

YUKE WANG

805-259-9421 ◇ yuke_wang@cs.ucsb.edu

EDUCATION

Ph.D. candidate, Computer Science. <i>University of California, Santa Barbara</i>	GPA: 3.86/4.00	09/2018 – Now
B.E. , Software Engineering. <i>University of Electronic Science and Technology of China</i>	GPA: 3.93/4.00 , Rank: 10/759	07/2018

SKILLS

Languages: Python, C/C++, CUDA, Linux Shell.
Tools: Tensorflow, Pytorch, Sikit-Learn, Vivado HLS, Intel FPGA SDK, NVProf.

EXPERIENCE HIGHLIGHT

Rich hands-on experience on: 1) designing and optimizing GPU kernels for Graph Neural Networks, such as GCN, GraphSage; 2) developing deep neural networks (**DNNs**), such as ResNet, and VGG.; 3) deploying **machine learning** techniques (random forest, logistic regression, etc.) for **bigdata analysis**; 4. working on **bigdata sets** such as ImageNet.

RESEARCH & PROJECTS

-
- 1. EGG: A Framework for Efficient Graph Neural Networks on GPUs** 01/2020 – Now
1) Develop a GPU-based GNN acceleration framework – EGG. It abstracts and parameterizes GNN execution paradigms (e.g., Sparse Matrix Multiplication, Message Passing), and automatically “crafts” an optimal execution based on several factors, such as the GNN architectures, dataset properties, and hardware resource constraints; 2) EGG also incorporates a novel low-cost workload scheduling scheme to exploit the data spatial and temporal locality, which can largely improve the GPU memory access efficiency under the large graph and high node-embedding dimensionality settings.
 - 2. Dynamic Sparsity-aware Neural Network Inference** 11/2019 – 01/2020
1) Develop a light-weighted auxiliary “little” module with random projection and weight quantization for probing Neural Network (NN) layerwise output sparsity to facilitate NN inference acceleration; 2) The proposed “little” module inference scheme can be easily applied to various types of neural networks, such as CNN, and LSTM. (e.g., on ResNet-18, our proposed method outperforms the state-of-the-art solutions with much higher FLOPs reduction, memory saving and model accuracy).
 - 3. Performance and Power Efficient Accelerator for GNNs** 09/2019 – 11/2019
Improve the aggregation step of GNNs accelerator through on-chip computation reuse; Implement an efficient sparse matrix transformation based on Locality-Sensitive Hashing and Row-Column Reordering to guide PE workload mapping.
 - 4. AccD: A Framework for Accelerating Distance-related Algorithms on Reconfigurable Hardware** 04/2019 – 09/2019
1) Build a compiler framework for accelerating distance-related algorithms (KNN, N-Body Simulation, etc.) on CPU-FPGA platform; 2) Develop a Domain-specific Language (DDSL) and algorithm-hardware co-design to simplify design workflow.
 - 5. Static CNN Optimization Based on Information Field** 02/2019 - 05/2019
Propose a new static analysis method for CNN structure- and kernel- level optimization; The proposed optimization can save 50% FLOPs and 90% parameters on most state-of-the-art CNNs.

6. TiAcc: A FPGA Design Framework for K-means

01/2019 – 04/2019

Build a comprehensive FPGA design framework for K-means acceleration; The proposed design framework combines an algorithmic and a hardware design optimization.

7. KPynq: A Work-Efficient Triangle-Inequality based K-means on FPGA

10/2018 – 01/2019

Develop a novel design for K-means acceleration on Xilinx Pynq-Z1 FPGA board; KPynq achieves speedup (up to $4.2\times$) and energy-efficiency (up to $218\times$) compared with the CPU baseline.

PUBLICATIONS

Liu Liu, Lei Deng, Zhaodong Chen, **Yuke Wang**, Shuangchen Li, Jingwei Zhang, Yihua Yang, Zhenyu Gu, Yufei Ding, Yuan Xie. ***Boosting Deep Neural Network Efficiency with Dual-Module Inference.*** [ICML-2020 Submitted]

Xing Hu, Xinfeng Xie, **Yuke Wang**, Abanti Basak, Mingyu Yan, Ling Liang, Lei Deng, Yufei Ding, Yuan Xie. ***Geom: a Hierarchical Spatial Architecture with Hybrid Dataflows for Graph Learning.***

Yuke Wang, Boyuan Feng, Gushu Li, Lei Deng, Yuan Xie, Yufei Ding. ***AccD: A Compiler-based Framework for Accelerating Distance-related Algorithms on CPU-FPGA Platforms.***

Yuke Wang, Boyuan Feng, Georgios Tzimpragos, Lei Deng, Yuan Xie, Yufei Ding. ***TiAcc: An Efficient Triangle-Inequality based K-means Hardware Accelerator Design on FPGA.*** [ICS-2020 Submitted]

Zhizhou Zhang, Boyuan Feng, **Yuke Wang**, Yufei Ding. ***Convolution Neural Network Architecture Optimization via Information Field.***

Yuke Wang, Zhaorui Zeng, Boyuan Feng, Lei Deng, and Yufei Ding. ***KPynq: A Work-Efficient Triangle-Inequality based K-means on FPGA.*** [FCCM-2019].

HONORS & AWARDS

2019 Summer GSR recipient in CS Department of UCSB	06/2019
Outstanding Graduates Award of UESTC	10/2017
First-class People's Scholarship (2/20 in the Elite Program)	10/2017
Interdisciplinary Contest In Modeling (ICM) [Honorable Mention]	04/2017
Suzhou Industrial Zone Scholarship (2/20 in the Elite Program)	04/2017
International Software Testing Qualifications Board (Certified Tester) [Foundational Level]	10/2016
First-class People's Scholarship (4/116)	04/2016

EXPERIENCE

Teaching Assistant of CS160 (Translation of Programming Languages)	Fall. 2019
Teaching Assistant of CS8 (Python Programming Language)	Summer. 2019
Teaching Assistant of CS16 (C++ Programming Language)	Winter. 2019

PROFESSIONAL ACTIVITIES

Reviewed Papers:

ASPLOS-2020, ASPLOS-2019 (Top computer system & architecture conference)
PPOPP-2020 (Top parallel programming & computer system conference)
OOPSLA-2019 (Top programming language conference)

Presentations

Invited talk at Computer Systems WIPS: KPynq: A Work-Efficient Triangle-Inequality based K-means on FPGA. Feb. 2019