Daize Dong

Personal Page | Google Scholar | dzdong2019@gmail.com

Research Interest

My research spans domains such as ML, NLP, and CV, and I have a strong passion for uncovering the intrinsic properties of neural networks with theoretical guarantees. My primary research interests include, but are not limited to:

1. **Representation Learning:** Enhancing abstract data representations to improve the model ability and prevent degradation.
2. **Model Architecture:** Discovering general structures to enhance model efficiency and achieve mathematical completeness.
3. **AI for Biology:** Leveraging AI to advance the scientific progress of human beings.

Education

University of Electronic Science and Technology of China *Sep. 2019 – Jul. 2023*
Bachelor of Computer Science & Mathematics and Applied Mathematics *GPA: 3.91/4.00*

Working Experience

Zhipu AI – *Junior Researcher* *Sep. 2024 – Present*
Mixture of Experts, Large-scale Multimodal & Language Model Pretraining

Research Experience

OpenGVLab, Shanghai Artificial Intelligence Laboratory – *Research Assistant* *Jul. 2023 – Aug. 2024*
Mixture of Experts, Large Language Models
Instructor: Dr. Xiaoye Qu. Supervisor: Prof. Yu Cheng

- Explored the pipeline for efficiently constructing large language models with the Mixture of Experts (MoE) structure.
- Further conducted research on enhancing the representation and structure of other conditional & dynamic networks.

Center for Artificial Intelligence Research and Innovation, Westlake University – *Research Assistant* *Apr. 2023 – Aug. 2024*
Graph Transformers, Molecule & Protein Generation, AI for Biology
Collaborator: Zhangyang Gao. Supervisor: Prof. Stan Z. Li

- Explored frontier graph networks and their applications for molecule and protein representation & generation.
- Conducted research on a self-supervised pretraining framework for modeling graph data using the pure transformer.

Data Intelligence Group, University of Electronic Science and Technology of China – *Research Intern* *Jul. 2022 – Mar. 2023*
Domain Adaptation, Transfer Learning, Computer Vision
Instructor: Prof. Wen Li

- Explored the theories and algorithms under unsupervised & self-supervised paradigms in transfer learning.
- Conducted research on enhancing the latent representation for domain adaptation through contrastive learning.

NLP Group, JD Explore Academy – *Independent Researcher* *Feb. 2022 – Oct. 2022*
Model Compression, Natural Language Processing
Collaborator: Shwai He. Instructor: Dr. Liang Ding. Supervisor: Prof. Dacheng Tao

- Explored parameter-efficient strategies for downstream fine-tuning and model compression methods.
- Conducted research on enhancing parameter efficiency for dynamic networks and adapters.

Projects

LLaMA-MoE: Building Mixture-of-Experts from LLaMA with Continual Pre-training. [Paper] [Code] *Jul. 2023 – Dec. 2023*
The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

- Worked as the core member for designing methods to convert large language models into Mixture of Experts (MoE).
- Explored and designed multiple methods to initialize MoE using parameters from pretrained dense models.
- Proposed a simple yet effective output-scaling strategy to recover model performance at initialization.

Publications

1. **A Graph is Worth K Words: Euclideanizing Graph using Pure Transformer.** [Paper]
   Zhangyang Gao*, **Daize Dong***, Cheng Tan, Jun Xia, Bozhen Hu, Stan Z. Li.
   The 41st International Conference on Machine Learning (ICML 2024).
2. **iDAT: inverse Distillation Adapter-Tuning.** [Paper]
   Jiacheng Ruan, Jingsheng Gao, Mingye Xie, **Daize Dong**, Suncheng Xiang, Ting Liu, Yuzhuo Fu.
   2024 IEEE International Conference on Multimedia and Expo (ICME 2024). ***(Oral)***
3. **PAD-Net: An Efficient Framework for Dynamic Networks.** [Paper]
   Shwai He, Liang Ding, **Daize Dong**, Boan Liu, Fuqiang Yu, Dacheng Tao.
   Proceedings of The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).
4. **SparseAdapter: An Easy Approach for Improving the Parameter-Efficiency of Adapters.** [Paper]
   Shwai He, Liang Ding, **Daize Dong**, Miao Zhang, Dacheng Tao.
   Findings of The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).
5. **SD-Conv: Towards the Parameter-Efficiency of Dynamic Convolution.** [Paper]
   Shwai He, Chenbo Jiang, **Daize Dong**, Liang Ding.
   IEEE/CVF Winter Conference on Applications of Computer Vision, 2023 (WACV 2023).

Preprints

1. **Demystifying the Compression of Mixture-of-Experts Through a Unified Framework.** [Paper] [Code]
   Shwai He*, **Daize Dong***, Liang Ding, Ang Li.
   Under Review at The Thirteenth International Conference on Learning Representations (ICLR 2025).
2. **Dynamic Data Mixing Maximizes Instruction Tuning for Mixture-of-Experts.** [Paper] [Code]
   Tong Zhu, **Daize Dong**, Xiaoye Qu, Jiacheng Ruan, Wenliang Chen, Yu Cheng.
   Under Review at 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025).
3. **DLO: Dynamic Layer Operation for Efficient Vertical Scaling of LLMs.** [Paper] [Code]
   Zhen Tan*, **Daize Dong***, Xinyu Zhao, Jie Peng, Yu Cheng, Tianlong Chen.
   Under Review at 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025).
4. **ExFusion: Efficient Transformer Training via Multi-Experts Fusion.**
   Jiacheng Ruan, **Daize Dong**, Xiaoye Qu, Tong Zhu, Ting Liu, Yuzhuo Fu, Yu Cheng.
   Under Review at IEEE Transactions on Circuits and Systems for Video Technology (IEEE-TCSVT).
5. **Blending and Aggregating the Target for Blended-Target Domain Adaptation.**
   Tong Chu*, **Daize Dong***, Jinhong Deng, Lixin Duan, Wen Li.
   Under Review at IEEE Transactions on Image Processing (IEEE-TIP).
6. **LLaMA-MoE v2: Exploring Sparsity of LLaMA from Perspective of Mixture-of-Experts with Post-Training.** [Paper] [Code]
   Xiaoye Qu, **Daize Dong**, Xuyang Hu, Tong Zhu, Weigao Sun, Yu Cheng.
   ArXiv Preprints.

Honors and awards

Excellent Student Scholarship *2020 – 2021*
University of Electronic Science and Technology of China

The Second Prize Scholarship *2019 – 2020*
University of Electronic Science and Technology of China

Relevant Courses

Deep Learning: Machine Learning, Artificial Intelligence, Deep Learning for Computer Vision, Deep Learning for Natural Language Processing, Knowledge Representation and Reasoning, Data Mining and Big Data Analysis.
Optimization & Algorithm: Optimization Theory and Methods, Introduction to Algorithms.
Mathematics: Differential Calculus, Linear Algebra, Probability Theory, Stochastic Process, Discrete Mathematics, Graph Theory, Multivariate Statistical Analysis, Causal Inference.
Computer Science: Computer Organization and Architecture, Compiler Principles, Computer Operating Systems, Database Principles and Applications, Information Retrieval, Software Engineering.

* Equal Contribution