Zhongzhi Yu

Email : zyu401@gatech.edu
Mobile : +1-832-709-3055
Google Scholar: https://scholar.google.com/citations?user=KjvcaBQAAAAJ&hl=en
Linkedin: https://www.linkedin.com/in/zhongzhi-yu

Education

- **Georgia Institute of Technology** — Atlanta, GA
  Ph.D. in Computer Science — *Jan. 2023 - Now*
- **Rice University** — Houston, TX
  Ph.D. in Electrical and Computer Engineering — *Aug 2020 - Dec 2022*
  Transferred to Georgia Institute of Technology with my advisor.
- **Columbia University** — New York, NY
  Master of Science in Electrical Engineering — *Aug 2017 - May 2019*
- **Zhejiang University** — Zhejiang, China
  Bachelor of Engineering in Opto-electronic Information Science and Engineering with honor — *Sep 2013 - June 2017*

Experiences

- **MIT-IBM Watson AI Lab** — Cambridge, MA
  Advisor: Yang Zhang — *May 2022 - Aug 2022*
  Research on developing modular models to equip existing ASR systems with multilingual scalability and low-resource adaptation ability.

Publications

- **Yu, Zhongzhi**, Yang Zhang, Kaizhi Qian, Cheng Wan, Yonggan Fu, Yongan Zhang, and Yingyan (Celine) Lin. "Master-ASR: Achieving Multilingual Scalability and Low-Resource Adaptation in ASR with Modular Learning." In International Conference on Machine Learning (ICML 2023).
- **Yu, Zhongzhi**, Shang Wu, Yonggan Fu, Shunyao Zhang, and Yingyan (Celine) Lin. "Hint-Aug: Drawing Hints from Foundation Vision Transformers Towards Boosted Few-Shot Parameter-Efficient Tuning." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023).
- **Yu, Zhongzhi**, Yonggan Fu, Jiayi Yuan, Haoran You, and Yingyan (Celine) Lin. "NetBooster: Empowering Tiny Deep Learning By Standing on the Shoulders of Deep Giants." In the 2023 60th ACM/IEEE Design Automation Conference (DAC 2023),
- **Yu, Zhongzhi**, Yonggan Fu, Sicheng Li, Chaojian Li, and Yingyan Lin. "MIA-Former: Efficient and Robust Vision Transformers via Multi-Grained Input-Adaptation." In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022), vol. 36, no. 8, pp. 8962-8970. 2022.
- **Yu, Zhongzhi**, Yonggan Fu, Shang Wu, Mengquan Li, Haoran You, and Yingyan Lin. "LDP: Learnable Dynamic Precision for Efficient Deep Neural Network Training and Inference." TinyML Research Symposium 2022.
- Fu, Yonggan*, Yongan Zhang*, **Zhongzhi Yu***, Sixu Li, Zhifan Ye, Chaojian Li, Cheng Wan, and Yingyan (Celine) Lin. "GPT4AIGChip: Towards Next-Generation AI Accelerator Design Automation via Large Language Models." In 2023 IEEE/ACM International Conference On Computer-Aided Design (ICCAD 2023). IEEE, 2023.
- You, Haoran*, Cheng Wan*, Yang Zhao*, **Zhongzhi Yu***, Yonggan Fu, Jiayi Yuan, Shang Wu, Shunyao Zhang, Yongan Zhang, Chaojian Li, Vivek Boominathan, Ashok Veeraraghavan, Ziyun Li, and Yingyan (Celine) Lin. "EyeCoD: eye tracking system acceleration via flatcam-based algorithm and accelerator co-design." In Proceedings of the 49th Annual International Symposium on Computer Architecture (ISCA 2022), pp. 610-622. 2022.
- **Yu, Zhongzhi**, and Yemin Shi. "Centralized Space Learning for open-set computer-aided diagnosis." In Scientific Reports (2023), 13(1), 1630.
- **Yu, Zhongzhi**, and Yemin Shi. "Kernel Quantization for Efficient Network Compression." IEEE Access 10 (2022): 4063-4071.
- Li, Chaojian, **Zhongzhi Yu**, Yonggan Fu, Yongan Zhang, Yang Zhao, Haoran You, Qixuan Yu, Yue Wang, and Yingyan Lin. "HW-NAS-Bench: Hardware-Aware Neural Architecture Search Benchmark." In the 9th International Conference on Learning Representations 2021 (ICLR 2021).
- Fu, Yonggan, Yang Zhang, Kaizhi Qian, Zhifan Ye, **Zhongzhi Yu**, Cheng-I Lai, and Yingyan Lin. "Losses Can Be Blessings: Routing Self-Supervised Speech Representations Towards Efficient Multilingual and Multitask Speech Processing." In Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022), 2022.

- You, Haoran, Zhanyi Sun, Huihong Shi, **Zhongzhi Yu**, Yang Zhao, Yongan Zhang, Chaojian Li, Baopu Li, and Yingyan Lin. "Vitcod: Vision transformer acceleration via dedicated algorithm and accelerator co-design." In 2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA 2023). IEEE, 2023.
- Fu, Yonggan, **Zhongzhi Yu**, Yongan Zhang, Yifan Jiang, Chaojian Li, Yongyuan Liang, Mingchao Jiang, Zhangyang Wang, and Yingyan Lin. "InstantNet: Automated Generation and Deployment of Instantaneously Switchable-Precision Networks." In 2021 58th ACM/IEEE Design Automation Conference (DAC 2021), pp. 757-762. IEEE, 2021.
- Fu, Yonggan, Yongan Zhang, Chaojian Li, **Zhongzhi Yu**, and Yingyan Lin. "A3C-S: Automated Agent Accelerator Co-Search towards Efficient Deep Reinforcement Learning." In 2021 58th ACM/IEEE Design Automation Conference (DAC 2021), pp. 13-18. IEEE, 2021.
- Li, Mengquan, **Zhongzhi Yu**, Yongan Zhang, Yonggan Fu, and Yingyan Lin. "O-HAS: Optical hardware accelerator search for boosting both acceleration performance and development speed." In 2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD 2021), pp. 1-9. IEEE, 2021.
- Fu, Yonggan, **Zhongzhi Yu**, Yongan Zhang, and Yingyan Lin. "Auto-agent-distiller: Towards efficient deep reinforcement learning agents via neural architecture search." arXiv preprint arXiv:2012.13091 (2020).
- Zhao, Guangyuan, Mohammad M. Kabir, Kimani C. Toussaint, Cuifang Kuang, Cheng Zheng, **Zhongzhi Yu**, and Xu Liu. "Saturated absorption competition microscopy." Optica 4, no. 6 (2017): 633-636.
- **Yu, Zhongzhi**, Shaocong Liu, Dazhao Zhu, Cuifang Kuang, and Xu Liu. "Parallel detecting super-resolution microscopy using correlation based image restoration." Optics Communications 404 (2017): 139-146.

Awards

- Won second place in University Best Demonstration at DAC 2023

Services

Served as Reviewer for AICAS 2022, NeurIPS 2022, AAAI 2023, CVPR 2023, ICML 2023, and NeurIPS 2023