ZHANG Gengyuan

Oettingenstr. 67 EU/103, 80538, Munich, Germany

gengyuanmax@gmail.com zhang@dbs.ifi.lmu.de gengyuanmax.github.io

EDUCATION

Ludwig Maximilian University of Munich (LMU) **Oct. 2021-Present**
Ph.D. student, Computer Science *Munich, Germany*
- Advisor: Prof. Dr. Volker Tresp

Technical University of Munich (TUM) **Oct. 2018-Jul. 2021**
M.Sc., Electrical Engineering and Information Technology *Munich, Germany*
- Grade: 1.3/1.0

Zhejiang University **Hangzhou, China**
B.Eng., Opto-Electronics Information Science and Engineering *Munich, Germany*
- Final grade: 3.73/4.00

Research EXPERIENCE

Department of Informatics, LMU **Oct. 2021 - Present**
Research Assistant *Munich, Germany*
- Research on multimodal learning and video understanding
- Taking on teaching assignments
  - Tutorial: Machine Learning
  - Master Seminar: Machine Learning with Knowledge Graph, Foundation Models in AI
  - Master Practical Course: Connecting Language to Vision

Agile Robots AG **Mar. 2020 - Nov. 2020**
Internship *Munich, Germany*
- Developed an automatic hand-to-eye camera calibration pipeline
- Deployed and tested robotic object localizing and grasping project

Department of Informatics, TUM **Sept. 2019 - Feb. 2020**
Student Assistant *Munich, Germany*
- Designed and implemented perception stack and perception world model of the robotic platforms RobMoSys
- Developed computer vision components including object detection, recognition

PUBLICATIONS

Gengyuan Zhang, Jisen Ren, Jindong Gu, and Volker Tresp. Multi-event video-text retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22113–22123, 2023.

Gengyuan Zhang*, **Mang Ling Ada Fok***, Yan Xia, Yansong Tang, Daniel Cremers, Philip Torr, Volker Tresp, and Jindong Gu. Localizing Events in Videos with Multimodal Queries. *arXiv preprint arXiv:2406.10079*, June 2024.

Gengyuan Zhang, Yurui Zhang, Kerui Zhang, and Volker Tresp. Can vision-language models be a good guesser? exploring vlms for times and location reasoning. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.

Gengyuan Zhang*, Jinhe Bi*, Jindong Gu, Yanyu Chen, and Volker Tresp. Spot! revisiting video-language models for event understanding. *arXiv preprint arXiv:2311.12919*, 2023.

Jindong Gu, Zhen Han, Shuo Chen, Ahmad Beirami, Bailan He, **Gengyuan Zhang**, Ruotong Liao, Yao Qin, Volker Tresp, and Philip Torr. A systematic survey of prompt engineering on vision-language foundation models. *arXiv preprint arXiv:2307.12980*, 2023.

Yao Zhang, Haokun Chen, Ahmed Frikha, Yezi Yang, Denis Krompass, **Gengyuan Zhang**, Jindong Gu, and Volker Tresp. Cl-crossvqa: A continual learning benchmark for cross-domain visual question answering. *arXiv preprint arXiv:2211.10567*, 2022.

QUALIFICATIONS

- Languages:
  - English: proficient (IELTS 7.5)
  - German: intermediate (TestDAF 4*4)
  - Chinese: native speaker