Jiani Huang

Contact Information

Phone: 858-232-8235
Email: jianih@seas.upenn.edu
Website: https://www.cis.upenn.edu/~jianih/

Research Interests

My research interests involve programming languages and machine learning. I am particularly interested in combining neural and symbolic approaches to develop accurate, reliable, and efficient machine learning solutions.

Education

University of Pennsylvania — Philadelphia, PA
Ph.D. Candidate, Advisor: Mayur Naik — *Aug. 2018 – June 2024*
University of California, San Diego — San Diego, CA
Bachelor of Science in Math - Computer Science — *Sep. 2014 – June 2018*

Conference Publications

- **Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning**
  Jiani Huang, Ziyang Li, Binghong Chen, Karan Samel, Xujie Si, Le Song, Mayur Naik
  Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021)
- **Generating Programmatic Referring Expressions via Program Synthesis**
  Jiani Huang, Calvin Smith, Osbert Bastani, Rishabh Singh, Aws Albarghouthi, Mayur Naik
  International Conference on Machine Learning (ICML 2020)

Workshop Publications

- **Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning**
  Jiani Huang, Ziyang Li, Binghong Chen, Karan Samel, Xujie Si, Le Song, Mayur Naik
  NeurIPS 2021 Workshop: Advances in Programming Languages and Neurosymbolic Systems (AIPLANS)
- **Numerical Reasoning over Legal Contracts via Relational Database**
  Jiani Huang, Ziyang Li, Ilias Fountalis, Mayur Naik
  NeurIPS 2021 Workshop: Workshop on Databases and AI (DBAI)
- **Generating Programmatic Referring Expressions via Program Synthesis**
  Jiani Huang, Calvin Smith, Osbert Bastani, Rishabh Singh, Aws Albarghouthi, Mayur Naik
  International Conference on Machine Learning (ICML 2020)

Research Experience

Research Assistant — May 2021 – Aug 2021
Relational AI — *Remote*

- **Lawgic:** Reasoning about Legal Contracts with Logic
  Numerical reasoning over text requires deep integration between the semantic understanding of the natural language context and the mathematical calculation of the symbolic terms. To incorporate the domain-specific knowledge and express mathematical calculation over structured data, I designed a database schema for natural language semantic representation and a set of rules for information extraction. Evaluating this method on the CUAD dataset shows that our approach has high correct answer coverage and reduces a significant amount of incorrect results even without any labels.

Research Assistant — June 2018 – Present
University of Pennsylvania — *Philadelphia, PA*

- **Scallop:** A Scalable Inference Engine for Probabilistic Logic Programs
  I designed a differentiable probabilistic datalog engine Scallop, which takes in a Datalog program and a distribution over its inputs, efficiently approximates the marginal probabilities for the query results. The engine enables end-to-end differentiable training of deep neural nets for machine learning applications that must satisfy logical rules. Compared to DeepProbLog, which performs exact inference but does not scale to real-world tasks, Scallop provides fast turnaround time with tunable accuracy.

- **VQA + Knowledge Graph:** Visual Question-Answering with Reasoning
  I combined the differentiable probabilistic Datalog engine, Scallop, with a perception model to answer questions involving images and common sense knowledge. The answers are generated in an interpretable and extensible manner by decoupling the perception and reasoning components. This architecture is more data-efficient and its predictions are more precise and generalizable compared to the modular network, a state-of-the-art neural architecture that performs complex reasoning by combining component neural networks for individual reasoning tasks.
- **Generating Programmatic Referring Expressions via Program Synthesis**
  I implemented a neural synthesizer that generates correct programmatic referring expressions in images by incorporating an in-loop interpreter into the reinforcement learning pipeline. The neural network improves synthesis efficiency by using perception to shrink the exponential search space while the interpreter ensures accuracy. Compared to the state-of-art neural symbolic synthesizer METAL, which is solely dependant on recurrent neural networks, our approach has significantly better performance and generalizability.

Undergraduate Research Assistant Jan. 2015 – June 2018
University of California, San Diego *San Diego, CA*

- Improving OCaml compiler error messages using type annotations
  - Mentor: Prof. Ranjit Jhala
- Implementing accurate and hardware friendly machine learning algorithms using hyper-dimension vectors.
  - Mentors: Prof. Mohsen Imani and Prof. Tajana Rosing

Professional Services

Reviewer: NeurIPS 2021-2022, ICML 2021-2022, ICML 2022 pre-training workshop

Teaching experience

Teaching Assistant Fall 2019
University of Pennsylvania *Philadelphia, PA*

- CIS 511: Computational Theory

Tutor
University of California, San Diego *San Diego, CA*

- CIS 20: Discrete Mathematics
- MATH 187A: Cryptography

Technical Skills

Languages: Python, Rust, Datalog, OCaml, Coq, Java, C/C++, Haskell, Verilog, Latex, JavaScript, HTML/CSS
Libraries: Pytorch, Pytorch Geometrics, YOLO, OpenCV, StanfordNLP
Developer Tools: Git, Docker, VS Code
Design: Adobe Photoshop, Adobe InDesign, Procreate