universität freiburg

SS25 Seminar

Learning with Limited Supervision

Dr. Simon Bultmann

Robot Learning Lab

24 April 2025

ellis unit | FREIBURG

Agenda

I. Organization: Enrollment, important dates and evaluation.

II. Robot Learning Lab: Our research interests and publications.

III. Topics: Seminar Papers.

?. Questions.

I.

Organization

Enrollment, important dates and evaluation criteria

Seminar

Objectives

- Learn to read and understand **scientific literature**.
- Familiarize with the **State-of-the-Art** (SOTA) in the field.
- Discover **limitations**, propose **improvements** and potential **future** work.
- Build knowledge from **related work,** prior and follow-ups.
- Improve **presentation** skills.
- Develop abilities for **synthesis** (diagram drawing, summarizing main ideas, …).

TL;DR:

Show us that you have a **solid** grasp of your topic.

DIE NATURWISSENSCHAFTEN

23. Jahrgang 29. November 1935 Heft 48

Die gegenwärtige Situation in der Quantenmechanik.

Von E. Schrödinger, Oxford.

Enrollment Procedure

Select **3 papers** in decreasing order of preference.

Register for the seminar in HISinOne.

Students selected based on HISinONE Priority.

Students assigned papers based on their preferences

Fill in our Google Form

By **28.04.2025**

Please check the course website for more information:

https://rl.uni-freiburg.de/teaching/ss25/seminar-limited-supervision

Important Dates

Event	Date	Time
Lecture 1: Introduction *	24.04.2025	10:00
HISinOne registration + Paper Selection	28.04.2025	
Place allocation	29.04.2025	
Paper assignment	30.04.2025	
Supervisor Meeting	06.2025	
Lecture 2: *How to do a good presentation* *	27.06.2025	10:00
Lecture 3: Block Seminar Presentations *	24.07.2025	9:00 - 17:00
Paper Summary submission	01.08.2025	< 23:59

*** Mandatory in-person attendance**

Evaluation Criteria

Evaluation	Due Date
Seminar Presentation	24.07.2025
Paper Summary	01.08.2025

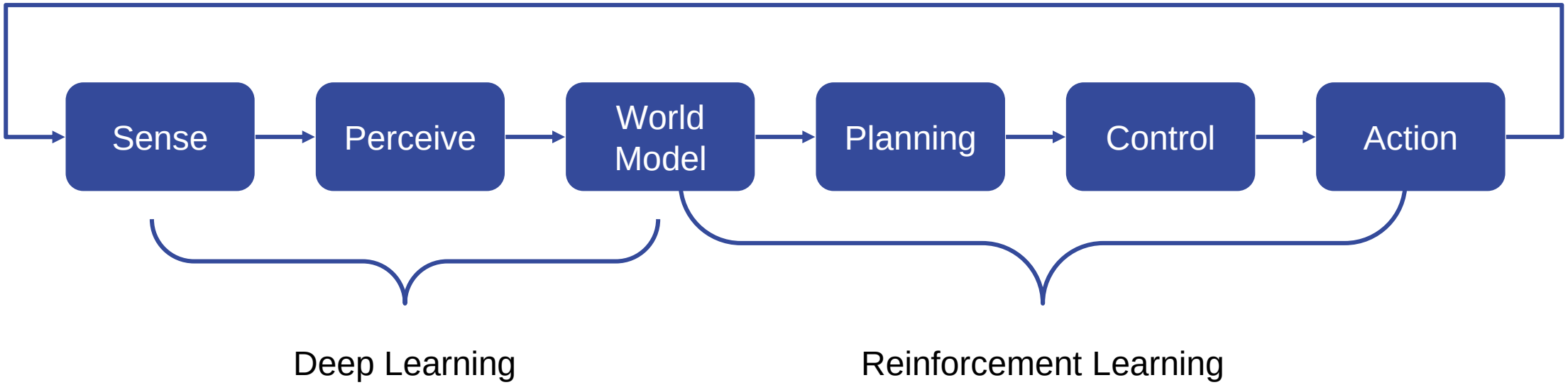
- **Presentation:** at most **20 min**.
- **Summary:** at most **7 pages** excluding bibliography and figures.
- **Final grade:**
  - Presentation (slides & delivery) + Summary + **Seminar Participation**.

II.

Robot Learning Lab

Our research interests and publications

Autonomous Robotics



Can we **learn** certain parts of this pipeline?

Robot Learning

Learning …

- … models of robots, tasks or environments
- … deep hierarchies/representations from sensor and motor representations to task representations
- … plans and control policies
- … methods for probabilistic inference from multi-modal data
- … structured spatio-temporal representations, e.g. low-dim. embeddings of Movements

How can we ensure **autonomous operation** of embodied AI systems

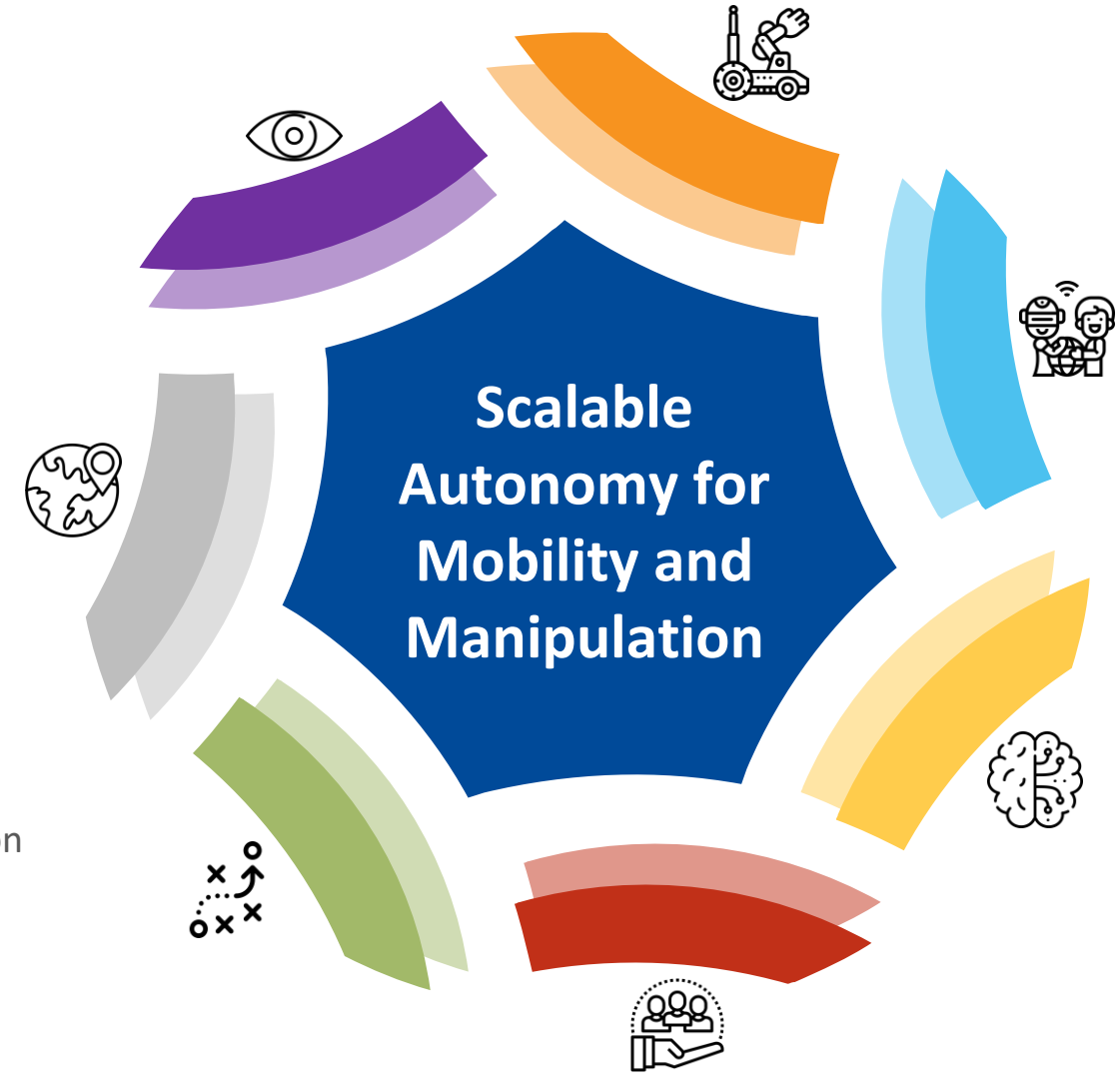
with **limited supervision** ?

Research Areas

Perception
- Recognition
- Depth estimation
- Motion estimation

State Estimation
- Tracking & Prediction
- SLAM
- Registration

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



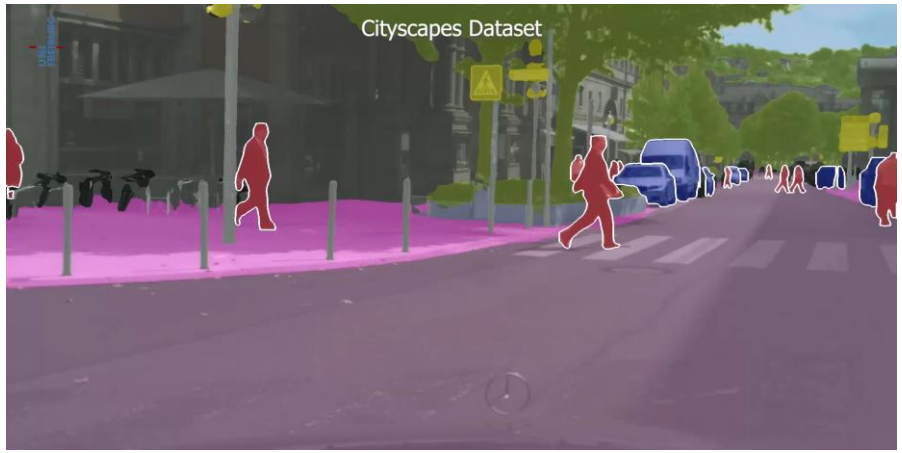
Mobile Manipulation
- Whole-body motion
- Long-horizon reasoning
- Planning for sensing

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

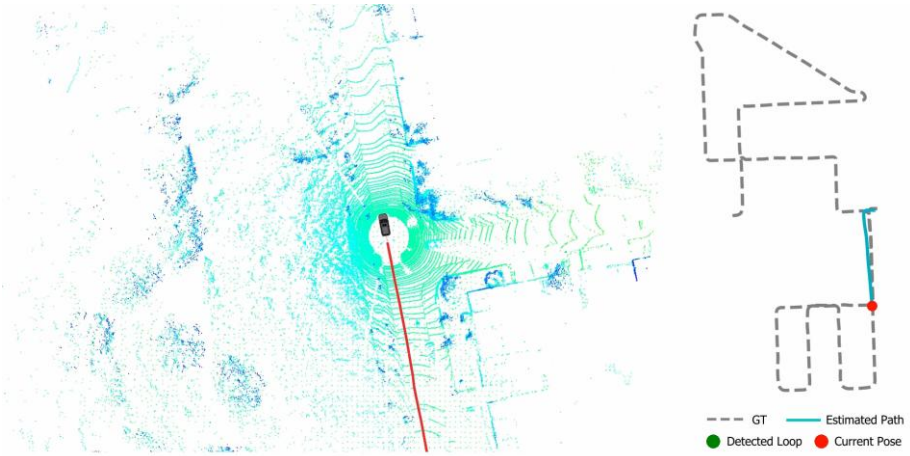
Learning Fundamentals
- Self-supervised learning
- Continual & Interactive learning
- Multimodal & Multitask learning

Responsible Robotics
- Fairness
- Explainabiity & Privacy
- Practical ethics

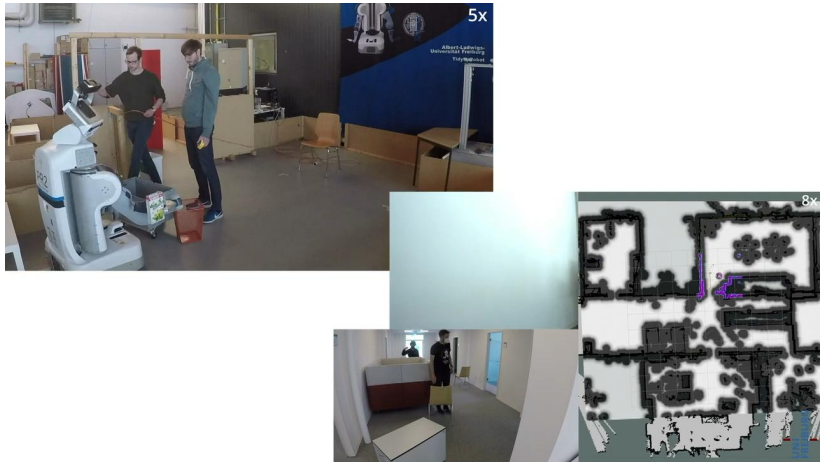
Many Seminal Works



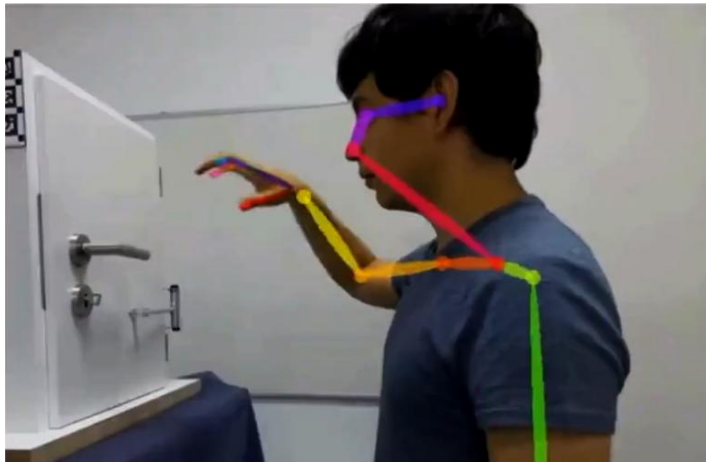
Scene Understanding



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

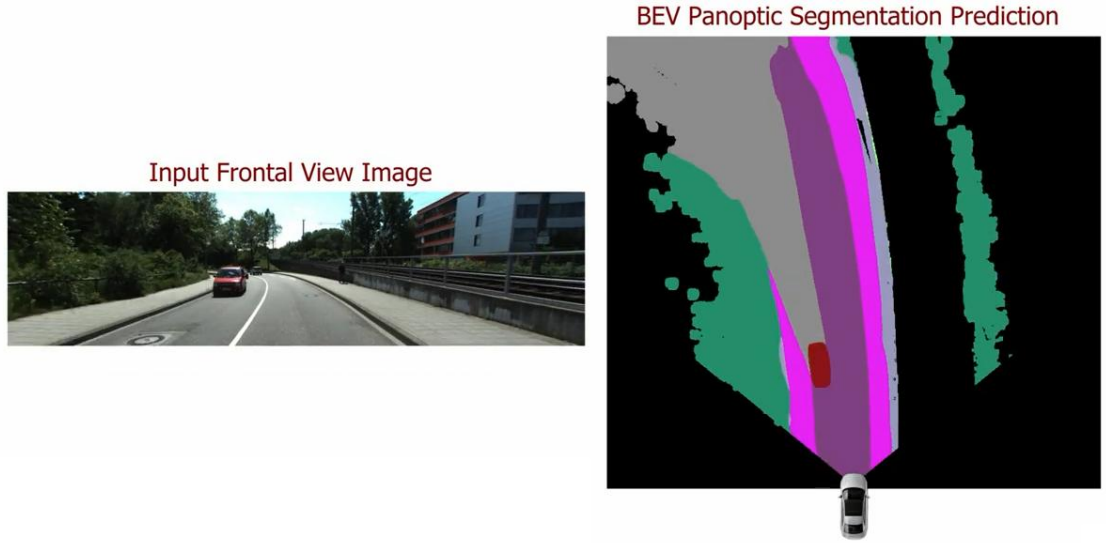
Mohan, Valada: RA-L'22
Mohan, Valada: CVPR'22

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

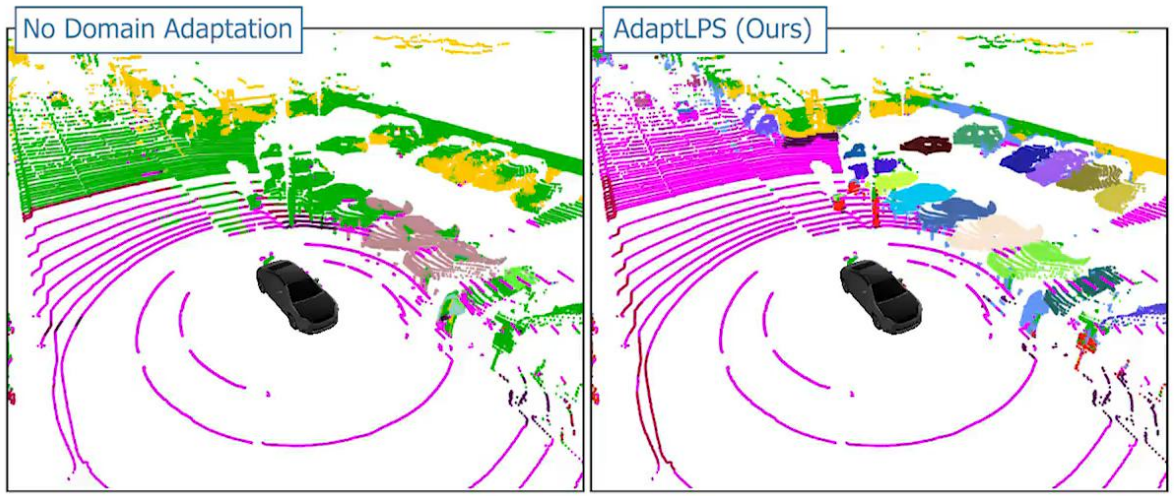


Predicting panoptic HD maps from monocular frontal view images

Robotic Perception - Mobility

Besic, Gosala, Cattaneo, Valada: RA-L'22

Unsupervised LiDAR Domain Adaptation



Sirohi, et al: ICRA'23

Panoptic Uncertainty Estimation

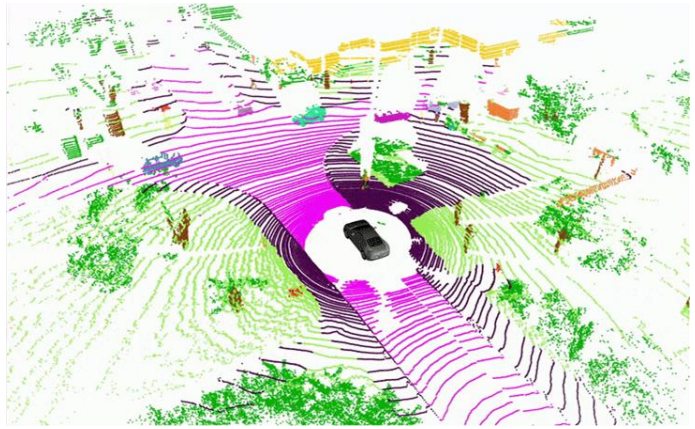
Robotic Perception - Mobility

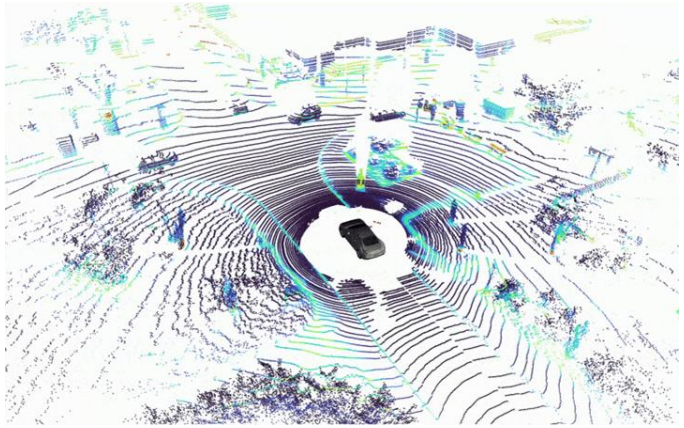
Besic, Gosala, Cattaneo, Valada: RA-L'22

Unsupervised LiDAR Domain Adaptation

Sirohi, et al: ICRA'23

Panoptic Uncertainty Estimation

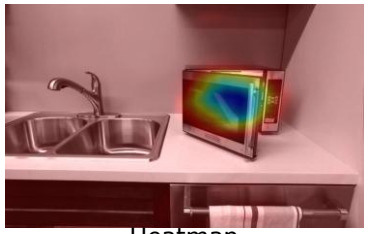




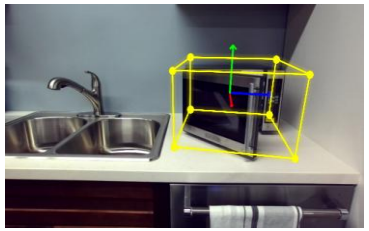
Robotic Perception - Manipulation

Heppert, et al.: CVPR'23

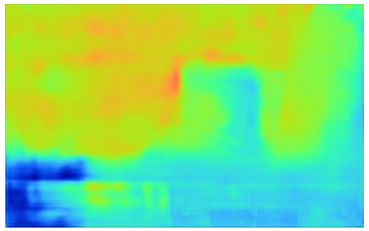
Single-Shot Reconstruction



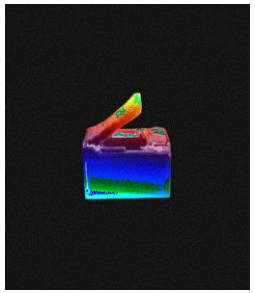
Heatmap

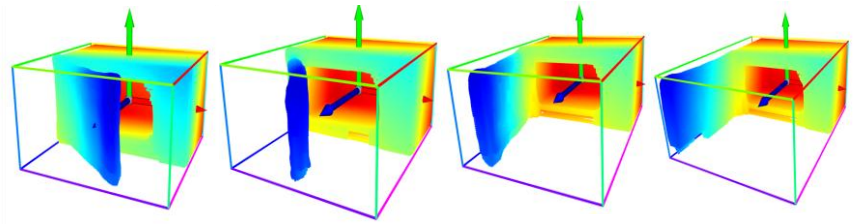


Poses and Bounding Box



Auxiliary Depth Prediction





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

Chisari et al., RA-L 2022

Robot Learning from Human Feedback

Exploiting interactive learning where a human teacher provides feedback during execution

Robotic Perception - Manipulation

Heppert, et al.: CVPR'23

Single-Shot Reconstruction

Heatmap

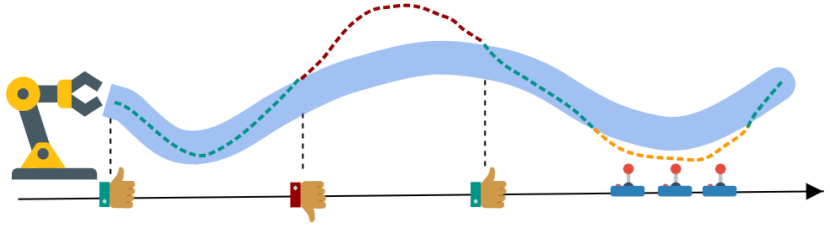
Poses and Bounding Box

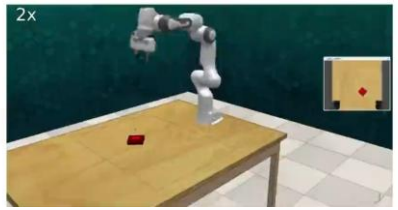
Auxiliary Depth Prediction

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

Chisari et al., RA-L 2022

Robot Learning from Human Feedback



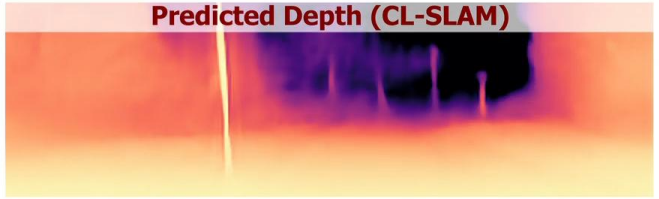


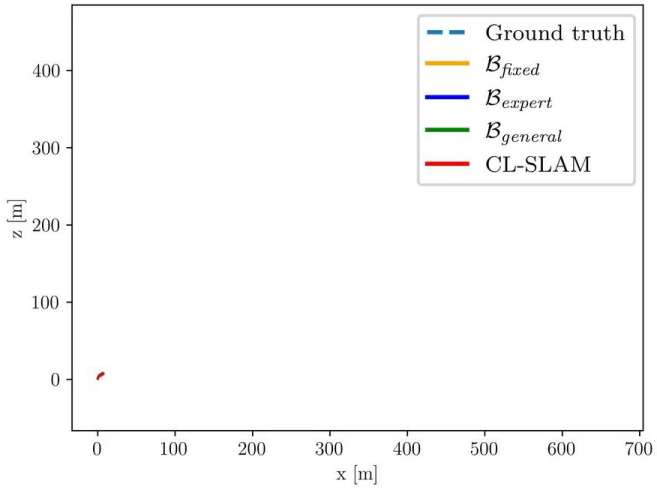
Exploiting interactive learning where a human teacher provides feedback during execution

Mapping and Localization

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

Continual SLAM





Collaborative Dynamic 3D Scene Graphs

Greve, Vödisch, Büchner, Valada: ICRA'24

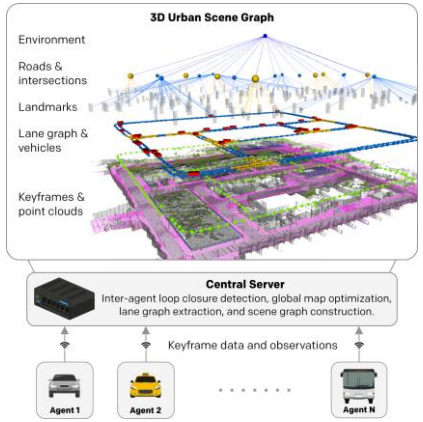
- Environment
- Roads & intersections
- Landmarks
- Vehicles on lane graph
- Keyframes & point clouds

Mapping and Localization

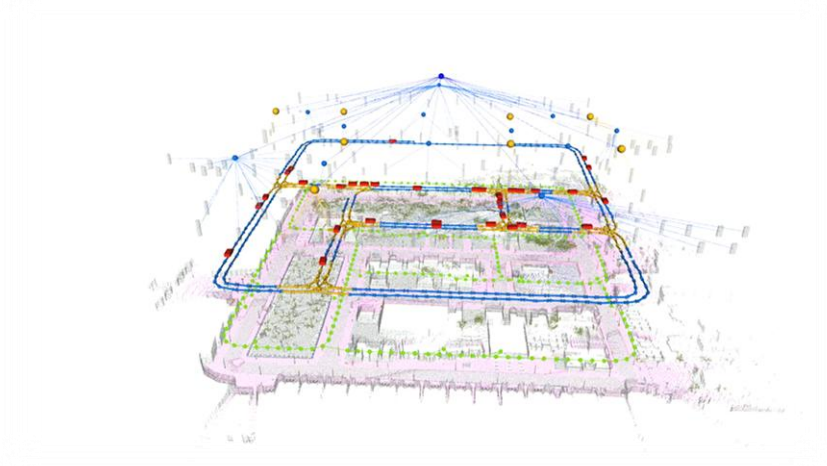
Vödisch, Cattaneo, Burgard, Valada: ISRR'22
Vödisch, Cattaneo, Burgard, Valada: CVPRw'23

Collaborative Dynamic 3D Scene Graphs

Greve, Vödisch, Büchner, Valada: ICRA'24



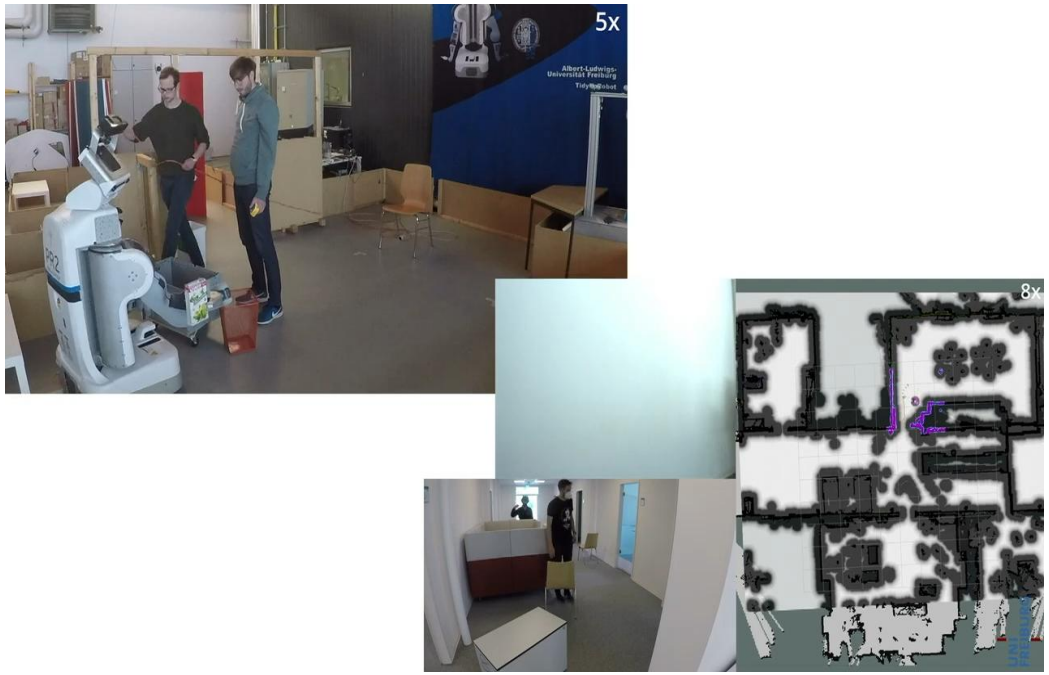
- Environment
- Roads & intersections
- Landmarks
- Vehicles on lane graph
- Keyframes & point clouds



Mobile Manipulation

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

Neural Navigation for Mobile Manipulation



Schmalstieg, Honerkamp, Welschehold, Valada : RA-L'23
Schmalstieg, Honerkamp, Welschehold, Valada : ISRR'22

Long-Horizon Object Search

Mobile Manipulation

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

Neural Navigation for Mobile Manipulation

Schmalstieg, Honerkamp, Welschehold, Valada : RA-L'23
Schmalstieg, Honerkamp, Welschehold, Valada : ISRR'22

Long-Horizon Object Search

Language-Grounded Learning

Honerkamp, Buechner, Despinoy, Welschehold, Valada: RA-L'25

MoMa-LLM

Werby, Huang, Buechner, Valada, Burgard : RSS'24

Language-Grounded Navigation

Speed 4X

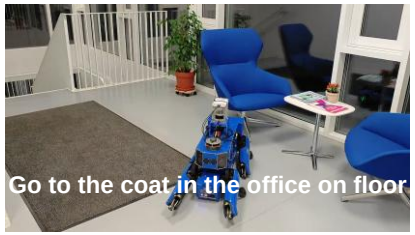
Language-Grounded Learning

Honerkamp, Buechner, Despinoy, Welschehold, Valada: under review

MoMa-LLM

Werby, Huang, Buechner, Valada, Burgard : RSS'24

Language-Grounded Navigation



Speed 4X

III.

Topics

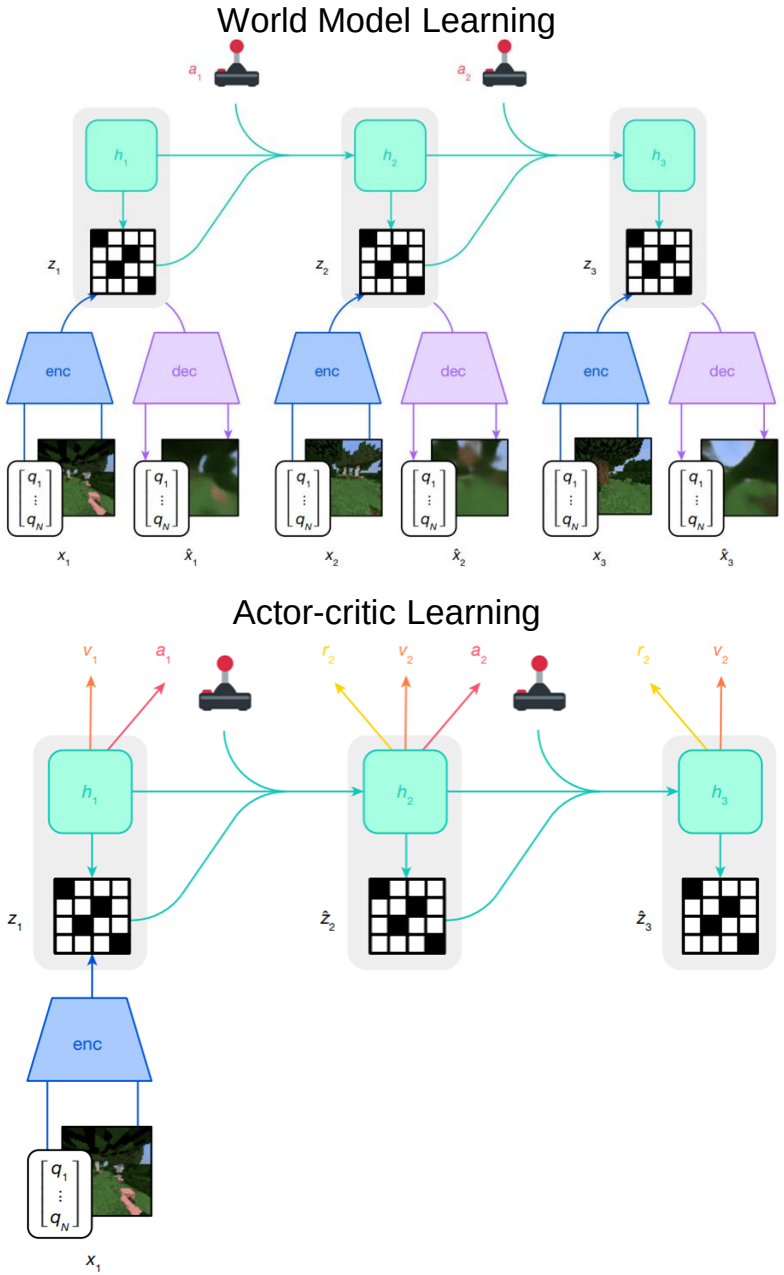
Seminar Papers

Supervisor: Iman Nematollahi

Mastering diverse control tasks through world models

https://www.nature.com/articles/s41586-025-08744-2

- **Generalist RL via World Models:** DreamerV3 leverages learned world models to imagine and plan future outcomes for efficient decision-making.
- **Robustness:** Solves 150+ tasks (including Minecraft) with fixed hyperparameters.
- **Scalability and Efficiency:** Performance improves predictably with model size and compute.

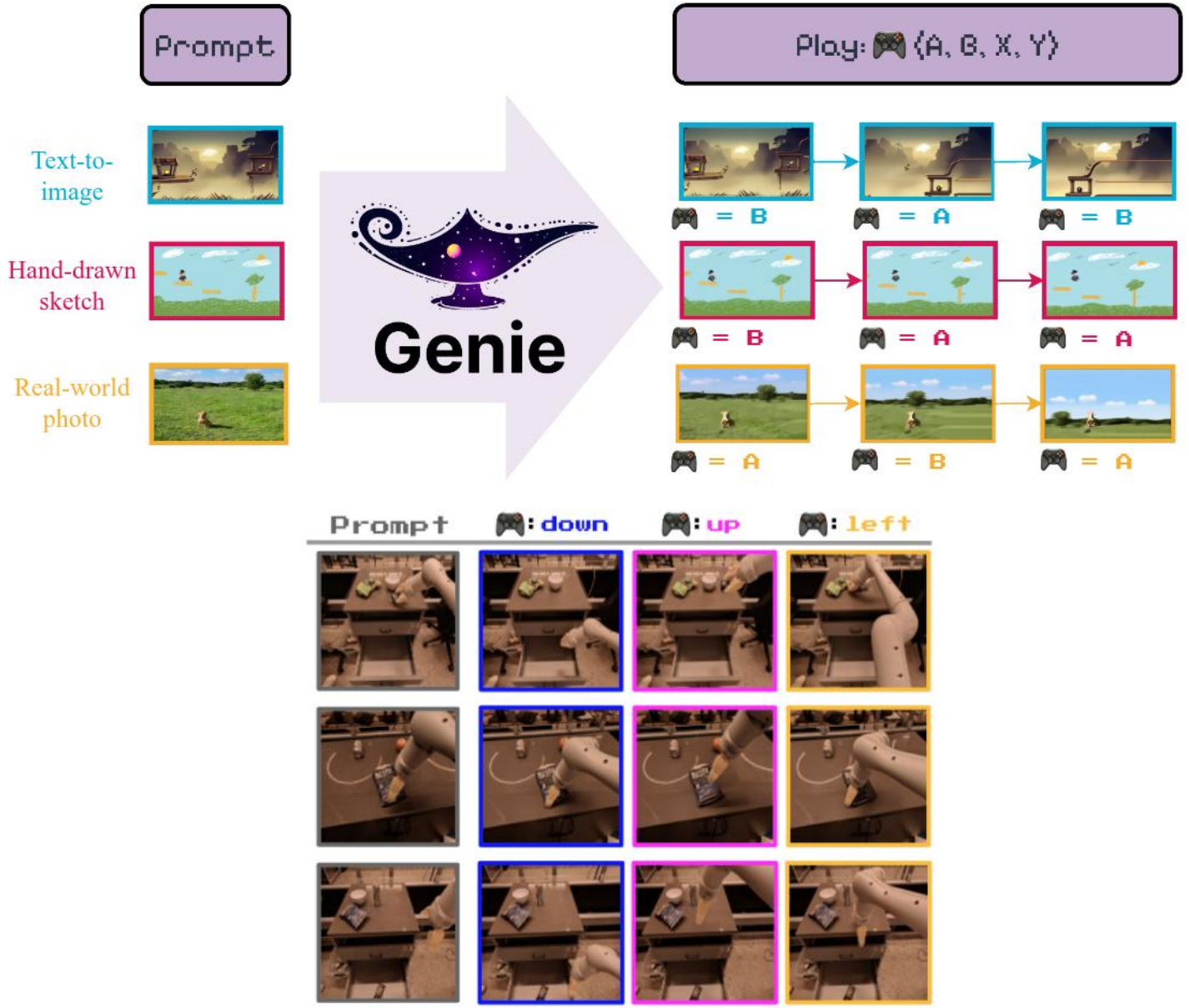


Supervisor: Iman Nematollahi

Genie: Generative Interactive Environments

https://arxiv.org/pdf/2402.15391

- **Video-only World Model:** Learns interactive environments from unlabeled Internet videos using latent actions.
- **Versatile Prompting:** Generates controllable environments from sketches, photos, or text-to-image outputs.
- **Unsupervised Action Learning:** Infers latent actions without ground-truth labels, enabling frame-level controllability.

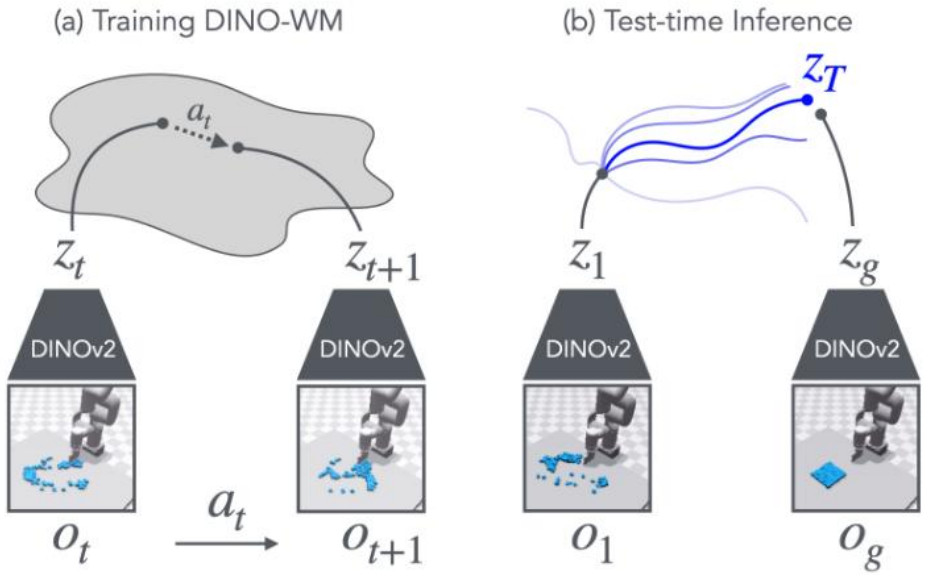


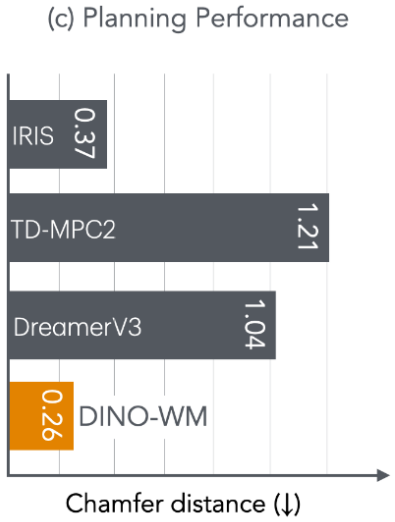
Supervisor: Iman Nematollahi

DINO-WM: World Models on Pre-trained Visual Features enable Zero-shot Planning

https://arxiv.org/pdf/2411.04983

- **Latent World Modeling:** Predicts future outcomes in DINOv2 patch embedding space; no need for pixel-level reconstructions.
- **Zero-shot Visual Planning:** Reaches goal images via model-predictive control without expert demos, rewards, or inverse models.
- **Simple & Scalable:** Trained on offline data using pretrained frozen visual encoders, enabling task-agnostic reasoning.



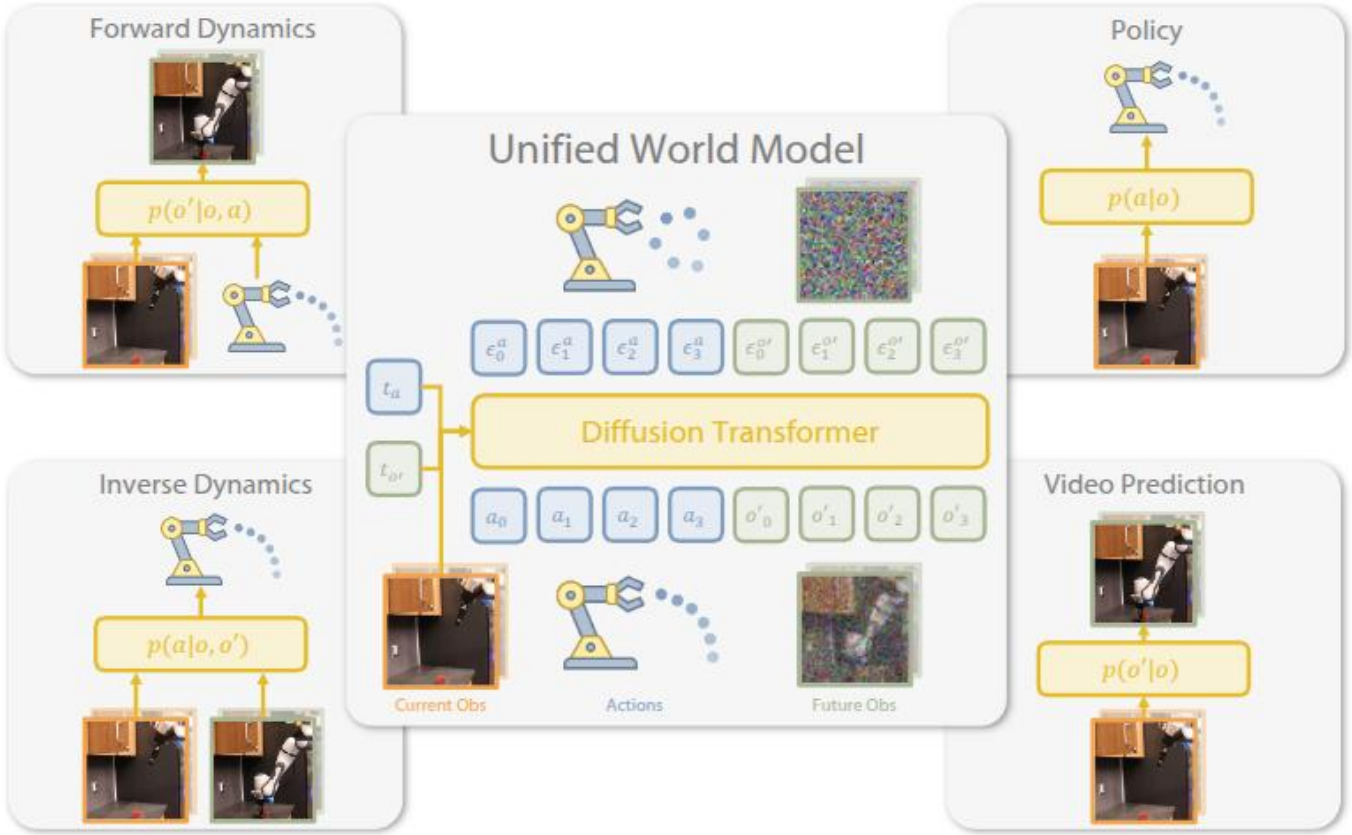


Supervisor: Iman Nematollahi

Unified World Models: Coupling Video and Action Diffusion for Pretraining on Large Robotic Datasets

https://arxiv.org/pdf/2504.02792

- **Unified Diffusion Framework:** Combines action and video diffusion in one model with independently controlled timesteps.
- **Versatile Inference:** Acts as a policy, dynamics model, inverse model, or video predictor by adjusting diffusion steps.
- **Learns from Videos and Actions:** Trains on both action-annotated robot data and action-free videos.

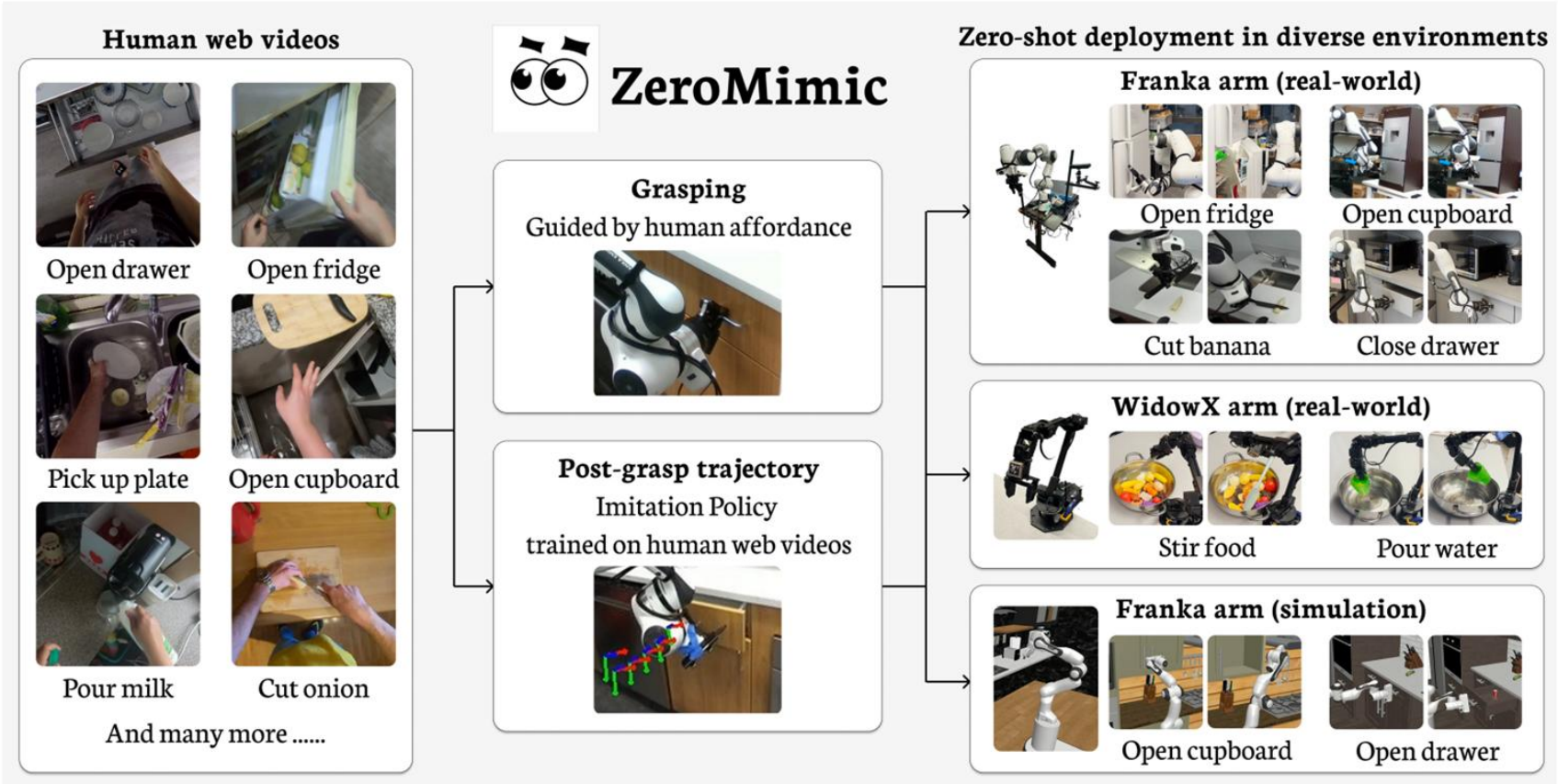


Supervisor: Nick Heppert

ZeroMimic: Distilling Robotic Manipulation Skills from Web Videos

https://arxiv.org/abs/2503.23877

- How to use human web videos to train a robot?
- Without robot-specific demonstrations or exploration

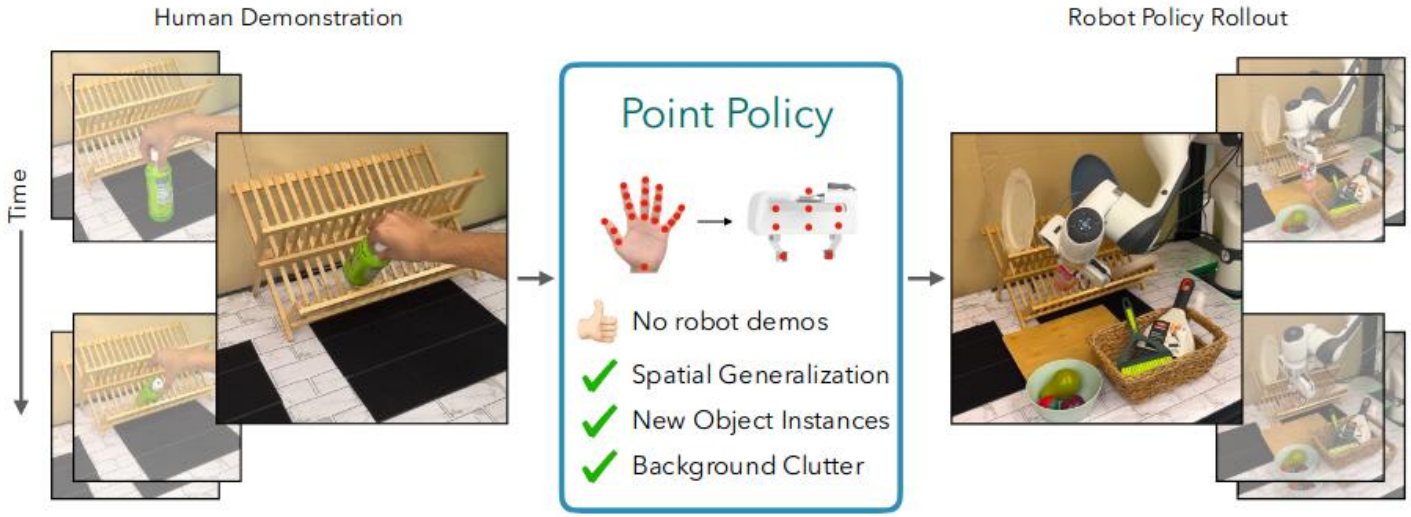


Supervisor: Nick Heppert

Point Policy: Unifying Observations and Actions with Key Points for Robot Manipulation

https://arxiv.org/abs/2502.20391

- Given human demonstrations in their own embodiment the task is to learn a manipulation policy
- Use keypoints as task abstraction to reduce the amount of demonstrations

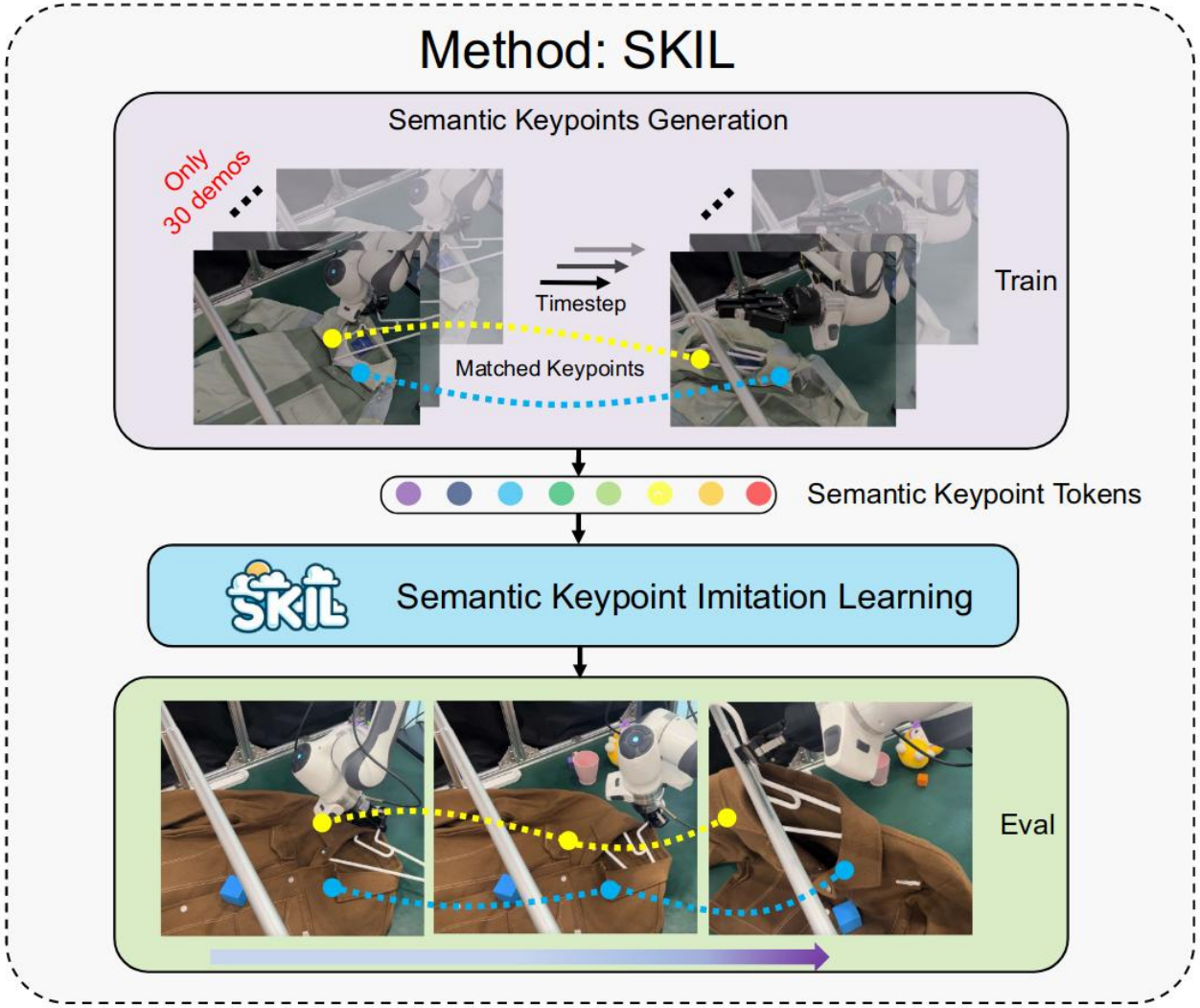


Supervisor: Nick Heppert

SKIL: Semantic Keypoint Imitation Learning for Generalizable Data-efficient Manipulation

https://arxiv.org/abs/2502.16932

- Use pre-trained semantic feature extractors (e.g. DINO) to automatically generate semantic and transferable keypoints.
- Learn tasks from less demos and allow category-level generalization

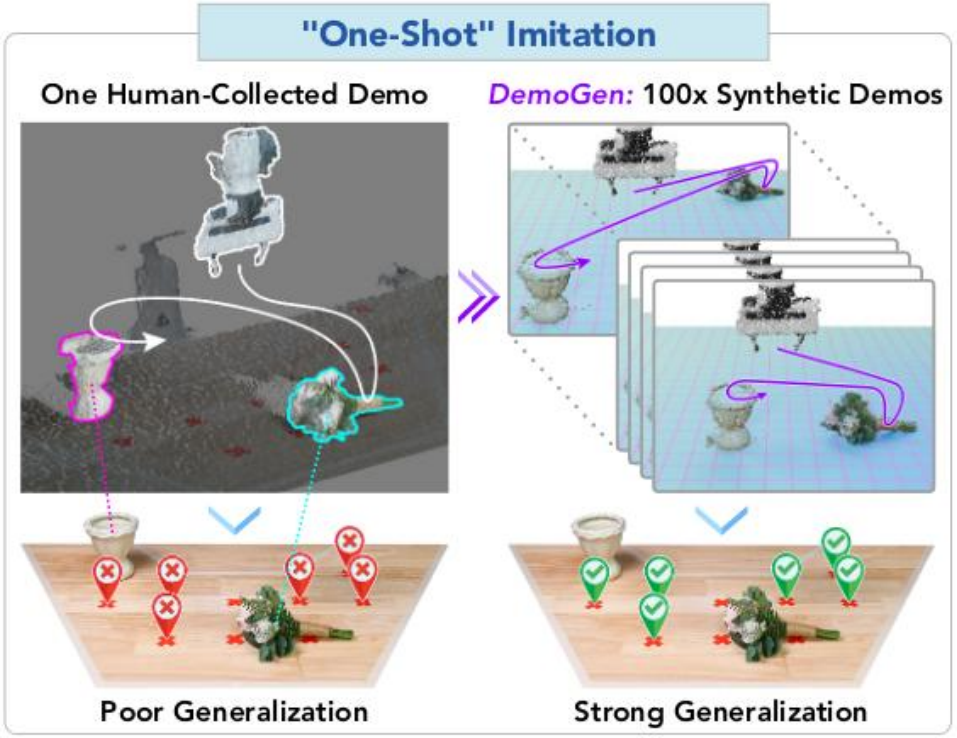


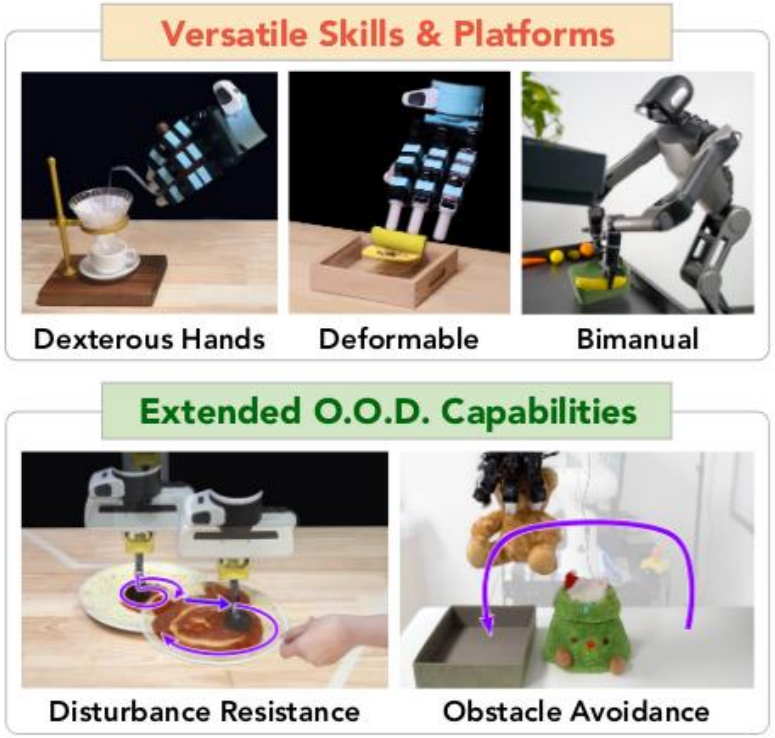
Supervisor: Nick Heppert

DemoGen: Synthetic Demonstration Generation for Data-Efficient Visuomotor Policy Learning

https://arxiv.org/abs/2502.16932

- Provide the robot only with a single demonstration
- Use simulation to collect varied data



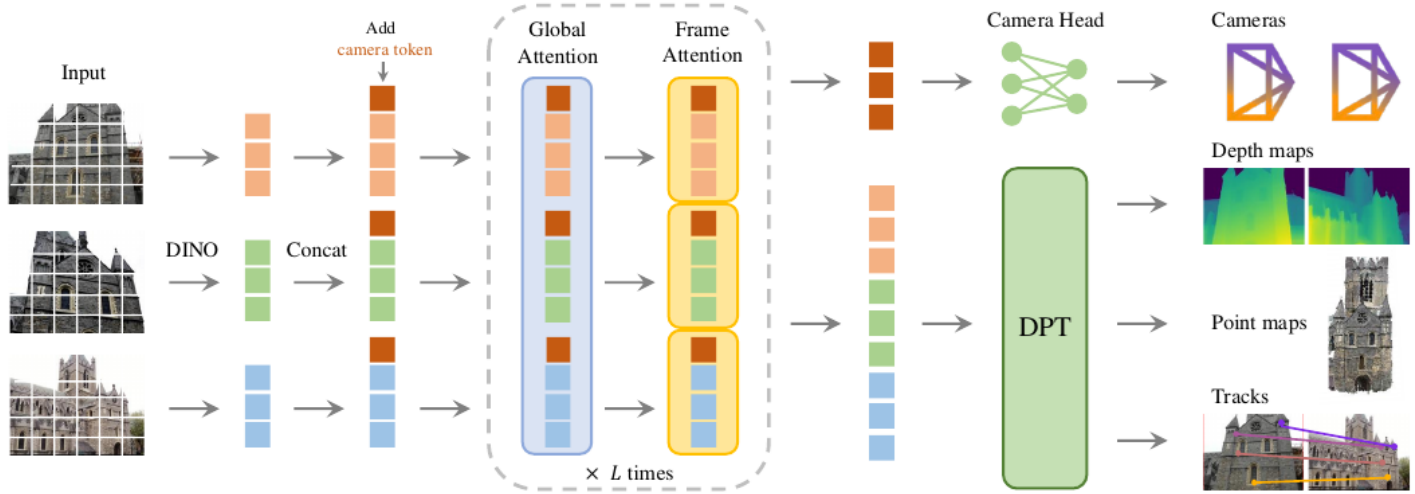


Supervisor: Simon Bultmann

VGGT: Visual Geometry Grounded Transformer

https://arxiv.org/pdf/2503.11651

- Efficient **feed-forward transformer**, that predicts camera parameters, depth maps, point maps, and 3D point tracks from 1-100+ images for **unified 3D scene understanding.**
- Processes **large image batches** in under a second, in a **single forward pass** without iterative geometry optimization.
- Also acts as **feature extractor** for various downstream **3D vision applications**.

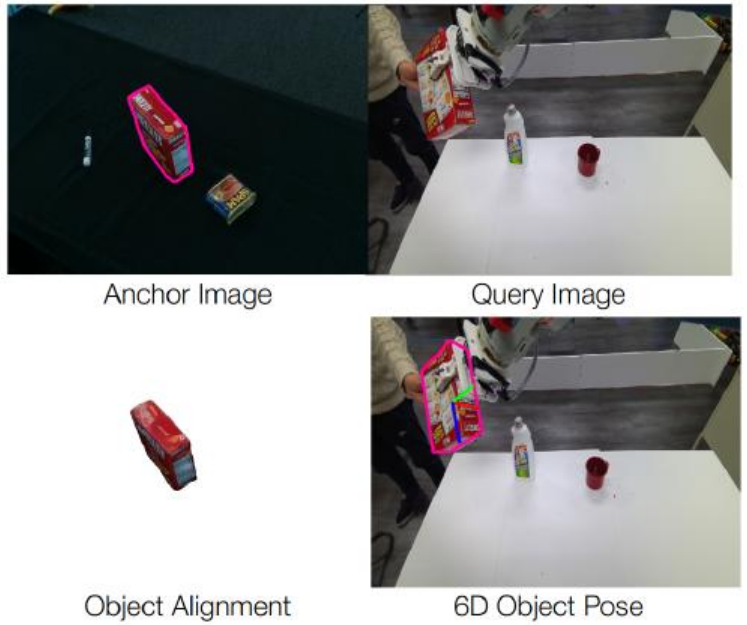


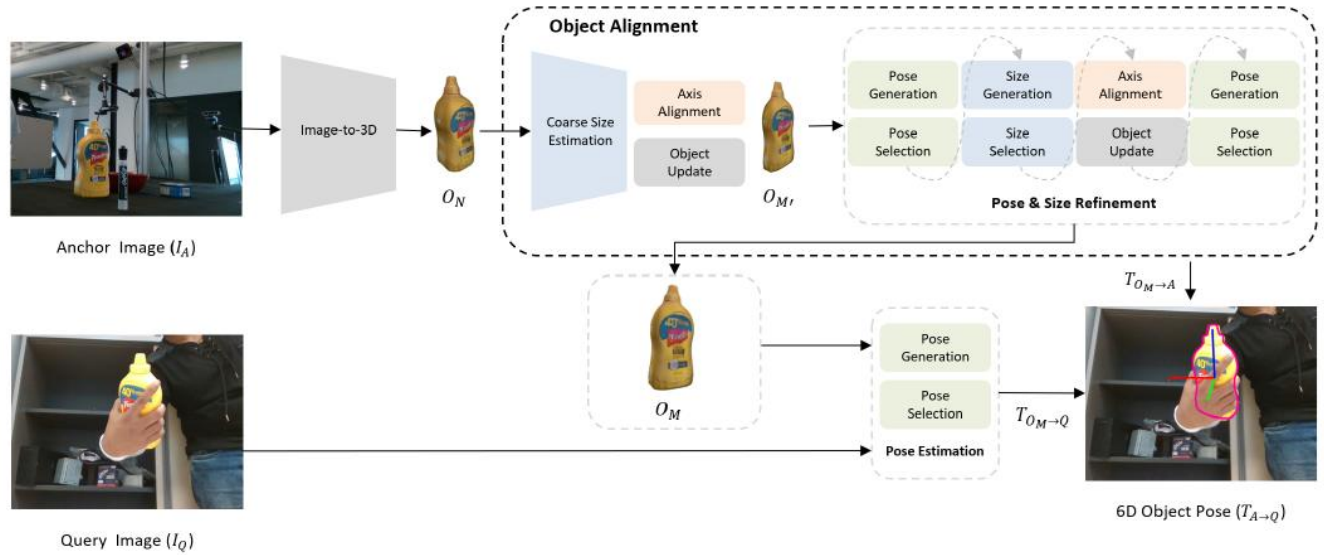
Supervisor: Simon Bultmann

Any6D: Model-free 6D Pose Estimation of Novel Objects

https://arxiv.org/pdf/2503.18673

- **Model-free 6D pose estimation** from a single RGB-D anchor image, without requiring CAD models or multi-view references.
- **Joint pose & size refinement** with render & compare approach enhances 2D-3D alignment and metric scale estimation.
- **Robust performance** in challenging scenarios, including occlusions, non-overlapping views, diverse lighting conditions, and large cross-environment variations.



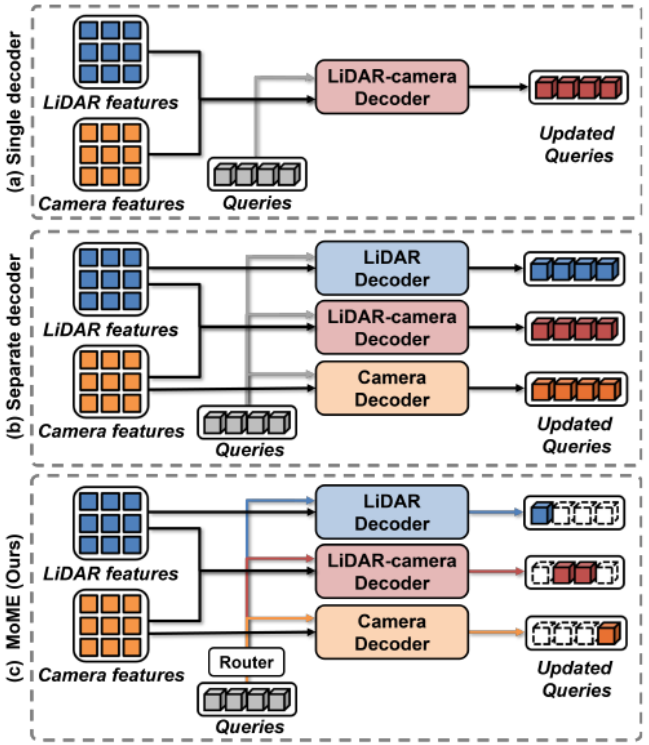


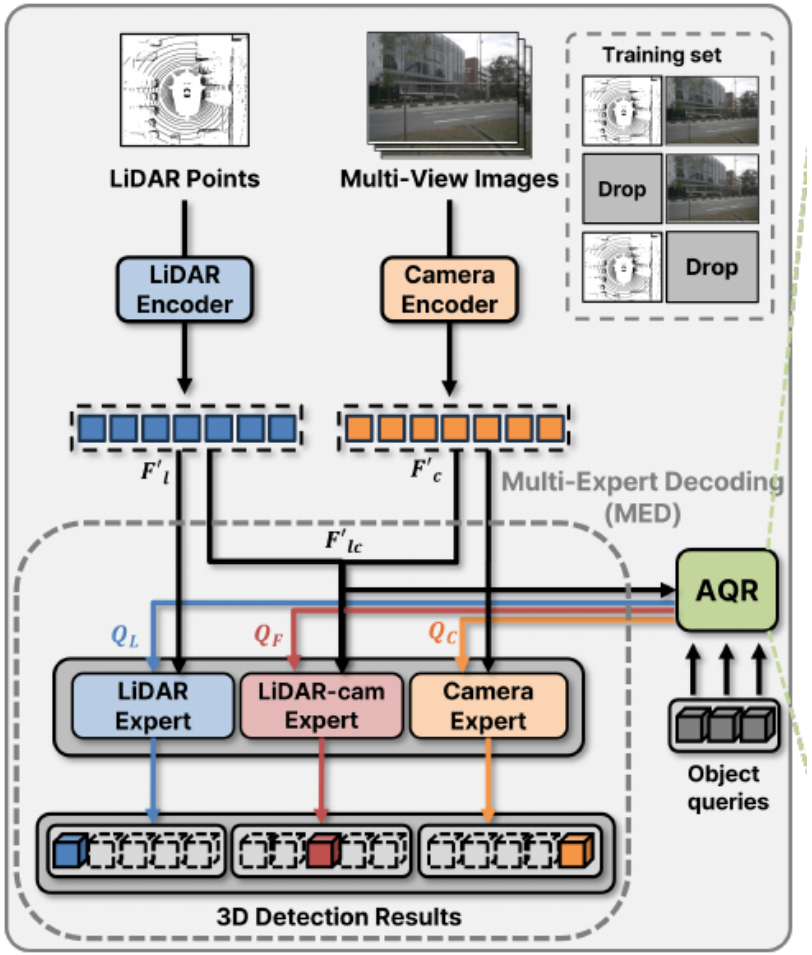
Supervisor: Simon Bultmann

Resilient Sensor Fusion under Adverse Sensor Failures via Multi-Modal Expert Fusion

https://arxiv.org/pdf/2503.19776

- Employs **three parallel decoders** for LiDAR, camera, and combined features.
- Utilizes an **Adaptive Query Router** (AQR) to dynamically **assign object queries to the most suitable expert** based on sensor input quality.
- SOTA results on nuScenes-R benchmark.



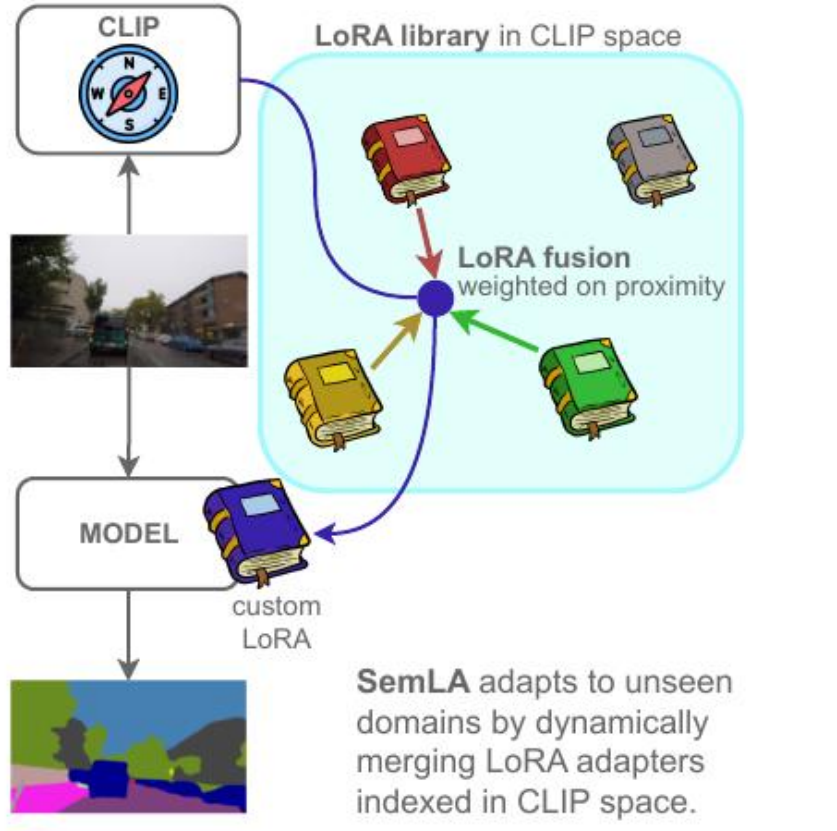


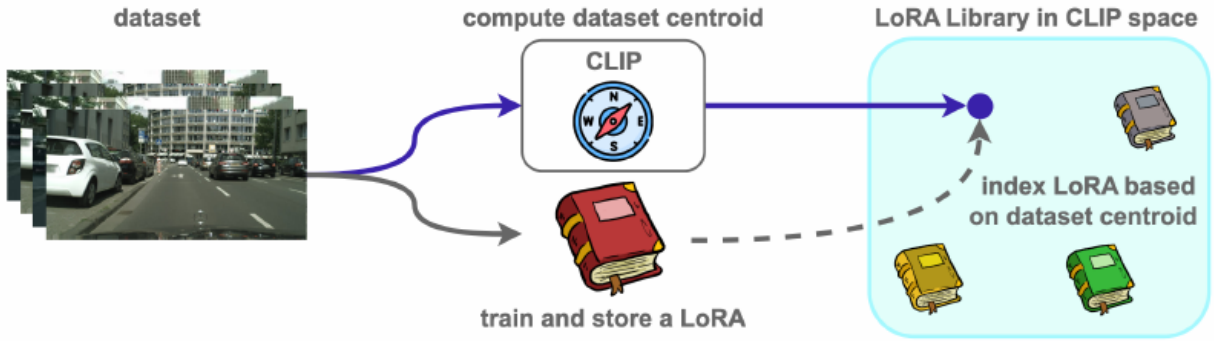
Supervisor: Simon Bultmann

Semantic Library Adaptation: LoRA Retrieval and Fusion for Open-Vocabulary Semantic Segmentation

https://arxiv.org/pdf/2503.21780

- SemLA enables **open-vocabulary** semantic segmentation **without retraining**, adapting to **new domains** during **inference**.
- Utilizes a library of **Low-Rank Adaptors** (LoRA) indexed by **CLIP embeddings**, merging relevant adapters based on domain similarity.
- Fusion of adapters enhances **explainability** by tracking adapter contributions.



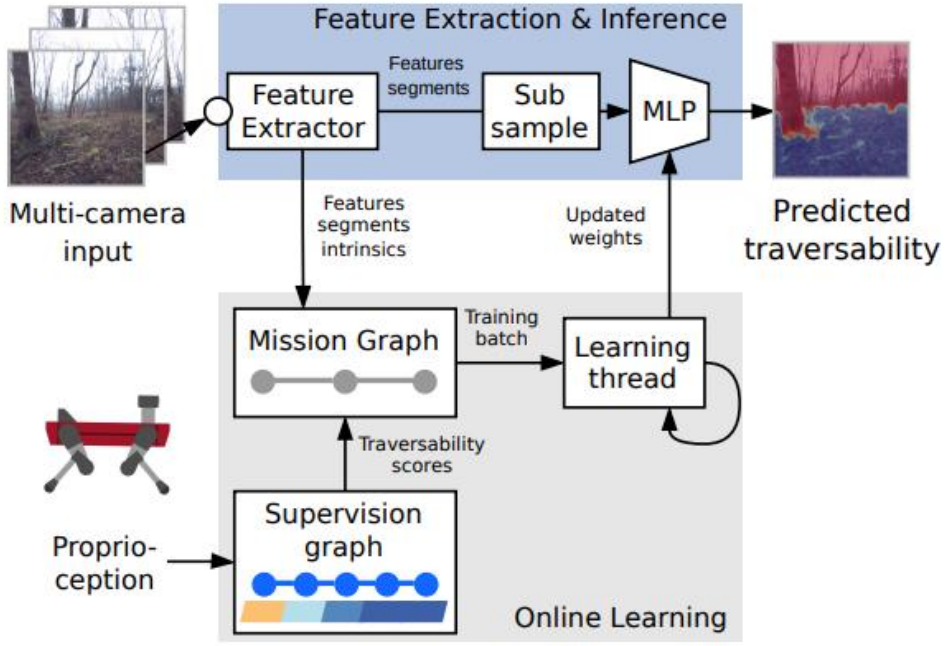


Supervisor: Iana Zhura

Wild Visual Navigation: Fast Traversability Learning via Pre-Trained Models and Online Self-Supervision

https://arxiv.org/pdf/2404.07110

- **Online self-supervised** traversability estimation with **continuous adaptation** to new environments
- Leverages **pre-trained visual models** (DiNO-ViT) to enhance traversability estimation, reducing need for labeled data
- Demonstrates **robust performance** in challenging terrains

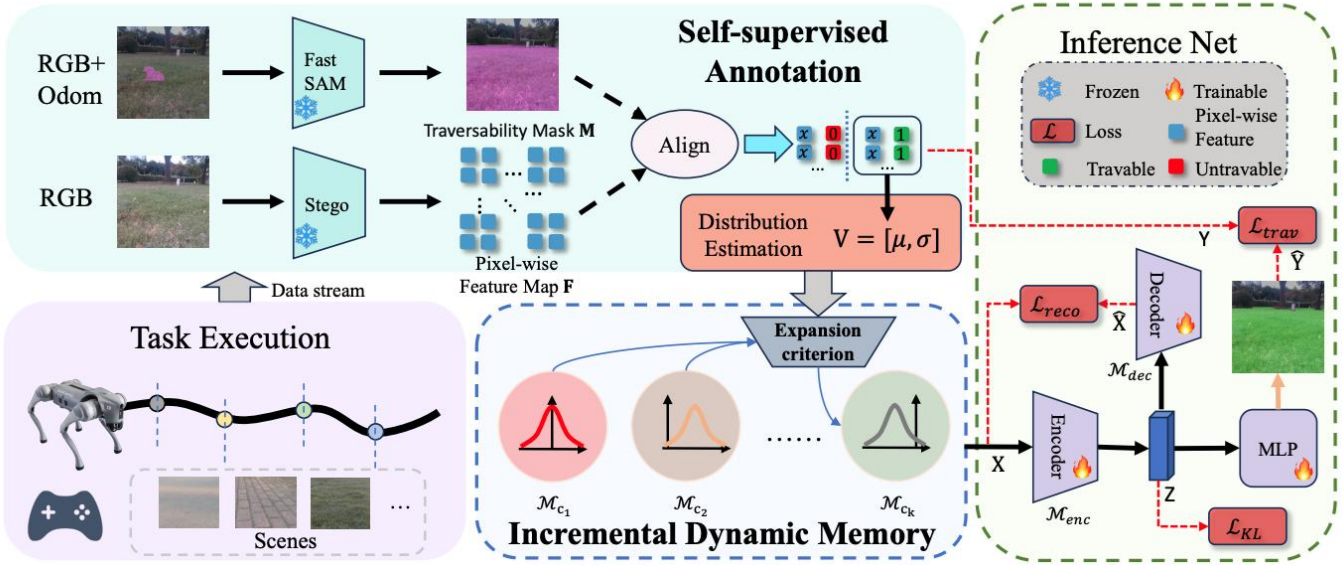


Supervisor: Iana Zhura

IMOST: Incremental Memory Mechanism with Online Self-Supervision for Continual Traversability Learning

https://arxiv.org/pdf/2409.14070

- **Self-supervised** traversability estimation
- **Incremental Memory**
- Utilizing **Fast SAM** and **Stego** for self-supervised annotations

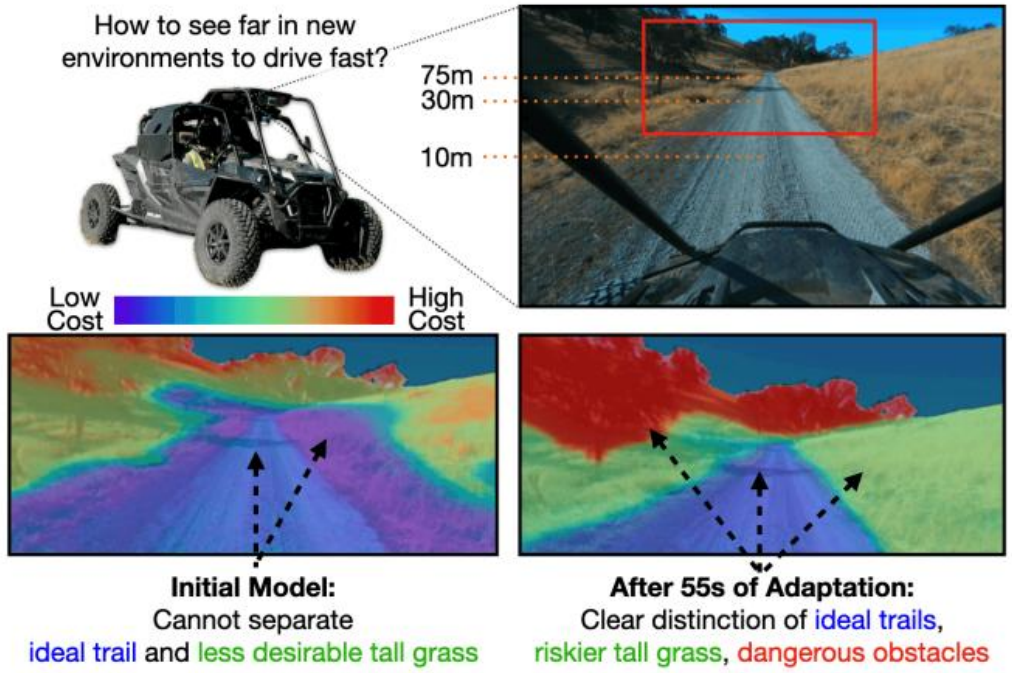


Supervisor: Iana Zhura

Learning-on-the-Drive: Self-supervised Adaptive Long-range Perception for High-speed Off Road Driving

https://arxiv.org/pdf/2306.15226

- **Online learning framework** for terrain estimation
- Learns from near-range **LiDAR** measurements
- **Cross-modal self-supervised learning** utilizes projected labels from 3D for images

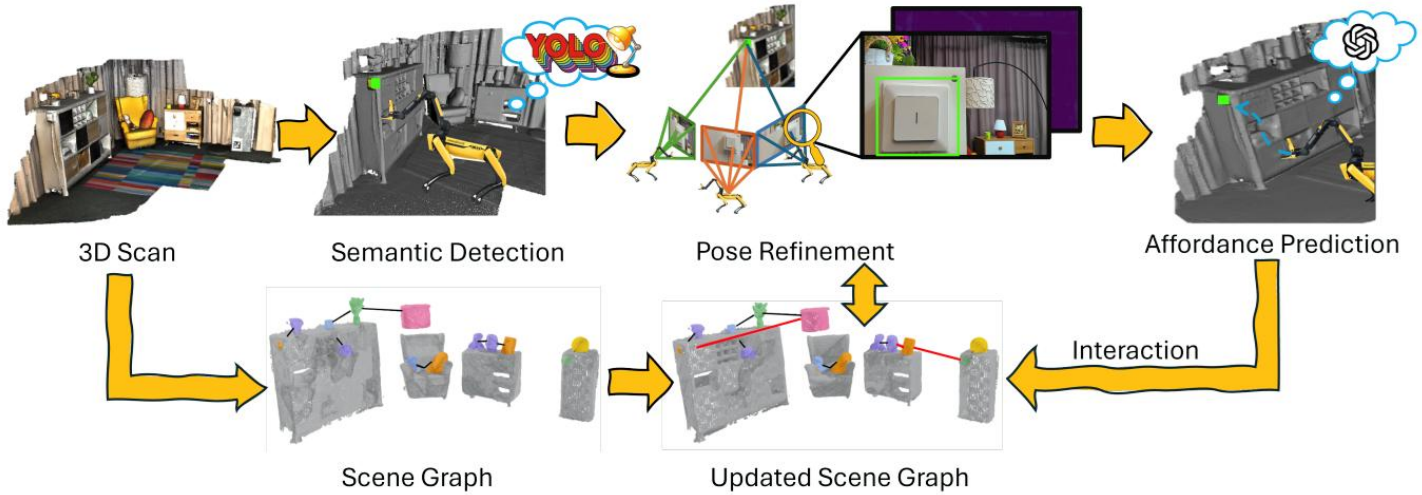


Supervisor: Iana Zhura

SpotLight: Robotic Scene Understanding through Interaction and Affordance Detection

https://arxiv.org/pdf/2409.14070

- **VLM-based affordance prediction** to estimate motion primitives for light switch interaction
- **Active perception:** learning through interaction and exploration of the environment
- Utilizes 3D scene graph representation



?

Questions

Questions or Comments

Simon Bultmann

Robot Learning Lab

bultmann@cs.uni-freiburg.de