

# Seminar Vision Systems MA-INF 4208

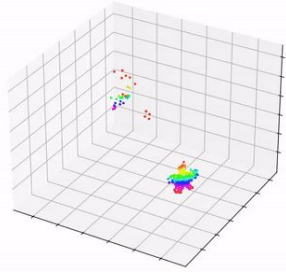
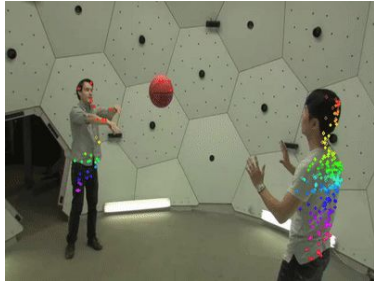
---

11.07.2025

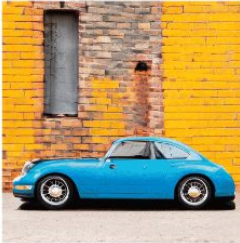
PROF. SVEN BEHNKE, ANGEL VILLAR-CORRALES

Contact: [villar@ais.uni-bonn.de](mailto:villar@ais.uni-bonn.de)

# The Age of Deep Learning



A living room with a fireplace at a wood cabin. Interior design.

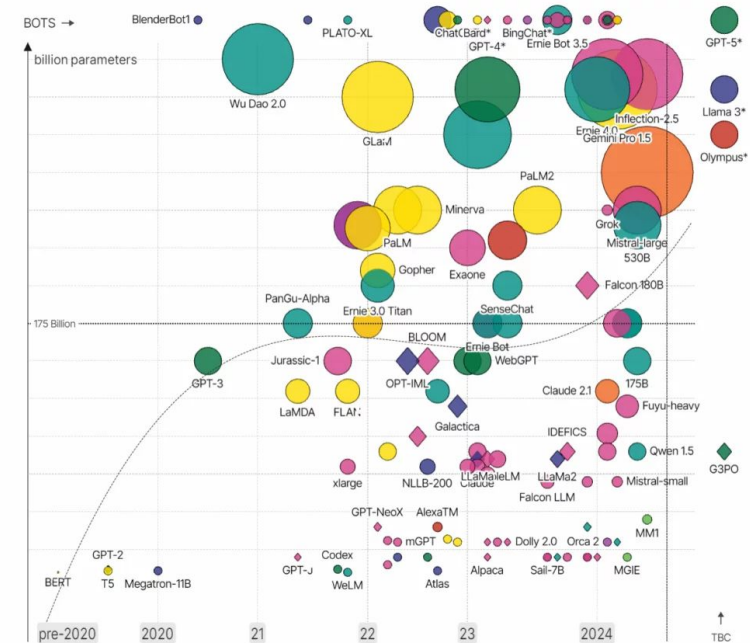
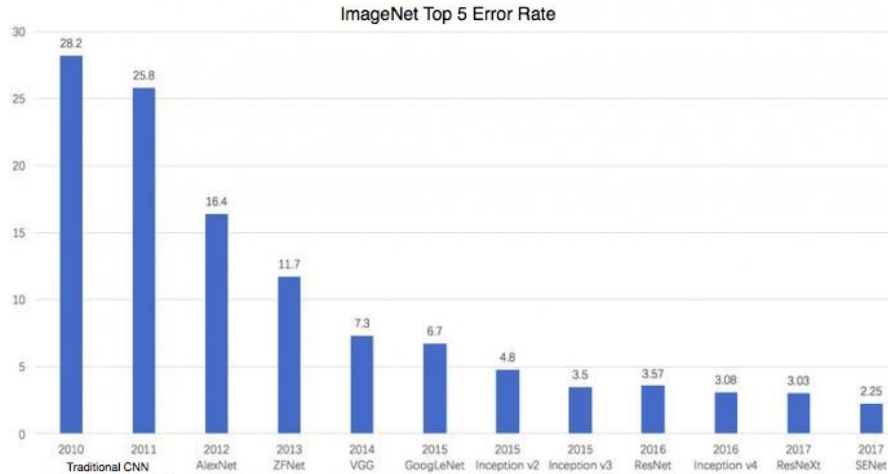


a blue Porsche 356 parked in front of a yellow brick wall.



Eiffel Tower, landscape photography

# The Age of Deep Learning



# The Age of Deep Learning

---



HUGGING FACE

Google



DAIMLER

amazon

SIEMENS



TOYOTA  
RESEARCH INSTITUTE



TESLA



Microsoft



# In this seminar...

---

- Acquire/improve ability to:
  - deal with scientific publications (e.g. papers)
  - write a scientific report
  - present a scientific topic to an audience
  - engage technical topics

 Important skills for Master Thesis!

# In this seminar

- Discuss trending topics in deep learning and computer vision
- We will cover the following topics
  - 3D Deep Learning
  - Representation Learning from Images & Video
  - Advances in Neural Network Architectures
  - World Models

## Seminar: Vision Systems MA-INF 4208

Prof. Dr. Sven Behnke, Angel Villar-Corrales

### 1 Paper List

#### 1. 3D Deep Learning

- a) Wang, Jianyuan, et al. *VGGT: Visual Geometry Grounded Transformer*. CVPR. 2025. [Link](#)
- b) Asim Mohammad, et al. *MEt3R: Measuring Multi-View Consistency in Generated Images*. CVPR. 2025. [Link](#)
- c) Li, Zhengqi, et al. *MegaSaM: Accurate, Fast, and Robust Structure and Motion from Casual Dynamic Videos*. CVPR. 2025. [Link](#)

#### 2. Representation Learning from Images & Video

- a) van Steenkiste, Sjoerd, et al. *Moving Off-the-Grid: Scene-Grounded Video Representations*. NeurIPS. 2024. [Link](#)
- b) Cijo, Jose, et al. *DINOv2 Meets Text: A Unified Framework for Image- and Pixel-Level Vision-Language Alignment*. CVPR. 2025. [Link](#)
- c) Tschannen, Michael, et al. *SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features*. ArXiv Preprint. 2025. [Link](#)

#### 3. Advances in Neural Network Architectures

- a) Braso, Guillem, et al. *Native Segmentation Vision Transformers*. ArXiv Preprint. 2025. [Link](#)
- b) Assran, Mahmoud, et al. *V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning*. ArXiv Preprint. 2025. [Link](#)
- c) Ma, Xin, et al. *Latte: Latent Diffusion Transformer for Video Generation*. TMLR. 2025. [Link](#)

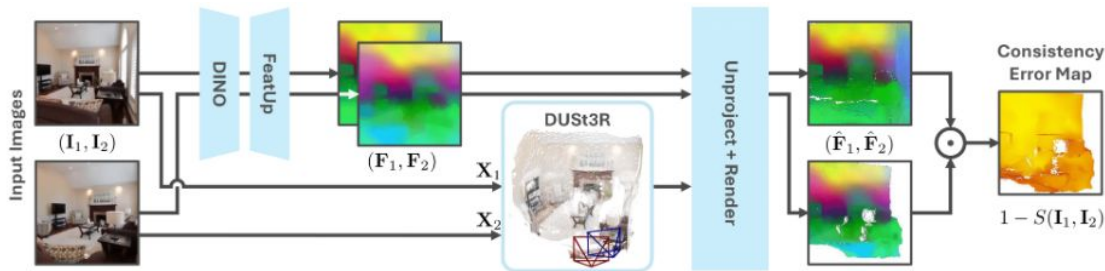
#### 4. World Models

- a) Bar, Amir, et al. *Navigation World Models*. CVPR. 2025. [Link](#)
- b) Zhou, Gaoyue, et al. *DINO-WM: World Models on Pre-trained Visual Features enable Zero-shot Planning*. ICLR. 2025. [Link](#)
- c) Gao, Shenyuan, et al. *AdaWorld: Learning Adaptable World Models with Latent Actions*. ICML. 2025. [Link](#)

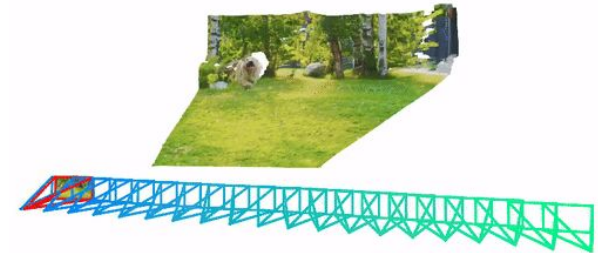
**Paper List:** <https://www.ais.uni-bonn.de/SS/SeminarVision/PaperList.pdf>

# 3D Deep Learning

- Learning 3D-aware representations of a scene given a set of posed images
- Applications such as:
  - Novel View Synthesis
  - Evaluating 3D consistency
  - Structure-from-Motion



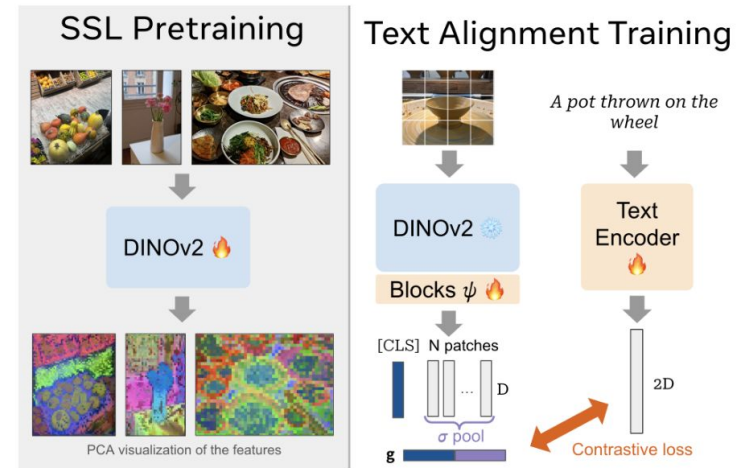
32 Views





# Representation Learning from Images & Video

- Learning representations from video data without annotations
- Applications such as:
  - Representation learning
  - Image-Language Grounding

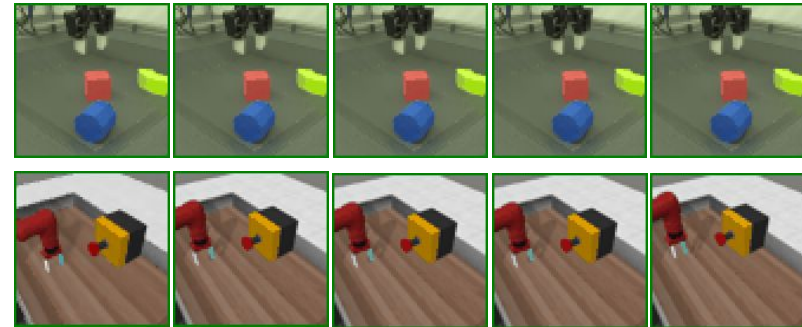
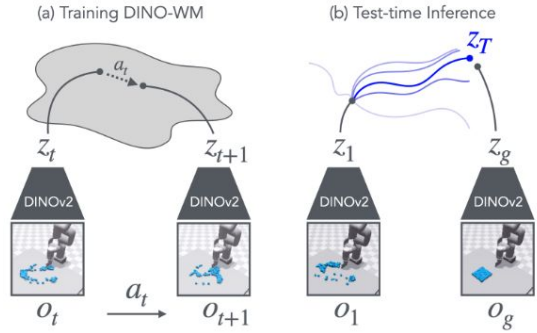






# World Models

- Models that learn to simulate environments and predict future possible outcomes
- Applications such as:
  - Navigation
  - Planning
  - Controllable video generation



# Get a Spot and Select your Topic

---

- Fill the following [form](#) no later than 14.07.2025
  - Your name & email
  - Matriculation number
  - Your three preferred papers
- Based on this form, I will and assign seminar spots and papers to review on Monday 18.07.2025
- Upon my confirmation:
  - Register in BASIS
  - Start working on your paper

**BASIS Registration opened until 07.08.2025!**

# Deliverables (preliminary dates)

---

- **Presentation:** Thursday 25.09.2025
  - 30 min presentation
  - 15 min discussion
- **Report:** Thursday 02.10.2025 (will be one week after presentations)
  - LaTeX template
  - 8-12 pages
  - Brief but readable and informative
  - BibTex citations

➡ Arrange a meeting with me  $\approx 1$  week before the presentation to check the preliminary materials for the presentation and report.

# Report

---

- Well structured:
  - Abstract
  - Introduction, methods, results, conclusion, ...
  - Tables and figures
  - Correct citations
- Your own scientific opinion:
  - What are the weak points of the paper?
  - What is missing?
  - Are comparisons fair and believable?
  - Possible future steps?

**We don't want a copy of the paper!**

# Grading

---



- 66.7%: Presentation
  - Quality of the presentation slides
  - Presentation skills
  - Ability to answer questions
- 33.3%: Report
  - Overall quality of the report
  - Critical thinking and own discussion
  - Understanding of the concept

# Seminar Alternative

## Seminar Cognitive Robotics: [Link](#)

- Same seminar format
- Papers more robotics related:
  - Grasping and Manipulation
  - **Robot vision & perception**
  - SLAM
  - Planning and Navigation

➤ Introductory meeting on **25.07.202**

UNIVERSITÄT **BONN** AIS

**Institute for Computer Science VI**  
**Autonomous Intelligent Systems**

[Home](#)  
[Persons](#)  
[Teaching](#)  
[Research](#)  
[Publications](#)  
[News](#)  
[Jobs](#)

**Seminar Cognitive Robotics (MA-INF 4211)**




Prof. Dr. Sven Behnke, Dr. Raphael Memmesheimer

First organizational meeting: 25.07.2025 11:00 c.t., Friedrich-Hirzebruch-Allee 6, Room 0.107  
 Registration of interest till 01.08.2025  
 BASIS registration: 08.09.2025  
 Seminar presentations on 26.09.2025 Room: Friedrich-Hirzebruch-Allee 6, Room 0.107

Cognitive robotics is an active research area at the border between artificial intelligence and robotics. It investigates and tries to implement in technical systems mental functions, which are associated with intelligence. This covers the perception of the environment, action planning, and learning.

In this seminar, we will cover research papers from the area of cognitive robotics. Details will be announced in the organizational meeting.

Some of our robots:

[University of Bonn](#), [Institute for Computer Science](#), [Computer Science VI - Intelligent Systems and Robotics](#) [Imprint](#) [Data Privacy Statement](#)



# Questions?



