Suozhi Huang

Undergraduate Student, China

huang-sz20@mails.tsinghua.edu.cn—suozhi.huang@gmail.com — +86-178-1022-7930 —https://hsz0403.github.io/ —

EDUCATION

Tsinghua University, Beijing, China   Sept 2020 — Jun 2025(expected)

Yao Class, Institute for Interdisciplinary Information Sciences

   Bachelor of Engineering in Computer Science and Technology (majored in Artificial Intelligence)

Cumulative GPA: 3.61/4.00

New York University, New York, USA   May 2023 — Sept 2023

Exchange visiting student

   Tandon school of Engineering

RESEARCH INTERESTS

LLM for formal math & code, LLM for reasoning.

PUBLICATIONS

Zijian Wu*, **Suozhi Huang***, Zhejian Zhou, Huaiyuan Ying, Jiayu Wang, Dahua Lin, Kai Chen. 'InternLM2.5-StepProver: Advancing Automated Theorem Proving via Expert Iteration on Large-Scale LEAN Problems', Tech report.

YuChen Gao, YuHang Yuan, **Suozhi Huang**, Nan Yao, Legeng Yu, YaoPeng Chen, Professor Qiang Zhang, Prof Xiang Chen. 'A Knowledge-Data Dual-Driven Framework for Predicting the Molecular Properties of Rechargeable Battery Electrolytes', accepted by Angewandte Chemie.

Suozhi Huang*, Juexiao Zhang*, Yiming Li, Chen Feng. 'ActFormer: Scalable Collaborative Perception via Active Queries', accepted by ICRA 2024. Webpage

ACADEMIC EXPERIENCE

Anima AI+Science Lab   Caltech, USA

Lean steps prediction analyst with external models in LeanCopilot

Advisor: Song Peiyang, Prof. Anima anandkumar   Aug 2024— Present

- Using Reprover generate searching tree dataset
- Train and validate steps preditor and progress prediction model with formal math problems' search data
- Integrate supplemental functions for LeanCopilot to provide additional predictions for a user-friendly environment

Shanghai AI Lab (PJlab)   Shanghai, China

Theorem proving(formal math) LLM training

Advisor: Jiayu Wang , Prof. Dahua Lin   Jun 2024—Oct 2024

- Built critic model for expert iteration on Lean dataset(Lean workbook & Lean github)
- Replaced Best first search method with critic model guided search in tactic selection
- Contruct a new comprehensive dataset by expert iteration searching
- Finished one SOTA paper in LLM for theorem proving

Department of Chemistry, Tsinghua.   Beijing, China

Machine Learning Study of Flash Point for Designing Highly Safe Battery Electrolytes (GNN for Molecules)

Advisor: Prof. Chen Xiang   Sept 2023 — Feb 2024

- Completed prediction of safety properties of electrolytes on pretrained molecular property prediction model Uni-Mol
- Implemented quantitative interpretation module for molecules' subgraph by applying MCTS on each's Shapley value to minimize the mutual information
- Run the baseline of electrolyte vector database, which contains billions of molecules with various predicted scientific molecular properties e.g. HOMO,LUMO
- Submitted a paper to Nature Computational Science as coauthor

Tandon School of Engineering, NYU   New York, USA

Scalable Collaborative Camera-only Perception based on Active BEV Queries (Vision Perception)

Advisor: Prof. Feng Chen, Dr. Yiming Li   Feb 2023 — Sept 2023

- Enhanced the general applicability of the collaborative autonomous driving dataset V2X-Sim by adapting it to the mmdet 3d framework

- Proposed a scalable, active paradigm in query-based camera-only collaboration task, which reduced the cost in feature transfer
- Implemented a pose-guided selection network in deformable attention in camera-based perception, which greatly reduced the number of BEV queries with effective selection
- Completed all experiments in multi-agent dataset, analyzed the results, and improved scalability
- Submitted a paper to ICRA 2024 as first author

INTERNSHIPS

Code LLM Pretraining   Beijing, China
Inf tech   April 2024 — Jun 2024

- Participated in code llm pretraining, especially data processing(deduplication) and evaluate with multiple benchmarks
- Built verified code benchmarks with models (like deepseek coder, qwen coder) and aligned the benchmarks to our own model.

AWARDS

Gold Award(Top 30) in 36th Chinese Physics Olympiad,   HangZhou, China, 2019
Freshman Scholarship at Tsinghua Univeristy   Beijing, China, 2020

SKILLS

- **Programming:** C++, Python
- **ML/AI:** Pytorch, ffmpeg, opencv, mmdet, unicore, etc.
- **Misc:** Quantum mechanics & Quantum computing