

Seminar: Vision Systems MA-INF 4208

Prof. Dr. Sven Behnke, Angel Villar-Corrales

1 Paper List

1. Foundation Models

- a) Pernias, Pablo, et al. *Wuerstehen: An Efficient Architecture for Large-Scale Text-to-Image Diffusion Models*. ICLR 2024. **Link**
- b) Yang, Lihe, et al. *Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data*. ArXiv Preprint 2024. **Link**
- c) Shvets, Mykhailo, et al. *Joint Depth Prediction and Semantic Segmentation with Multi-View SAM*. CVPR 2024. **Link**
- d) Brohan, Anthony, et al. *Rt 2: Vision-language-action models transfer web knowledge to robotic control*. ArXiv Preprint 2023. **Link**

2. Advances in Network Architectures and Learning Algorithms

- a) Liu, Yue, et al. *VMamba: Visual State Space Model*. ArXiv Preprint. 2024. **Link**
- b) Weinzaepfel, Philippe, et al. *CroCo v2: Improved Cross-view Completion Pre-training for Stereo Matching and Optical Flow*. CVPR. 2023. **Link**
- c) Xu, Jilan, et al. *Learning Open-vocabulary Semantic Segmentation Models From Natural Language Supervision*. CVPR. 2023. **Link**

3. Neural Rendering and Applications

- a) Yang, Jiawei, et al. *EmerNeRF: Emergent Spatial-Temporal Scene Decomposition via Self-Supervision*. ArXiv Preprint 2023. **Link**
- b) Kim, Chung, et al. *GARField: Group Anything with Radiance Fields*. ArXiv Preprint 2024. **Link**
- c) Yao-Chih Lee, et al. *Fast View Synthesis of Casual Videos* ArXiv Preprint 2023. **Link**

4. Unsupervised Learning from Objects and Videos

- a) Fan, Ke, et al. *Unsupervised Open-Vocabulary Object Localization in Videos*. ICCV. 2023. **Link**

- b) ~~Seitzer, Maximilian, et al. *DyST: Towards Dynamic Neural Scene Representations on Real World Videos*. ArXiv Preprint. 2023. [Link](#)~~
- c) Jabri Allan et al *DORSal: Diffusion for Object-centric Representations of Scenes et al.* ICLR 2024. [Link](#)
- d) Bao Zhipeng et al. *Object Discovery from Motion-Guided Tokens* CVPR 2023 [Link](#)