Yifei Liu

Room 336, New CS Building, Stony Brook, NY 11794-2424 • Cell: +1 631-710-8377
yifeliu@cs.stonybrook.edu • https://www.fsl.cs.stonybrook.edu/~yifei/ • linkedin.com/in/yifei-liu/
Visa: F-1 Student • Timezone: US/Eastern

Education

Stony Brook University — Stony Brook, NY
- Ph.D. in Computer Science (Advisor: Prof. Erez Zadok) — 08/2019 – Present
- CGPA: 3.91 / 4.0
- *Relevant Courses*: CSE506 Operating Systems; CSE512 Machine Learning; CSE532 Theory of Database Systems; CSE548 Analysis of Algorithms

Huazhong University of Science and Technology — Wuhan, China
- M.Sc. in Computer System Architecture (Advisor: Prof. Ke Zhou) — 09/2016 – 06/2019

Huazhong Agricultural University — Wuhan, China
- B.Eng. in Computer Science and Technology — 09/2012 – 06/2016

Experience

File systems and Storage Lab (FSL), Stony Brook University — Stony Brook, NY
Graduate Research Assistant — 05/2020 – Present
- Apply model checking and state-space exploration to verify Linux file systems thoroughly and automatically
- Design and benchmark multi-tier caching systems with intelligent MRC point selection to identify good cache configurations effectively

Stony Brook University — Stony Brook, NY
Graduate Teaching Assistant — 08/2019 – 05/2020
- CSE376 Advanced Systems Programming in Unix/C (S'21, S'20) Prof. Erez Zadok
- CSE306 Operating Systems (F'19) Prof. Eugene Stark

Wuhan National Laboratory for Optoelectronics — Wuhan, China
Master's Student/Research Assistant — 09/2016 – 06/2019
- Used deep learning hash to design and implement a metadata system for integration of high-precision and low-latency content-based semantic queries in storage systems
- Proposed a framework for assessing image "dark data" based on a novel semantic hash ranking (SHR) algorithm
- Performed theoretical analysis on hash-based graphs to facilitate rank algorithms and graph database operations

Tencent — Shenzhen, China
Backend Developer Intern — 12/2015 – 08/2016
- Predicted disk failures with disk data collected via machine learning algorithms to achieve high precision and recall
- Built infrastructure for collecting long-term disk S.M.R.A.T. data from over 10,000 servers in Tencent data centers
- Tested cache replacement policies on Tencent application traces

Selected Publications

Journal Articles

[1] Ke Zhou, Yangtao Wang, Yu Liu, Yujuan Yang, Yifei Liu, Guoliang Li, Lianli Gao, and Zhili Xiao. "A Framework for Image Dark Data Assessment." *World Wide Web*, 2020.

[2] Yu Liu, Yangtao Wang, Ke Zhou, Yujuan Yang, and Yifei Liu. "Semantic-aware Data Quality Assessment for Image Big Data." *Future Generation Computer Systems*, 2020.

Conference and Workshop Papers

[1] Wei Su, Yifei Liu, Gomathi Ganesan, Gerard Holzmann, Scott Smolka, Erez Zadok and Geoff Kuenning. "Model-Checking Support for File System Development." In *the 13th ACM Workshop on Hot Topics in Storage and File Systems (HotStorage)*, Virtual, 2021.

[2] Yu Liu, Hong Jiang, Yangtao Wang, Ke Zhou, Yifei Liu, and Li Liu. "Content Sifting Storage: Achieving Fast Read for Large-scale Image Dataset Analysis." In *the 57th Design Automation Conference (DAC)*, San Francisco, CA, 2020.

[3] Yu Liu, Yangtao Wang, Ke Zhou, Yujuan Yang, Yifei Liu, Jingkuan Song, and Zhili Xiao. "A Framework for Image Dark Data Assessment." In *the 3rd APWeb-WAIM joint conference on Web and Big Data (APWeb-WAIM)*, Chengdu, China, 2019. **(Best Paper Runner-Up)**

Patents

[1] Ke Zhou, Yifei Liu, Yu Liu, Yangtao Wang, and Yujuan Yang. A kind of image inquiry method and system based on contents semantic metadata. Chinese patent CN110413807B, Filed June, 2019. Granted April, 2021.

Talks

- *Model-Checking Support for File System Development*, ACM HotStorage 2021, Virtual. (Joint talk w/ Wei Su)
- *OS Support for File System Model Checking*, Computer Science Graduate Research Day 2021, Stony Brook, NY.

Skills

Programming Languages

- Familiar ($\geq$ 4 years of experience): C, C++, Python
- Intermediate (1 ~ 3 years): MATLAB, Bash, SQL, Java, Cypher
- Basic ($\leq$ 1 year): JavaScript, Prolog

Technologies

- **Databases:** MySQL (3 years), Neo4j (2 years), DB2 (< 1 year), HBase (< 1 year)
- **File and Storage:** Linux VFS (2 years), OpenStack Swift (2 years), HDFS (< 1 year)
- **Operating Systems:** Linux (6 years), Linux kernel development (1 year)
- **Big Data:** Hadoop (1 year), Spark (< 1 year)
- **Tools:** VSCode (5 years), Git (3 years), Vim (3 years), Makefile (3 years), LaTeX (3 years), GDB (2 years), Elasticsearch (1 year), CMake (1 year), Scikit-learn (1 year)

Human Languages

- Chinese (Native), English (Fluent)

Projects

- **MCFS:** A model checking framework to verify Linux file systems thoroughly and efficiently
- **MTCache:** Point selection to identify optimal multi-tier caching configurations effectively
- **SwiftGraph:** A system middleware for content-based semantic queries in storage systems
- **Disk Failure Prediction:** Prediction of disk failures in Tencent Cloud

Contest Awards

- Finalist, Interdisciplinary Contest in Modeling (MCM/ICM), USA, 2015.
- First Prize, National Postgraduate Mathematic Contest in Modeling, China, 2014.
- First Prize, Contemporary Undergraduate Mathematical Contest in Modeling, China, 2014.