Wei Liu

Curriculum Vitae

Education

2024– **University of Southern California**, *Ph.D.*
- ○ Major: Computational Biology and Bioinformatics
- ○ Awards: Graduate School Fellowship

2021–2024 **Fudan University**, *M.S.*
- ○ Major: Applied Mathematics
- ○ Research Interests: Protein Language Model, Homologous Protein Search
- ○ Awards: Outstanding Graduate of Fudan University

2017–2021 **Southeast University**, *B.Eng.*
- ○ Major: Automation (Pattern Recognition)
- ○ Grades: 88.4/100, 3.8/4.0, 9/114
- ○ Awards: Haila Scholarship (2/114)

Competition awards

2023 **Kaggle — CAFA 5 Protein Function Prediction**, *Rank 1 in the Leaderboard*

2021 **China Computer Federation Bioinformatics Conference Challenge 2**, *Rank 3*

2018 **ACM-ICPC Asian Regional Competition Nanjing Station**, *Bronze Medal*

Research experiences

2021-2024 **Graduate Student Researcher**, *Fudan University (Shanghai, China)*, Advisor: Shanfeng Zhu

Research topics

2024 **PLMSearch**, **Nature communication**, Invited to present on **WAIC 2023** (**W**orld **A**rtificial **I**ntelligence **C**onference), **RECOMB 2024**, **ISMB 2024**
- ○ <u>*Wei Liu*</u>, *Ziye Wang, Ronghui You, Chenghan Xie, Hong Wei, Yi Xiong, Jianyi Yang* and Shanfeng Zhu*. PLMSearch: Protein language model powers accurate and fast sequence search for remote homology.* [DOI], [WebServer], [GitHub]
- ○ With deep representations from a pre-trained **P**rotein **L**anguage **M**odel to predict similarity, PLMSearch can search millions of query-target protein pairs in seconds like MMseqs2 while increasing the sensitivity by more than threefold, and is comparable to state-of-the-art structure search methods with only sequences as input.

2024 **COMEBin**, **Nature communication**
- ○ *Ziye Wang, Ronghui You, Haitao Han,* <u>*Wei Liu*</u>, *Fengzhu Sun* and Shanfeng Zhu*. Effective binning of metagenomic contigs using contrastive multi-view representation learning.* [DOI], [GitHub]
- ○ Verify the two downstream tools **Resistance Gene Identifier** (RGI version 6.0.2) and **antiSMASH** (version 6.1.1) and polish the manuscript.

Research funding application

2023   **Photosynthetic Fund Phase III**, *Large-scale Protein Function Prediction*
- Parallel optimization of the existing NetGO framework to power large-scale function prediction.

2023   **Wudao Research Funding of Beijing Academy of Artificial Intelligence**, *Protein Language Model Pre-training*
- We offer a unified contrastive learning paradigm as well as a prompt-guided multi-task pre-training framework to assist SimPLM in learning from multi-modal protein similarity at the same time.

Internship experience

2024–   **Biomap**, *Beijing (Remote)*, Algorithm Intern
- Improve existing MSA (Multiple Sequence Alignment) generation methods through PLMSearch, a better homologous protein search tool, and further improve protein structure prediction methods (2024.2-)

2020–2021   **ByteDance**, *Shanghai*, Real-Time Communications Client Department
- Research and development engineer, responsible for the development of the basic components (2020.10-2021.03)
  - Reconstruction of log report module
  - Task scheduling and execution status reporting
  - Thread deadlock detection