Yang Miao

yangmiaogz@gmail.com • y9miao.github.io • y9miao

Educational Background

INSAIT, Researcher in Computer Vision and Robotics, 2024–now

ETH Zurich, Master in Robotics, Systems and Control, 2021–2024 **GPA: 5.83/6**

Harbin Institute of Tech., B.Eng. in Electrical Engineering 2016–2020 **GPA: 96.9/100 (Top 1/289)**

UC Berkeley, Visiting student 2019 **GPA: 4.0/4.0**

Research Interests

3D/4D Generation, Embodied AI, Robotic World Model, Digital Twin, Vision–Language Models, Scene Understanding

Research Projects as the Main Contributor

INSAIT, Institute for Computer Science, Artificial Intelligence and Technology 2024-now

Main Supervisors: Prof. Luc Van Gool, Dr. Danda Paudel

• **Ongoing Project**: *3D Generation of Fine-grained Articulated Objects for Embodied AI*, **Target: ICML 2026**

Motivation: The goal is to generate 3D articulated assets for fine-grained robotic interaction. Existing approaches to articulated object generation primarily focus on articulation prediction and part-level geometry, while the modeling of interior structures and interactions with internal objects remains largely unexplored. This limitation hinders fine-grained robot–environment interaction modeling (e.g., opening a cabinet and retrieving an object from an interior shelf). This project aims to construct a benchmark and a generation framework that produces articulated objects with detailed interior structures and contained objects aligned with textual descriptions.

Future Work: Based on the work, the plan is to build a robotic world model enabling video synthesis of fine-grained robot–environment interactions, including articulated object manipulation.

• **LangHOPS**: *Language Grounded Hierarchical Open-Vocabulary Part Segmentation*, **NeurIPS 2025** Link

Description: We proposed LangHOPS, the first MLLM-based framework for open-vocabulary object–part instance segmentation. LangHOPS can jointly segment hierarchical object and part instances from open-vocabulary categories. By grounding object–part hierarchies in language space and integrating the MLLM into the object-part parsing pipeline, LangHOPS achieves superior performance in the cross-dataset and zero-shot settings.

• **Articulate3D**: *Holistic Understanding of 3D Scenes as Universal Scene Description*, **ICCV 2025** Link

Description: We proposed Articulate3D, an expertly curated dataset in the Universal Scene Description (USD) format, with high-quality annotations of articulation on indoor scenes. We also proposed USDNet, a learning-based model together with a novel baseline capable of predicting part segmentation along with motion attributes, including motion type, articulated and interactable parts, and motion parameters.

ETH Zurich, Computer Vision and Geometry Lab 2022-2024

Main Supervisors: Prof. Marc Pollefeys, Dr. Dániel Baráth, Dr. Iro Armeni, Dr. Francis Engelmann

• **SceneGraphLoc**: *Cross-Modal Coarse Visual Localization on 3D Scene Graphs*, **ECCV 2024** Link

Description: We proposed a novel challenge of cross-modal localization of a query image within 3D scene graphs. We leveraged multiple modalities for object embedding in the scene graph and applied contrastive learning for a shared embedding space for query images and 3D scene graph objects. Our Method outperformed existing cross-modal localization methods by a large margin, and achieved competitive results with existing image-based methods but with significantly less memory usage and computation.

• **Panoptic SLAM** *with Semantic and Geometric Consistency (Master Thesis)*, **IROS 2024 Oral** Link

Description: We developed algorithms for incremental 3D panoptic mapping with RGB-D frames. The method

outperformed existing 2D-3D semantic-instance mapping methods with estimated trajectory. We also improved localization performance of ORB-SLAM3 and Voxgraph with panoptic information.

Supervision of Student Projects

• **Ongoing Student Project**: *3D Generation of Simulation-ready Articulated Objects* , **Target: NeurIPS 2026**

Student: Xiaoye Wang
Main Supervisors: **Yang Miao**, Runyi Yang
Senior Supervisors: Dr. Lei Sun, Dr. Danda Paudel

Motivation: Generating simulation-ready 3D articulated objects with accurate physical properties is a fundamental challenge with significant downstream impact in robotics. However, it remains unclear whether autoregressive or diffusion-based approaches are better suited for this task. Diffusion models (e.g., Trellis), trained on large-scale 3D datasets, excel at producing high-quality geometry but typically lack explicit modeling of physical properties such as material, density, and articulation, which are essential for simulation. Conversely, pretrained multimodal large language models (MLLMs) can infer physical properties from visual and semantic cues but are not designed for detailed 3D geometry generation. This project investigates the integration of autoregressive and diffusion-based foundation models to generate high-quality, physically grounded, and simulation-ready articulated objects.

• **SceneKey**: *Multi-Modal Embedding for Fine-Grained 3D Scene Retrieval*, **Submitted to CVPR 2026**

Student: Luka Milivojevic, Xiaoye Wang
Main Supervisors: **Yang Miao**, Sayan Deb Sarkar
Senior Supervisors: Dr. Nikola Popovic, Prof. Iro Armeni, Dr. Danda Paudel

Description: 3D scene retrieval is an important task in computer vision and graphics, enabling applications such as digital twins, interior design, and content-based 3D search. Real-world 3D environments are typically captured across diverse modalities, including RGB images, point clouds, CAD models, floor plans, and textual descriptions. We present SceneKey, a method for embedding above scene-level modality into a sequence of fine-grained feature tokens suitable for cross-modal 3D scene retrieval. We validate SceneKey on several 3D scene retrieval tasks using ScanNet and 3RScan, achieving dominant performance, particularly in fine-grained retrieval scenarios.

List of Publications

Main publications

• **Y. Miao**, JN. Zaech, X. Wang, F. Despinoy, D. Paudel, Luc Van Gool, *LangHOPS: Language Grounded Hierarchical Open-Vocabulary Part Segmentation*, **NeurIPS 2025**
• AM Halacheva*****, **Y. Miao***, JN. Zaech, X. Wang, D. Paudel, Luc Van Gool, *Articulate3D: Holistic Understanding of 3D Scenes as Universal Scene Description*, **ICCV 2025** (***** for equal contribution)
• **Y. Miao**, F. Engelmann, O. Vysotska, F. Tombari, M. Pollefeys, D. Baráth, *SceneGraphLoc: Cross-Modal Coarse Visual Localization on 3D Scene Graphs*, **ECCV 2024**
• **Y. Miao**, I. Armeni, M. Pollefeys, D. Baráth, *Volumetric Semantically Consistent 3D Panoptic Mapping*, **IROS 2024 (Oral Presentation)**
• **Y. Miao**, C. Li, Z. Li, Y. Yang, X. Yu, *A Novel Algorithm of Ship Structure Modeling and Target Identification Based on Point Cloud for Automation in Bulk Cargo Terminals*, **Measurement and Control 2021**

Other Publications

• D. Zhang, Y. Fu, R. Yang, **Y. Miao**, T. Qian, X. Zheng, G. Sun, A. Chhatkuli, X. Huang, Yu. Jiang, L. Van Gool and D. Paudel, *EgoNight: Towards Egocentric Vision Understanding at Night with a Challenging Benchmark*, arxiv 2510.06218 (submitted to ICLR 2026)

Other Responsibilities

• Co-organizer of OpenSun3D workshop
• Reviewer of IROS, IJCV, CVPR, ECCV, ICCVW