

Hongkuan Zhou

(323)-352-4237 | hongkuaz@usc.edu | tedzhouhk.github.io/about

EDUCATION

University of Michigan-Shanghai Jiao Tong University Joint Institute <i>B.S. in Electrical Computer Engineering</i>	Shanghai, China Aug. 2013 – May 2017
University of Southern California <i>M.S. in Computer Engineering</i>	Los Angeles, US Aug. 2017 – May 2019
University of Southern California <i>Ph.D. in Computer Engineering</i> Advisor: Professor Viktor K. Prasanna	Los Angeles, US Aug. 2019 – Present

WORKING EXPERIENCE

Research Assistant <i>FPGA/Parallel Computing Lab, USC</i>	Dec. 2017 – Present <i>Los Angeles, US</i>
<ul style="list-style-type: none">Developed the knowledge base for a dynamic compiler to achieve near-ASIC performance on reconfigurable processors in the SDH project.Developed novel minibatch sampling based training algorithms for Graph Neural Networks (GNNs) which scale to large graphs. Designed and implemented optimized parallel algorithms on CPU and GPU. Processed and published two node classification datasets from raw data in social networks.Developed universal GNN inference framework for real-time and large scale GNN applications using channel pruning to meet requirements of various inference scenarios.Working on the model hardware co-design of various GNN applications such as knowledge graph reasoning and fraud detection.	
Applied Scientist Intern <i>Amazon Web Service</i>	May 2021 – Aug. 2021 <i>Palo Alto, US</i>
<ul style="list-style-type: none">Designed and implemented TGL – a general framework for Temporal Graph Neural Networks training on large-scale dynamic graphs.Developed novel scheduling algorithm to optimize the performance of TGL on multiple-GPUs.	

PUBLICATIONS

- First Author:** SeDyT: A General Framework for Multi-Step Event Forecasting via Sequence Modeling on Dynamic Entity Embeddings, *ACM International Conference on Information and Knowledge Management (CIKM)*, 2021
- First Author:** Accelerating Large Scale Real-Time GNN Inference using Channel Pruning, *International Conference on Very Large Data Bases (VLDB)*, 2021
- Co-First Author:** Accurate, Efficient and Scalable Training of Graph Neural Networks, *Journal of Parallel and Distributed Computing (JPDC)*, 2020
- Co-First Author:** GraphSAINT: Graph Sampling Based Inductive Learning Method, *International Conference on Learning Representations (ICLR)*, 2020
- First Author:** Design and Implementation of Knowledge Base for Runtime Management of Software Defined Hardware, *IEEE High Performance Extreme Computing Conference (HPEC)*, 2019 (**Best Student Paper Nominee**)
- Co-First Author:** Accurate, Efficient and Scalable Graph Embedding, *33rd International Parallel and Distributed Processing Symposium (IPDPS)*, 2019

TECHNICAL STRENGTHS

Computer Languages: C/C++, Python, Matlab, Verilog, MySQL
Developer Tools: Git, Docker, Tensorflow, PyTorch, Bash, L^AT_EX
Language: Chinese, English, Japanese (N2 level)