universität freiburg

Robot Learning Seminar WS 2023

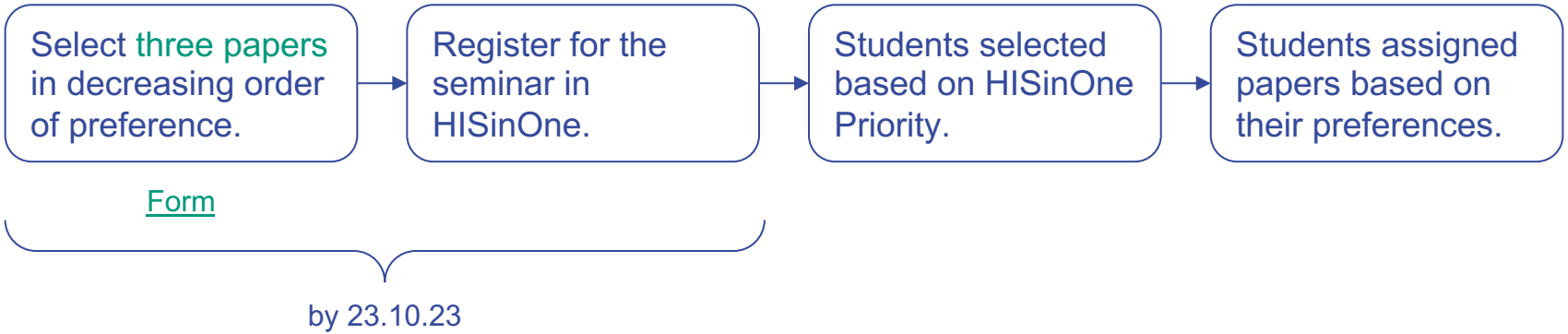
Julia Hindel

Robot Learning Lab

20 October 2023

ellis | UNIT FREIBURG

Enrollment Procedure



https://rl.uni-freiburg.de/teaching/ws23/seminar-robot-learning

Important Dates

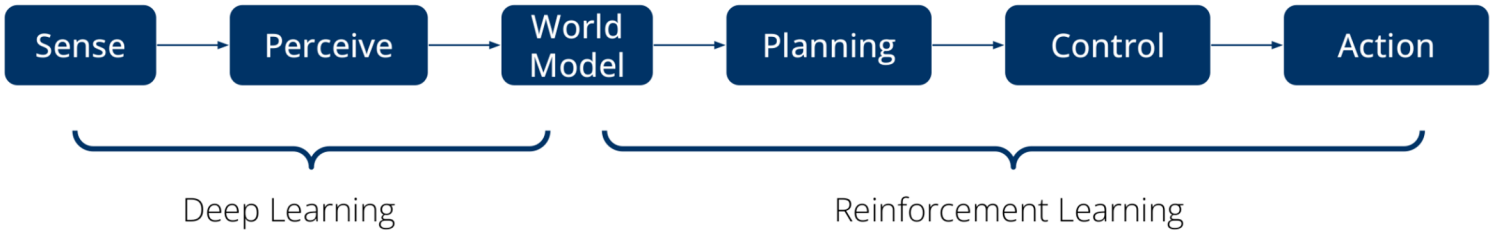
Event	Date
Lecture 1	20/10/2023
Paper selection + registration	23/10/2023
Place allocation	26/10/2023
Paper allocation	30/10/2023
Supervisor Meeting	12/2023
Lecture 2	12/01/2024
Seminar Presentation	09/02/2024
Paper Summary	23/02/2024

Evaluation

Evaluation	Due Date
Seminar Presentation	09/02/2024
Paper Summary	23/02/2024

- Presentation: At most 20 minutes
- Summary: At most 10 pages excluding bibliography
- Final grade: Presentation + Summary + Seminar participation

Autonomous Robotics



Can we learn certain parts of this pipeline?

Research Areas

Perception
- Recognition
- Depth estimation
- Motion estimation

State Estimation
- Tracking & Prediction
- SLAM
- Registration

Motion Planning
- Hierarchical learning
- Reinforcement learning
- Learning from demonstration

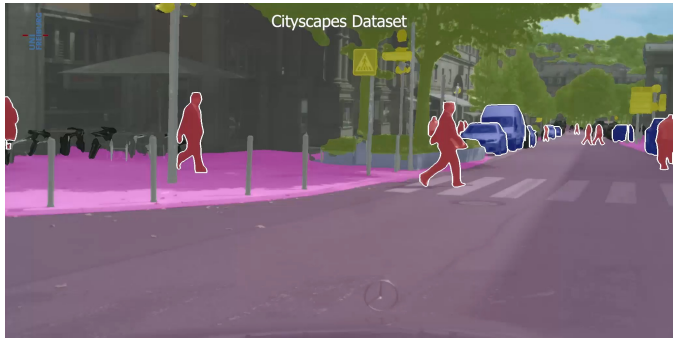
Mobile Manipulation
- Whole-body motion
- Long-horizon reasoning
- Planning for sensing

Human-Robot Interaction
- Socially-compliant behavior
- Human-robot collaboration
- Behavior adaptation & safety

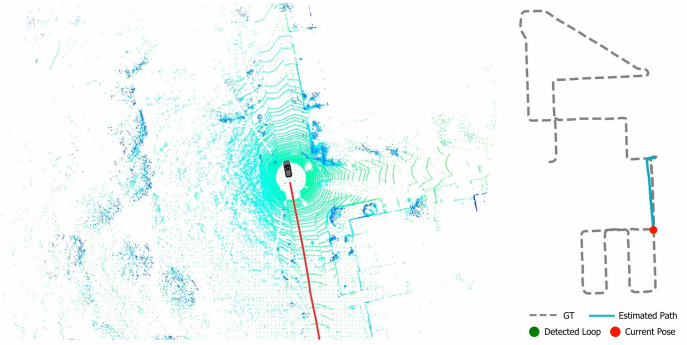
Learning Fundamentals
- Self-supervised learning
- Continual & Interactive learning
- Multimodal & Multitask learning

Responsible Robotics
- Fairness
- Explainabiity & Privacy
- Practical ethics

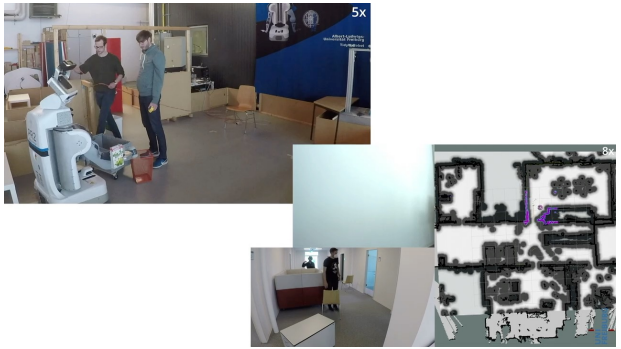
Many Seminal Works



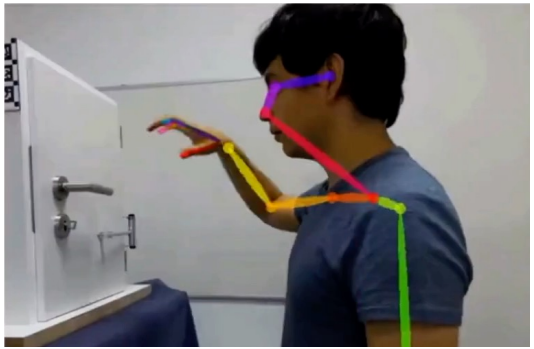
Scene Understanding



Simultaneous Localization and Mapping



Motion Planning

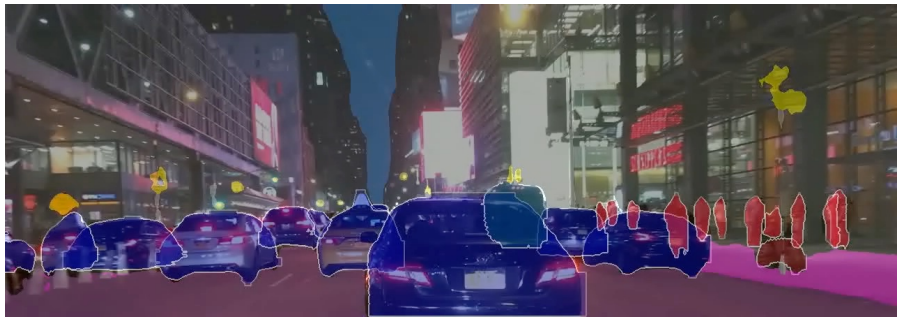


Learning from Demonstration

Robotic Perception - Mobility

Mohan, Valada: RA-L'22
Mohan, Valada: CVPR'22

Amodal Panoptic Segmentation



Enabling robots to perceive objects as a whole regardless of partial occlusion

Gosala, Valada: RA-L'22
Gosala, Petek, Drews, Burgard, Valada: CVPR'23

Bird's-Eye-View Panoptic Maps

Predicting panoptic HD maps from monocular frontal view images

Robotic Perception - Mobility

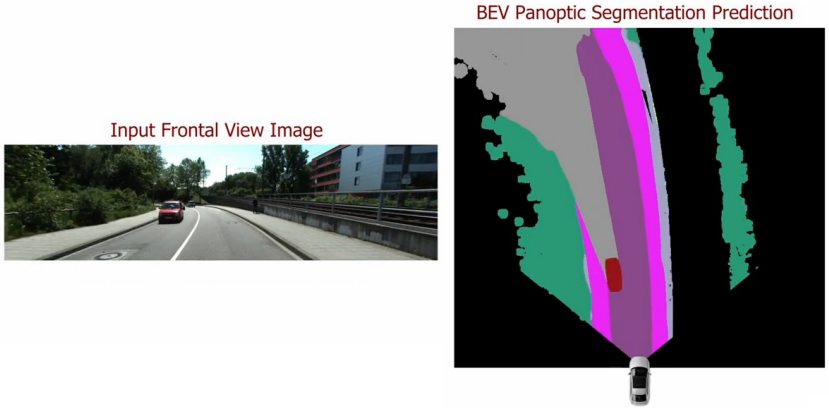
Mohan, Valada: RA-L'22
Mohan, Valada: CVPR'22

Amodal Panoptic Segmentation

Enabling robots to perceive objects as a whole regardless of partial occlusion

Gosala, Valada: RA-L'22
Gosala, Petek, Drews, Burgard, Valada: CVPR'23

Bird's-Eye-View Panoptic Maps

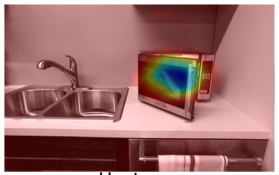


Predicting panoptic HD maps from monocular frontal view images

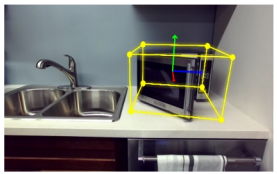
Robotic Perception - Manipulation

Heppert, et al.: CVPR'23

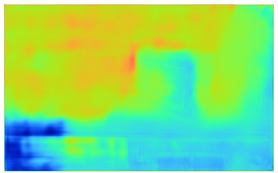
Single-Shot Reconstruction



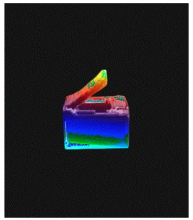
Heatmap

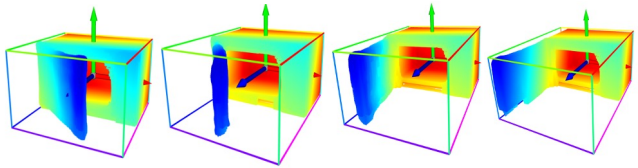


Poses and Bounding Box



Auxiliary Depth Prediction





Category-independent reconstruction and pose estimation of articulated objects from latent codes

Von Hartz, et al.: RA-L'23

Bayesian Keypoints for Manipulation

Learning scale-invariant compact representation for mobile manipulation

Robotic Perception - Manipulation

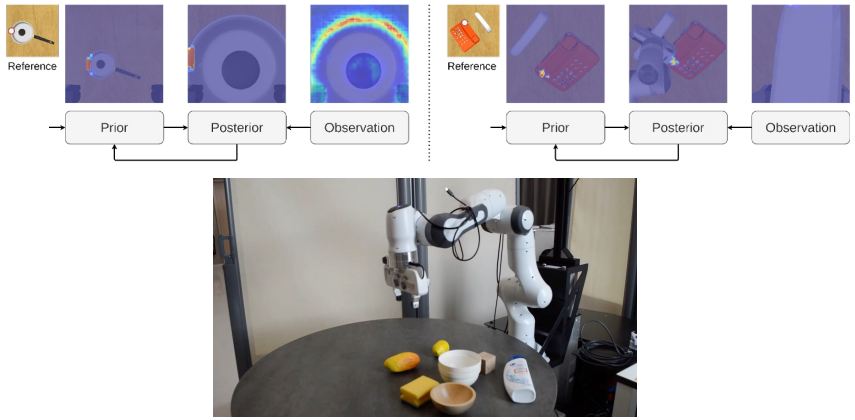
Heppert, et al.: CVPR'23

Single-Shot Reconstruction

Category-independent reconstruction and pose estimation of articulated objects from latent codes

Von Hartz, et al.: RA-L'23

Bayesian Keypoints for Manipulation

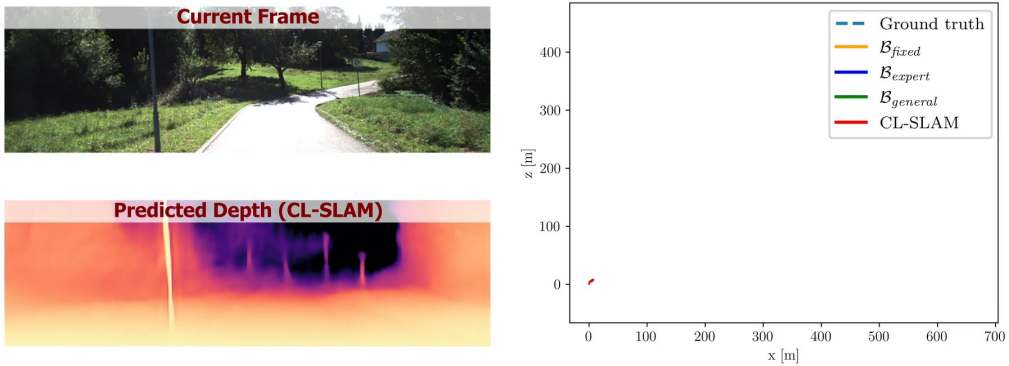


Learning scale-invariant compact representation for mobile manipulation

Mapping and Localization

Continual SLAM

Vödisch, Cattaneo, Burgard, Valada: ISRR'22
Vödisch, Cattaneo, Burgard, Valada: CVPRw'23



Continual Depth Estimation and Segmentation

Vödisch, Petek, Burgard, Valada: RSS'23

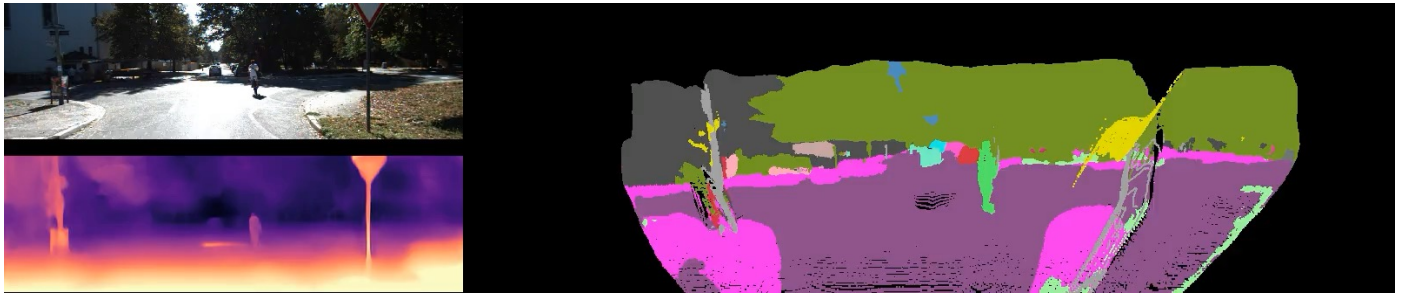
Mapping and Localization

Continual SLAM

Vödisch, Cattaneo, Burgard, Valada: ISRR'22
Vödisch, Cattaneo, Burgard, Valada: CVPRw'23

Continual Depth Estimation and Segmentation

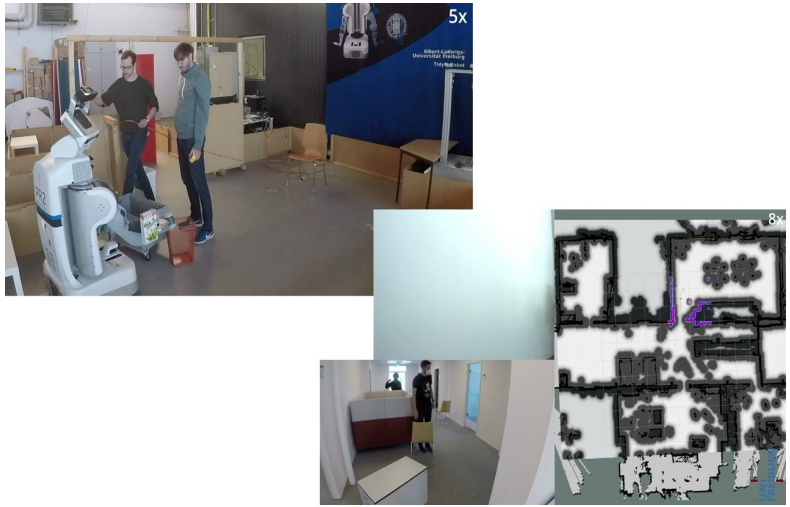
Vödisch, Petek, Burgard, Valada: RSS'23



Mobile Manipulation

Honerkamp, Welschehold, Valada: RA-L'21
Honerkamp, Welschehold, Valada: T-RO'23

Neural Navigation for Mobile Manipulation



Schmalstieg, Honerkamp, Welschehold, Valada : under review
Schmalstieg, Honerkamp, Welschehold, Valada : ISRR'22

Long-Horizon Object Search

Research Areas

Perception
- Recognition
- Depth estimation
- Motion estimation

State Estimation
- Tracking & Prediction
- SLAM
- Registration

Motion Planning
- Hierarchical learning
- Reinforcement learning
- Learning from demonstration

Mobile Manipulation
- Whole-body motion
- Long-horizon reasoning
- Planning for sensing

Human-Robot Interaction
- Socially-compliant behavior
- Human-robot collaboration
- Behavior adaptation & safety

Learning Fundamentals
- Self-supervised learning
- Continual & Interactive learning
- Multimodal & Multitask learning

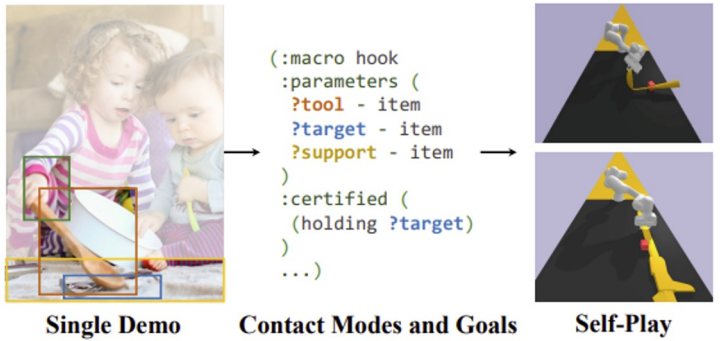
Responsible Robotics
- Fairness
- Explainabiity & Privacy
- Practical ethics

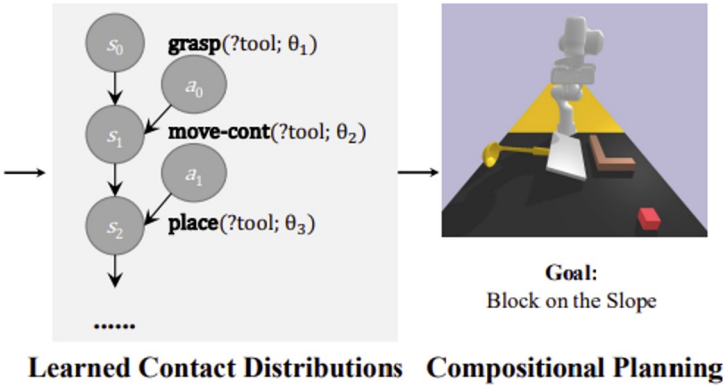
Seminar Topics

Learning Reusable Manipulation Strategies

Supervisor: Adrian Röfer

- Learning of tool usage from a single demonstration
- Extract high-level task structure from contact switches
- Learn samplers to fill the gaps in the structure
- Simulated evaluation

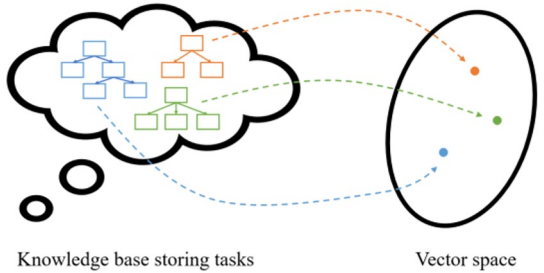




Behavior-Tree Embeddings for Robot Task-Level Knowledge

Supervisor: Adrian Röfer

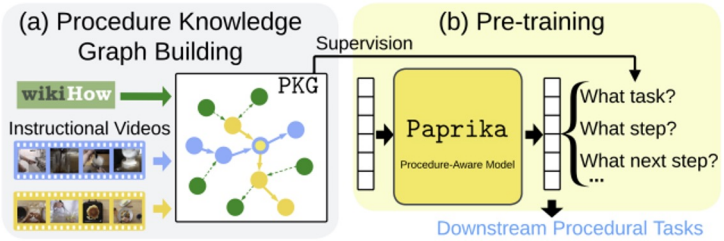
- Behavior Trees are a powerful mechanism to structure complex long-horizon behaviors
- Cannot be used in learning due to lack of fixed-size representation
- Learning of vector representation of BTs based on word-embeddings
- Statistical evaluation of knowledge-transfer, task-similarity, and relatedness of subtasks

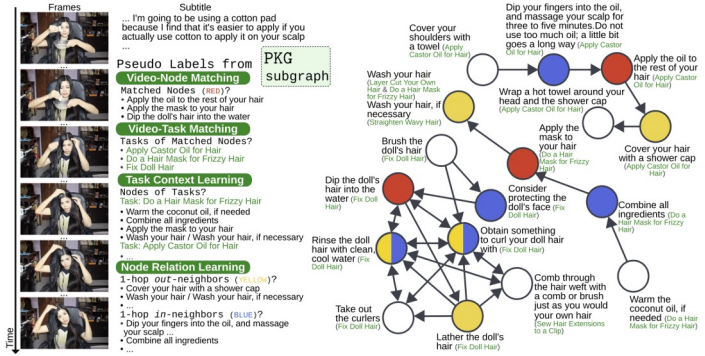


Procedure-Aware Pretraining for Instructional Video Understanding

Supervisor: Adrian Röfer

- Trying to generate procedural graphs from instructional videos
- Matching descriptions from WikiHow with videos
- Representation improves video labeling
- Structural representation can be used to relate tasks and track task progress



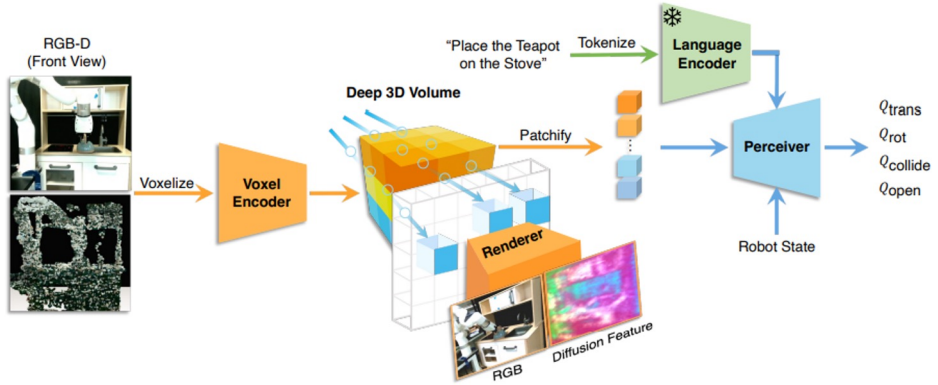


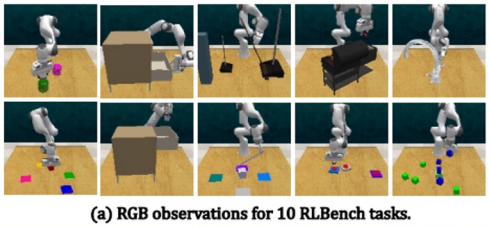
GNFactor: Multi-Task Real Robot Learning with Generalizable Neural

Feature Fields

Supervisor: Nick Heppert

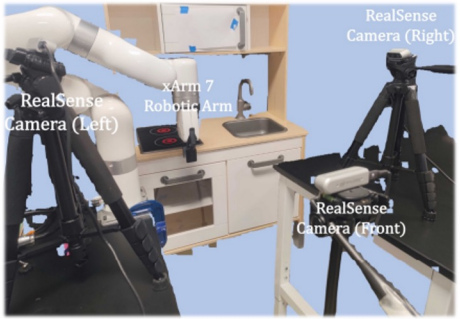
- Multi-Task imitation learning on a voxel grid
- Distill pre-trained semantic features from 2D foundation models into a NeRF
- Use PerAct to train policy
  - Embed language
- Simulation and real world evaluation





(a) RGB observations for 10 RLBench tasks.

(b) Sampled views for GNF training in simulation.

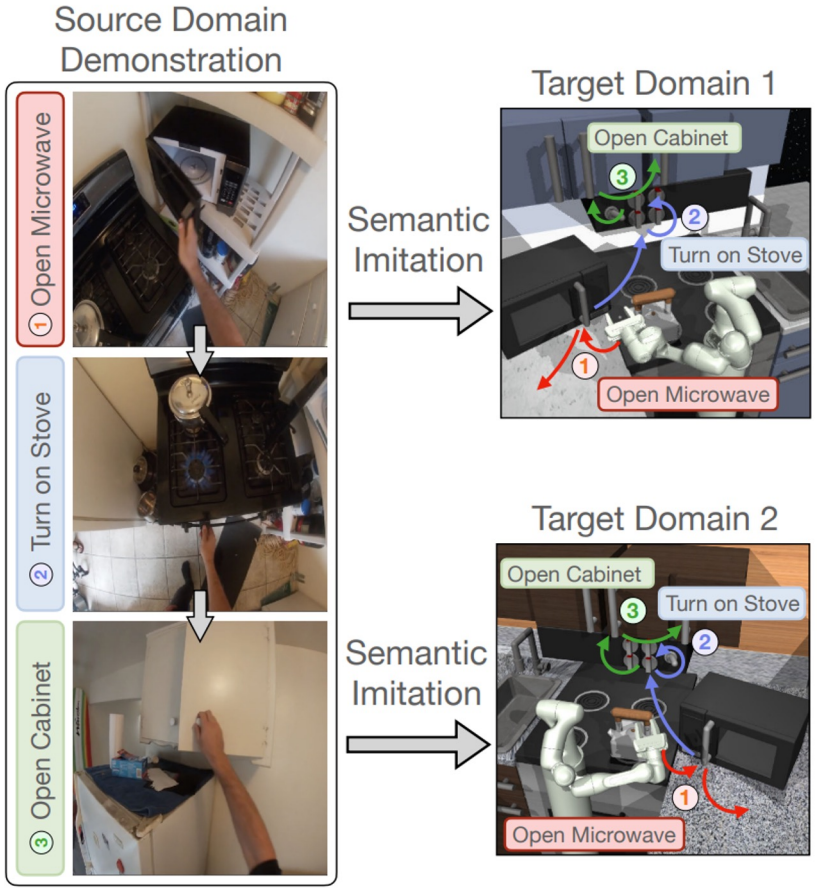


(c) Real robot setup.

Cross-Domain Transfer via Semantic Skill Imitation

Supervisor: Nick Heppert

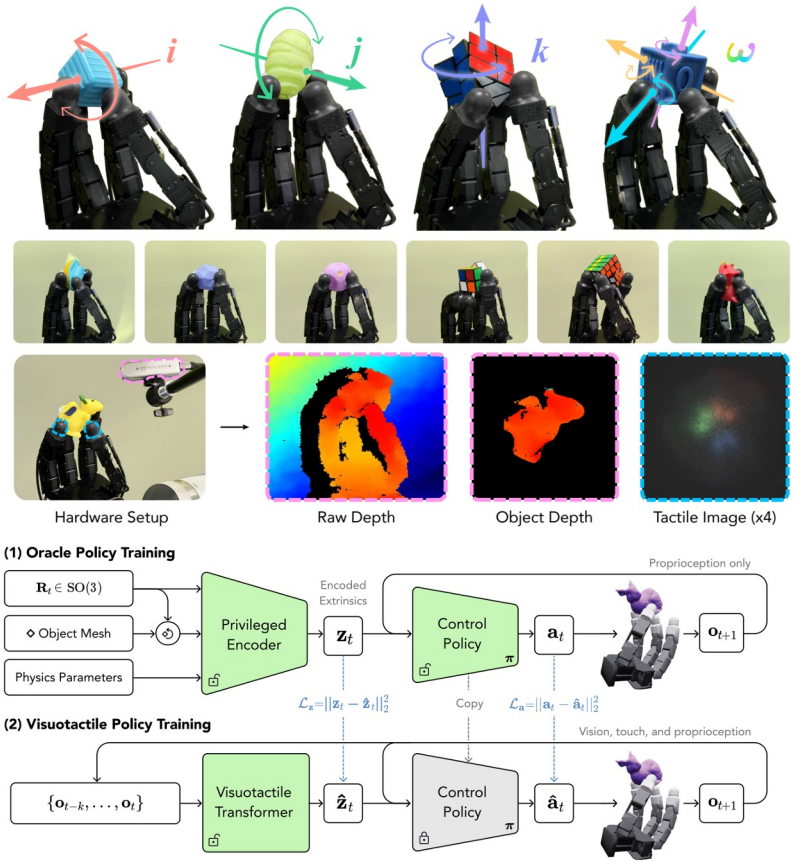
- Classically: imitate low-level actions
- Here: high-level semantic skills
- Use prior experience to learn semantic “skills”
- Enables cross-domain transfer → learning from different sources



General In-Hand Object Rotation with Vision and Touch

Supervisor: Nick Heppert

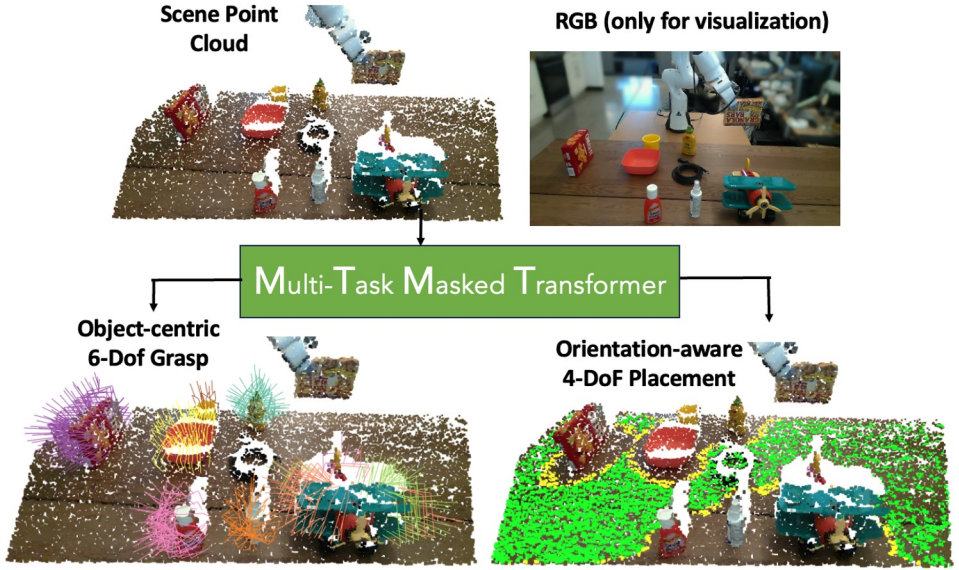
- Rotate object around axis
- Multi-modality: vision and touch
- Underlying technique: rapid domain adaptation
  - learn encoding + policy based on full information
  - learn to encode with observation



M2T2: Multi-Task Masked-Transformer for Object-centric Pick and Place

Supervisor: Jan Ole von Hartz

- Transformer model, predicting object grasp and placement from raw point cloud.
- Zero-shot Sim2Real transfer.
- Can be language-conditioned.



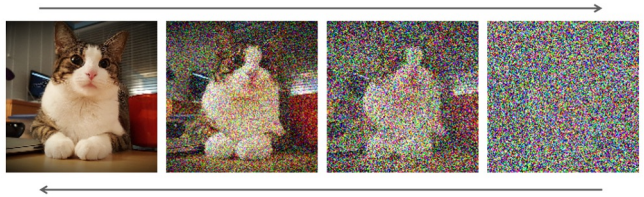
Semantic-SAM Segment and Recognize Anything at Any Granularity

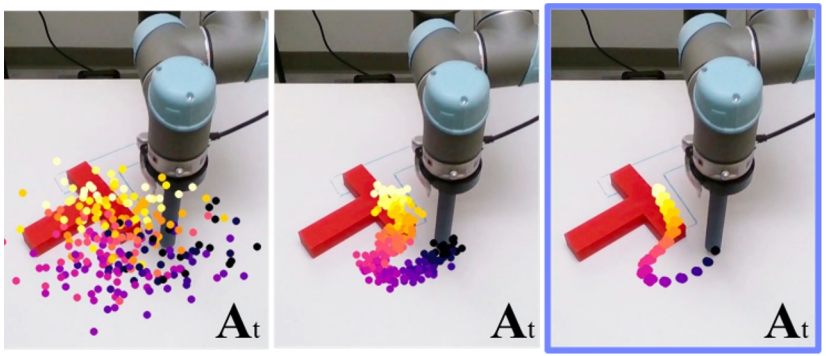
Supervisor: Jan Ole von Hartz

- Segmentation Model: instance, semantic, panoptic and part, multi-level.

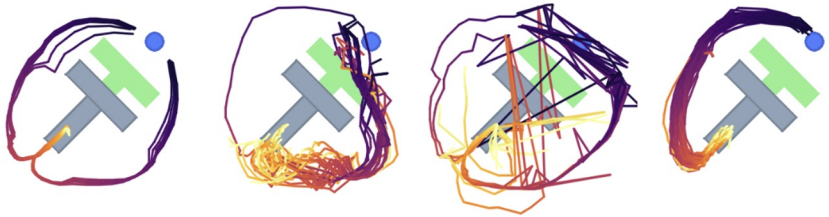
Diffusion Policy: Visuomotor Policy Learning via Action Diffusion

Supervisor: Jan Ole von Hartz





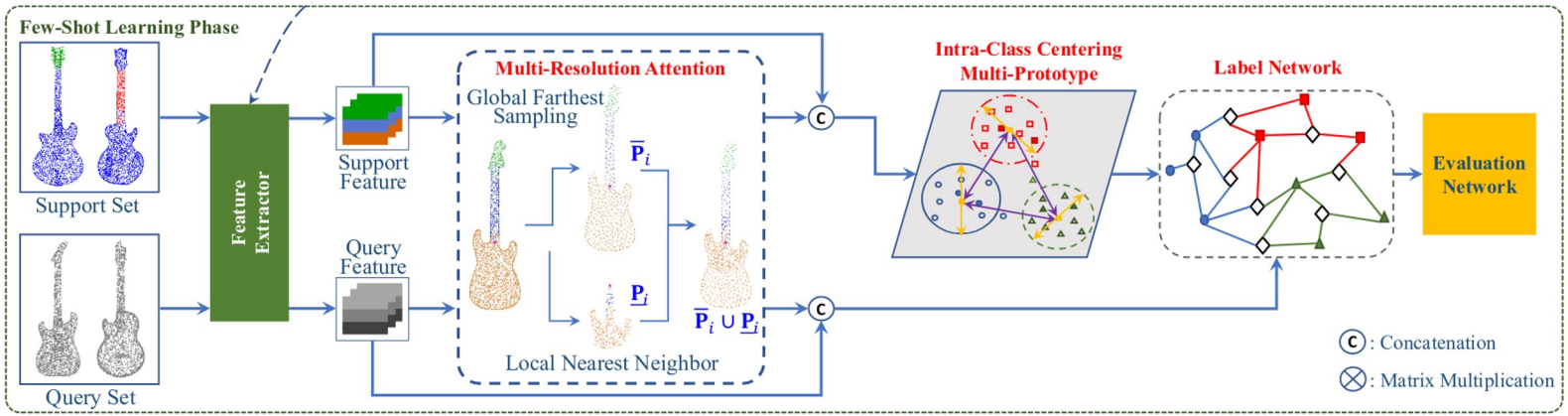
- Diffusion. Policy. Exciting!



Few-Shot Point Cloud Semantic Segmentation via Contrastive Self-Supervision and Multi-Resolution Attention

Supervisor: Julia Hindel

- Few-shot learning: training with limited labeled samples (support set) with the help of unlabeled data (query set)
- Self-supervised pre-training of feature extractor.
- Class prototype extraction and label propagation.

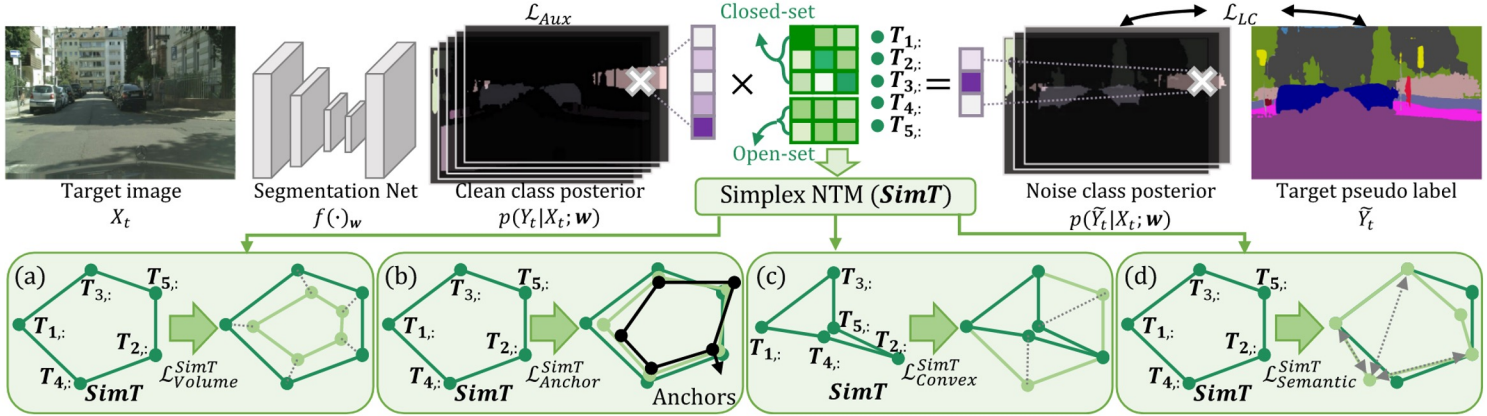


Handling Open-Set Noise and Novel Target Recognition in Domain

Adaptive Semantic Segmentation

Supervisor: Julia Hindel

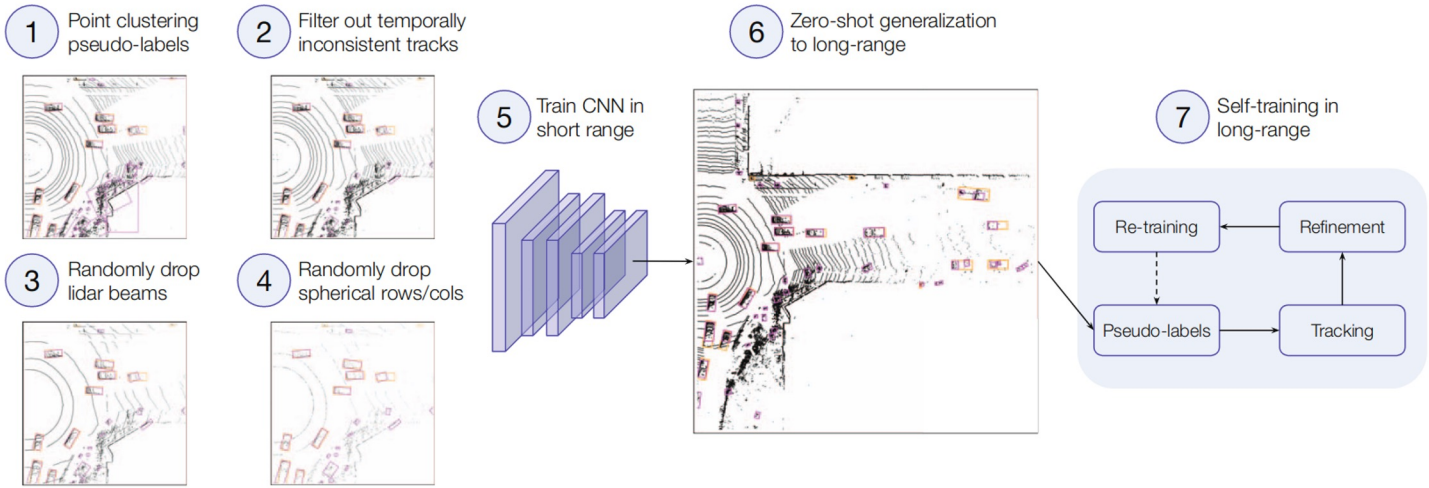
- Open-set domain adaptation: target domain with unknown classes, black-box source model.
- Self-training with pseudo-labels
- Learn noise transition matrix (NTM) with constraints



Towards Unsupervised Object Detection from LiDAR Point Clouds

Supervisor: Julia Hindel

- Near-range clustering, augmentation and temporal consistency
- Self-training on selected pseudo-labels

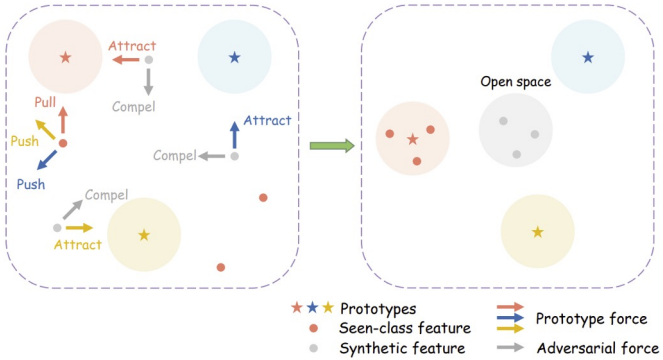


Open-set Semantic Segmentation for Point Clouds via Adversarial

Prototype Framework

Supervisor: Rohit Mohan

- Open set: Detect new classes but maintain performance on seen-class points.
- Feature extraction, prototypical constraint, feature adversarial modules
- The modules do the following:
  - Learn prototype for seen classes
  - Implicit distribution estimation of unseen-classes
  - Synthesize unseen-class features
- Adversarial mechanism to improve discriminability

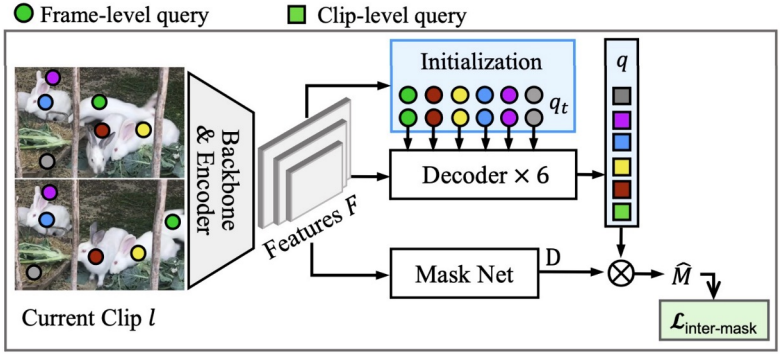


MDQE: Mining Discriminative Query Embeddings to Segment

Occluded Instances on Challenging Videos

Supervisor: Rohit Mohan

- Video instance segmentation with per-clip labels fail with crowded scenes
- Problem: Instance queries cannot encode discriminative embeddings -> poor query based performance
- Object Queries initialization with spatial contextual information and inter-frame object motion
- Inter-instance mask repulsion loss

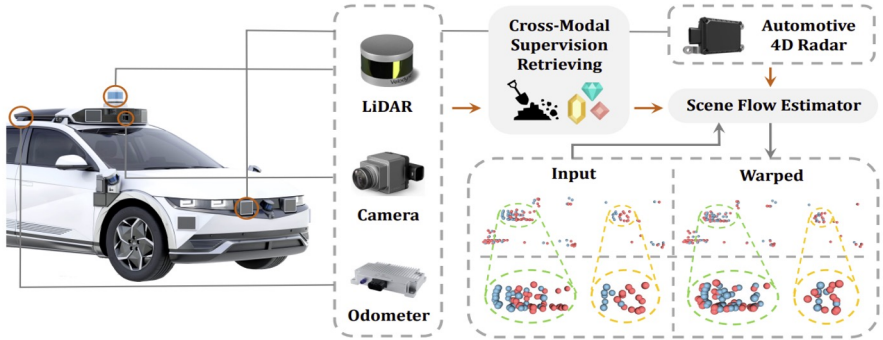


Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal

Supervision

Supervisor: Rohit Mohan

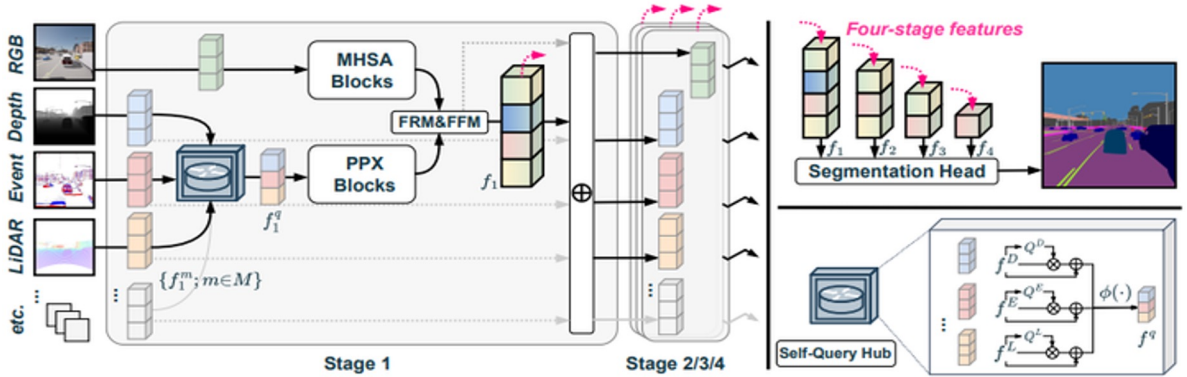
- Scene flow annotations are costly and requires tedious and intensive human labour -> Low-fidelity data further complicates procurement of annotations
- Self-supervised learning via cross-modal supervision utilizing co-located sensing
- Multitask model architecture with multiple cross-modal constrained losses



Delivering Arbitrary-Modal Semantic Segmentation

Supervisor: Iana Zhura

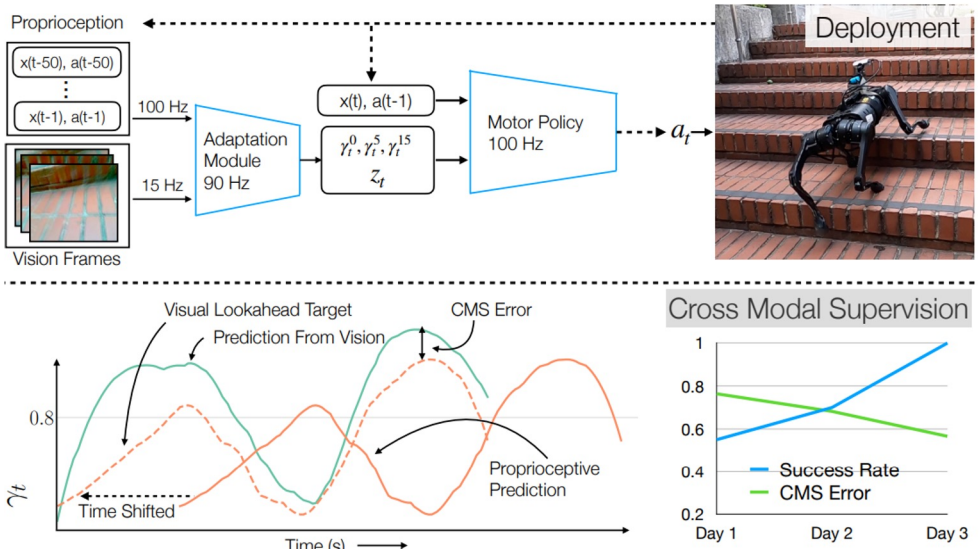
- Main idea: extract effective information from any modality for superior semantic segmentation
- The RGB image is gradually processed by Multi-Head Self-Attention (MHSA) blocks.
- Feature maps from other modalities will be merged via the proposed Self-Query Hub.
- Features are fused and forwarded to decoder for segmentation prediction.



Learning Visual Locomotion with Cross-Modal Supervision

Supervisor: Iana Zhura

- CMS uses time shifted proprioception to supervise vision.
- Policies takes an estimate of terrain in front and below the robot for visual adaptation to complex terrain
- Then, the current proprioceptive state and terrain information, is used to predict the target actions.

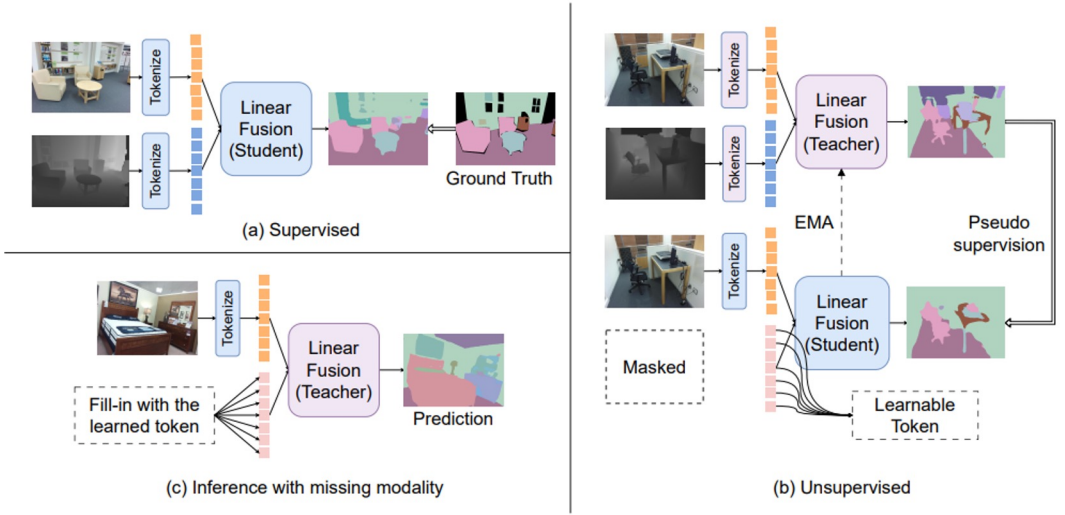


Missing Modality Robustness in Semi-Supervised Multi-Modal

Semantic Segmentation

Supervisor: Iana Zhura

- Robust to missing modality during training.
- Mask modality and replace with learnable tokens.
- Multi-modal mean (EMA) teacher generates a segmentation prediction which is used to supervise student.
- The learned token is used during inference if any modality is missing.



Questions

Announcement: Open positions

- We have open positions for a hiwi, good opportunity to work on practical robotics and get to know the lab.
- We have multiple MSc project and thesis topics related to many directions of robot learning.

https://rl.uni-freiburg.de/open-positions