Shuyang Liu

Department of Computer Science
Huazhong University of Science and Technology

shuyang-liu.github.io
liushuyang860@gmail.com

Education

•B.E. in Computer Science *Sept. 2020 - Jun. 2024*
Huazhong University of Science and Technology (HUST) GPA: 3.92/4.00

Research Experience

•Exploiting Code Symmetries for Learning Program Semantics *Advisor: Kexin Pei, Suman Jana*
Columbia University 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 to construct PDGs based on data/control dependencies between statements.
- –Applied 6 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; boolean exchange; unused statement; and switch to if.
- –Evaluated baselines for Java method name prediction, which performs an "extreme summarization" of the function behavior.

•Automatic identification of Bug Inducing Commits *Advisor: Ming Wen*
With CASTLE Lab, HKUST Oct. 2022 - Present

- –Systematically validated bug-fixing and associated bug-inducing commits for 237 bugs across 5 large open-source Java projects. Employed bisection method for precise identification.
- –Replicated SZZ Unleashed using the extensive defects4j dataset, establishing a foundational baseline for performance benchmarking.
- –Reproduced "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 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 Keras, 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 (Pytorch, Tensorflow)
Tools: Git, Linux, LaTeX, MySQL, JavaParser, and Tree-sitter
Relevant Coursework: Algorithm, Data Structure, Machine Learning, Compilers, and Software Engineering
Areas of Interest: Software Engineering, Security, Programming Language, and Machine Learning

Publication

Symmetry-Preserving Program Representations for Learning Code Semantics
Kexin Pei, Weichen Li*, Qirui Jin*, **Shuyang Liu**, Scott Geng, Lorenzo Cavallaro, Junfeng Yang, Suman Jana
Accepted by the 7th Symposium on Machine Programming (MAPS), 2023.
Submitted to the 12th International Conference on Learning Representations (ICLR), 2024. Under Review.