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**