HENGGANG CUI

1856 Fairhill Rd, Allison Park, PA, 15101
(651)403-3366 ⋄ cuihenggang@gmail.com

INTERESTS

Autonomous driving, deep learning, computer vision, and large-scale machine learning systems.

SKILLS

Programming languages	Python, C++, C, Java, CUDA
Machine learning softwares	TensorFlow, PyTorch, Spark, Matlab, Caffe, GraphLab
Big data softwares	MPI, TBB, HBase, OpenTSDB, Hadoop

EDUCATION

Carnegie Mellon University, Pittsburgh, PA.

Ph.D., Electrical and Computer Engineering (GPA: 3.93) May, 2017

- Advisor: Greg Ganger
- Research Topic: Large-Scale Machine Learning Systems

Master of Science, Electrical and Computer Engineering May, 2016

Tsinghua University, Beijing, China

Bachelor of Science, Electronic Information Science and Technology July 2012

WORK EXPERIENCE

Uber Advanced Technologies Group, Pittsburgh, PA.

Senior Autonomy Engineer June, 2017 to *now*

- Conducting research and development for autonomous driving.
- Designing and owning the production deep learning model for predicting the trajectories of the vehicle actors on the road. The model predicts multiple trajectories for each actor with probabilities, uncertainty estimation, and vehicle kinematics constraints.
- Works patented and published at *ICRA '19*, *NeurIPS '18 MLITS Workshop*, and *NeurIPS '19 "ML for Autonomous Driving" Workshop*.

PHD THESIS RESEARCH

Thesis: Exploiting Application Characteristics for Efficient System Support for Large-Scale Machine Learning

- **Committee**: Greg Ganger, Phil Gibbons, Garth Gibson, Eric Xing, and Derek Murray
- **Thesis Subprojects**: As follows:

GeePS: Specialized Parameter Server for Deep Learning on GPUs

Published at EuroSys'16, and open-sourced at https://github.com/cuihenggang/geeps

- Designed GeePS, a distributed parameter server system for distributed deep learning on GPU machines.
- Achieved good scalability from single-machine Caffe (13× more throughput with 16 machines), by overlapping communication with computation.
- Supported DNNs that cannot fit in GPU memory, by swapping data to/from CPU memory in the background.

IterStore: Efficient Parameter Server for Iterative ML

Published at SoCC'14 and open-sourced at https://github.com/cuihenggang/iterstore

- Observed the iterativeness characteristic of many ML applications: same access sequence every iterations.
- Proposed five parameter server optimizations, achieving up to 50× speedup.

INTERNSHIP

HP Labs, advised by Kimberly Keeton May 2014 to Aug 2014

Aperture: Ingest-Time Transformation for Big Time Series Data

Published at SoCC'15, with patent US20170322987A1

- Designed a database for big time series data analytics with low query latencies.

PUBLICATIONS

1 **Henggang Cui**, Vladan Radosavljevic, Fang-Chieh Chou, Tsung-Han Lin, Thi Nguyen, Tzu-Kuo Huang, Jeff Schneider, Nemanja Djuric. "Multimodal Trajectory Predictions for Autonomous Driving using Deep Convolutional Networks." In *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

2 Fang-Chieh Chou, Tsung-Han Lin, **Henggang Cui**, Vladan Radosavljevic, Thi Nguyen, Tzu-Kuo Huang, Matthew Niedoba, Jeff Schneider, Nemanja Djuric. "Predicting Motion of Vulnerable Road Users using High-Definition Maps and Efficient ConvNets." In *NeurIPS Workshop on Machine Learning for Intelligent Transportation Systems*, 2018.

3 **Henggang Cui**, Hao Zhang, Gregory R. Ganger, Phillip B. Gibbons, and Eric P. Xing. "GeePS: Scalable Deep Learning on Distributed GPUs with a GPU-Specialized Parameter Server." In *ACM European Conference on Computer Systems (EuroSys)*, 2016.

4 Aaron Harlap, **Henggang Cui**, Wei Dai, Jinliang Wei, Gregory R. Ganger, Phillip B. Gibbons, Garth A. Gibson, and Eric P. Xing. "Addressing the Straggler Problem for Iterative Convergent Parallel ML." In *ACM Symposium on Cloud Computing (SoCC)*, 2016.

5 **Henggang Cui**, Kimberly Keeton, Indrajit Roy, Krishnamurthy Viswanathan, and Gregory R. Ganger. "Using Data Transformations for Low-latency Time Series Analysis." In *ACM Symposium on Cloud Computing (SoCC)*, 2015.

6 Jinliang Wei, Wei Dai, Aurick Qiao, Qirong Ho, **Henggang Cui**, Gregory R. Ganger, Phillip B. Gibbons, Garth A. Gibson, Eric P. Xing. "Managed Communication and Consistency for Fast Data-Parallel Iterative Analytics." In *ACM Symposium on Cloud Computing (SoCC)*, 2015.

7 **Henggang Cui**, Alexey Tumanov, Jinliang Wei, Lianghong Xu, Wei Dai, Jesse Haber-Kucharsky, Qirong Ho, Gregory R. Ganger, Phillip B. Gibbons, Garth A. Gibson, and Eric P. Xing. "Exploiting Iterative-ness for Parallel ML Computations." In *ACM Symposium on Cloud Computing (SoCC)*, 2014.

8 **Henggang Cui**, James Cipar, Qirong Ho, Jin Kyu Kim, Seunghak Lee, Abhimanu Kumar, Jinliang Wei, Wei Dai, Gregory R. Ganger, Phillip B. Gibbons, Garth A. Gibson, and Eric P. Xing. "Exploiting Bounded Staleness to Speed Up Big Data Analytics." In *USENIX Annual Technical Conference (ATC)*, 2014.

9 Qirong Ho, James Cipar, **Henggang Cui**, Jin Kyu Kim, Seunghak Lee, Phillip B. Gibbons, Garth A. Gibson, Gregory R. Ganger, and Eric P. Xing. "More Effective Distributed ML via a Stale Synchronous Parallel Parameter Server." In *Neural Information Processing Systems (NIPS)*, 2013.

10 **Henggang Cui**, Danielle Rasooly, Moises R. N. Ribeiro, and Leonid Kazovsky. "Optically Cross-Braced Hypercube: A Reconfigurable Physical Layer for Interconnects and Server-Centric Datacenters." In *Optical Fiber Communication Conference and Exposition (OFC/NFOEC)*, 2012.

11 Dan Li, **Henggang Cui**, Yan Hu, Yong Xia, and Xin Wang. "Scalable Data Center Multicast using Multi-class Bloom Filter." In *19th IEEE International Conference on Network Protocols (ICNP)*, 2011.

PATENTS

1 Nemanja Djuric, Vladan Radosavljevic, Thi Duong Nguyen, Tsung-Han Lin, Jeff Schneider, **Henggang Cui**, Fang-Chieh Chou, and Tzu-Kuo Huang. "Object Motion Prediction and Autonomous Vehicle Control". *US20190049970A1.*

2 **Henggang Cui**, Kimberly Keeton, Indrajit Roy, Krishnamurthy Viswanathan, and Haris Volos. "Processing a query using transformed raw data". *US20170322987A1.*

PROFESSION SERVICES

Program Committee: *NeurIPS "Machine Learning for Autonomous Driving" Workshop, CVPR "Precognition: Seeing through the Future" Workshop, ACM Symposium on Cloud Computing '17 and '18, Computing Conference '17*

Reviewer: *IEEE International Conference on Robotics and Automation '19, IEEE Robotics and Automation Letters '19, IEEE Big Data '19, Cluster Computing '19, IEEE Transactions on Network Science and Engineering '18*