Chuanrui WANG

✉ dragon.wcr12@gmail.com

Education

M.Sc. in Computer Science — **Montréal, Canada**
- **Mila / University of Montréal** — *Sep. 2022 - Jul. 2024*
- **Supervisor: Jian Tang**
- **Courses:** Machine Learning(A+), Representation Learning(A+), Structured Optimization(A+), AGI(A+)
- **Courses(Audit):** Probabilistic Graphical Models, Reinforcement Learning and Optimal Control

M.Sc. in System Engineering — **Beijing, China**
- **Beihang University, Dual Degree Program** — *Aug. 2020 - Jun. 2022*
- **Overall GPA:** 3.64/4.0

Ingénieur Généraliste — **Lille, France**
- **École Centrale de Lille, Dual Degree Program** — *Aug. 2018 - Jun. 2020*
- **Overall GPA:** 3.70/4.0

B.Sc. in Mathematics and Applied Mathematics — **Beijing, China**
- **Beihang University** — *Aug. 2015 - Jun. 2018*
- **Major Courses GPA:** 3.77/4.0, **Rank:** 2/120

Publications

PDB-Struct: A Comprehensive Benchmark for Structure-based Protein Design
- **Chuanrui Wang**, Bozitao Zhong, Zuobai Zhang, Narendra Chaudhary, Sanchit Misra, Jian Tang
- In **Neurips 2023 workshops** on AI for Drug Discovery/Machine Learning in Structural Biology.

A Systematic Study of Joint Representation Learning on Protein Sequences and Structures
- Zuobai Zhang, **Chuanrui Wang**, Minghao Xu, Vijil C., Aurelie L., Payel Das, Jian Tang
- In submission to **AAAI 2024**.

Protein Sequence and Structure Co-Design with Equivariant Translation
- Chence Shi, **Chuanrui Wang**, Jiarui Lu, Bozitao Zhong, Jian Tang
- In **ICLR 2023** (poster).

KnowReQA: A Knowledge-aware Retrieval Question Answering System
- **Chuanrui Wang**, Jun Bai, Xiaofeng Zhang, Cen Yan, Yuanxin Ouyang, Wenge Rong, Zhang Xiong
- In **KSEM 2022**.

A New Mesh Smoothing Method based on A Neural Network
- Yufei Guo, **Chuanrui Wang**, Zhe Ma, Xuhui Huang, Kewu Sun, Rongli Zhao
- In **Comput Mech 2021 (Journal)**.

Colossal-AI: A Unified Deep Learning System for Large-Scale Parallel Training
- Zhengda Bian, Hongxin Liu, Boxiang Wang, Haichen Huang, Yongbin Li, **Chuanrui Wang**, Fan Cui, Yang You
- In **arXiv preprint**. (29k+ stars on GitHub)

Current Interests

- My current research mainly focuses on **AI for science**, especially in the field of **drug discovery**, **protein structure prediction** and **protein design**.
- I specialize in **generative models** and **geometric deep learning**, exploring innovative methods within these areas.
- Prior to joining Mila, my work encompassed NLP, particularly in question answering systems and knowledge graphs, as well as developing large-scale model training systems.

Research Experiences

Research in Generative Models and Protein Design

Advisor: Jian Tang, Associate Professor, Mila *Jun. 2022 - Present*

- Developed a novel protein sequence-structure co-design method with rapid one-shot inference, overcoming the slow inference issues of autoregressive and diffusion generative models. We demonstrated its high sampling quality and efficiency, alongside flexible conditioning in tasks such as antibody CDR design and general protein design. The work was accepted at ICLR 2023, where I also presented a poster session.
- Improved the evaluation of protein sequence design models by creating two innovative metrics: refoldability, which uses structure prediction models as proxies for wet-lab experiments, and stability, to evaluate the likelihood landscapes of generated proteins. Furthermore, we introduced the PDB-Struct benchmark with carefully selected datasets to facilitate the comparison of recent protein design methods, thereby aiding protein scientists in model selection. This work was accepted by NeurIPS 2023 workshops focusing on AI for Drug Discovery and for Structural Biology.
- In progress: Designing a diffusion-based generative model that simultaneously designs sequences and structures, which differs from previous models that operate in a sequential manner. The challenge lies in designing a diffusion process that operates within combination of discrete space and continuous $SE(3)$ space.

Research in Protein Representation Learning and Protein Folding

Advisor: Jian Tang, Associate Professor, Mila *Sep. 2022 - May. 2023*

- Investigated protein representation learning by integrating sequence and structural data through a state-of-the-art protein language model and various structural encoders.
- Proposed three fusion strategies and six pre-training objectives for constructing and training such joint learning model. The enhanced model significantly outperformed sequence-based and structure-based models in function prediction tasks. These findings are detailed in a paper under review for AAAI 2024.
- Improved protein folding models using pseudo-data backtranslated from an inverse protein folding model, aiming to overcome the scarcity of labeled sequence-structure pair data. Experiments yielded enhanced results with antibody datasets; however, we have yet to replicate this success with general protein datasets, possibly constrained by the availability of computational resources.

Large-Scale Model Training System with Efficient Parallelization Techniques

Advisor: Yang You, Professor, National University of Singapore *Aug. 2021 - Feb. 2022*

- Developed Colossal-AI, a distributed machine learning framework. The framework enabled hybrid parallel training with minimum pytorch code modification, enhancing the scalability of AI model training.
- Realized the 1d tensor parallel architecture, conducted scaling experiments, and achieved efficient pre-training of a VIT-base/32 model in 0.6 hours, showcasing the system's effectiveness.

Research in Natural Language Processing and Pre-trained Language Models

Advisor: Wenge Rong, Professor, Beihang University *Nov. 2020 - Aug. 2021*

- Proposed a Hierarchical Lexicon Embedding Structure to better solve Chinese named entity recognition (NER) task and outperformed baselines in three commonly used Chinese NER dataset (MSRA, Resume and Weibo).
- Proposed a Question Answer System that fuses Knowledge Graph information using an attention-based filter mechanism, improving detail and context in natural language responses.

Straightforward Neural Network Method for Mesh Smoothing

Advisor: Yufei Guo, PhD Researcher, X Lab *Sep. 2020 - Nov. 2020*

- Pioneered a deep learning approach for mesh smoothing that imitates traditional optimization-based methods, enabling real-time parameter adjustments.
- As the first deep learning approach for mesh smoothing problem, the model achieved a fourfold increase in speed compared to optimization-based methods, maintaining comparable mesh quality.

Skills

- **Computer Skills:** Python, Pytorch, C, SQL, Linux, Matlab, Office
- **English Skills: IELTS:** 7(L8.5, R8, W6.5, S6); **GRE:** V156, Q169, W3.0
- **Other Languages: Mandarin:** Native; **French:** DELF B2