

# Wei WEN

412-944-8906 | [weiwen.web@gmail.com](mailto:weiwen.web@gmail.com) | <http://www.pittnuts.com/> | <https://github.com/wenwei202>

## EDUCATION

---

<b>Duke University</b>	<i>Durham, NC, United States</i>	<i>08/2017-12/2019 (Expected)</i>
Ph.D. Electrical and Computer Engineering		Supervisor: Dr. Hai Li
Research Area: Deep Learning & Neuromorphic Computing		
<b>University of Pittsburgh</b>	<i>Pittsburgh, PA, United States</i>	<i>09/2014-08/2017 (Transferred to Duke)</i>
Ph.D. Electrical and Computer Engineering	GPA: 3.952	Supervisor: Dr. Hai Li
<b>Beihang University</b>	<i>Beijing, China,</i>	<i>09/2006-07/2010, 09/2010-01/2013</i>
B.S., M.S. Electronic and Information Engineering	Rank: 10/170 (B.S.)	

## SELECTED PUBLICATION

---

- **Wei Wen**, Cong Xu, Feng Yan, Chunpeng Wu, Yandan Wang, Yiran Chen, Hai Li, “*TernGrad: Ternary Gradients to Reduce Communication in Distributed Deep Learning*”, the 31st Annual Conference on Neural Information Processing Systems (*NIPS*), 2017. (**Oral, 40/3240=1.2%**)
- **Wei Wen**, Cong Xu, Chunpeng Wu, Yandan Wang, Yiran Chen, Hai Li, “*Coordinating Filters for Faster Deep Neural Networks*”, Proceedings of the IEEE International Conference on Computer Vision (*ICCV*), 2017.
- **Wei Wen**, Chunpeng Wu, Yandan Wang, Yiran Chen, Hai Li, “*Learning Structured Sparsity in Deep Neural Networks*”, the 30th Annual Conference on Neural Information Processing Systems (*NIPS*), 2016.
- Jongsoo Park, Sheng Li, **Wei Wen**, Ping Tak Peter Tang, Hai Li, Yiran Chen, Pradeep Dubey, “*Faster CNNs with Direct Sparse Convolutions and Guided Pruning*”, the 5th International Conference on Learning Representations (*ICLR*), 2017.
- Chunpeng Wu, **Wei Wen**, Tariq Afzal, Yongmei Zhang, Yiran Chen, Hai Li, “*A Compact DNN: Approaching GoogLeNet-Level Accuracy of Classification and Domain Adaptation*”, *CVPR*, 2017.
- Yandan Wang, **Wei Wen**, Linghao Song, Hai Li, “*Classification Accuracy Improvement for Neuromorphic Computing Systems with One-level Precision Synapses*”, *ASP-DAC*, 2017. (**Best Paper Award**)
- **Wei Wen**, Chunpeng Wu, Yandan Wang, Kent Nixon, Qing Wu, Mark Barnell, Hai Li, Yiran Chen, “*A New Learning Method for Inference Accuracy, Core Occupation, and Performance Co-optimization on TrueNorth Chip*”, the 53rd Design Automation Conference (*DAC*), 2016. (**Best Paper Nomination**)
- **Wei Wen**, Chi-Ruo Wu, Xiaofang Hu, Beiye Liu, Tsung-Yi Ho, Xin Li, Yiran Chen, “*An EDA Framework for Large Scale Hybrid Neuromorphic Computing Systems*”, the 52nd Design Automation Conference (*DAC*), 2015. (**Best Paper Nomination**).

## INDUSTRY EXPERIENCE

---

<b>Microsoft Research, AI &amp; Research, Redmond, WA, USA</b>	<i>05/2017-07/2017</i>
Summer Intern, Supervisor: Yuxiong He & Fang Liu	
• Machine Reading Comprehension and Recurrent Neural Networks.	
<b>HP Labs, Platform Architecture Group, Palo Alto, CA, USA</b>	<i>06/2016-09/2016</i>
Summer Intern, Supervisor: Dr. Paolo Faraboschi and Dr. Cong Xu	
• Benchmarked Distributed Deep Learning Systems.	
<b>Agricultural Bank of China, Software Development Center, Beijing, China</b>	<i>07/2013-07/2014</i>
Software Developer Employee, Supervisor: Mr. Lei Fan	
• Developed web services for online bank transactions.	
<b>Microsoft Research, Mobile and Sensing Systems Group, Beijing, China</b>	<i>04/2013-06/2013</i>
Research Intern, Supervisor: Dr. Guobin Shen	
• Worked on a mobile system that dynamically generates frontal views for the user even when the user is at a slant viewing angle.	
<b>Tencent Inc., Advertising Platform and Products Division, Beijing, China</b>	<i>07/2012-09/2012</i>
Summer Intern, Supervisor: Mr. Yanan Zhao	
• Developed MVC-framework-based advertising websites	

## **ACADEMIA EXPERIENCE**

---

***TernGrad: Ternary Gradients to Reduce Communication in Distributed Deep Learning*** 01/2017-05/2017

- Quantizing floating gradients of SGD to ternary levels to speedup distributed training of Deep Neural Networks.

***Lower-rank Deep Convolutional Neural Networks*** 09/2016-03/2017

- Exploring to improve low-rank approximation methods to obtain faster deep neural networks.

***Learning Structured Sparsity in Deep Neural Networks*** 12/2015-05/2016

- Group-lasso Regularization: Enforced structured sparsity constraints to learn the number of filters, channels, neurons and layers in deep neural nets.
- Caffe CPU & GPU Implementation: 3× speedup in GPUs and 5× speedup in CPUs.

***TrueNorth Cognitive Learning*** 09/2015-12/2015

- Developed a new learning method for spiking neural networks in IBM TrueNorth chip;

***Brain-inspired Computing Systems*** 09/2014-04/2015

- Spectral Clustering: proposed Iterative Spectral Clustering algorithm to group connections of large-scale sparse neural networks into small clusters, so that connections can be locally and densely realized by Brain-inspired Computing Systems.

## **SKILLS**

---

- Machine learning: TensorFlow, Caffe (3 years)
- Languages: C/C++/CUDA C (5 years), Python & numpy (2 years)
- Linux, Bash Shell, git and svn (4 years)
- Android Development (with [Google Play](#) publications)

## **SELECTED HONORS & AWARDS**

---

- NIPS Travel Award 2017
- Best Paper Award, 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 2007