Gyeong-In Yu

Ph.D. Student (Computer Science & Engineering)

Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Korea, 08826

+82-10-9511-9531 | gyeongin@snu.ac.kr | gyeongin.github.io | gyeongin | gyeonginyu

Research Interests

My research interest lies in the intersection of computer systems and machine learning, with a focus on systems for machine learning. More specifically, I am primarily working on software techniques to improve machine learning in the datacenter, including both inference and training.

Education

Seoul National University — *Seoul, Korea*

Ph.D. in Computer Science and Engineering — *Mar. 2017 - present*

- Advisor: Prof. Byung-Gon Chun

Seoul National University — *Seoul, Korea*

B.S. in Computer Science and Engineering and B.A. in Economics — *Mar. 2012 - Feb. 2017*

Research Experience

Research Assistant at Seoul National University — *Seoul, Korea*

Advisor: Prof. Byung-Gon Chun — *Jun. 2015 - present*

Research Intern at Microsoft AI and Research — *Redmond, WA*

Mentor: Dr. Matteo Interlandi, Dr. Saeed Amizadeh — *Jun. 2018 - Sep. 2018*

Research Intern at Microsoft Research Asia — *Beijing, China*

Mentor: Dr. Ming Wu — *Jun. 2017 - Sep. 2017*

Publications

Google Scholar: `https://scholar.google.com/citations?user=RwhPHaEAAAAJ`

Conference Publications

1. Woosuk Kwon*, **Gyeong-In Yu***, Eunji Jeong, Byung-Gon Chun (*equal contribution). Nimble: Lightweight and Efficient GPU Task Scheduling for Deep Learning. To appear at *34th Conference on Neural Information Processing Systems (NeurIPS 2020) (Spotlight)*.
2. Supun Nakandala, Karla Saur, **Gyeong-In Yu**, Konstantinos Karanasos, Carlo Curino, Markus Weimer, Matteo Interlandi. A Tensor Compiler Approach for One-size-fits-all ML Prediction Serving. To appear at *14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2020)*.
3. Woo-Yeon Lee, Yunseong Lee, Joo Seong Jeong, **Gyeong-In Yu**, Joo Yeon Kim, Ho Jin Park, Beomyeol Jeon, Wonwook Song, Gunhee Kim, Markus Weimer, Brian Cho, Byung-Gon Chun. Automating System Configuration of Distributed Machine Learning. *39th IEEE International Conference on Distributed Computing Systems (ICDCS 2019)*, July 2019.
4. Soojeong Kim, **Gyeong-In Yu**, Hojin Park, Sungwoo Cho, Eunji Jeong, Hyeonmin Ha, Sanha Lee, Joo Seong Jeong, Byung-Gon Chun. Parallax: Sparsity-aware Data Parallel Training of Deep Neural Networks. *14th European Conference on Computer Systems (EuroSys 2019)*, March 2019.

5. Eunji Jeong, Sungwoo Cho, **Gyeong-In Yu**, Joo Seong Jeong, Dongjin Shin, Byung-Gon Chun. JANUS: Fast and Flexible Deep Learning via Symbolic Graph Execution of Imperative Programs. *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2019)*, February 2019.
6. Eunji Jeong*, Joo Seong Jeong*, Soojeong Kim, **Gyeong-In Yu**, Byung-Gon Chun (*equal contribution). Improving the Expressiveness of Deep Learning Frameworks with Recursion. *13th European Conference on Computer Systems (EuroSys 2018)*, April 2018.

Other Publications

1. Supun Nakandala, **Gyeong-In Yu**, Markus Weimer, Matteo Interlandi. Compiling Classical ML Pipelines into Tensor Computations for One-size-fits-all Prediction Serving. *Systems for ML Workshop at 33rd Conference on Neural Information Processing Systems (NeurIPS)*, December 2019.
2. Ahnjae Shin, Dong-Jin Shin, Sungwoo Cho, Do Yoon Kim, Eunji Jeong, **Gyeong-In Yu**, Byung-Gon Chun. Stage-based Hyper-parameter Optimization for Deep Learning. *Systems for ML Workshop at 33rd Conference on Neural Information Processing Systems (NeurIPS)*, December 2019.
3. Eunji Jeong, Sungwoo Cho, **Gyeong-In Yu**, Joo Seong Jeong, Dong-Jin Shin, Taebum Kim, Byung-Gon Chun. Speculative Symbolic Graph Execution of Imperative Deep Learning Programs. *ACM SIGOPS Operating Systems Review (OSR)*, July 2019.
4. Eunji Jeong, Sungwoo Cho, **Gyeong-In Yu**, Joo Seong Jeong, Dongjin Shin, Byung-Gon Chun. Demonstration of JANUS: Fast and Flexible Deep Learning via Symbolic Graph Execution of Imperative Programs. *Demonstration at Second Conference on Machine Learning and Systems (MLSys)*, April 2019.
5. **Gyeong-In Yu**, Saeed Amizadeh, Byung-Gon Chun, Markus Weimer, Matteo Interlandi. Making Classical Machine Learning Pipelines Differentiable: A Neural Translation Approach. *Systems for ML Workshop at 32nd Conference on Neural Information Processing Systems (NeurIPS)*, December 2018.
6. Soojeong Kim, Eunji Jeong, Joo Seong Jeong, **Gyeong-In Yu**, Hojin Park, Byung-Gon Chun. Auto-Parallelizing Deep Learning for Multi-machine, Multi-GPU Environments. *Workshop on AI Systems at 26th ACM Symposium on Operating Systems Principles (SOSP)*, October 2017.
7. Byung-Gon Chun, Brian Cho, Beomyeol Jeon, Joo Seong Jeong, Gunhee Kim, Joo Yeon Kim, Woo-Yeon Lee, Yun Seong Lee, Markus Weimer, **Gyeong-In Yu**. Dolphin: Runtime Optimization for Distributed Machine Learning. *ML Systems Workshop at 33rd international conference on machine learning (ICML)*, June 2016.

Teaching Experience

Fall 2019	**Theory and Lab of IoT, AI, and Big Data**, Seoul National University	*Teaching Assistant*
Fall 2018	**Big Data Analytics and Deep Learning Systems**, Seoul National University	*Teaching Assistant*
Spring 2017	**Operating Systems**, Seoul National University	*Teaching Assistant*