

赵海亮

浙江大学软件学院百人计划研究员，博士生导师
hliangzhao.me Google Scholar 浙江大学主页

hliangzhao@zju.edu.cn

浙江宁波、杭州

研究简介

研究面向智能服务的算法与系统，当前关注学习增强系统、智能体服务、生成式人工智能运行时与云边智能。研究连接在线优化、机器学习与系统设计，相关方法已在国产智能计算芯片及企业级人工智能服务中开展落地验证。

工作经历

浙江大学软件学院 2024 年 8 月至今
百人计划研究员，博士生导师

教育与访问经历

浙江大学计算机科学与技术学院 2019 至 2024 年
工学博士，计算机科学与技术；导师：邓水光教授
新加坡南洋理工大学 PDCL 实验室 2022 年 9 月至 2023 年 9 月
访问博士生；合作导师：Xueyan Tang 教授
武汉理工大学 2015 至 2019 年
工学学士

研究方向

学习增强在线算法；智能计算系统；智能体服务；生成式人工智能系统；云边智能；服务计算与资源调度。

代表荣誉

- 华为技术有限公司“火花奖”，2026 年。
- 浙江省优秀博士学位论文提名，2026 年。
- SICI 最佳论文奖，2025 年。
- CCF 服务计算专委博士学位论文激励计划，2025 年。
- 浙江大学启真优秀青年学者，2024 年。
- 浙江大学优秀博士学位论文，2024 年。
- IEEE ICWS 最佳学生论文奖，2019 年。

代表科研项目

- 主持，国家重点研发计划课题：全要素互联结构自适应的多模态数据治理技术体系，2024–2027 年。
- 主持，国家自然科学基金青年科学基金项目：基于软硬件协同的服务调度与资源管理优化方法研究，2026–2028 年。
- 参与，浙江省自然科学基金重大项目：面向生成式人工智能的异构硬件资源优化与计算加速理论与方法，2025–2027 年。
- 主持，快手推理场景下典型 GPU 服务的吞吐量提高，2024–2025 年。

《计算机工程与科学》青年编委; NeurIPS、IJCAI-ECAI、AAAI、PAKDD 等会议审稿人; IEEE Transactions on Services Computing、IEEE Transactions on Mobile Computing、Future Generation Computer Systems、ACM TKDD 和 IEEE Internet of Things Journal 等期刊审稿人。

代表论文

2026

1. Wenzhuo Qian, **Hailiang Zhao***, Ziqi Wang, Zhipeng Gao, Jiayi Chen, Zhiwei Ling, and Shuiguang Deng*. ARMOR: Missing-Aware Multimodal Fusion for Unified Microservice Incident Management *ASE, accepted*.
2. Peng Chen, **Hailiang Zhao***, Xueyan Tang, Yixuan Wang, and Shuiguang Deng*. Towards Optimal Robustness in Learning-Augmented Paging *ICML, Spotlight*.
3. Jiaji Zhang, Ruichao Sun, **Hailiang Zhao***, Jiaju Wu, Peng Chen, Hao Li, Yuying Liu, Kingsum Chow, Gang Xiong, and Shuiguang Deng*. SegQuant: A Semantics-Aware and Generalizable Quantization Framework for Diffusion Models *CVPR*.
4. Jiaji Zhang, **Hailiang Zhao***, Guoxuan Zhu, Ruichao Sun, Jiaju Wu, Xinkui Zhao, Hanlin Tang, Weiyi Lu, Kan Liu, Tao Lan, Lin Qu, and Shuiguang Deng*. Shiva-DiT: Residual-Based Differentiable Top-k Selection for Efficient Diffusion Transformers *arXiv preprint*.
5. Zhiwei Ling, **Hailiang Zhao***, Chao Zhang, Xiang Ao, Ziqi Wang, Cheng Zhang, Zhen Qin, Xinkui Zhao, Kingsum Chow*, Yuanqing Wu, and MengChu Zhou. Adaptive Dual-Weighting Framework for Federated Learning via Out-of-Distribution Detection *arXiv preprint*.
6. Yu Tang[†], **Hailiang Zhao^{†*}**, Chuansheng Lu, Yifei Zhang, Kingsum Chow*, Shuiguang Deng, and Rui Shi. Morphis: SLO-Aware Resource Scheduling for Microservices with Time-Varying Call Graphs *arXiv preprint*.
7. **Hailiang Zhao**, Ziqi Wang, Daojiang Hu, Zhiwei Ling, Wenzhuo Qian, Jiahui Zhai, Yuhao Yang, Zhipeng Gao, Mingyi Liu, Kai Di, Xinkui Zhao, Zhongjie Wang, Jianwei Yin, MengChu Zhou, and Shuiguang Deng. Industrial Data-Service-Knowledge Governance: Toward Integrated and Trusted Intelligence for Industry 5.0 *arXiv preprint*.

2025

8. Jiahong Yu[†], Ziqi Wang[†], **Hailiang Zhao***, Wei Zhai, Xueqiang Yan, and Shuiguang Deng. TrackTeller: Temporal Multimodal 3D Grounding for Behavior-Dependent Object References *arXiv preprint*.
9. Ziqi Wang, **Hailiang Zhao***, Yuhao Yang, Daojiang Hu, Cheng Bao, Mingyi Liu, Kai Di, Schahram Dustdar, Zhongjie Wang, and Shuiguang Deng. OmniFuser: Adaptive Multimodal Fusion for Service-Oriented Predictive Maintenance *arXiv preprint*.
10. Shuiguang Deng*, **Hailiang Zhao***, Ziqi Wang, Guanjie Cheng, Peng Chen, Wenzhuo Qian, Zhiwei Ling, Jianwei Yin, Albert Y. Zomaya, and Schahram Dustdar. Agentic Services Computing *arXiv preprint*.
11. Peng Chen, Jiaji Zhang, **Hailiang Zhao***, Yirong Zhang, Jiahong Yu, Xueyan Tang, Yixuan Wang, Hao Li, Jianping Zou, Gang Xiong, Kingsum Chow, Shuibing He, and Shuiguang Deng*. Toward Robust and Efficient ML-Based GPU Caching for Modern Inference *arXiv preprint*.
12. Ziqi Wang, **Hailiang Zhao***, Cheng Bao, Wenzhuo Qian, Yuhao Yang, Xueqiang Sun, and Shuiguang Deng*. XFMNet: Decoding Cross-Site and Nonstationary Water Patterns via Stepwise Multimodal Fusion for Long-Term Water Quality Forecasting *arXiv preprint*.
13. Wenzhuo Qian, **Hailiang Zhao***, Tianlv Chen, Jiayi Chen, Ziqi Wang, Kingsum Chow, and Shuiguang Deng*. Learning Unified System Representations for Microservice Tail Latency Prediction *arXiv preprint*.
14. Yinuo Deng[†], **Hailiang Zhao^{†*}**, Dongjing Wang, Peng Chen, Wenzhuo Qian, Jianwei Yin, Schahram Dustdar, and Shuiguang Deng*. PeerSync: Accelerating Containerized Model Inference at the Network Edge *IEEE Transactions on Services Computing*.
15. Peng Chen, **Hailiang Zhao***, Jiaji Zhang, Xueyan Tang, Yixuan Wang, and Shuiguang Deng*. Robustifying Learning-Augmented Caching Efficiently without Compromising 1-Consistency *NeurIPS*.

16. **Hailiang Zhao**, Xueyan Tang*, Peng Chen, Jianwei Yin, and Shuiguang Deng. Data-Locality-Aware Task Assignment and Scheduling for Distributed Job Executions *IEEE Transactions on Services Computing*.
17. **Hailiang Zhao**, Ziqi Wang, Guanjie Cheng*, Wenzhuo Qian, Peng Chen, Jianwei Yin, Schahram Dustdar, and Shuiguang Deng*. Online Workload Scheduling for Social Welfare Maximization in the Computing Continuum *IEEE Transactions on Services Computing*.
18. Cheng Zhang, Yinuo Deng, **Hailiang Zhao***, Tianlv Chen, and Shuiguang Deng*. Tail-Learning: Adaptive Learning Method for Mitigating Tail Latency in Autonomous Edge Systems *ACM Transactions on Autonomous and Adaptive Systems*.
19. Zhiwei Ling, Yachen Chang, **Hailiang Zhao***, Xinkui Zhao, Kingsum Chow*, and Shuiguang Deng. CADRef: Robust Out-of-Distribution Detection via Class-Aware Decoupled Relative Feature Leveraging *CVPR*.

2024

20. **Hailiang Zhao**, Xueyan Tang*, Peng Chen, and Shuiguang Deng. Learning-Augmented Algorithms for the Bahncard Problem *NeurIPS*.
21. Shuiguang Deng*, **Hailiang Zhao***, Bingbing Huang, Cheng Zhang, Feiyi Chen, Yinuo Deng, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. Cloud-Native Computing: A Survey from the Perspective of Services *Proceedings of the IEEE, Cover Paper*.
22. **Hailiang Zhao**, Shuiguang Deng*, Zhengzhe Xiang, Xueqiang Yan, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. Scheduling Multi-Server Jobs with Sublinear Regrets via Online Learning *IEEE Transactions on Services Computing*.

2023

23. **Hailiang Zhao**, Shuiguang Deng*, Feiyi Chen, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. Learning to Schedule Multi-Server Jobs with Fluctuated Processing Speeds *IEEE Transactions on Parallel and Distributed Systems*.

2022

24. Shuiguang Deng, **Hailiang Zhao**, Zhengzhe Xiang, Cheng Zhang, Rong Jiang, Ying Li*, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. Dependent Function Embedding for Distributed Serverless Edge Computing *IEEE Transactions on Parallel and Distributed Systems*.
25. **Hailiang Zhao**, Shuiguang Deng*, Zijie Liu, Zhengzhe Xiang, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. DPoS: Decentralized, Privacy-Preserving, and Low-Complexity Online Slicing for Multi-Tenant Networks *IEEE Transactions on Mobile Computing*.
26. **Hailiang Zhao**, Shuiguang Deng*, Zijie Liu, Jianwei Yin, and Schahram Dustdar. Distributed Redundant Placement for Microservice-based Applications at the Edge *IEEE Transactions on Services Computing*.

2020

27. Shuiguang Deng, **Hailiang Zhao**, Weijia Fang*, Jianwei Yin, Schahram Dustdar, and Albert Y. Zomaya. Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence *IEEE Internet of Things Journal*.

2019

28. **Hailiang Zhao**, Shuiguang Deng*, Cheng Zhang, Wei Du, Qiang He, and Jianwei Yin. A Mobility-aware Cross-edge Computation Offloading Framework for Partitionable Applications *IEEE ICWS, Best Student Paper Award*.