Guanzhe Hong

+1 (765) 491-5569
hong288@purdue.edu
www.linkedin.com/in/guanzhe-hong-551349136

Education

Purdue University Aug 2018 – Present (Anticipated: 2025 July)

School of Electrical and Computer Engineering

Doctor of Philosophy

University of Toronto June 2018

Department of Mathematics

Faculty of Arts and Science

Mathematics Major

University of Toronto June 2017

Department of Electrical and Computer Engineering

Faculty of Applied Science & Engineering

Bachelor of Applied Science (B.ASc)

Research Interests and Methodology

My research focuses on the Foundations of Deep Learning, relying on tools from computer science, signal processing, machine learning, and statistics.

I typically follow a "physics"-like procedure in my research: perform controlled experiments, formulate hypotheses, build theories that reflect the real situation and the hypotheses, and make practical recommendations based on the theoretical and empirical observations.

My current interests revolve around developing a deeper mechanistic understanding of large language models, and relying on such insights to improve their generalization performance, efficiency and intepretability. For instance, my recent works revolve around the study of reasoning circuits in LLMs. In the past, I developed mathematical theories for how feature-based knowledge distillation affects the generalization performance of the student neural networks, and how the pretraining label granularity influences deep neural network generalization in the transfer learning setting.

Experience

Student Researcher June 2024 - Present

Google Research *Google*

- Host: Dr. Rina Panigrahy
- Researching how large language models reason

Student Researcher Oct 2022 - March 2023

Google Research *Google*

- Hosts: Dr. Yin Cui, Dr. Enming Luo
- Researched the influence of pretraining label granularity on deep neural network generalization

Research Assistant May 2020 - Present

Intelligent Imaging Lab *Purdue University*

- Advisor: Prof. Stanley Chan
- Advisory Committee: Prof. Xiaojun Lin, David Inouye, Greg Buzzard

Teaching Assistant 2018 Sept. - 2019 May

ECE595 Machine Learning I *Purdue University*

- Aimed at ECE graduate students
- Developing assignments and serving as TA for the course

Teaching Assistant 2019 Sept. - Dec.

ECE595 Machine Learning II *Purdue University*

- Developing assignments for the course

Projects

How transformers solve propositional logic problems: a mechanistic analysis

- We study the hidden mechanisms behind how tiny transformers and large language models solve simple in-context propositional logic problems.
    - For *tiny transformers* trained only on the logic problems, we are able to identify certain "planning" and "reasoning" circuits in the network that necessitate cooperation between the attention blocks to implement the desired logic.
    - We study how *pretrained LLMs*, namely *Mistral-7B* and *Gemma-2-9B*, solve the logic problems. We characterize their *reasoning circuits* through causal intervention experiments, providing necessity and sufficiency evidence for the circuits. We found that the two LLMs' latent reasoning strategies are surprisingly *similar*, and *human-like*.
    - This work is, to our knowledge, the first to characterize the circuit employed by LLMs in the wild for solving non-trivial in-context logic problems that require *latent multi-hop reasoning*.
    - Submitted to ICLR. Preprint: https://arxiv.org/abs/2411.04105

How pretraining label granularity influences generalization

- We theoretically study how the granularity of pretraining labels affects the *generalization* of deep neural networks in image classification tasks.
    - Core research question: Pretraining deep neural networks (DNNs) at a high label granularity is a common practice in deep learning, as it is believed to benefit generalization. Is it really so, and if so, why?
    - Empirical confirmation: Pretraining on the leaf labels of *ImageNet21k* produces better transfer results on ImageNet1k than pretraining on other coarser granularity levels, which supports the common practice used in the community. Experiments on *iNaturalist 2021* yield similar results.
    - Theoretical results: We prove that pretraining at reasonably large label granularities tend to benefit generalization, by "simulating" the whole pipeline of how a neural network learns.
        - We define a *hierarchical multi-view* property of the data distribution, capturing the hierarchy of (visual) features in natural data.
        - We precisely characterize the *feature-learning* process of a *nonlinear* convolutional neural network trained from *random initialization* via *stochastic gradient descent*.
        - The data properties, in conjunction with the characterization of the feature-learning process, yields a correspondence between representation complexity and label complexity. This representation-complexity result leads to our conclusion on the network's generalization performance.
    - Published in TMLR: https://openreview.net/pdf?id=FojAV72owK

Student-teacher learning

- Developing a theoretical understanding of *feature-based knowledge distillation* using (deep) linear neural networks, and empirical understanding of it using nonlinear neural networks.
    - Setting: teacher learns on data with *clean* input signals, while the student learns on *noisy* input signals. It is a common setting in computational imaging.
    - Theory: utilized tools from LASSO analysis and SGD dynamics of deep linear networks.
    - Experiments: conducted systematic experiments on Cifar-10 using ResNet-18 to investigate the operating regime of the distillation method.

    - We arrived at three conclusions: (1) whether the student is trained to convergence; (2) how knowledgeable the teacher is on the clean-input problem; (3) how the teacher decomposes its knowledge in its hidden features. Lack of proper control in any of the three factors leads to failure of the student-teacher learning method.
- Published in CVPR 2021: https://openaccess.thecvf.com/content/CVPR2021/html/Hong_Student-Teacher_Learning_From_Clean_Inputs_to_Noisy_Inputs_CVPR_2021_paper.html

Turbulence mitigation

- Mainly responsible for providing theoretical justifications for parameter designs of the non-local reference generation method in the overall turbulence mitigation pipeline.
    - Supervisor: Professor Stanley Chan.
    - See https://arxiv.org/abs/1905.07498

Publications and Preprints

Guanzhe Hong, Nishanth Dikkala, Enming Luo, Cyrus Rashtchian, Xin Wang, Rina Panigrahy. "How Transformers Solve Propositional Logic Problems: A Mechanistic Analysis". arXiv preprint arXiv:2411.04105, 2024.

Guanzhe Hong, Nishanth Dikkala, Enming Luo, Cyrus Rashtchian, Xin Wang, Rina Panigrahy. "How Transformers Reason: A Case Study on a Propositional Logic Problem". In *NeurIPS MATH-AI workshop*, 2024.

Guanzhe Hong, Yin Cui, Ariel Fuxman, Stanley H Chan, and Enming Luo. "Why Fine-grained Labels in Pretraining Benefit Generalization?". In *Transactions on Machine Learning Research*, 2024.

Guanzhe Hong, Zhiyuan Mao, Xiaojun Lin, Stanley Chan. "Student-Teacher Learning from Clean Inputs to Noisy Inputs". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021

Nicholas Chimitt, Zhiyuan Mao, **Guanzhe Hong**, and Stanley H Chan, "Rethinking atmospheric turbulence mitigation," arXiv preprint arXiv:1905.07498, 2019.

Awards & Honors

Magoon Excellence in Teaching Award 2021

- Given to outstanding teaching assistants and instructors at Purdue University

Dean's Honors List 2014 - 2017

- Dean's Honors List gives special recognition to students who have demonstrated academic excellence in an individual session at the University of Toronto

Services

Conference Reviewer: CVPR, ECCV, ICASSP

Journal Reviewer: IEEE Transactions on Computational Imaging

Technical Skills

Languages: English, Chinese (Mandarin)

Programming Languages: Python, Matlab, C/C++, Java, Verilog, Simulink

Libraries: Pytorch, Tensorflow