Wei Liu

Curriculum Vitae

Education

2021–2023 **Fudan University**, *Postgraduate*
- Major: Applied Mathematics
- Research Interests: Protein Language Model, Homologous Protein Search, Protein Function Prediction
- English: TOFEL 94 (26+25+21+22)
- Major courses
  - Neural Networks and Deep Learning (A)
  - Academic English (A)
  - Computer vision (A)

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

Competition awards

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

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

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

Major research skills

Python Maching learning (PyTorch, Keras, TensorFlow); Processing and analysis (Pandas, Seaborn); Web Crawler (Request, Scrapy); Webserver (Flask); etc.

C++ ACM algorithms; Main language used during the internship @ Bytedance;

Tools Linux; Git; Docker; etc.

Research experiences

2021-2023 **Postgraduate**, *Shanfeng Zhu's lab*, Institution of Science and Technology for Brain-Inspired Intelligence (ISTBI)

Research topics

2023- **Homologous protein search**, *Under peer review of* **Nature Communications**, Presention on **WAIC 2023** (**W**orld **A**rtificial **I**ntelligence **C**onference)

- ○ We propose PLMSearch (**P**rotein **L**anguage **M**odel), a homologous protein search method with only sequences as input. With deep representations from a pre-trained protein language model to predict similarity, PLMSearch can capture the remote homology information hidden behind the sequences.
- ○ 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.
- ○ Webserver: `https://issubmission.sjtu.edu.cn/PLMSearch/`
- ○ Github: `https://github.com/maovshao/PLMSearch/`
- ○ BioRxiv: `https://doi.org/10.1101/2023.04.03.535375`

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.

2022-2023 **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.
- ○ We also employ a prompt fine-tuning module to provide downstream tasks the on-demand flexibility of utilizing respective protein similarity.
- ○ SimPLM outperforms state-of-the-art general protein language models in rich functional prediction and remote homology prediction tests.

Internship experience

2020–2021 **ByteDance**, *Shanghai*, RTC (**R**eal-**T**ime **C**ommunications) 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