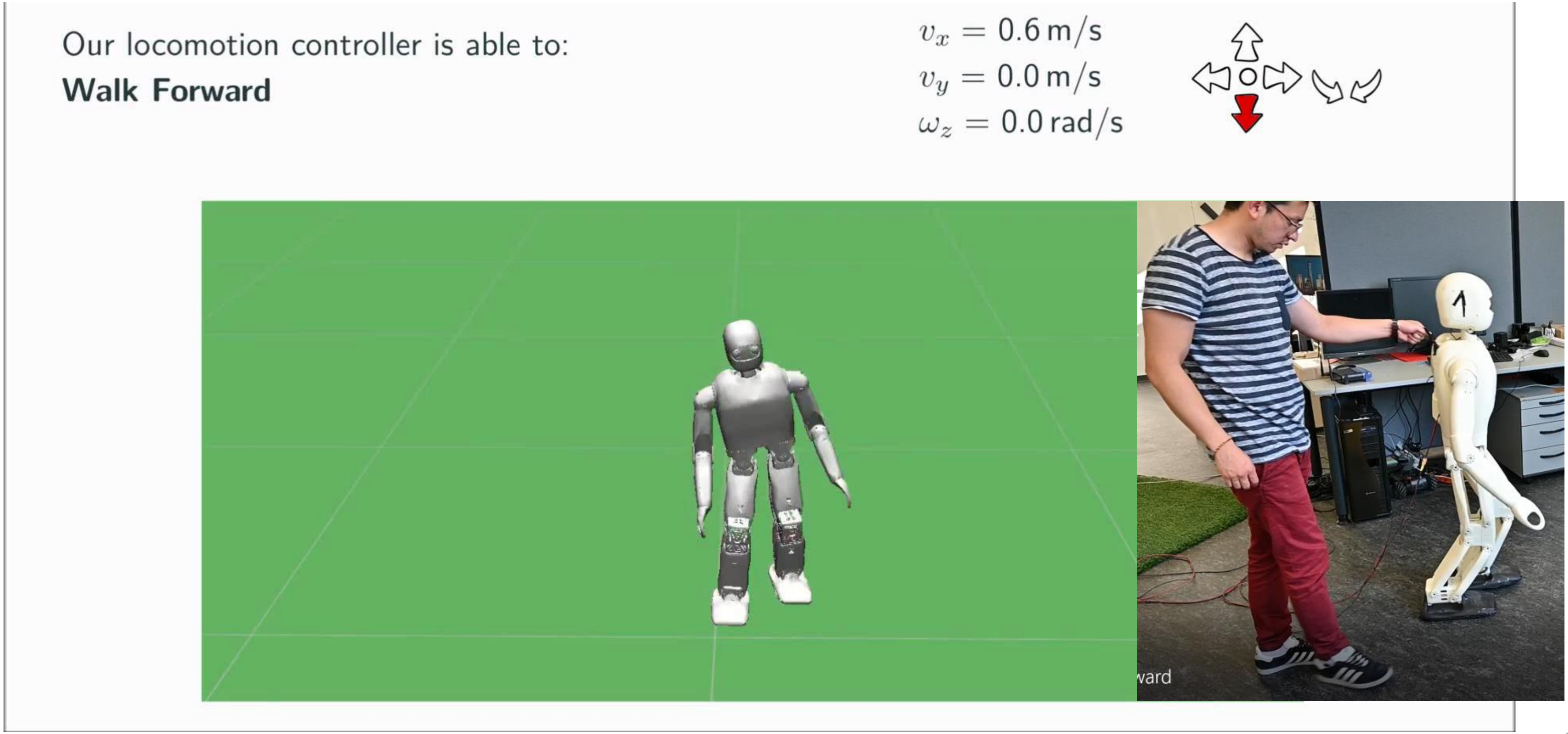
- Start with small velocities
- Increase range of sampled velocities



[Rodriguez and Behnke, ICRA 2021]

Learned Omnidirectional Gait

- Target velocity can be changed continuously

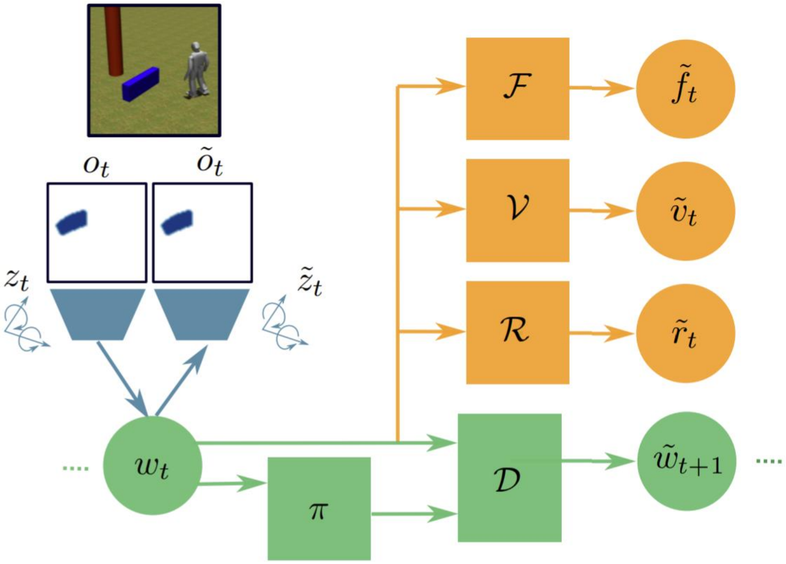


[Rodriguez and Behnke, ICRA 2021]

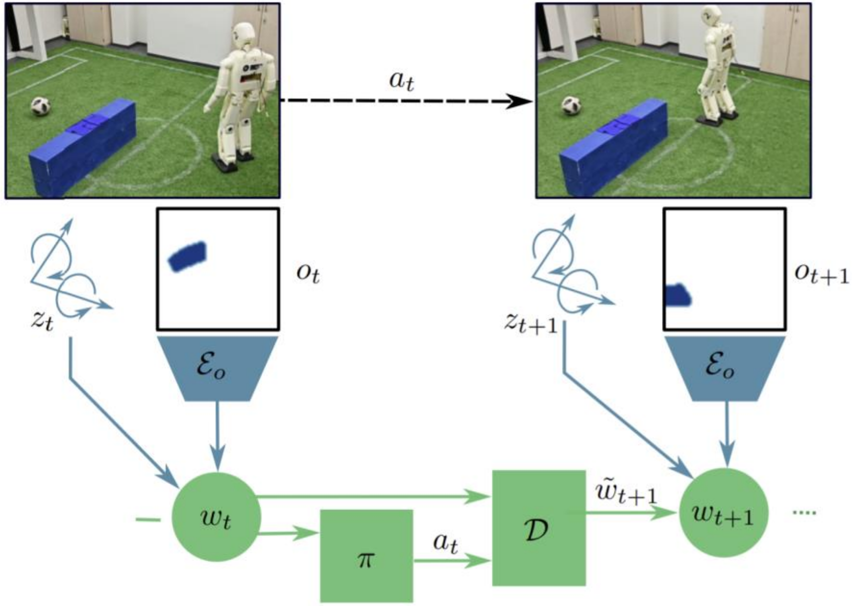
Learning Mapless Humanoid Navigation

- Visual (RGB images) and nonvisual observations to learn a control policy and an environment dynamics model
- Anticipate terminal states of success and failure

Training



Inference



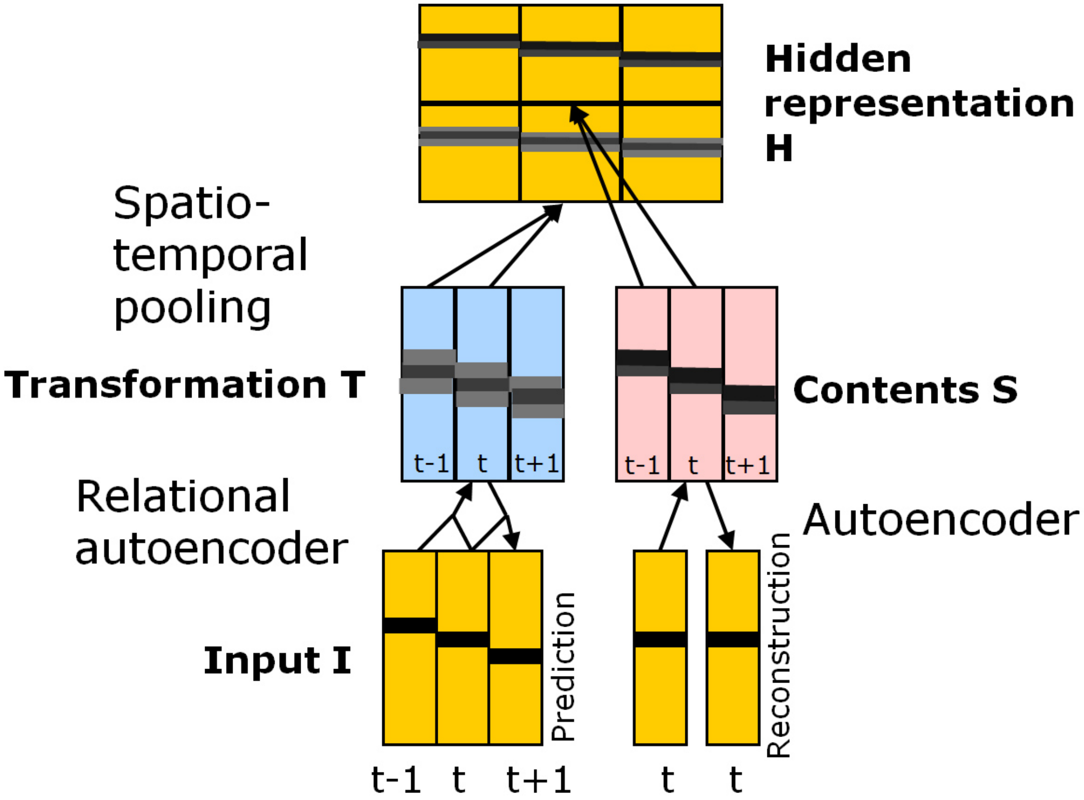
[Brandenburger et al. IROS 2021]

Learning Mapless Humanoid Navigation

[Brandenburger et al. IROS 2021]

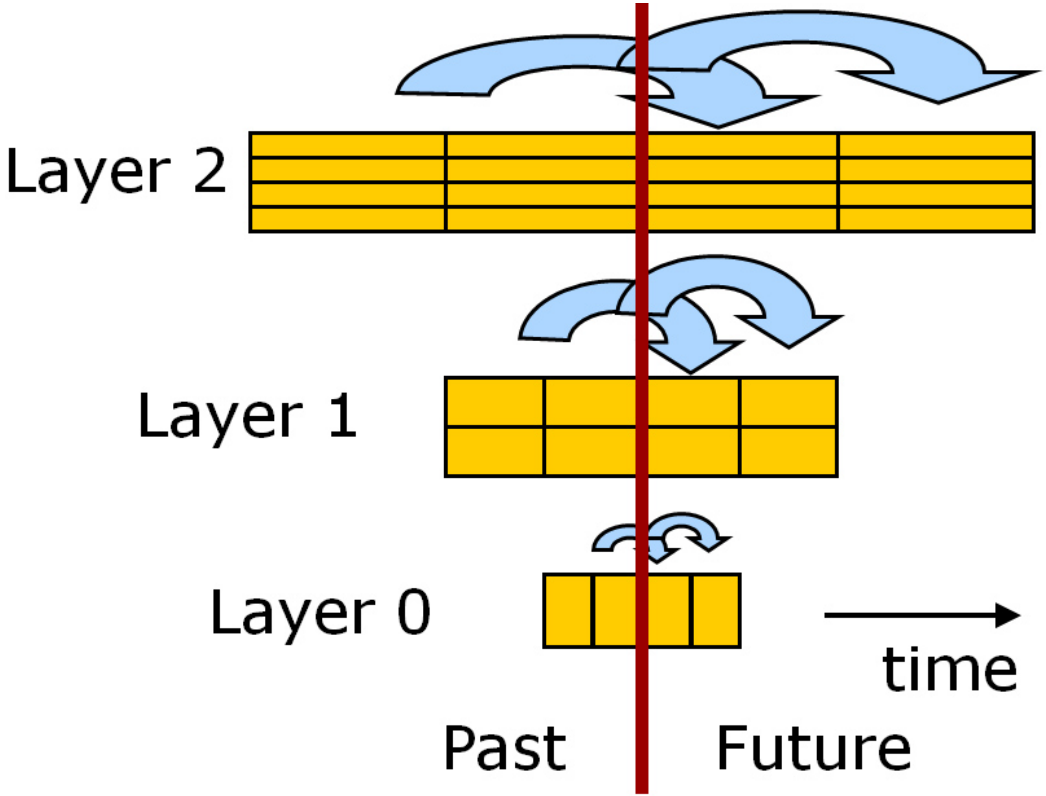
Learning of Hierarchical Representations for Prediction

- Local learning module



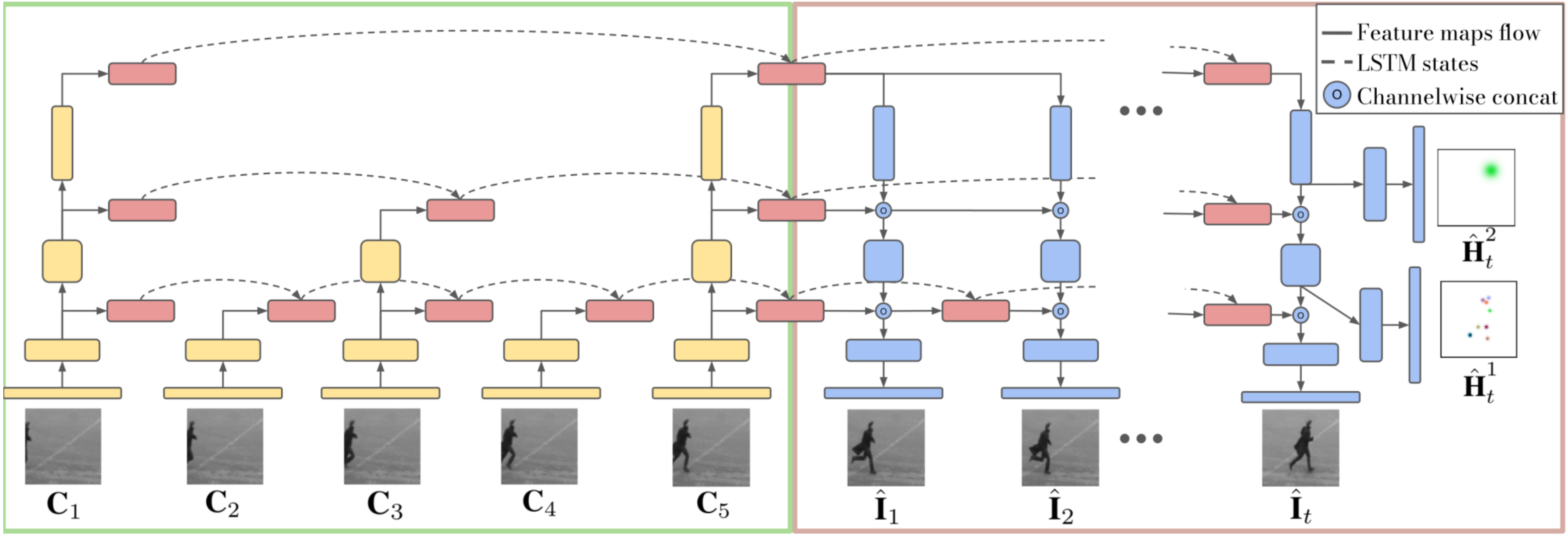
Learning of Hierarchical Representations for Prediction

- Coarser, more abstract predictions for longer time horizons in higher layers



MSPred: Video Prediction at Multiple Spatio-Temporal Scales

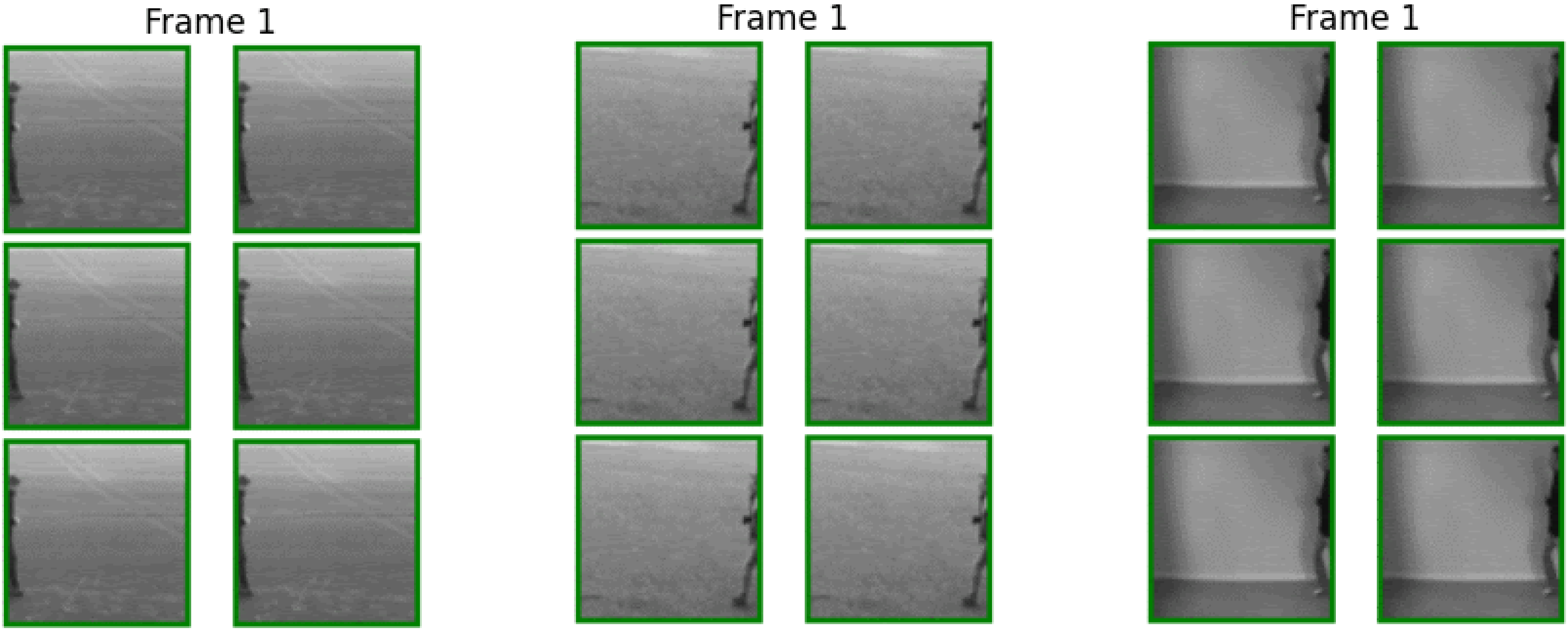
- Coarser, more abstract predictions for longer time horizons in higher layers
- Predict image itself, human pose joint keypoints, and human body position



[Villar-Corrales et al., BMVC 2022]

MSPred: Video Prediction at Multiple Spatio-Temporal Scales

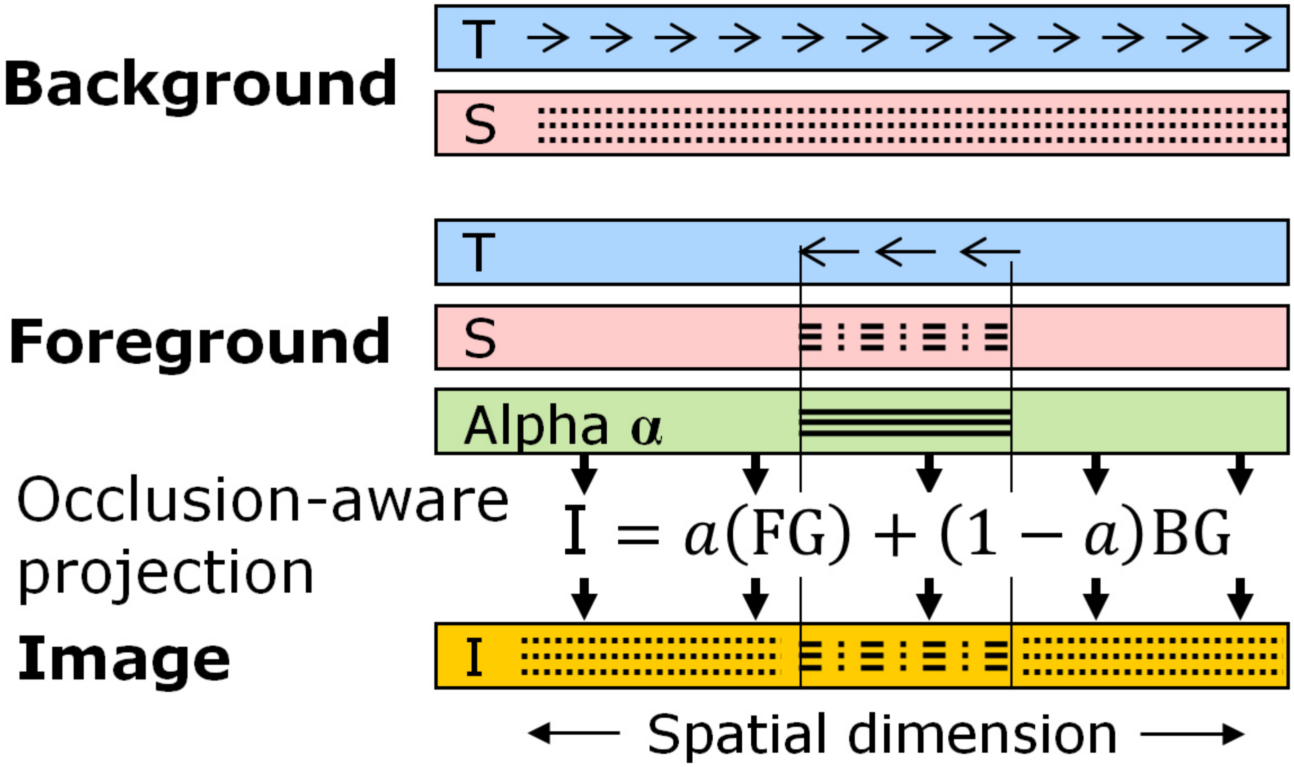
- Coarser, more abstract predictions for longer time horizons in higher layers
- Predict image itself, human pose joint keypoints, and human body position



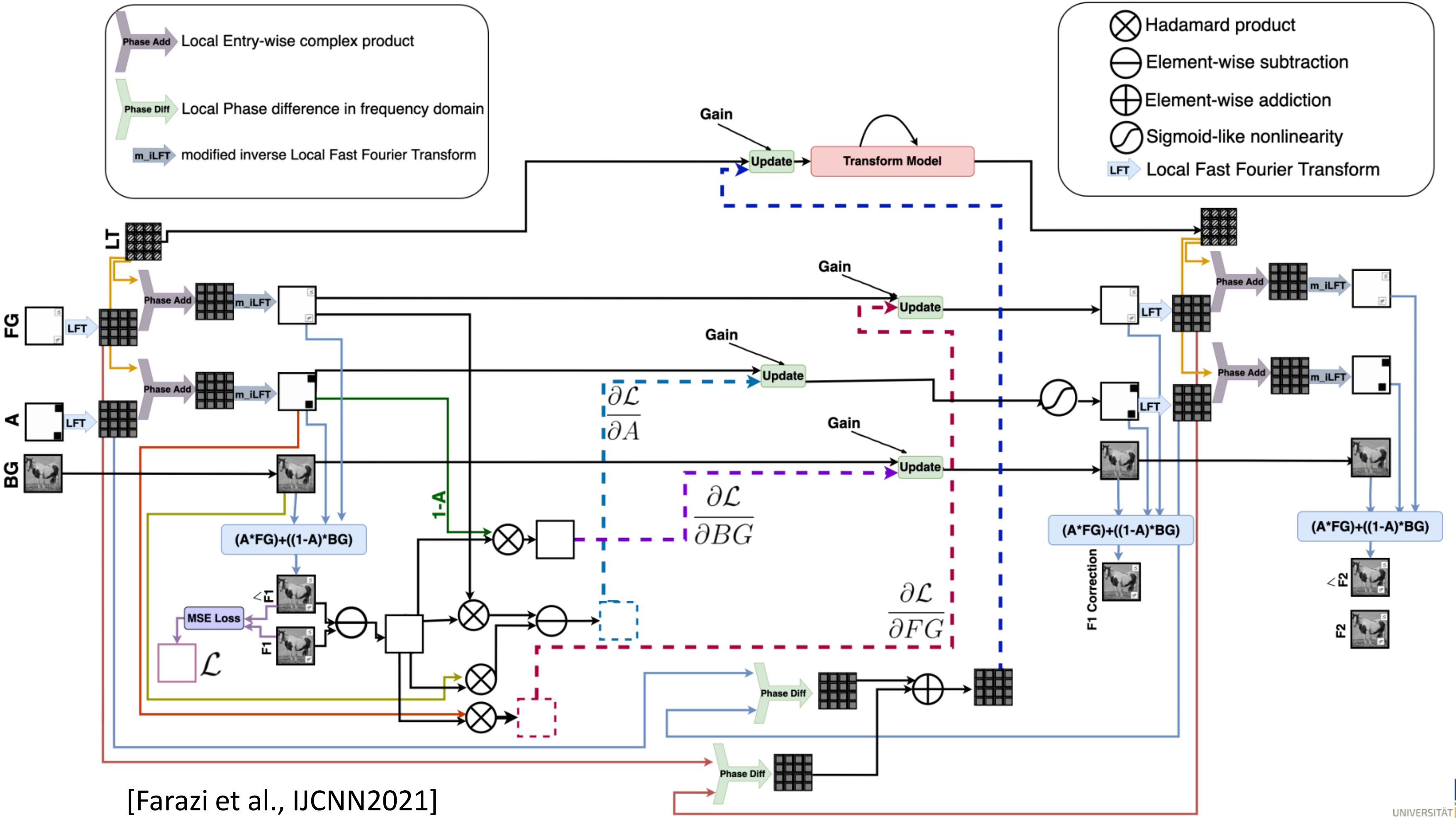
[Villar-Corrales et al., BMVC 2022]

Depth-layered Models for Prediction

- Modelling occlusions



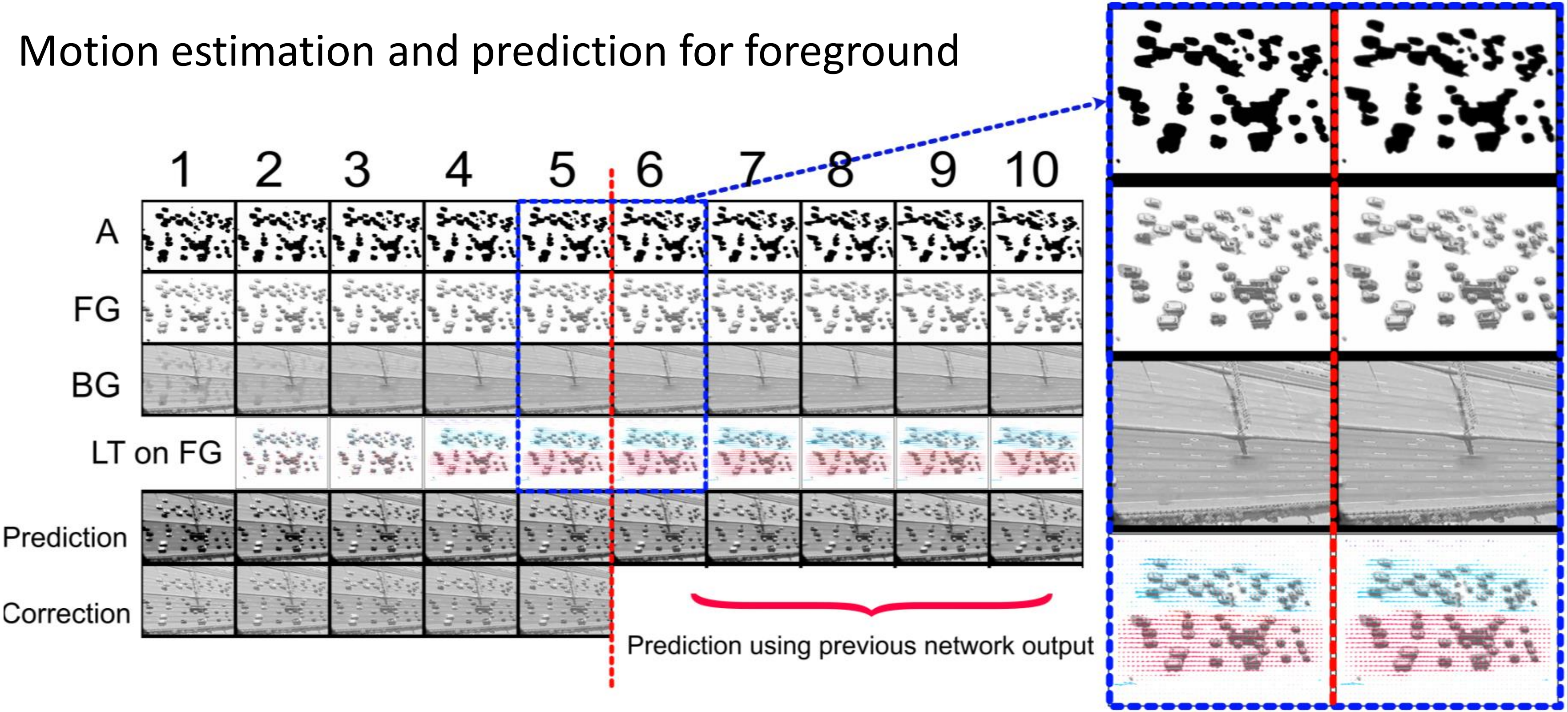
Local Frequency Domain Transformer Networks: Motion Segmentation



[Farazi et al., IJCNN2021]

Local Frequency Domain Transformer Networks: Motion Segmentation

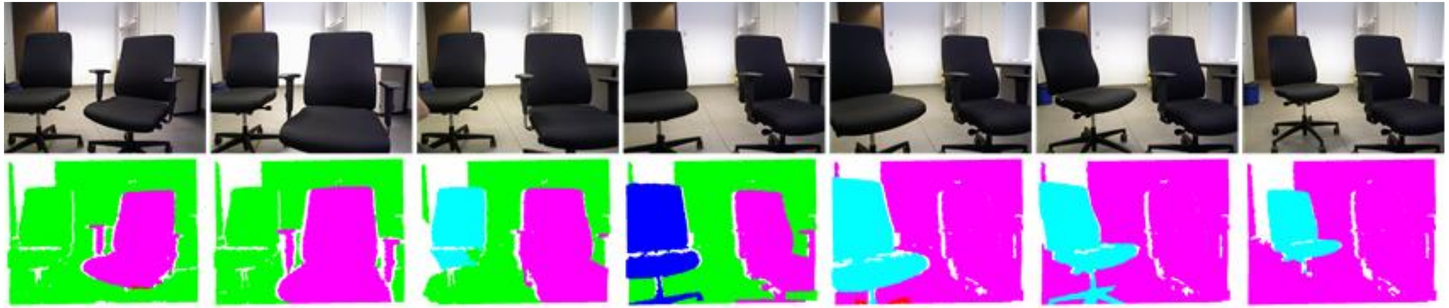
- Unsupervised foreground/background segmentation
- Motion estimation and prediction for foreground



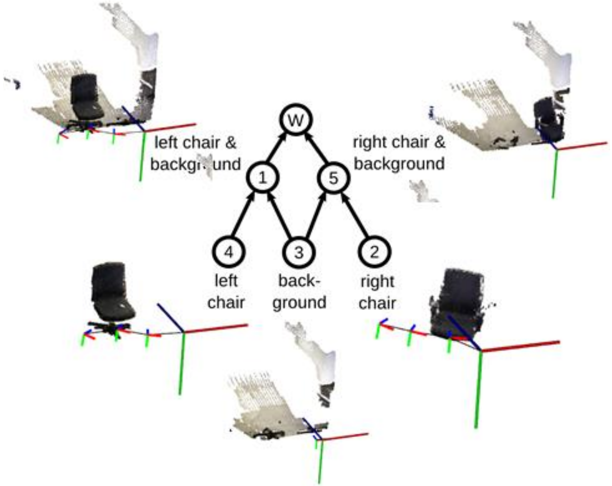
[Farazi et al., IJCNN2021]

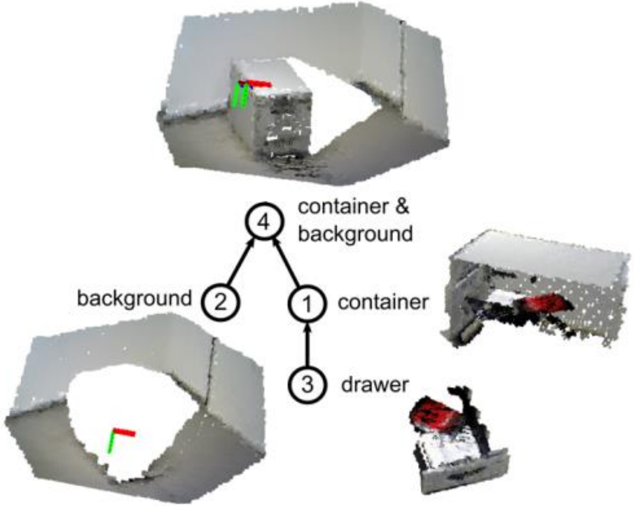
Hierarchical Object Discovery trough Motion Segmentation

- Simultaneous object modeling and motion segmentation



- Inference of a segment hierarchy



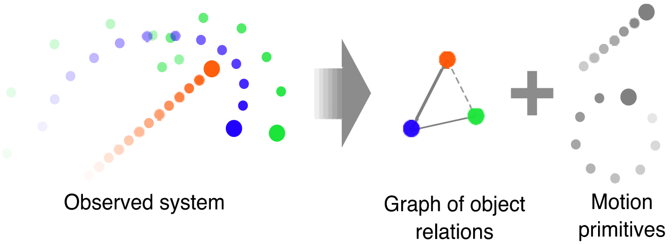


[Stückler, Behnke: IJCAI 2013]

Fourier-based Video Prediction through Relational Object Motion

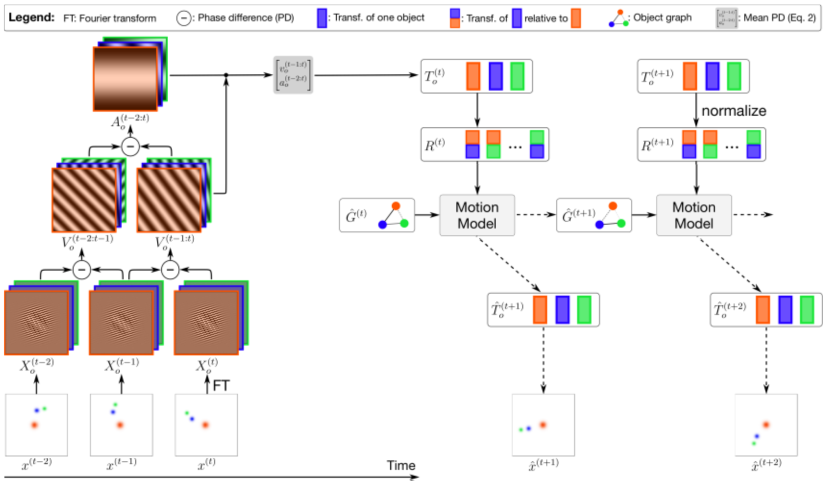
- Model relative object movement

Disentangling Motion and Object Relations

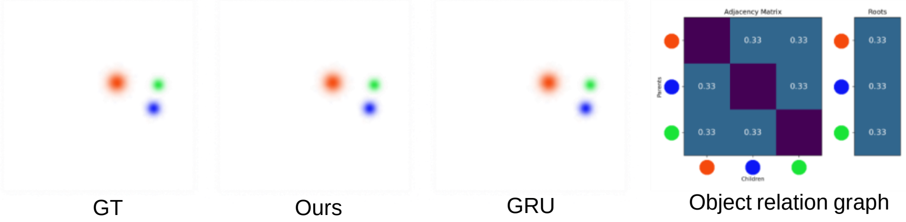


- Star, planet, moon data set
- Infer object relation graph

Method



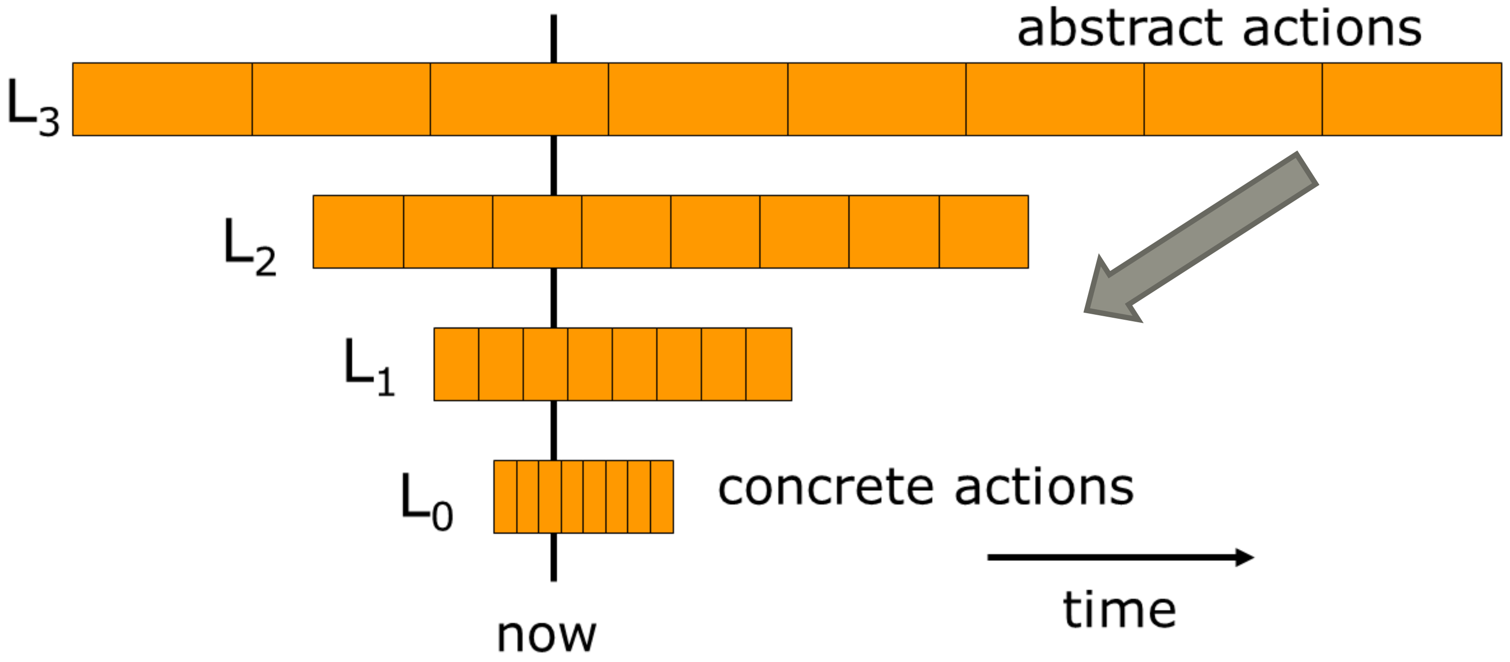
Results



[Mosbach and Behnke, ESANN 2021]

Hierarchical Planning in the Now

- Use predicted state on different layers of abstraction for planning
- Coarse-to-fine planning makes actions more concrete as they come closer to execution
- Plan consists of few steps on each layer



Centauro Robot

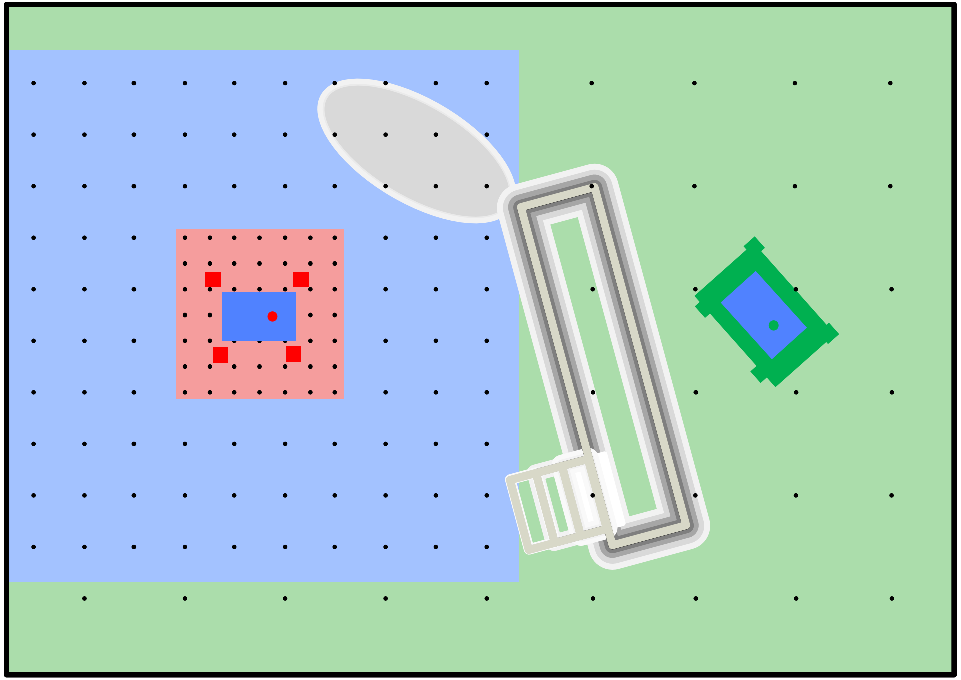
- Serial elastic actuators
- 42 main DoFs
- Schunk hand
- 3D laser
- RGB-D camera
- Color cameras
- Two GPU PCs

[Tsagarakis et al., IIT 2017]

UNIVERSITÄT BONN AIS

Hybrid Driving-Stepping Locomotion Planning: Abstraction

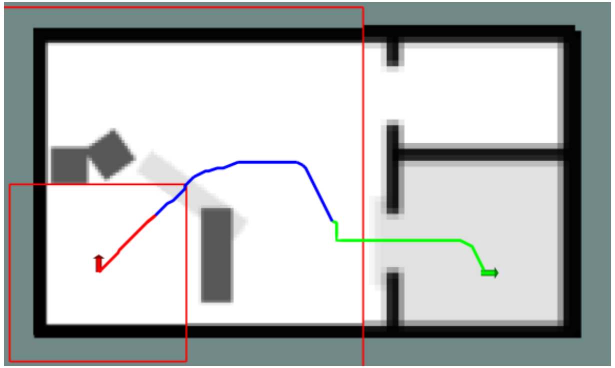
- Planning in the here and now
- Far-away details are abstracted away

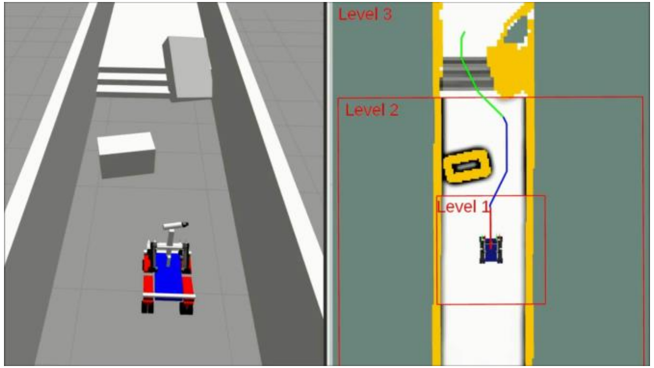


[Klamt and Behnke, IROS 2017, ICRA 2018]

Hybrid Driving-Stepping Locomotion Planning: Abstraction

Level	Map Resolution	Map Features	Robot Representation	Action Semantics
1	• 2.5 cm • 64 orient.	• Height		• Individual Foot Actions
2	• 5.0 cm • 32 orient.	• Height • Height Difference		• Foot Pair Actions
3	• 10 cm • 16 orient.	• Height • Height Difference • Terrain Class		• Whole Robot Actions

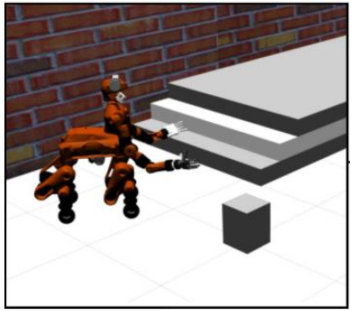




[Klamt and Behnke, IROS 2017, ICRA 2018]

Learning Cost Functions of Abstract Representations

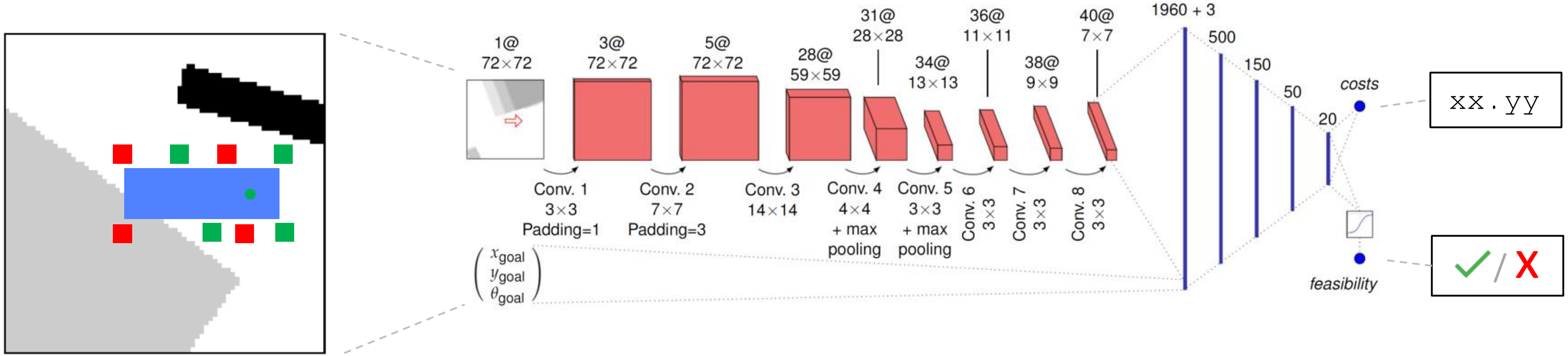
Planning problem



[Klamt and Behnke, ICRA 2019]

Abstraction CNN

- Predict feasibility and costs of local detailed planning

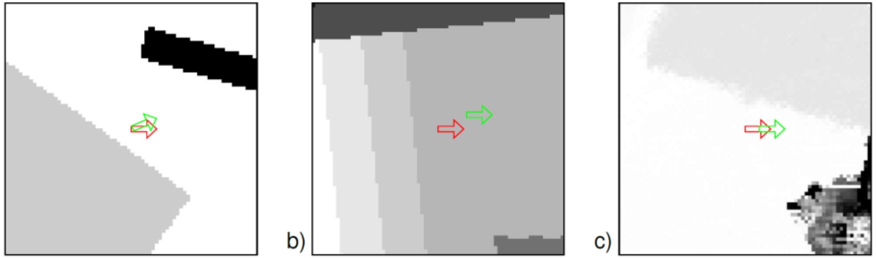


Training data
- generated with random obstacles, walls, staircases
- *costs* and *feasibility* from detailed A*-planner
- ~250.000 tasks

[Klamt and Behnke, ICRA 2019]

Learned Cost Function: Abstraction Quality

- CNN predicts feasibility and costs better than manually tuned geometric heuristics

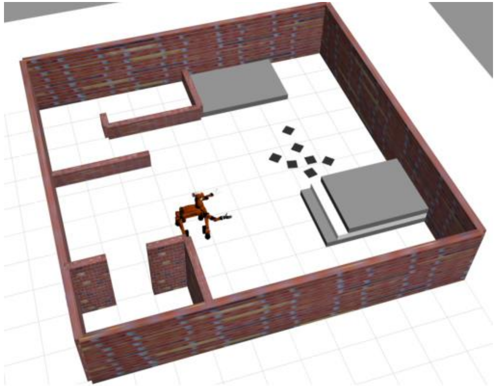


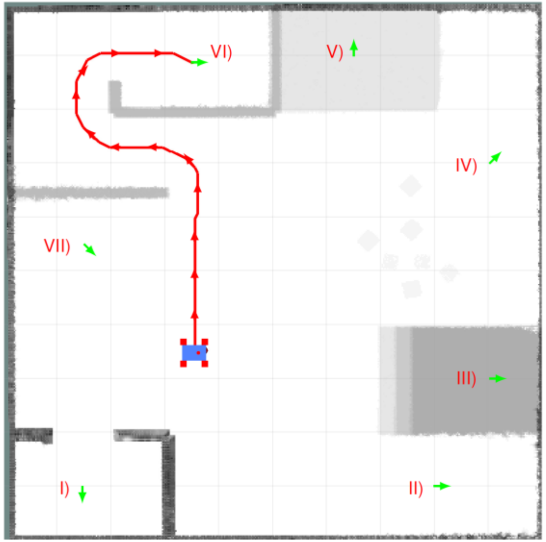
	random	*simulated*	*real*
feasibility correct, man.tuned	79.27%	65.35%	69.77%
Error($\mathcal{C}_{\mathrm{a,man.tuned}}$)	0.057	0.021	0.103
feasibility correct, CNN	95.04%	96.69%	92.62%
Error($\mathcal{C}_{\mathrm{a,CNN}}$)	0.027	0.013	0.081

[Klamt and Behnke, ICRA 2019]

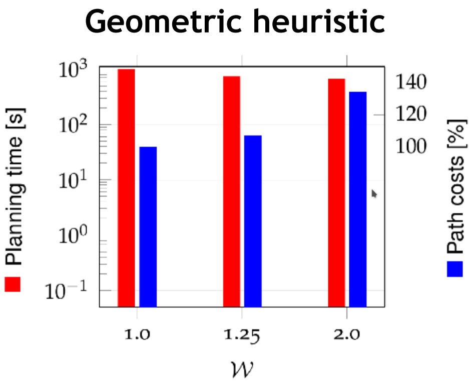
Experiments - Planning Performance

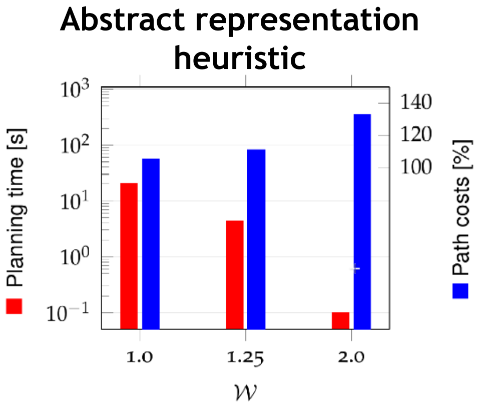
- Learned heuristics accelerates planning, without increasing path costs much





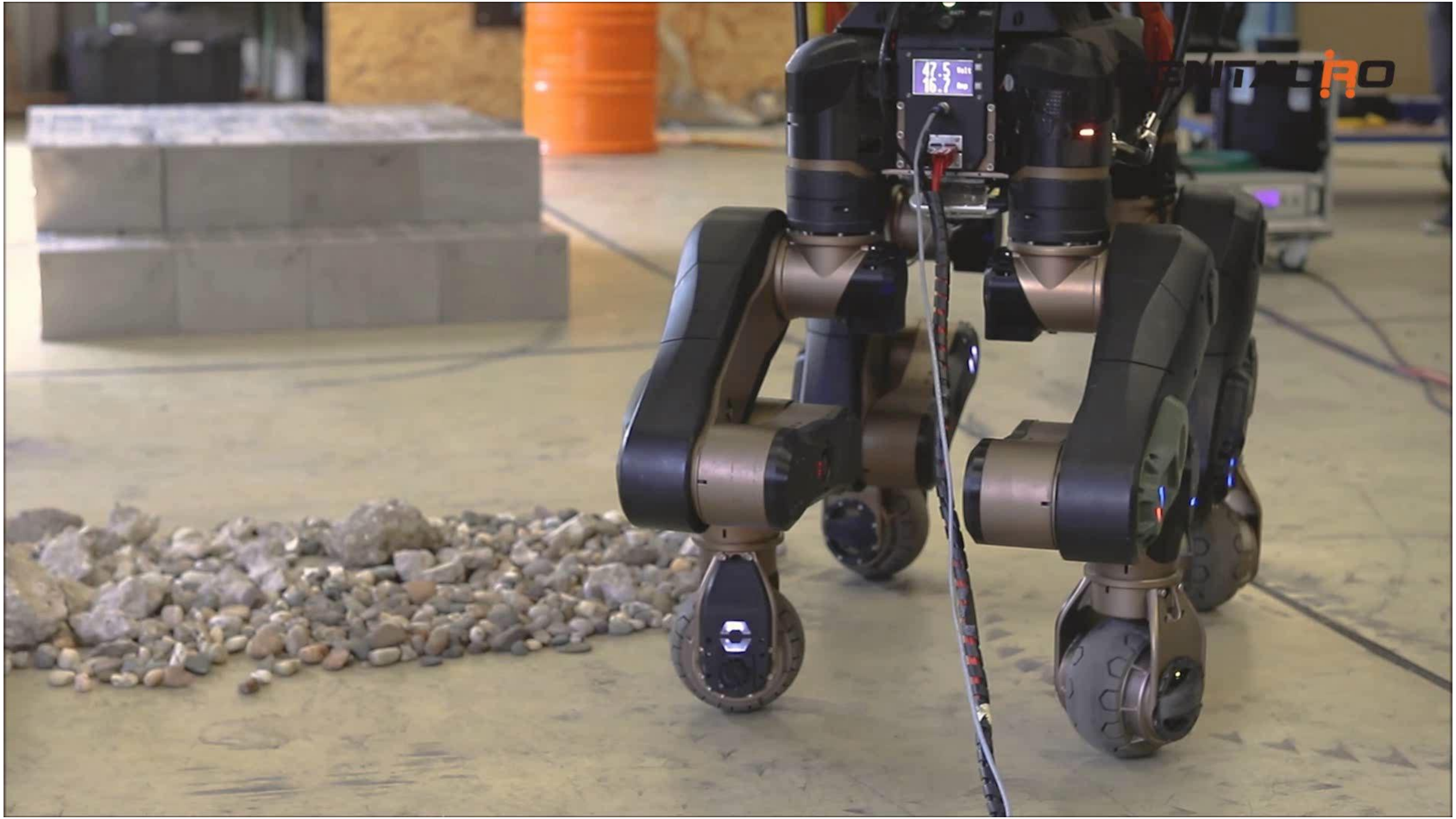
Heuristic preprocessing: 239 sec





[Klamt and Behnke, ICRA 2019]

CENTAURO Evaluation @ KHG: Locomotion Tasks



[Klamt et al. RAM 2019]

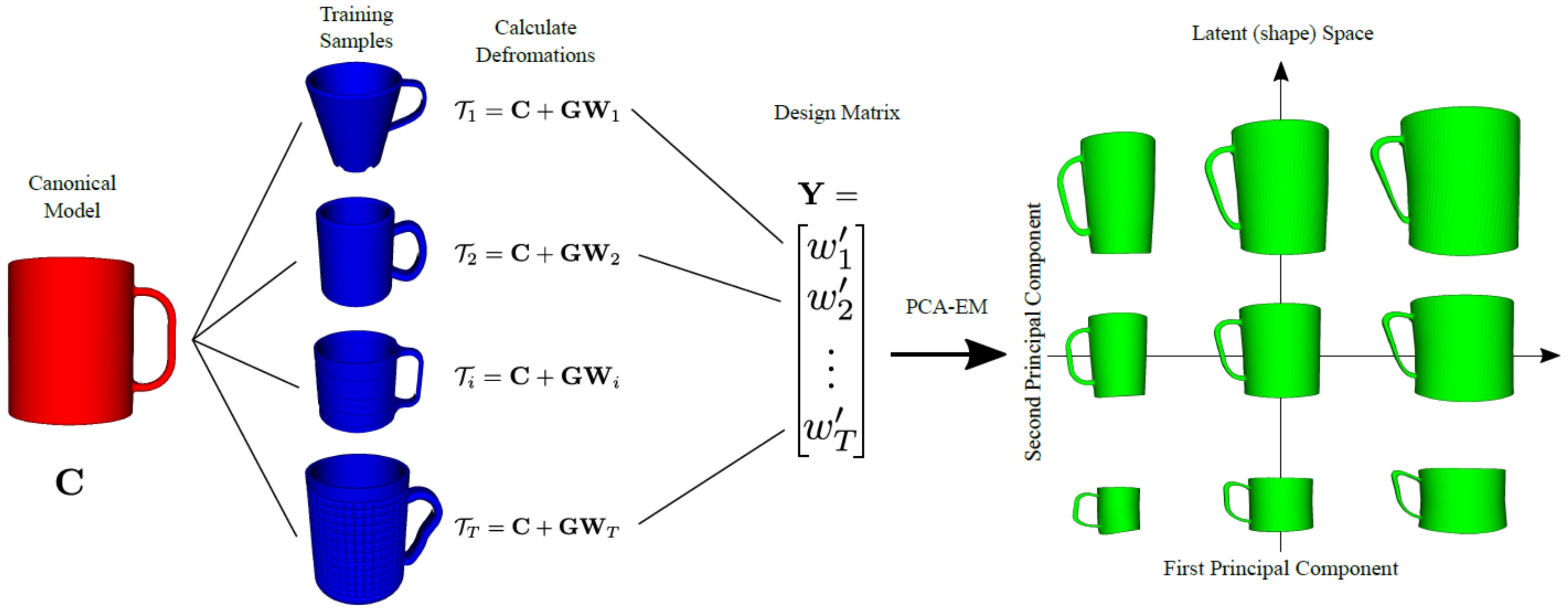
Transfer of Manipulation Skills

Knowledge Transfer

[Rodriguez and Behnke ICRA 2018]

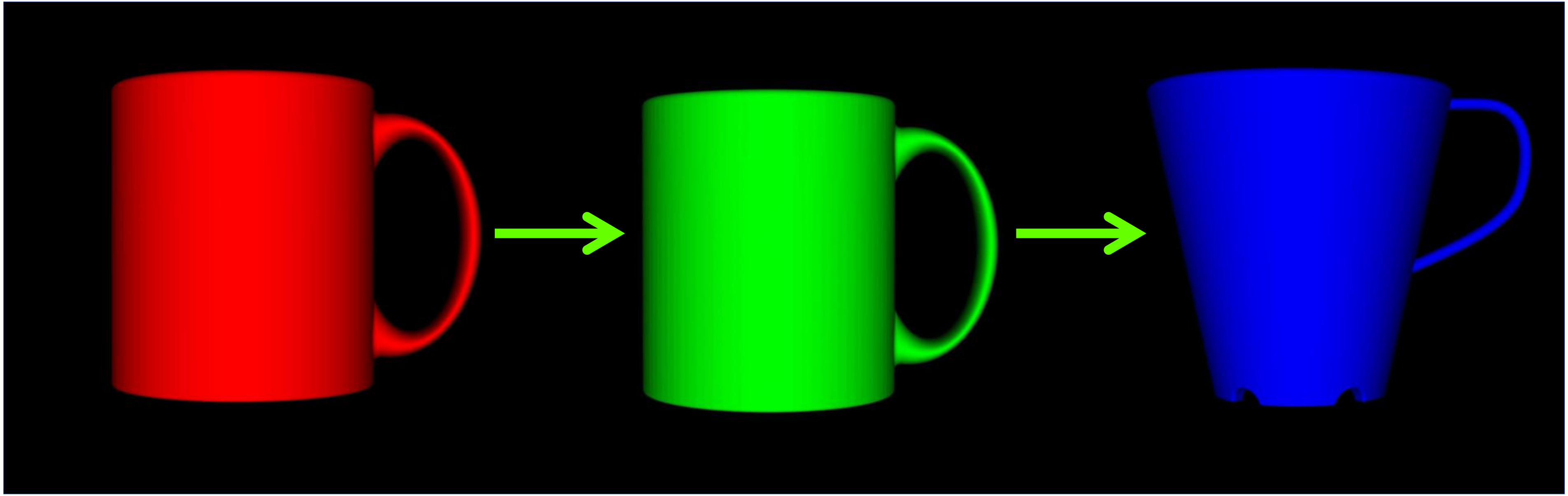
Learning a Latent Shape Space

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



[Rodriguez and Behnke ICRA 2018]

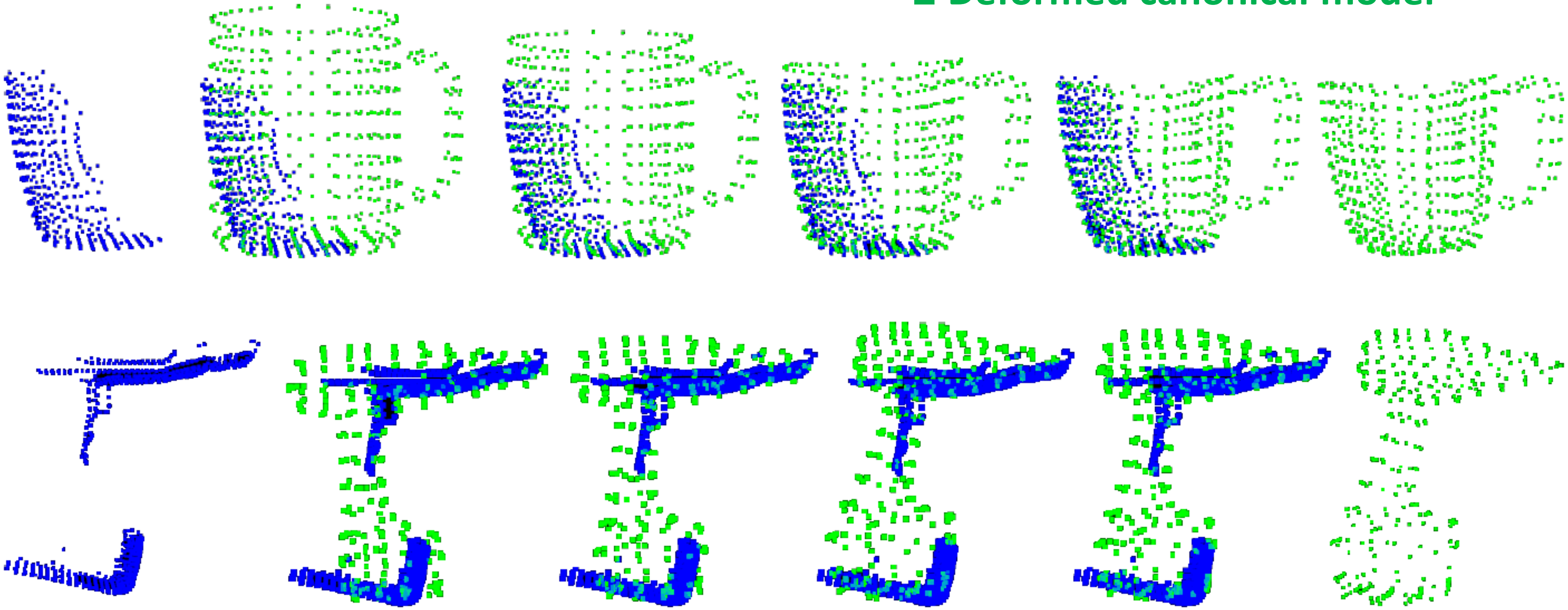
Interpolation in Shape Space



[Rodriguez and Behnke ICRA 2018]

Shape-aware Non-rigid Registration

- Partial view of novel instance
- Deformed canonical model

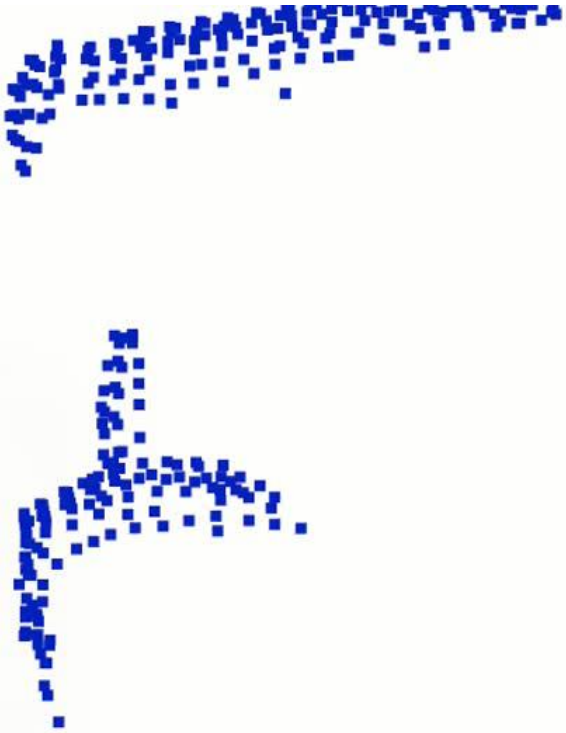


[Rodriguez and Behnke ICRA 2018]

Shape-aware Registration for Grasp Transfer

- Full point cloud

- Partial view

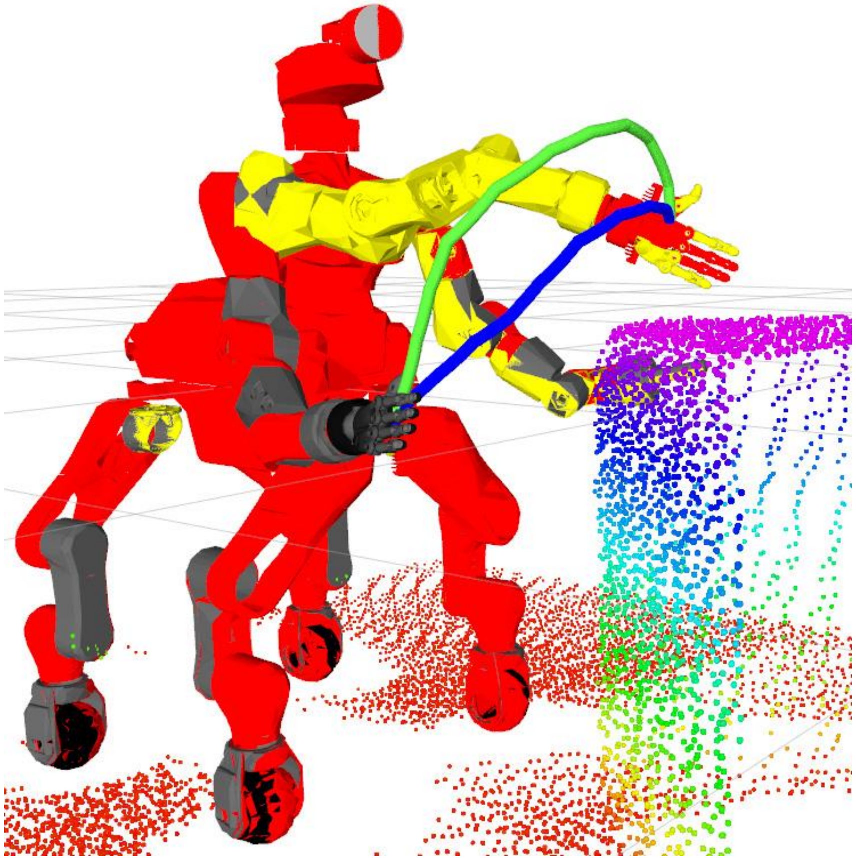


[Rodriguez and Behnke ICRA 2018]

Collision-aware Motion Generation

Constrained Trajectory Optimization:

- Collision avoidance
- Joint limits
- Time minimization
- Torque optimization

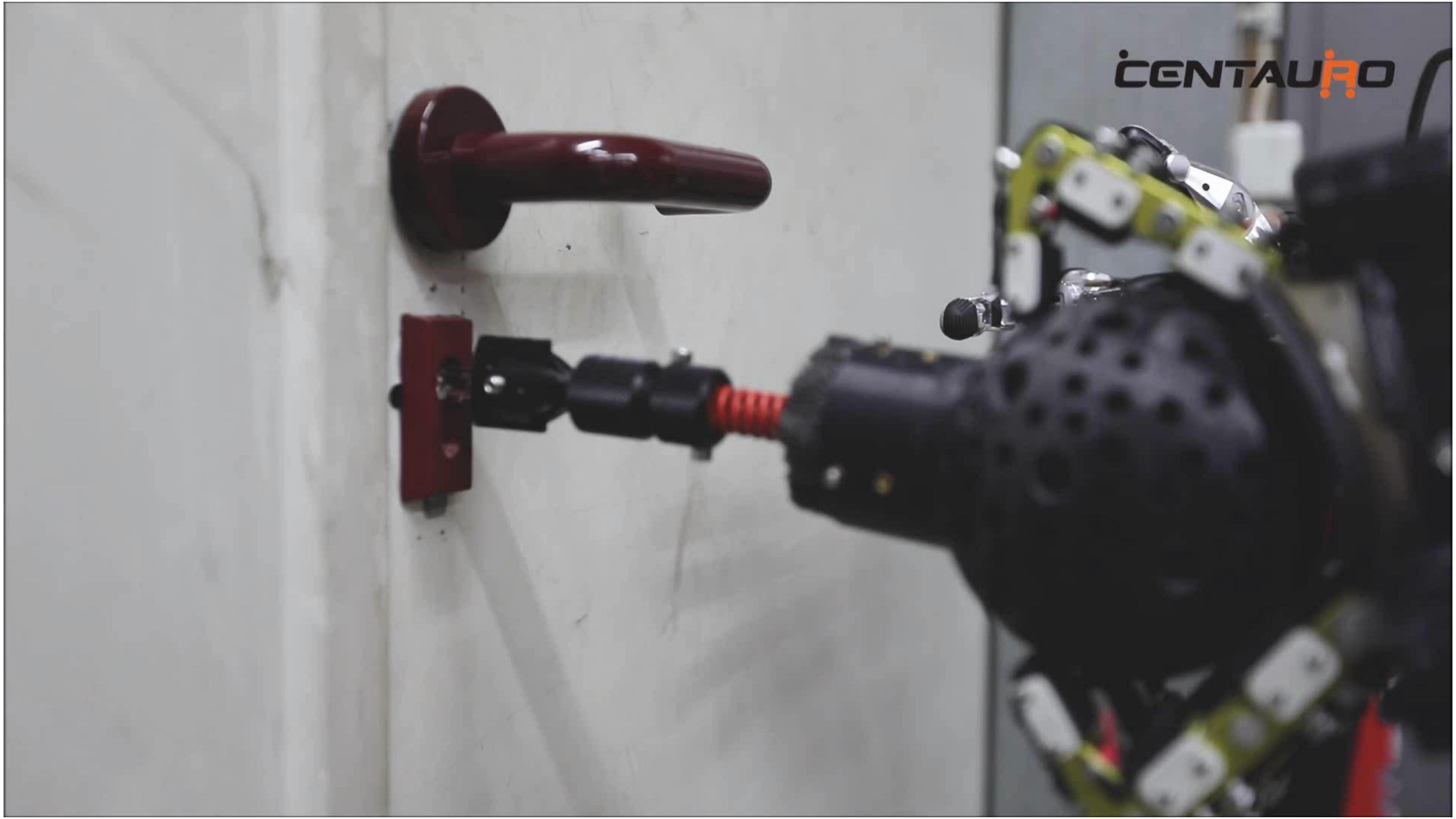


[Pavlichenko et al., IROS 2017]

Grasping an Unknown Power Drill and Fastening Screws

[Rodriguez and Behnke ICRA 2018]

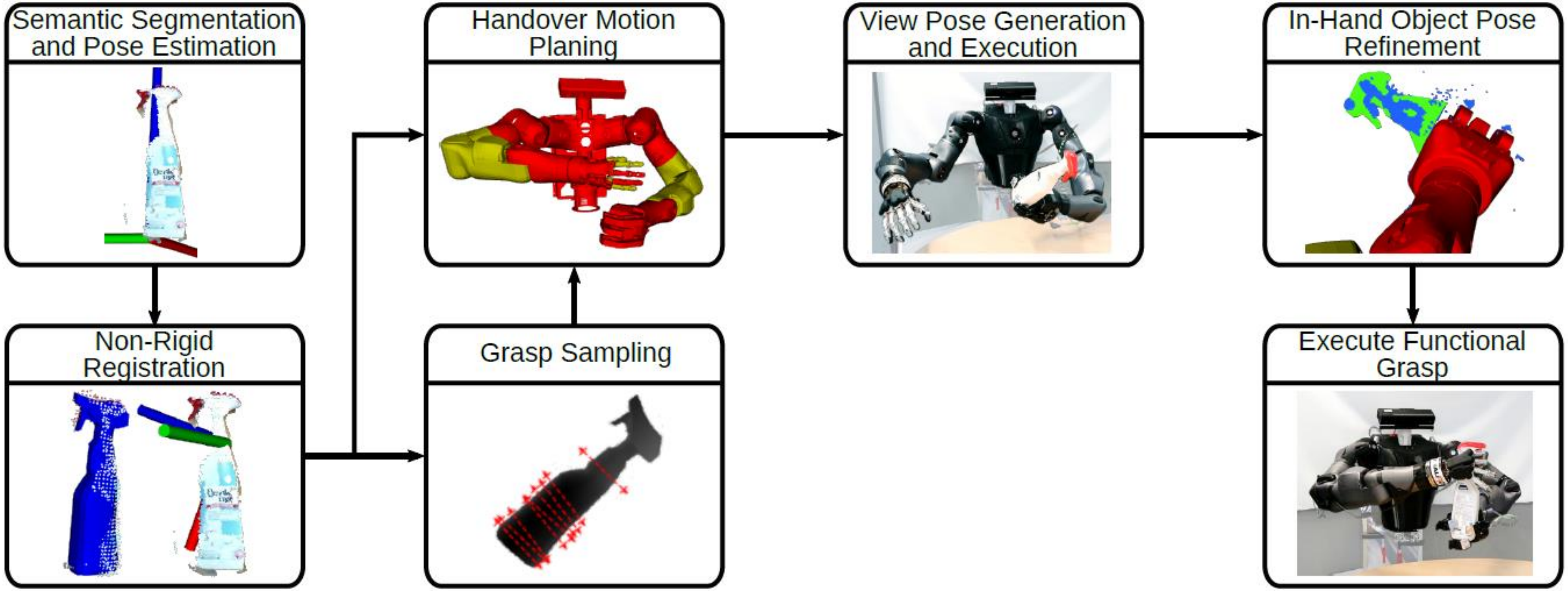
CENTAURO: Complex Manipulation Tasks



[Klamt et al. RAM 2019]

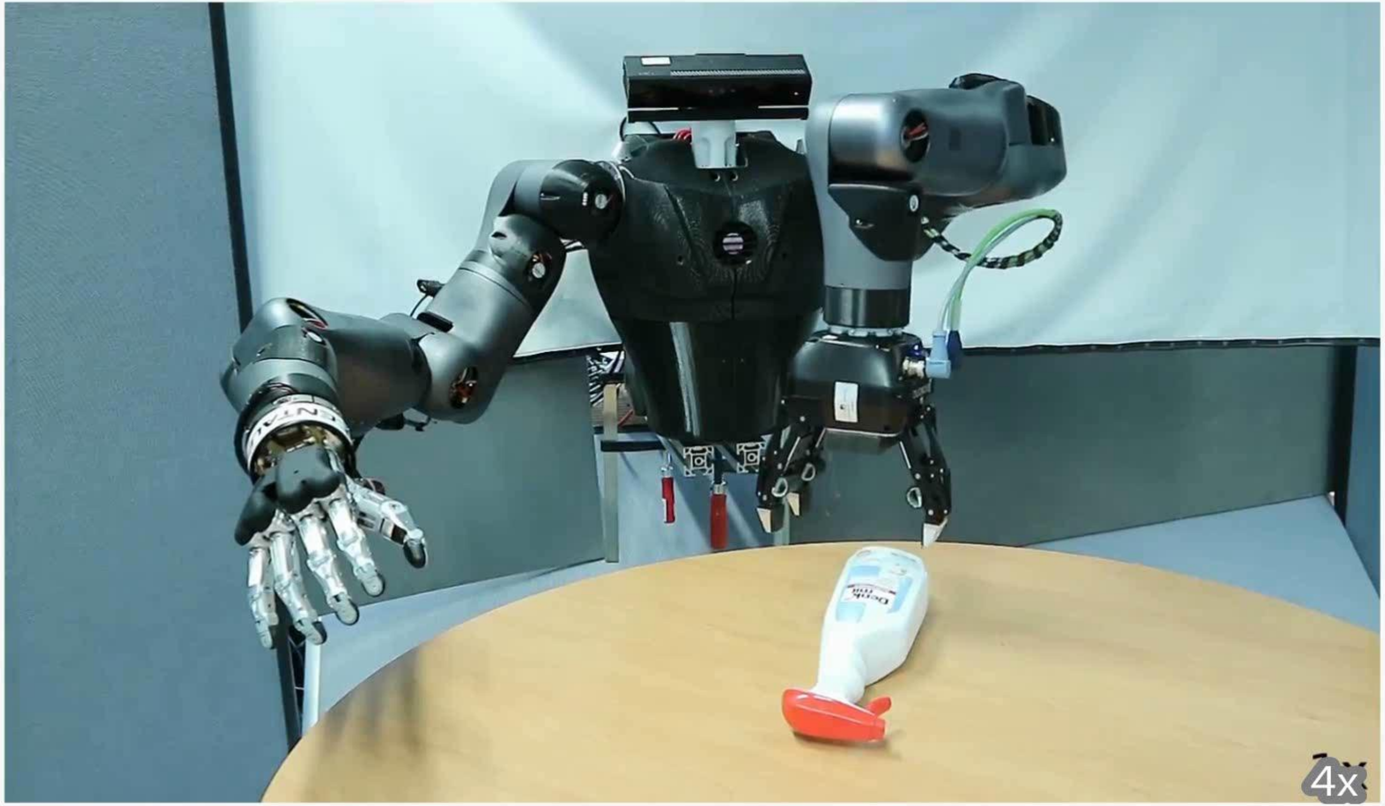
Regrasping for Functional Grasp

- Direct functional grasps not always feasible
- Pick up object with support hand, such that it can be grasped in a functional way



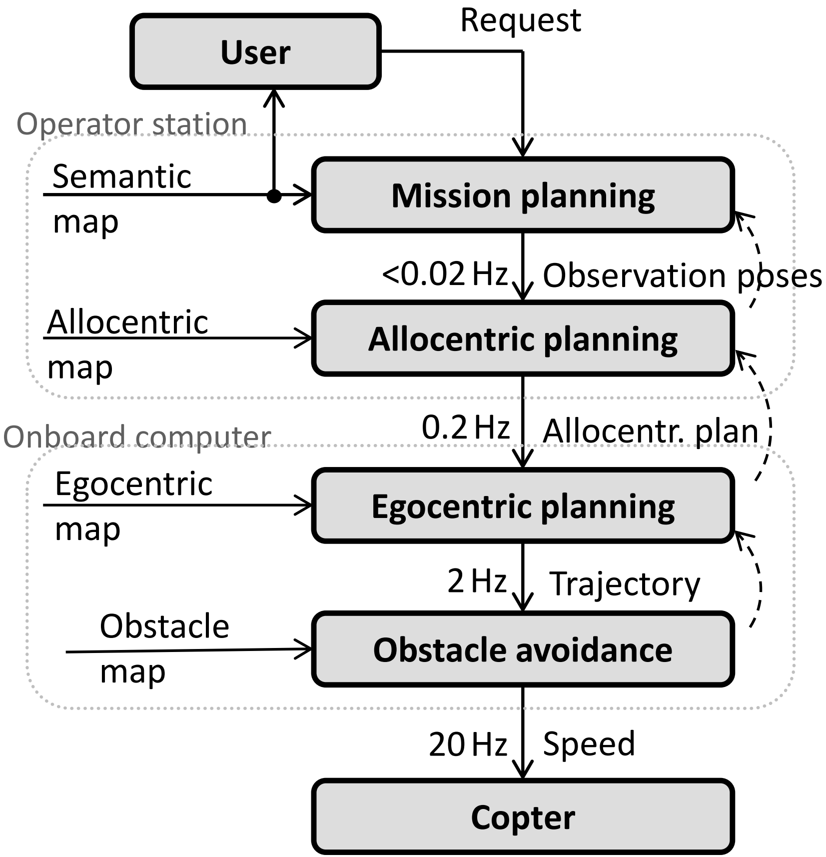
[Pavlichenko et al. Humanoids 2019]

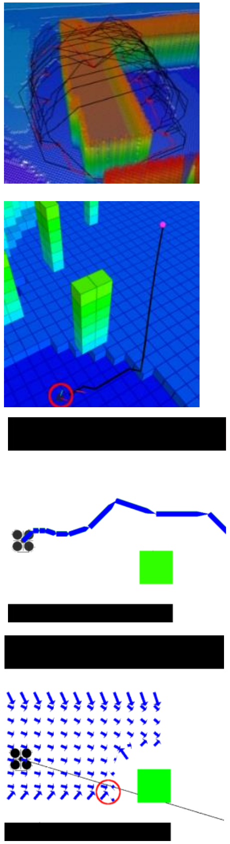
Regrasping Experiments



[Pavlichenko et al. Humanoids 2019]

Micro Aerial Vehicles: Hierarchical Navigation





Mission plan

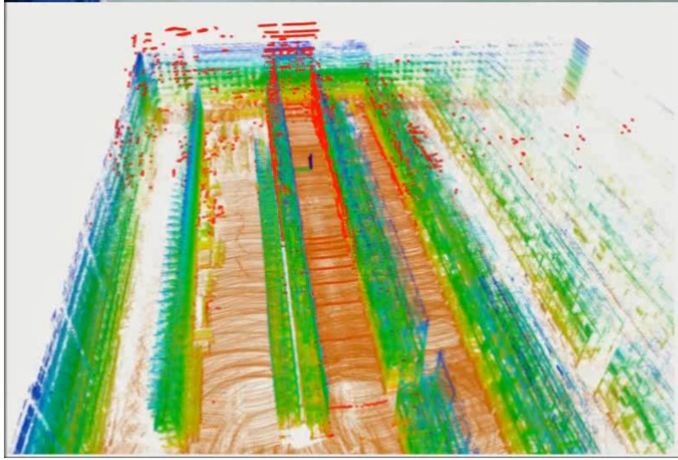
Allocentric planning

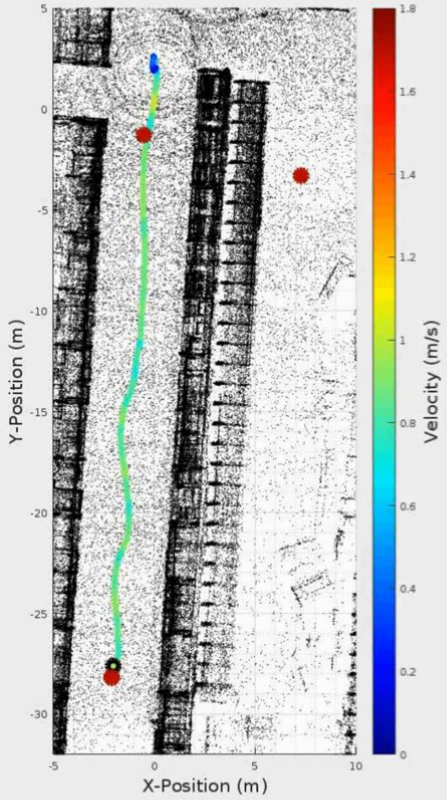
Egocentric planning

Obstacle avoidance

[Droeschel et al. JFR 2016]

InventAIRy: Autonomous Navigation in a Warehouse





[Beul et al. RA-L 2018]

InventAIRy: Detected Tags in Shelf

[Beul et al. RA-L 2018]

German Rescue Robotics Center

DRZ
DEUTSCHES RETTUNGSROBOTIK ZENTRUM

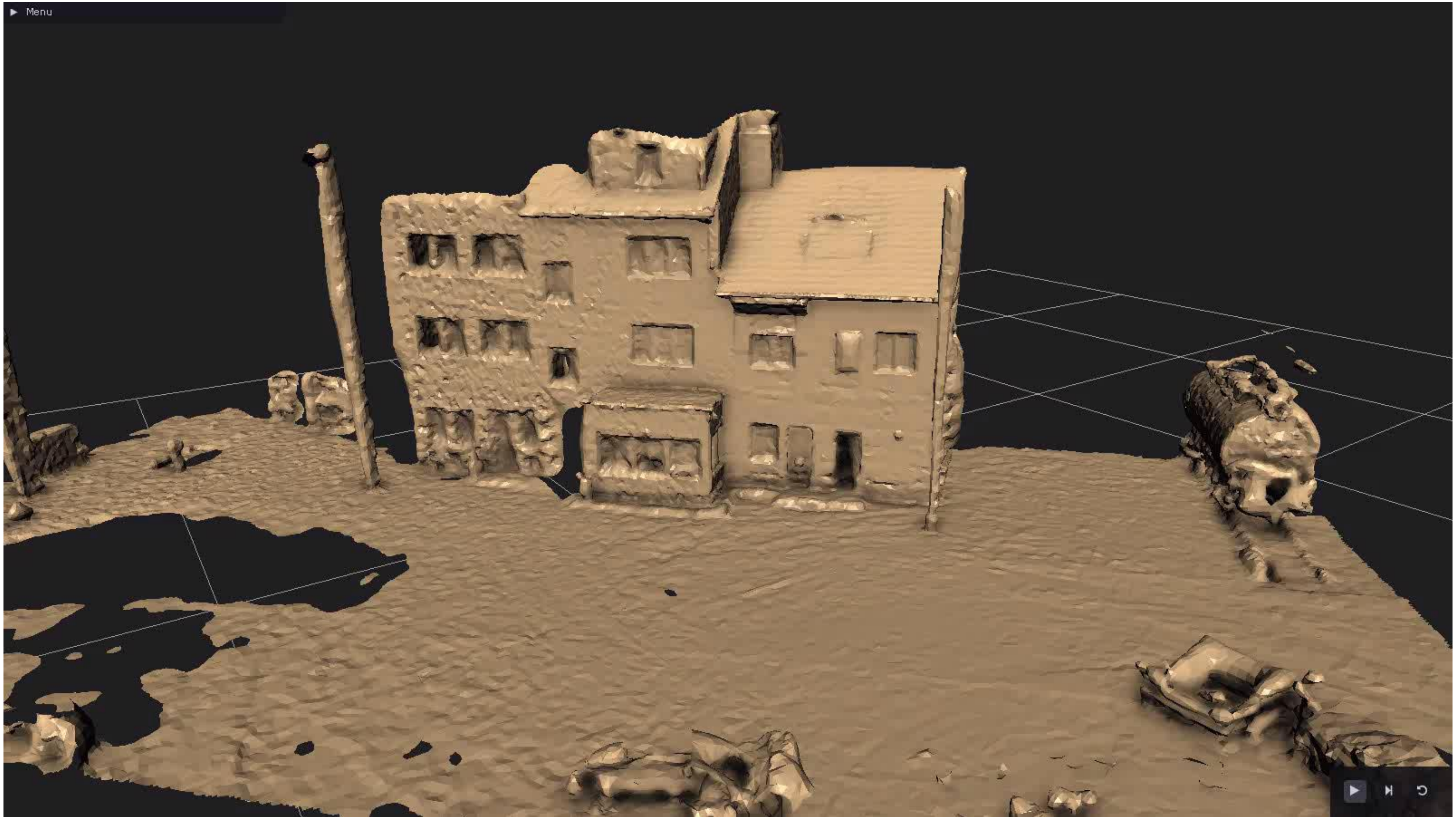
Initial demonstrator

- Basis: DJI Matrice 600 Pro
- Sensors: Velodyne VLP 16, FLIR Boson, 2x FLIR BlackFly S
- Tiltable sensor head

Current demonstrator

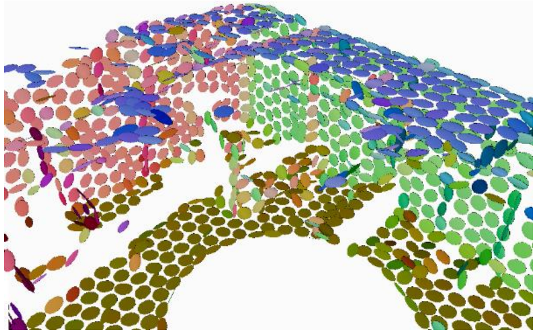
- Basis: DJI Matrice 210 v2
- Sensors: Ouster OS-0, FLIR AGX, 2× Intel RealSense D455
- IP43 water resistance

Modeling the Brandhaus Dortmund

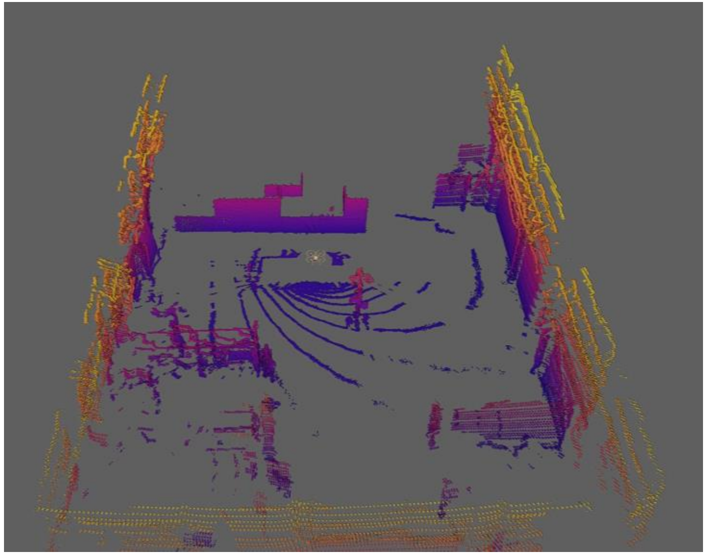


[Rosu et al. SSRR 2019]

Real-time LiDAR Odometry with Continuous-time Trajectory Optimization



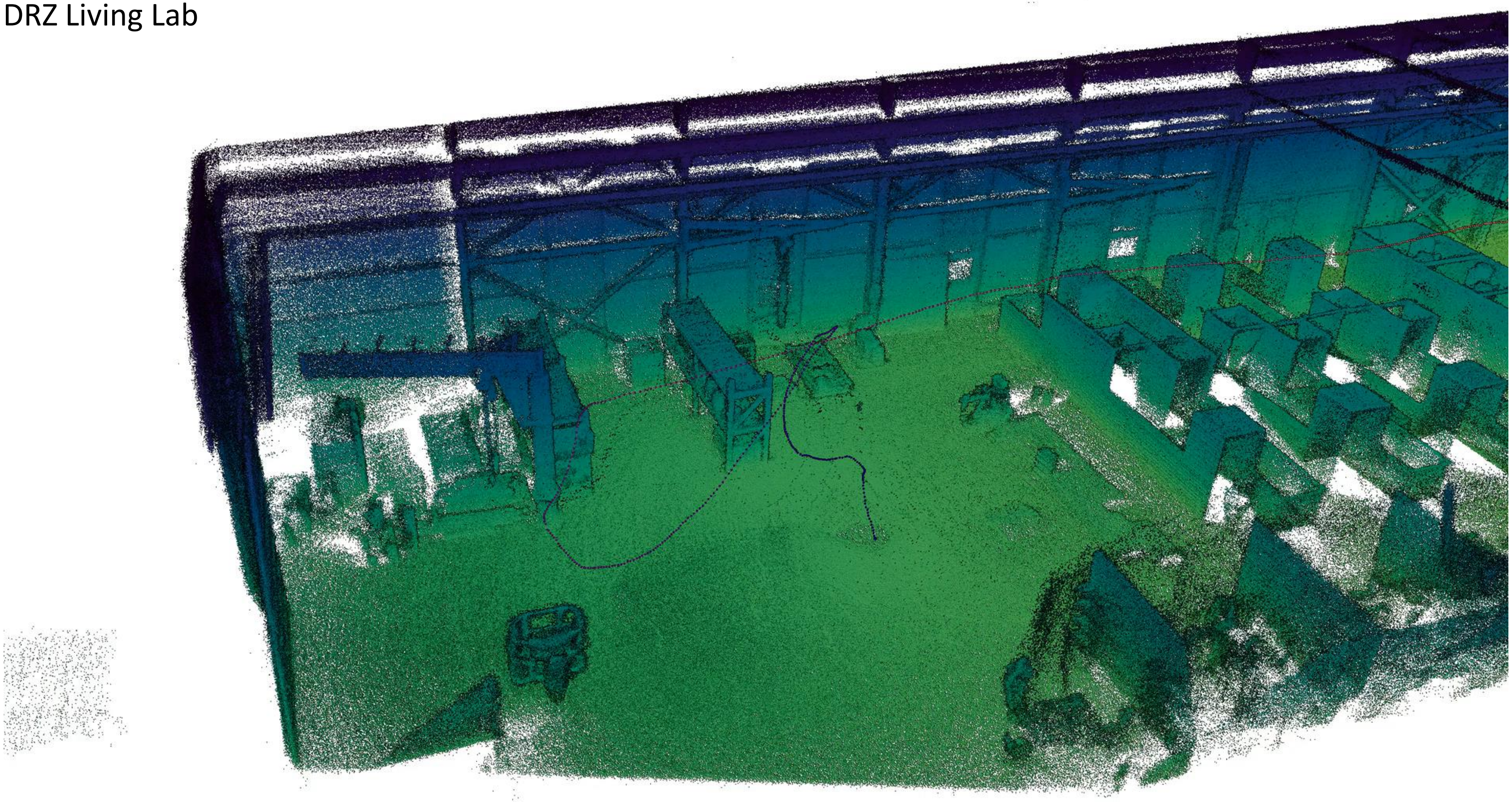
- Simultaneous registration of multiple multiresolution surfel maps using Gaussian mixture models and temporally continuous B-spline
- Accelerated by sparse permutohedral voxel grids and adaptive choice of resolution
- Real-time onboard processing 16-20 Hz
- Open-Source
  https://github.com/AIS-Bonn/lidar_mars_registration



[Quenzel and Behnke, IROS 2021]

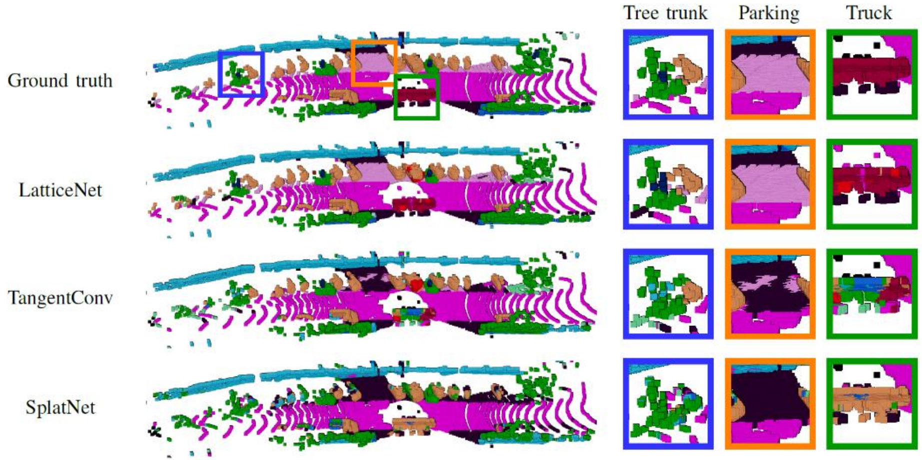
3D LiDAR Mapping

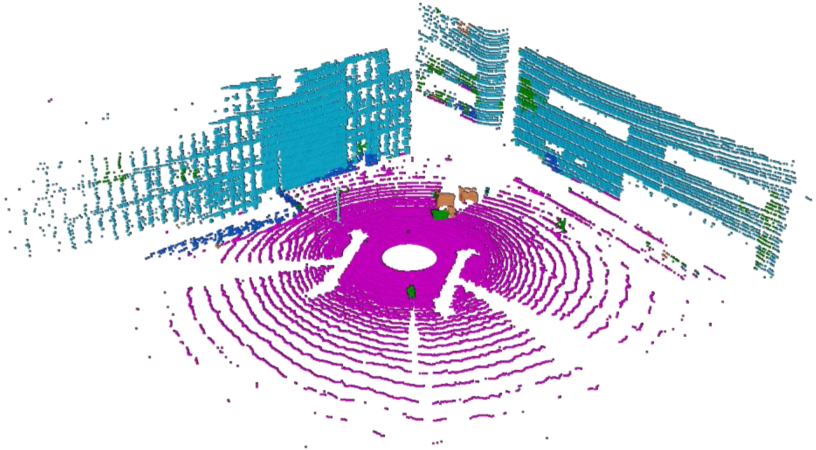
DRZ Living Lab



[Quenzel and Behnke, IROS 2021]

Semantic Perception: LiDAR Segmentation



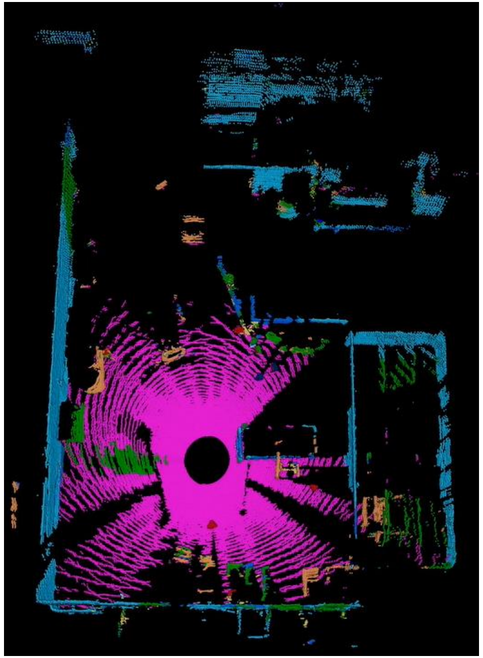


- LatticeNet segmentation of 3D point clouds based on sparse permutohedral grid
- Hierarchical information aggregation through U-Net architecture
- LatticeNet is real-time capable and achieves excellent results in benchmarks

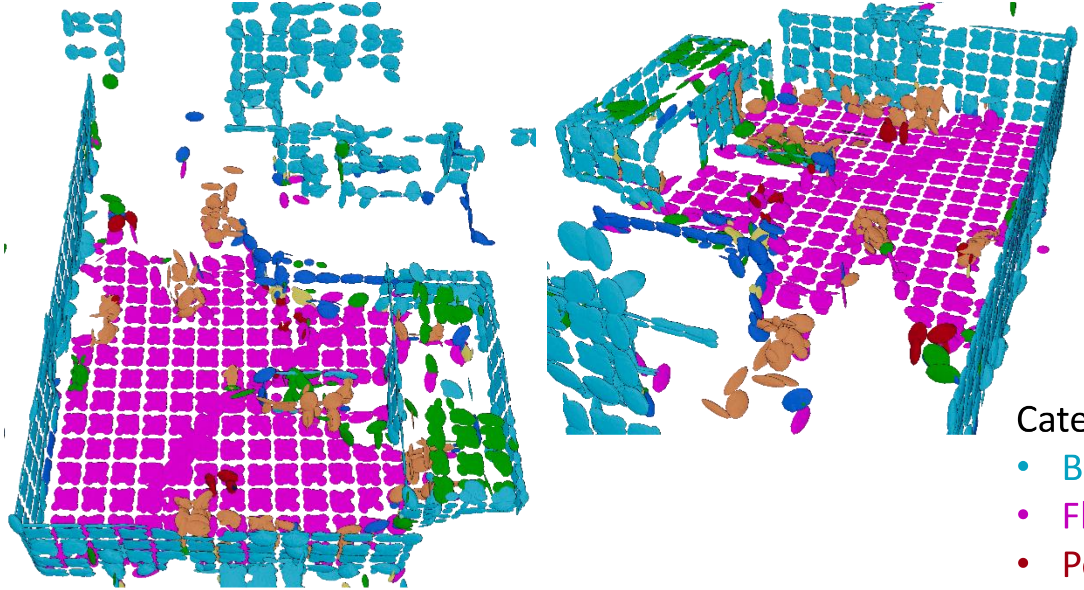
[Rosu et al., RSS 2020]

Semantic Fusion: 3D LiDAR Mapping

Minimax-Viking fire house



Segmented point cloud



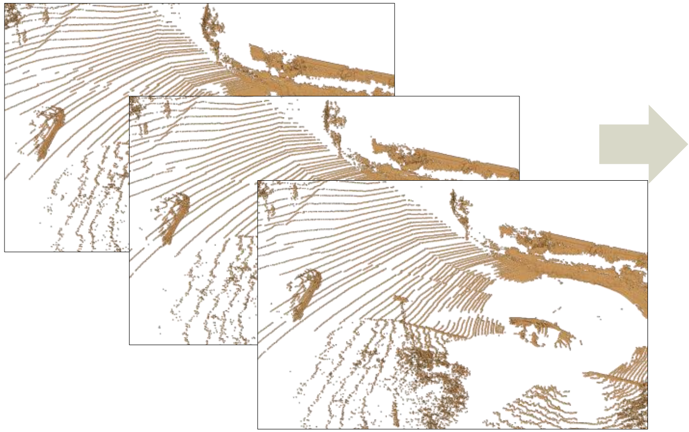
Semantic multiresolution surfel map

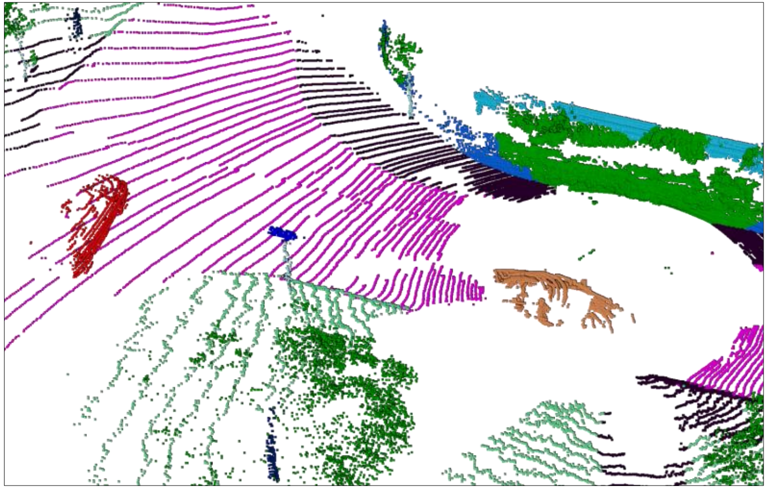
Categories:

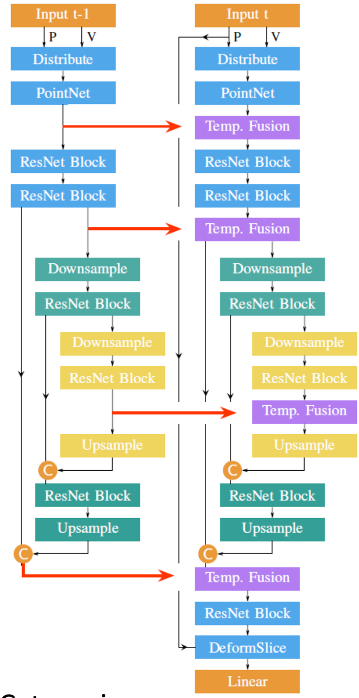
- Building
- Floor
- Persons
- Vehicles
- Fence
- Vegetation

Semantic Fusion: Temporal LatticeNet

- Semantic segmentation of sequences of 3D point clouds
- Integration of recurrent connections
- Trained on three scans of SemanticKITTI
- Distinguishing moving from parking vehicles





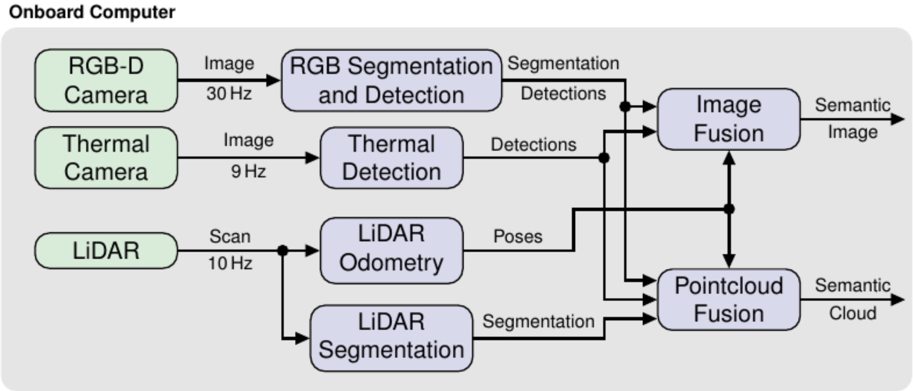


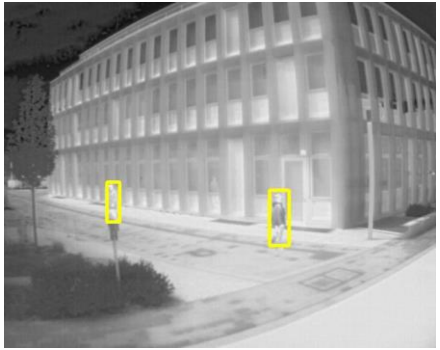
Categories:
- Street
- Moving Vehicle
- Parking Vehicle
- Vegetation

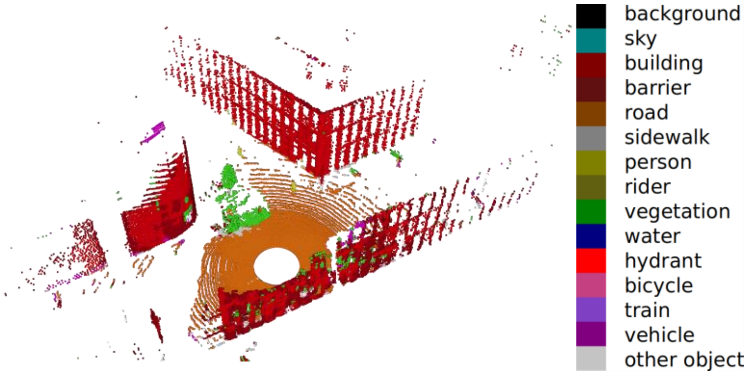
[Rosu et al. Autonomous Robots 2021]

Onboard Multimodal Semantic Fusion

- Real-time semantic segmentation and object detection (≈9Hz) with EdgeTPU / iGPU
  - SalsaNext for LiDAR
  - DeepLabv3 for RGB images
  - SSD MobileDet for Thermal/RGB
- Late-fusion for
  - Point cloud
  - Image segmentation



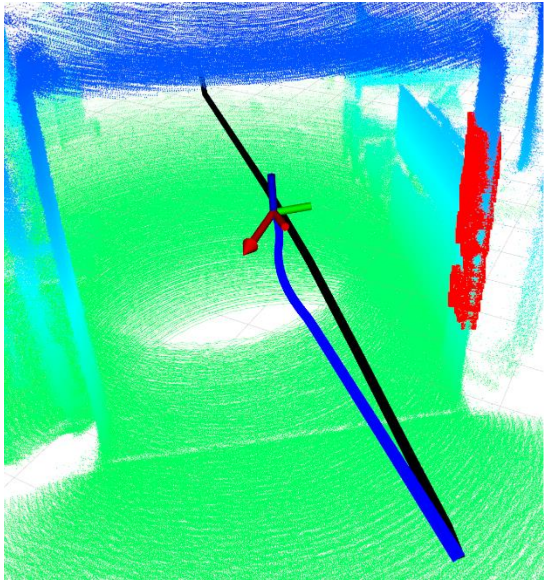


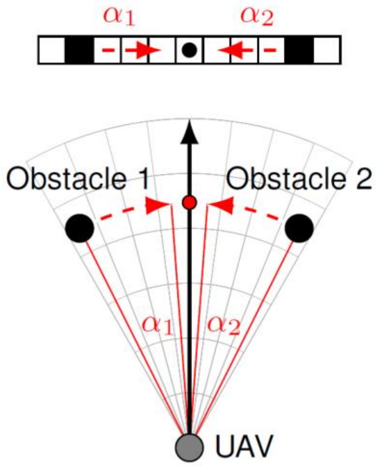


[Bultmann et al. ECMR 2021, RAS 2022]

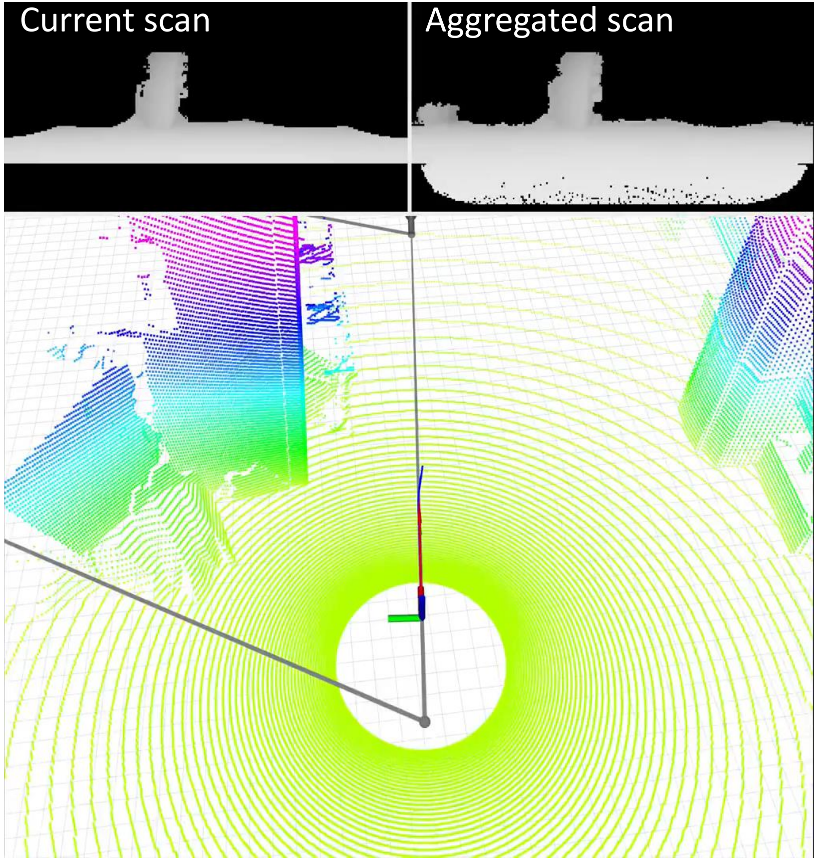
Predictive Angular Potential Field-based Obstacle Avoidance

- Aggregate LiDAR scans in range image
- Adjust direction using angular potential field
- Predict trajectory and range image
- Scale velocity based on time-to-contact





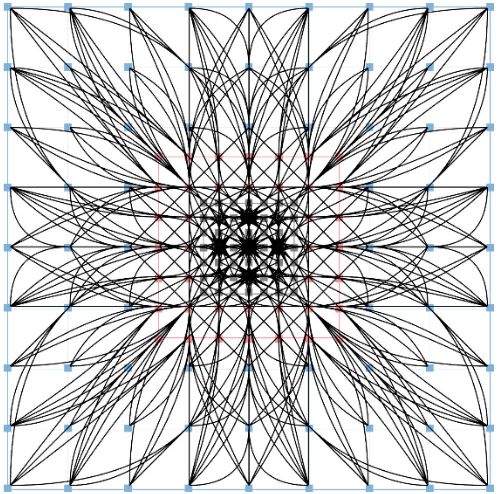
Angular Potential Field

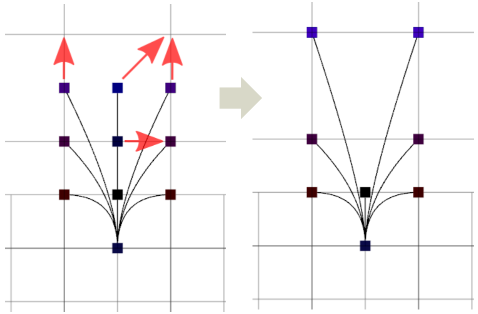


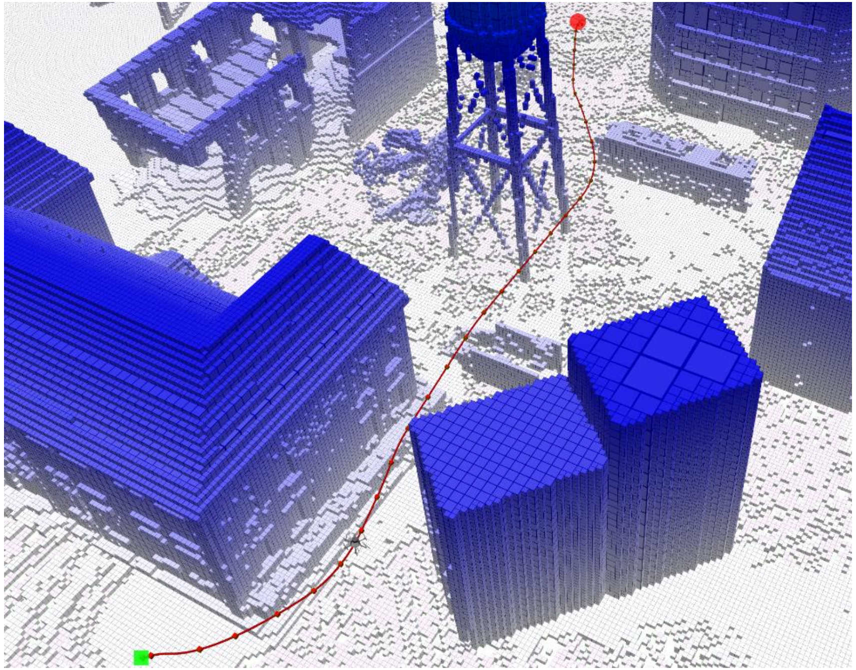
[Schleich and Behnke, IROS 2022]

Dynamic 3D Navigation Planning

- Positions and velocities in sparse local multiresolution grid
- Adaptation of movement primitives to grid
- Optimization of flight time and control costs
- 1 Hz replanning



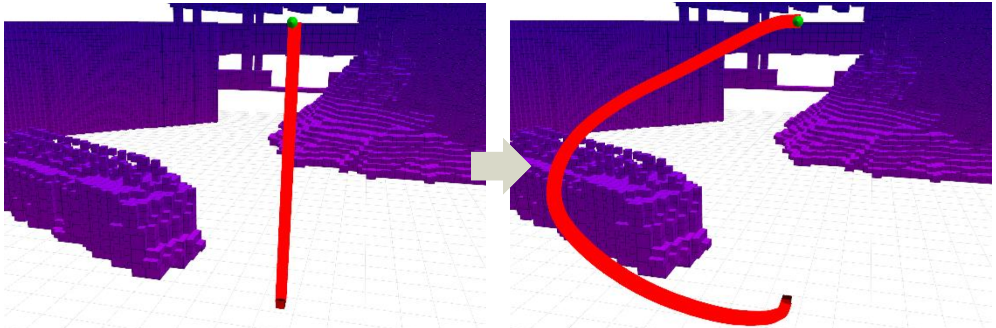




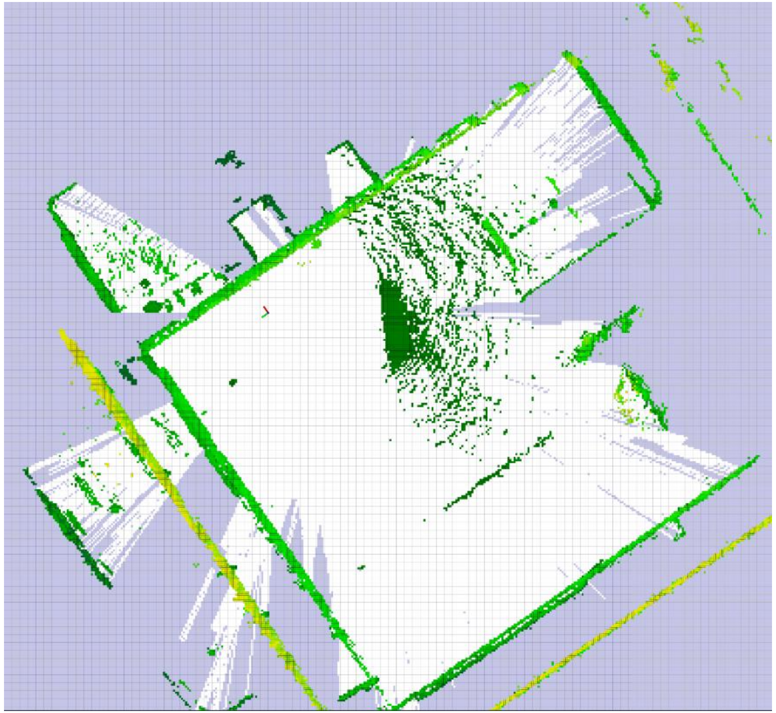
[Schleich and Behnke, ICRA 2021]

Planning with Visibility Constraints

- Extra costs for flight through unmapped volumes
- Consideration of sensor frustum:
  - Coupling of vertical and horizontal motion
  - Preferred forward flight with limited rotational speed



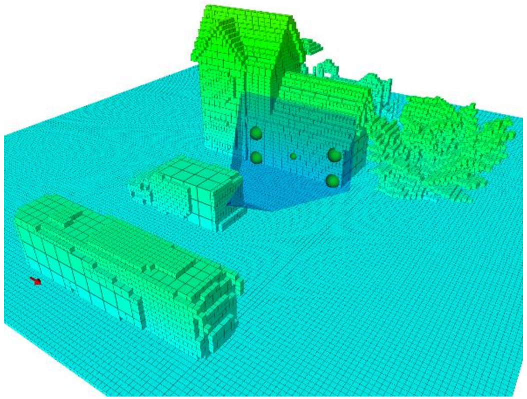
Initial trajectory   Optimized trajectory

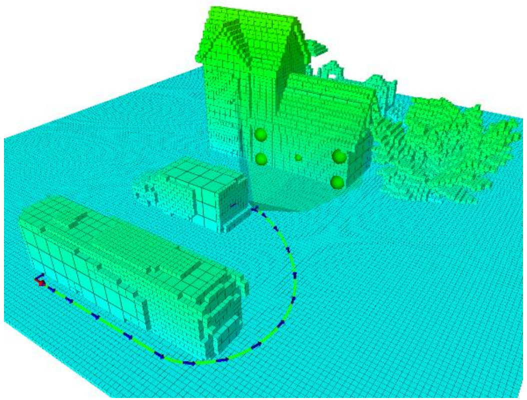


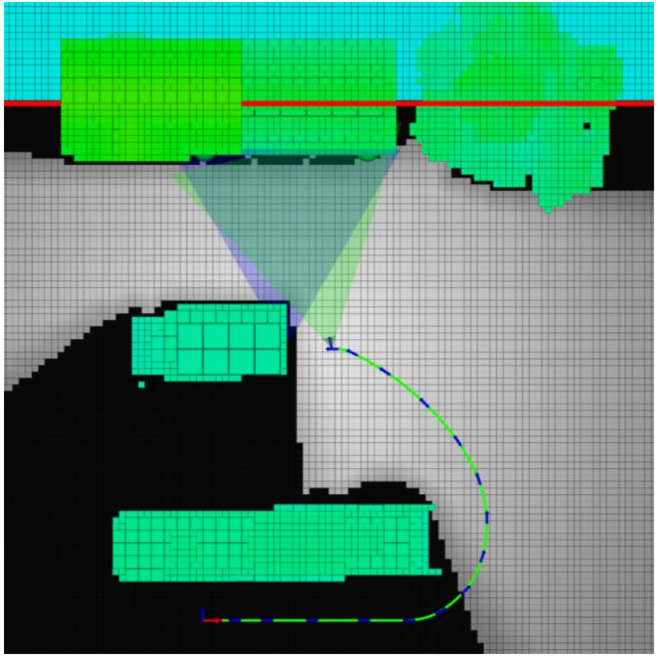
Obstacle map

Observation Pose Planning

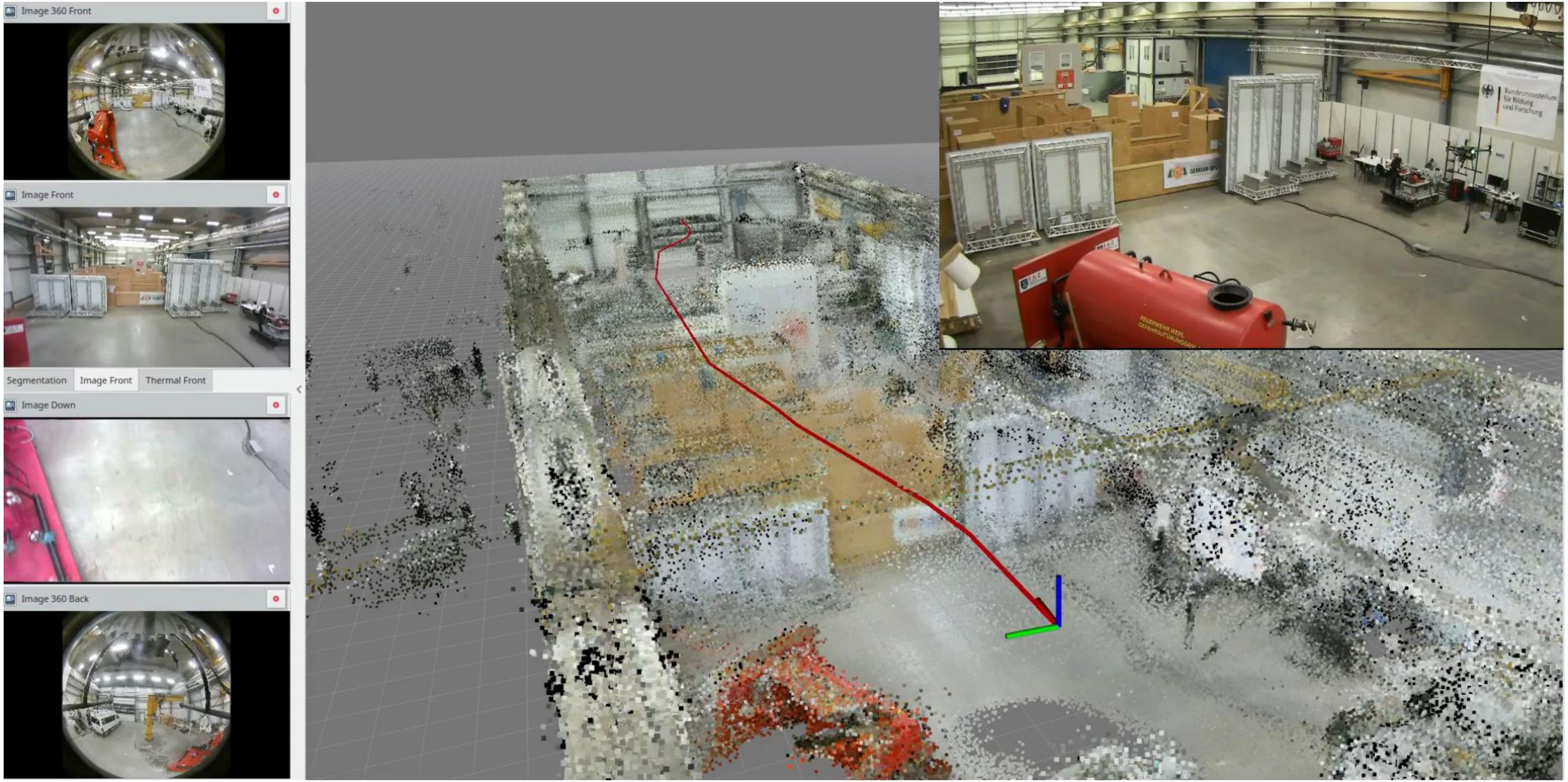
- Planning of observation poses with line of sight to the target object despite occlusions
- Target objects are defined by position, line of sight and distance
- Optimization of observation poses with regard to visibility quality and accessibility


Initial observation pose


Optimized path


Top-down view

Autonomous Flight without GNSS



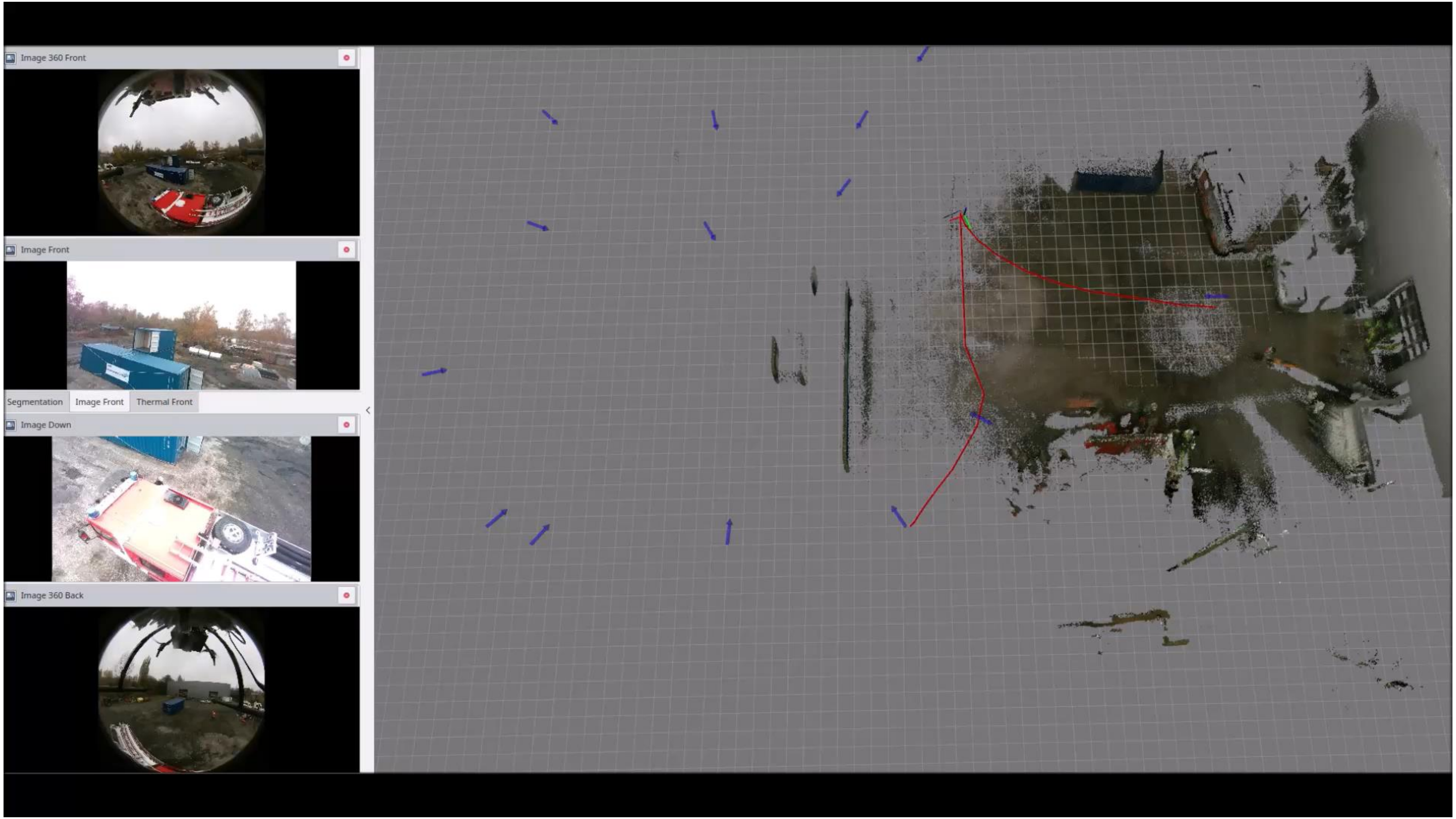
DRZ Dortmund

Exploration

- Definition of target area w.r.t. satellite images or maps
- Simple exploration patterns (spirals, meanders, …)
- Collision check
- TSP to determine segment sequence
- Continuous replanning

Campus Poppelsdorf

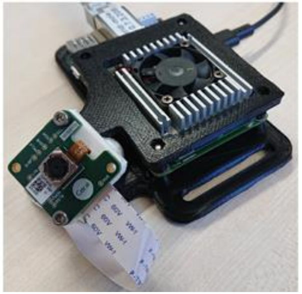
Autonomous Exploration

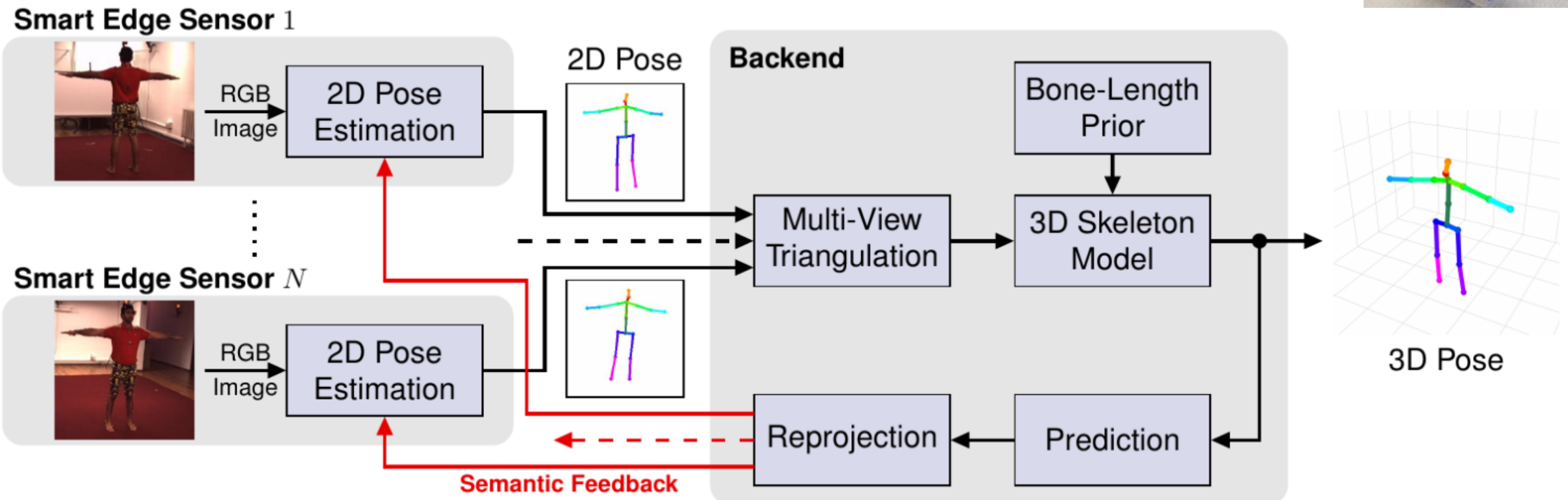


DRZ Dortmund

Real-Time Multi-View 3D Human Pose Estimation using Semantic Feedback to Smart Edge Sensors

- Triangulation and skeleton model to recover 3D pose
- Semantic feedback channel for bidirectional communication between backend and sensors

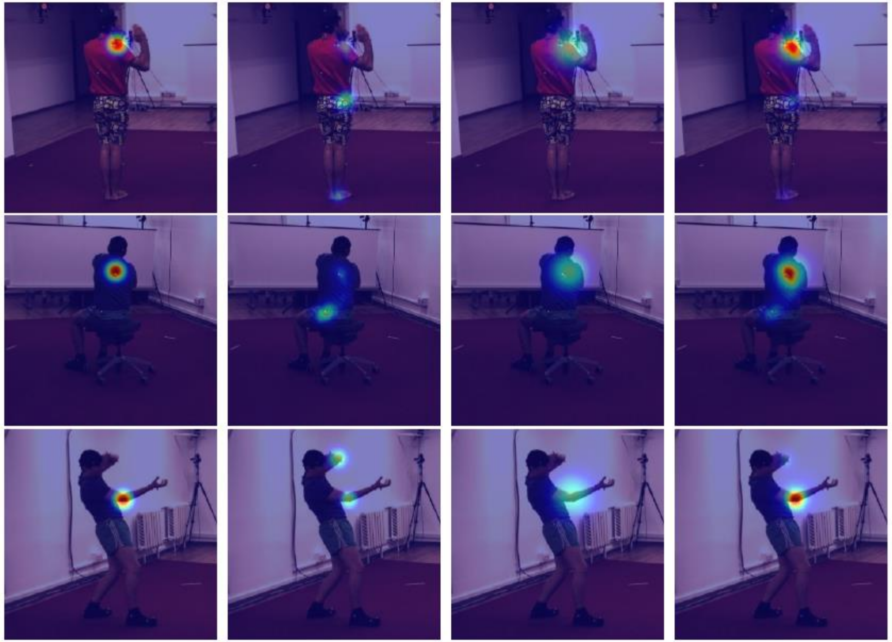




[Bultmann and Behnke, RSS 2021]

Real-Time Multi-View 3D Human Pose Estimation using Semantic Feedback to Smart Edge Sensors

- Feedback heatmap is rendered from feedback skeleton and fused with detection on sensors
- Feedback heatmap helps to recover from incorrect or imprecise 2D joint detections
- Examples:
  - Occluded left wrist (rows 1 and 2)
  - Confusion of left and right elbow (row 3)

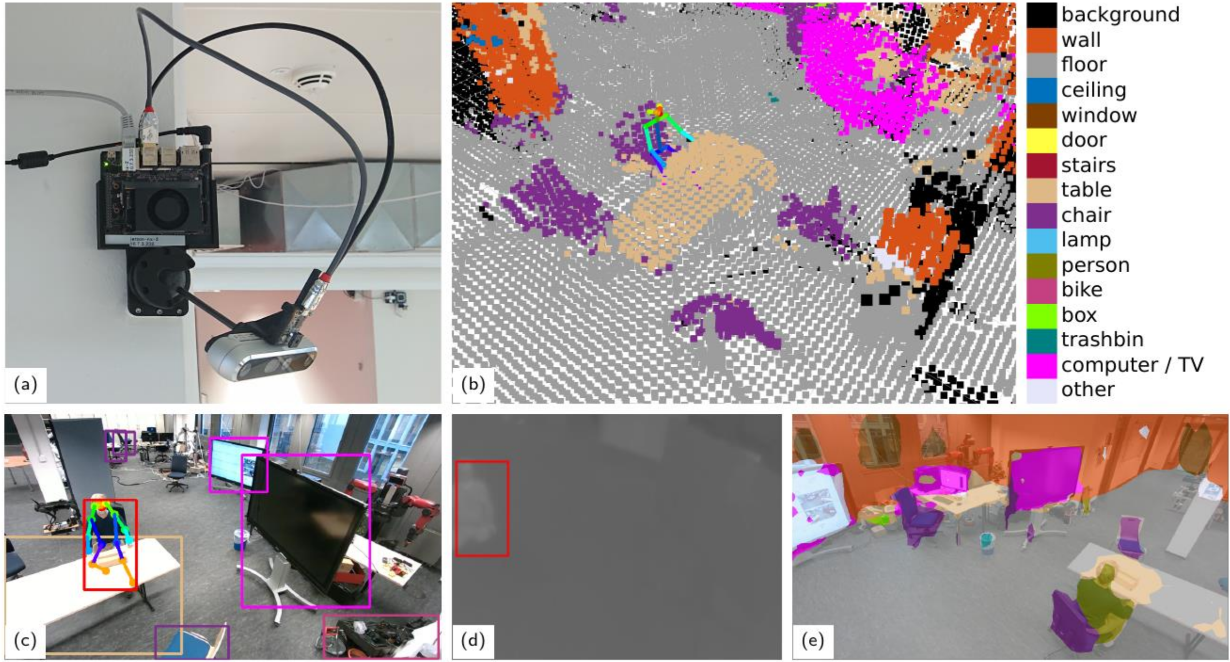


(a) ground-truth (b) detected (c) feedback (d) fused

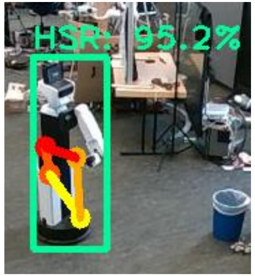
[Bultmann and Behnke, RSS 2021]

Semantic Perception with Smart Edge Sensor Network

- Object detection and semantic segmentation of RGB images
- Person detection in IR images
- Semantic labelling of RGB-D point clouds
- Pose estimation for mobile robot and chairs



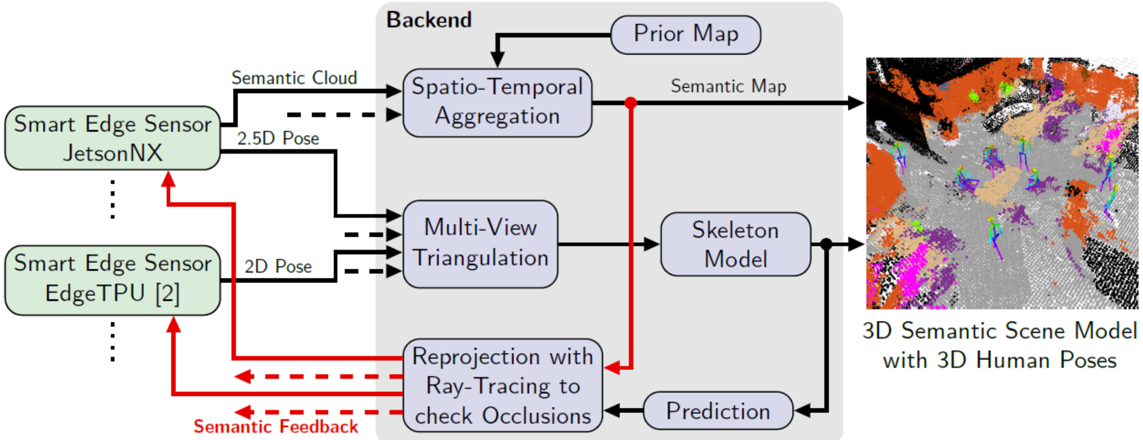
(a) Smart Edge Sensor with Jetson NX (b) 3D semantic scene model,
(c) RGB and (d) thermal detections, (e) semantic segmentation

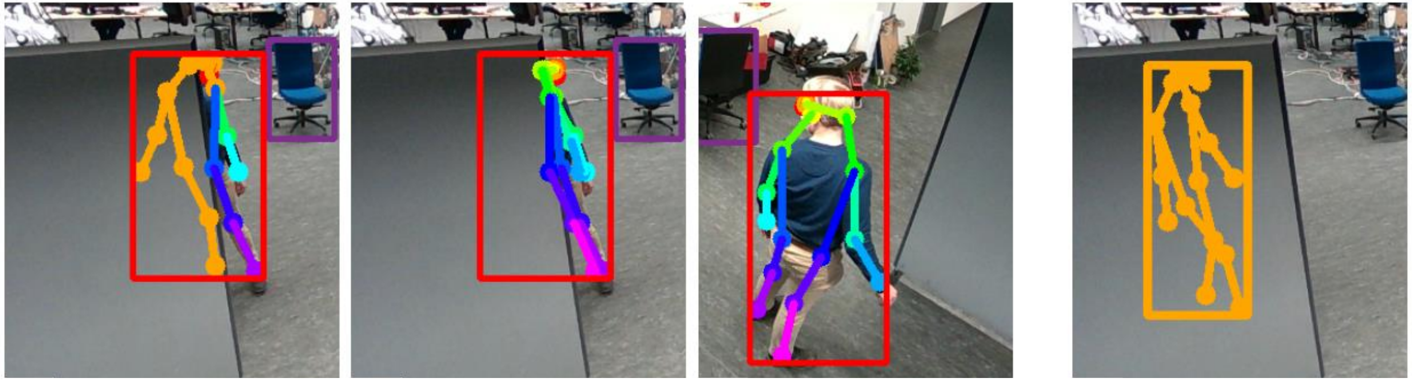


[Bultmann and Behnke: IAS 2022]

3D Human Pose Estimation with Occlusion Feedback

- Heavy occlusion causes the pose estimation to collapse to the visible side only
- With occlusion feedback occluded joint detections can be discarded and the local model is completed



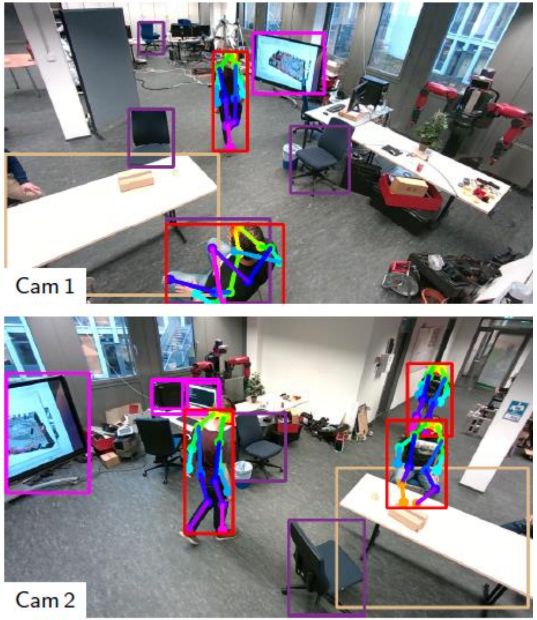


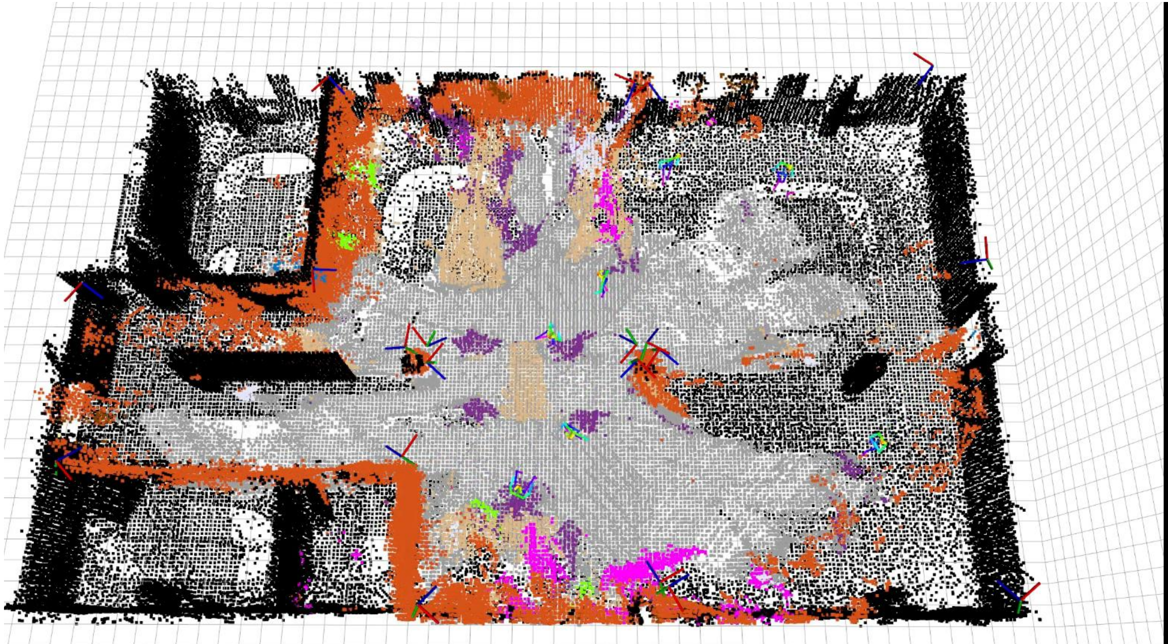
With occlusion feedback | W/o occlusion feedback | Unoccluded reference | Fully occluded

[Bultmann and Behnke: IAS 2022]

Evaluation in Real-World Multi-Person Scenes

- 20 smart edge sensors (4 Jetson NX, 16 Edge TPU), covering 12×22 m area
- Experiments with 8 persons moving through the scene





The sensor network provides a complete 3D semantic scene view and estimates dynamic 3D poses of multiple persons in real time.

[Bultmann and Behnke: IAS 2022]

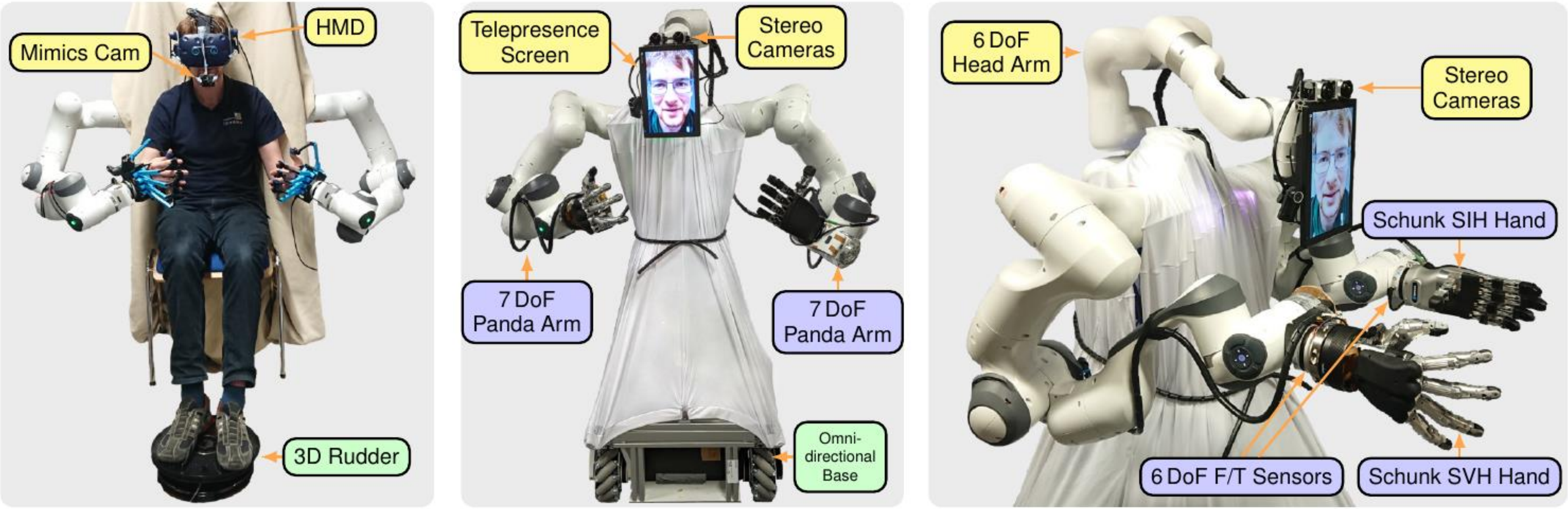
ANA Avatar XPRIZE Competition

- Requires mobility, manipulation, human-human interaction
- Focuses on the immersion in the remote environment and the presence of the remote operator

NimbRo Avatar

ANA AVATAR XPRIZE

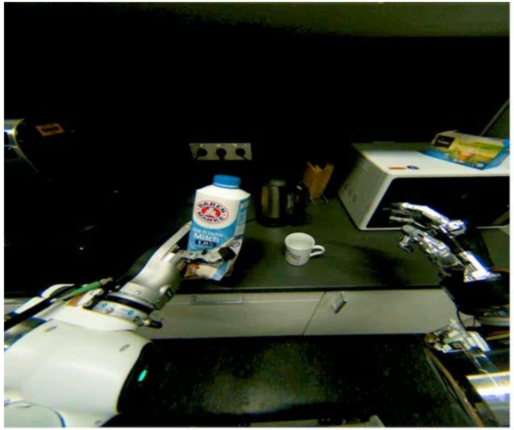
- Two-armed avatar robot designed for teleoperation with immersive visualization & force feedback
- Operator station with HMD, exoskeleton and locomotion interface



[Schwarz et al. IROS 2021]

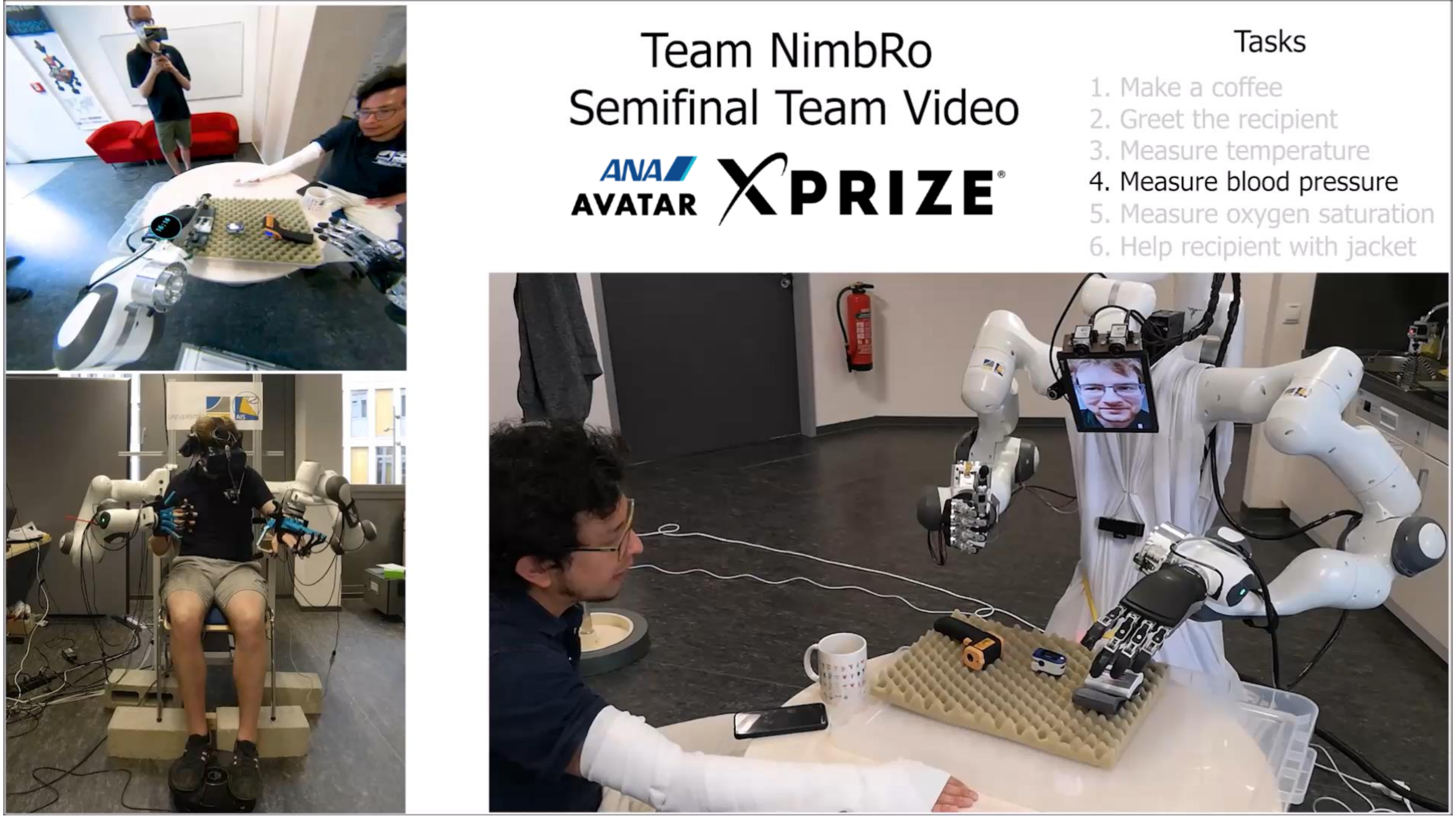
Team NimbRo Semifinal Submission

ANA AVATAR XPRIZE



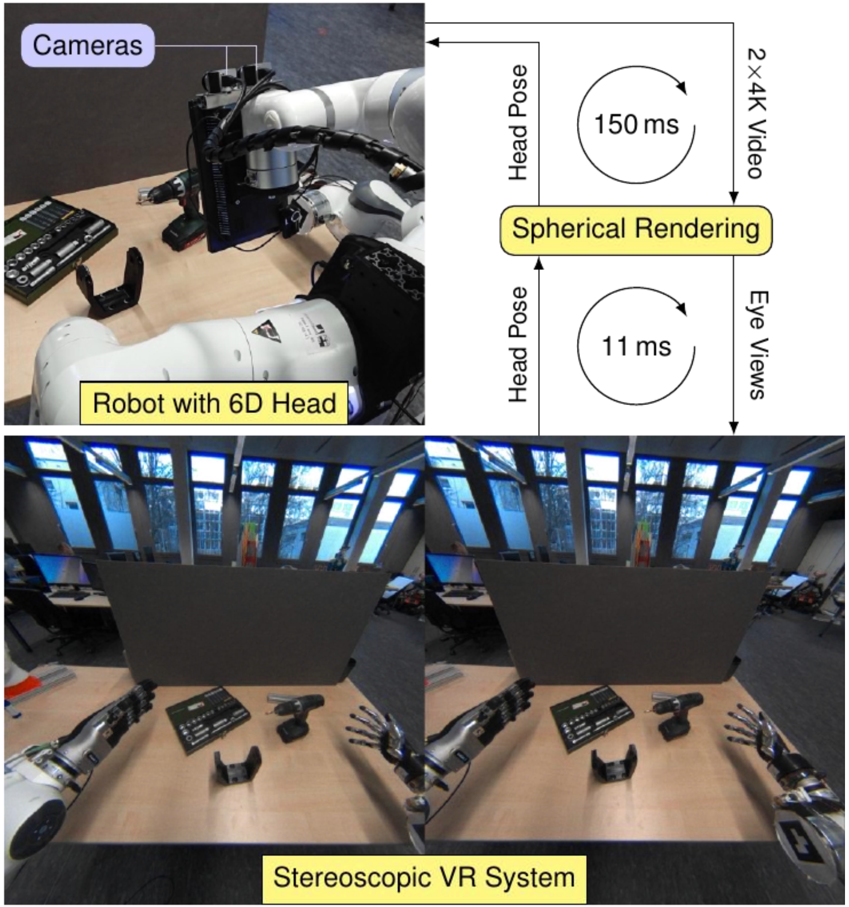
[Schwarz et al. IROS 2021]

UNIVERSITÄT BONN AIS

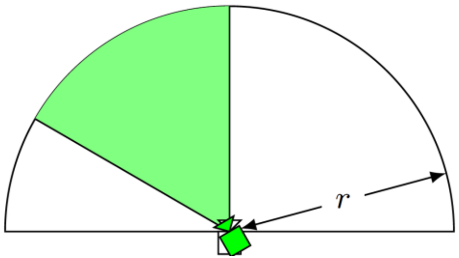


[Schwarz et al. IROS 2021]

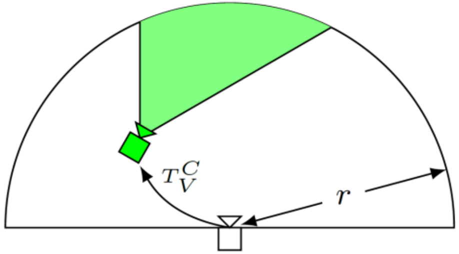
NimbRo Avatar: Immersive Visualization



- 4K wide-angle stereo video stream
- 6D neck allows full head movement
  - Very immersive
- Spherical rendering technique hides movement latencies
  - Assumes constant depth



Exact for pure rotations



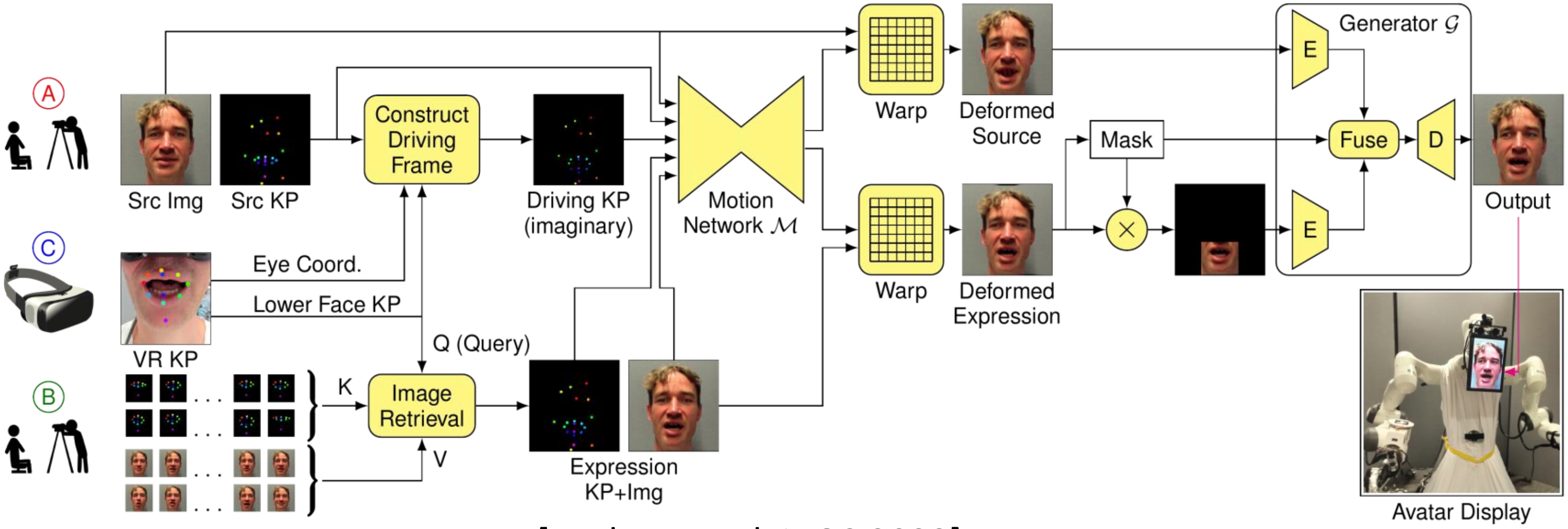
Distortions for translations

[Schwarz and Behnke Humanoids 2021]

NimbRo Avatar: Operator Face Animation

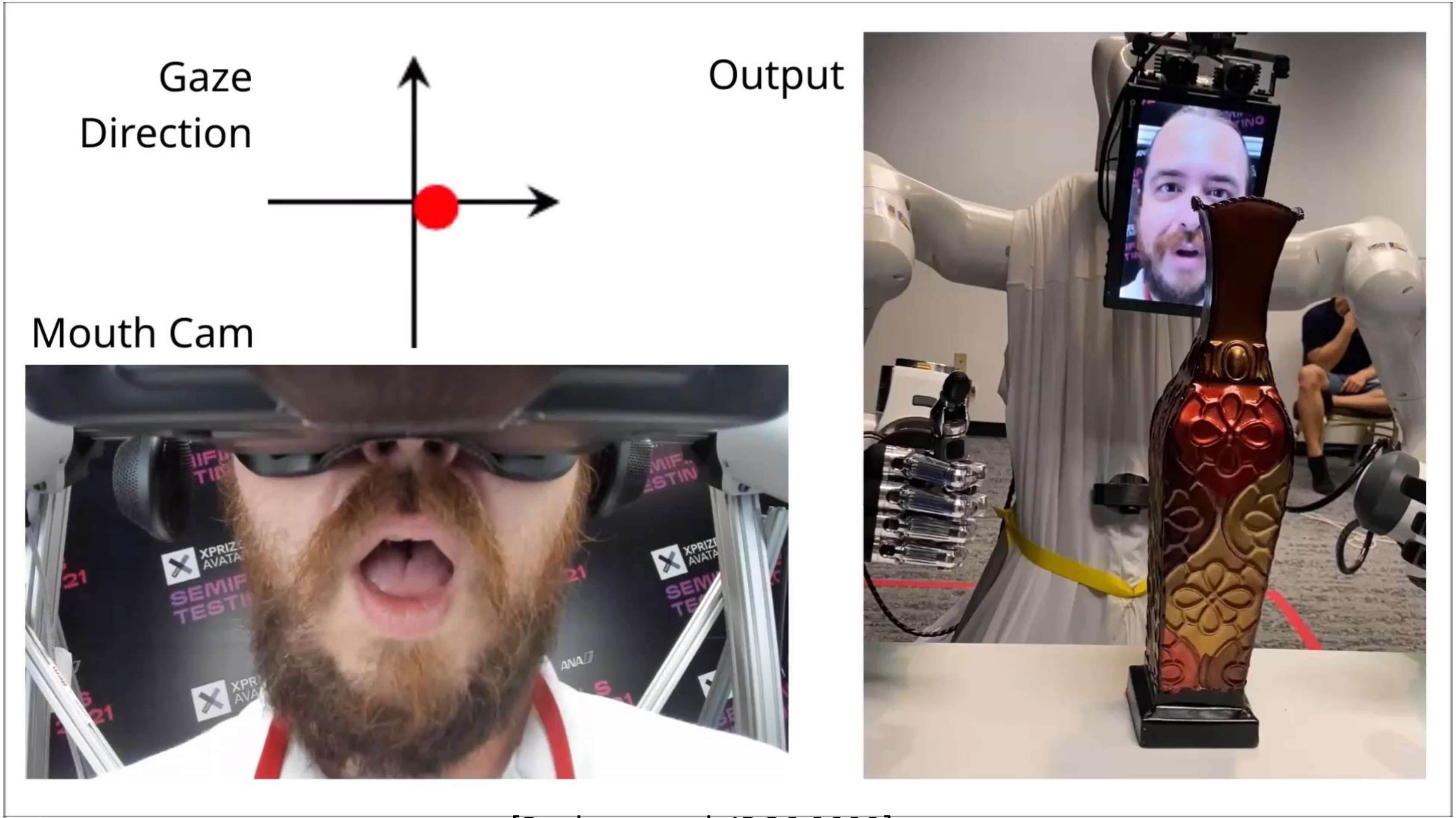
- Operator images without HMD
- Capture mouth and eyes
- Estimate gaze direction and facial keypoints

- Generate animated operator face using a warping neural network



[Rochow et al. IROS 2022]

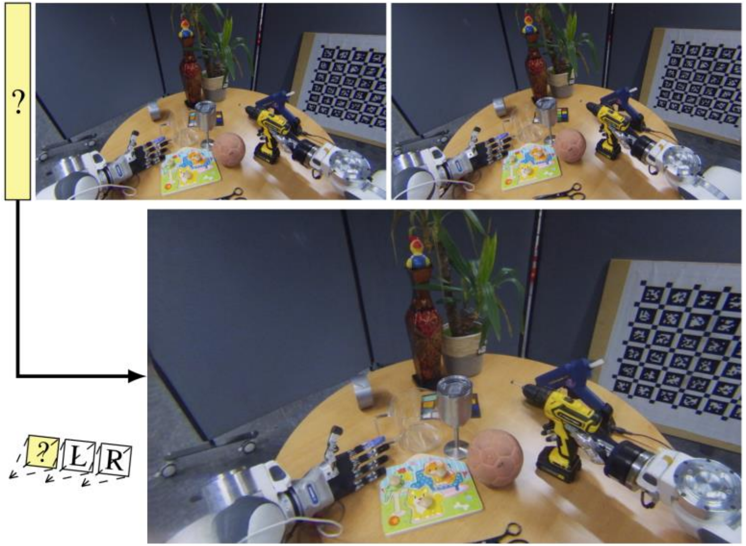
NimbRo Avatar: Operator Face Animation

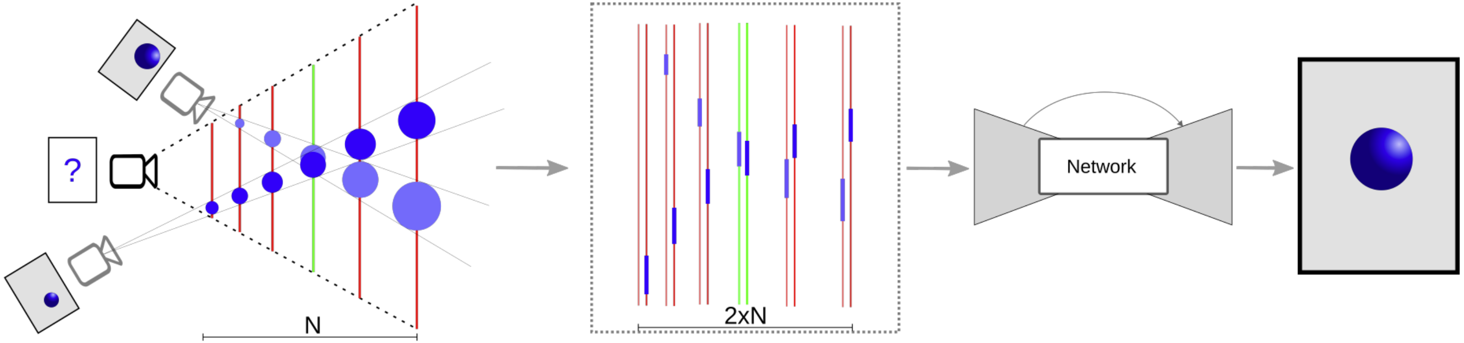


[Rochow et al. IROS 2022]

FaDIV-Syn: Fast Depth-Independent View Synthesis

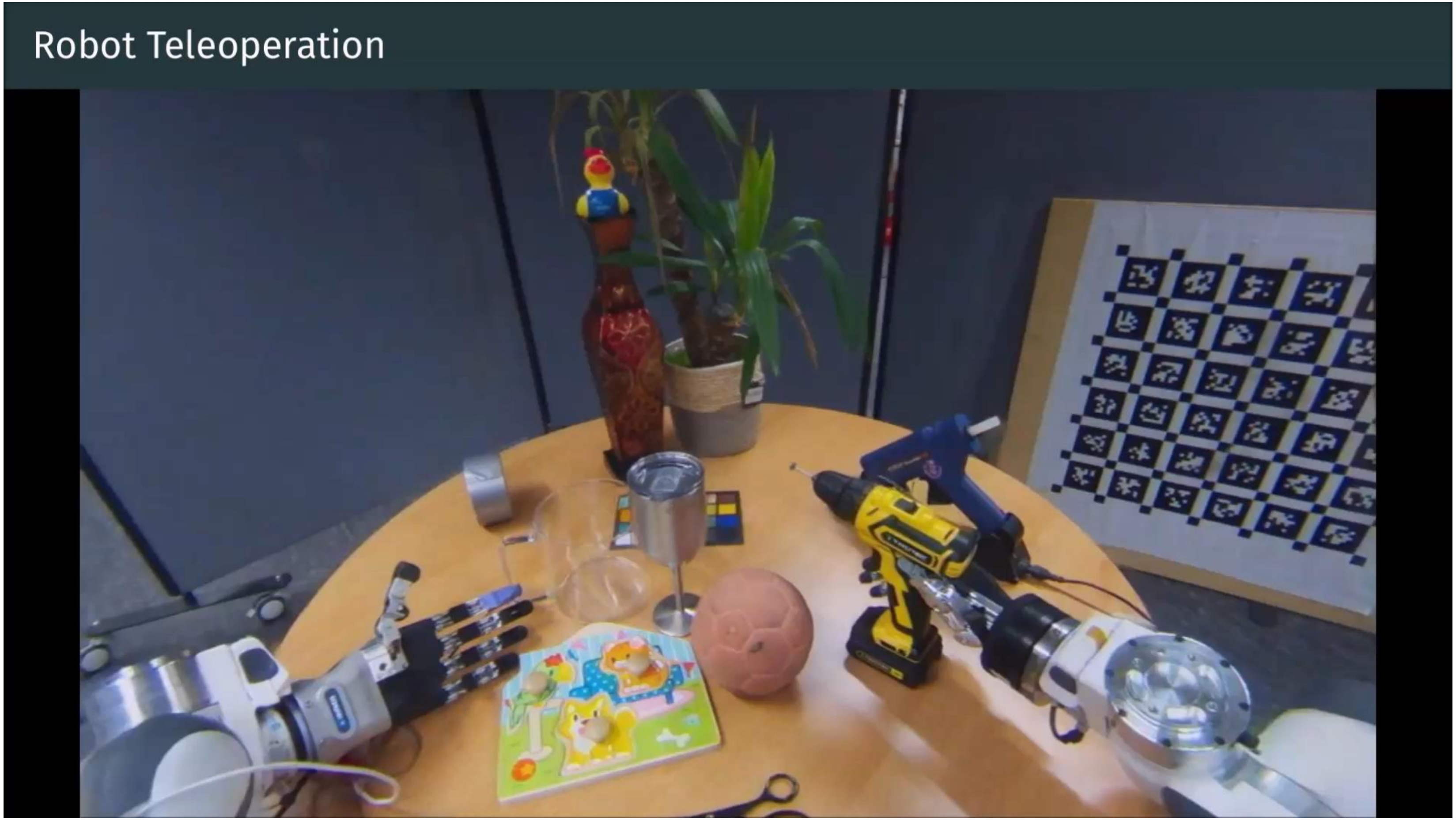
- Two input views
- Generate novel view from different pose
- Does not require depth
- Handles occlusions, transparency, reflectance, moving objects, …





[Rochow et al. RSS 2022]

FaDIV-Syn: Fast Depth-Independent View Synthesis



[Rochow et al. RSS 2022]

Conclusions

- Developed capable robotic systems for challenging scenarios
  - Plant reconstruction
  - Bin picking
  - Humanoid soccer
  - Disaster response (UGV, UAV)
- Challenges include
  - 4D semantic perception
  - High-dimensional motion planning
- Promising approaches
  - Prior knowledge (inductive bias)
  - Shared experience (fleet learning)
  - Shared autonomy (human-robot)
  - Instrumented environments