# ZIYUE LUO

ziyueluocs.github.io ⋄ luo.1457@osu.edu ⋄ +1 (614) 607-1890

## ACADEMIC BACKGROUND

**Ph.D., Computer Science** *September 2018 - September 2022*
The University of Hong Kong, Hong Kong
Supervisor: Prof. Chuan Wu

**B.E., Computer Science and Technology** *September 2014 - June 2018*
Hongyixuetang Class, Wuhan University, China

## PROFESSIONAL EXPERIENCE

**Postdoctoral Researcher** *December 2022 - present*
Dept. of Electrical and Computer Engineering,
The Ohio State University, US

**Research Intern** *July 2019 - September 2020*
Platform for AI (PAI),
Alibaba Group, China

## RESEARCH INTERESTS

▷ Modeling, optimization, design, and implementation of cloud computing systems and distributed machine learning systems
▷ Design, and optimization of distributed machine learning algorithms

## PUBLICATIONS

Ziyue Luo, Jia Liu, Myungjin Lee, Ness B. Shroff. "Prediction-Assisted Online Distributed Deep Learning Workload Scheduling in GPU Clusters," to appear in IEEE INFOCOM, London, United Kingdom, May 19-22, 2025.

Ziyue Luo, Yixin Bao, Chuan Wu. "Optimizing Task Placement and Online Scheduling for Distributed GNN Training Acceleration in Heterogeneous Systems," in IEEE/ACM Transactions on Networking, vol. 32, no. 5, pp. 3715-3729, October 2024.

Bingqian Du, Jun Liu, Ziyue Luo, Chuan Wu, Qiankun Zhang, Hai Jin. "Expediting Distributed GNN Training with Feature-only Partition and Optimized Communication Planning," in IEEE INFOCOM, Vancouver, Canada, May 20-23, 2024.

Ye Tian, Zhen Jia, Ziyue Luo, Yida Wang, Chuan Wu. "DiffusionPipe: Training Large Diffusion Models with Efficient Pipelines," in the Seventh Conference on Machine Learning and Systems (MLSys), Santa Clara, USA, May 13-16, 2024.

Zhe Zhang, Ziyue Luo, Chuan Wu. "Two-level Graph Caching for Expediting Distributed GNN Training," in IEEE INFOCOM, New York, USA, May 17-20, 2023.

Ziyue Luo, Yixin Bao, Chuan Wu. "Optimizing Task Placement and Online Scheduling for Distributed GNN Training Acceleration," in IEEE INFOCOM, online, May 2-5, 2022.

Ziyue Luo, Xiaodong Yi, Guoping Long, Shiqing Fan, Chuan Wu, Wei Lin. "Efficient Pipeline Planning for Expedited Distributed DNN Training," in IEEE INFOCOM, online, May 2-5, 2022.

Jingyan Jiang, Ziyue Luo, Chenghao Hu, Zhaoliang He, Zhi Wang, Shutao Xia, Chuan Wu. "Joint Model and Data Adaptation for Cloud Inference Serving," in the 42th IEEE Real-time Systems Symposuim (RTSS), Dortmund, Germany, December 7-10, 2021.

Xiaodong Yi, Ziyue Luo, Chen Meng, Mengdi Wang, Guoping Long, Chuan Wu, Jun Yang, Wei Lin. "Fast Training of Deep Learning Models over Multiple GPUs," in ACM/IFIP Middleware, Delft, The Netherlands, December 7-11, 2020.

Xiaodong Yi, Shiwei Zhang, Ziyue Luo, Guoping Long, Lansong Diao, Chuan Wu, Zhen Zheng, Jun Yang, Wei Lin. "Optimizing Distributed Training Deployment in Heterogeneous GPU Clusters," in ACM CoNEXT, Barcelona, Spain, December 1-4, 2020.

Ziyue Luo, Chuan Wu. "An Online Algorithm for VNF Service Chain Scaling in Datacenters," in IEEE/ACM Transactions on Networking, vol. 28, No. 3, pp. 1061-1073, June 2020.

Ziyue Luo, Chuan Wu, Zongpeng Li, Wei Zhou. "Scaling Geo-distributed Network Function Chains: A Prediction and Learning Framework," in IEEE Journal on Selected Areas in Communications (JSAC) Special issue on Network Softwarization & Enablers, vol. 37, no. 8, pp. 1838-1850, August 2019.

Ziyue Luo, Zongpeng Li, Chuan Wu. "Online Cloud Resource Allocation and Pricing with Server Speed Scaling," in IEEE ICC 2018, Kansas City, MO, USA, May 20-24, 2018.

## PROFESSIONAL SERVICE

**PC Member** *2025*
IEEE ICDCS

**Reviewer**
IEEE/ACM Transactions on Networking
IEEE Journal on Selected Areas in Communications
IEEE Transactions on Cloud Computing
IEEE Transactions on Cognitive Communications and Networking
IEEE INFOCOM

## HONORS AND AWARDS

▷ INFOCOM Student Conference Grant *2022*
▷ HKU Foundation Postgraduate Fellowship *2018*
▷ Third-Class Scholarship for Outstanding Students *2016*
▷ National Scholarship *2015*
▷ First-Class Scholarship for Outstanding Students *2015*

## TEACHING EXPERIENCE

At the University of Hong Kong:
▷ COMP7404B Computational intelligence and machine learning *2019*
▷ ENGG1330 Computer programming I *2018*