Ziyang Li

https://liby99.github.io/

Email : liby99@seas.upenn.edu
Mobile : +1-858-699-3237

Education

- **University of Pennsylvania** — Philadelphia, PA
  Ph.D. Computer and Information Science, Advisor: Mayur Naik; GPA: 4.0 — *Jul 2019 – Present*
- **University of California – San Diego** — La Jolla, CA
  B.S. Computer Science (3.9/4.0); B.S. Mathematics (3.7/4.0); GPA: 3.6 — *Sep 2015 – Jun 2019*

Research Interest

My research spans the fields of programming languages (PL) and machine learning (ML). I am particularly interested in neurosymbolic approaches which integrates traditional symbolic reasoning frameworks with machine learning models, jointly improving natural perception and reasoning capabilities. I design the programming language, Scallop, which is a neurosymbolic language based on Datalog, that can be integrated with PyTorch framework, supporting differentiable logical reasoning. I apply neurosymbolic methods to various domains including computer vision (CV), natural language processing (NLP), program analysis, computer security, robotics, and etc.

Publications

- **Understanding the Effectiveness of Large Language Models in Detecting Security Vulnerabilities**
  Avishree Khare, Saikat Dutta, Ziyang Li, Alaia Solko-Breslin, Rajeev Alur, Mayur Naik
  [arXiv]
- **LASER: Neuro-Symbolic Learning of Semantic Video Representations**
  Jiani Huang, Ziyang Li, Mayur Naik, David Jacobs, Sernam Lim
  [arXiv]
- **Relational Programming with Foundation Models**
  Ziyang Li, Jiani Huang, Jason Liu, Felix Zhu, Eric Zhao, William Dodds, Neelay Velingker, Rajeev Alur, Mayur Naik
  AAAI Conference on Artificial Intelligence (AAAI 2024)
- **Improved Logical Reasoning of Language Models via Differentiable Symbolic Programming**
  Jiani Huang, Hanlin Zhang*, Ziyang Li, Mayur Naik, Eric Xing*
  ACL-Findings 2023
- **Scallop: a Language for Neurosymbolic Programming**
  Ziyang Li, Jiani Huang, Mayur Naik
  Programming Language Design and Implementation (PLDI 2023). [Code] [arXiv]
- **Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning**
  Jiani Huang, Ziyang Li*, Binghong Chen, Karan Samel, Mayur Naik, Le Song, Xujie Si*
  NeurIPS 2021. [Code]
- **ARBITRAR: User-Guided API Misuse Detection**
  Ziyang Li, Aravind Machiry, Binghong Chen, Mayur Naik, Ke Wang, Le Song
  IEEE Security and Privacy (S&P) 2021. [Paper] [Code] [Talk]
- **HOPPITY: Learning Graph Transformations to Detect and Fix Bugs in Programs**
  Elizabeth Dinella, Hanjun Dai, Ziyang Li, Mayur Naik, Le Song, Ke Wang
  International Conference on Learning Representations (ICLR) 2020, **Spotlight**. [Paper]

Workshop Papers:

- **Scallop: From Probabilistic Deductive Databases to Scalable Differentiable Reasoning**
  Jiani Huang, Ziyang Li*, Binghong Chen, Karan Samel, Mayur Naik, Le Song, Xujie Si*
  AIPLANS Workshop @ NeurIPS 2021
- **Numerical Reasoning over Legal Contracts via Relational Database**
  Jiani Huang, Ziyang Li, Ilias Fountalis, Mayur Naik
  DBAI Workshop @ NeurIPS 2021

In Submission:

- **Beyond Differentiability: Neurosymbolic Learning with Black-Box Programs**
  Alaia Solko-Breslin, Ziyang Li, Neelay Velingker, Rajeev Alur, Mayur Naik
  In Submission
- **DISCRET: A Self-Interpretable Framework for Treatment Effect Estimation**
  Yinjun Wu, Neelay Velingker, Ziyang Li, Kan Chen, Mayank Keoliya, Mayur Naik, Qi Long, Eric Wong, Emily Getzen, Ravi Parikh
  In Submission
- **Bityr: A Pluggable Framework for Learning Types from Binaries**
  Ziyang Li, Anton Xue, Hanjun Dai, Rajeev Alur, Aravind Machiry, Mayur Naik
  In Submission
- **Weighted Contrastive Learning for Scene Graph Generation**
  Yinjun Wu, Adam Stein, Ziyang Li, Mayur Naik
  In Submission

Research Experiences

- **Research Assistant, UPenn PEARL** — University of Pennsylvania
  Advisor: Mayur Naik — *July 2019 – Present*
  - **Scallop, a differentiable logical reasoning engine**: I build a scalable and flexible probabilistic DataLog engine, oriented towards machine learning applications. Comparing to existing tools like ProbLog, we support customizable provenance structures for scalable differentiable reasoning. We apply Scallop to many machine learning applications such as CV, NLP, and robotics. [Website] [SSFT22 Tutorial]
  - **Binary Analysis with Machine Learning**: We propose Bityr, a binary analysis framework for binary type inference with graph neural networks. We analyze the binary and construct symbolic dataflow graphs, which provide rich information for binary type inference. We also demonstrate flexibility to integrate user feedback through our graph based neural models.
  - **Find API Misuses with User Interaction**: We build Arbitrar, an Active Learning based method to involve human-in-the-loop to find API Misuses bugs in large code corpus. We demonstrated that we can find the bug given a target API with only a few rounds of user interactions, where we learn the correct usage from scratch. We showed that Arbitrar is much more effective than existing tools like APISan which has a huge amount of false positives.
  - **Learn to Find and Fix JavaScript Bugs**: We present Hoppity, a GNN based learning method to find and fix JavaScript bugs, trained on a huge dataset consists of Github commits. Hoppity correctly detects and fixes bugs in 9,490 out of 36,361 programs in an end-to-end fashion. Given the bug location and type of the fix, Hoppity also outperforms the baseline approach by a wide margin.
- **Undergraduate Research Assistant, UCSD PL** — Univerysity of California – San Diego
  Advisor: Sorin Lerner — *Sep 2018 – Jun 2019*
  - **Proverbot9001**: Learn to prove theorems in Coq automatically using machine learning and neural networks.
- **Undergraduate Research Assistant, UCSD VISCOMP** — Univerysity of California – San Diego
  Advisor: Ravi Ramamoorthi — *Mar 2018 – Jun 2019*
  - **Spherical Harmonics**: I developed a Spherical Harmonics based real time realistic lighting for AR applications.

- **Undergraduate Research Intern, UCSD Design Lab** Univerysity of California – San Diego
  Advisor: Scott Klemmer *Jun 2017 – Jun 2018*
  - **Galileo**: We built a platform for people to design and run life-style related experiments with community. We teach users to design and conduct double blind experiments, recruit experiment candidates, and generate experiment reports.

Working Experiences

- **Relational AI** *Research Intern* Virtual, *May 2021 – August 2021*
- **Visa, Inc.** *Research Intern, Mentor: Ke Wang* Virtual, *May 2020 – July 2020*
- **Coursera, Inc.** *Front-end Engineer Intern* Mountain View, CA, *Jun 2018 - Sep 2018*
- **Deep Media, Ltd.** *Full-stack Engineer Intern* Shenzhen, China, *Sep 2016 – Jan 2017*
- **Yobs Technology** *Full-stack Engineer Intern* Los Angeles, CA, *Jan 2016 – Sep 2016*
- **Easyhin** *Front-end Engineer Intern* Shenzhen, China, *Aug 2015 – Sep 2015*

Fellowships

- **KPCB Fellows 2018** *Engineering Fellows* San Francisco, *June 2018*

Teaching Experiences

- **Teaching Assistant** *CIS 547, Software Analysis* University of Pennsylvania, *Fall 2020, Fall 2021*
- **Tutor** *CSE 190, Virtual Reality Technology* University of California – San Diego, *Spring 2019*
- **Tutor** *CSE 165, 3D User Interaction* University of California – San Diego, *Winter 2019*
- **Tutor** *CSE 130, Programming Language* University of California – San Diego, *Fall 2018*
- **Tutor** *CSE 163, Advanced Computer Graphics* University of California – San Diego, *Spring 2018*
- **Tutor** *CSE 167, Intro to Computer Graphics* University of California – San Diego, *Winter 2018*
- **Tutor** *CSE 12, Data Structure* University of California – San Diego, *Winter 2017*

Selected Side-Projects

- **Probabilistic DataLog Engine**: A probabilistic datalog engine with high performance optimizations oriented towards machine learning applications. Written in Rust.
- **Under-constrained Symbolic Execution Engine**: High performance under-constrained symbolic execution engine for LLVM IR written in Rust. Used in Arbitrar.
- **LLVM IR Binding for Rust**: Safe LLVM Binding for Rust. Used in Arbitrar. [Github]
- **AoSoA Storage**: Array-of-struct-of-array storage system for high performance parallel computing with Kokkos and Cabana. Oriented for physics simulation applications. Used by UPenn CG Group. [Github]
- **MPM-RS**: Material point method written in Rust. [Github]
- **Geometry Sketchpad**: Geometry sketching GUI application written in Rust. [Github]
- **Menhera**: A TypeScript-like functional programming language compiler written in OCaml. [Github]
- **Fourier Depth of Field**: Fourier transform based depth of field analysis for RayTracer. [Github]
- **Neon Ping Pong**: VR Ping Pong Game written in C++. [Website] [Video]
- **Space Escape**: VR Room Escape Puzzle Game settled in Space Station. Developed in Unity. [Website] [Video]
- **Rotamina**: Character animator and simulator with GUI. Written in C++. [Github]

Skills

- **Languages**: Rust, C++/C, TypeScript/JavaScript, Python, C#, OCaml, Java, Haskell, Coq
- **Libraries/Tools**: PyTorch, Unity, Rocket, React, ExpressJs, Asp.net
- **Design**: Adobe Photoshop, Premiere, Illustrator, Blender, Cinema 4D

Last update: Jan 3, 2024