universität freiburg

SS24 Seminar
Learning with Limited Supervision

José Arce

Robot Learning Lab

17 April 2024

ellis 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 **22.04.2023**

Please check the course website for more information:

https://rl.uni-freiburg.de/teaching/ss24/seminar-limited-supervision

Important Dates

Event	Date	Time
Lecture 1: Introduction *	17.04.2024	10:00
HISinOne registration + Paper Selection	22.04.2024	
Place allocation	26.04.2024	
Paper assignment	30.04.2024	
Supervisor Meeting	06.2024	
Lecture 2: *How to do a good presentation* *	21.06.2024	10:00
Lecture 3: Block Seminar Presentations *	19.07.2024	9:00 - 17:00
Paper Summary submission	02.08.2024	< 23:59

*** Mandatory in-person attendance**

Evaluation Criteria

Evaluation	Due Date
Seminar Presentation	19.07.2024
Paper Summary	02.08.2024

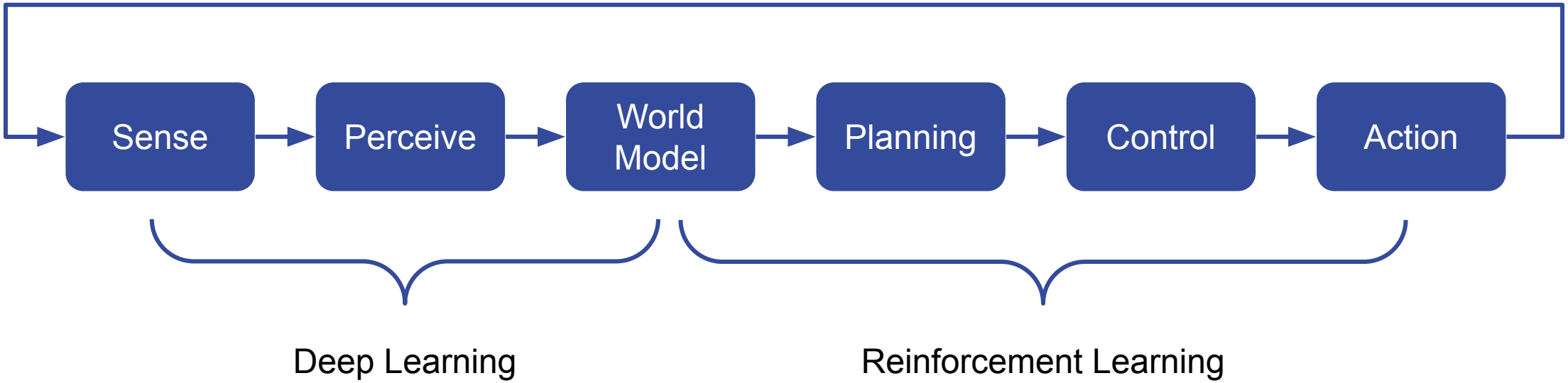
- **Presentation:** at most **20 min**.
- **Summary:** at most **7 pages** excluding bibliography and figures.
- **Final grade:**
  - Presentation (slides & delivery) + Summary + **Seminar Participation**.

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** ?

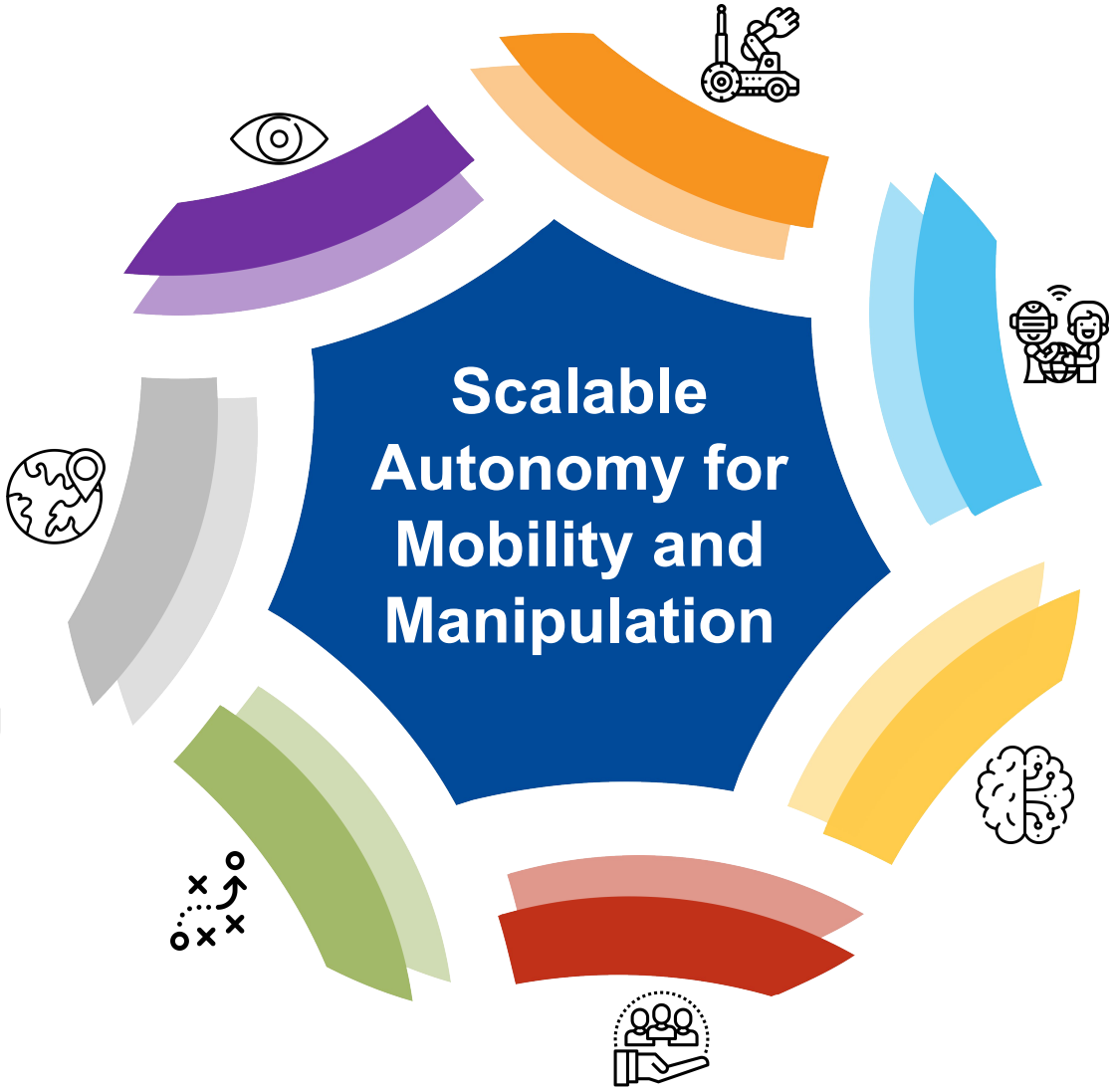
Research Areas

Perception
- Recognition
- Depth Estimation
- Motion Estimation

State Estimation
- Tracking & Prediction
- SLAM
- Registration

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

Responsible Robotics
- Fairness
- Explainability & Privacy
- Practical Ethics

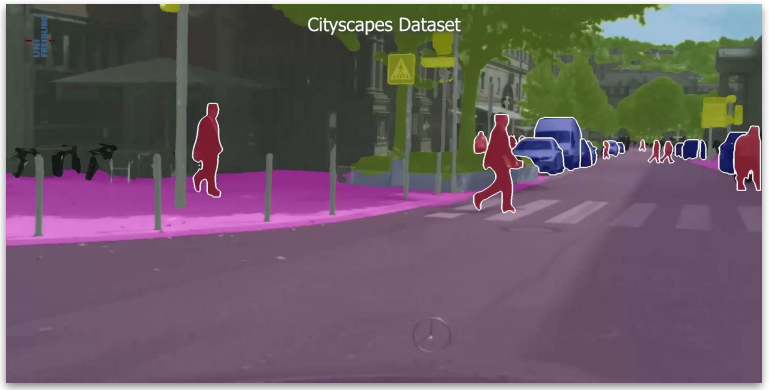


Mobile Manipulation
- Whole-Body Motion
- Long-Horizon Reasoning
- Planning for Sensing

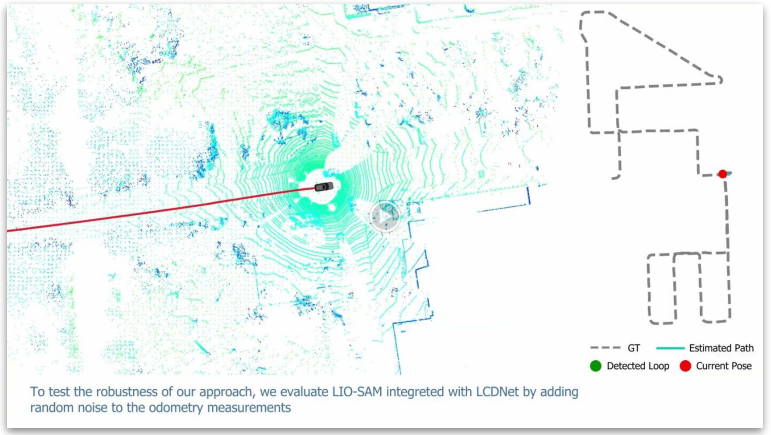
Human-Robot Interaction
- Socially-Compliant Behavior
- Human-Robot Collaboration
- Behavior Adaptation & Safety

Learning Fundamentals
- Socially-Supervised Learning
- Continual & Interactive Learning
- Multimodal & Multitask Learning

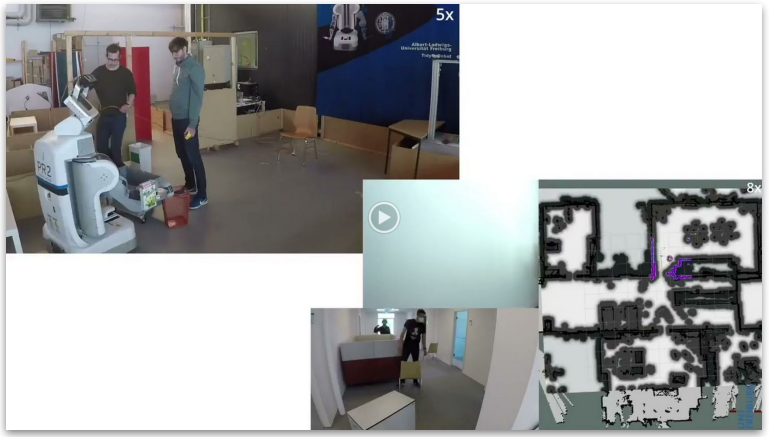
Many Seminal Works



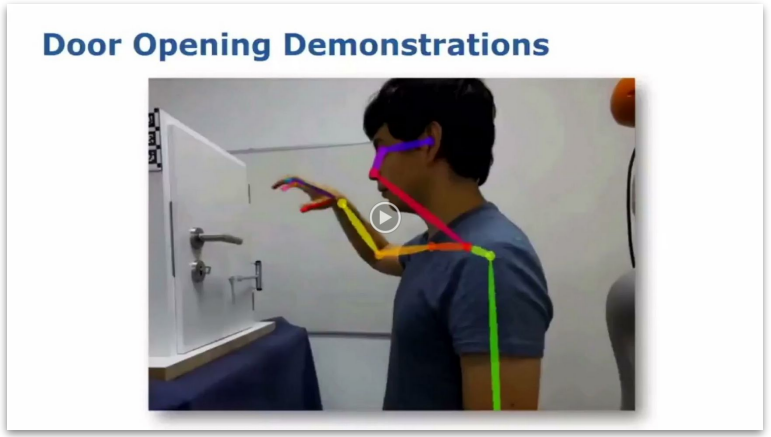
Scene Understanding



Simultaneous Localization and Mapping

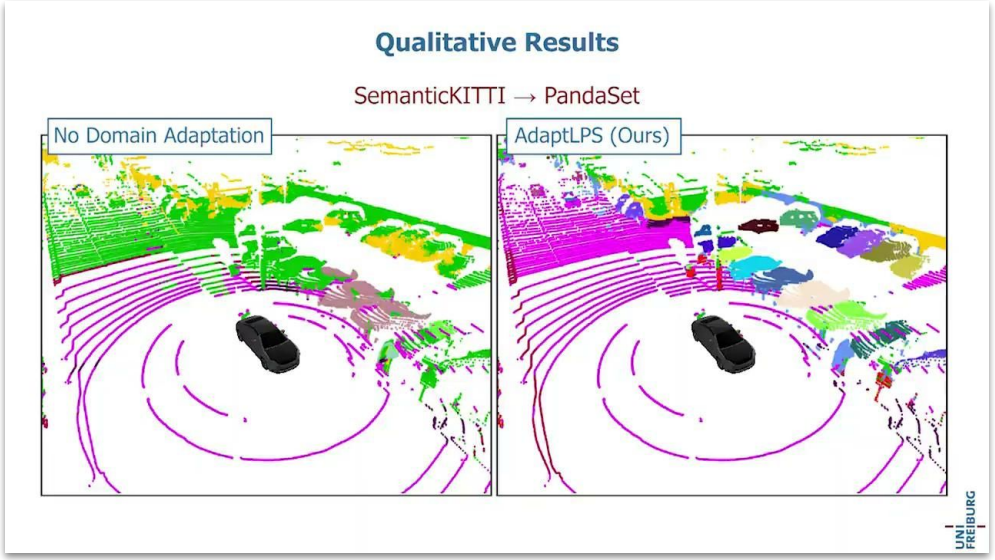


Motion Planning



Learning from Demonstrations

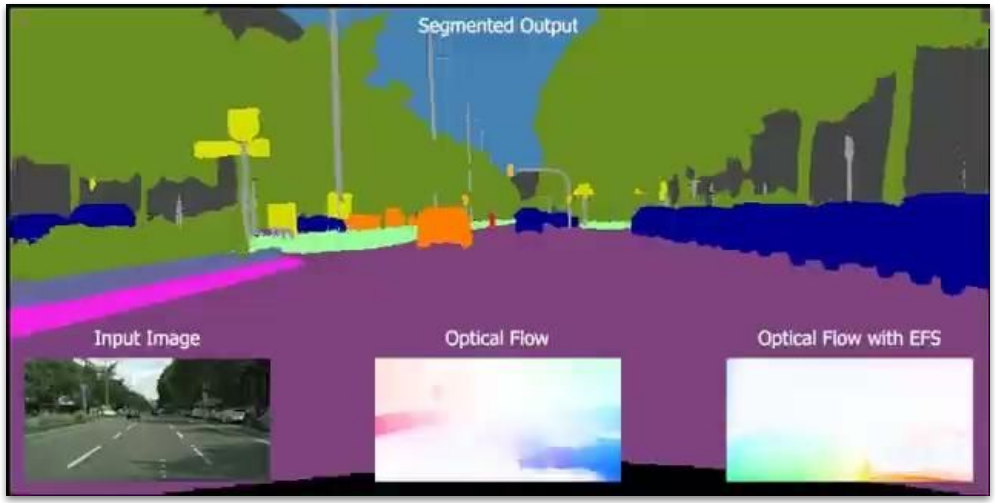
Robotic Perception — Mobility



Unsupervised LiDAR Domain Adaptation

Besic, Gosala, Cattaneo, Valada

RA-L ‘22



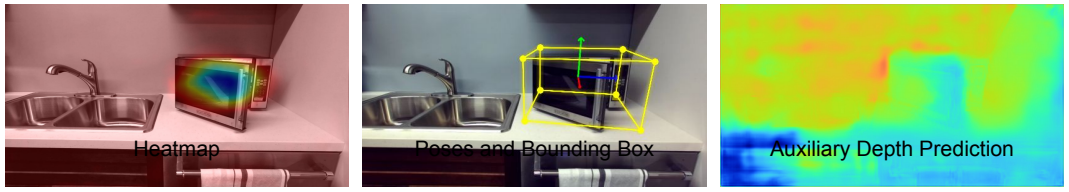
Semantic Motion Segmentation

Vertens, Valada, Burgard

ICRA ‘17

Robotic Perception — Manipulation

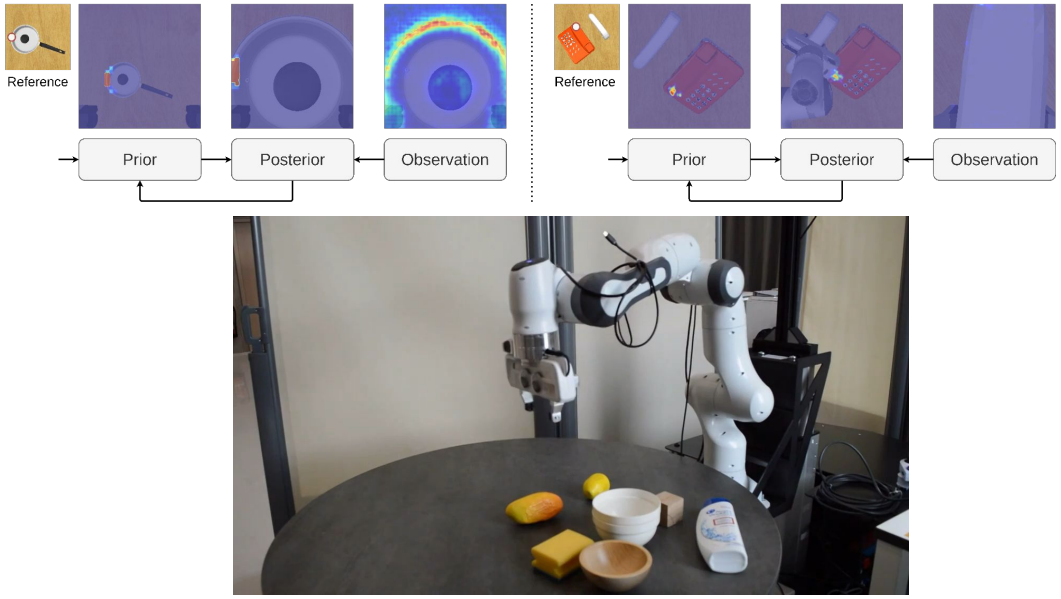
Single-Shot Reconstruction



Category and Joint Agnostic Reconstruction of ARTiculated Objects

Heppert, et al
CVPR ‘23

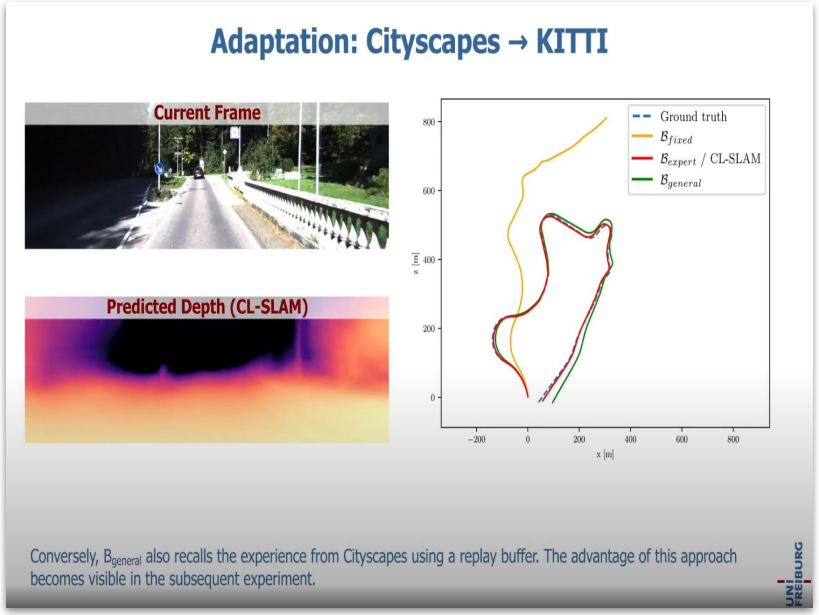
Learning scale-invariant compact representations for mobile manipulation



Bayesian Scene Keypoints for Deep Policy Learning in Robotic Manipulation

von Hartz, et al
RA-L ‘23

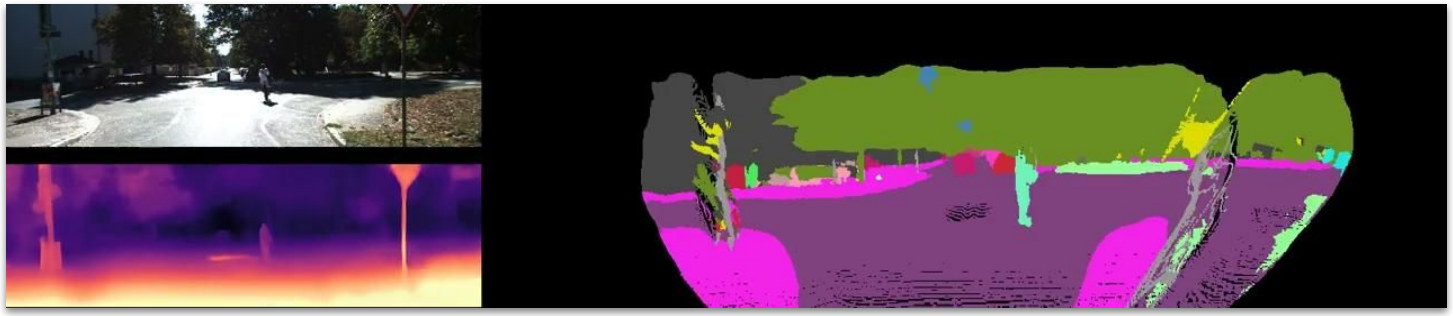
Mapping and Localization



Continual SLAM

Vödisch, Cattaneo, Burgard, Valada

ISSR ‘22



Continual Depth Estimation and Segmentation

Vödisch, Petek, Burgard, Valada

RSS ‘23

III.

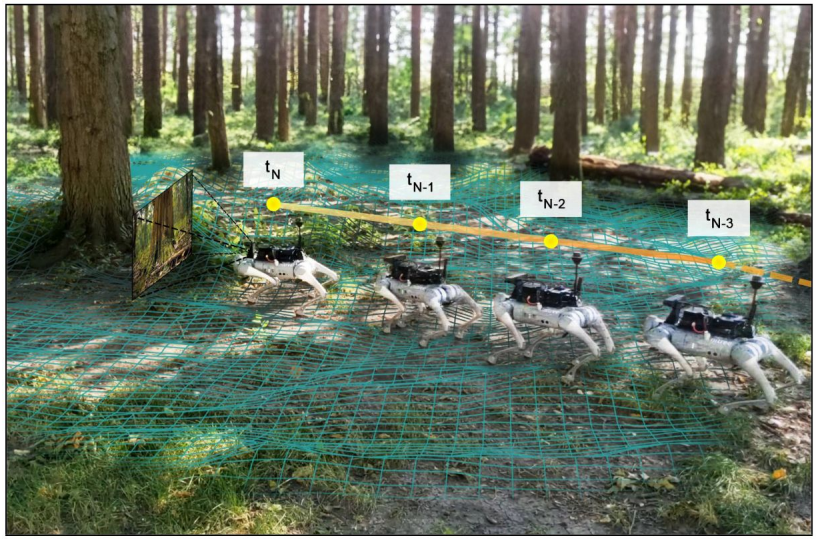
Topics

Seminar Papers

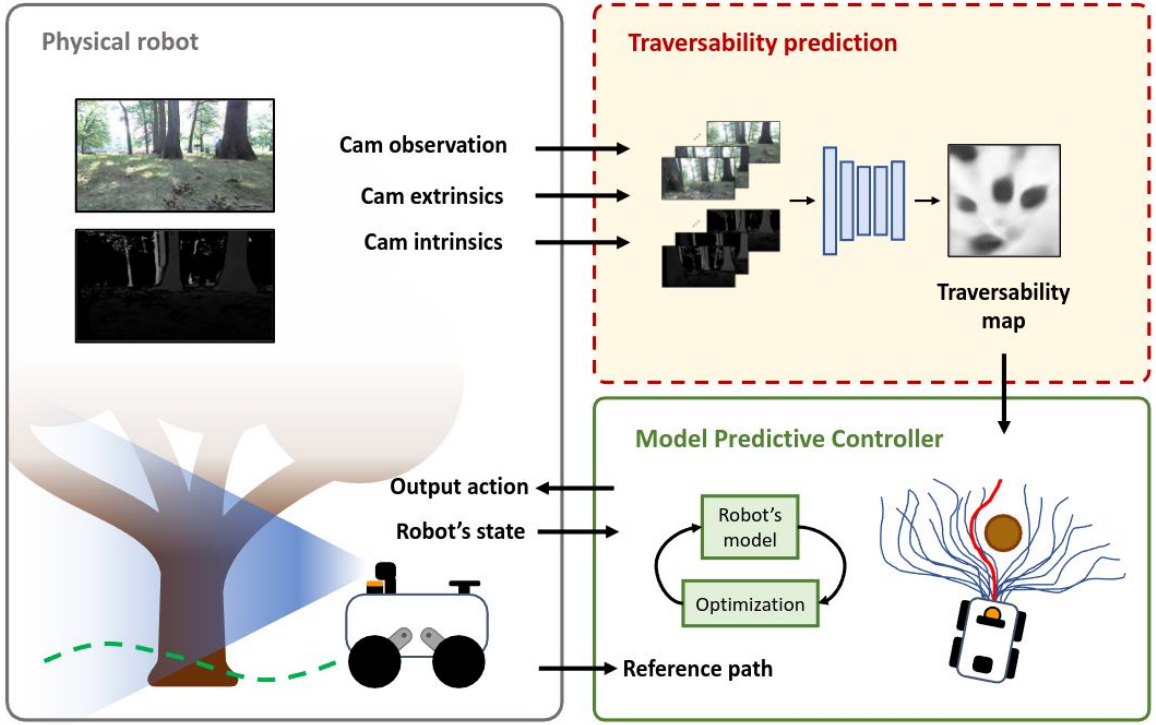
Supervisor: José Arce

WayFASTER: a Self-Supervised Traversability Prediction for Increased Navigation Awareness

https://arxiv.org/pdf/2402.00683.pdf



- Predicts terrain **traversability** from RGB-D + GNSS inputs in a BEV map.
- Uses a Receding Horizon Estimator (**RHE**) to estimate the poses and label the data.
- The traversability index depends on whether the **actual motion** follows the **control sent**.

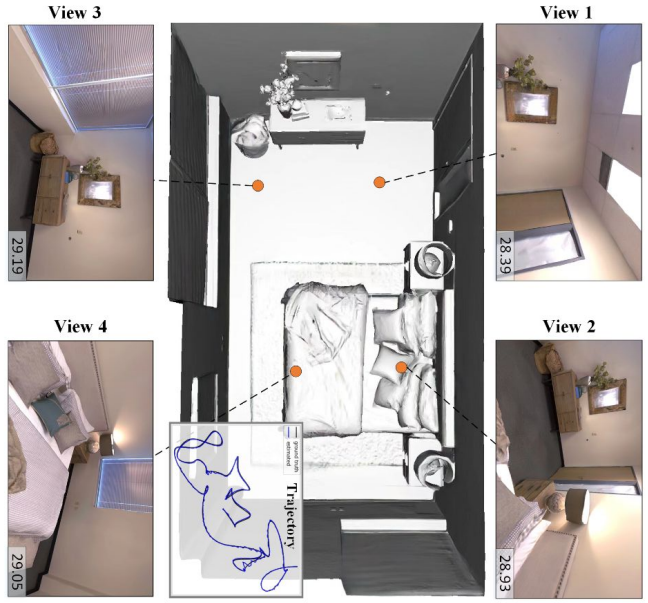


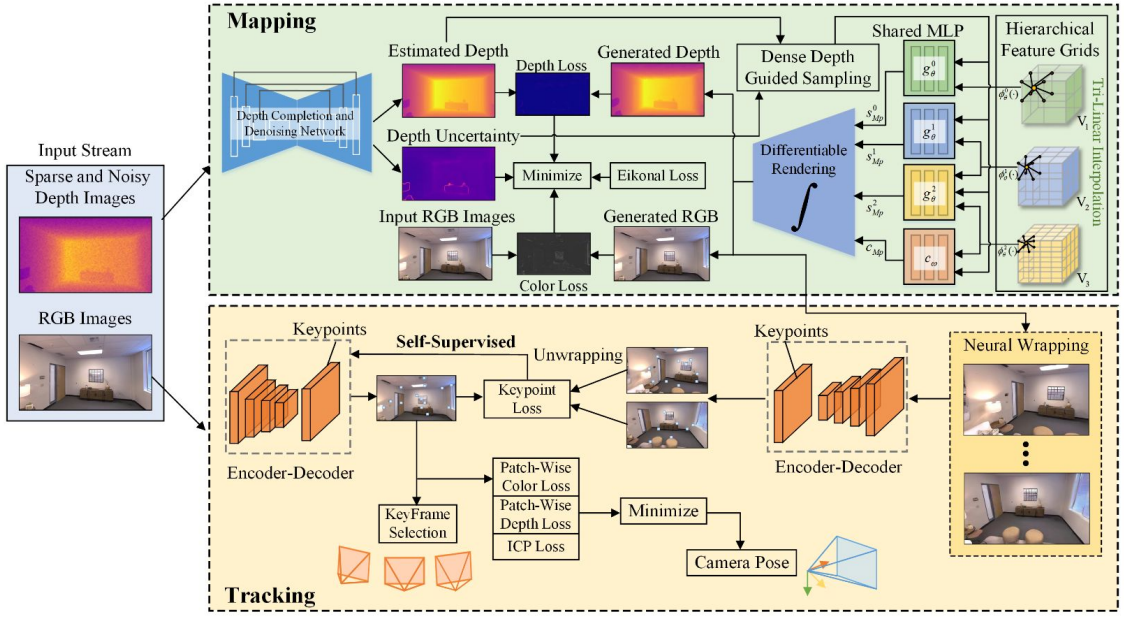
Supervisor: José Arce

NeSLAM: Neural Implicit Mapping and Self-Supervised Feature Tracking With Depth Completion and Denoising

https://arxiv.org/pdf/2403.20034.pdf

- Predicts the **ego-motion**, denoises the input **depth** and generates a **NeRF** map from RGB-D inputs.
- SS in the form of comparing **rendered images** from the NeRF with the **input** and tracking keypoints.

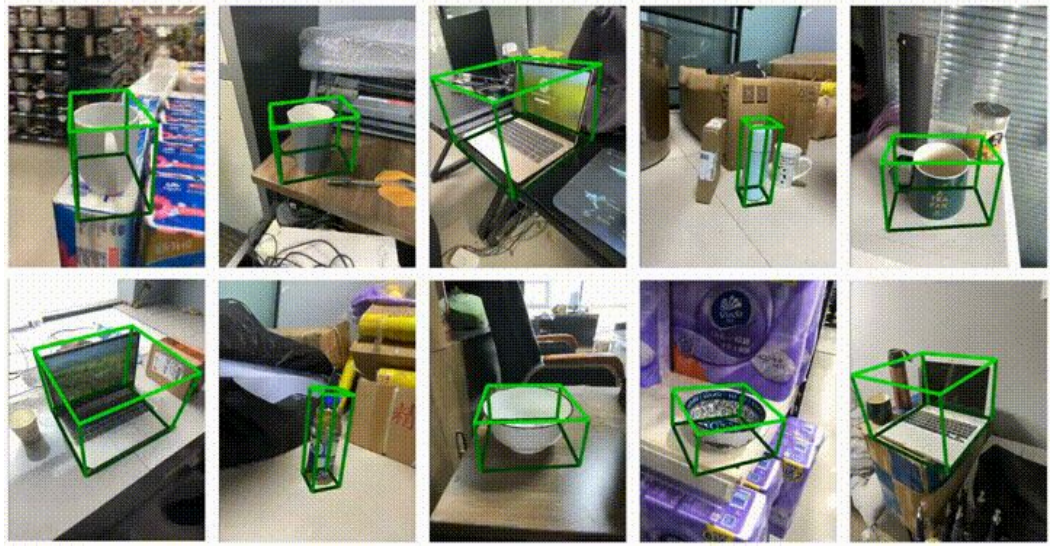




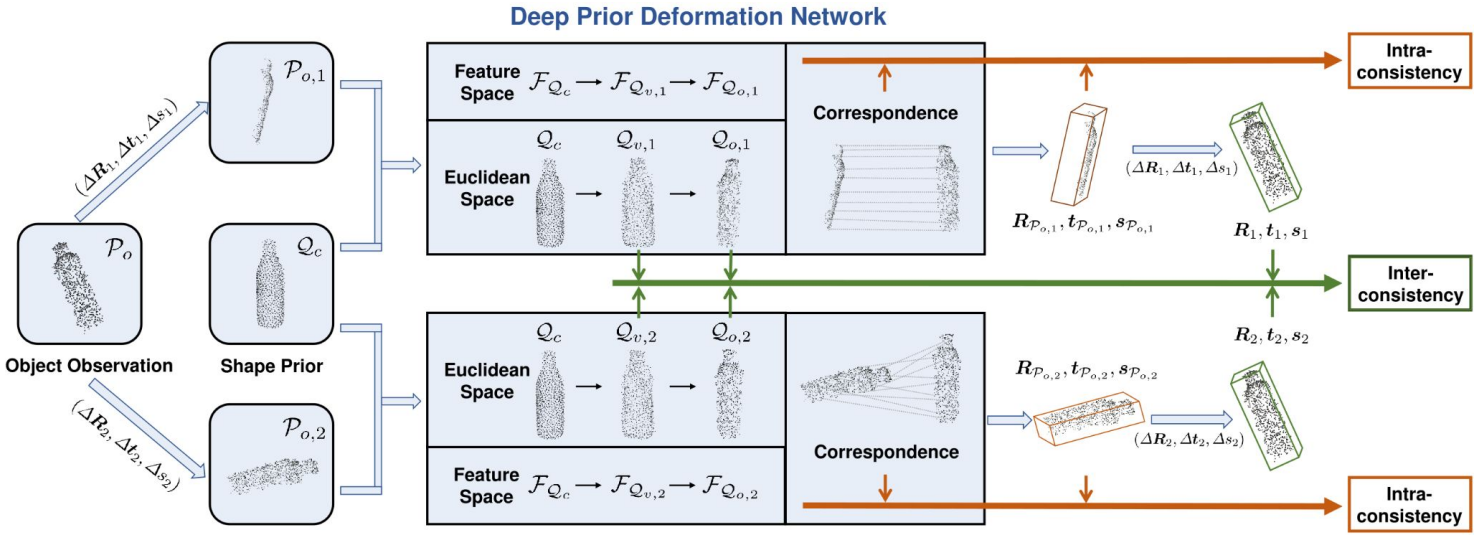
Supervisor: José Arce

Category-Level 6D Object Pose and Size Estimation using Self-Supervised Deep Prior Deformation Networks

https://arxiv.org/pdf/2207.05444.pdf



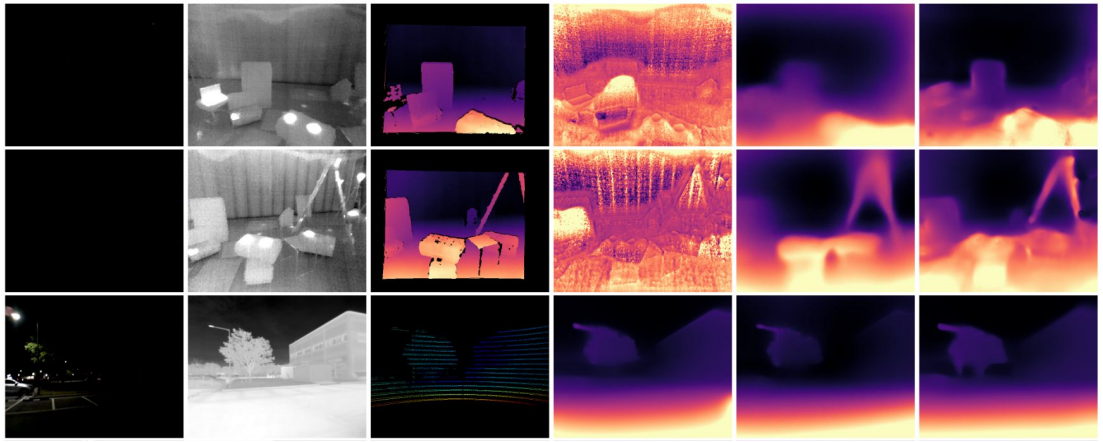
- Estimates rigid transformation (**R**, **t**, s) for an object from an RGB-D scan.
- Takes a **shape prior** as input in the form of an object's point cloud.
- Combines a supervised loss on a small set of synthetic data and a self-supervised loss comprised of **intra- and inter-consistency** terms.



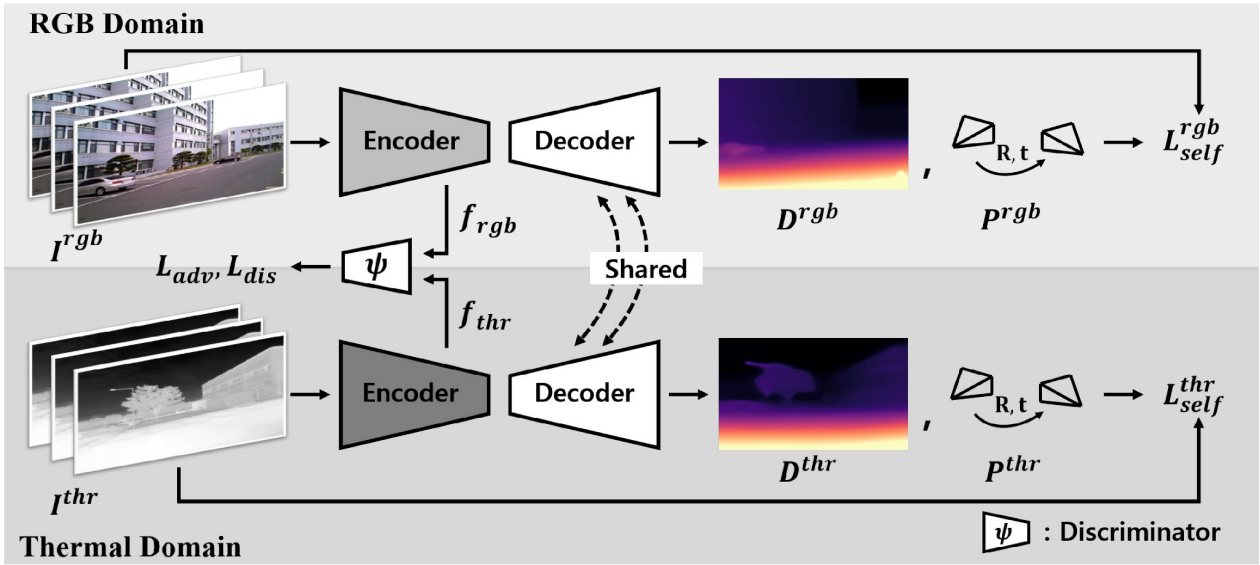
Supervisor: José Arce

Self-Supervised Monocular Depth Estimation From Thermal Images via Adversarial Multi-Spectral Adaptation

OpenAccess



- Thermal images have **low contrast** and **poor features**.
- **Unsupervised Domain Adaptation** from RGB to Thermal.
- Domain-specific encoders (RGB / Thermal).
- Domain-shared decoder.
- SS adversarial feature-level discriminator.
- SS self reconstruction loss.



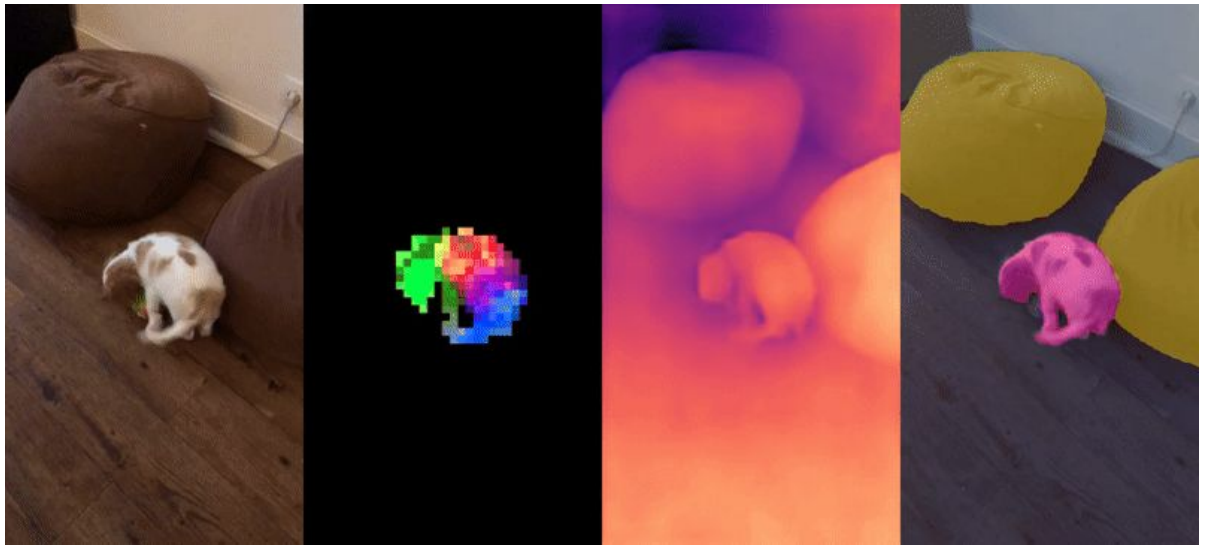
Supervisor: Niclas Vödisch

DINOv2: Learning Robust Visual Features without Supervision

https://arxiv.org/pdf/2304.07193.pdf

(Vehicles) (Birds / Airplanes)

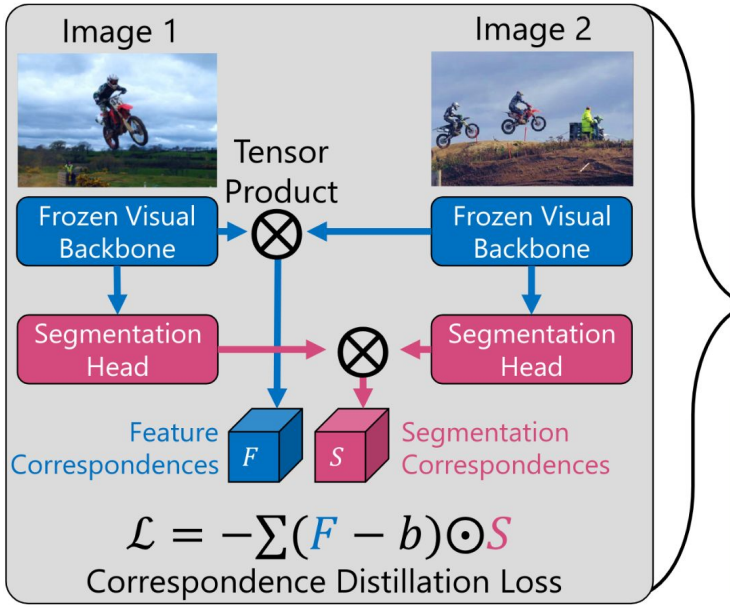
- Unsupervised **visual foundation model** generating semantically rich image features.
- Powerful visual transformer (ViT) architecture trained on an automatically curated dataset.
- Utilized as **task-agnostic pretraining strategy** for various downstream applications.



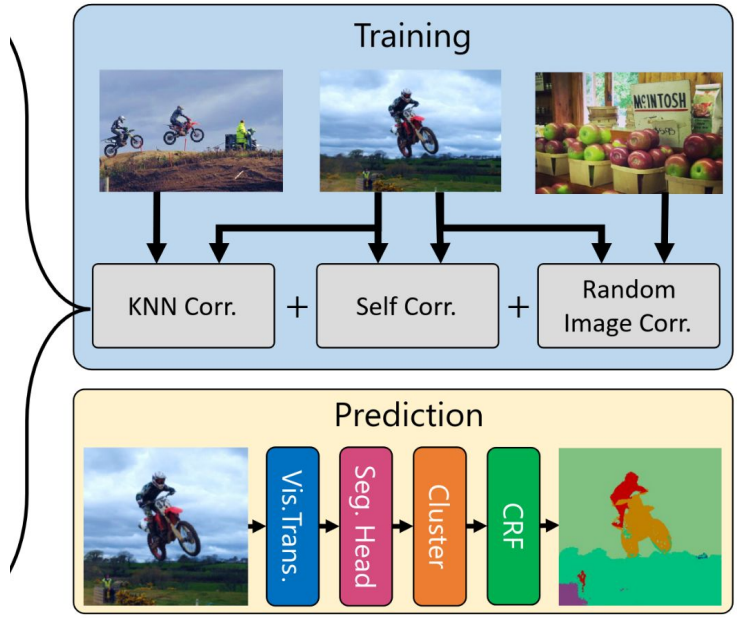
Supervisor: Niclas Vödisch

STEGO: Unsupervised Semantic Segmentation by Distilling Feature Correspondences

https://arxiv.org/pdf/2203.08414.pdf



- Semantic **segmentation without human annotations** inspired by self-supervised feature learning.
- Exploit descriptive image **features from visual foundation model**.
- Correspondence distillation loss for **contrastive learning** aligning visual features and image segmentation.

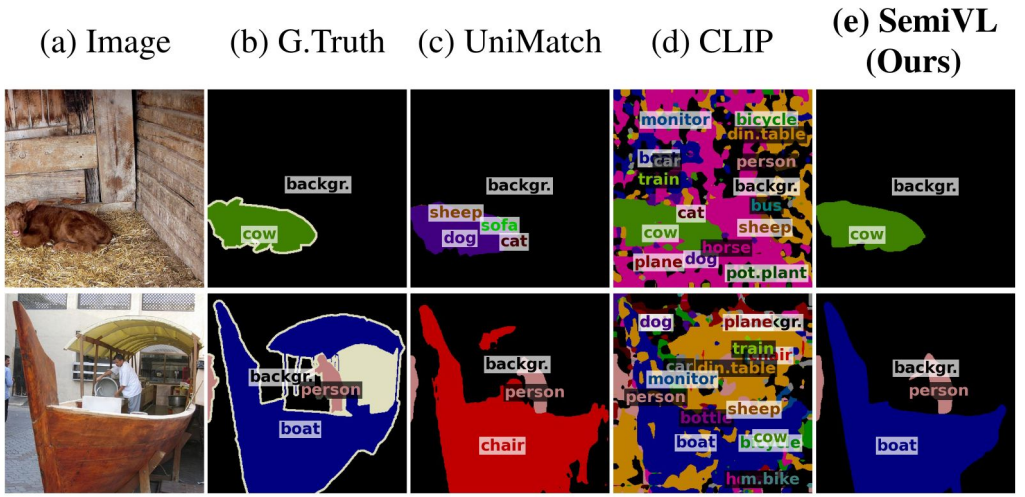


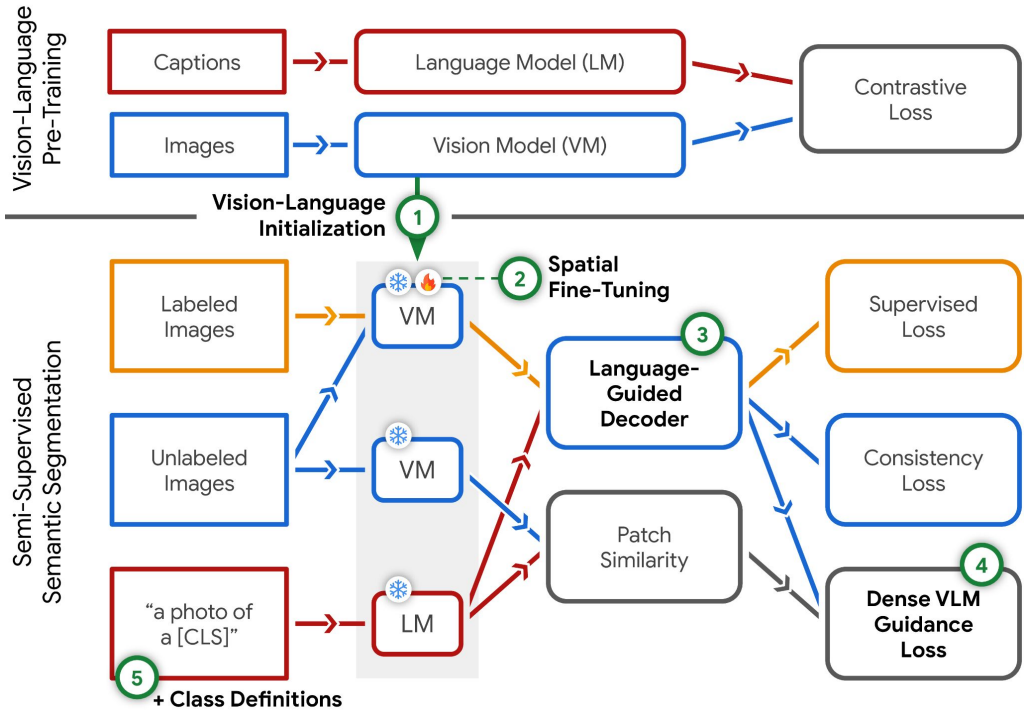
Supervisor: Niclas Vödisch

SemiVL: Semi-Supervised Semantic Segmentation with Vision-Language Guidance

https://arxiv.org/pdf/2311.16241.pdf

- **Efficient semantic segmentation** aiming to reduce the amount of human annotations.
- Exploit semantically rich **features from frozen visual-language foundation models** (VLMs).
- **Dense VLM guidance loss** to address label inconsistencies between dataset definitions.



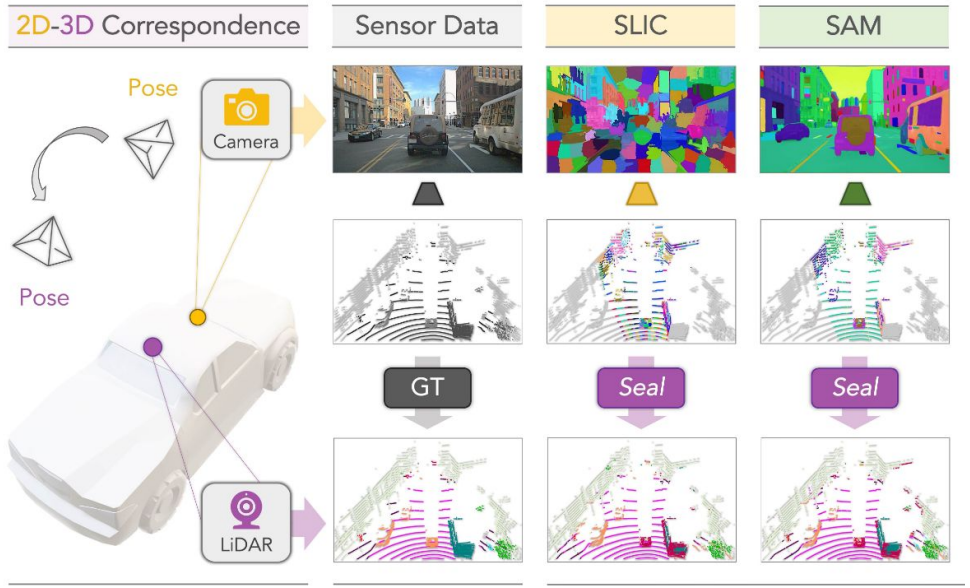


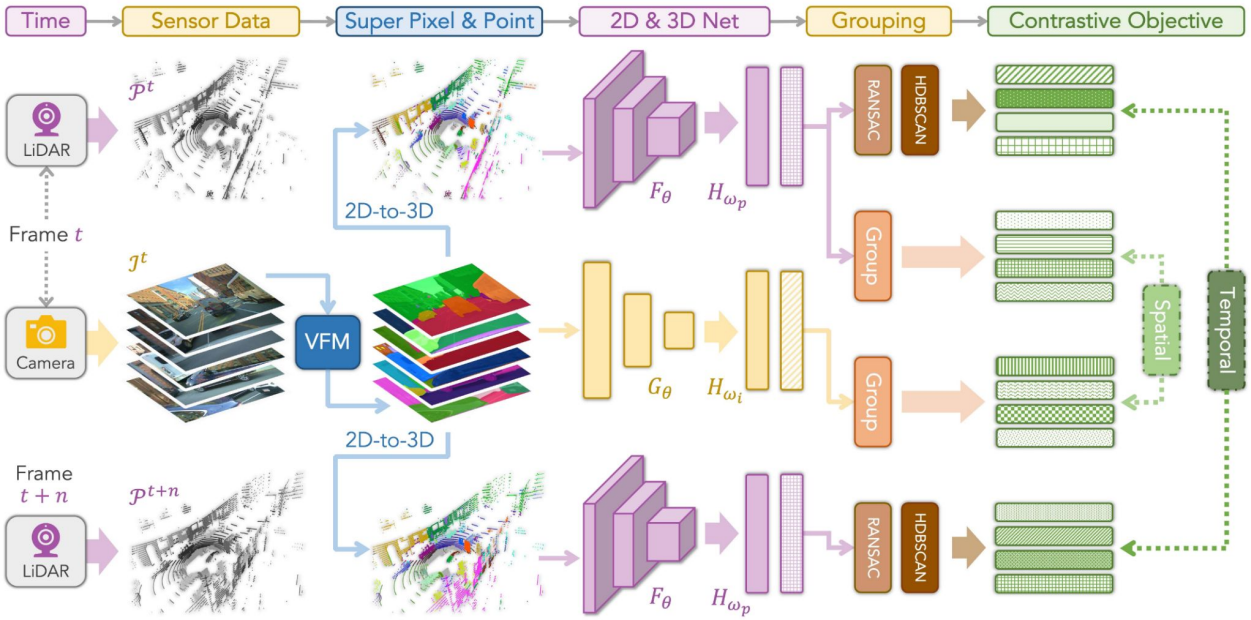
Supervisor: Niclas Vödisch

Seal: Segment Any Point Cloud Sequences by Distilling Vision Foundation Models

https://arxiv.org/pdf/2306.09347.pdf

- Exploit **semantic predictions from visual foundation models** to annotate LiDAR point clouds.
- **Temporal-spatial contrastive learning** between paired LiDAR and camera features.
- **Temporal consistency regularization** between point segments at different timestamps.

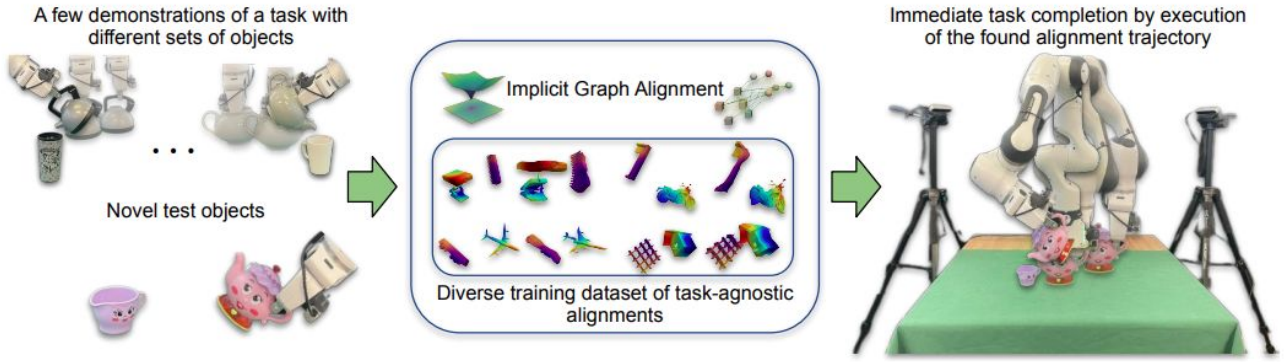




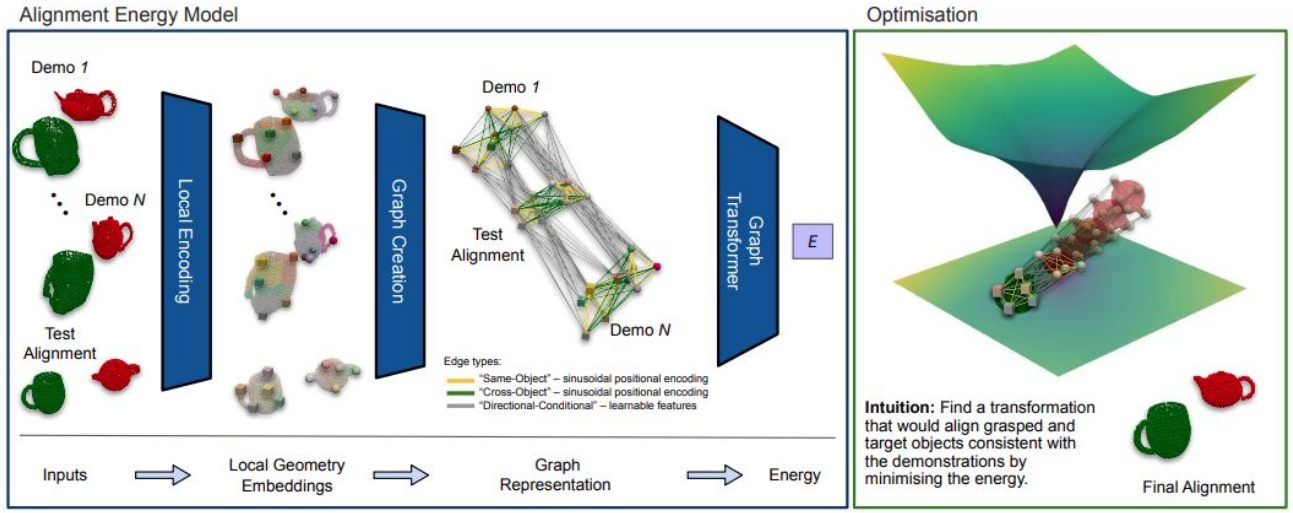
Supervisor: Adrian Röfer

Few-Shot In-Context Imitation Learning via Implicit Graph Alignment

https://arxiv.org/pdf/2310.12238.pdf



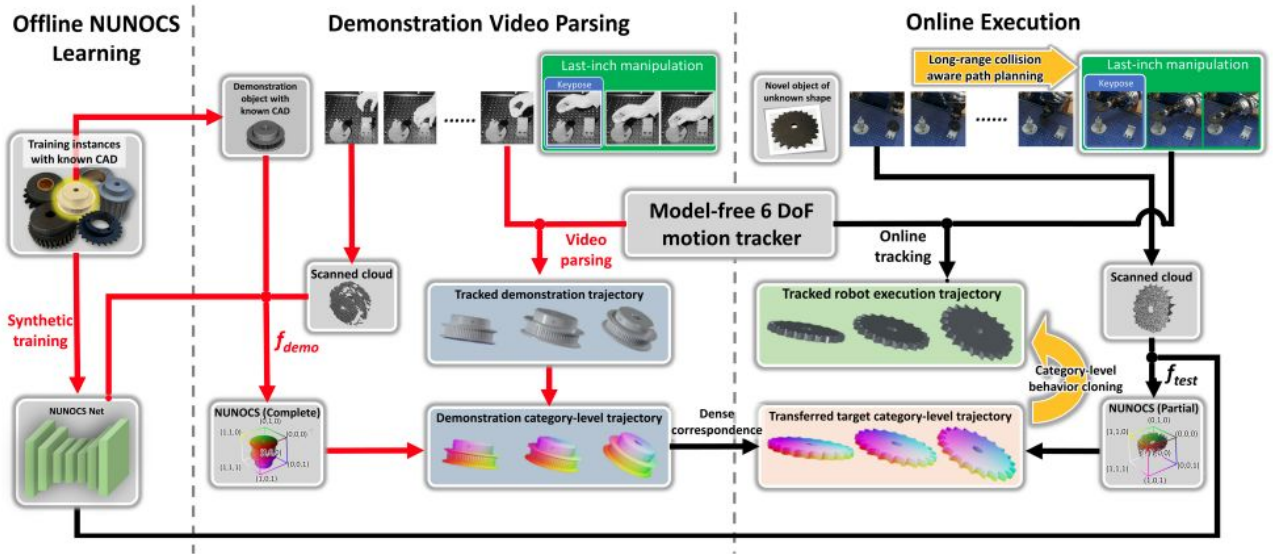
- Learn to reproduce relative object trajectories 1-Shot
- Self-supervised learning of feature embedding to track object points
- Embedding allows Zero-Shot transfer to new object instances

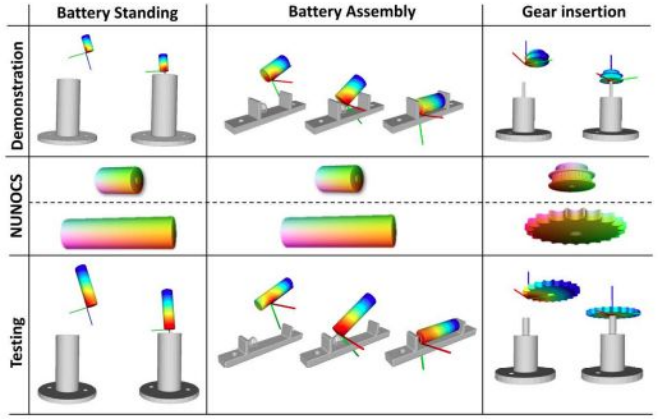


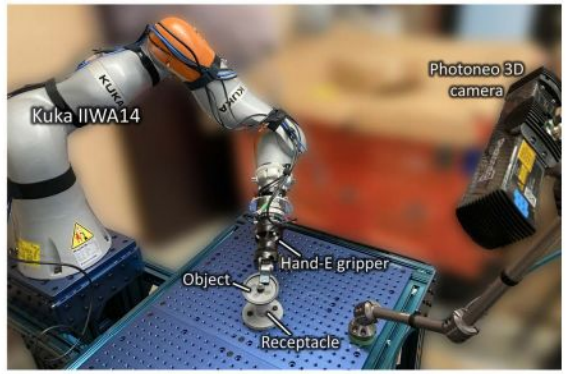
Supervisor: Adrian Röfer

You Only Demonstrate Once: Category-Level Manipulation from Single Visual Demonstration

https://arxiv.org/pdf/2201.12716.pdf







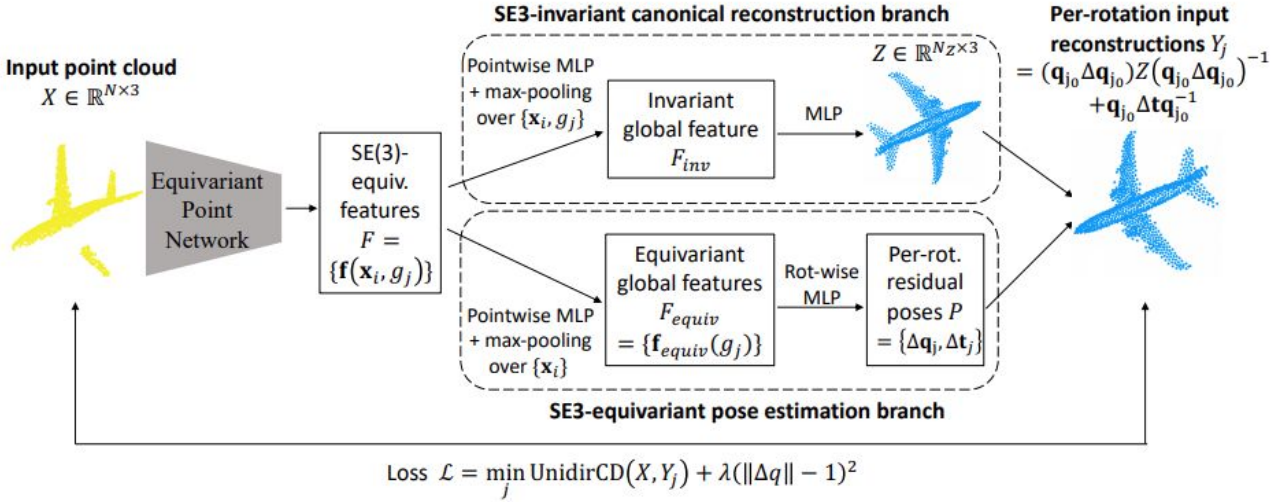
- Full robotic system learning from single demonstration and reproducing with other instances
- Learning of categorical representation from synthetic data
- Full 6-DoF tracking of new objects and control for “the last inch”

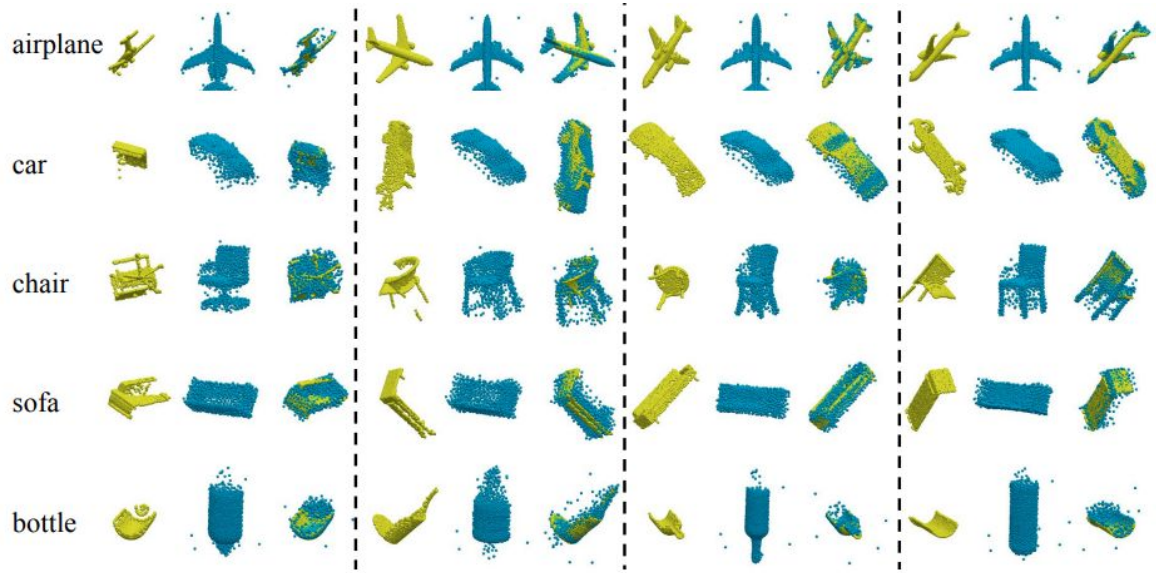
Supervisor: Adrian Röfer

Leveraging SE(3) Equivariance for Self-Supervised Category-Level Object Pose Estimation

https://arxiv.org/pdf/2111.00190.pdf

- Category-Level estimation of 6D poses of unseen objects by disentangling pose and shape
- Equivariant-Feature embedding as disentanglement technique
- Technique is label-free, requiring no CAD models or GT pose information

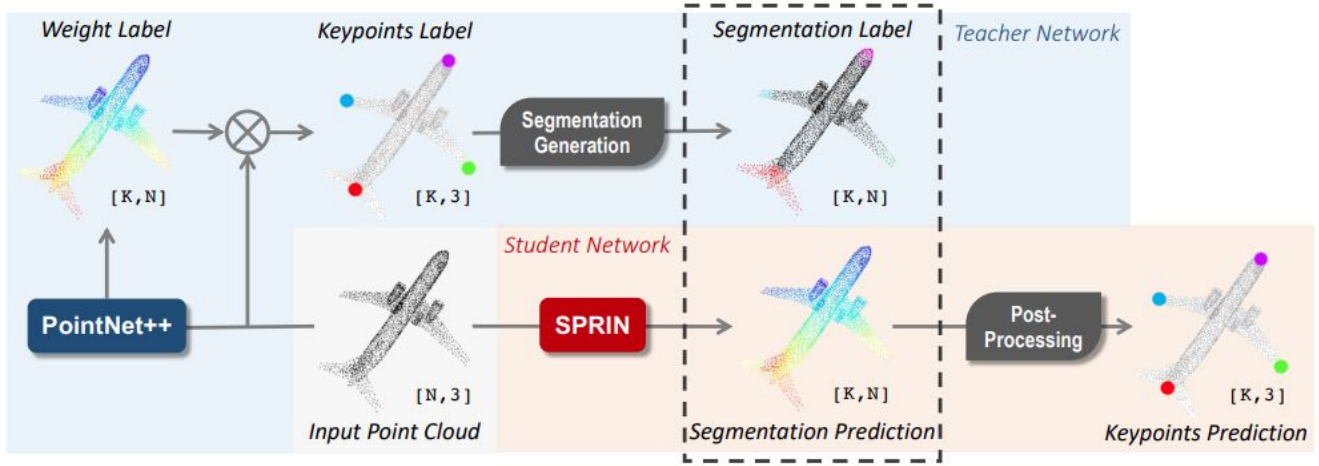




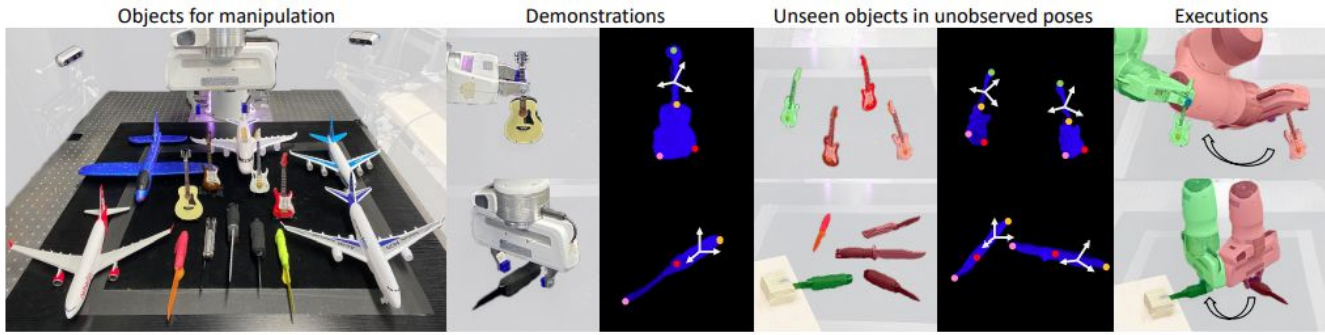
Supervisor: Adrian Röfer

USEEK: Unsupervised SE(3)-Equivariant 3D Keypoints for Generalizable Manipulation

https://arxiv.org/pdf/2209.13864.pdf



- Unsupervised learning of category-level keypoints for object manipulation
- Teacher-Student networks to disentangle keypoint detection and equivariance-prediction
- Demonstrate approach's applicability to robotic manipulation with multiple pick-and-place tasks

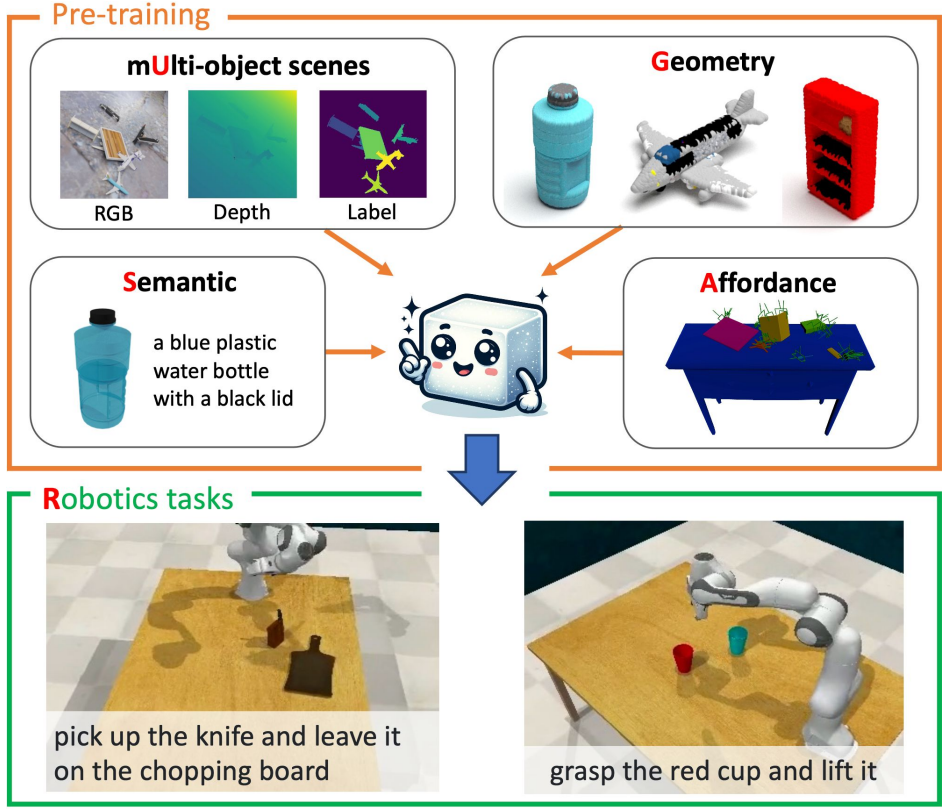


Supervisor: Jan Ole von Hartz

SUGAR: Pre-training 3D Visual Representations for Robotics

https://arxiv.org/pdf/2404.01491.pdf

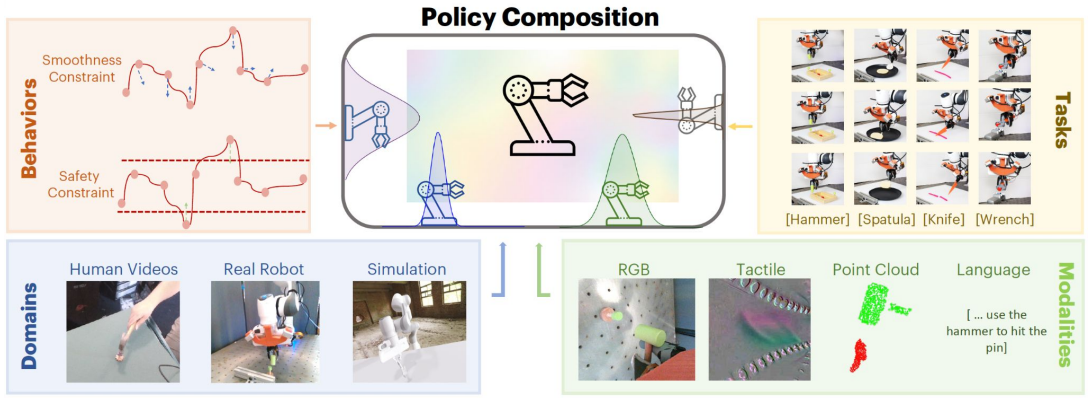
- 3D pre-training framework leveraging
  - cross-modal knowledge distillation for semantic learning,
  - masked point modeling for geometry understanding,
  - grasping pose synthesis for object affordance,
  - 3D instance segmentation for cluttered scenes.
- Self-supervised training in simulation.
- Evaluate on zero-shot 3D object recognition, referring expression grounding, and language-driven manipulation.



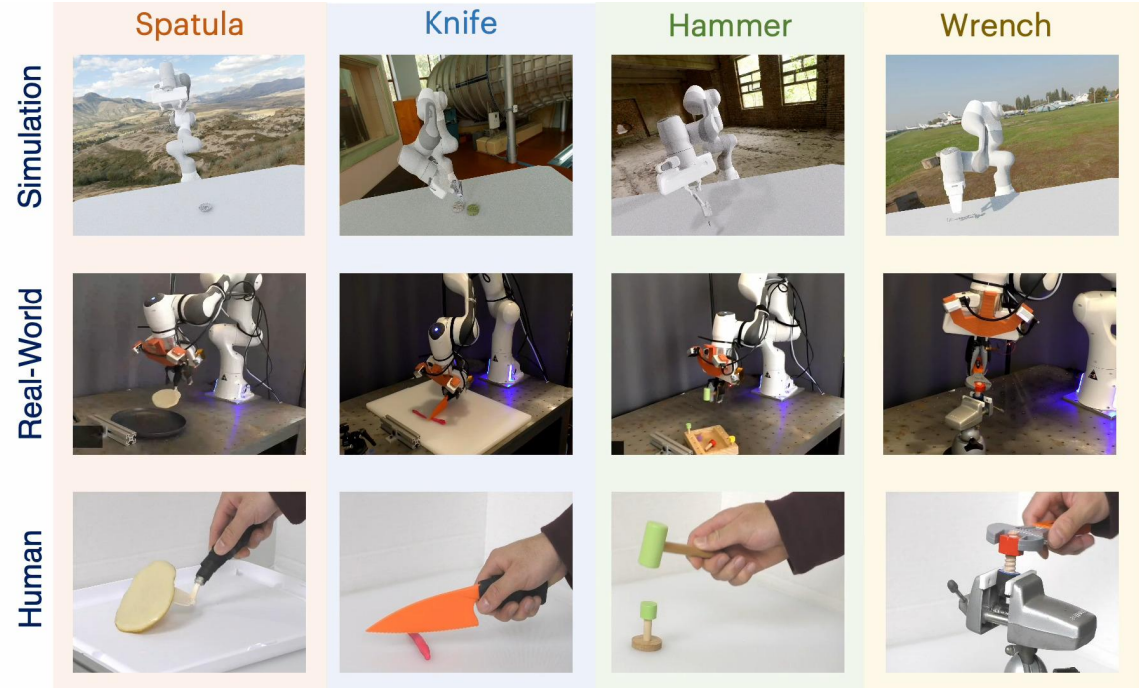
Supervisor: Jan Ole von Hartz

PoCo: Policy Composition from and for Heterogeneous Robot Learning

https://arxiv.org/pdf/2402.02511.pdf



- Use diffusion models to combine
  - multiple modalities (color, depth, tactile, …),
  - domains (simulation, real robots, human).
- Learns generalizable tool-using policies.

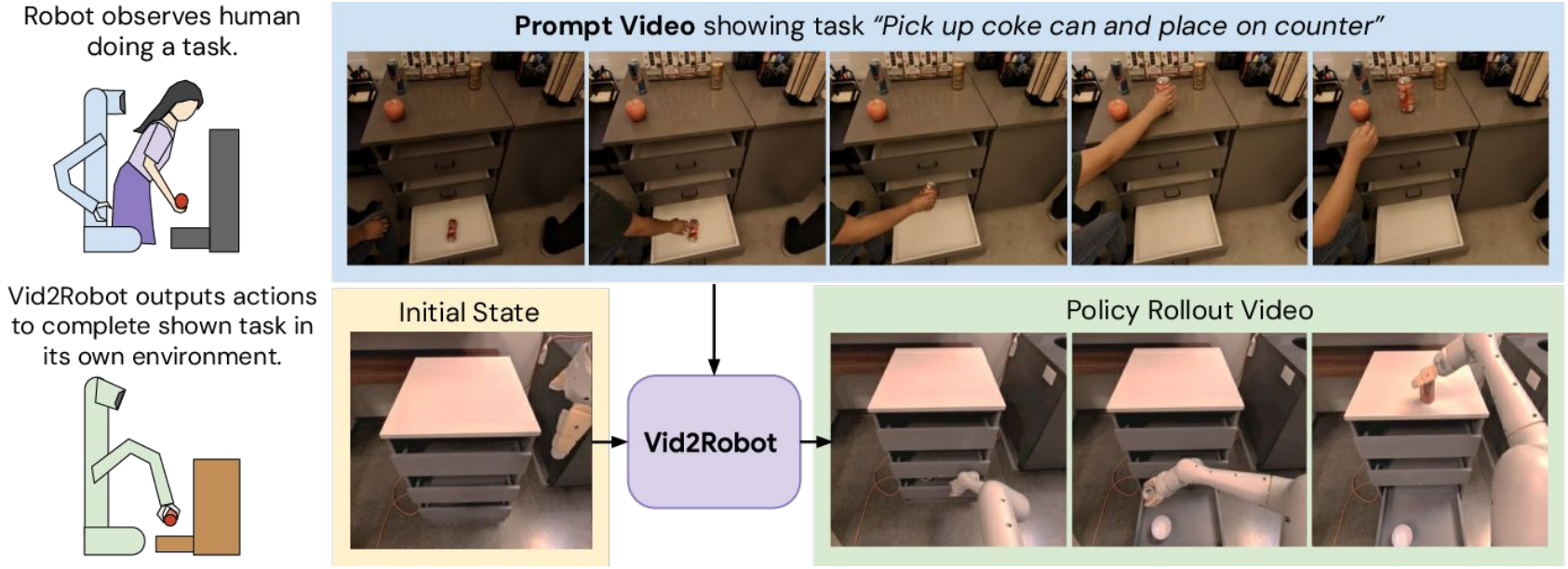


Supervisor: Jan Ole von Hartz

Vid2Robot: End-to-end Video-conditioned Policy Learning with Cross-Attention Transformers

https://arxiv.org/pdf/2403.12943.pdf

- Infer robot task from human prompt video.
- Then predict action sequence to solve task.
- Trained on mixed dataset of robot-robot und human-robot video pairs.

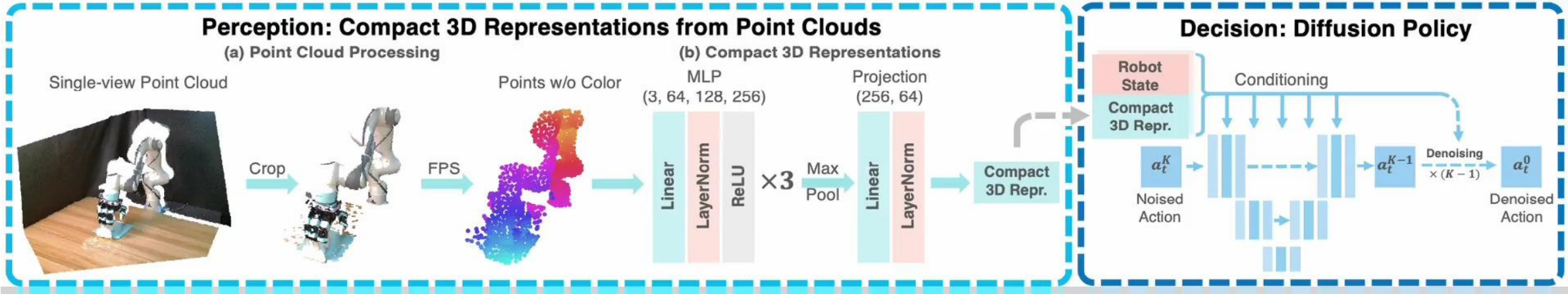


Supervisor: Jan Ole von Hartz

3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations

https://arxiv.org/pdf/2403.03954.pdf

- Point encoder extracting a compact 3D representation from sparse point clouds.
- Used to learn tasks from just 10 demos.

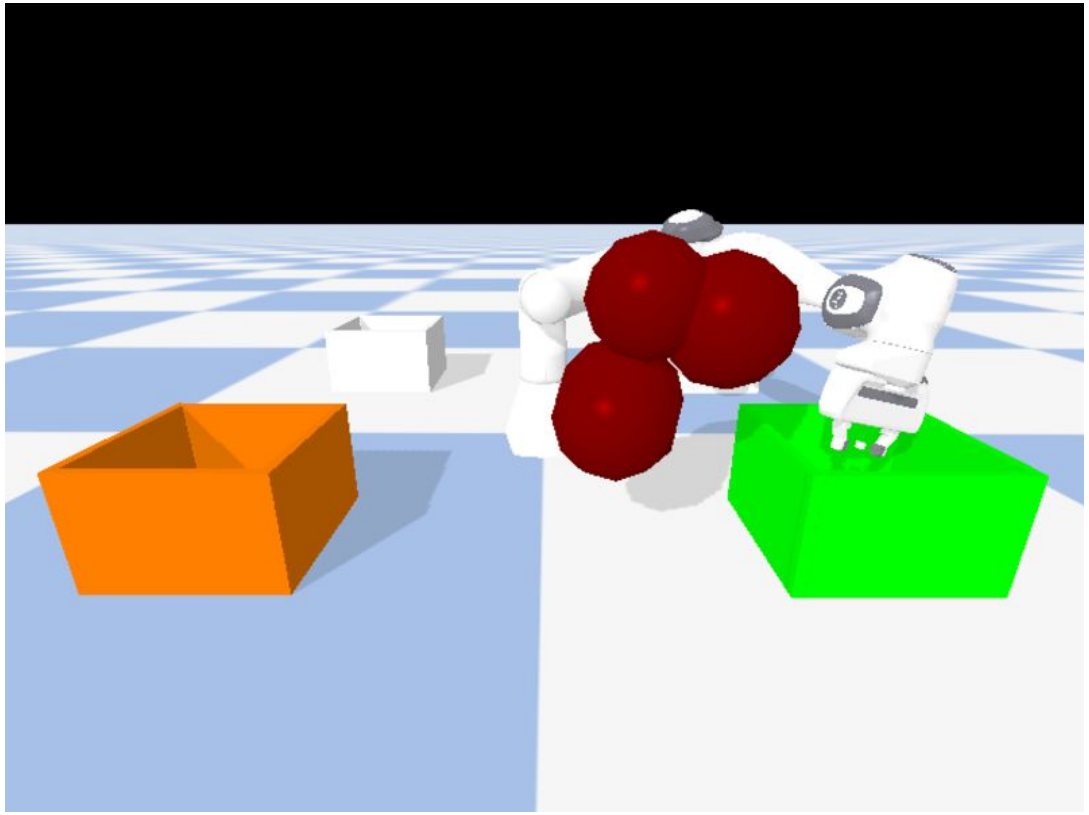


Supervisor: Jan Ole von Hartz

Hierarchical Policy Blending As Optimal Transport

https://arxiv.org/pdf/2212.01938.pdf

- High-level look-ahead planner that blends a set of low-level, reactive Riemannian motion policies via optimal transport.
- Eg. combine goal-reaching (green box) with obstacle avoidance (red blobs).
- Danger: lots of math ahead!



?

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.

Please check our website for information on how to apply:

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

Questions or Comments

José Arce

Robot Learning Lab

arceyd@cs.uni-freiburg.de