Yifei Liu

New CS Building, Room 336, Stony Brook University, Stony Brook, NY 11794-2424
yifeliu@cs.stonybrook.edu • +1 631-710-8377 • http://www.fsl.cs.stonybrook.edu/~yifei/

RESEARCH INTERESTS

File/Storage Systems, Operating Systems, and Machine Learning for Systems.

EDUCATION

Stony Brook University — Stony Brook, NY

- Ph.D. in Computer Science — 08/2019 – Current
  - 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
  - *Visa Status*: F-1 Student

Huazhong University of Science and Technology — Wuhan, China

- Master of Science (M.Sc.) in Computer System Architecture — 09/2016 – 06/2019
  - 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

- Bachelor of Engineering (B.Eng.) in Computer Science and Technology — 09/2012 – 06/2016
  - 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), CS Department, SBU — Stony Brook, NY
Graduate Research Assistant — 05/2020 – Current

Advisor: Prof. Erez Zadok

- Conduct research on file/storage systems and operating systems.

Wuhan National Laboratory for Optoelectronics — Wuhan, China
Master's Student/Research Assistant — 09/2016 – 06/2019

Advisor: Prof. Ke Zhou

- 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 Java Garbage Collection parameters

Chinasoft International Ltd — Guangzhou, China
Front-end Student Intern — 07/2015 – 08/2015

- Collaborated with other interns to develop a catering website using jQuery, AJAX, Java Servlet

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

Supervisors: Prof. Jianxiao Liu and Prof. Ruifang Zhai

- 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)

TEACHING

Graduate Teaching Assistant, Stony Brook University 08/2019 – 05/2020

- CSE376 Advanced Systems Programming in Unix/C, Spring 2020. Professor: Erez Zadok
- CSE306 Operating Systems, Fall 2019. Professor: Eugene Stark

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 Papers

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Pujuan Shi, Yifei Liu, Yihang Fang, Chengda Lin, 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. CN110413807A, June 2019. (Under substantive examination)

[2] Ke Zhou, Yu Liu, Yujuan Yang, Hua Wang, Chunhua Li, Yangtao Wang, and Yifei Liu. A kind of dark data value appraisal procedure of image based on similitude Hash. CN110390352A, June 2019. (Under substantive examination)

CONTEST AWARDS

- Second Prize, National Postgraduate Mathematic Contest in Modeling, China 12/2016
- Finalist, Interdisciplinary Contest in Modeling (MCM/ICM), USA 04/2015
  One of 52 awardees in the world. First winner of my university. (**52/9773 ~0.53%**)

- First Prize, National Postgraduate Mathematic Contest in Modeling, China 12/2014
  Won the highest award of this contest, (**120/4900 ~2.4%**)
- First Prize, Contemporary Undergraduate Mathematical Contest in Modeling, China 11/2014
  Won the highest award of this contest, (**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 2017 – 2019
- First-class Academic Scholarship 2016 – 2019

Huazhong Agricultural University

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

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

Softwares and Frameworks

- Parallel Computing: Hadoop, Spark
- File and Storage: OpenStack Swift, HDFS, Redis, Extended File System (ext)
- Databases: HBase, MySQL, Neo4j, DB2
- Deep learning platforms: Caffe, TensorFlow
- Tools: Linux, Git, Vim, GCC, GDB, pdb, CMake, Scikit-learn, Sqoop, LaTeX

Human Languages

- Chinese (Native), English (Fluent)

PROFESSIONAL TRAINING

Machine Learning - Stanford Univeristy (2016), Coursera

[*CV updated on Dec 2020*]