# 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

## Biography

I am a Ph.D. student in Computer Science at Stony Brook University advised by Prof. Erez Zadok and a Research Assistant at the File systems and Storage Lab (FSL). My research interests include file/storage systems, operating systems, and machine learning for systems. I am currently seeking an SDE or research internship in the summer of 2022.

## Education

**Stony Brook University** — Stony Brook, NY
- Ph.D. in Computer Science — 08/2019 – Present
- Advisor: Prof. Erez Zadok
- CGPA: 3.91 / 4.0
- *Courses*: CSE505 Computing with Logic; CSE506 Operating Systems; CSE512 Machine Learning; CSE532 Theory of Database Systems; CSE544 Probability and Statistics for Data Science; CSE548 Analysis of Algorithms

**Huazhong University of Science and Technology** — Wuhan, China
- M.Sc. in Computer System Architecture — 09/2016 – 06/2019
- Advisor: Prof. Ke Zhou
- CGPA: 85.02 / 100
- *Thesis*: Research on Metadata Organization Approach for Image Storage Systems towards Content-based Semantic Similarity Query

**Huazhong Agricultural University** — Wuhan, China
- B.Eng. in Computer Science and Technology — 09/2012 – 06/2016
- Advisors: Prof. Jianxiao Liu and Prof. Ruifang Zhai
- CGPA: 3.43 / 4.00 Major GPA: 3.79 / 4.00 Rank: 9/118
- *Thesis*: Predicting Disk Failures based on Machine Learning Methods

## Experience

**File systems and Storage Lab (FSL), Stony Brook University** — Stony Brook, NY
*Graduate Research Assistant* — 05/2020 – Present
- Design and benchmark multi-tier caching systems with intelligent MRC analysis
- Apply model checking to verify file systems formally and automatically

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

**Institute of Computing Technology, Chinese Academy of Sciences** Beijing, China
*Intern/Visiting Student* 11/2015 – 12/2015

- Explored the correlation between Apache Spark and JVM GC parameters

**Huazhong Agricultural University** Wuhan, China
*Undergraduate Research Assistant* 11/2014 – 06/2015

- Used Bayesian Network Reasoning to propose a web service recommendation approach for organizing and recommending a set of correlated services
- Designed an image line detection algorithm for automatically measuring character parameters of the rapeseed plant
- Developed a GUI for counting the number of cotton cells in microscopic images (for the Huazhong Agricultural University's College of Plant Science & Technology)

## 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 (FGCS)*, 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 *Proceedings of 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 *Proceedings of the 57th Design Automation Conference (DAC)*, San Francisco, CA, 2020.

[3] Yangtao Wang, Yu Liu, Yifei Liu, Ke Zhou, Yujuan Yang, Jiangfeng Zeng, Xiaodong Xu, and Zhili Xiao. "Analysis and Management to Hash-Based Graph and Rank." In *Proceedings of the 3rd APWeb-WAIM joint conference on Web and Big Data (APWeb-WAIM)*, Chengdu, China, 2019.

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

[5] Jianxiao Liu, Zonglin Tian, Yifei Liu, and Liang Zhao. "Research of Web Service Recommendation Using Bayesian Network Reasoning." In *Proceedings of the 15th International Conference on Services Computing (SCC)*, Seattle, WA, 2018.

[6] Pujuan Shi, Yihang Fang, Chengda Lin, Yifei Liu and Ruifang Zhai. "A new line detection algorithm - Automatic measurement of character parameter of rapeseed plant by LSD." In *Proceedings of the 4th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Istanbul, Turkey, 2015.

**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.

[2] Ke Zhou, Yu Liu, Yujuan Yang, Hua Wang, Chunhua Li, Yangtao Wang, and Yifei Liu. Method for valuation of image dark data based on similarity hashing. US patent US20200410304A1, Filed June 2019. (Under examination)

## 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), Windows 10 (4 years), MacOS (2 years), Linux kernel development (1 year)
- **Parallel Computing:** 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)

**Contest Awards**

- Finalist, Interdisciplinary Contest in Modeling (MCM/ICM), USA, 2015.
  - One of 52 winning teams in the world. First awardee of my university. (Rate: **52/9773 ~0.53%**)
- National First Prize, National Postgraduate Mathematic Contest in Modeling, China, 2014.
  - Won the highest award of this contest. (Rate: **120/4900 ~2.4%**)
- National First Prize, Contemporary Undergraduate Mathematical Contest in Modeling, China, 2014.
  - Won the highest award of this contest. (Rate: **293/22233 ~1.3%**)

**Academic Awards**

**Huazhong University of Science and Technology**

- APWeb-WAIM Best Paper Runner-Up 2019
- Outstanding Graduate Spring 2019
- Merit Graduate Student (2 times) 2017 – 2019
- First-class Academic Scholarship (3 times) 2016 – 2019

**Huazhong Agricultural University**

- Outstanding Graduate Spring 2016
- BioMarker Scholarship Fall 2015
- First Prize for Excellent Study Spring 2015
- Merit Undergraduate Student (2 times) 2014 – 2016

**Certifications** Machine Learning - Stanford Univeristy (2016), Coursera