### universitätfreiburg

## SS25 Seminar Learning with Limited Supervision

Dr. Simon Bultmann

Robot Learning Lab

24 April 2025



## Agenda

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

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

III. Topics: Seminar Papers.

?. Questions.

# 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.

#### left 48. 1 SCHRÖDINGER: Die gegenwärtige Situation in der Quantenmechanik GER: Die gegenwärtige Situation in der Quantenmechanik wollte ücker Zahl esagt ingen ssage e des Jwert klas-iben? fälle, der gibt hern, thnit-selbe sich ürde, Elek-man iteren CHRÖDINGER: Die gegenwärtige Situation in der Ouantenmechanik komm erster ndere r der r der en ken lessen sind. i Bild tende rückt hiche ver-heorie t, den zerein-er es ignet, ungen l aus-lacht, belfs, Ent-t prä-acht-t, was acht-t, was rrund-tende DIE NATURWISSENSCHAFTEN tz zu den le übriger 23. Jahrgang 20. November 1035 Heft 48 as Privileg sind: die Die gegenwärtige Situation in der Quantenmechanik. Impuls-tes, die Von E. Schrödinger. Oxford Inhaltsühersicht sbewegung. besondere Gebilde, das sich mit der Zeit verändert, das ver § 1. Die Physik der Modelle. chiedene Zustände annehmen kann; und wenn § 2. Die Statistik der Modellvariablen in der Ouar in Zustand durch die nötige Zahl von Bestin sie haben ene Werte. ie im Laufe mungsstücken bekannt gemacht ist, so sind nicht tenmechanik Beispiele für Wahrscheinlichkeitss nur alle anderen Stücke in diesem Augenblick mit Kann man der Theorie ideale Gesamtheite egeben (wie oben am Dreieck erläutert), sondern ehalten sie z ebenso alle Stücke, der genaue Zustand, zi rum auch die Variablen wirklich verwaschen eder bestimmten späteren Zeit; ähnlich wie die Beschaffenheit eines Dreiecks an der Basis seine ied von den er bewußte Wechsel des erkenntnistheoreti schen Standpunktes. Die w-Funktion als Katalog der Erwartung Beschaffenheit an der Spitze bestimmt. Es ge-hört mit zum inneren Gesetz des Gebildes, sich in resses Aber bald Hälfte klares nuben igen? v.v.) stets i ihre ntlich ; der n der ent-i Vor-tralen inten-ie der Vesen } mes ents i der i kt. In ischau-it, bei-it" um lenken aß die Dinge t dann radio-Grade r Zeit-eht, in Kern orie des Messens, erster Teil in der Die w-Funktion als Beschreibung des Zu estimmter Weise zu verändern, das heißt, wenn es in einem bestimmten Anfangsz and sich selb . des Messens, zweiter Teil. fhebung der Verschränkung. Das Er überlassen wird, eine bestimmte Folge von Zu eicht noch ständen kontinuierlich zu durchlaufen, deren jede ie ich fest gebnis abhängig vom Willen des Experimen es zu ganz bestimmter Zeit erreicht. Das ist seine n Angel Natur, das ist die Hypothese, die man, wie ich aß Modelle Lin Deispiel, Fortiestzung des Beispiels: alle möglichen Messungen sind eindeutig verschränkt. Fort Die Anderung der Verschränkung mit der Zeit. Bedenken gegen die Sonderstellung der Zeit. FJ. Naturprinzip oder Rechenkunstgriff? oben sagte, auf Grund intuitiver Imagination setzt der Natur Natürlich ist man nicht so einfältig zu denke daß solchermaßen zu erraten sei, wie es auf de das glaub Welt wirklich zugeht. Um anzudeuten, daß ma usgespielt verwendet § 1. Die Physik der Modelle. das nicht denkt, nennt man den präzisen Denk In der zweiten Hälfte des vorigen Jahrhunderts behelf, den man sich geschaffen hat, gern ein gative der war aus den großen Erfolgen der kinetischen Gastheorie und der mechanischen Theorie der Bild oder ein Modell. Mit seiner nachsichtslose ndern auch Klarheit, die ohne Willkür nicht herbeizuführe rminierung. th wie ischen von mente Wahr-n wir i stets tellter zu er-e ent-t Zeit schein-ne die Wärme ein Ideal der exakten Naturbeschreibung ist, hat man es lediglich darauf abgeschen, daf eine ganz bestimmte Hypothese in ihren Folger eschrieben iablen der achsen, das als Krönung jahrhundertengen Forschens und Erfüllung jahrtausendealter geprüft werden kann, ohne neuer Willkür Raur rden. Fol-Hoffnung einen Höhepunkt bildet und das klas sische heißt. Dieses sind seine Züge. zu geben während der langwierigen Rechnungen durch die man Folgerungen ableitet. Da hat mat andes geht Von den Naturobiekten deren beobachtetes wohlaus gebundene Marschroute und errechnet eigentlic Verhalten man erfassen möchte, bildet man, genur, was ein kluger Hans aus den Daten direk atzes von stützt auf die experimentellen Daten, die man herauslesen würde! Man weiß dann wenigsten icen lassen besitzt, aber ohne der intuitiven Imagination zu wehren, eine Vorstellung, die in allen Details sweise den wo die Willkür steckt und wo man zu bessern ha wenn's mit der Erfahrung nicht stimmt: in de Geschwingenau ausgearbeitet ist, viel genauer als irgenddere Grup-bleibt dann Ausgangshypothese, im Modell. Dazu muß ma che Erfahrung in Anschung ihres begrenzten stets bereit scin. Wenn bei vielen verschieder Umfangs je verbürgen kann. Die Vorstellung in artigen Experimenten das Naturobjekt sich wirk lige Stücke lich so benimmt wie das Modell, so freut man sich ihrer absoluten Bestimmtheit gleicht einem mathetimmtheit matischen Gebilde oder einer geometrischen Figur, und denkt, daß unser Bild in den wesentliche werden in welche aus einer Anzahl von Bestimmungsstücker Zügen der Wirklichkeit gemäß ist. nmt es be hen Modell ganz und gar berechnet werden kann; wie z. B. an annt sein. einem neuartigen Experiment oder bei Verfeine einem Dreieck eine Seite und die zwei ihr an-liegenden Winkel, als Bestimmungsstücke, den rung der Meßtechnik nicht mehr, so ist nicht ø am besten sagt, daß man sich nicht freut. Denn im Grund hen Mechaablen dafür dritten Winkel, die anderen zwei Seiten, die drei ist das die Art, wie allmählich eine immer besser Höhen, den Radius des eingeschriebenen Kreises Anpassung des Bildes, das heißt unserer Gedanke isw. mit bestimmen. Von einer geometrischen an die Tatsachen gelingen kann Die klassische Methode des präzisen Figur unterscheidet sich die Vorstellung ihrem Wesen nach bloß durch den wichtigen Umstand, hat den Hauptzweck, die unvermeidliche Willkür daß sie auch noch in der Zeit als vierter Dimension in den Annahmen sauber isoliert zu halten, ich ebenso klar bestimmt ist wie jene in den drei Dimensionen des Raumes. Das heißt es handelt für den historischen Anpasungsprozeß an die sich (was ja selbstverständlich ist) stets um ein fortschreitende Erfahrung. Vielleicht liegt der

## **Enrollment Procedure**

Select <u>3 papers</u> in decreasing order of preference.	Register for t seminar in HISinOne.	:he	Students selected based on HISinONE Priority.	Students assigned papers based on their preferences
Fill in our <u>Google Form</u>				
			By <u>28.04.2</u>	<u>.025</u>

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

universität freiburg

8

#### **Robot Learning Lab**

### **Autonomous Robotics**



Can we learn certain parts of this pipeline?

## Robot Learning Lab 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
- Explainability & Privacy
- Practical ethics

## Many Seminal Works



Scene Understanding



Motion Planning



Simultaneous Localization and Mapping



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



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

#### Bird's-Eye-View Panoptic Maps



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

Predicting panoptic HD maps from monocular frontal view images

#### universitätfreiburg

Input Frontal View Image

## **Robotic Perception - Mobility**

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

Unsupervised LiDAR Domain Adaptation



#### Sirohi, et al: ICRA'23

Panoptic Uncertainty Estimation





24. April 2025

## **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



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



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



universitätfreiburg

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

Continual SLAM



## **Mapping and Localization**



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

Greve, Vödisch, Büchner, Valada: ICRA'24



universitätfreiburg

#### Collaborative Dynamic 3D Scene Graphs

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



Long-Horizon Object Search



## Language-Grounded Learning

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

#### MoMa-LLM

Task: I am hungry. Find me something for breakfast.

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

#### Language-Grounded Navigation



universität freiburg

Speed 4X

3x

## Language-Grounded Learning

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

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

#### MoMa-LLM

Task: I am hungry. Find me something for breakfast.

Language-Grounded Navigation







iniversität freibur



Go to the toilet in the bathroom on floor 2

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 pixellevel 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.



Chamfer distance (↓)

## 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.



## 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



## 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



## 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



## 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



## 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**.



Add

Global

Frame



Camera Head

Cameras

Tracks

# 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, nonoverlapping views, diverse lighting conditions, and large cross-environment variations.



Query Image  $(I_Q)$ 

## 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.





## 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.



train and store a LoRA

dataset

index LoRA based on dataset centroid

## 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



## 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 selfsupervised annotations



## 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



# 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

universität freiburg

Simon Bultmann | Robot Learning Lab | 24. April 2025