

EDUCATION

- **M.S.E, Tsinghua University, Beijing** Sep 2023 - June 2025
Research focus: Multimodal LLM, Satellite & Street View, Health & Socioeconomics
- **M.S.E, Johns Hopkins University, Baltimore** Sep 2022 - June 2023
Compressed Sensing(A), Intro to Data Science(A-), Deep Learning(A-), Data Science for BME(A)
- **B.E, Shanghai Jiao Tong University, Shanghai** Sep 2018 - June 2022
Signal Processing, Complex Analysis, Electromagnetics, Embedded Systems

RESEARCH EXPERIENCE

- **ChestLIP-LLM: Understanding Health Records and Chest X-ray Images via Multi-modal LLM** Aug 2024 -
(Ongoing Project) Beijing
 - Chest X-rays are among the most common and cost-effective medical imaging exams. However, diagnosing chest X-rays can be challenging. To address this, we propose a novel approach that utilizes contrastive pre-training (CLIP) to cluster X-ray images based on similar diagnostic labels. By incorporating chain of thought (CoT) prompts and visual instruction tuning with clinical knowledge, our method aims to enhance the accuracy of large language models (LLMs) in diagnosis, and provide more thorough, structured explanations.
- **M-CityLLM: Improving City Related Tasks via Multimodal Large Language Model** Feb 2024 -
(Ongoing Project) In Submission of ICLR, Shared First Author Beijing
 - We propose a multimodal LLM specifically pre-trained and fine-tuned for city relevant tasks such as city object detection, city scene classification, geo-location reasoning, health and socioeconomics indicators prediction. A vision-language dataset for instruct tuning is built using millions of satellite and street view images sampled from cities around the world. By satellite-streetview fusion pre-training and downstream fine-tuning, we improve city relevant tasks performance.
- **Health CLIP: Depression Rate Prediction Using Satellite and Street View Images** Sep 2023 - Feb 2024
Accepted by WWW'24 Health Day, First Author Beijing
 - As the pressure of society keeps growing, depression has become a severe problem in modern cities, therefore finding a way to estimate depression rate is of significance to relieve the problem. We adopt GPT4-vision to generate health related captions for satellite and street view images, then fine-tune CLIP model with the encapsulated image-text pairs, pushing CLIP's image encoder to extract health related features. In the end, we concatenate the features with neural network to predict the depression rate in New York City.
- **Explaining Infant and Maternal Mortality in Africa via Night Light and Satellite Imagery** May 2024 -
(Ongoing Project) Beijing
 - We try to use regression algorithm and machine learning method to explain what are the factors that affecting infant and maternal mortality rate in Africa, where only a small proportion of data are available with long interval.
- **CityBench: Evaluating the Capabilities of Large Language Model as World Model** Feb 2024 - June 2024
Under Review of NeurIPS Benchmark, Shared Second Author Beijing
 - We propose a simulator based global scale benchmark to evaluate the performance of large language model on various urban tasks, such as geospatial prediction, mobility prediction and street navigation task. I am in charge of vision part of this project, focusing on testing various open-source and closed-source multi-modal LLMs' ability on geo-location and socio-economy prediction via street view and satellite images.
- **LCSim: A Large-Scale Controllable Traffic Simulator** March 2024 - June 2024
Under Review of NeurIPS Benchmark, Second Author Beijing
 - We propose LCSim, a large-scale controllable traffic simulator. LCSim provides map tools for constructing unified HD map descriptions from either open-source datasets including Waymo and Argoverse or publicly available data sources like OpenStreetMap to scale up the simulation scenarios. Diffusion-based traffic simulation is integrated into the simulator to achieve realistic and controllable microscopic traffic flow modeling.

WORKSHOP EXPERIENCE

•LLM for Computational Social Science

In Charge of Public Health Section

Jan 2024

Beijing

- Share and summarize papers on pandemic, health prediction and HCI for public health persuasion.

•Multimodal LLM

In Charge of Evaluation and Foundation Model Section

April 2024

Beijing

- Share and summarize papers on MLLM valuation, video-LLMs and audio-LLMs.

COURSE PROJECTS

•Semantic Segmentation Analysis for Improving Automated Driving

May 2023

Keywords: Deep learning, UNet++, Deeplab V3+

- Train and evaluate several state-of-the-art segmentation methods to establish benchmarks and analyze effectiveness for Urban scene analysis and segmentation. Apply supervised trained models on real-world segmentation for videos.

•Endoscopic Instrument Segmentation For Da Vinci Robotic Surgery System

Apr 2023 - May 2023

Keywords: Deep learning, Unsupervised learning

- Train a network that takes each RGB frame as an input and predicts a pixel-wise segmentation mask that labels the target instrument type and background tissue. Use an unsupervised pre-training method and compare the performance of training on a small labeled dataset with/without pre-training.

TECHNICAL SKILLS AND INTERESTS

Coding Languages: Python, C++, Matlab

Areas of Interest: Multimodal LLM, Urban Imagery, Health and Public Health, Socioeconomics

SCHOLARSHIP AND AWARDS

•Graduate Scholarship for Academic Master Student

Tsinghua University, 2023-2025

•Undergraduate Research Outstanding Award

Shanghai Jiao Tong University, 2019-2020

•Undergraduate C-class scholarship

Shanghai Jiao Tong University, 2018-2019

ACTIVITIES

•Volunteer for COVID-19

2021

•Soccer championship and golden boots winner in Tianjin, Shanghai, Eastern U.S. and Beijing

2016, 2020, 2023