Wei Wen

412-944-8906 | weiwen.web@gmail.com | http://www.pittnuts.com/ | https://github.com/wenwei202

EDUCATION

Ph.D. in Electrical and Computer Engineering, Duke University, USA, 08/2014-12/2019 (Expected)

Advisor: Dr. Hai Li

Research Interest: Deep Learning & Machine Learning & Neuromorphic Computing

M.S. in Electronic and Information Engineering, Beihang University, China, 09/2010-01/2013

B.S. in Electronic and Information Engineering, Beihang University, China, 09/2006-07/2010

RESEARCH STATEMENT

My research is Machine Learning and its applications in Computer Vision and Natural Language Processing. Recently, I focus on understanding learning algorithms, structural learning for accurate & efficient deep neural networks, and optimization algorithms for distributed deep learning. I was invited to give talks in UC Berkeley, Cornell University, NIPS 2017 oral, etc. I worked with Facebook Research, Microsoft Research, Intel Labs and HP Labs, where I incorporated my research into AI productions, including Facebook AI Infra and AML, Microsoft Bing, Intel Nervana & SkimCaffe, etc. I am a contributor of PyTorch/Caffe2.

INDUSTRIAL RESEARCH & ENGINEERING

Google Brain, Mountain View, CA, USA 05/2019-08/2019

Research Intern.

Facebook Research, Caffe2 and Applied Machine Learning, Menlo Park, CA, USA 05/2018-08/2018

Research Intern. Mentor: Yangqing Jia

- Personalization and Distributed Machine Learning.

Microsoft Research, Business AI, Redmond, WA, USA 05/2017-07/2017

Research Intern. Mentor: Yuxiong He

- Machine Reading Comprehension and Recurrent Neural Networks.

HP Labs, Platform Architecture Group, Palo Alto, CA, USA 06/2016-09/2016

Research Intern. Mentor: Cong Xu & supervisor: Paolo Faraboschi

- Distributed Deep Learning Systems.

Agricultural Bank of China, Software Engineer Employee, Beijing, China 07/2013-07/2014

Microsoft Research Asia, Research Intern, Mobile and Sensing Systems Group, Beijing, China 04/2013-06/2013

Tencent Inc., Software Engineer Intern, Advertising Platform and Products Division, Beijing, China 07/2012-09/2012

SELECTED PUBLICATIONS

- **W. Wen**, Y. Wang, F. Yan, C. Xu, Y. Chen, H. Li, "*SmoothOut: Smoothing Out Sharp Minima for Generalization in Large-Batch Deep Learning*", **Preprint**, 2018
- **W. Wen,** C. Xu, F. Yan, C. Wu, Y. Wang, Y. Chen, H. Li, "*TernGrad: Ternary Gradients to Reduce Communication in Distributed Deep Learning*", ***NIPS***, 2017. (***Oral, 40/3240=1.2%***) (Integrated into PyTorch/Caffe2)
- **W. Wen**, Y. He, S. Rajbhandari, M. Zhang, W. Wang, F. Liu, B. Hu, Y. Chen, H. Li, "*Learning Intrinsic Sparse Structures within Long Short-Term Memory*", ***ICLR***, 2018.
- **W. Wen,** C. Xu, C. Wu, Y. Wang, Y. Chen, H. Li, "*Coordinating Filters for Faster Deep Neural Networks*", ***ICCV***, 2017.
- **W. Wen,** C. Wu, Y. Wang, Y. Chen, H. Li, "*Learning Structured Sparsity in Deep Neural Networks*", ***NIPS***, 2016.
- J. Park, S. Li, **W. Wen**, P. T. P. Tang, H. Li, Y. Chen, P. Dubey, "*Faster CNNs with Direct Sparse Convolutions and Guided Pruning*", ***ICLR***, 2017.
- Y. Wang, **W. Wen,** L. Song, H. Li, "*Classification Accuracy Improvement for Neuromorphic Computing Systems with One-level Precision Synapses*", ***ASP-DAC***, 2017. (***Best Paper Award***)

SELECTED HONORS & AWARDS

- ICLR Travel Award 2018
- Graduate Student Conference Travel Fellowship, Duke ECE 2017
- NIPS Travel Award 2017
- Best Paper Award, Asia and South Pacific Design Automation Conference (ASP-DAC), IEEE 2017
- NIPS Travel Award 2016
- Best Paper Nomination, Design Automation Conference (DAC), IEEE 2016
- Best Paper Nomination, Design Automation Conference (DAC), IEEE 2015
- National Scholarship (3/233), Ministry of Education China 2009
- Second Prize, National College Physics Competition China 2007

SKILLS

- Machine learning: TensorFlow, PyTorch, Caffe
- Languages: C/C++/CUDA C, Python
- Android Development (with Google Play publications)

TEACHING

- Teach Assistant, CEE 690/ECE 590: Introduction to Deep Learning, Duke University, Fall 2018
- Teach Assistant, STA561/COMPSCI571/ECE682: Probabilistic Machine Learning, Duke University, Spring 2019

TALKS

- UC Berkeley, Scientific Computing and Matrix Computations Seminar, "On Matrix Sparsification and Quantization for Efficient and Scalable Deep Learning", 10/10/2018
- Cornell University, AI Seminar, "Efficient and Scalable Deep Learning", 10/05/2018
- NIPS 2017, TernGrad: Ternary Gradients to Reduce Communication in Distributed Deep Learning, 12/6/2017
- Alibaba DAMO Academy, "Deep Learning in Cloud-Edge AI Systems", SunnyVale, CA, 06/28/2018

SERVICE

- Paper reviewer, NeurIPS, CVPR, ICLR, TNNLS, TCAD, Neurocomputing, etc.
- Activity volunteer, Machine Learning for Girls, FEMMES (Female Excelling More in Math, Engineering, and Science) Capstone at Duke University, 02/2018
- Conference volunteer, Embedded Systems Week (ESWEEK), Pittsburgh, PA, USA, 10/2016