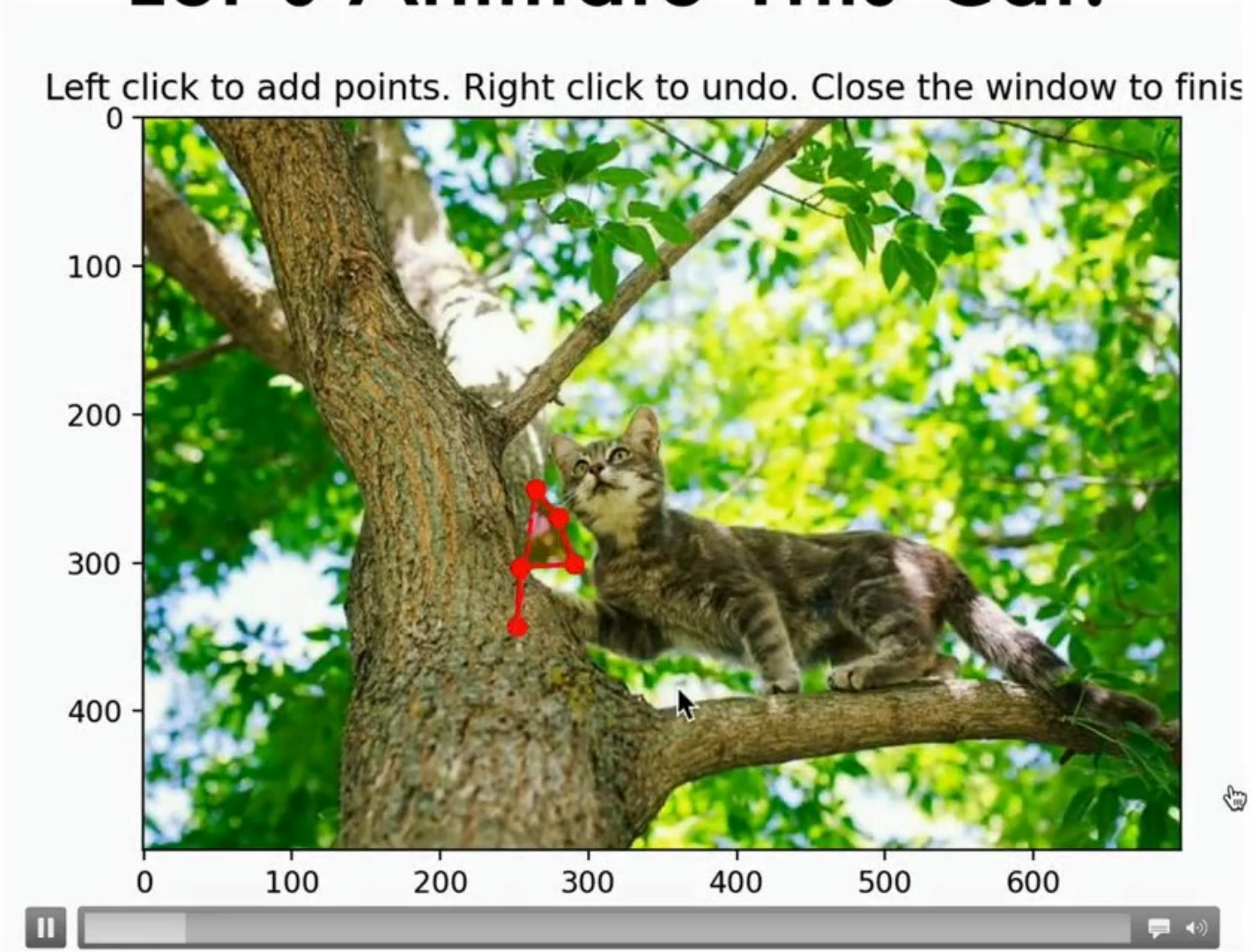
Modern Computer Vision Methods

Preliminary Meeting

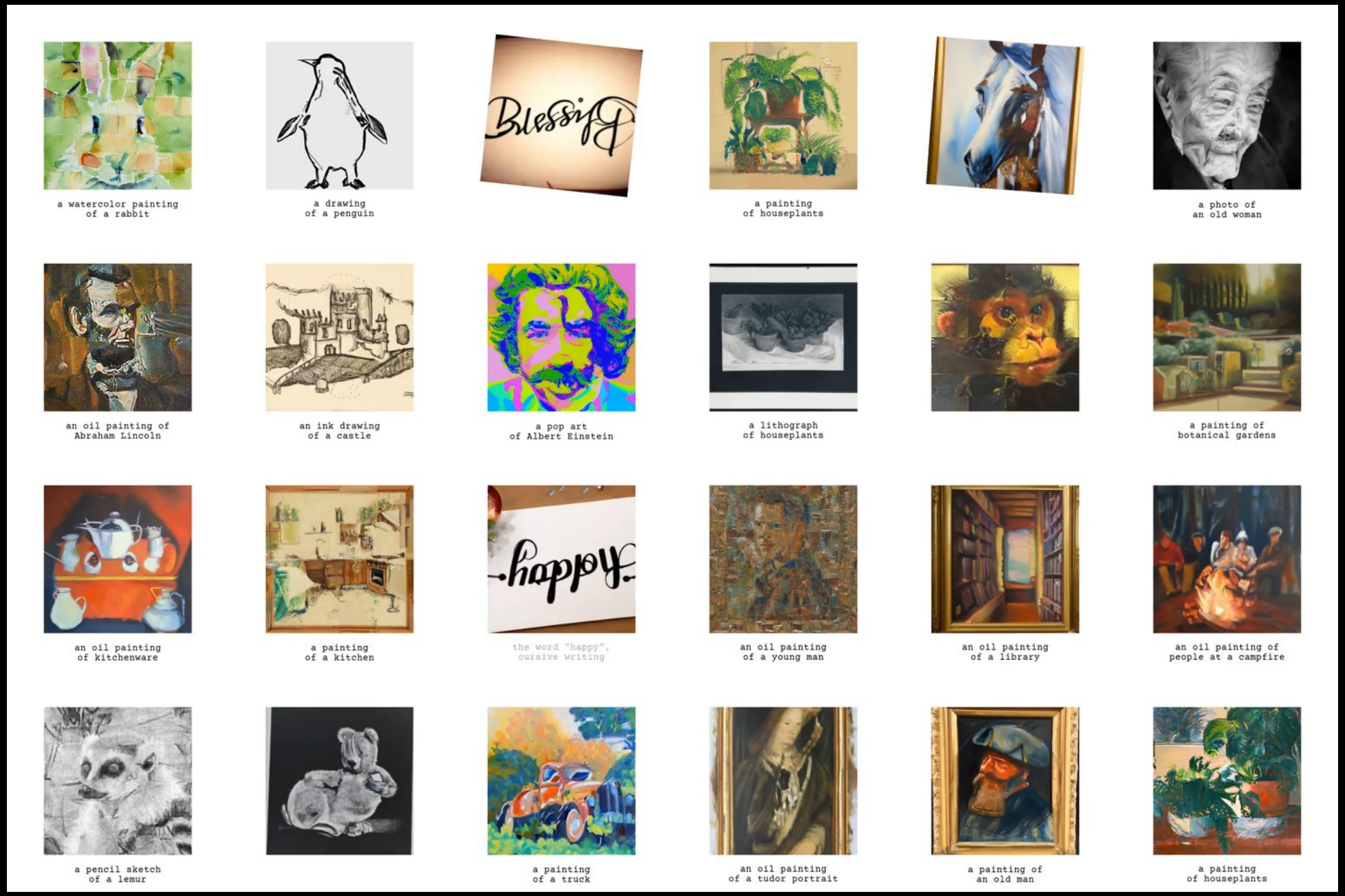
for WS 2025/26 [IN2107]

Dr. Benjamin Busam, Boody Elskhawy, Florian Eichinger, Junwen Huang, Manoj Kumar, Mert Karaoglu, Mert Kiray, Michael Greza, Qilin Zhang, Weihang Li

Let's Animate This Cat!



[Burgert, Xu, Xian, Pilarski, Clausen, He, Ma, Deng, Li, Mousavi, Ryoo. Go-with-the-flow: Motion-controllable video diffusion models using real-time warped noise. CVPR 2025]



[Daniel Geng, Inbum Park, Andrew Owens. Visual Anagrams: Generating Multi-View Optical Illusions with Diffusion Models. CVPR 2024]

Video Generation Prompt: “A fish swimming into a coffee shop and trying to order.”

[Wang, Ma, Liu, Hou, Xu, Wang, Juefei-Xu, Luo, Zhang, Hou, Vajda. LinGen: Towards High-Resolution Minute-Length Text-to-Video Generation with Linear Computational Complexity. CVPR 2025]

Goals

- Scientifically Learning about…
  - State-of-the-art Computer Vision
  - Current research challenges and applications
  - Communicate / discuss on most recent advantages with expert scientists
  - Hands-on experience with available code bases

- Skill training of…
  - Reading / understanding of a scientific work
  - Get overview of scientific field through literature research
  - Research talk in front of an audience, related Q&A

Seminar Contents

Most recent advances in Computer Vision field on

- Object Detection & Tracking
- 6D Object / Camera Pose Estimation
- Robotic Grasping / 3D Manipulation
- Generative Image / Video / Scene Synthesis
- 3D Scene Understanding / Reconstruction
- Multi-View Reconstruction
- Sensor Fusion / Multi-modal Imaging
- Universal Text & Vision Models

Presentation

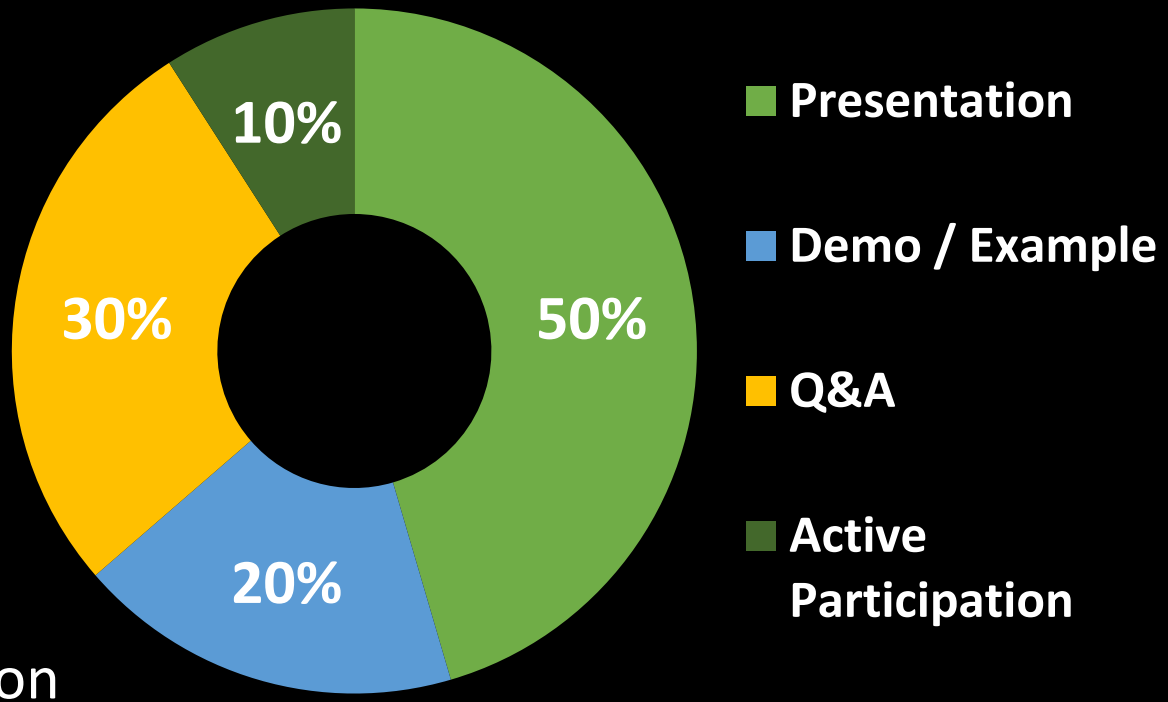
- Presentation: 20 minutes + 10-15 minutes Q&A
- Content should cover
  - Introduction / Relevance of Problem
  - Context / Related Work
  - Main Contribution(s)
  - Experimental Results
    - Hands-on experience with code
  - Discussion
  - Future Work
- Presentation should be self-contained
- Attend all talks + active participation in other discussions

Seminar Schedule

- 8 sessions + 1 intro + 1 presentation training
- 2 presentations per session
- Invited Talk(s): Renown computer vision researchers

- Topic assignment
  - Indicate preferences
  - Matching to maximize global happiness

Evaluation Criteria



- Quality of Presentation
  - Scientific Content of the Talk + Preparation
  - Quality of the Slides
  - Putting the Topic in Context (Related Work)
- Examples / Hands-on Code
- Scientific Discussion (Q&A)
- Independent Interaction / Active Participation in the Course

Some more Examples…

… with modern 3D computer vision applications

LangSplat: 3D Language Gaussian Splatting

- A 3D language field is learned by grounding CLIP language features into a set of 3D language Gaussians.

Rendered RGB Video

Visualization of Learned Language Feature[1]

[1]Different colors represent different language features.

[Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, Hanspeter Pfister. LangSplat: 3D Language Gaussian Splatting. CVPR 2024]

LangSplat: 3D Language Gaussian Splatting

- A 3D language field is learned by grounding CLIP language features into a set of 3D language Gaussians.

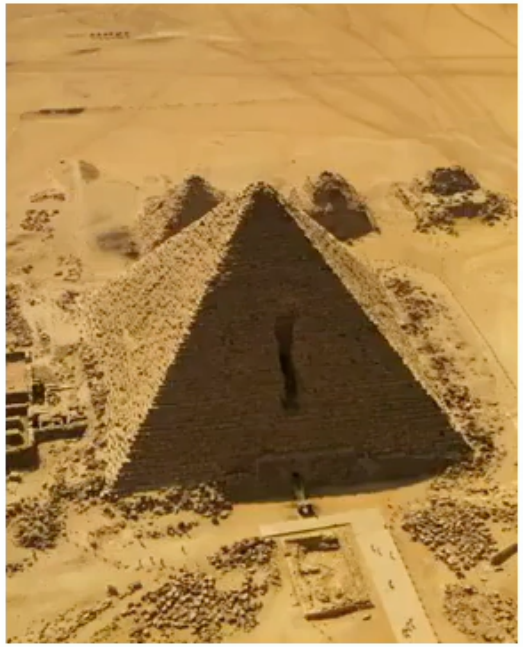
Rendered RGB Video

Visualization of Learned Language Feature[1]

[1]Different colors represent different language features.

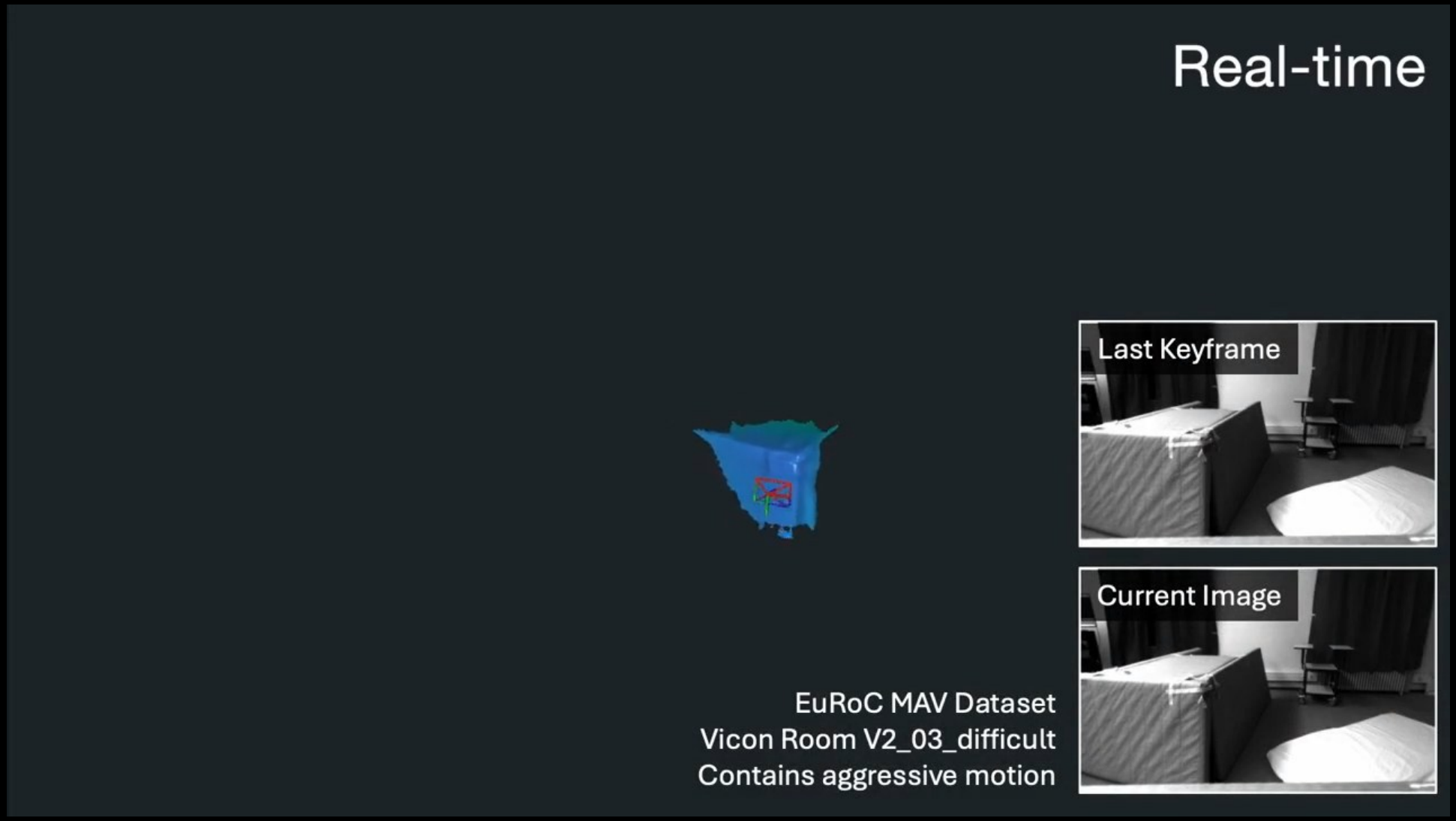
[Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, Hanspeter Pfister. LangSplat: 3D Language Gaussian Splatting. CVPR 2024]

32 Views



[Wang, Chen, Karaev, Vedaldi, Rupprecht, Novotny. VGGT: Visual Geometry Grounded Transformer. CVPR 2025]

Real-time Navigation & Reconstruction



[Murai, Dexheimer, Davison. MASt3R-SLAM: Real-time dense SLAM with 3D reconstruction priors. CVPR 2025.]

Application

2 stage process

- Register in TUM Matching System
  https://docmatching.in.tum.de/

- Apply via Form: Submit motivation + background info to increase your chances
  https://forms.gle/V3Hzu1C6fYnbUWMo6

  Include:
  - Name, E-Mail, Study Program, Semester
  - Motivation + previous experience in Computer Vision (and related field)
  - (not mandatory): Submit your latest CV + transcript of records

- Deadline: 22nd of July 2025

- 16 Students will be selected

Questions

E-Mail us on

mcvm@camp.cit.tum.de

Your MCM Team:

Dr. Benjamin Busam, Boody Elskhawy, Florian Eichinger, Junwen Huang

Manoj Kumar, Mert Karaoglu, Mert Kiray, Michael Greza, Qilin Zhang, Weihang Li

Web:

https://www.cs.cit.tum.de/camp/teaching/seminars/modern-computer-vision-methods-ws-2025-26/