Zhikai Li

Research Interest: AI4Science, LLM, Computer Vision
zhikaili1150@outlook.com | (+86) 18978109619 | lizk1150.github.io
Tongji University, 1239 Siping Road, Shanghai, China

Education

Tongji University Sep 2020 – Jul 2024
B.Eng in Software Engineering, Machine Intelligence Track *Shanghai, China*

GPA: 89.64/100 (Rank 30%)

Award: Third-class Scholarship for Outstanding Student at Tongji University

Main Core: Machine Learning, Computer Vision, Multi-agent Reinforce Learning, SLAM, Software Engineering, Data Structure, Computer Network, Algorithm Analysis and Design, Operating System

Imperial College London Sep 2024 – Dec 2025
M.Sc in Applied Computational Science and Engineering (Expected Dec 2025) *London, UK*

Main Core: Modern programming methods, Modelling dynamical processes, Numerical methods, Applying computational science, Advanced programming, Parallel programming, Inversion and optimisation, Machine learning

Publication

[1] SaprotHub: Making Protein Modeling Accessible to All Biologists Nature Methods(Under Review)
*Jin Su, **Zhikai Li**, et al. (2024)* *Westlake University*

- Developed ColabSaprot and SaprotHub to support scientific research, allowing biologists to easily train and use Protein Language Models. SaprotHub is widely used for protein-related tasks, with wet lab experiments validating its results.
- Conducted research on model compression and parameter-efficient fine-tuning for Protein Language Model deployment, performing LoRA experiments to keep performance loss within 2% while fine-tuning less than 4% of the model's parameters.

[2] ESM-Ezy: A deep learning strategy for the mining of novel multicopper oxidaseswith superior properties Nature Communications(Under Review)
*Hui Qian, Yuxuan Wang, Xibin Zhou, Tao Gu, Hui Wang, Hao Lyu, **Zhikai Li**, et al. (2024)* *Westlake University*

- Utilized the ESM protein language model (Transformer-based) as the backbone, integrating attention and classifier layers.
- Fine-tuned specific model parameters using a cross-entropy loss function to extract high-quality, discriminative protein representations, which were then used to compute Euclidean distances for retrieval.
- Used UMAP to visualize the high-dimensional protein representations, and analyzed the model's classification performance.

Research Project

Zero-Shot Sketch-based 3D Model Retrieval based on CLIP Feb 2024 – June 2024
Undergraduate thesis at Tongji University, supervised by Prof. Shuang Liang *Tongji University, Shanghai, China*

- Implemented a baseline combining the Multi-view CNN module with CLIP's Pre-trained Image Encoder to leverage its zero-shot learning capability and cross-modal alignment ability.
- Optimized the pipeline with PEFT(LoRA, Visual Prompt) and loss function improvements(AM-Softmax, Triplet-Center Loss), leading to experiments that outperformed traditional methods.
- Developed a web interface for model usage, where users input a sketch to retrieve the most relevant 3D models.

Protein-Molecule Pair Prediction base on pre-trained LLMs Nov 2023 – Jan 2024
Developer, supervised by Prof. Fajie Yuan *Westlake University, Hangzhou, China*

- Preprocessed raw data (amino acid sequences, SMILES) by performing data cleaning, feature engineering, sampling, splitting, and data distribution analysis to construct a high-quality dataset.
- Utilized ESM (Protein Encoder) and ChemBERTa (Molecule Encoder) for representation extraction, and fine-tuned the dual-tower model with contrastive learning to enhance cross-modal alignment capability.

Activity

Tonji University's Abroad Communication Program July 2023

- Visited UPM in Madrid, KIT in Karlsruhe and SAP in Waldorf, gaining valuable insights into international cutting-edge research and industry practices.
- Explored the world-leading Computational Biology research center at UPM and actively participated in academic discussions with the laboratory researchers about their specific research domains

Skill

Programing: PyTorch Lightning, WandB, Linux, Git, Docker

Language: English(IELTS band 7(perfect score in reading), capable of reading literature, presenting, and research discussion), Cantonese(proficient), Mandarin(proficient)