universität freiburg

SS26 Seminar
Learning with Limited Supervision

Prof. Dr. Abhinav Valada
Dr. Simon Bultmann, Dr. Iman Nematollahi, Adrian Röfer, Yao Lu

Robot Learning Lab

23 April 2026

ellis unit FREIBURG

Agenda

I. Organization: Enrollment, important dates, tasks, and evaluation.

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

III. Topics: Seminar Topics and Papers.

?. Questions.

I.

Organization

Enrollment, important dates, tasks, and evaluation criteria

Enrollment Procedure

Register for the seminar in HISinOne.

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

Fill in our Google Form

Students selected based on HISinONE Priority.

Students assigned topics based on their preferences

By **27.04.2026**

Please check the course website for more information:

https://rl.uni-freiburg.de/teaching/ss26/seminar-limited-supervision

Important Dates

Event	Date	Time
Lecture 1: Introduction *	23.04.2026	14:00
HISinOne registration + Paper Selection	27.04.2026	
Place allocation	29.04.2026	
Topic assignment	05.05.2026	
Supervisor Meeting	Mid June	
Lecture 2: *How to do a good presentation* *	25.06.2026	16h
Lecture 3: Block Seminar Presentations *	24.07.2026	9h - 17h
Summary submission	02.08.2026	23h59

*** Mandatory in-person attendance**

Seminar

Objectives

- Learn to conduct **literature research.**
- 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

Seminar

Structure Overview

- Form groups of three students
- Each group is assigned a **topic** and **two papers**
  - starting points for literature research
- **Group:** Familiarize with **topic** and identify (at least **one**) **additional** relevant **paper**(s) for your topic
  - add an additional angle to approach the topic
  - assign one paper per member for deep technical reading
  - **Deadline: June 14**, Supervisor meeting in the following week
- **Individual Student:** familiarize yourself with assigned paper
- **Individual Student**: condense important aspects of selected papers to slides and written summary
- **Group**: Merge to one presentation, discussing connections and distinctions between the individual works

Evaluation Criteria

Evaluation	Due Date
Literature Review (group)	14.06.2026
Seminar Presentation (group)	24.07.2026
Paper Summary (individual)	02.08.2026

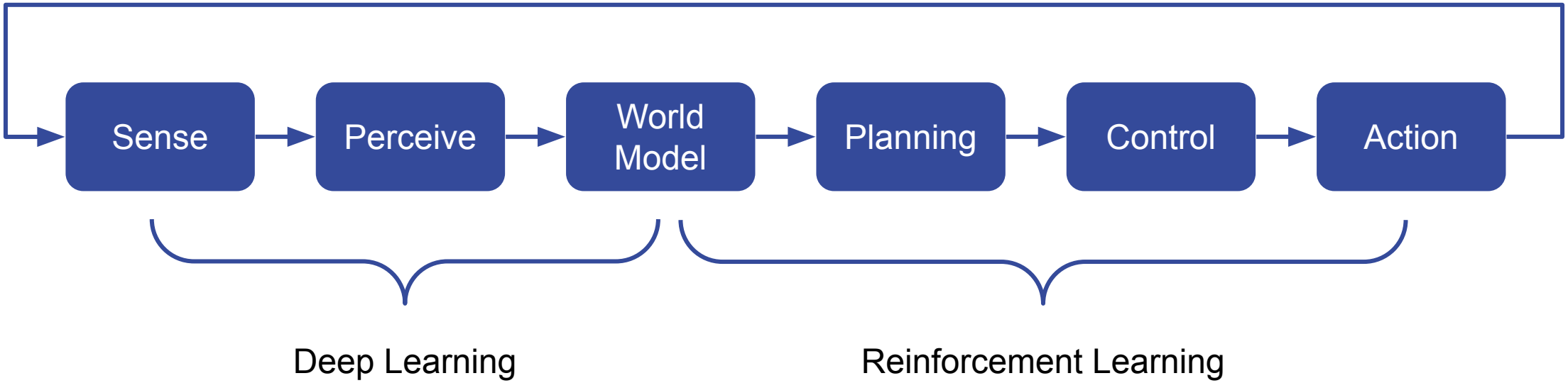
- **Presentation:** at most **30 min** per **group**. (10 min per group member)
- **Summary:** at most **7 pages** excluding bibliography and figures.
- **Final grade:**
  - Literature Review (relevance and justification of additional paper(s)) **(group)**
  - Presentation (slides & delivery) **(group)**
  - Seminar Participation **(individual)**
  - Summary **(individual)**

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** ?

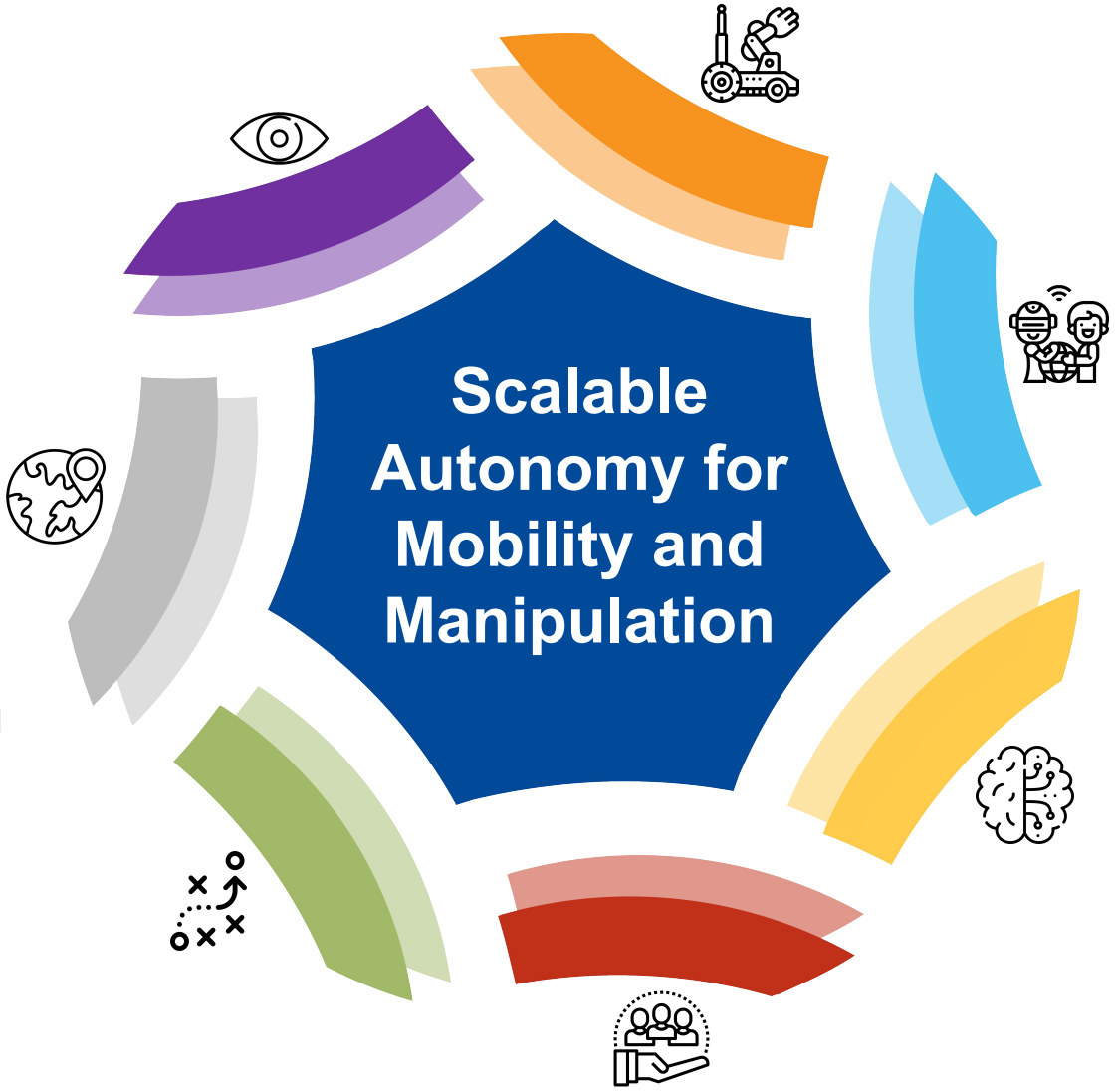
Research Areas

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

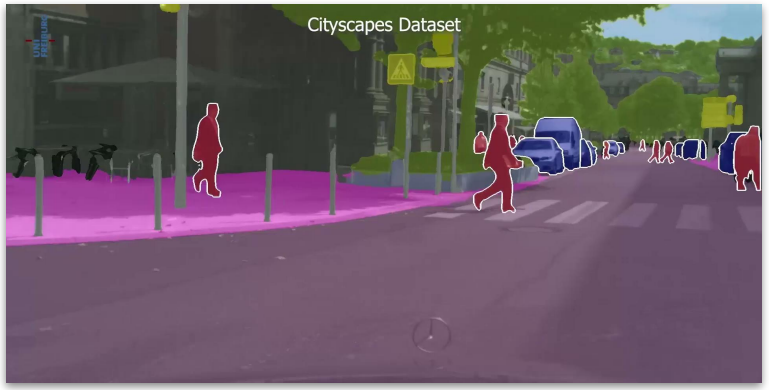


Mobile Manipulation
- Whole-Body Motion
- Long-Horizon Reasoning
- Planning for Sensing

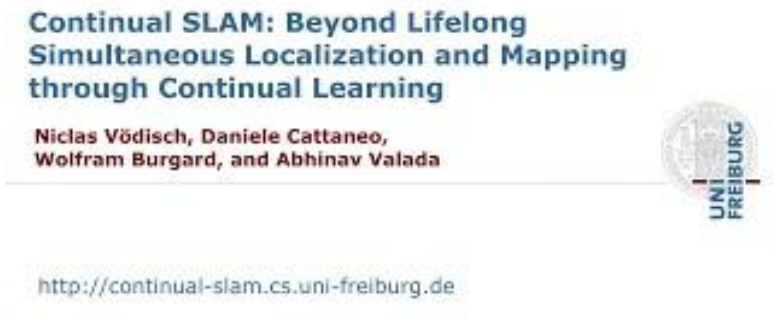
Human-Robot Interaction
- Socially-Compliant Behavior
- Human-Robot Collaboration
- Behavior Adaptation & Safety

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

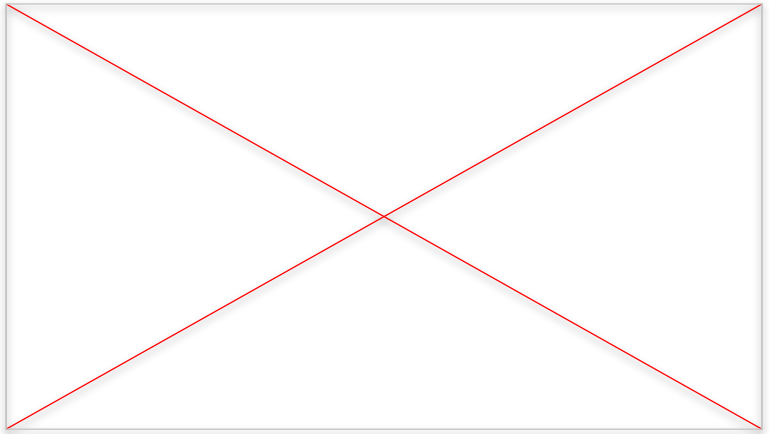
Many Seminal Works



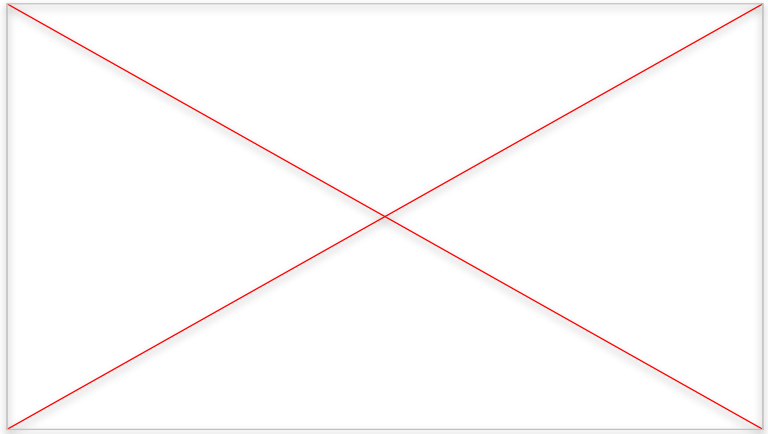
Scene Understanding



Simultaneous Localization and Mapping

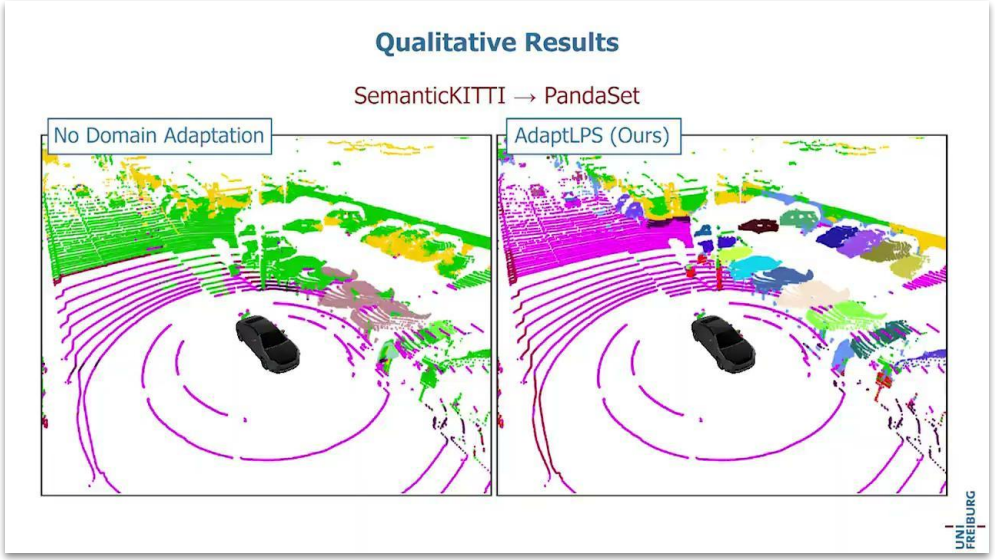


Motion Planning



Learning from Demonstrations

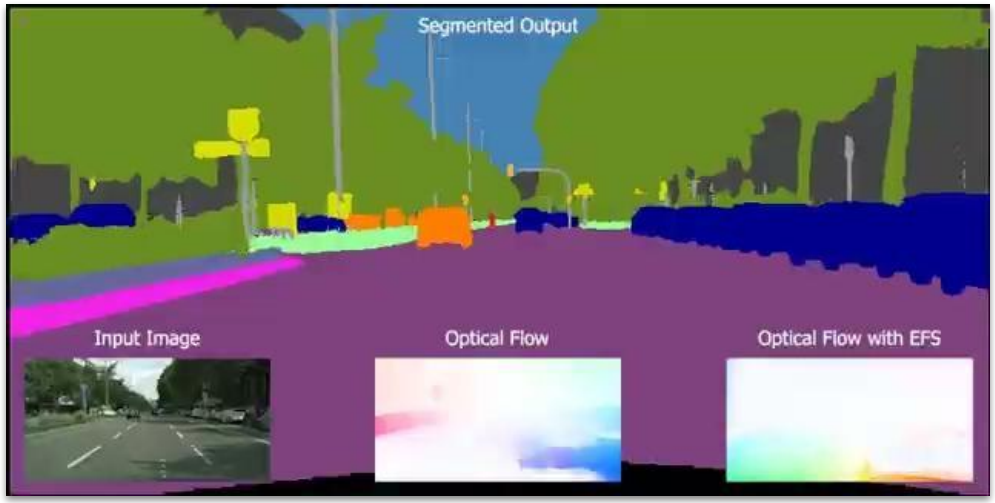
Robotic Perception — Mobility



Unsupervised LiDAR Domain Adaptation

Besic, Gosala, Cattaneo, Valada

RA-L ‘22

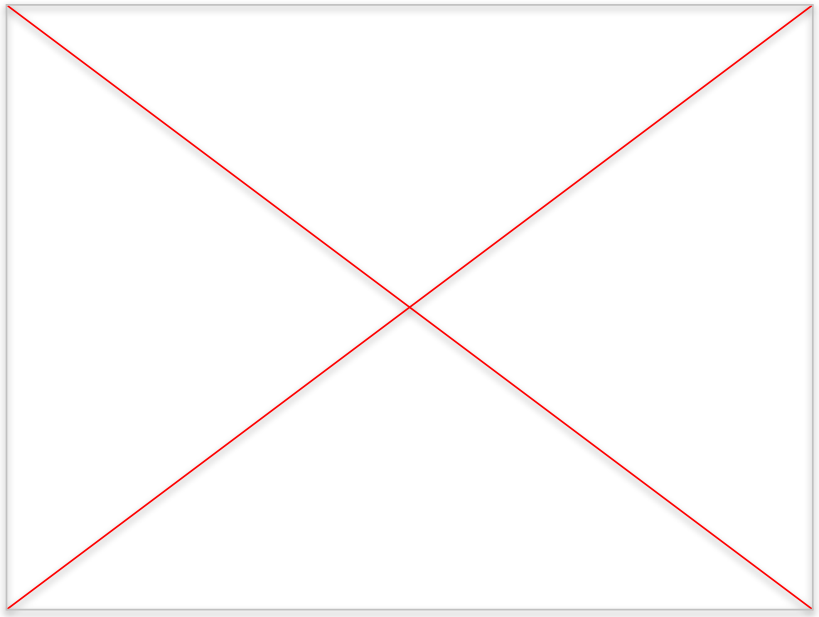


Semantic Motion Segmentation

Vertens, Valada, Burgard

ICRA ‘17

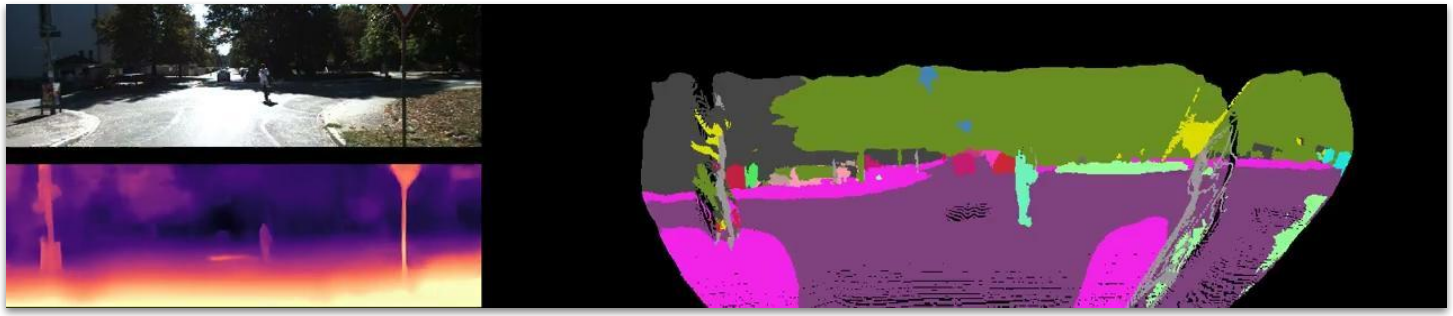
Mapping and Localization



Continual SLAM

Vödisch, Cattaneo, Burgard, Valada

ISSR ‘22



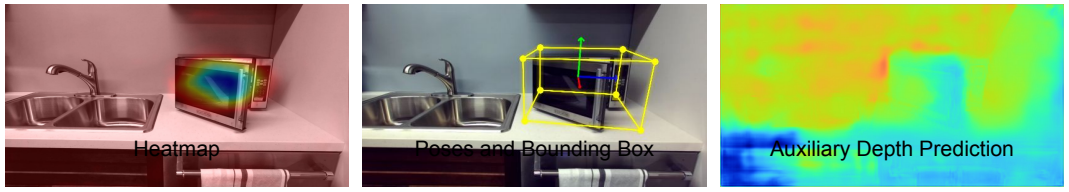
Continual Depth Estimation and Segmentation

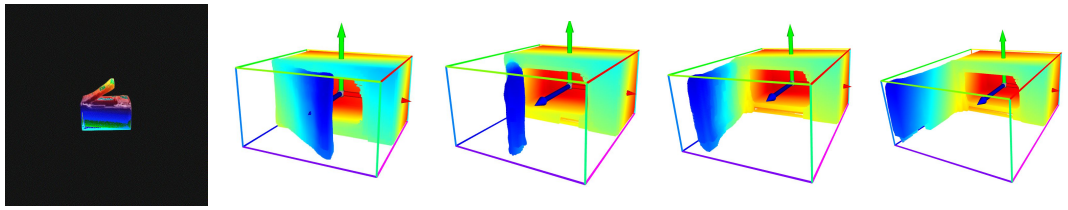
Vödisch, Petek, Burgard, Valada

RSS ‘23

Robotic Perception — Manipulation

Single-Shot Reconstruction

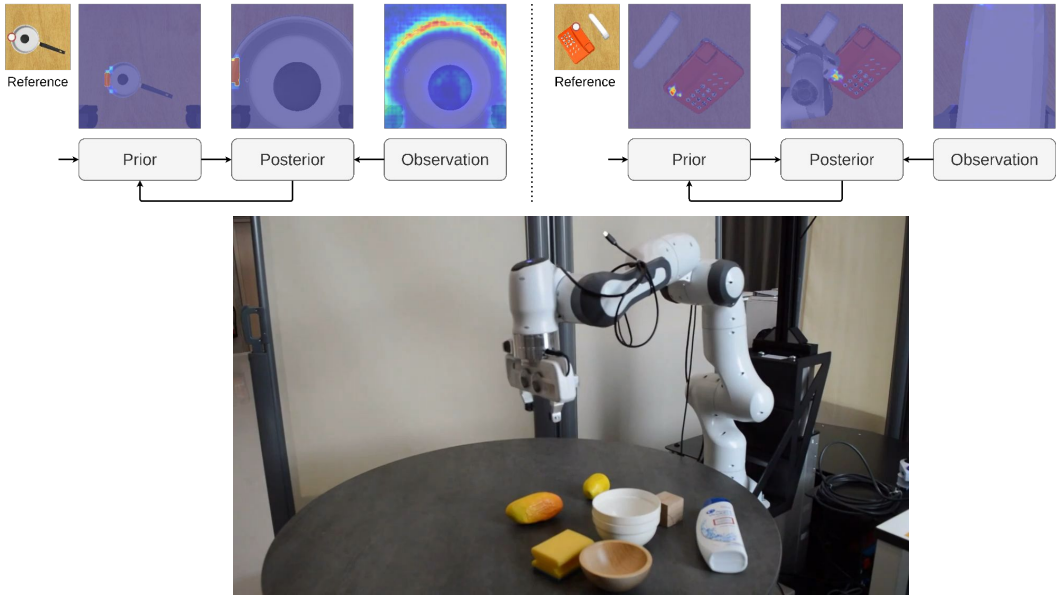




Category and Joint Agnostic Reconstruction of ARTiculated Objects

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

III.

Topics

Seminar Papers

Supervisor: Yao Lu

Text Adaption of Self-Supervised Vision Foundation Models

https://arxiv.org/abs/2412.16334 , publisher
https://arxiv.org/abs/2411.19331 , publisher

- **Motivation**: self-supervised vision foundation models learn outstanding visual features, but have no language grounding -> unusable for open-vocabulary task.
- **Paper 1:** *dino.txt* (CVPR 2025): systematic application of the LiT (Locked Image Tuning) strategy on DINOv2
- **Paper 2:** *Talk2DINO* (ICCV 2025): in contrast to dino.txt, Talk2DINO focuses on adapting embeddings from other foundation model (CLIP) to DINO vision space

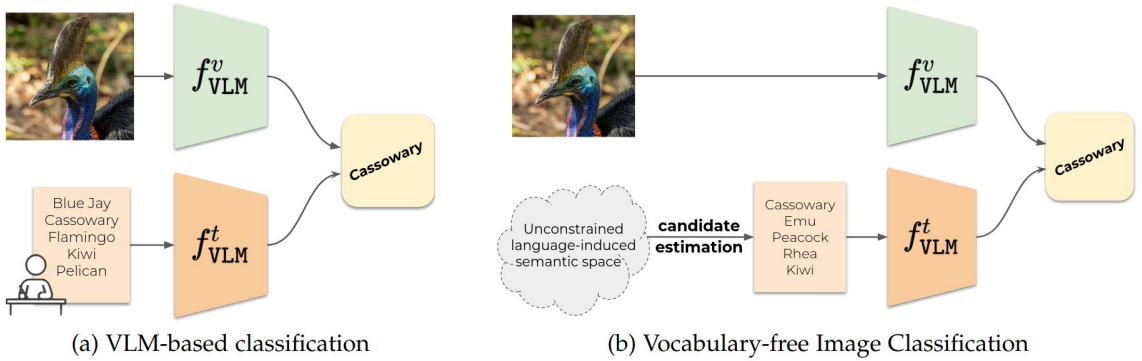
Example output of adding language grounding to DINO [Jose et al. CVPR 2025]

Supervisor: Yao Lu

Auto-Vocabulary Semantic Segmentation: Open Vocabulary Without Text Prompts

https://arxiv.org/abs/2312.04539 , publisher
https://arxiv.org/abs/2404.10864 , publisher

- **Motivation**: best of both world: traditional semantic segmentation needs predefined label set, open-vocabulary does not limit semantic class, but require text prompts as input
- **Paper 1:** *CASED* (IEEE TPAMI 2026): combines pretrained vision-languange model with external dataset
- **Paper 2:** *Autoseg* (CVPR 2025): pipeline of multiple foundation models



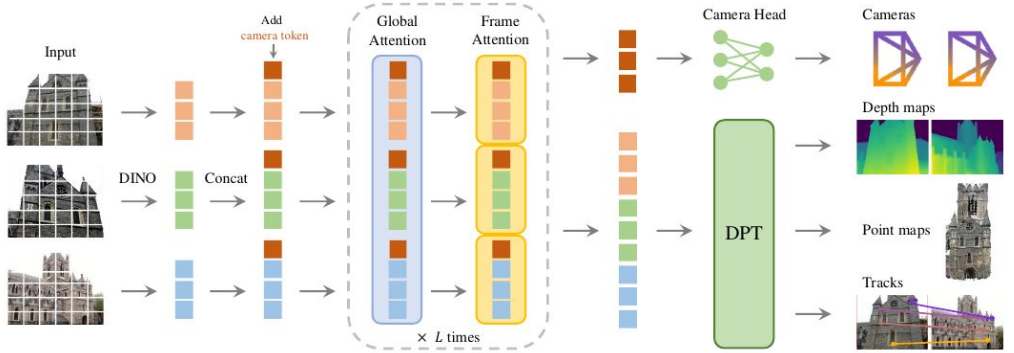
Overview of task differences between Open-Vocabulary and Auto-Vocabulary Classification [Conti et al. IEEE TPAMI 2026]

Supervisor: Simon Bultmann

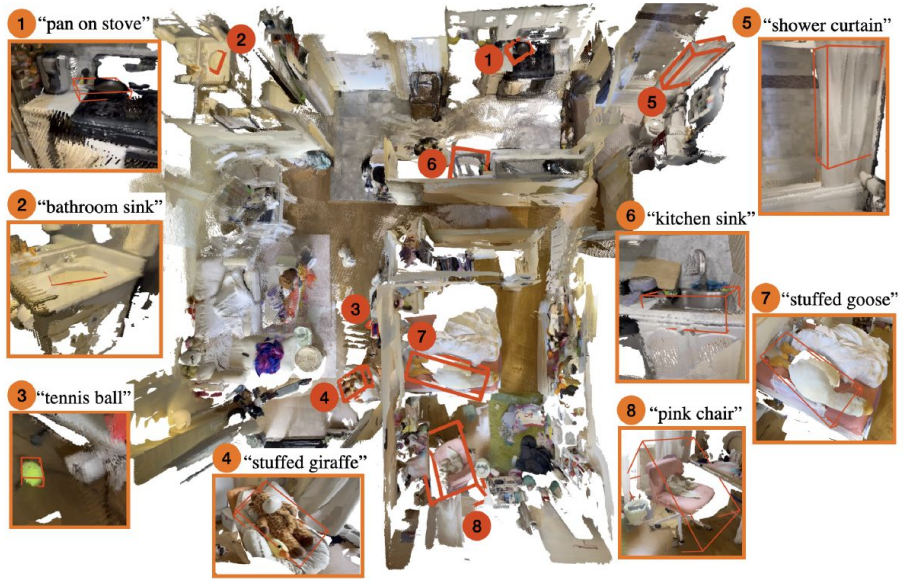
Feed-forward 3D Geometry Foundation Models

https://arxiv.org/abs/2601.19887
https://arxiv.org/abs/2603.08254

- **Motivation**: Recent transformer-based geometry foundation models (e.g. VGGT) enable downstream tasks like SLAM, open-vocabulary mapping, and dynamic scene understanding.
- **Paper 1:** *VGGT-SLAM 2.0* (preprint): Builds **real-time SLAM** system (incl. open-vocabulary objects) on top of **pretrained VGGT** geometry model
- **Paper 2:** *DynamicVGGT* (preprint): Extends feed-forward geometry models to **dynamic scenes**, predicting both current + future point maps



VGGT pipeline [Wang et al., CVPR 2025]



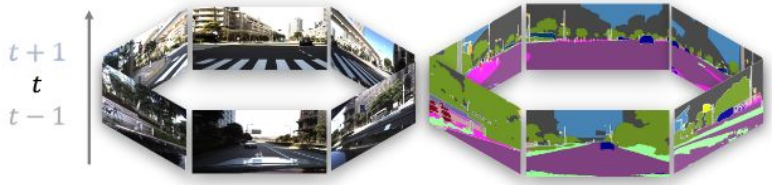
VGGT-SLAM 2.0 map example [Maggio & Carlone, arxiv 2601.19887]

Supervisor: Simon Bultmann

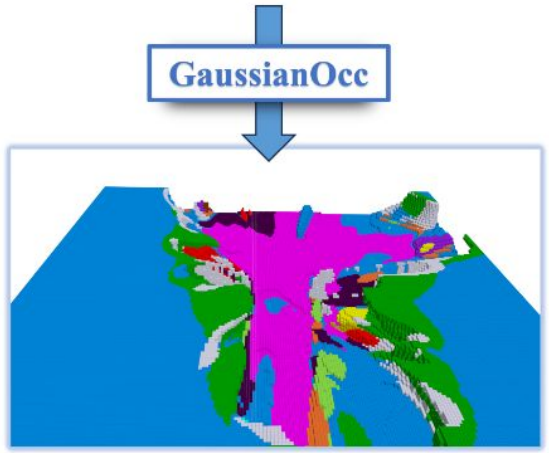
Label-Efficient Semantic 3D Occupancy Prediction

https://arxiv.org/abs/2408.11447, publisher
https://arxiv.org/abs/2511.15396

- **Motivation**: Dense 3D occupancy (**geometry + semantics**) from **surround-view images** is a powerful scene representation for autonomous driving. However, dense **3D voxel labels** (LiDAR-based) are **expensive**
- **Paper 1:** *GaussianOcc* (ICCV 2025): Fully self-supervised 3D occupancy with Gaussian Splatting, using **rendering**-based **2D supervision**
- **Paper 2:** *ShelfOcc* (accepted for CVPR 2026): **3D supervision** via geometry foundation models (no LiDAR)



Surround image sequence, camera extrinsic and corresponding 2D semantic annotation



Surround view 3D occupancy estimation

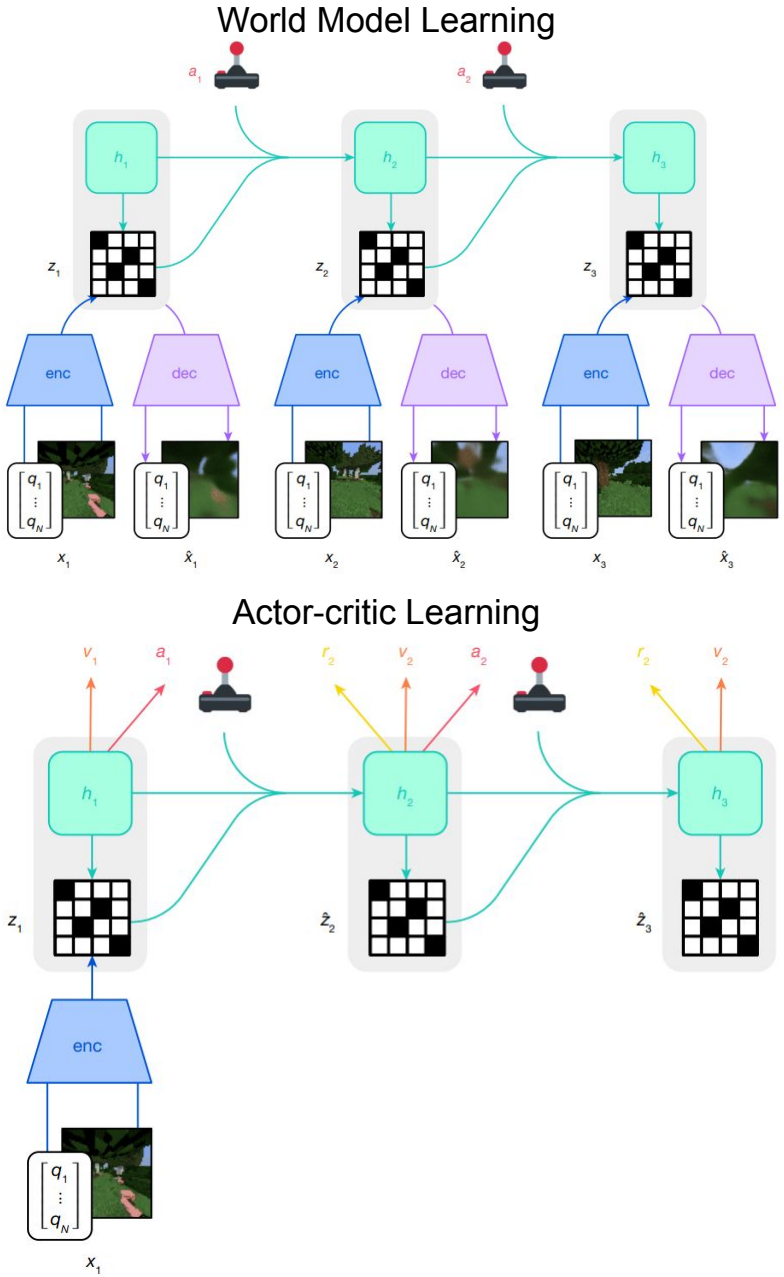
3D occupancy estimation from surround-view images using volume rendering [Gan et al. ICCV 2025]

Supervisor: Iman Nematollahi

Learning from Imagination: World Models for Scalable Control

https://arxiv.org/abs/2509.24527
https://arxiv.org/abs/2603.19312

- **Core idea:** Learn a predictive latent model of environment dynamics and train agents “in imagination” from offline or unlabeled data
- **Training Agents Inside of Scalable World Models (Dreamer 4):** Scales world models to complex tasks (e.g., Minecraft) and trains policies purely inside the learned simulator, enabling long-horizon control without real interaction
- **LeWorldModel (LeWM):** Proposes a simple, stable JEPA-based world model trained end-to-end from pixels with minimal losses, enabling efficient planning and learning of physical structure



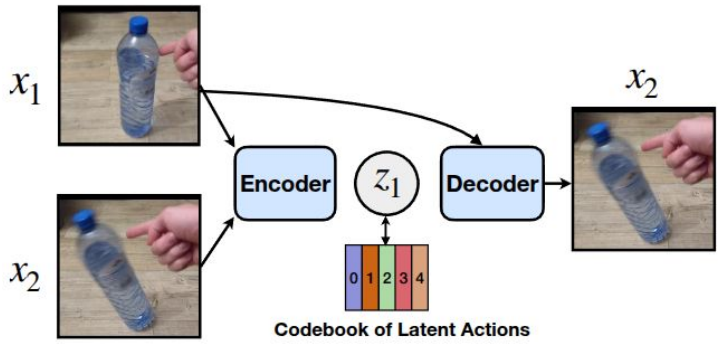
Supervisor: Iman Nematollahi

Learning to Act without Actions: Latent Action Spaces from Video

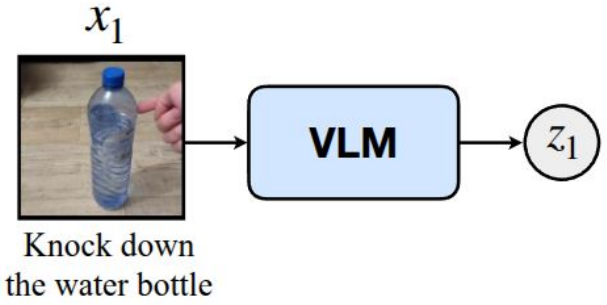
https://arxiv.org/abs/2410.11758
https://arxiv.org/abs/2505.08787

- **Core idea:** Learn action representations implicitly from videos or observations, reducing reliance on labeled actions and enabling cross-domain skill transfer
- **LAPA (Latent Action Pretraining):** Learns latent action spaces from videos, allowing agents to acquire control-relevant structure without explicit action supervision
- **UniSkill:** Learns transferable skill representations across embodiments, enabling imitation of human videos despite different morphologies (cross-embodiment learning)

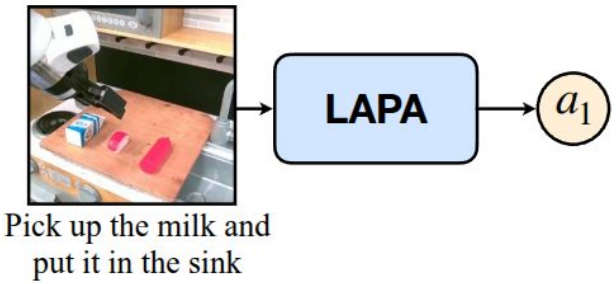
1. Latent Action Quantization



2. Latent Pretraining



3. Action Finetuning

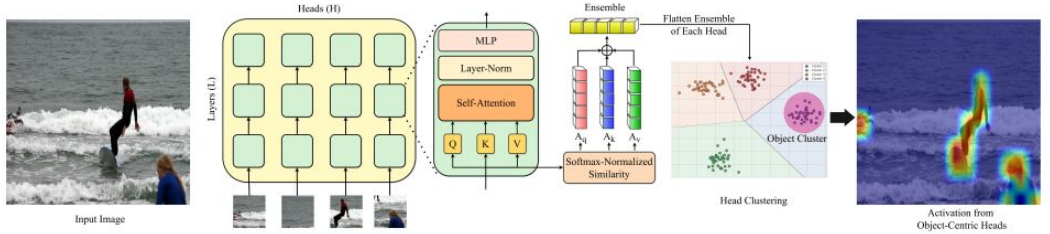


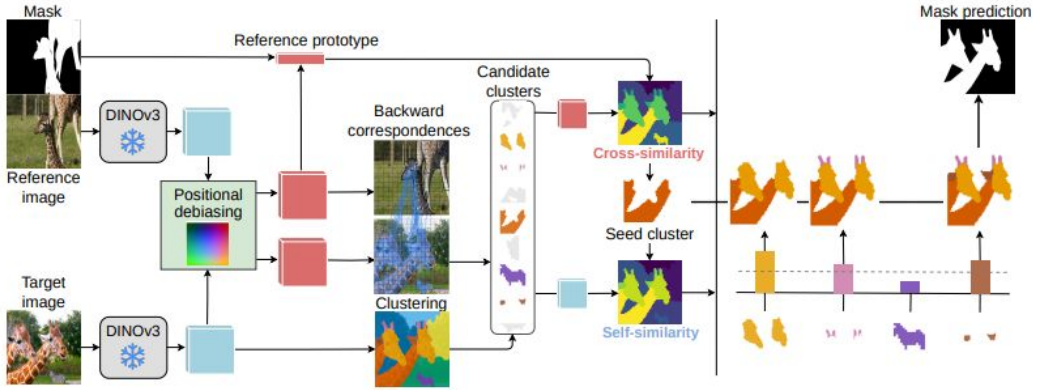
Supervisor: Adrian Röfer

Training-Free Segmentation methods on ViT-Features

https://arxiv.org/pdf/2603.26127
https://arxiv.org/pdf/2603.28480

- **Core idea:** Deep Vision Features carry a surprising amount of semantic information. Instead of training models, try to utilize the feature-space statistically.
- **Finding Distributed Object-Centric Properties in Self-Supervised Transformers:** Training free object segmentation by exploiting similarity maps across the ViT.
- **INSID3: Training-Free In-Context Segmentation with DINOv3:** Promtable training free object segmentation by clustering and model debiasing.



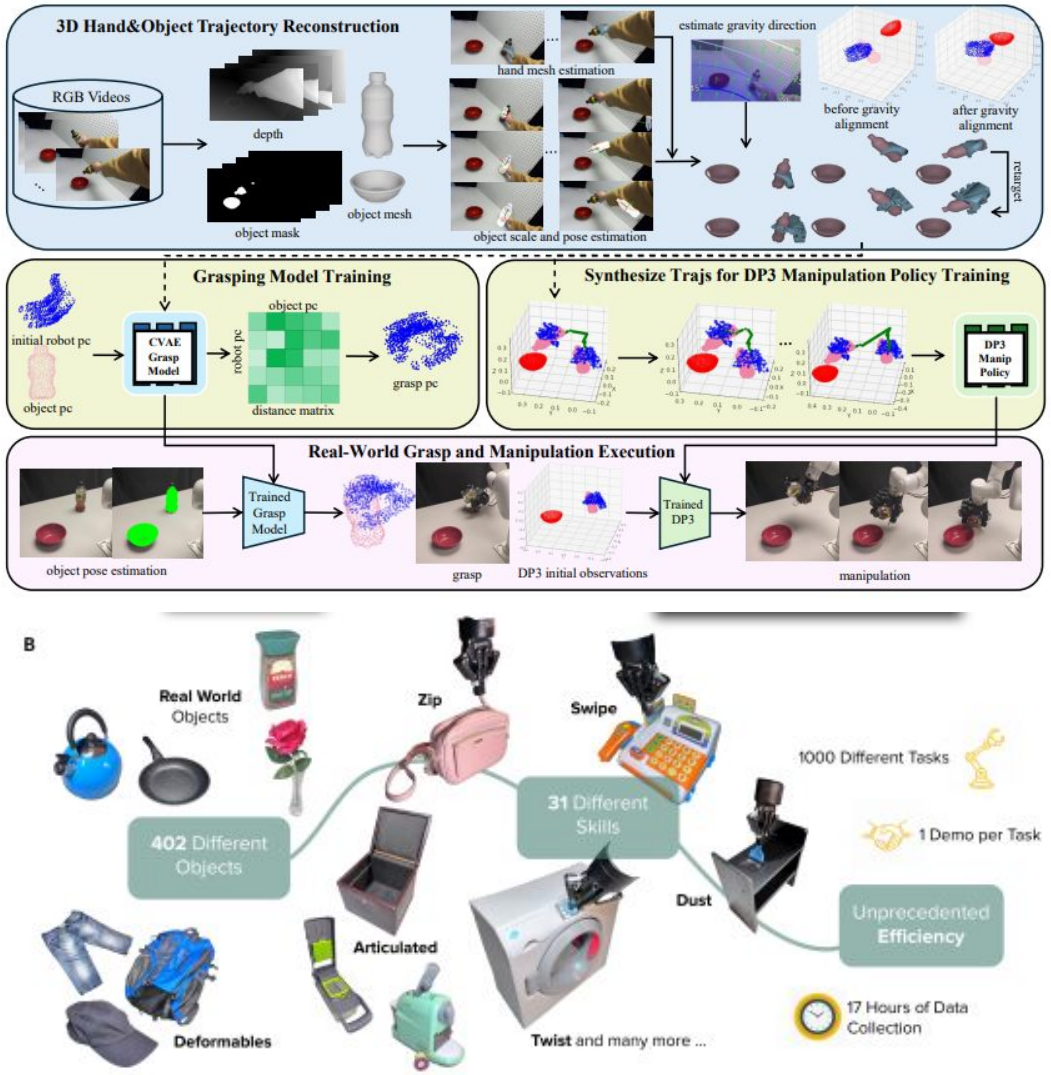


Supervisor: Adrian Röfer

Efficient Robotic Imitation Learning

https://arxiv.org/pdf/2602.09013
https://arxiv.org/pdf/2511.10110

- **Core idea:** Deep behavior cloning and VLAs require large amounts of training data which is costly to collect. Here we study approaches requiring a handful of demonstrations.
- **Dexterous Manipulation Policies from RGB Human Videos via 3D Hand-Object Trajectory Reconstruction:** Integrated method for generating training data for grasping and motion from human videos.
- **Learning a Thousand Tasks in a Day:** Enable fast learning by prototype retrieval and re-targeting motions to the execution context.



?

Questions