

Fuzhao XUE

+65 83434586 | xuefuzhao@outlook.com | <https://xuefuzhao.github.io> | Singapore

EDUCATION

Doctor of Philosophy (Ph.D.)

Jul. 2025 (expected)
Singapore

- National University of Singapore
- Majored in Computer Science
- Supervisor: Yang You
- President's Graduate Fellowship
- Google Ph.D. Fellowship
- AAAI 2023 Distinguished Paper Award (Co-author)
- Published papers at ICML, NeurIPS, ACL, AAAI, CVPR
- Internship at Google Brain and worked with Yi Tay and Mostafa Dehghani
- Internship at NVIDIA AI and worked with Linxi (Jim) Fan and Yuke Zhu

Master of Engineering (M.Eng.)

Jul. 2021
Singapore

- Nanyang Technological University
- Majored in Computer Science and Engineering
- Supervisor: Eng-Siong Chng and Aixin Sun
- Thesis: Refining latent multi-view graph for relation extraction
- GPA: 5.0/5.0
- Published paper at AAAI, ICML and ICASSP.

RESEARCH INTEREST

- **Machine Learning:** Efficient Foundation Model, Conditional Computation, Adaptive Computation
- **Natural Language Processing:** Large Language Model Pre-training and Instruction Tuning
- **High Performance Computing:** Model Parallelism and Large-scale Deep Learning System for Long Sequence.

INDUSTRY RESEARCH EXPERIENCE

NVIDIA AI | Work Remotely from Singapore *Research Intern*

Mar. 2023-Dec. 2023

In charge of long range foundation model for In-context Reinforcement Learning Foundation model. Working with Linxi(Jim) Fan and Yuke Zhu.

- Implementing algorithm distillation to enable RL Foundation Model to handle unseen tasks via in-context learning.
- Designing long range foundation model to speedup the RL foundation model training and inference.

Google Brain | Work Remotely from Singapore *Student Researcher*

Jul. 2022-Nov 2022

In charge of adaptive computation with adaptive model depth and elastic input sequence. Worked with Yi Tay and Mostafa Dehghani. Part of the work is accepted at ICML 2023.

- Implemented universal transformer and PonderNet on vision transformer based on Scenic and made vision universal transformer open source.
- Proposed AdaTape to achieve adaptive computation with elastic input sequence, which is a novel strategy that enables dynamic computation in neural networks via adaptive tape token. (ICML 2023)

ACADEMIC RESEARCH EXPERIENCE

National University of Singapore | Singapore *Ph.D. Candidate*

Jul. 2021-Present

In charge of efficient and effective general training framework for both computer vision and natural language processing to scale transformer. Published at ICML 2023, ACL 2023, AAAI 2022.

- Studied the relationship between transformer configuration and training objectives. (ICML 2023)
- Proposed sequence parallelism to train transformer with longer sequence from system perspective. (ACL 2023)
- Proposed to go wider instead of deeper, compressing along depth by parameter sharing and scaling along width by mixture-of-experts to construct a parameter-efficient framework. (AAAI 2022)

Nanyang Technological University | Singapore *M.Eng. Student*

Jul. 2020-Jun. 2021

In charge of design of improving dialogue-level relation extraction by adaptive graph pooling proposed. Published paper at AAAI 2021, ICASSP 2022.

- Identified indicative words using Dynamic Time Wrapping Pooling in an unsupervised manner with high accuracy.
- Improved the state-of-the-art of dialogue-level relation extraction by 6% on DialogRE.

National University of Singapore | Singapore *Research Intern*

Oct. 2019-Apr. 2020

In charge of designing hybrid speech recognition framework with Pytorch and Kaldi. Published paper at ICML 2020.

- Used Pytorch underlying operations to implement acoustic models such as DNN, LSTM, SRU, RRN and RTN.
- Modeled relational thinking via Deep Graph Process, reducing WER by 10% relatively on Chime2&5 and SWBD.

SKILLS

- **Programming:** Python, C/C++, C#, Matlab
- **Frameworks:** JAX, TensorFlow, PyTorch, Keras, Theano, Scikit-learn, pandas
- **Advanced:** T5x, Flaxformer, Scenic, SeqIO, Deepspeed, Megatron, Huggingface

SERVICES

- **Conference 2024:** ICLR Reviewer
- **Conference 2023:** NeurIPS Reviewer, ACL PC Member, CVPR Reviewer
- **Conference 2022:** ICML Reviewer, SIGIR PC Member, EMNLP Reviewer
- **Journal:** TKDE Reviewer

PUBLICATION

2023

- **Fuzhao Xue**, Yao Fu, Wangchunshu Zhou, Zangwei Zheng, Yang You “*To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis*” Published at *Neural Information Processing Systems (NeurIPS) 2023*
- **Fuzhao Xue**, Jianghai Chen, Aixin Sun, Xiaozhe Ren, Zangwei Zheng, Xiaoxin He, Yongming Chen, Xin Jiang, Yang You “*Adaptive Computation with Elastic Input Sequence*” Published at International Conference on Machine Learning (**ICML**) **2023**
- **Fuzhao Xue**, Valerii Likhoshesterov, Anurag Arnab, Neil Houlsby, Mostafa Dehghani, Yang You “*A Study on Transformer Configuration and Training Objective*” Published at International Conference on Machine Learning (**ICML**) **2023**
- Shenggui Li*, **Fuzhao Xue***, Yongbin Li, Yang You “*Sequence Parallelism: Long Sequence Training from System Perspective*” Published at Annual Meeting of the Association for Computational Linguistics (**ACL**) **2023 (equal contribution)**
- Zangwei Zheng, Pengtai Xu, Xuan Zou, Da Tang, Zhen Li, Chenguang Xi, Peng Wu, Leqi Zou, Yijie Zhu, Ming Chen, Xiangzhuo Ding, **Fuzhao Xue**, Ziheng Qing, Youlong Cheng, Yang You “*CowClip: Reducing CTR Prediction Model Training Time from 12 hours to 10 minutes on 1 GPU*” Published at Association for the Advancement of Artificial Intelligence (**AAAI**) **2023 Distinguished Paper Award**
- Zangwei Zheng, Xiaozhe Ren, **Fuzhao Xue**, Yang Luo, Xin Jiang, Yang You “*Response Length Perception and Sequence Scheduling: An LLM-Empowered LLM Inference Pipeline*” *arXiv preprint arXiv:2305.13144*
- Xiao Liu, Jian Zhang, Heng Zhang, **Fuzhao Xue**, Yang You “*Hierarchical Dialogue Understanding with Special Tokens and Turn-level Attention*” Published at International Conference on Learning Representations (**ICLR**) **2023 Tiny Paper**

2022

- **Fuzhao Xue**, Ziji Shi, Futao Wei, Yong Liu, Yang You “*Go Wider Instead of Deeper*” Published at Association for the Advancement of Artificial Intelligence (**AAAI**) **2022**
- **Fuzhao Xue**, Aixin Sun, Hao Zhang, Eng-Siong Chng “*An Embarrassingly Simple Model for Dialogue Relation Extraction*” Published at International Conference on Acoustics, Speech, & Signal Processing (**ICASSP**) **2022**
- Andrew Koh, **Fuzhao Xue**, Eng Siong Chng “*Automated Audio Captioning using Transfer Learning and Reconstruction Latent Space Similarity Regularization*” Published at International Conference on Acoustics, Speech, & Signal Processing (**ICASSP**) **2022**
- Jinjie Ni, Tom Young, Vlad Pandealea, **Fuzhao Xue**, Vinay Adiga, Erik Cambria “*Recent Advances in Deep Learning-based Dialogue Systems*” Published at **Artificial Intelligence Review**
- Wangbo Zhao, Kai Wang, Xiangxiang Chu, **Fuzhao Xue**, Xinchao Wang, and Yang You “*Modeling Motion with Multi-Modal Features for Text-Based Video Segmentation*” Published at IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), **2022**
- **Fuzhao Xue**, Xiaoxin He, Xiaozhe Ren, Yuxuan Lou, Yang You “*One Student Knows All Experts Know: From Sparse to Dense*” *arXiv preprint arXiv:2201.10890*

2021

- **Fuzhao Xue**, Aixin Sun, Hao Zhang, Eng-Siong Chng “*GDPNet: Refining Refining Latent Multi-View Graph for Relation Extraction*” Published at Association for the Advancement of Artificial Intelligence (**AAAI**) **2021**
- Xiaoxin He, **Fuzhao Xue**, Xiaozhe Ren, Yang You “*Large-Scale Deep Learning Optimizations: A Comprehensive Survey*” *arXiv preprint arXiv:2111.00856*
- Yuxuan Lou, **Fuzhao Xue**, Zangwei Zheng, Yang You “*Cross-token Modeling with Conditional Computation*” *arXiv preprint arXiv:2109.02008*
- Xingmei Wang, Jiayang Meng, Bin Wen, **Fuzhao Xue** “*RACP: A network with Attention Corrected Prototype for Few-shot Speaker Recognition using Indefinite Distance Metric*” Published at **Neurocomputing**

2020

- Hengguan Huang, **Fuzhao Xue**, Hao Wang, Ye Wang * "*Deep Graph Random Process for Relational-Thinking-Based Speech Recognition*" Published at International Conference on Machine Learning (**ICML**) **2020**
- Xingmei Wang, **Fuzhao Xue**, Wei Wang, Anhua Liu. "*A network model of speaker identification with new feature extraction methods and asymmetric BLSTM*" Published at **Neurocomputing**

2019

- Xingmei Wang, Anhua Liu, Yu Zhang, **Fuzhao Xue**. "*An Underwater Acoustic Target Recognition: A Combination of Multi-dimensional Fusion Features and Modified Deep Neural Network*" Published at **Remote Sensing**.