universität freiburg

WS 24/25 Seminar
Robot Learning

Markus Käppeler

Robot Learning Lab

18 October 2024

ellis UNIT FREIBURG

robot learning

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.

Inhaltsübersicht.

§ 1. Die Physik der Modelle.
§ 2. Die Statistik der Modellvariablen in der Quantenmechanik.
§ 3. Beispiele für Wahrscheinlichkeitsvoraussagen.
§ 4. Kann man der Theorie ideale Gesamtheiten unterlegen?
§ 5. Sind die Variablen wirklich verwaschen?
§ 6. Der bewußte Wechsel des erkenntnistheoretischen Standpunktes.
§ 7. Die ψ-Funktion als Katalog der Erwartung.
§ 8. Theorie des Messens, erster Teil.
§ 9. Die ψ-Funktion als Beschreibung des Zustandes.
§ 10. Theorie des Messens, zweiter Teil.
§ 11. Die Aufhebung der Verschränkung. Das Ergebnis abhängig vom Willen des Experimentators.
§ 12. Ein Beispiel.
§ 13. Fortsetzung des Beispiels: alle möglichen Messungen sind eindeutig verschränkt.
§ 14. Die Änderung der Verschränkung mit der Zeit. Bedenken gegen die Sonderstellung der Zeit.
§ 15. Naturprinzip oder Rechenkunstgriff?

§ 1. Die Physik der Modelle.

In der zweiten Hälfte des vorigen Jahrhunderts war aus den großen Erfolgen der kinetischen Gastheorie und der mechanischen Theorie der Wärme ein Ideal der exakten Naturbeschreibung hervorgewachsen, das als Krönung jahrhundertelangen Forschens und Erfüllung jahrtausendealter Hoffnung einen Höhepunkt bildet und das klassische heißt. Dieses sind seine Züge.

Von den Naturobjekten, deren beobachtetes Verhalten man erfassen möchte, bildet man, gestützt auf die experimentellen Daten, die man besitzt, aber ohne der intuitiven Imagination zu wehren, eine Vorstellung, die in allen Details genau ausgearbeitet ist, *viel* genauer als irgendwelche Erfahrung in Ansehung ihres begrenzten Umfangs je verbürgen kann. Die Vorstellung in ihrer absoluten Bestimmtheit gleicht einem mathematischen Gebilde oder einer geometrischen Figur, welche aus einer Anzahl von *Bestimmungsstücken* ganz und gar berechnet werden kann; wie z. B. an einem Dreieck eine Seite und die zwei ihr anliegenden Winkel, als Bestimmungsstücke, den dritten Winkel, die anderen zwei Seiten, die drei Höhen, den Radius des eingeschriebenen Kreises usw. mit bestimmen. Von einer geometrischen Figur unterscheidet sich die Vorstellung ihrem Wesen nach bloß durch den wichtigen Umstand, daß sie auch noch in der *Zeit* als vierter Dimension ebenso klar bestimmt ist wie jene in den drei Dimensionen des Raumes. Das heißt es handelt sich (was ja selbstverständlich ist) stets um ein Gebilde, das sich mit der Zeit verändert, das verschiedene *Zustände* annehmen kann; und wenn ein Zustand durch die nötige Zahl von Bestimmungsstücken bekannt gemacht ist, so sind nicht nur alle anderen Stücke in diesem Augenblick mit gegeben (wie oben am Dreieck erläutert), sondern ganz ebenso alle Stücke, der genaue Zustand, zu jeder bestimmten späteren Zeit; ähnlich wie die Beschaffenheit eines Dreiecks an der Basis seine Beschaffenheit an der Spitze bestimmt. Es gehört mit zum inneren Gesetz des Gebildes, sich in bestimmter Weise zu verändern, das heißt, wenn es in einem bestimmten Anfangszustand sich selbst überlassen wird, eine bestimmte Folge von Zuständen kontinuierlich zu durchlaufen, deren jeden es zu ganz bestimmter Zeit erreicht. Das ist seine Natur, das ist die Hypothese, die man, wie ich oben sagte, auf Grund intuitiver Imagination setzt.

Natürlich ist man nicht so einfältig zu denken, daß solchermaßen zu erraten sei, wie es auf der Welt wirklich zugeht. Um anzudeuten, daß man das nicht denkt, nennt man den präzisen Denkbehelf, den man sich geschaffen hat, gern ein *Bild* oder ein *Modell*. Mit seiner nachsichtslosen Klarheit, die ohne Willkür nicht herbeizuführen ist, hat man es lediglich darauf abgesehen, daß eine ganz bestimmte Hypothese in ihren Folgen geprüft werden kann, ohne neuer Willkür Raum zu geben während der langwierigen Rechnungen, durch die man Folgerungen ableitet. Da hat man gebundene Marschroute und errechnet eigentlich nur, was ein kluger Hans aus den Daten direkt herauslesen würde! Man weiß dann wenigstens, wo die Willkür steckt und wo man zu bessern hat, wenn's mit der Erfahrung nicht stimmt: in der Ausgangshypothese, im Modell. Dazu muß man stets bereit sein. Wenn bei vielen verschiedenartigen Experimenten das Naturobjekt sich wirklich so benimmt wie das Modell, so freut man sich und denkt, daß unser Bild in den wesentlichen Zügen der Wirklichkeit gemäß ist. Stimmt es bei einem neuartigen Experiment oder bei Verfeinerung der Meßtechnik nicht mehr, so ist nicht gesagt, daß man sich *nicht* freut. Denn im Grunde ist das die Art, wie allmählich eine immer bessere Anpassung des Bildes, das heißt unserer Gedanken, an die Tatsachen gelingen kann.

Die klassische Methode des präzisen Modells hat den Hauptzweck, die unvermeidliche Willkür in den Annahmen sauber isoliert zu halten, ich möchte fast sagen wie der Körper das Keimplasma, für den historischen Anpassungsprozeß an die fortschreitende Erfahrung. Vielleicht liegt der

Nw. 1935 52

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 **21.10.2024**

Please check the course website for more information:

https://rl.uni-freiburg.de/teaching/ws24/robotlearning/

Important Dates

Event	Date	Time
Lecture 1: Introduction *	18.10.2024	13:00
HISinOne registration + Paper Selection	21.10.2024	
Place allocation	24.10.2024	
Paper assignment	28.10.2024	
Supervisor Meeting	12.2024	
Lecture 2: *How to do a good presentation* *	10.01.2025	13:00
Lecture 3: Block Seminar Presentations *	07.02.2025	9:00 - 17:00
Paper Summary submission	21.02.2025	< 23:59

* **Mandatory in-person attendance**

Evaluation Criteria

Evaluation	Due Date
Seminar Presentation	07.02.2025
Paper Summary	21.02.2025

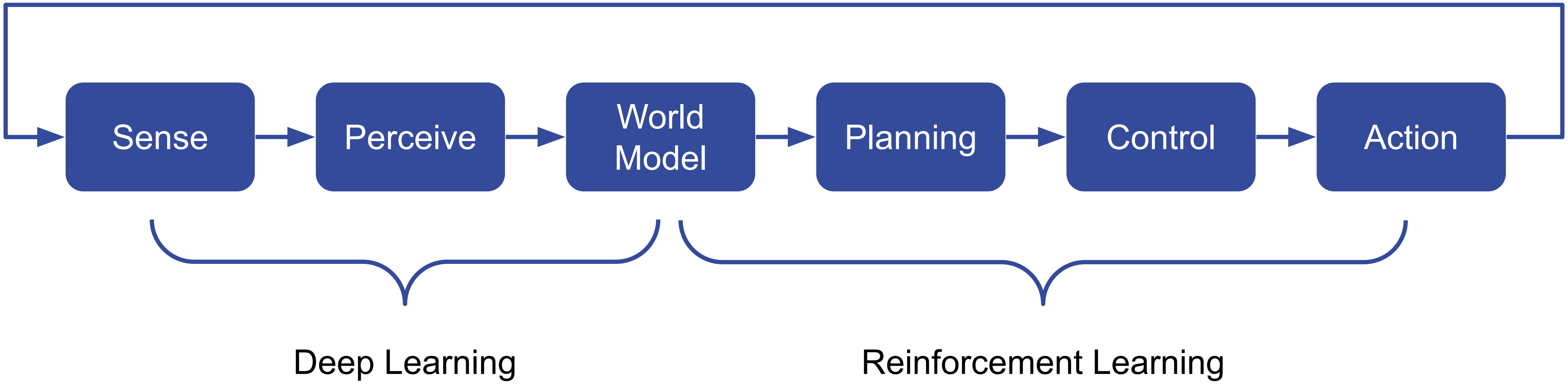
- **Presentation:** at most **20 min**.
- **Summary:** at most **7 pages** excluding bibliography and figures.
- **Final grade:**
  - Presentation (slides & delivery) + Summary + **Seminar Participation**.

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?

Robot Learning Lab

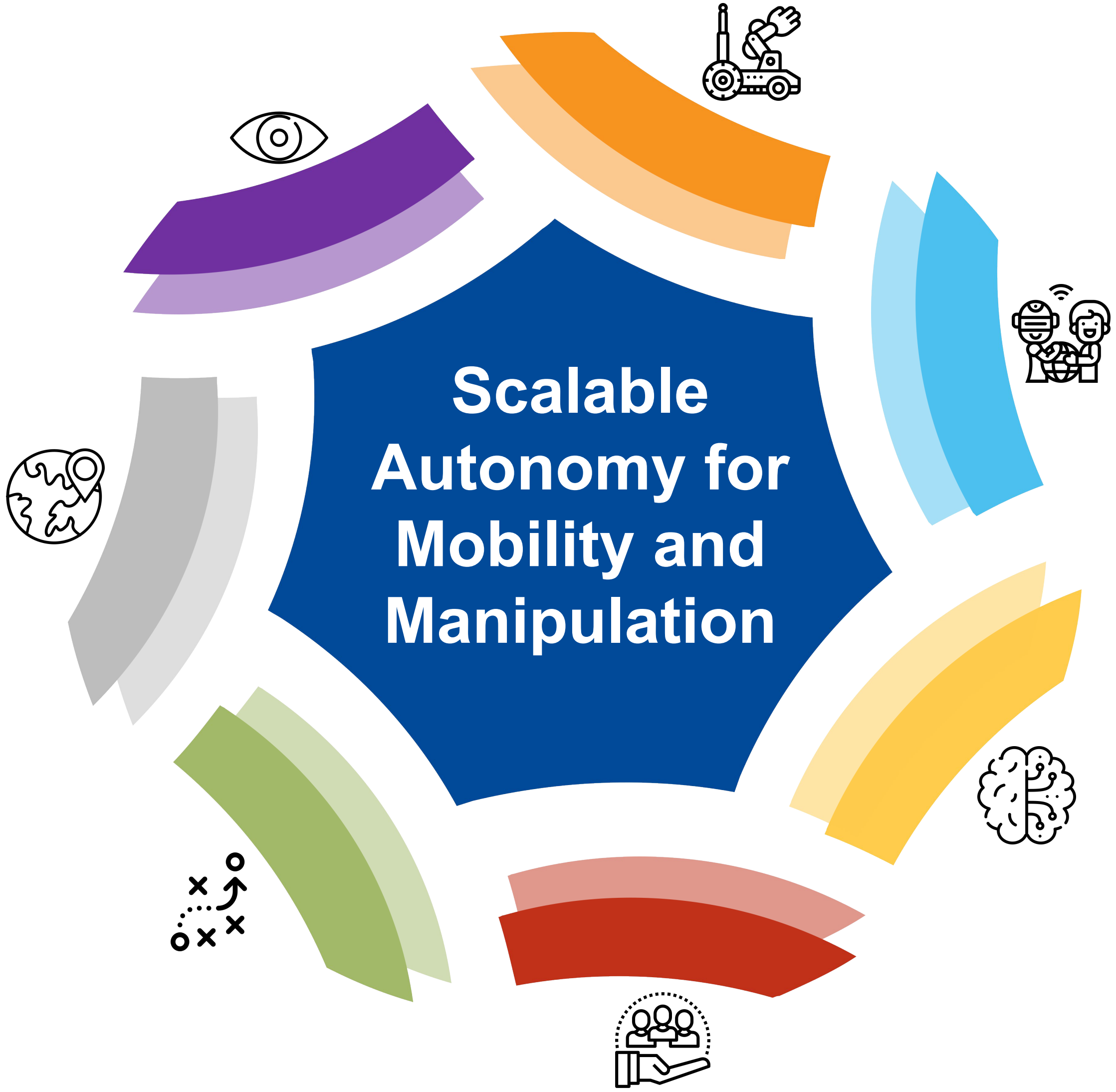
Research Areas

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

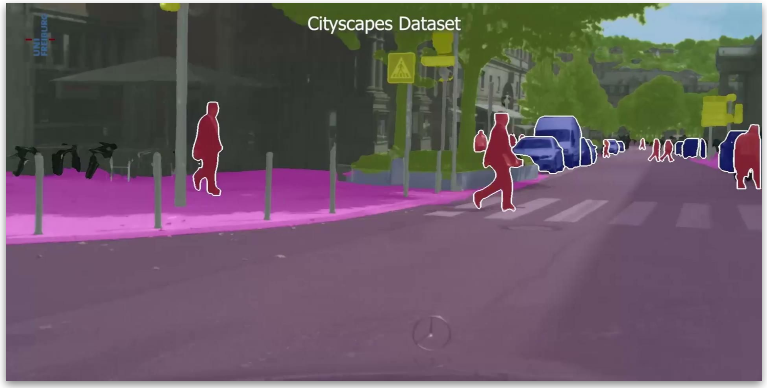


Mobile Manipulation
- Whole-Body Motion
- Long-Horizon Reasoning
- Planning for Sensing

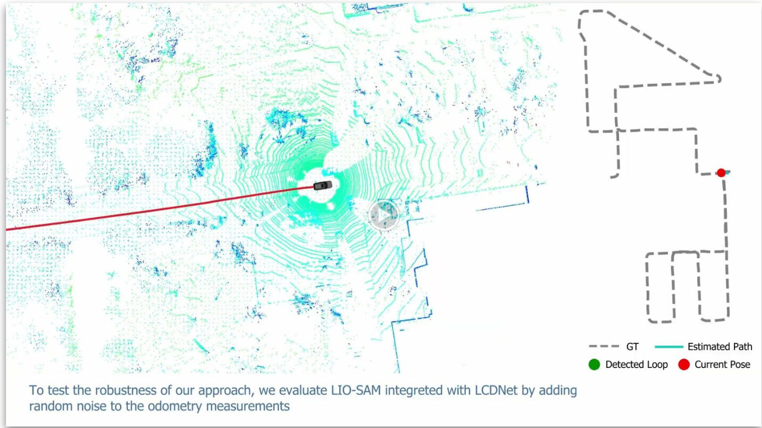
Human-Robot Interaction
- Socially-Compliant Behavior
- Human-Robot Collaboration
- Behavior Adaptation & Safety

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

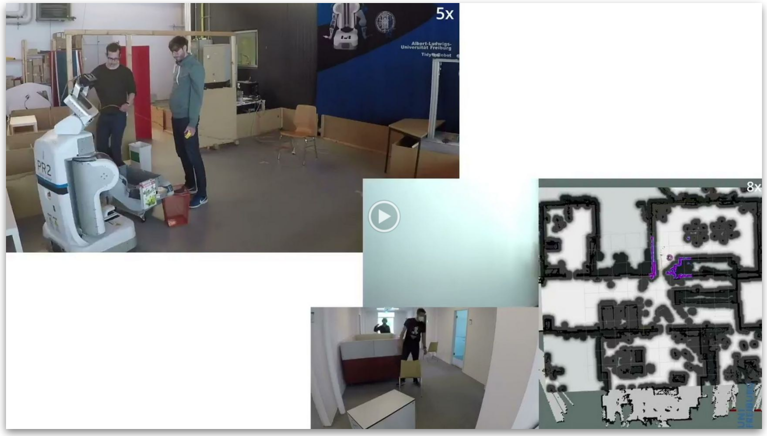
Many Seminal Works



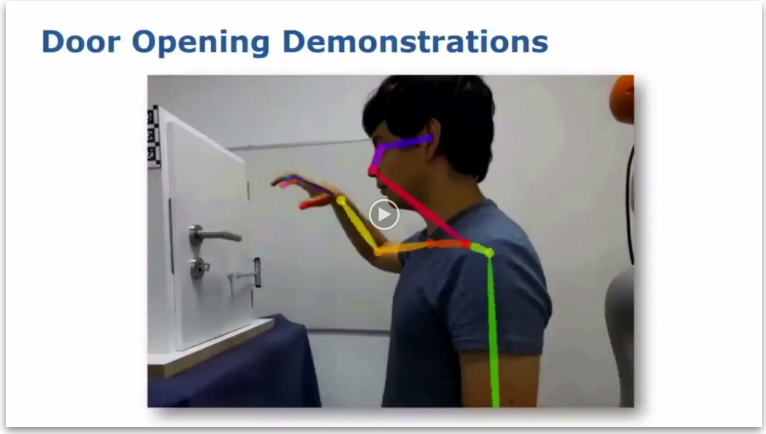
Scene Understanding



Simultaneous Localization and Mapping

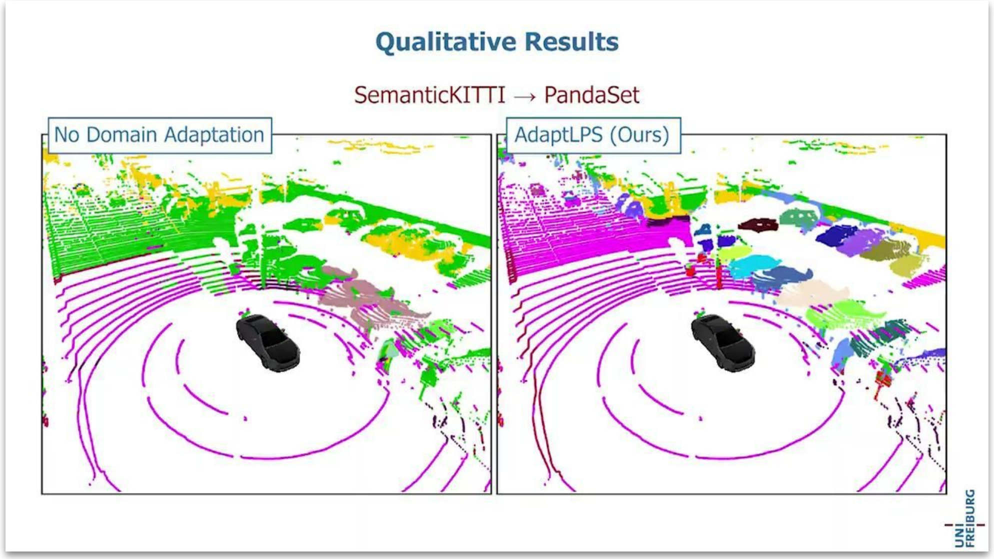


Motion Planning

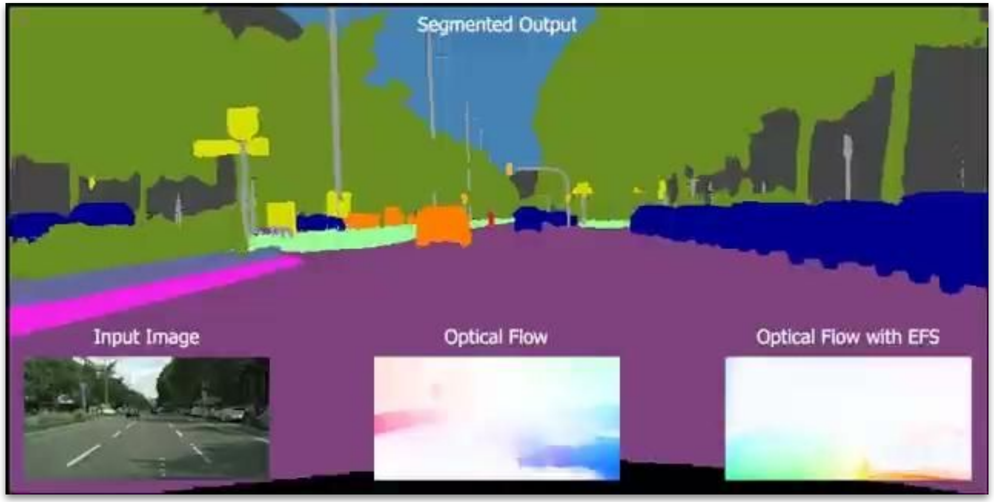


Learning from Demonstrations

Robotic Perception — Mobility



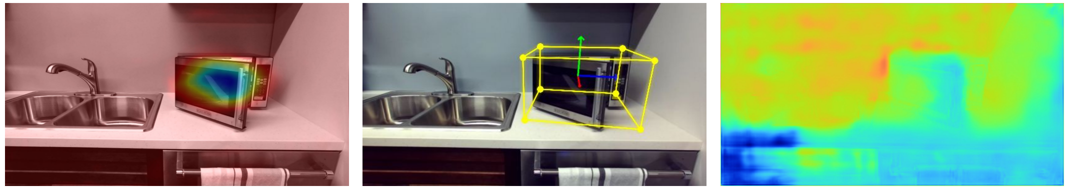
Unsupervised LiDAR Domain Adaptation
Besic, Gosala, Cattaneo, Valada
RA-L '22



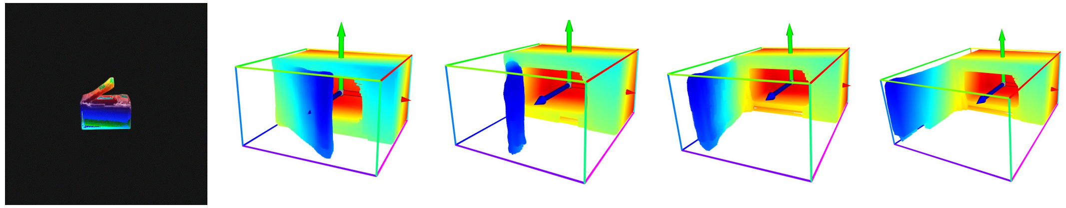
Semantic Motion Segmentation
Vertens, Valada, Burgard
ICRA '17

Robotic Perception — Manipulation

Single-Shot Reconstruction



Heatmap | Poses and Bounding Box | Auxiliary Depth Prediction

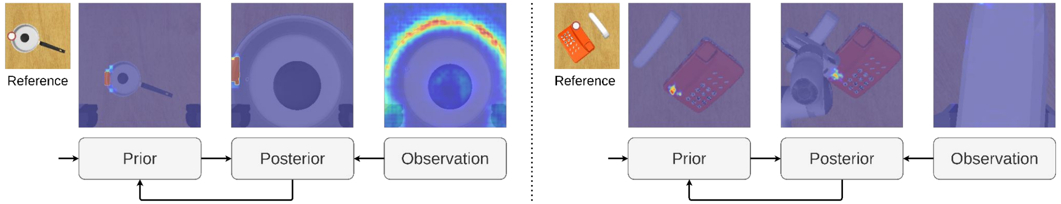


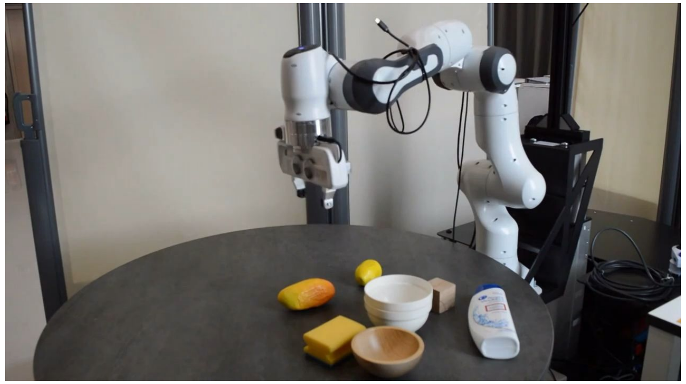
Category and Joint Agnostic Reconstruction of ARTiculated Objects

Heppert, et al

CVPR ‘23

Learning scale-invariant compact representations for mobile manipulation



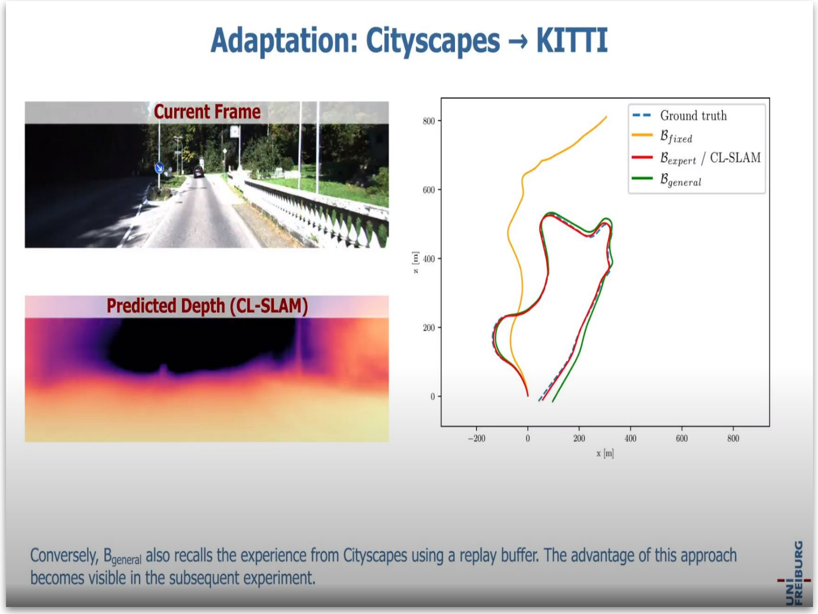


Bayesian Scene Keypoints for Deep Policy Learning in Robotic Manipulation

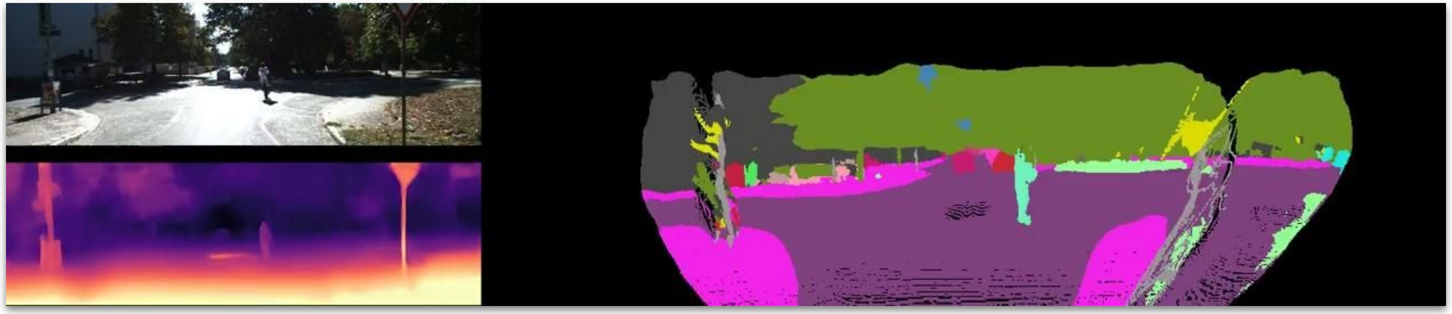
von Hartz, et al

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

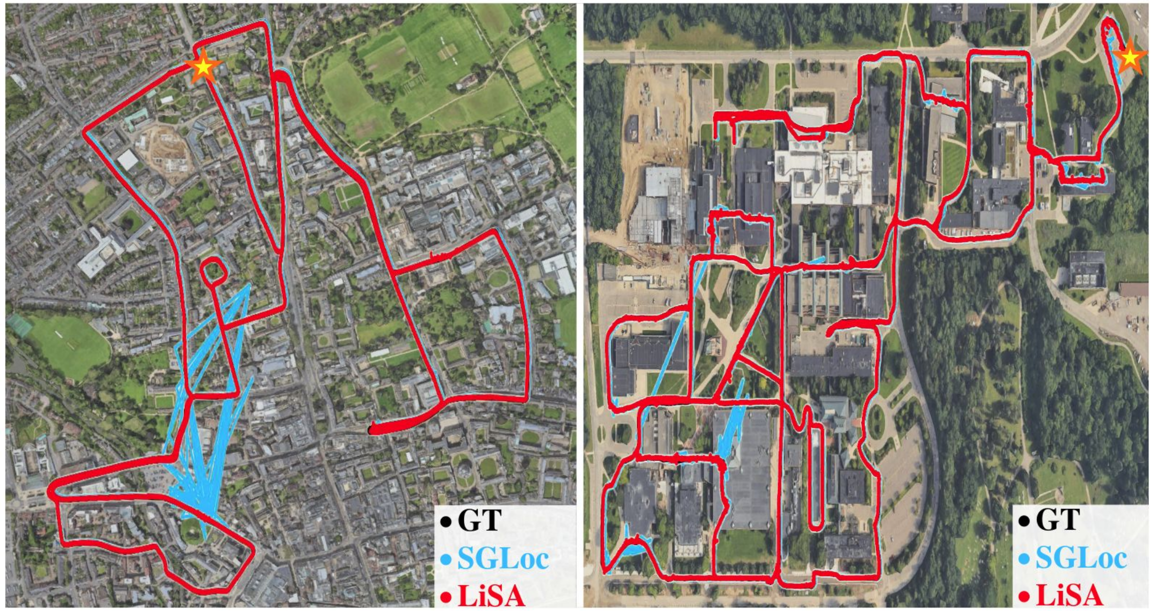
Seminar Papers

Supervisor: Niclas Vödisch

LiSA: LiDAR Localization with Semantic Awareness

https://openaccess.thecvf.com/content/CVPR2024/papers/Yang_LiSA_LiDAR_Localization_with_Semantic_Awareness_CVPR_2024_paper.pdf

- Pose estimation of a LiDAR point cloud in a global scene map.
- Learning-based scene coordinate regression to find point correspondences between LiDAR scan and 3D map.
- LiSA exploits semantic point information to robustly handle dynamic objects and textureless regions.

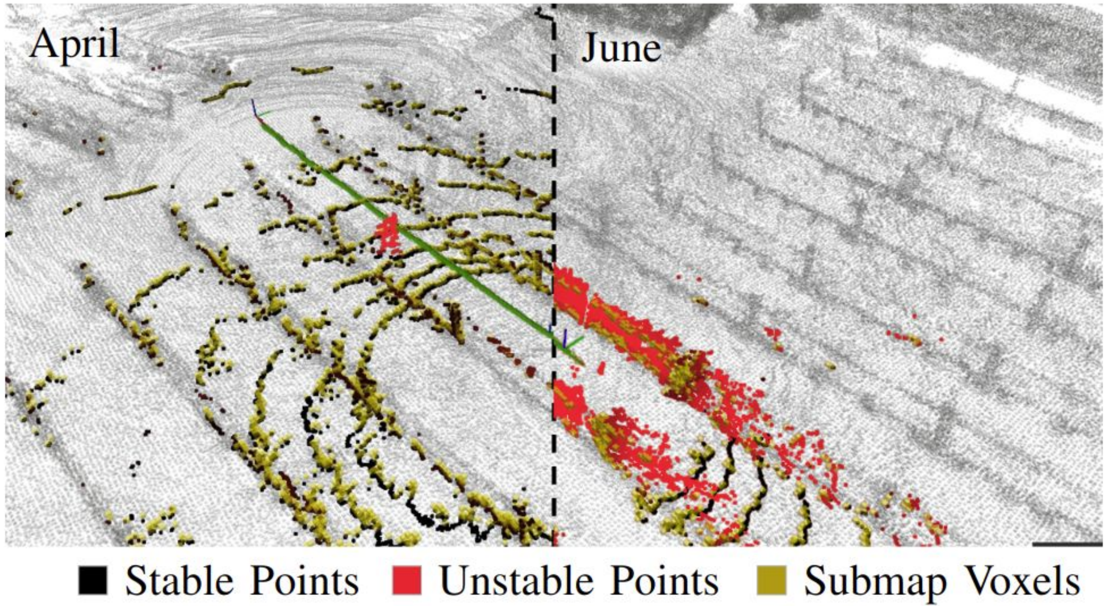


Supervisor: Niclas Vödisch

Generalizable Stable Points Segmentation for 3D LiDAR Scan-to-Map Long-Term Localization

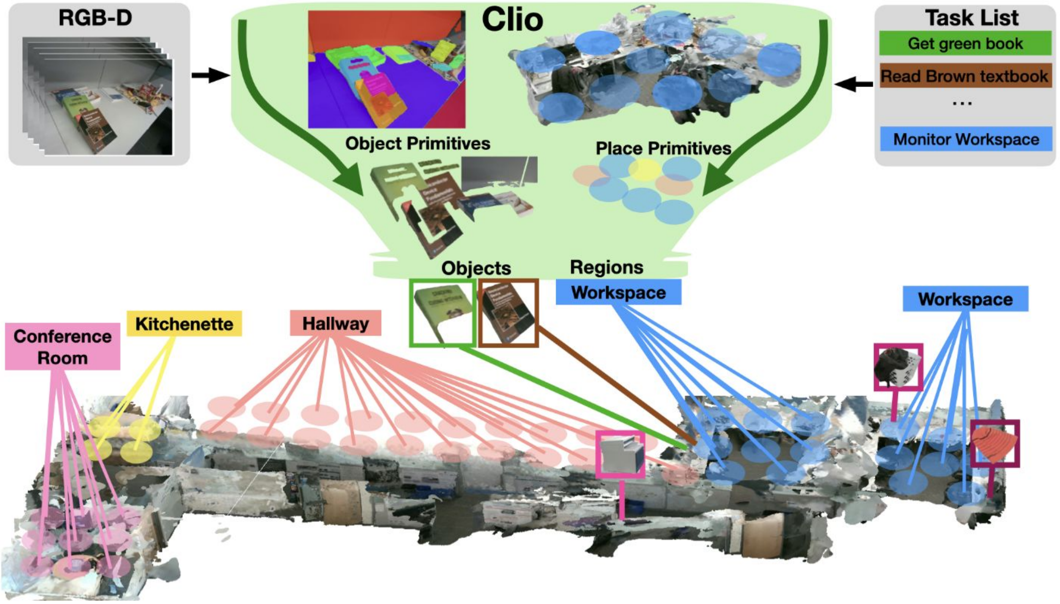
https://www.ipb.uni-bonn.de/wp-content/papercite-data/pdf/hroob2024ral.pdf

- 3D scenes change over time, e.g., seasonal changes or human-made structural differences.
- Learning-based prediction whether a 3D point is temporally static or unstable.
- Reject unstable points during scan-to-map point cloud registration.



Clio: Real-time Task-Driven Open-Set 3D Scene Graphs

https://arxiv.org/pdf/2404.13696



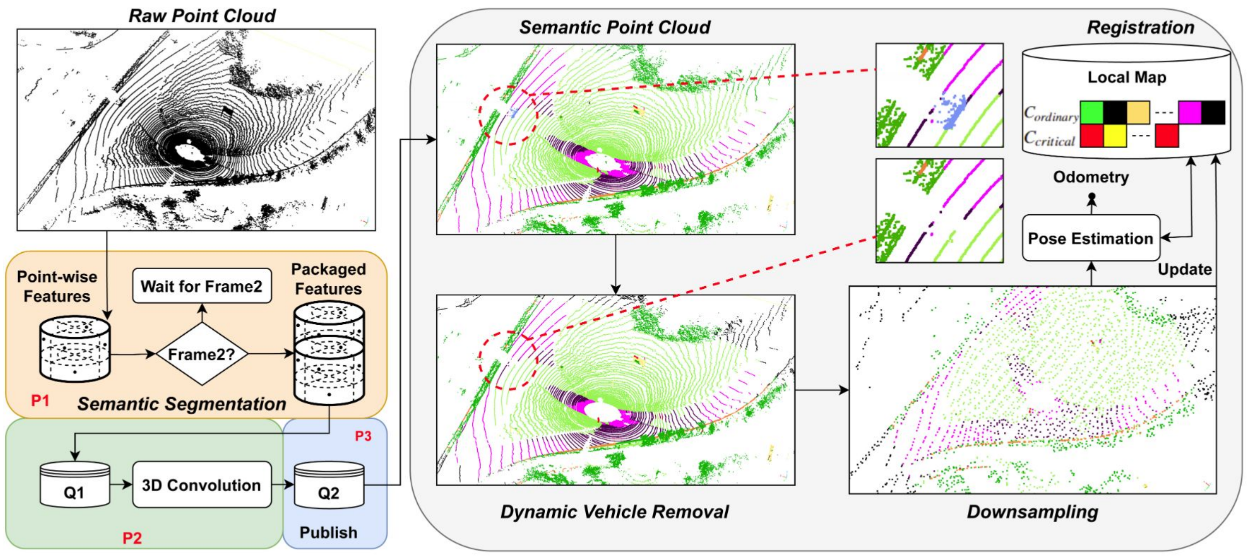
- Scene graphs are an efficient representation to store hierarchical dependencies in complex environments, e.g., office buildings.
- Clio relies on vision-language foundation models to gather open-set semantic understanding
- Experiments include real-time scene graph generation with object pick-up tasks.

Supervisor: Niclas Vödisch

SAGE-ICP: Semantic Information-Assisted ICP

https://arxiv.org/pdf/2310.07237

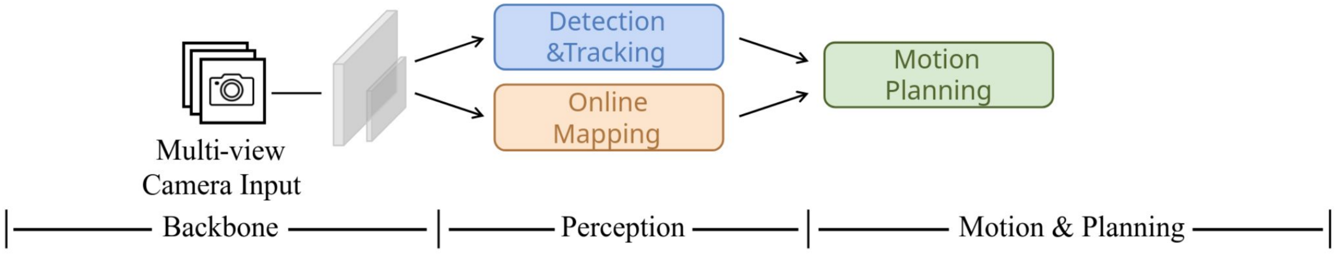
- The iterative closest point (ICP) algorithm regresses the transformation between two point clouds.
- SAGE-ICP extends the point-to-point approach with semantic information to avoid wrong correspondences.
- Semantics are further used for adaptive voxelization, i.e., keep details where required (e.g., poles) but reject where unnecessary (e.g., road surface).



SparseDrive: End-to-End Autonomous Driving via Sparse Scene Representation

https://arxiv.org/abs/2405.19620

- Fully differentiable autonomous driving network that leverages perception outputs to optimize towards the ultimate goal of planning.
- Unified sparse perception networks to avoid computationally expensive bird's-eye view (BEV) features.
- Joint prediction and planning with a planning selection strategy that takes possible collisions with other agents into account.

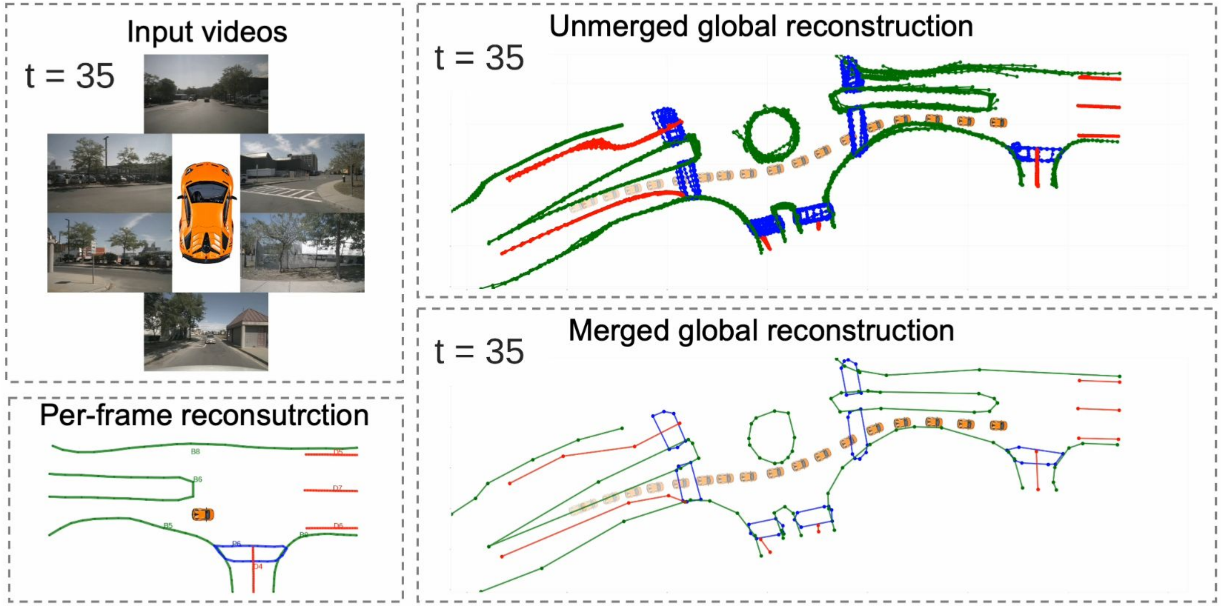


Supervisor: Markus Käppeler

MapTracker: Tracking with Strided Memory Fusion for Consistent Vector HD Mapping

https://map-tracker.github.io/

- Consistent vector HD mapping system to enhance the safety and stability of self-driving cars.
- Formulate vectorized mapping as a tracking task and use memory latents from previous frames to get a consistent mapping over time.
- Rasterized bird's-eye view (BEV) features and vector latents (queries) of individual map elements as latent representations for the sequence memory.

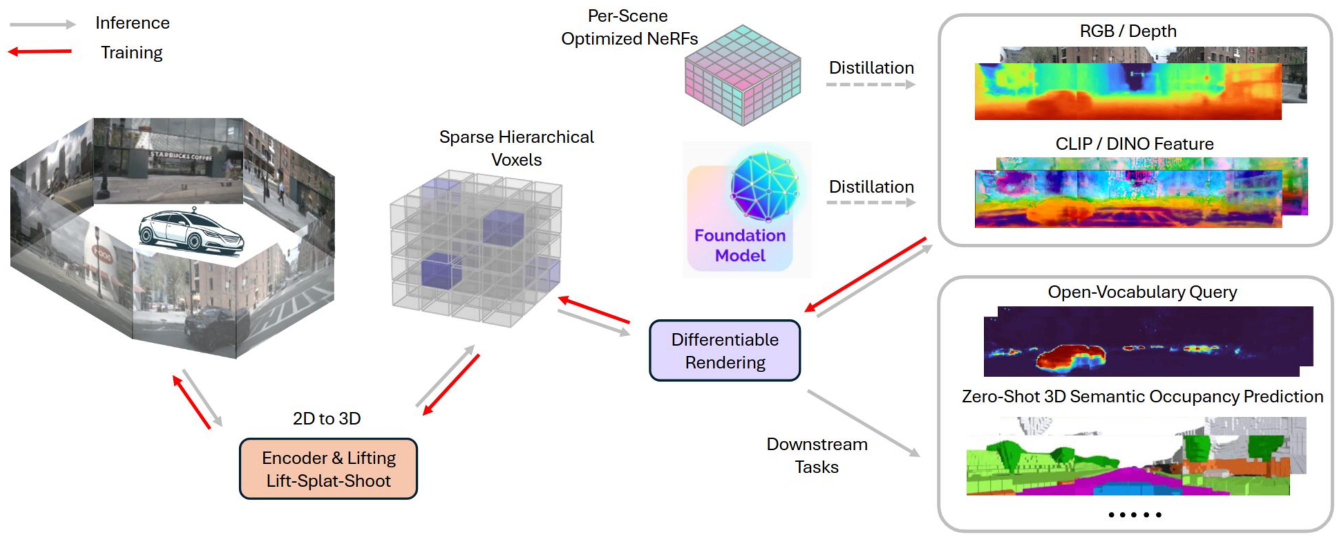


Supervisor: Markus Käppeler

DistillNeRF: Perceiving 3D Scenes from Single-Glance Images by Distilling Neural Fields and Foundation Model Features

https://distillnerf.github.io/

- Understanding 3D environments is crucial for autonomous driving.
- Network predicts a generalizable 3D neural scene representation from 2D camera images.
- Self-supervised training via differentiable rendering to reconstruct RGB+depth from a per-scene optimized NeRF and feature maps from a vision foundation model.

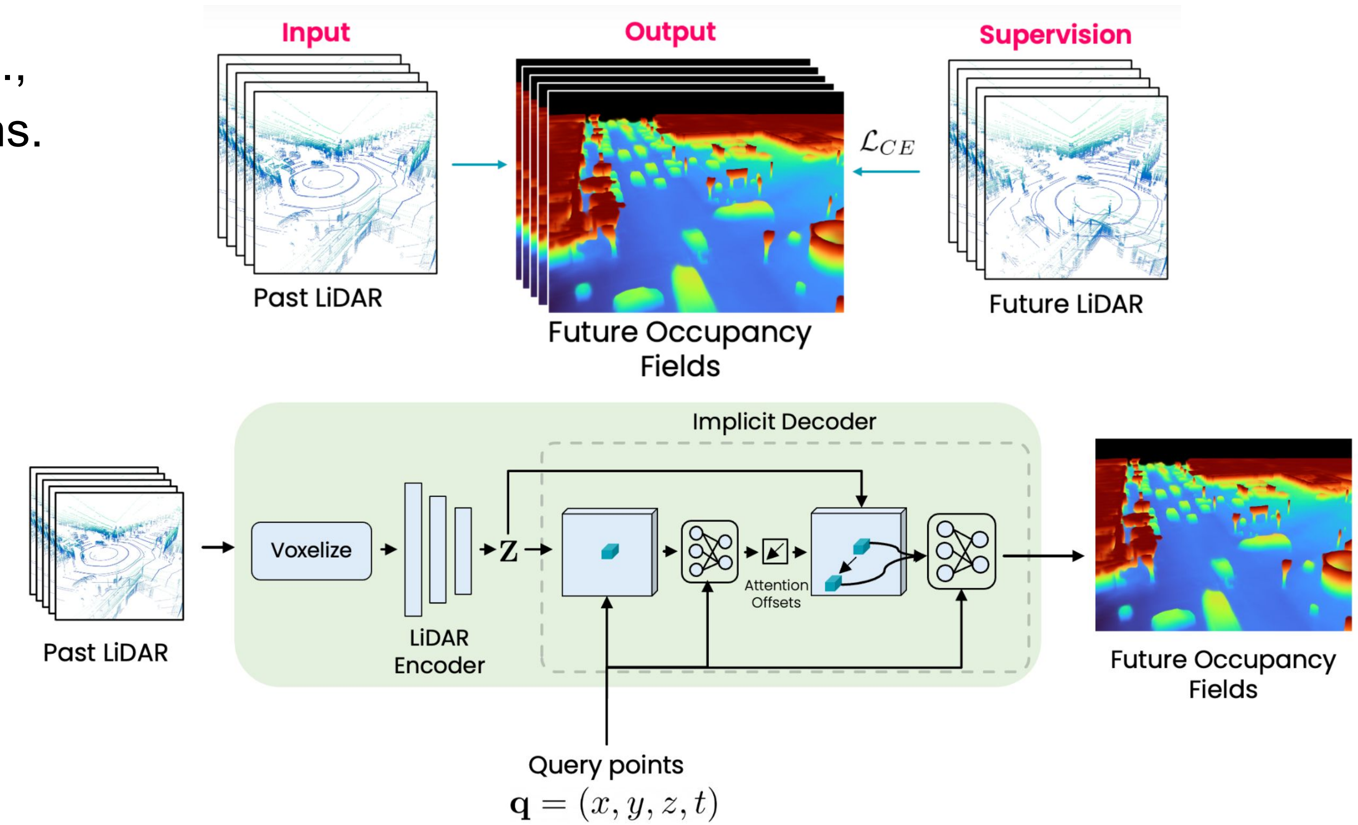


Supervisor: Markus Käppeler

UnO: Unsupervised Occupancy Fields for Perception and Forecasting

https://waabi.ai/uno/

- Supervised approaches leverage annotated object labels to learn a model of the world, e.g., with object detections and trajectory predictions.
- However, annotations are expensive and typically limited to a set of predefined categories.
- UnO learns to perceive and forecast a continuous 4D (spatio-temporal) occupancy field with self-supervision from LiDAR data.

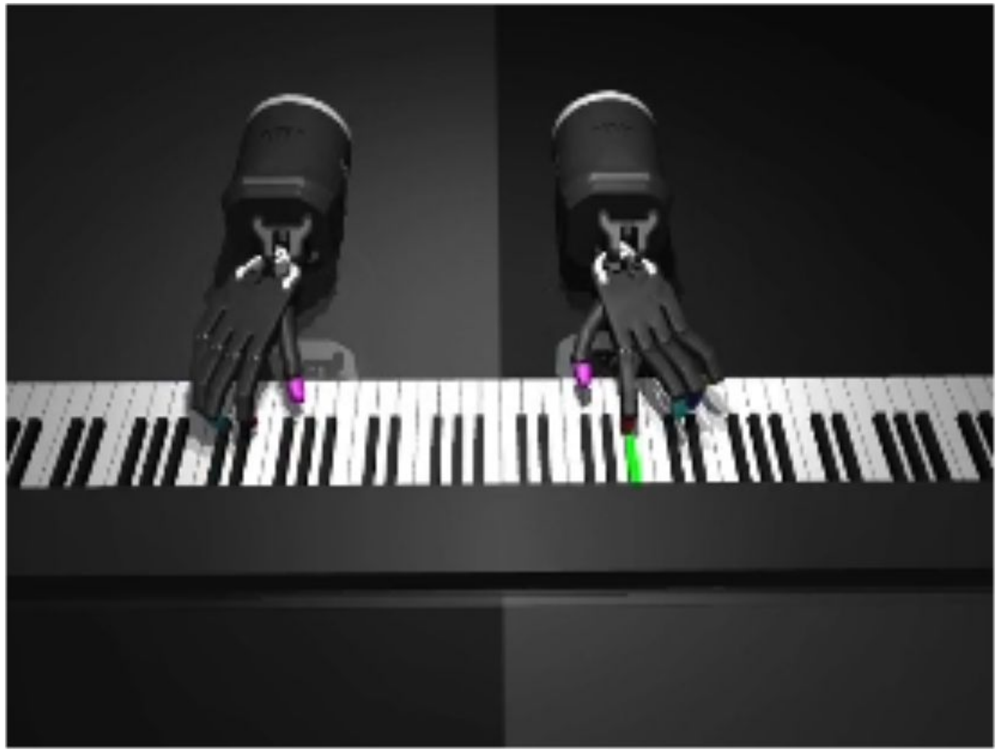


Supervisor: Eugenio Chisari

PianoMime: Learning a Generalist, Dexterous Piano Player from Internet Demonstrations

https://pianomime.github.io/

- The internet is a source of large-scale demonstrations for training robot agents. Youtube is full of videos of professional pianists playing a wide range of songs.
- This work leverages youtube videos as demonstrations to train a generalist piano-playing agent capable of playing any arbitrary song.



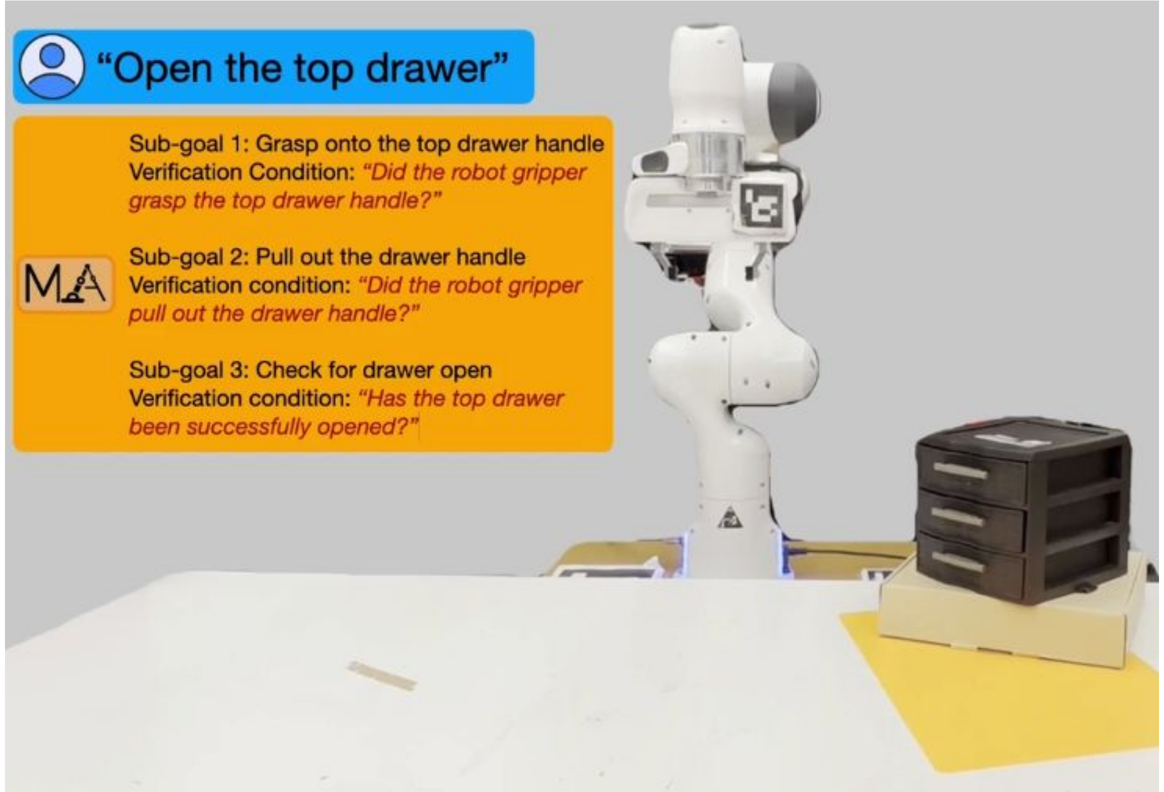
Supervisor: Eugenio Chisari

Manipulate-Anything: Automating Real-World Robots using Vision-Language Models

https://robot-ma.github.io/

- The quality, quantity, and diversity of robot demonstration data represent a serious bottleneck to the progress of robotics research

- Manipulate-Anything is a scalable automated generation method for real-world robotic manipulation. It can operate in real-world environments without any privileged state information or hand-designed skills.

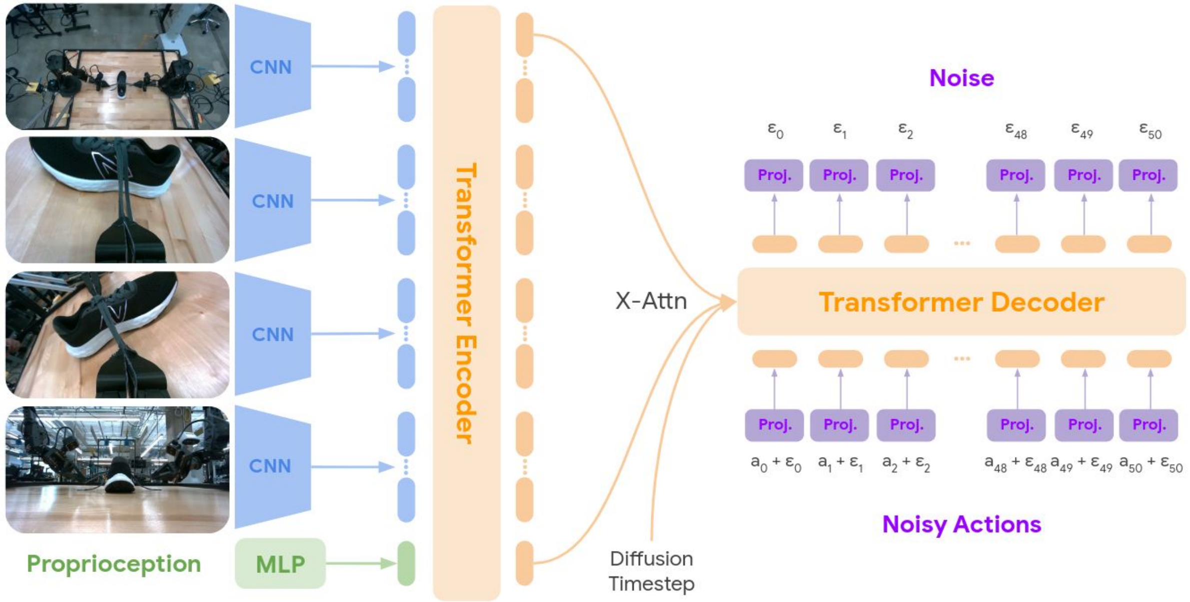


Supervisor: Eugenio Chisari

ALOHA Unleashed: A Simple Recipe for Robot Dexterity

https://aloha-unleashed.github.io/

- This work addresses the question of how far can we push imitation learning for challenging dexterous manipulation tasks.
- The authors demonstrate how to solve challenging bimanual manipulation tasks involving deformable objects by combining diffusion policies with a scalable transformer architecture.

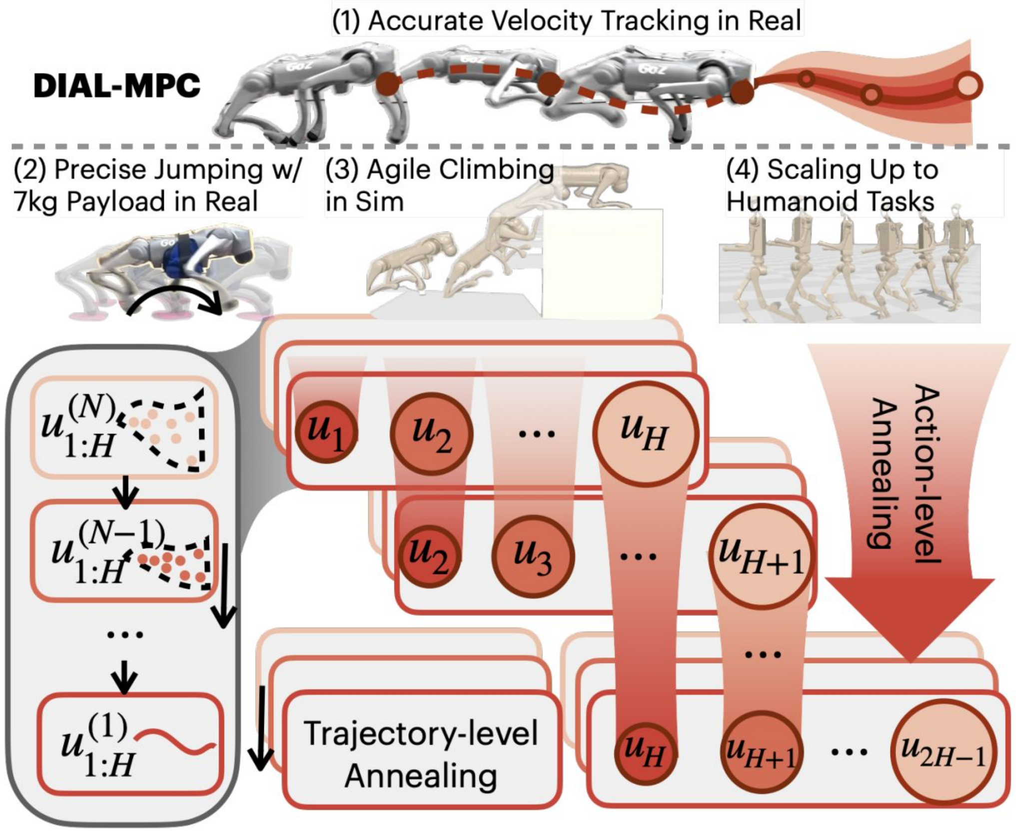


Supervisor: Eugenio Chisari

Full-Order Sampling-Based MPC for Torque-Level Locomotion Control via Diffusion-Style Annealing

https://lecar-lab.github.io/dial-mpc/

- Due to high dimensionality and non-convexity, real-time optimal control using full-order dynamics models for legged robots is challenging.
- This work introduces DIAL-MPC, a sampling-based MPC framework with a novel diffusion-style annealing process.
- DIAL-MPC is the first training-free method that optimizes over full-order quadruped dynamics in real-time.

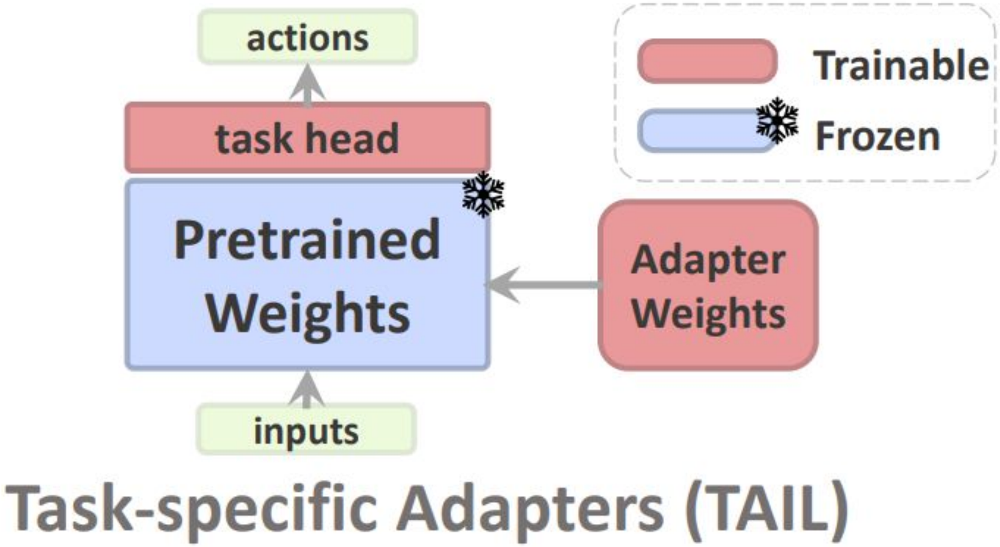


Supervisor: Nick Heppert

TAIL: Task-specific Adapters for Imitation Learning with Large Pretrained Models

https://arxiv.org/abs/2310.05905

- Pre-training of robotic control policies that allows continual adaptation for new task demonstration (imitation learning)
- Exploring different fine-tuning techniques
- Using Low-Rank Adaption (LoRA) only a few parameters needs to be trained (1%)



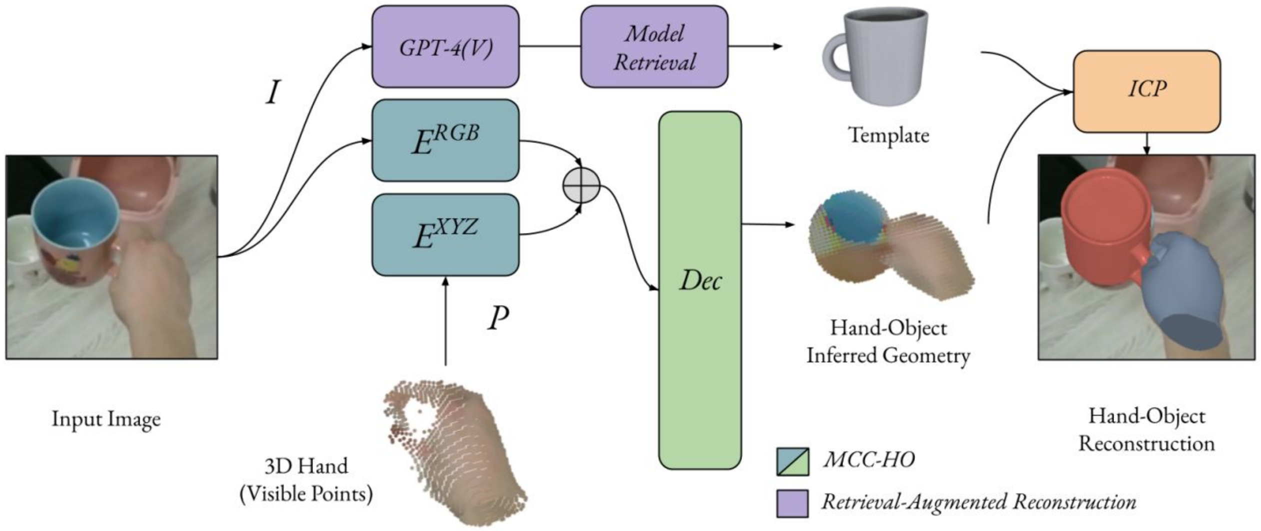
Supervisor: Nick Heppert

Reconstructing Hand-Held Objects in 3D

https://arxiv.org/pdf/2404.06507

- Reconstructing objects held in hands is a challenging task since objects are small and occluded but the hand gives a prior on location and scale of the object

- The proposed method reconstructs hand and objects jointly followed by a retrieval-augmented reconstruction.

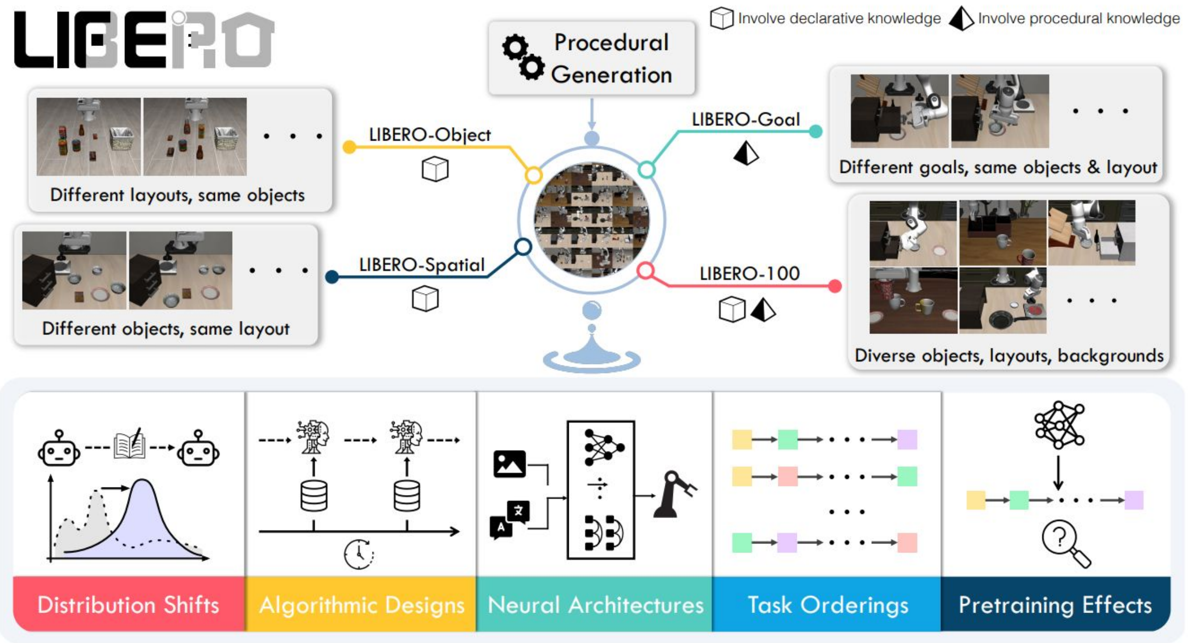


Supervisor: Nick Heppert

LIBERO: Benchmarking Knowledge Transfer for Lifelong Robot Learning

https://libero-project.github.io/

- A novel benchmark for lifelong learning for robot manipulation using procedural generation
- Overview of different lifelong learning techniques and interesting results on the new benchmark.

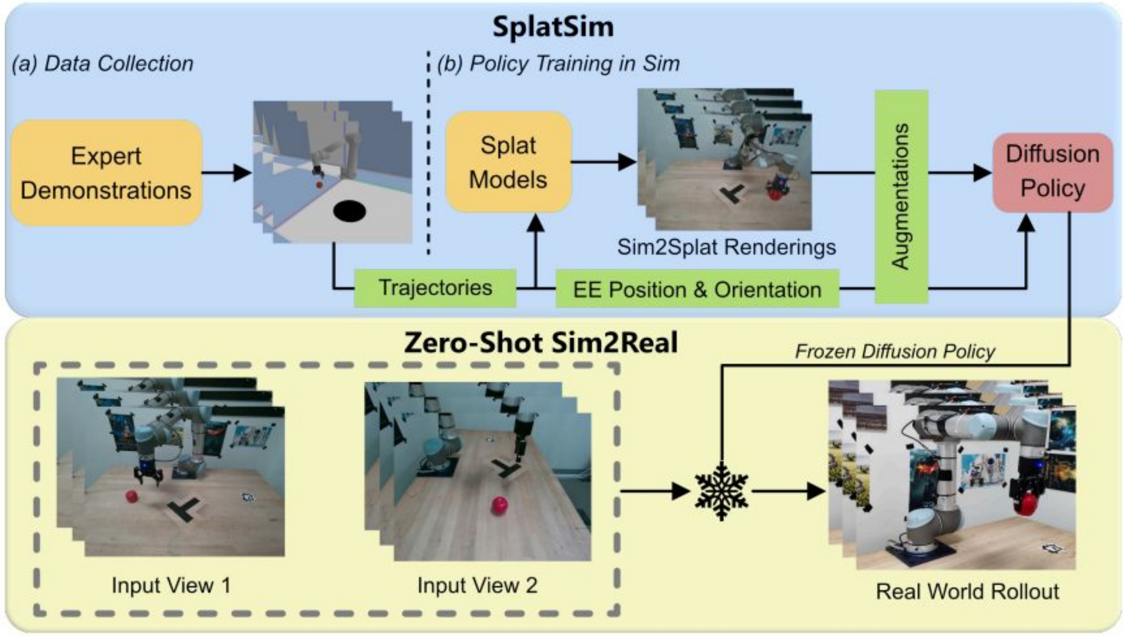


Supervisor: Nick Heppert

SplatSim: Zero-Shot Sim2Real Transfer of RGB Manipulation Policies Using Gaussian Splatting

https://splatsim.github.io/

- Sim2Real for RGB-based policies are prone to overfitting.
- SplatSim replaces meshes with object-centric Gaussian Splats, allowing fast and realistic data synthesis resulting in almost the same success rate when trained on real data.



?

Questions

Open Positions

- 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

Markus Käppeler

Robot Learning Lab

kaeppelm@cs.uni-freiburg.de