

MENG LI

Tech Lead ◊ Staff Research Scientist ◊ Facebook Reality Lab On-Device AI Research

meng_li@utexas.edu ◊ (512)468-9081

CAREER INTERESTS

Efficient deep learning for mobile and edge devices

- Hardware-aware neural network optimization (neural architecture search, quantization, pruning), neural network/accelerator co-optimization

Deep learning-based visual/speech understanding and enhancement

- Visual understanding: classification, object detection, semantic segmentation, pose estimation
- Image enhancement: neural denoising, super resolution
- Speech: automatic speech recognition, wake word

WORKING EXPERIENCE

Facebook Reality Lab, Facebook Inc.

Tech Lead, Staff Research Scientist

Sep. 2018 - Present

Menlo Park, CA

- Founded the efficient AI research and engineering team and expanded to more than 8 research scientists
- Led the cutting-edge researches and productization to enable efficient on-device AI for AR/VR devices
- Led the research and development of NN optimization tool stacks, including *neural architecture search*, *quantization*, *pruning*, and *knowledge distillation*, for hardware-aware model optimization
- Led path-finding researches on co-designing NNs (e.g., vision transformers) and accelerator
- Delivered accurate yet efficient models for key AR/VR applications, including *hand tracking*, *eye tracking*, *foveated rendering*, *person segmentation*, *neural denoising*, *video understanding*, *wake word*, etc
- Published papers in top-tier conferences and deliver SOTA results on *Imagenet classification*, *pose estimation*, and *semantic segmentation*

Machine Learning Research Group, Arm Inc.

Research Intern

May. 2016 - Sep. 2016, May 2017 - Sep. 2017

San Jose, CA

- Worked on research projects on privacy-preserving neural network training, including federated learning with non-IID data and PrivyNet with split network architectures

EDUCATION

University of Texas at Austin, TX

Sep. 2013 - Jul. 2018

- Ph.D. (*with honors*), Department of Electrical and Computer Engineering
- Advisor: David Z. Pan
- GPA: 4.0/4.0
- Citation (self-excluding): 1078

Peking University, Beijing, China

Sep. 2009 - Jul. 2013

- B.S., School of Microelectronics
- Advisor: Ru Huang
- GPA: 3.86/4.00 (3.92/4.00 for major)

HONORS AND AWARDS

- UT Austin Margarida Jacome Outstanding Dissertation Prize 2019
- EDAA Outstanding Dissertations Award, European Design and Automation Association (EDAA) 2019
- UT Austin Nominee of ACM Doctoral Dissertation Award, University of Texas at Austin 2019
- First Place, ACM Student Research Competition Grand Final (Graduate Category), Association for Computing Machinery (ACM) 2018
- Best Paper Award, ACM Great Lake Symposium on VLSI (GLSVLSI) 2018
- Best Poster Award, Florida Institute for Cybersecurity (FICS) Research Conference on Cybersecurity, University of Florida 2018
- Best Poster (Presentation) Award, ACM/SIGDA Asia and South Pacific Design Automation Conference (ASPDAC) Student Research Forum 2018
- Gold Medal, ACM/SIGDA Student Research Competition 2017
- Best Paper Award, IEEE International Symposium on Hardware Oriented Security and Trust (HOST) 2017
- Cockrell School Graduate Student Fellowship, University of Texas at Austin 2013 - 2017

PUBLICATION

Summary

- 60 research publications (11 first author, 49 co-author) in leading conferences/journals and 3 patents, 1928 total citations
- Full publication list available through Google Scholar:
<https://scholar.google.com/citations?user=lvdRkEkAAAAJ&hl=en&authuser=1>

Recent Publications

- Jiaqi Gu, Hyoukjun Kwon, Dilin Wang, Wei Ye, **Meng Li**, Yu-Hsin Chen, Liangzhen Lai, Vikas Chandra, and David Z Pan, “Multi-Scale High-Resolution Vision Transformer for Semantic Segmentation.”, CVPR 2022 (submitted)
- Haichuan Yang, Yuan Shangguan, Dilin Wang, **Meng Li**, Pierce Chuang, Xiaohui Zhang, Ganesh Venkatesh, Ozlem Kalinli, and Vikas Chandra, “Omni-sparsity DNN: Fast Sparsity Optimization for On-Device Streaming E2E ASR via Supernet.”, ICASSP 2022
- Chengyue Gong, Dilin Wang, **Meng Li**, Xinlei Chen, Zhicheng Yan, Yuandong Tian, Qiang Liu, and Vikas Chandra, “NASViT: Neural Architecture Search for Efficient Vision Transformers with Gradient Conflict aware Supernet Training.”, ICLR 2022
- Chengyue Gong, Dilin Wang, **Meng Li**, Vikas Chandra, and Qiang Liu, “Improve vision transformers training by suppressing over-smoothing.”, arXiv preprint arXiv:2104.12753 (2021)
- Yongan Zhang, Yonggan Fu, Weiwen Jiang, Chaojian Li, Haoran You, **Meng Li**, Vikas Chandra, and Yingyan Lin. “DNA: Differentiable Network-Accelerator Co-Search.” ISLPED 2021
- Dilin Wang, Chengyue Gong*, **Meng Li***, Qiang Liu, and Vikas Chandra. ”AlphaNet: Improved Training of Supernet with Alpha-Divergence.” ICML 2021 (**Long Oral**).
- Yonggan Fu, Han Guo, **Meng Li**, Xin Yang, Yining Ding, Vikas Chandra, and Yingyan Lin. “CPT: Efficient Deep Neural Network Training via Cyclic Precision.” ICLR 2021 (**Spotlight**)
- Dilin Wang, **Meng Li**, Chengyue Gong, and Vikas Chandra. “AttentiveNAS: Improving Neural Architecture Search via Attentive Sampling.” CVPR 2021
- Chengyue Gong, Dilin Wang, **Meng Li**, Vikas Chandra, and Qiang Liu. “KeepAugment: A Simple Information-Preserving Data Augmentation Approach.” CVPR 2021

- Hsin-Pai Cheng, Tunhou Zhang, Shiyu Li, Feng Yan, **Meng Li**, Vikas Chandra, Hai Li, and Yiran Chen. “Nasgem: Neural architecture search via graph embedding method.” AAAI 2021
- Hsin-Pai Cheng, Feng Liang, **Meng Li**, Bowen Cheng, Feng Yan, Hai Li, Vikas Chandra, and Yiran Chen. “ScaleNAS: One-Shot Learning of Scale-Aware Representations for Visual Recognition.” arXiv preprint arXiv:2011.14584 (2020)
- Lei Yang, Zheyu Yan, **Meng Li**, Hyoukjun Kwon, Liangzhen Lai, Tushar Krishna, Vikas Chandra, Weiwen Jiang, and Yiyu Shi. “Co-exploration of neural architectures and heterogeneous asic accelerator designs targeting multiple tasks.” ACM/IEEE Design Automation Conference (DAC) 2020
- Dilin Wang, **Meng Li**, Lemeng Wu, Vikas Chandra, and Qiang Liu. “Energy-aware neural architecture optimization with fast splitting steepest descent.” NeurIPS EMC2 Workshop 2019
- **Meng Li**, Yilei Li, Pierce Chuang, Liangzhen Lai, and Vikas Chandra. “Improving Efficiency in Neural Network Accelerator Using Operands Hamming Distance optimization.” NeurIPS EMC2 Workshop 2019
- Dilin Wang, **Meng Li**, Lemeng Wu, Vikas Chandra, and Qiang Liu. “Energy-aware neural architecture optimization with fast splitting steepest descent.” arXiv preprint arXiv:1910.03103 (2019)
- Yue Zhao, **Meng Li**, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra, “Federated Learning with Non-IID Data”, arXiv preprint arXiv:1806.00582 (2018)
- **Meng Li**, Liangzhen Lai, Naveen Suda, Vikas Chandra, and David Z. Pan, “PrivyNet: A Flexible Framework for Privacy-Preserving Deep Neural Network Training”, arXiv preprint arXiv:1709.06161 (2017)
- **Meng Li** and David Z. Pan, “A Synergistic Framework for Hardware IP Privacy and Integrity Protection”, PhD Thesis (published by Springer, UT Austin Margarida Jacome Outstanding Dissertation Prize, EDAA Outstanding Dissertation Award, ACM Student Research Competition Grand Final First Place)

Patents

- Ganesh Venkatesh, Liangzhen Lai, Pierce Chuang, **Meng Li**, and Vikas Chandra. “System and method for supporting alternate number format for efficient multiplication.” US Patent App. 16/511,085
- Ganesh Venkatesh, Liangzhen Lai, Pierce Chuang, and **Meng Li**. “System and method for performing small channel count convolutions in energy-efficient input operand stationary accelerator.” US Patent App. 16/511,544
- Ganesh Venkatesh, Liangzhen Lai, Pierce Chuang, and **Meng Li**. “Systems and methods for reading and writing sparse data in a neural network accelerator.” US Patent App. 16/509,138