

HAORAN YOU

Tel.: +1-281-236-1978 Email: haoran.you@gatech.edu Profile: [Homepage](#); [Linkedin](#); [Github](#)

EDUCATION

- **Georgia Institute of Technology** Atlanta, Georgia
Transferred CS Ph.D. supervised by Prof. Yingyan Lin Dec. 2022 - Present
- **Rice University** Houston, Texas
ECE M.S. & Ph.D. supervised by Prof. Yingyan Lin Sep. 2019 - Dec. 2022
- **Huazhong University of Science and Technology** Wuhan, China
ECE B.S. Advanced Class of 2015 (Overall GPA: 90.01/100; Ranking: 2/30) Sep. 2015 - Jul. 2019

RESEARCH INTERESTS

- Efficient and Automated Machine Learning (ICML, NeurIPS, ICLR, AAAI, CVPR, ECCV, TNNLS, TMLR)
- Algorithm and Hardware Accelerator Co-Design (ISCA, HPCA, MICRO, ICCAD, ISLPED)

INDUSTRY EXPERIENCE

- **Research Intern at Adobe Research, Seattle, WA** May. 2024 - Aug. 2024
Mentors: Dr. Sohrab Amirghodsi, Yuqian Zhou, Connelly Barnes, Eli Shechtman
Internship focus: Efficient GenAI Research
- **Startup Founder at CREATE-X & VentureLab, Georgia Tech, Atlanta, GA** May. 2023 - Aug. 2023
Embark on an entrepreneurial experience in the startup launch program
Co-founders: Dr. Yang Zhao, Yingyan Lin. Idea: EyeCoD [IEEE Micro's Top Pick]
- **Research Intern at Meta Reality Lab (Team Now: Meta GenAI), Burlingame, CA** May. 2022 - Dec. 2022
Mentors: Dr. Peizhao Zhang, Yunyang Xiong, Xiaoliang Dai, Bichen Wu, Peter Vajda
Publication during internship: Caslting-ViT [CVPR'23]
- **Research Intern at Baidu Research, Sunnyvale, CA** May. 2021 - Apr. 2022
Under the instruction of Dr. Baopu Li
Publications during internship: ShiftAddNAS [ICML'22], SuperTickets [ECCV'22]

PUBLICATIONS

- **H. You, C. Li, P. Xu, Y. Fu, Y. Wang, X. Chen, R.G. Baraniuk, Z. Wang, Y. Lin.**
Selected as an ICLR Spotlight ! [Link]
Drawing Early-Bird Tickets: Towards More Efficient Training of Deep Networks. ICLR 2020 Spotlight Oral
- **H. You, Z. Sun, H. Shi, Z. Yu, Y. Zhao, Y. Zhang, C. Li, B. Li, Y. Lin.**
Selected as the Meta Faculty Research Award of 2022 ! [Link]
ViTCoD: Vision Transformer Acceleration via Dedicated Algorithm and Accelerator Co-Design. HPCA 2023
- **H. You, Y. Zhao, Z. Yu, C. Wan, Y. Fu, C. Li, S. Zhang, S. Wu, J. Yuan, Y. Zhang, Vivek B., Ashok V., Z. Li, Y. Lin.**
Selected as the IEEE Micro's Top Pick of 2023 ! [Link]
EyeCoD: Eye Tracking System Acceleration via FlatCam-Based Algorithm and Accelerator Co-Design. ISCA 2022
- **H. You, Y. Guo, Y. Fu, W. Zhou, H. Shi, X. Zhang, S. Kundu, A. Yazdanbakhsh, Y. Lin.**
ShiftAddLLM: Accelerating Pretrained LLMs via Post-Training Multiplication-Less Reparameterization. Preprint

- **H. You**, Y. Fu, Z. Wang, A. Yazdanbakhsh, Y. Lin.
When Linear Attention Meets Autoregressive Decoding: Towards More Effective and Efficient Linearized LLMs. **ICML 2024**
- **H. You**, H. Shi, Y. Guo, Y. Lin.
ShiftAddViT: Mixture of Multiplication Primitives Towards Efficient Vision Transformer. **NeurIPS 2023**
- **H. You**, Y. Xiong, X. Dai, B. Wu, P. Zhang, H. Fan, P. Vajda, Y. Lin.
Castling-ViT: Compressing Attention via Switching Towards Linear-Angular Attention at ViT Inference. **CVPR 2023**
- **H. You**, B. Li, Z. Sun, X. Ouyang, Y. Lin
SuperTickets: Drawing Task-Agnostic Lottery Tickets from Supernets via Jointly Searching and Pruning. **ECCV 2022**
- **H. You**, B. Li, H. Shi, Y. Lin
ShiftAddNAS: Hardware-Inspired Search for More Accurate and Efficient Neural Networks. **ICML 2022**
- **H. You**, Z. Lu, Z. Zhou, Y. Fu, Y. Lin.
Early-Bird GCNs: Graph-Network Co-Optimization Towards More Efficient GCN Training and Inference. **AAAI 2022**
- **H. You**, T. Geng, Y. Zhang, A. Li, Y. Lin.
GCoD: Graph Convolutional Network Acceleration via Dedicated Algorithm and Accelerator Co-Design. **HPCA 2022**
- **H. You**, X. Chen, Y. Zhang, C. Li, S. Li, Z. Liu, Z. Wang, Y. Lin
ShiftAddNet: A Hardware-Inspired Deep Network. **NeurIPS 2020**
- **H. You**, R. Balestriero, Z. Lu, Y. Kou, Y. Lin, R.G. Baraniuk.
Max-Affine Spline Insights Into Deep Network Pruning. **TMLR 2022**
- **H. You**, Y. Cheng, T. Cheng, C. Li, P. Zhou.
Bayesian Cycle-Consistent GANs via Marginalizing Latent Sampling. **IEEE TNNLS**
- Y. Fu, **H. You**, Y. Zhao, Y. Wang, C. Li, K. Gopalakrishnan, Z. Wang, Y. Lin.
FracTrain: Fractionally Squeezing Bits Both Temporally and Spatially for Efficient DNN Training. **NeurIPS 2020**
- H. Shi, **H. You**, Y. Zhao, Z. Wang, Y. Lin.
NASA: Neural Architecture Search and Acceleration for Hardware Inspired Hybrid Networks. **ICCAD 2022**
- Y. Zhang, **H. You**, Y. Fu, T. Geng, A. Li, Y. Lin
G-CoS: GNN-Accelerator Co-Search Towards Both Better Accuracy and Efficiency. **ICCAD 2021**
- C. Li, T. Chen, **H. You**, Z. Wang, Y. Lin.
HALO: Hardware-Aware Learning to Optimize. **ECCV 2020**
- Y. Zhao, Z. Li, **H. You**, Y. Fu, Y. Zhang, C. Li, C. Wan, S. Wu, X. Ouyang, V. Boominathan, A. Veeraraghavan, Y. Lin.
Won the first place in University Best Demonstration at DAC 2022 ! [Link]
i-FlatCam: A 253 FPS, 91.49 μ J/Frame Ultra-Compact Intelligent Lensless Camera Eye Tracking System. **VLSI 2022**
- Y. Zhao, X. Chen, Y. Wang, C. Li, **H. You**, Y. Fu, Y. Xie, Z. Wang, Y. Lin.
SmartExchange: Trading Higher-cost Memory Storage/Access for Lower-cost Computation. **ISCA 2020**
- T. Geng, C. Wu, Y. Zhang, C. Tan, C. Xie, **H. You**, M. Herbordt, Y. Lin, A. Li
I-GCN: A GCN Accelerator with Runtime Locality Enhancement through Islandization. **MICRO 2021**
- Z. Yu, Z. Wang, Y. Li, **H. You**, R. Gao, X. Zhou, S. Bommu, Y. Zhao, Y. Lin
EDGE-LLM: Enabling Efficient LLM Adaptation on Edge Devices via Layerwise Unified Compression and Adaptive Layer Tuning & Voting. **DAC 2024**
- X. Chen, Y. Zhao, Y. Wang, P. Xu, **H. You**, C. Li, Y. Fu, Y. Lin, Z. Wang
SmartDeal: Re-Modeling Deep Network Weights for Efficient Inference and Training. **IEEE TNNLS**

- Y. Zhang, Y. Fu, W. Jiang, C. Li, **H. You**, M. Li, V. Chandra, Y. Lin
DIAN: differentiable accelerator-network co-search towards maximal DNN efficiency. **ISLPED 2021**
- C. Li, Z. Yu, Y. Fu, Y. Zhang, Y. Zhao, **H. You**, Q. Yu, Y. Wang, C. Hao, Y. Lin.
HW-NAS-Bench: Hardware-Aware Neural Architecture Search Benchmark. **ICLR 2021 Spotlight Oral**
- Y. Fu, Z. Ye, J. Yuan, S. Zhang, S. Li, **H. You**, Y. Lin.
Gen-NeRF: Efficient and Generalizable Neural Radiance Fields via Algorithm-Hardware Co-Design. **ISCA 2023**
- S. Li, C. Li, W. Zhu, B. Yu, Y. Zhao, C. Wan, **H. You**, H. Shi, Y. Lin
Instant-3D: Instant Neural Radiance Fields Training Towards Real-Time AR/VR 3D Reconstruction. **ISCA 2023**
- Z. Yu, Y. Fu, S. Wu, M. Li, **H. You**, Y. Lin.
LDP: Learnable Dynamic Precision for Efficient Deep Neural Network Training and Inference. **TinyML 2022**

SELECTED AWARDS

- Haoran received the NeurIPS 2023 Scholar Award ! Dec. 2023
- Haoran won First Place in [ACM Student Research Competition \(SRC\) at ICCAD](#) ! Nov. 2023
- ICCAD 2023 Student Travel Grant ! Oct. 2023
- Haoran was selected as one of the [ML and Systems Rising Stars of 2023](#) ! May. 2023
- Haoran won the [Best Poster Award](#) at the SCS Poster Competition of GaTech ! Apr. 2023
- Haoran was selected by GaTech to receive the [Outstanding Graduate Research Assistant Award](#) ! Mar. 2023
- [EyeCoD \(ISCA'22\)](#) was selected for being the [IEEE Micro's Top Pick of 2023](#) ! Jan. 2023
- HPCA 2023 Student Travel Grant ! 2022 - 2023
- Haoran got PhD Growth Annual Rock Star Award in EIC Lab ! 2022 - 2023
- [ViTCoD \(HPCA'23\)](#) was selected for [Meta Faculty Research Award of 2022](#) ! Nov. 2022
- Our [i-FlatCam](#) won First Place in [University Demo Best Demonstration at DAC Conference](#) ! Jul. 2022
- ISCA 2022 Student Trave Grant ! Jun. 2022
- Samsung Scholarship ! 2017 - 2018
- International Interdisciplinary Contest in Modeling (ICM), Meritorious Winner ! Apr. 2018
- Selected as *Outstanding Undergraduates in Term of Academic Performance*, (HUST 1%) ! Dec. 2017
- The Eighth Chinese Mathematics Competition (CMC), First Price (China 5%) ! Nov. 2016

CONFERENCE PRESENTATION

- The Second CoCoSys Annual Review, 2024 [\[Link\]](#)
- The 37th Conference on Neural Information Processing Systems (NeurIPS), 2023 [\[Link\]](#)
- Semiconductor Research Corporation (SRC) TECHCON, 2023 [\[Link\]](#)
- The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023 [\[Link\]](#)
- The 50th International Symposium on Computer Architecture (ISCA), 2023 [\[Link\]](#)
- The 29th International Symposium on High-Performance Computer Architecture (HPCA), 2023 [\[Link\]](#)
- European Conference on Computer Vision (ECCV), 2022 [\[Link\]](#)
- The 49th International Symposium on Computer Architecture (ISCA), 2022 [\[Link\]](#)
- The 39th International Conference on Machine Learning (ICML), 2022 [\[Link\]](#)
- The 28th International Symposium on High-Performance Computer Architecture (HPCA), 2022 [\[Link\]](#)
- The 36th AAAI Conference on Artificial Intelligence (AAAI), 2022 [\[Link\]](#)
- The 8th International Conference on Learning Representations (ICLR), 2020 [\[Link\]](#)
- The 34th Conference on Neural Information Processing Systems (NeurIPS), 2020 [\[Link\]](#)

INVITED TALKS

- Invited Speaker at Georgia Tech ECE 8803 SW/HW Co-Design Course, 2024 [\[Link\]](#)
- Invited Speaker at Google, 2023 [\[Link\]](#)
- Invited Speaker at University of Rochester IntelliArch Lab, 2023 [\[Link\]](#)
- Invited Speaker at ISCA AutoDL Workshop, 2023 [\[Link\]](#)
- Invited Speaker at UT Austin Energy-Aware Computing Group, 2023 [\[Link\]](#)
- Invited Speaker at the Tutorial in Asian Conference on Computer Vision, 2022 [\[Link\]](#)
- Invited Speaker at the Ken Kennedy Institute Data Science Conference, 2021 [\[Link\]](#)

MENTORING EXPERIENCE

- Zhanyi Sun (B.S. Student), Published two 2nd author HPCA'23 and ECCV'22 papers, Next move: MS@CMU
- Zhihan Lu (B.S. Student), Published two 2nd author AAI'22 and TMLR journal paper, Next move: MS@CMU
- Shang Wu (M.S. Student), Published co-authored ISCA'22/TMLR, Next move: PhD@Northwestern University
- Xu Ouyang (M.S. Student), Published one co-authored ECCV'22 paper, Next move: PhD@University of Virginia
- Zijian Zhou (B.S. Student), Published one co-authored AAI'22 paper, Next move: MS@Rice University
- Yichao Fu (M.S. Student), Published one 2nd authored ICML'24 paper, Next move: PhD@UCSD
- Yi Sun (B.S. Student), Submitted a co-authored paper, Next move: MS@CMU
- Yutong Kou (B.S. Student), Published a co-authored paper, Next move: PhD@CAS
- Yipin Guo (M.S. Student), Published one co-1st author NeurIPS'23 paper, Current position: MS@ZJU
- Wei Zhou (B.S. Student), Submitted a co-authored paper, Next move: MS@GaTech

REVIEWER SERVICES

- 2024: ICLR; AAI, CVPR, ICML, ECCV, NeurIPS
- 2023: ICLR; CVPR; MLSys; ICML; ICCV; NeurIPS; IEEE TCAS-II
- 2022: CVPR; ICML; ICLR; NeurIPS; ECCV; AAI; JCST
- 2021: ICML; NeurIPS; ICLR; MLSys; IEEE TNNLS
- 2020: ICML; ICLR; CVPR

TEACHING EXPERIENCE

- | | |
|---|-------------|
| • TA at CS 4803/8803: Efficient Machine Learning, Georgia Tech | Spring 2024 |
| • TA at ELEC 220: Fundamentals of Computer Engineering, Rice University | Fall 2021 |
| • TA at ELEC 220: Fundamentals of Computer Engineering, Rice University | Spring 2021 |
| • TA at ELEC 220: Fundamentals of Computer Engineering, Rice University | Fall 2020 |

VOLUNTEERING SERVICE

- Student organization helper and website developer for the [ICCAD HALO Workshop](#)
- Student organization helper and website developer for the [EECS Rising Stars Workshop of 2023](#)