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 lie at the intersection of programming languages and machine learning. My thesis research focuses on the design, implementation, and applications of Scallop, a general-purpose **neurosymbolic** language and compiler framework. The neurosymbolic paradigm serves to bridge the otherwise complementary worlds of deep learning and symbolic reasoning. My research goal is to develop accurate, explainable, and efficient AI solutions for a wide variety of fields ranging from natural language processing, and computer vision, to medical applications.

Education

University of Pennsylvania — Philadelphia, PA
Ph.D. Candidate, Computer Science (Advisor: Mayur Naik) — *Aug. 2018 – Dec. 2024*

University of California, San Diego — San Diego, CA
Bachelor of Science in Math - Computer Science — *Sep. 2014 – June. 2018*

Publications

Pre-prints

- **LASER: A Neuro-Symbolic Framework for Learning Spatial-Temporal Scene Graphs with Weak Supervision**
  Jiani Huang, Ziyang Li, Mayur Naik, Ser-Nam Lim
  ArXiv:2304.07647 (Preprint)

Monographs

- **Neurosymbolic Programming in Scallop: Principles and Practice**
  Ziyang Li, **Jiani Huang**, Mayur Naik
  Foundations and Trends® in Programming Languages, 2024

Major Conference Publications

- **Relational Programming with Foundation Models**
  Ziyang Li, **Jiani Huang**, Jason Liu, Felix Zhu, Eric Zhao, William Dodds, Neelay Velingker, Rajeev Alur, Mayur Naik
  Association for the Advancement of Artificial Intelligence (AAAI 2024)
- **Scallop: A Language for Neurosymbolic Programming**
  Ziyang Li*, **Jiani Huang***, Mayur Naik
  Conference on Programming Language Design and Implementation (PLDI 2023)
- **Improved Logical Reasoning of Language Models via Differentiable Symbolic Programming**
  Jiani Huang*, Hanlin Zhang*, Ziyang Li, Mayur Naik, Eric Xing
  Findings of the Association for Computational Linguistics (ACL 2023)
- **Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning**
  Jiani Huang, Ziyang Li, Binghong Chen, Karan Samel, Xujie Si, Le Song, Mayur Naik
  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)

Other Conference Publications

- **FACH: FPGA-based Acceleration of Hyperdimensional Computing by Reducing Computational Complexity**
  Mohsen Imani, Sahand Salamat, Saransh Gupta, Jiani Huang, Tajana Rosing
  IEEE Asia and South Pacific Design Automation Conference (ASP-DAC 2019)

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
  ICML 2020 Workshop: Bridge Between Perception and Reasoning: Graph Neural Networks and Beyond

Research Experience

Visiting Researcher June 2022 – Dec 2023
Meta

* **LASER**: A Neuro-Symbolic Framework for Learning Spatial-Temporal Scene Graphs with Weak Supervision
  LASER is a neuro-symbolic approach that learns semantic video representations by leveraging logic specifications that can capture rich spatial and temporal properties in video data. In particular, we formulate the problem in terms of alignment between raw videos and specifications. The alignment process efficiently trains low-level perception models to extract a fine-grained video representation that conforms to the desired high-level specification. We demonstrate that our method not only learns fine-grained video semantics but also outperforms existing baselines on downstream tasks such as video retrieval.

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 engine Scallop, which takes in a Datalog program and a distribution over its inputs, and 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 dependent 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

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

Mentoring Experience

Mentor June 2023 – Aug. 2023
University of Pennsylvania *Philadelphia, PA*
* Mentored undergraduate students Jason Liu, Liam Dodds, Eric Zhao leading to a publication at AAAI 2024.

Mentor Nov. 2023 – Present
University of Pennsylvania *Philadelphia, PA*
* Currently mentoring Masters student Siyang Zhang on a neurosymbolic text-to-video generation project.

Academic Service

Reviewer: CVPR 2023, ICLR 2023-2024, NeurIPS 2021-2023, ICML 2021-2024, ACL ARR 2023 October, ICML pre-training workshop 2022

Technical Skills

Languages: Scallop, 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