My Research: Develop empirical methods to make AI faster and more efficient.

High Performance Computing, Machine Learning, Optimization, AI Infrastructure, ML Systems.

Basic Information of Yang You:

Presidential Young Professor, National University of Singapore

Homepage: https://www.comp.nus.edu.sg/~youy/

Email: yangyou@nus.edu.sg

Phone: +65 86488197

(a) Education

- Ph.D. in Computer Science, UC Berkeley, 08/2015 - 08/2020

Nominated for ACM Doctoral Dissertation Award (top 2 of 81 Berkeley PhD Graduates)

- MS in Computer Science, Tsinghua University, 08/2012 - 07/2015

Ranked 1st among 134 students in the selection for Outstanding Graduates

Siebel Scholarship (top 5 of 134 students): highest-value scholarship in Tsinghua CS

- BS in Computer Science, China Agricultural University, 08/2009 - 07/2012

Ranked 1st among 52 students in the Honors Program (the most selective program)

Ranked top 0.3% in 2009 Gaokao (Nationwide College Entrance Exam) in my province

Finished the 4-year undergraduate program in 3 years (with highest honors)

(b) Appointments

- Presidential Young Professor, Computer Science, NUS, 01/2021 - present

a prestigious and highly competitive tenure-track position for promising early-career faculty

- Assistant Professor (Tenure-Track), Computer Science, NUS, 08/2020 - 01/2021

(c) Startup

- Founder & Chairman of HPC-AI Tech: raised 70 million USD, 08/2021 - present

Selected Awards

- I am the corresponding author, and my student is the first author.
    - AAAI'23 Distinguished Paper Award (Top 0.14% of all submissions)
    - ACL'23 Outstanding Paper Award (Top 0.8% of all submissions)
    - ICLR'24 Oral Paper (Top 1.2% of all submissions)
    - CVPR'22 Oral Paper (Top 4% of all submissions)
- I am the first author.
    - ICDM'19 Best Paper Candidate (Top 0.8% of all submissions)
    - ICPP'18 Best Paper Award (Top 0.3% of all submissions), plenary presentation
    - IPDPS'15 Best Paper Award (Top 0.8% of all submissions), plenary presentation
    - SC'19 Best Student Paper Finalist (Supercomputing)
- [12/2025] NeurIPS Oral Paper (77 out of 21575 submissions: Top 0.36%)
- [11/2025] Best Mentor Award, ByteDance Scholarship Program
- [09/2025] World's Top 2% Scientists by Stanford/Elsevier (2023, 2024, 2025)
- [04/2024] Intelligent Computing Innovator of the Year by MIT Technology Review (China)
- [10/2023] Top 100 Most Influential Chinese by Forbes China
- [08/2022] Singapore Maritime R&D Grant Award on Maritime AI (~1 million USD)
- [09/2021] IEEE-CS Early Career Researcher Award for Excellence in HPC
- [07/2021] ACM SIGHPC Outstanding Doctoral Dissertation Award (Honorable Mention)
- [04/2021] Forbes 30 under 30 Asia list (https://www.forbes.com/profile/you-yang/)
- [01/2021] Presidential Young Professor (early career track for exceptional young talents)
- [10/2020] ACM Doctoral Dissertation Award Nomination by UC Berkeley: 2 out of 81 Berkeley EECS PhD students graduated in 2020
- [09/2020] Top Reviewer of International Conference on Machine Learning (ICML)
- [04/2020] Lotfi A. Zadeh Prize: recognizes a Berkeley graduating PhD student who has made outstanding contributions to soft computing and its applications
- [11/2017] ACM George Michael Memorial HPC Fellowship: the only PhD fellowship on ACM homepage in 2017
- [07/2015] Outstanding Graduate of Tsinghua University (ranked top 1 of 134 students)
- [07/2015] Outstanding Graduate of Beijing (ranked top 1 of 134 students, top 4 got the awards)
- [07/2015] Outstanding Graduate of Tsinghua CS department (ranked top 1 of 134 students, top 20 got the awards)

- [07/2015] Best Thesis Award of Tsinghua University (10 out of 134 students: 7%)
- [10/2014] Siebel Scholar (top 5 of 134 students): highest-value scholarship (35,000 USD) in Tsinghua CS Department
- [09/2011] National Scholarships of China (ranked top 1 of 52 students, top 2 got the award)
- [09/2010] National Scholarships of China (ranked top 1 of 52 students, top 2 got the award)

Top-5 Publications by Yang You

---

1. [ICLR'20] Yang You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.-J. Hsieh. "Large Batch Optimization for Deep Learning: Training BERT in 76 minutes."

International Conference on Learning Representations

1370 citations. The method has been used by Google, Meta, and Microsoft and included in MLPerf.

I am the first author and lead the project.

2. [ICLR'23] Z. Qin, K. Wang, Z. Zheng, J. Gu, X. Peng, Z. Xu, D. Zhou, L. Shang, B. Sun, X. Xie, Yang You. "InfoBatch: Lossless Training Speed Up by Unbiased Dynamic Data Pruning"

International Conference on Learning Representations

ORAL paper: 1.2% acceptance rate

I am the corresponding author, and my student is the first author.

3. [ICPP'18] Yang You, Z. Zhang, C.-J. Hsieh, J. Demmel, K. Keutzer. "ImageNet Training in Minutes"

47th International Conference on Parallel Processing.

Best Paper Award (1 out of 313 submissions: Top 0.3%); The most cited HPC conference paper (HPDC, ICS, ICPP, IPDPS, PPOPP, SC, etc.) published between 2018 and 2020 (according to Google Scholar).

I am the first author and lead the project.

4. [ACL'23] Yang Luo, Xiaozhe REN, Zangwei Zheng, Xin Jiang, ZHUO JIANG and Yang You.

"CAME: Confidence-guided Adaptive Memory Efficient Optimization"

Annual Meeting of the Association for Computational Linguistics

Outstanding Paper Award: 39 out of 4562 submissions, Top 0.86%

I am the corresponding author, and my student is the first author.

5. [ICPP'23] Li, Shenggui, Jiarui Fang, Zhengda Bian, Hongxin Liu, Yuliang Liu, Haichen Huang, Boxiang

Wang, and Yang You.

"Colossal-AI: A unified deep learning system for large-scale parallel training." ICPP 2023.

Colossal-AI got 41,000 GitHub stars. It has been used by 100+ big enterprises globally including IBM, Meta, Oracle, HPE, Walmart, Alibaba, etc.

https://github.com/hpcaitech/ColossalAI

I am the corresponding author, and my student is the first author.

Students Supervised

Graduated PhD Students:

- **PENG XIANGYU** – First-Author of CVPR Oral Paper (Top 4%)
- **SHEN CHENHUI** – Alibaba Scholarship
- **WANG KAI** – AI Singapore PhD Fellowship
- **XUE FUZHAO** – First-Author of ICML (3x)/NeurIPS/AAAI Papers, Google PhD Fellowship
- **ZHENG ZANGWEI** – First-Author of AAAI Distinguished Paper (Top 0.14%)
- **QIN ZIHENG** – First-Author of ICLR Oral Paper (Top 1.2%), Research Achievement Award
- **LIN RUIXI** – Dean's Graduate Fellowship
- **ZHAO WANGBO** – First-Author of ICLR/CVPR (3x)/NeurIPS Papers, Google PhD Fellowship, Research Achievement Award

Active PhD Students:

- **CHENG SHENGGAN** – First-Author of ASPLOS/PPoPP Papers
- **HUANG HAICHEN** – Key Contributor of Colossal-AI project
- **LIU YONG** – First-Author Papers of ICLR/ICML/NeurIPS (2x)/CVPR, Dean's Graduate Research Excellence Award
- **LIU ZIMING** – “Stars of Tomorrow” Award from Microsoft Research, Bytedance PhD fellowship, Research Achievement Award
- **LOU YUXUAN** – Key Contributor of Colossal-AI project
- **LUO YANG** – First-Author of ACL Outstanding Paper (Top 0.8%), Research Achievement Award
- **SONG YANG** – First-Author of an EMNLP paper
- **WANG ZIQIAO** – First-Author of a NeurIPS Paper
- **ZHANG GENG** – Key Contributor of VideoSys project
- **ZHANG YIQI** – Presidential Graduate Fellowship
- **ZHAO XUANLEI** – First-Author of ICLR (2x)/ICML/MLSys Papers

- **ZHOU Pengfei** – First-Author of a NeurIPS Paper, Research Achievement Award
- **ZHU ZIRUI** – First-Author of a WWW paper

Master's Student	**Position After Leaving NUS**
Raphaël Perot	Ph.D. Student at ETH Zürich
Bian Zhengda	Chief Technology Officer (CTO), Startup
Mai Siqi	Vice President (VP), Startup
Liang Bowen	Engineer at a Fortune 500 Company
Liu Hongxin	Engineering Leader at a Tech Company
Li Yongbin	Chief Operating Officer (COO), Startup

Academic Services

PC Vice Co-Chairs of IEEE BigData 2025 (Track: Foundation Models for Big Data)

Senior Area Chair: AAAI’26

Area Chair: CVPR'24, SC'24

Senior PC member: IJCAI'17, AAAI’26

Reviewer of USA NSF SBIR/STTR funding program

Reviewer:

ICLR’26, ASPLOS'26, NeurIPS'25, ICML'25, ICLR'25, ASPLOS’25, ACL’25, CVPR’25, ICCV’25, NeurIPS'23, ICML'23, TPDS, MLSys'23, JMLR, SIGMOD'23, NeurIPS'22, JPDC, CVPR'22, ISC'22, AAAI'22, ICLR'22, IJCAI'22, NeurIPS'21, ICML'21, JEA'21, NSF SBIR/STTR, AAAI'21, SC'21, ISC'21, IPDPS'21, ICLR'21, TNNLS'20, NeurIPS'20, HiPC'20, IEEE Access'20, TKDE'20, ICML'20, TBD'20, IPDPS'20, T-SP'19, TOPC'19, IBM'19, ICPP'18, CCGRID'18, JMLR'17

Organizer for Workshop on Dataset Distillation Challenges at ECCV 2024

Session Chair: ICPP'21, ICPP'22, AAAI'23

PhD Thesis Committee:

SHEN QIANLI (NUS Computer Science PhD)

JI BO (NUS Computer Science PhD)

JIANG MINGLIANG (NUS Information Science PhD)

FAN XIAOFENG (NUS Computer Science PhD)

DU JIAWEI (NUS Computer Science PhD)

QIAO RUI (NUS Computer Science PhD)

Teaching

[2020 Fall]

NUS CS6285: Bridging System and Deep Learning
https://www.comp.nus.edu.sg/~youy/teach/CS6285.pdf

[2021 Spring]

NUS CS5260: Neural Networks and Deep Learning II
https://www.comp.nus.edu.sg/~youy/teach/CS5260.pdf

[2022 Spring]

NUS CS5260: Neural Networks and Deep Learning II
https://www.comp.nus.edu.sg/~youy/teach/CS5260Y2022.pdf

[2022 Fall]

NUS CS5242: Neural Networks and Deep Learning I
https://www.comp.nus.edu.sg/~youy/teach/CS5242Y2022.pdf

[2022 Nov]

Tutorial on Open Data Science Conference (ODSC)

How to Use Large AI Models at Low Costs

https://odsc.medium.com/how-to-use-large-ai-models-at-low-costs-5aef8bc1c67e

[2022 Nov]

Tutorial on Supercomputing conference (SC)

Colossal-AI: Scaling Large AI Models on Distributed Systems and Supercomputers

https://sc22.supercomputing.org/presentation/?id=tut129&sess=sess211

[2023 Feb]

Tutorial on AAAI Conference on Artificial Intelligence (AAAI)

Large-scale Deep Learning Optimization Techniques

https://aaai-23.aaai.org/aaai23tutorials/

[2023 Feb]

Tutorial on AAAI Conference on Artificial Intelligence (AAAI)

Colossal-AI: Scaling AI Models in Big Model Era

https://aaai-23.aaai.org/aaai23tutorials/

[2023 Feb]

Tutorial on Principles and Practice of Parallel Programming (PPoPP)

Large-scale Deep Learning Optimization Techniques

https://conf.researchr.org/home/PPoPP-2023

[2023 July]

Colossal-AI: Breakthroughs in Efficient AI

International Conference on Machine Learning (ICML)

https://icml.cc/Expo/Conferences/2023/talk%20panel/25681

[2023 July]

Large-scale Deep Learning Optimization Techniques

Conference on Computer Vision and Pattern Recognition (CVPR)

https://cvpr.thecvf.com/Conferences/2023/tutorial-list

[2023 December]

Contributing to an Efficient and Democratized Large Model Era

Conference on Neural Information Processing Systems (NeurIPS)

https://nips.cc/virtual/2023/tutorial/73956

[2023 Spring]

NUS CS5260: Neural Networks and Deep Learning II
https://docs.google.com/document/d/1onycbNTlsW-XflfFIFPO9fV_NYGhJkM36JpSlWg4aC0/edit?usp=sharing

[2024 Spring]

NUS CS5260: Neural Networks and Deep Learning II

https://docs.google.com/document/d/15sa6LMw0sx66dxNrjlmAS0cMQtbS1Sm9mJSN7ND1T6E/edit?usp=sharing

[2024 Spring]

NUS CS5242: Neural Networks and Deep Learning I

https://docs.google.com/document/d/11keR5h-2KTGq7HdWBo8YCLURd0JbTch30JhzX2KkrA4/edit?usp=sharing

[2024 Fall]

NUS CS5242: Neural Networks and Deep Learning I

https://docs.google.com/document/d/17wn8qyiFRqYzgO3cc60-Bs_YyJ5f_y1bXCFuO5rPLX0/edit?usp=sharing

[2025 Spring]

NUS CS5260: Neural Networks and Deep Learning II

https://docs.google.com/document/d/1IBtGeyvYCnOwWxgaOGy6c29FWC-UlNJT3KIwD7wQxQ0/edit?usp=sharing

[2025 March]

Open-Sora: Training a Commercial-Level Video Generation Model with $200k

NVIDIA GTC conference

https://www.nvidia.com/en-us/on-demand/session/gtc25-s72728/

Publications after Yang You got his PhD (after August 2020)

1.

CORE rankings: A [SC'21]

Bian, Zhengda, Shenggui Li, Wei Wang, and Yang You.

"Online evolutionary batch size orchestration for scheduling deep learning workloads in GPU clusters."

In Proceedings of the International Conference for High-Performance Computing, Networking, Storage and Analysis, pp. 1-15. 2021.

I am the corresponding author, and my student is the first author.

2.

Impact Factor: 8.793 [TNNLS'21]

Ma, Jie, Jun Liu, Qika Lin, Bei Wu, Yaxian Wang, and Yang You.

"Multitask Learning for Visual Question Answering."

IEEE Transactions on Neural Networks and Learning Systems (2021).

I am the corresponding author, and my student is the first author.

3.

Largest supercomputing conference in Europe [ISC'21]

Han, Ruobing, James Demmel, and Yang You.

"Auto-precision scaling for distributed deep learning."

In International Conference on High-Performance Computing, pp. 79-97. Springer, Cham, 2021.

I am the corresponding author, and my student is the first author.

4.

CORE rankings: A* [ICLR'22]

Liu, Yong, Xiangning Chen, Minhao Cheng, Cho-Jui Hsieh, and Yang You.

"Concurrent Adversarial Learning for Large-Batch Training."

International Conference on Learning Representations 2022.

I am the corresponding author, and my student is the first author.

5.

CORE rankings: A [ICPP'22]

Boxiang Wang, Qifan Xu, Zhengda Bian, Yang You.

"Tesseract: Parallelize the Tensor Parallelism Efficiently".

International Conference on Parallel Processing (ICPP)

I am the corresponding author, and my student is the first author.

6.

CORE rankings: A* [AAAI'22]

Xue, Fuzhao, Ziji Shi, Futao Wei, Yuxuan Lou, Yong Liu, and Yang You.

"Go wider instead of deeper."

AAAI Conference on Artificial Intelligence 2022.

I am the corresponding author, and my student is the first author.

7.

CORE rankings: A* [ACL'22]

Shen, Chenhui, Liying Cheng, Ran Zhou, Lidong Bing, Yang You, and Luo Si.

"MReD: A Meta-Review Dataset for Controllable Text Generation."

Annual Meeting of the Association for Computational Linguistics 2022.

8.

Impact Factor: 3.816 [IJAR'22]

Gong, Chaoyu, Zhi-gang Su, Pei-hong Wang, Qian Wang, and Yang You.

"Evidential instance selection for K-nearest neighbor classification of big data."

International Journal of Approximate Reasoning 138 (2021): 123-144.

I am the corresponding author, and my student is the first author.

9.

Wang, Kai, Shuo Wang, Jianfei Yang, Xiaobo Wang, Baigui Sun, Hao Li, and Yang You.

"Mask Aware Network for Masked Face Recognition in the Wild."

IEEE/CVF International Conference on Computer Vision (workshop), pp. 1456-1461. 2021.

I am the corresponding author, and my student is the first author.

10.

Impact Factor: 6.093 [TKDE'22]

Gong, Chaoyu, Zhi-gang Su, Pei-hong Wang, Qian Wang, and Yang You.

"A Sparse Reconstructive Evidential-Nearest Neighbor Classifier for High-dimensional Data."

IEEE Transactions on Knowledge and Data Engineering (2022).

I am the corresponding author, and my student is the first author.

11.

Impact Factor: 9.602 [ESWA'22]

Gong, Chaoyu, Zhi-gang Su, Pei-hong Wang, and Yang You.

"Distributed evidential clustering toward time series with big data issues."

Expert Systems with Applications 191 (2022): 116279.

I am the corresponding author, and my student is the first author.

12.

CORE rankings: A* [CVPR'22]

Wang, Kai, Bo Zhao, Xiangyu Peng, Zheng Zhu, Shuo Yang, Shuo Wang, Guan Huang, Hakan Bilen, Xinchao Wang, and Yang You.

"CAFE: Learning to Condense Dataset by Aligning Features."

IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.

I am the corresponding author, and my student is the first author.

13.

CORE rankings: A* [CVPR'22]

Peng, Xiangyu, Kai Wang, Zheng Zhu, and Yang You.

"Crafting Better Contrastive Views for Siamese Representation Learning."

IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.

CVPR Oral Acceptance Rate is ~4%

I am the corresponding author, and my student is the first author.

14.

CORE rankings: A* [CVPR'22]

Wang, Kai, Shuo Wang, Zhipeng Zhou, Xiaobo Wang, Xiaojiang Peng, Baigui Sun, Hao Li, and Yang You.

"An efficient training approach for very large scale face recognition."

IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.

I am the corresponding author, and my student is the first author.

15.

CORE rankings: A* [CVPR'22]

Liu, Yong, Siqi Mai, Xiangning Chen, Cho-Jui Hsieh, and Yang You.

"Towards Efficient and Scalable Sharpness-Aware Minimization."

IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.
I am the corresponding author, and my student is the first author.

16.
CORE rankings: A* [CVPR'22]
Zhao, Wangbo, Kai Wang, Xiangxiang Chu, Fuzhao Xue, Xinchao Wang, and Yang You.
"Modeling Motion with Multi-Modal Features for Text-Based Video Segmentation"
IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.
I am the corresponding author, and my student is the first author.

17.
CORE rankings: A* [ICDE'22]
Gong, Chaoyu, Yongbin Li, Yong Liu, Pei-hong Wang, Yang You.
"Joint Evidential K-Nearest Neighbor Classification."
In IEEE International conference on data engineering (ICDE), 2022.
I am the corresponding author, and my student is the first author.

18.
CORE rankings: A* [ICDE'22]
Gong, Chaoyu, Yongbin Li, Di Fu, Yong Liu, Pei-hong Wang, Yang You.
"Self-reconstructive evidential clustering for high-dimensional data."
In IEEE International conference on data engineering (ICDE), 2022.
I am the corresponding author, and my student is the first author.

19.
CORE rankings: A [ICS'22]
Du, Jiangsu, Jiazhi Jiang, Yang You, Dan Huang, and Yutong Lu.
"Handling heavy-tailed input of transformer inference on GPUs."
In Proceedings of the 36th ACM International Conference on Supercomputing, pp. 1-11. 2022.

20.

21.

CORE rankings: A [ICPP'23]

Li, Shenggui, Jiarui Fang, Zhengda Bian, Hongxin Liu, Yuliang Liu, Haichen Huang, Boxiang Wang, and Yang You. "Colossal-AI: A unified deep learning system for large-scale parallel training." ICPP 2023.

Colossal-AI got 41,000 GitHub stars. It has been used by 100+ big enterprises globally including IBM, Oracle, HPE, Walmart, etc.

https://github.com/hpcaitech/ColossalAI

I am the corresponding author, and my student is the first author.

22.

CORE rankings: A* [NeurIPS'22]

Yong Liu, Siqi Mai, Minhao Cheng, Xiangning Chen, Cho-Jui Hsieh, Yang You.

Random Sharpness-Aware Minimization

Conference on Neural Information Processing Systems 2022.

I am the corresponding author, and my student is the first author.

23.

CORE rankings: A* [EMNLP'22]

Chenhui Shen, Liying Cheng, Lidong Bing, Yang You and Luo Si

Rethinking Structure-Controllability in Controllable Summarization.

Conference on Empirical Methods in Natural Language Processing 2022.

24.

CORE rankings: A* [TPDS'22]

Jiarui Fang, Zilin Zhu, Shenggui Li, Hui Su, Yang Yu, Jie Zhou, Yang You

Parallel Training of Pre-trained Models via Chunk-based Dynamic Memory Management

https://ieeexplore.ieee.org/abstract/document/9940581

25.

CORE rankings: A* [AAAI'23]

Zangwei Zheng, Pengtai Xu, Xuan Zou, Da Tang, Zhen Li, Chenguang Xi, Peng Wu, Leqi Zou, Yijie Zhu, Ming Chen, Xiangzhuo Ding, Fuzhao Xue, Ziheng Qing, Youlong Cheng, Yang You

CowClip: Reducing CTR Prediction Model Training Time from 12 hours to 10 minutes on 1 GPU

https://arxiv.org/abs/2204.06240

Distinguished Paper Award (12 of all 8777 submissions: 0.14%)

https://drive.google.com/file/d/1bPxI5_NMlZbXQod7rCbH1EkGgt_bRjTe/view?usp=sharing

I am the corresponding author, and my student is the first author.

26.

CORE rankings: A* [AAAI'23]

Xin Hu, Lingling Zhang, Jun Liu, Jinfu Fan, Yang You, Yaqiang Wu.

GPTR: Gestalt-Perception Transformer for Diagram Object Detection.

https://arxiv.org/pdf/2212.14232.pdf

27.

CORE rankings: A* [ICLR'23]

Jianfei Yang, Xiangyu Peng, Kai Wang, Zheng Zhu, Jiashi Feng, Lihua Xie, Yang You

Divide to Adapt: Mitigating Confirmation Bias for Domain Adaptation of Black-Box Predictors

https://arxiv.org/pdf/2205.14467.pdf

Selected as **ICLR spotlight**

28.

CORE rankings: A* [ICML'23]

Fuzhao Xue, Jianghai Chen, Aixin Sun, Xiaozhe Ren, Zangwei Zheng, Xiaoxin He, Yongming Chen, Xin Jiang, Yang You

A Study on Transformer Configuration and Training Objective

I am the corresponding author, and my student is the first author.

29.

CORE rankings: A* [ICML'23]

Fuzhao Xue, Valerii Likhosherstov, Anurag Arnab, Neil Houlsby, Mostafa Dehghani, Yang You

Adaptive Computation with Elastic Input Sequence

I am the corresponding author, and my student is the first author.

30.

CORE rankings: A* [CVPR'23]

Jianyang Gu, Kai Wang, Hao Luo, Chen Chen, Wei Jiang, Yuqiang Fang, Shanghang Zhang, Yang You, Jian Zhao

MSINet: Twins Contrastive Search of Multi-Scale Interaction for Object ReID

31.

CORE rankings: A* [CVPR'23]

Shuo Yang, Xu Zhao Pan, Kai Wang, Yang You, Hongxun Yao, Tongliang Liu, Min Xu

BiCro: Noisy Correspondence Rectification for Multi-modality Data via Bi-directional Cross-modal Similarity Consistency

32.

CORE rankings: A [IPDPS'23]

Qifan Xu, Shenggui Li, Chaoyu Gong, Yang You

An Efficient 2D Method for Training Super-Large Deep Learning Models

I am the corresponding author, and my student is the first author.

33.

CORE rankings: A* [ACL'23]

Yang Luo, Xiaozhe REN, Zangwei Zheng, Xin Jiang, ZHUO JIANG and Yang You

CAME: Confidence-guided Adaptive Memory Efficient Optimization

Outstanding Paper Award, 39 out of 4562 submissions, 0.86%

I am the corresponding author, and my student is the first author.

34.

CORE rankings: A* [ACL'23]

Shenggui Li, Fuzhao Xue, Chaitanya Baranwal, Yongbin Li and Yang You

Make Training Transformer with Infinite Long Sequence Possible from System Perspective

I am the corresponding author, and my student is the first author.

35.

Impact Factor: 8.793 [TNNLS'23]

Huang, Linqing; Zhao, Wangbo; Liu, Yong; Yang, Duo; Liew, Alan; You, Yang

An Evidential Multi-target Domain Adaptation Method Based on Weighted Fusion for Cross-domain Pattern Classification

I am the corresponding author, and my student is the first author.

36.

CORE rankings: A [SC'23]

Ziming Liu, Shenggan Cheng, Haotian Zhou, Yang You

WPipe: Harnessing Wave-Like Pipeline Parallelism for Enhanced Large Model Training Efficiency

The International Conference for High Performance Computing, Networking, Storage, and Analysis

I am the corresponding author, and my student is the first author.

37.

CORE rankings: A* [ICCV'23]

Dataset Quantization

Daquan Zhou, Kai Wang, Jianyang Gu, Xiangyu Peng, Dongze Lian, Yifan Zhang, Yang You, Jiashi Feng

International Conference on Computer Vision 2023

38.

CORE rankings: A* [ICCV'23]

DREAM: Efficient Dataset Distillation by Representative Matching

Yanqing Liu, Jianyang Gu, Kai Wang, Zheng Zhu, Wei Jiang, Yang You

International Conference on Computer Vision 2023

I am the corresponding author, and my student is the first author.

39.

CORE rankings: A* [ICCV'23]

Preventing Zero-Shot Transfer Degradation in Continual Learning of Vision-Language Models

Zangwei Zheng, Mingyuan MA, Kai Wang, Ziheng Qin, Xiangyu Yue, Yang You

International Conference on Computer Vision 2023

I am the corresponding author, and my student is the first author.

40.

CORE rankings: A* [NeurIPS'23]

Does Graph Distillation See Like Vision Dataset Counterpart?

Beining Yang, Kai Wang, Qingyun Sun, Cheng Ji, Xingcheng Fu, Hao Tang, Yang You, Jianxin Li

Conference on Neural Information Processing Systems

41.

CORE rankings: A* [NeurIPS'23]

Response Length Perception and Sequence Scheduling: An LLM-Empowered LLM Inference Pipeline

Zangwei Zheng, Xiaozhe Ren, Fuzhao Xue, Yang Luo, Xin Jiang, Yang You

Conference on Neural Information Processing Systems

I am the corresponding author, and my student is the first author.

42.

CORE rankings: A* [NeurIPS'23]

To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis

Fuzhao Xue, Yao Fu, Wangchunshu Zhou, Zangwei Zheng, Yang You

Conference on Neural Information Processing Systems

I am the corresponding author, and my student is the first author.

43.

CORE rankings: A* [EMNLP'23]

A Hierarchical Encoding-Decoding Scheme for Abstractive Multi-document Summarization

Chenhui Shen, Liying Cheng, Xuan-Phi Nguyen, Yang You, Lidong Bing

The 2023 Conference on Empirical Methods in Natural Language Processing

I am the corresponding author, and my student is the first author.

44.

CORE rankings: A* [EMNLP'23]

Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization

Chenhui Shen, Liying Cheng, Xuan-Phi Nguyen, Yang You, Lidong Bing

The 2023 Conference on Empirical Methods in Natural Language Processing

45.

CORE rankings: A* [ICLR'24]

InfoBatch: Lossless Training Speed Up by Unbiased Dynamic Data Pruning

Ziheng Qin, Kai Wang, Zangwei Zheng, Jianyang Gu, Xiangyu Peng, Zhaopan Xu, Daquan Zhou, Lei Shang, Baigui Sun, Xuansong Xie, Yang You

https://arxiv.org/abs/2303.04947

The Twelfth International Conference on Learning Representations

ORAL paper 1.2% acceptance rate

I am the corresponding author, and my student is the first author.

46.

CORE rankings: A* [ICLR'24]

Towards Lossless Dataset Distillation via Difficulty-Aligned Trajectory Matching

Ziyao Guo, Kai Wang, George Cazenavette, Hui Li, Kaipeng Zhang, Yang You

The Twelfth International Conference on Learning Representations

I am the corresponding author, and my student is the first author.

47.

CORE rankings: A* [ICLR'24]

AutoChunk: Automated Activation Chunk for Memory-Efficient Deep Learning Inference

Xuanlei Zhao, Shenggan Cheng, Guangyang Lu, Jiarui Fang, Haotian Zhou, Bin Jia, Ziming Liu, Yang You

The Twelfth International Conference on Learning Representations

I am the corresponding author, and my student is the first author.

48.

CORE rankings: A* [AAAI'24]

Jianyang Gu, Kai Wang, Wei Jiang, Yang You

Summarizing Stream Data for Memory-Constrained Online Continual Learning

The 38th Annual AAAI Conference on Artificial Intelligence

49.

CORE rankings: A* [WWW'24]

Helen: Optimizing CTR Prediction Models with Frequency-wise Hessian Eigenvalue Regularization

Zirui Zhu, Yong Liu, Zangwei Zheng, Huifeng Guo, Yang You

The Web Conference (formerly known as the International World Wide Web Conference, abbreviated as WWW)

20.2% acceptance rate (406 out of 2008 submissions)

I am the corresponding author, and my student is the first author.

50.

[MLSys'24]

HeteGen: Efficient Heterogeneous Parallel Inference for Large Language Models on Resource-Constrained Devices

Xuanlei Zhao, Bin Jia, Haotian Zhou, Ziming Liu, Shenggan Cheng, Yang You

7th Conference on Machine Learning and Systems (MLSys)

I am the corresponding author, and my student is the first author.

51.

CORE rankings: A* [PPoPP'24]

FastFold: Optimizing AlphaFold Training and Inference on GPU Clusters

Shenggan Cheng, Xuanlei Zhao, Guangyang Lu, Jiarui Fang, Tian Zheng, Ruidong Wu, Xiwen Zhang, Jian Peng, Yang You

I am the corresponding author, and my student is the first author.

52.

CORE rankings: A* [ICML'24]

Navigating Complexity: Toward Lossless Graph Condensation via Expanding Window Matching

Yuchen Zhang, Tianle Zhang, Kai Wang, Ziyao Guo, Yuxuan Liang, Xavier Bresson, Wei Jin, Yang You

I am the corresponding author, and my student is the first author.

53.

CORE rankings: A* [ICML'24]

DiffAug: Enhance Unsupervised Contrastive Learning with Domain-Knowledge-Free Diffusion-based Data Augmentation

Zelin Zang, Hao Luo, Kai Wang, Panpan Zhang, Fan Wang, Stan Z. Li, Yang You

I am the corresponding author, and my student is the first author.

54.

CORE rankings: A* [ICML'24]

OpenMoE: An Early Effort on Open Mixture-of-Experts Language Models

Fuzhao Xue, Zian Zheng, Yao Fu, Jinjie Ni, Zangwei Zheng, Wangchunshu Zhou, Yang You

I am the corresponding author, and my student is the first author.

55.

CORE rankings: A* [ICML'24]

GliDe with a CaPE: A Low-Hassle Method to Accelerate Speculative Decoding

Cunxiao Du, Jing Jiang, Xu Yuanchen, Jiawei Wu, Sicheng Yu, Yongqi Li, Shenggui Li, Kai Xu, Liqiang Nie, Zhaopeng Tu, Yang You

I am the corresponding author, and my student is the first author.

56.

CORE rankings: A* [ECCV'24]

Dataset Growth

Ziheng Qin, Zhaopan Xu, YuKun Zhou, Kai Wang, Zangwei Zheng, Zebang Cheng, Hao Tang, Lei Shang, Baigui Sun, Radu Timofte, Xiaojiang Peng, Hongxun Yao, Yang You

I am the corresponding author, and my student is the first author.

57.

CORE rankings: A* [EMNLP'24]

How Does the Textual Information Affect the Retrieval of Multimodal In-Context Learning?

Yang Luo, Zangwei Zheng, Zirui Zhu, Yang You

I am the corresponding author, and my student is the first author.

58.

CORE ranking: A* [ACM SoCC'24]

ParaGAN: A Scalable Distributed Training Framework for Generative Adversarial Networks

Ziji Shi, Jialin Li, Yang You

I am the corresponding author, and my student is the first author.

59.

CORE rankings: A* [NeurIPS'24]

SpeedLoader: An I/O efficient scheme for heterogeneous and distributed LLM operation

Yiqi Zhang, Yang You

I am the corresponding author, and my student is the first author.

60.

CORE rankings: A* [NeurIPS'24]

MixEval: Fast and Dynamic Human Preference Approximation with LLM Benchmark Mixtures

Jinjie Ni, Fuzhao Xue, Xiang Yue, Yuntian Deng, Mahir Shah, Kabir Jain, Graham Neubig, Yang You

I am the corresponding author, and my student is the first author.

61.

CORE rankings: A* [NeurIPS'24]

Toward Reliable, Reproducible, and Practical Human Evaluation Protocol for Text-to-Video Models

Tianle Zhang, Langtian Ma, Yuchen Yan, Yuchen Zhang, Yue Yang, Ziyao Guo, Wenqi Shao, Kai Wang, Yang You, Yu Qiao, Ping Luo, Kaipeng Zhang

62.

CORE rankings: A* [NeurIPS'24]

Dynamic Tuning Towards Parameter and Inference Efficiency for ViT Adaptation

Wangbo Zhao, Jiasheng Tang, Yizeng Han, Yibing Song, Kai Wang, Gao Huang, Fan Wang, Yang You

I am the corresponding author, and my student is the first author.

63.

CORE rankings: A* [CVPR'24]

Efficient Dataset Distillation via Minimax Diffusion

Jianyang Gu, Saeed Vahidian, Vyacheslav Kungurtsev, Haonan Wang, Wei Jiang, Yang You, Yiran Chen

64.

CORE ranking: A* [ASPLOS'25]

Concerto: Automatic Communication Optimization and Scheduling for Large-Scale Deep Learning

Shenggan Cheng*, Shengjie Lin*, Lansong Diao, Hao Wu, Siyu Wang, Chang Si, Ziming Liu, Xuanlei Zhao, Jiangsu Du, Wei Lin, Yang You

I am the corresponding author, and my student is the first author.

65.

CORE rankings: A* [PPoPP'25]

WeiPipe: Weight Pipeline Parallelism for Communication-Effective Long-Context Large Model Training

Junfeng Lin, Liu Ziming, Yang You, Jun Wang, Weihao Zhang, Rong Zhao

66.

CORE rankings: A* [ICLR'25]

Dynamic Diffusion Transformer

Wangbo Zhao, Yizeng Han, Jiasheng Tang, Kai Wang, Yibing Song, Gao Huang, Fan Wang, Yang You

The Thirteenth International Conference on Learning Representations

I am the corresponding author, and my student is the first author.

67.

CORE rankings: A* [ICLR'25]

Real-Time Video Generation with Pyramid Attention Broadcast

Xuanlei Zhao, Xiaolong Jin, Kai Wang, Yang You

The Thirteenth International Conference on Learning Representations

I am the corresponding author, and my student is the first author.

68.

CORE rankings: A* [ICLR'25]

Can a Large Language Model be a Gaslighter?

Wei Li, Luyao Zhu, Yang Song, Ruixi Lin, Rui Mao, Yang You

The Thirteenth International Conference on Learning Representations

I am the corresponding author, and my student is the first author.

69.

CORE rankings: A* [ICLR'25]

MixEval-X: Any-to-any Evaluations from Real-world Data Mixture

Jinjie Ni, Yifan Song, Deepanway Ghosal, Bo Li, David Junhao Zhang, Xiang Yue, Fuzhao Xue, Yuntian Deng, Zian Zheng, Kaichen Zhang, Mahir Shah, Kabir Jain, Yang You, Michael Shieh

The Thirteenth International Conference on Learning Representations

70.

CORE rankings: A* [ICLR'25]

Train Small, Infer Large: Memory-Efficient LoRA Training for Large Language Models

Jun Zhang, Jue WANG, Huan Li, Lidan Shou, Ke Chen, Yang You, Guiming Xie, Xuejian Gong, Kunlong Zhou

The Thirteenth International Conference on Learning Representations

71.

CORE rankings: A* [CVPR'25]

A Stitch in Time Saves Nine: Small VLM is a Precise Guidance for accelerating Large VLMs

Wangbo Zhao, Yizeng Han, Jiasheng Tang, Zhikai Li, Yibing Song, Kai Wang, Zhangyang Wang, Yang You

I am the corresponding author, and my student is the first author.

72.

CORE rankings: A* [CVPR'25]

Emphasizing Discriminative Features for Dataset Distillation in Complex Scenarios

Kai Wang, Zekai Li, Zhi-Qi Cheng, Samir Khaki, Ahmad Sajedi, Shanmukha Ramakrishna Vedantam, Konstantinos N. Plataniotis, Alexander G Hauptmann, Yang You

I am the corresponding author, and my student is the first author.

73.

CORE rankings: A* [CVPR'25]

A Closer Look at Time Steps is Worthy of Triple Speed-Up for Diffusion Model Training

Kai Wang, Mingjia Shi, YuKun Zhou, Zekai Li, Xiaojiang Peng, Zhihang Yuan, Yuzhang Shang, Hanwang Zhang, Yang You

I am the corresponding author, and my student is the first author.

74.

CORE rankings: A* [ICML'25]

Info-Coevolution: An Efficient Framework for Data Model Coevolution

Ziheng Qin, Hailun Xu, Wei Chee Yew, Qi Jia, Yang Luo, Kanchan Sarkar, Danhui Guan, Kai Wang, Yang You

I am the corresponding author, and my student is the first author.

75.

CORE rankings: A* [ICML'25]

Unsupervised Learning for Class Distribution Mismatch

Pan Du, Wangbo Zhao, Xinai Lu, Nian Liu, Zhikai Li, Chaoyu Gong, Suyun Zhao, Hong Chen, Cuiping Li, Kai Wang, Yang You

I am the corresponding author, and my student is the first author.

76.

CORE rankings: A* [ICML'25]

DSP: Dynamic Sequence Parallelism for Multi-Dimensional Transformers

Xuanlei Zhao, Shenggan Cheng, Chang Chen, Zangwei Zheng, Ziming Liu, Zheming Yang, Yang You

I am the corresponding author, and my student is the first author.

77.

CORE rankings: A* [ICML'25]

MERIT: Maximum-normalized Element-wise Ratio for Language Model Large-batch Training

Yang Luo, Zangwei Zheng, Ziheng Qin, Zirui Zhu, Yong Liu, Yang You

I am the corresponding author, and my student is the first author.

78.

CORE rankings: A* [ICML'25]

SeedLoRA: A Fusion Approach to Efficient LLM Fine-Tuning

Yong Liu, Di Fu, Shenggan Cheng, Zirui Zhu, Yang Luo, Minhao Cheng, Cho-Jui Hsieh, Yang You

I am the corresponding author, and my student is the first author.

79.

CORE rankings: A* [ACL'25]

DavIR: Data Selection via Implicit Reward for Large Language Models

Haotian Zhou, Tingkai Liu, Qianli Ma, Yufeng Zhang, Jianbo Yuan, Pengfei Liu, Yang You, Hongxia Yang

80.

[COLM'25]

Ensemble Debiasing Across Class and Sample Levels for Fairer Prompting Accuracy

Ruixi Lin, Ziqiao Wang, Yang You

I am the corresponding author, and my student is the first author.

81.

CORE rankings: A* [NeurIPS'25]

StarTrail: Concentric Ring Parallelism for Efficient Near-Infinite-Context Transformer Model Training

Ziming Liu, Shaoyu Wang, Shenggan Cheng, Zhongkai Zhao, Kai Wang, Xuanlei Zhao, James Demmel, Yang You

I am the corresponding author, and my student is the first author.

82.

CORE rankings: A* [NeurIPS'25]

Drag-and-Drop LLMs: Zero-Shot Prompt-to-Weights

Zhiyuan Liang, Dongwen Tang, Yuhao Zhou, Xuanlei Zhao, Mingjia Shi, Wangbo Zhao, Zekai Li, Peihao Wang, Konstantin Schürholt, Damian Borth, Michael M. Bronstein, Yang You, Zhangyang Wang, Kai Wang

83.

CORE rankings: A* [NeurIPS'25]

Neural-Driven Image Editing

Pengfei Zhou, Jie Xia, Xiaopeng Peng, Wangbo Zhao, Zilong Ye, Zekai Li, Suorong Yang, Jiadong Pan, Yuanxiang Chen, Ziqiao Wang, Kai Wang, Qian Zheng, Xiaojun Chang, Gang Pan, Shurong Dong, Kaipeng Zhang, Yang You

I am the corresponding author, and my student is the first author.

84.

CORE rankings: A* [NeurIPS'25]

Scaling Up Parameter Generation: A Recurrent Diffusion Approach

Kai Wang, Dongwen Tang, Wangbo Zhao, Konstantin Schürholt, Zhangyang Wang, Yang You

I am the corresponding author, and my student is the first author.

85.

CORE rankings: A* [NeurIPS'25]

ElasticMM: Efficient Multimodal LLMs Serving with Elastic Multimodal Parallelism

Zedong Liu, Shenggan Cheng, Guangming Tan, Yang You, Dingwen Tao

NeurIPS Oral Paper (Top 0.36%, 77 out of 21575 submissions)

86.

CORE rankings: A* [NeurIPS'25]

REPA Works Until It Doesn't: Early-Stopped, Holistic Alignment Supercharges Diffusion Training

Ziqiao Wang, Wangbo Zhao, Yuhao Zhou, Zekai Li, Zhiyuan Liang, Mingjia Shi, Xuanlei Zhao, Pengfei Zhou, Kaipeng Zhang, Zhangyang Wang, Kai Wang, Yang You

I am the corresponding author, and my student is the first author.

87.

CORE rankings: A* [NeurIPS'25]

Sparse MeZO: Less Parameters for Better Performance in Zeroth-Order LLM Fine-Tuning

Yong Liu, Zirui Zhu, Chaoyu Gong, Minhao Cheng, Cho-Jui Hsieh, Yang You

I am the corresponding author, and my student is the first author.

87.

CORE rankings: A* [PPoPP'26]

HelixPipe: Efficient Distributed Training of Long Sequence Transformers with Attention Parallel Pipeline Parallelism

Geng Zhang, Shenggan Cheng, Xuanlei Zhao, Ziming Liu, Yang You

I am the corresponding author, and my student is the first author.

89.

CORE rankings: A* [AAAI'26]

MDK12-Bench: A Comprehensive Evaluation of Multimodal Large Language Models on Multidisciplinary Exams

Pengfei Zhou, Xiaopeng Peng, Fanrui Zhang, xu Zhao Pan, Jiaxin Ai, Yansheng Qiu, Wangbo Zhao, Jiajun Song, Chuanhao Li, Weidong Tang, Zhen Li, Haoquan Zhang, Zizhen Li, Xiaofeng Mao, Yukang Feng, Jianwen Sun, Kai Wang, Xiaojun Chang, Wenqi Shao, Yang You, Kaipeng Zhang

90.

IEEE TPAMI

DyDiT++: Dynamic Diffusion Transformers for Efficient Visual Generation

IEEE Transactions on Pattern Analysis and Machine Intelligence

Wangbo Zhao, Yizeng Han, Jiasheng Tang, Kai Wang, Hao Luo, Yibing Song, Gao Huang, Fan Wang, Yang You

91.

Publications before Yang You got his PhD

1.

[TPDS'19] You, Yang, Zhao Zhang, Cho-Jui Hsieh, James Demmel, and Kurt Keutzer. "Fast deep neural network training on distributed systems and cloud TPUs."

IEEE Transactions on Parallel and Distributed Systems 30, no. 11 (2019): 2449-2462.

2. CORE rankings: A

[SC'19] You, Yang, Jonathan Hseu, Chris Ying, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh.

"Large-batch training for LSTM and beyond."

In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1-16. 2019.

2. CORE rankings: A*

[ICDM'19] You, Yang, Yuxiong He, Samyam Rajbhandari, Wenhan Wang, Cho-Jui Hsieh, Kurt Keutzer, and James Demmel.

"Fast LSTM Inference by Dynamic Decomposition on Cloud Systems."

In 2019 IEEE International Conference on Data Mining (ICDM), pp. 748-757.

Best Paper Candidate.

3.

[KAIS'20] Y. You, Y. He, S. Rajbhandari, W. Wang, C.-J. Hsieh, K. Keutzer, J. Demmel.

Fast LSTM by dynamic decomposition on cloud and distributed systems

Journal of Knowledge and Information Systems.

4. CORE rankings: A*

[ICLR'20] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.-J. Hsieh.

Large Batch Optimization for Deep Learning: Training BERT in 76 minutes.

International Conference on Learning Representations

https://openreview.net/pdf?id=Syx4wnEtvH

The method has been used by Google, NVIDIA, Microsoft and included in MLPerf.

5. CORE rankings: A

[ICPP'18] Y. You, Z. Zhang, C.-J. Hsieh, J. Demmel, K. Keutzer.

ImageNet Training in Minutes

47th International Conference on Parallel Processing. August 13th - 16th, Eugene, USA.

https://arxiv.org/pdf/1709.05011.pdf

Best Paper Award (1 out of 313 submissions: 0.3%); The most cited HPC conference paper (HPDC, ICS, ICPP, IPDPS, PPOPP, SC, etc.) published between 2018 and 2020 (according to Google Scholar).

6. CORE rankings: A

[ICS'18] Y. You, J. Demmel, C.-J. Hsieh, R. Vuduc.

Accurate, Fast and Scalable Kernel Ridge Regression on Parallel and Distributed Systems

ACM International Conference on Supercomputing (ICS), June 12-15, Beijing, China. 18.7% (36/193) acceptance rate

7. CORE rankings: A

[SC'17] Y. You, A. Buluc, J. Demmel.

Scaling Deep Learning on GPU and Knights Landing Clusters

International Conference for High Performance Computing, Networking, Storage and Analysis (Supercomputing), November 12-17, Denver, USA. 18.7% (61/327) acceptance rate

8. CORE rankings: A

[ICPP'17] Y. You, J. Demmel.

Runtime Data Layout Scheduling for Machine Learning Dataset

46th International Conference on Parallel Processing. 28.4% (60/211) acceptance rate.

9.

[TPDS'16] Y. You, J. Demmel, K. Czechowski, L. Song, R. Vuduc.

Design and Implementation of a Communication-Optimal Classifier for Distributed Kernel Support Vector Machines

IEEE Transactions on Parallel and Distributed Systems, h5-index=76, DOI: 10.1109/TPDS.2016.2608823

10. CORE rankings: A*

[NeurIPS'16] Y. You, X. Lian, J. Liu, H. Yu, I. Dhillon, J. Demmel, C.-J. Hsieh.

Asynchronous Parallel Greedy Coordinate Descent,

2016 Conference on Neural Information Processing Systems, Dec 05-10, Barcelona, Spain. 22.7% (568/2500) acceptance rate

11.

[JPDC'16] Y. You, H. Fu, D. Bader, G. Yang.

Designing and Implementing a Heuristic Cross-Architecture Combination for Graph Traversal,

Journal of Parallel and Distributed Computing, h5-index=36, DOI: 10.1016/j.jpdc.2016.05.007

12. CORE rankings: A

[IPDPS'15] Y. You, J. Demmel, K. Czechowski, L. Song, R. Vuduc.

CA-SVM: Communication-Avoiding Support Vector Machines on Distributed Systems.

https://www2.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-9.pdf

Best Paper Award (4 out of 496 submissions: 0.8%) of IEEE International Parallel and Distributed Processing Symposium, May 25-29, Hyderabad, INDIA.

13.

[JPDC'15] Y. You, H. Fu, S. Song, A. Randles, D. Kerbyson, A. Marquez, G. Yang, A. Hoisie. Scaling Support Vector Machines on the Modern HPC Platforms

Journal of Parallel and Distributed Computing, h5-index=36, DOI: 10.1016/j.jpdc.2014.09.005

14. CORE rankings: A

[IPDPS'14] Y. You, S. Song, H. Fu, A. Marquez, M. Dehnavi, K. Barker, K. Cameron, A. Randles, G. Yang.

MIC-SVM: Designing A Highly Efficient Support Vector Machine For Advanced Modern Multi-Core and Many-Core Architectures.

IEEE Parallel and Distributed Processing Symposium, May 19-23, Phoenix, USA. 21% (114/541) overall acceptance rate; 17.5% acceptance rate for software track. DOI: 10.1109/IPDPS.2014.88

15. CORE rankings: A

[ICPP'14] Y. You, D. Bader, M. Dehnavi.

Designing a Heuristic Cross-Architecture Combination for Breadth-First Search

43rd International Conference on Parallel Processing, Sep 9-12, Minneapolis, USA. 36% (54/150) acceptance rate. DOI: 10.1109/ICPP.2014.16

16. A top journal for HPC

[IJHPCA'14] Y. You, H. Fu, S. Song, M. Dehnavi, L. Gan, X. Huang, G. Yang.

Evaluating the Many-core and Multi-core architectures through accelerating LWC stencil on Multi-core and Many-core architectures.

International Journal of High Performance Computing Application (2013 SCI IF=1.625), 21% (5/24) acceptance rate. DOI: 10.1177/1094342014524807

17.

Y. You, I. Gitman, B. Ginsburg.

Scaling SGD Batch Size to 32K for ImageNet Training.

NeurIPS workshop.