Huan Zhang

✉ huan@huan-zhang.com
UCLA, Los Angeles, CA 95005
https://huan-zhang.com

Education

Ph.D. in Computer Science (expected in 2020), **UCLA (2018-), UC Davis (2014-2018)**.
•*Advisor:* Prof. Cho-Jui Hsieh (2015-present) and Prof. Venkatesh Akella (2014-2018)
•*Area:* Safety in artificial intelligence and provable robustness analysis of machine learning.

M.S. in Computer Engineering, UC Davis (2012-2014).
•*Advisor:* Prof. Venkatesh Akella (akella@ucdavis.edu)
•*Area:* computer architecture and parallel computing.

Bachelor of Engineering, Zhejiang University (2008 - 2012), China.
•*Major:* Information Engineering

Work Experience

DeepMind, London, UK, Jun 2019 - Nov 2019.
•*Mentor: Pushmeet Kohli (pushmeet@google.com), Krishnamurthy (Dj) Dvijotham (dvij@google.com)*
Research scientist internship on machine learning fairness and robustness.

Microsoft Research, Redmond, WA, Jun 2018 - Sep 2018.
•*Mentor: Pengchuan Zhang (penzhan@microsoft.com) and Lin Xiao (Lin.Xiao@microsoft.com)*
Research internship on generative adversarial networks and neural network robustness verification.

Amazon A9.com Search Lab, Palo Alto, CA, Nov 2017 - Mar 2018.
•*Mentor: Inderjit Dhillon (isd@a9.com)*
Research internship on deep learning based product search query suggestion system for amazon.com.

IBM T.J. Watson Research Center, Yorktown Heights, NY, Jun 2017 - Nov 2017; Apr 2018 - Jun 2018,
•*Mentor: Jinfeng Yi (jinfengy@us.ibm.com), Pin-Yu Chen (Pin-Yu.Chen@ibm.com).*
Research internship on the safety and robustness of deep neural networks.

Selected Publications (* indicates equal contribution)

[1] **Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations**, *NeurIPS 2020*, Huan Zhang*, Hongge Chen*, Chaowei Xiao, Bo Li, Duane Boning, Cho-Jui Hsieh.

[2] **Provable, Scalable and Automatic Perturbation Analysis on General Computational Graphs**, *NeurIPS 2020*, Kaidi Xu*, Zhouxing Shi*, Huan Zhang*, Yihan Wang, Minlie Huang, Kai-Wei Chang, Bhavya Kailkhura, Xue Lin, Cho-Jui Hsieh.

[3] **An Efficient Adversarial Attack for Tree Ensembles.**, *NeurIPS 2020*, Chong Zhang, Huan Zhang, Cho-Jui Hsieh.

[4] **Reducing Sentiment Bias in Language Models via Counterfactual Evaluation.**, *Findings in EMNLP, 2020*, Po-Sen Huang*, Huan Zhang*, Ray Jiang, Robert Stanforth, Johannes Welbl, Jack Rae, Vishal Maini, Dani Yogatama, Pushmeet Kohli.

[5] **On ℓ_p-norm Robustness of Ensemble Decision Stumps and Trees.**, *ICML 2020*, Yihan Wang, Huan Zhang, Hongge Chen, Duane Boning and Cho-Jui Hsieh.

[6] **Towards Stable and Efficient Training of Verifiably Robust Neural Networks**, *ICLR 2020*, Huan Zhang, Hongge Chen, Chaowei Xiao, Sven Gowal, Robert Stanforth, Bo Li, Duane Boning, Cho-Jui Hsieh.

[7] **Robustness Verification of Tree-based Models**, *NeurIPS 2019*, Hongge Chen*, Huan Zhang*, Si Si, Yang Li, Duane Boning, Cho-Jui Hsieh. (*Equal contribution).

[8] **A Convex Relaxation Barrier to Tight Robustness Verification of Neural Networks**, *NeurIPS 2019*, Hadi Salman, Greg Yang, Huan Zhang, Cho-Jui Hsieh, Pengchuan Zhang.

[9] **Provably Robust Deep Learning via Adversarially Trained Smoothed Classifiers**, *NeurIPS 2019*, Hadi Salman, Greg Yang, Jerry Li, Pengchuan Zhang, Huan Zhang, Ilya Razenshteyn, Sebastien Bubeck.

[10] **RecurJac: An Efficient Recursive Algorithm for Bounding Jacobian Matrix of Neural Networks and Its Applications**, *AAAI 2019*, Huan Zhang, Pengchuan Zhang, Cho-Jui Hsieh.

[11] **Robust Decision Trees Against Adversarial Examples**, *ICML 2019*, Hongge Chen, Huan Zhang, Duane Boning, Cho-Jui Hsieh.

[12] **The Limitations of Adversarial Training and the Blind-Spot Attack**, *ICLR 2019*, Huan Zhang*, Hongge Chen*, Zhao Song, Duane Boning, inderjit dhillon, Cho-Jui Hsieh (* Equal contribution).

[13] **Efficient Neural Network Robustness Certification with General Activation Functions**, *NeurIPS 2018*, Huan Zhang*, Tsui-Wei Weng*, Pin-Yu Chen, Cho-Jui Hsieh, Luca Daniel. (* Equal contribution).

[14] **Towards Fast Computation of Certified Robustness for ReLU Networks**, *ICML 2018*, Tsui-Wei Weng*, Huan Zhang*, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Duane Boning, Inderjit S. Dhillon, Luca Daniel. (* Equal contribution).

[15] **Attacking Visual Language Grounding with Adversarial Examples: A Case Study on Neural Image Captioning**, *ACL 2018*, Hongge Chen*, Huan Zhang*, Pin-Yu Chen, Jinfeng Yi, Cho-Jui Hsieh.

[16] **Is Robustness the Cost of Accuracy? Lessons Learned from 18 Deep Image Classifiers**, *ECCV 2018*, Dong Su*, Huan Zhang*, Hongge Chen, Jinfeng Yi, Pin-Yu Chen, Yupeng Gao. (* Equal contribution).

[17] **Towards Robust Neural Networks via Random Self-ensemble**, *ECCV 2018*, Xuanqing Liu, Minhao Cheng, Huan Zhang, Cho-Jui Hsieh.

[18] **ZOO: Zeroth Order Optimization based Black-box Attacks to Deep Neural Networks without Training Substitute Models**, *10th ACM Workshop on Artificial Intelligence and Security, 2017*, Pin-Yu Chen*, Huan Zhang*, Yash Sharma, Jinfeng Yi and Cho-Jui Hsieh (* Equal contribution).

[19] **GPU-acceleration for Large-scale Tree Boosting**, *SysML Conference 2018*, Huan Zhang, Si Si, Cho-Jui Hsieh.

[20] **Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent**, *NIPS 2017*, Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, Ji Liu.

[21] **Gradient Boosted Decision Trees for High Dimensional Sparse Output**, *ICML 2017*, Si Si, Huan Zhang, Sathiya Keerthi, Dhruv Mahajan, Inderjit Dhillon, Cho-Jui Hsieh.

[22] **HogWild++: A New Mechanism for Decentralized Asynchronous Stochastic Gradient Descent**, *ICDM 2016*, Huan Zhang, Cho-Jui Hsieh, Venkatesh Akella.

[23] **Fixing the Convergence Problems in Parallel Asynchronous Dual Coordinate Descent**, *ICDM 2016*, Huan Zhang, Cho-Jui Hsieh.

[24] **A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order**, *NIPS 2016*, Xiangru Lian, Huan Zhang, Cho-Jui Hsieh, Yijun Huang, Ji Liu.

[25] **Sublinear Time Orthogonal Tensor Decomposition**, *NIPS 2016*, Zhao Song, David P. Woodruff, Huan Zhang (alphabetical order).

Open Source Projects

LightGBM on GPU, https://github.com/huanzhang12/lightgbm-gpu.
I developed a GPU accelerated algorithm and integrated it into LightGBM, a popular tree boosting package by Microsoft with high efficiency on large-scale datasets. I am a maintainer of LightGBM official repository.

auto_LiRPA, https://github.com/KaidiXu/auto_LiRPA.
I am leading the development of `auto_LiRPA`, a Pytorch based library for provable perturbation analysis of general neural network architectures. It is an essential tool for safe machine learning with provable guarantees.

List of Awards

IBM PhD Fellowship, 2018-2019.

Student Travel Award, NeurIPS, ICML, ICLR, ICDM.

Meritorious Winner, The U.S. Mathematical Contest in Modeling, 2010.

Academia Service

Program Committee/Reviewer, NeurIPS 2016-2020; ICLR 2019-2021; ICML 2019, 2020; CVPR 2020, 2021; AAAI 2020, 2021.