Jiajun Zhu

3200106048@zju.edu.cn · (+86) 173-009-89120 · zhuconv.github.io

Education

The University of Texas at Austin *Sept. 2025 (Expected) - Unknown*

Incoming Ph.D. in Electrical and Computer Engineering

Zhejiang University *Sept. 2020 - Jun. 2025 (Expected)*

B.S. in Mathematics

- Major GPA: 3.90/4.00 (top 1/64 in sophomore year)
- Math Average: 91.3
- Core Courses:
  - Ordinary Differential Equation: 95
  - Mathematical Analysis II/III: 93
  - Real Variable Analysis: 93
  - Mathematical Software: 99
  - Mathematical Statistics: 92
  - Point Set Topology: 93
  - Advanced Algebra II: 95
  - Abstract Algebra: 90
  - Geometry: 90

Publication

1. **Jiajun Zhu**, Yuehao Wang, Peihao Wang, Ruisi Cai, Siqi Miao, Liliang Ren, Pan Li, Zhangyang Wang. Interpretable and Robust State Space Models with Stochastic Selection, *The Forty-second International Conference on Machine Learning (ICML)*, 2025. Submitted.
2. **Jiajun Zhu**, Peihao Wang, Ruisi Cai, Jason D. Lee, Pan Li, Zhangyang Wang. Rethinking Addressing in Language Models via Contextualized Equivariant Positional Encoding, *The Forty-second International Conference on Machine Learning (ICML)*, 2025. Submitted.
3. **Jiajun Zhu**, Siqi Miao, Rex Ying, Pan Li. Towards Understanding Sensitive and Decisive Patterns in Explainable AI: A Case Study of Model Interpretation in Geometric Deep Learning, *Nature Machine Intelligence*, 2025.
4. Peihao Wang, Ruisi Cai, Yuehao Wang, **Jiajun Zhu**, Pragya Srivastava, Zhangyang Wang, Pan Li, Understanding Bottlenecks of State Space Models through the Lens of Recency and Over-smoothing, *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.
5. Yifei Sun, Qi Zhu, Yang Yang, Chunping Wang, Tianyu Fan, **Jiajun Zhu**, Lei Chen. Fine-tuning Graph Neural Networks by Preserving Graph Generative Patterns, *The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

Selected Research

Interpretable State-Space Model with Stochastic Selective Mechanism *Oct. 2024 - Feb. 2025*

Supervised by Prof. Zhangyang Wang *University of Texas at Austin*

- Proposed injecting noise based on the information bottleneck (IB) principle to enhance the interpretability and robustness of the selective state-space model against attacks.
- Implemented SSIB with two types of stochastic models within the S6 framework: Gaussian-SSIB and Bernoulli-SSIB.
- Demonstrated that both SSIB variants offer inherent interpretability, increased robustness to corrupted input, and improved generation performance on standard language benchmarks.

Enhanced Transformer with Contexualized Equivariant Position Encoding *Jun. 2024 - Sept. 2024*

Supervised by Prof. Zhangyang Wang *University of Texas at Austin*

- Introduced TAPE, a novel framework for dynamically layer-updated positional embeddings in transformers, adapting to content and surpassing the limitations of fixed long-term decay in traditional positional embeddings.
- Proposed the principles of permutation invariance and orthogonal equivariance to enhance the generalization of positional embeddings, and designed an enhanced Transformer with modules that integrate positional information into both the attention and feedforward layers in alignment with these principles.
- Demonstrated that TAPE excels in language modeling and downstream tasks such as arithmetic reasoning and long-context retrieval, achieving strong performance in both pretraining from scratch and parameter-efficient fine-tuning.

Interpretability of Geometric Deep Learning for Scientific Tasks *Jan. 2023 - Feb. 2024*

Supervised by Prof. Pan Li *Georgia Institute of Technology*

- Proposed the definition of two critical concepts in the domain of interpretability: sensitive patterns and decisive patterns, highlighting their misalignment, an aspect previously overlooked by researchers.

- Adapted 12 interpretability techniques from graph neural networks to geometric deep learning models, which are widely employed in scientific tasks, and benchmarked their performance.
- Derived key insights from empirical evidence to guide the effective and appropriate application of two major categories of interpretability techniques: post-hoc methods and self-interpretable methods.

Ongoing Research

Scaling Latent Reasoner to Infinite Recurrent Depth — *Feb. 2025 - Now*

Supervised by Prof. Zhangyang Wang — *University of Texas at Austin*

- Proposed improving optimization for recurrent LLMs by replacing truncated backpropagation with implicit gradient methods, ensuring constant computation and memory as depth scales.

Full Experience

LLM-based Agent for Automatic Cell Type Annotations — *May. 2024 - Dec. 2024*

Supervised by Prof. Zhiting Hu — *University of California San Diego*

- Proposed a framework empowering LLM-based agents to generate hypotheses, conduct experiments, perform evaluations, and iteratively refine hypotheses based on evaluation outcomes.
- Validated the approach through benchmarking on cell type annotation datasets and conducted a user study utilizing a streamlined user interface designed for scientists.

Interactive Reasoning of Visual Language Models — *Dec. 2023 - Apr. 2024*

Supervised by Prof. Yaochu Jin — *Westlake University*

- Proposed a paradigm enabling interaction with visual language models (VLMs) through visual referencing, specifically utilizing "click and segment" actions to improve interactivity and reference accuracy.
- Enhanced the reasoning capabilities of VLMs for image segmentation by fine-tuning them on a custom-built multimodal dataset.

Mitigating Structural Divergence in Fine-tuning Graph Neural Networks — *Aug. 2022 - Oct. 2022*

Supervised by Prof. Yang Yang — *Zhejiang University*

- Provided a theoretical analysis using Taylor decomposition to guide the design of our method, breaking down the module output from a complete matrix into a linear combination of coefficients and bases.
- Validated the effectiveness of our method by implementing five baseline approaches and benchmarking performance across eight molecular datasets.

Graph Neural Networks for Electronic Property Prediction — *Jan. 2022 - Jun. 2022*

Supervised by Prof. Renjun Xu — *Zhejiang University*

- Collected 100k+ data of electronic density and band structure from material database.
- Adapted CGCNN and MEGNet models for accurate electronic property prediction.

Additional Information

- Programming languages: Python, C/C++, CUDA.
- Software & Frameworks: LaTeX, Git, PyTorch, PyTorch Geometric, Transformers, Triton, Equinox.
- Interests: Guitar (performed at the school's New Year party), Skiing, Basketball, Snooker.