Shuyang Liu

Department of Computer Science
Huazhong University of Science and Technology

+86-18397675942
liushuyang860@gmail.com

Education

•B.E. in Computer Science *2020-24*

Huazhong University of Science and Technology (HUST) GPA: 3.92/4.00

Scholarships: Excellent Student Cadre (2020), Model Student of Academic Records (2021)

Research Experience

•Symmetry-Preserving Program Representations for Learning Code Semantics

Advisor: Kexin Pei, Suman Jana Mar 2023 - Aug 2023

Using a rigorous group-theoretic framework, we introduce a novel variant of self-attention that preserves program symmetries, demonstrating its effectiveness in generalization and robustness through detailed experimental evaluations across different binary and source code analysis tasks.

- –Employed JavaParser on Java ASTs to analyze control and data dependencies for source code.
- –Applied four types of semantic-preserving source transformations: variable renaming; statement permutation, which I extended existing work beyond only two-instruction permutation to all possible PDG automorphisms; loop exchange; and boolean exchange.
- –Reproduced three baselines for Java method name prediction: code2vec, code2seq, and GGNN.

•Automatic identification of Bug Inducing Commits via Static Analysis *Oct 2022 - Present*

Ongoing, with CASTLE Research Group, HKUST

- –Systematically validated bug-fixing and associated bug-inducing commits for 237 bugs across five large open-source Java projects. Employed the bisection method for precise identification.
- –Replicated the SZZ Unleashed algorithm using the extensive defects4j dataset, establishing a foundational baseline for performance benchmarking.
- –Reproduced the "Reducing the Search Space of BICs using Failure Coverage" approach. Augmented the research by intersecting its outcomes with those of the SZZ Unleashed, further enriching the comparative analysis.

Academic Projects

•Real-time Traffic Sign Recognition via Optimized CNN and Advanced Augmentation *Advisor: Terence Sim*

National University of Singapore (NUS) School of Computing 2022 Summer Workshop May 2022 - Aug 2022

Developed a real-time traffic sign recognition system utilizing the GTSRB Dataset, encompassing more than 50,000 images from real-world scenarios.

- –Constructed a Convolutional Neural Network (CNN) with TensorFlow's Sequential model, iteratively fine-tuning hyperparameters to distill the most performant configuration.
- –Implemented a suite of innovative data augmentation techniques, notably Random Erasing, strategically erasing regions within input images to heighten model robustness and generalize to diverse scenarios.

The system demonstrated a remarkable F1-score of 98.5%, outperforming human eye recognition benchmarks and affirming its efficacy in real-world traffic sign identification.

Technical Skills and Interests

Languages: C/C++, Java, Python
Libraries : C++ STL, Python Libraries (Pytorch, Tensorflow)
Tools: Git, Linux, LaTeX, MySQL, JavaParser, and Tree-sitter
Relevent Coursework: Algorithm, Data Structure, Machine Learning, Compilers, and Software Engineering
Areas of Interest: Program Analysis, Security, Software Engineering, and Machine Learning

Publication

[Under Review] [Oakland S&P 2024] Kexin Pei, Weichen Li*, Qirui Jin*, Shuyang Liu, Scott Geng, Lorenzo Cavallaro, Junfeng Yang, Suman Jana
Symmetry-Preserving Program Representations for Learning Code Semantics