Huan Zhang

Post-doctoral Fellow, Carnegie Mellon University (CMU)

✉ *huan@huan-zhang.com*
5000 Forbes Ave
Pittsburgh, PA 15213

Work Experience

Post-Doctoral Fellow at Carnegie Mellon University Pittsburgh, PA, Jan 2021 - Present.

•*Supervisor: J. Zico Kolter (zkolter@cs.cmu.edu)*

I am currently a postdoctoral researcher at CMU working on formal verification of machine learning and building trustworthy artificial intelligence.

Internship at Google DeepMind London, UK, Jun 2019 - Nov 2019.

•*Mentor: Po-Sen Huang (posenhuang@google.com), Pushmeet Kohli (pushmeet@google.com), Krishnamurthy (Dj) Dvijotham (dvij@google.com)*

Research scientist internship on machine learning fairness and robustness.

Internship at 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 formal verification of neural networks.

Internship at Amazon A9.com 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.

Internship at 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.

Internship at Nokia Bell Labs Murray Hill, NJ, June 2015 - Aug 2015, June 2013 - Sep 2013.

•*Supervisor: Noriaki Kaneda and Young-Kai Chen (ykchen@bell-labs.com)*

Internship on building a high speed optical communication system.

Education

Ph.D. in Computer Science **UCLA** (December 2020).

•*Advisor:* Prof. Cho-Jui Hsieh (chohsieh@cs.ucla.edu)

•*Area:* Trustworthy artificial intelligence and formal verification methods for machine learning.

M.S. in Computer Engineering **UC Davis** (September 2014).

•*Advisor:* Prof. Venkatesh Akella (akella@ucdavis.edu)

•*Area:* computer architecture and parallel computing.

Bachelor of Engineering **Zhejiang University** (June 2012).

•*Major:* Information Engineering

Publications

Google Scholar Profile: `https://scholar.google.com/citations?user=LTa3GzEAAAAJ`

Citations: **7517**; h-index: **36**; i10-index: **43** (as of Aug 5, 2022)

▪ *Conference papers* (***** indicates **co-first** authors)

1. A Branch and Bound Framework for Stronger Adversarial Attacks of ReLU Networks. **Huan Zhang***, Shiqi Wang*, Kaidi Xu, Yihan Wang, Suman Jana, Cho-Jui Hsieh and Zico Kolter, *International Conference on Machine Learning (ICML)*, 2022.

2. Linearity Grafting: Relaxed Neuron Pruning Helps Certifiable Robustness. Tianlong Chen*, **Huan Zhang***, Zhenyu Zhang, Shiyu Chang, Sijia Liu, Pin-Yu Chen and Zhangyang Wang, *International Conference on Machine Learning (ICML)*, 2022.

3. COPA: Certifying Robust Policies for Offline Reinforcement Learning against Poisoning Attacks. Fan Wu, Linyi

Li, Chejian Xu, **Huan Zhang**, Bhavya Kailkhura, Krishnaram Kenthapadi, Ding Zhao and Bo Li, *International Conference on Learning Representations (ICLR)*, 2022.

4. Beta-CROWN: Efficient Bound Propagation with Per-neuron Split Constraints for Complete and Incomplete Neural Network Verification. Shiqi Wang*, **Huan Zhang***, Kaidi Xu*, Xue Lin, Suman Jana, Cho-Jui Hsieh and Zico Kolter, *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

5. Training Certifiably Robust Neural Networks with Efficient Local Lipschitz Bounds. Yujia Huang, **Huan Zhang**, Yuanyuan Shi, Zico Kolter and Anima Anandkumar, *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

6. Robustness Between the Worst and Average Case. Leslie Rice, Anna Bair, **Huan Zhang**, and Zico Kolter, *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

7. Fast Certified Robust Training via Better Initialization and Shorter Warmup. Zhouxing Shi*, Yihan Wang*, **Huan Zhang**, Jinfeng Yi and Cho-Jui Hsieh, *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

8. Robust Reinforcement Learning on State Observations with Learned Optimal Adversary. **Huan Zhang***, Hongge Chen*, Duane Boning, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*, 2021.

9. Fast and Complete: Enabling Complete Neural Network Verification with Rapid and Massively Parallel Incomplete Verifiers. Kaidi Xu*, **Huan Zhang***, Shiqi Wang, Yihan Wang, Suman Jana, Xue Lin, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*, 2021.

10. Double Perturbation: On the Robustness of Robustness and Counterfactual Bias Evaluation. Chong Zhang, Jieyu Zhao, **Huan Zhang**, Kai-Wei Chang, Cho-Jui Hsieh, *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.

11. Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations. **Huan Zhang***, Hongge Chen*, Chaowei Xiao, Bo Li, Duane Boning, Cho-Jui Hsieh, *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

12. Automatic Perturbation Analysis for Scalable Certified Robustness and Beyond. Kaidi Xu*, Zhouxing Shi*, **Huan Zhang***, Yihan Wang, Minlie Huang, Kai-Wei Chang, Bhavya Kailkhura, Xue Lin, Cho-Jui Hsieh, *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

13. An Efficient Adversarial Attack for Tree Ensembles. Chong Zhang, **Huan Zhang**, Cho-Jui Hsieh, *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

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

15. On ℓ_p-norm Robustness of Ensemble Decision Stumps and Trees. Yihan Wang, **Huan Zhang**, Hongge Chen, Duane Boning and Cho-Jui Hsieh, *International Conference on Machine Learning (ICML)*, 2020.

16. Towards Stable and Efficient Training of Verifiably Robust Neural Networks. **Huan Zhang**, Hongge Chen, Chaowei Xiao, Sven Gowal, Robert Stanforth, Bo Li, Duane Boning, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*, 2020.

17. Robustness Verification for Transformers. Zhouxing Shi, **Huan Zhang**, Kai-Wei Chang, Minlie Huang, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*, 2020.

18. MACER: Attack-free and Scalable Robust Training via Maximizing Certified Radius. Runtian Zhai, Chen Dan, Di He, **Huan Zhang**, Boqing Gong, Pradeep Ravikumar, Cho-Jui Hsieh, Liwei Wang, *International Conference on Learning Representations (ICLR)*, 2020.

19. Robustness Verification of Tree-based Models. Hongge Chen*, **Huan Zhang***, Si Si, Yang Li, Duane Boning, Cho-Jui Hsieh., *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

20. A Convex Relaxation Barrier to Tight Robustness Verification of Neural Networks. Hadi Salman, Greg Yang, **Huan Zhang**, Cho-Jui Hsieh, Pengchuan Zhang, *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

21. Provably Robust Deep Learning via Adversarially Trained Smoothed Classifiers. Hadi Salman, Greg Yang, Jerry Li, Pengchuan Zhang, **Huan Zhang**, Ilya Razenshteyn, Sebastien Bubeck, *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

22. The Limitations of Adversarial Training and the Blind-Spot Attack. **Huan Zhang***, Hongge Chen*, Zhao Song, Duane Boning, Inderjit Dhillon, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*,

2019.

23. Query-Efficient Hard-label Black-box Attack: An Optimization-based Approach. Minhao Cheng, Thong Le, Pin-Yu Chen, **Huan Zhang**, Jinfeng Yi, Cho-Jui Hsieh, *International Conference on Learning Representations (ICLR)*, 2019.

24. Structured Adversarial Attack: Towards General Implementation and Better Interpretability. Kaidi Xu, Sijia Liu, Pu Zhao, Pin-Yu Chen, **Huan Zhang**, Quanfu Fan, Deniz Erdogmus, Yanzhi Wang, Xue Lin, *International Conference on Learning Representations (ICLR)*, 2019.

25. Robust Decision Trees Against Adversarial Examples. Hongge Chen, **Huan Zhang**, Duane Boning, Cho-Jui Hsieh, *International Conference on Machine Learning (ICML)*, 2019.

26. Evaluating Robustness of Deep Image Super-Resolution Against Adversarial Attacks. Jun-Ho Choi, **Huan Zhang**, Jun-Hyuk Kim, Cho-Jui Hsieh, Jong-Seok Lee, *International Conference on Computer Vision (ICCV)*, 2019.

27. Second Rethinking of Network Pruning in the Adversarial Setting. Shaokai Ye, Kaidi Xu, Sijia Liu, Hao Cheng, Jan-Henrik Lambrechts, **Huan Zhang**, Aojun Zhou, Kaisheng Ma, Yanzhi Wang, Xue Lin, *International Conference on Computer Vision (ICCV)*, 2019.

28. RecurJac: An Efficient Recursive Algorithm for Bounding Jacobian Matrix of Neural Networks and Its Applications. **Huan Zhang**, Pengchuan Zhang, Cho-Jui Hsieh, *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

29. AutoZOOM: Autoencoder-based Zeroth Order Optimization Method for Attacking Black-box Neural Networks. Chun-Chen Tu, Paishun Ting, Pin-Yu Chen, Sijia Liu, **Huan Zhang**, Jinfeng Yi, Cho-Jui Hsieh, Shin-Ming Cheng, *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

30. Efficient Neural Network Robustness Certification with General Activation Functions. **Huan Zhang***, Tsui-Wei Weng*, Pin-Yu Chen, Cho-Jui Hsieh, Luca Daniel., *Advances in Neural Information Processing Systems (NIPS)*, 2018.

31. Towards Fast Computation of Certified Robustness for ReLU Networks. Tsui-Wei Weng*, **Huan Zhang***, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Duane Boning, Inderjit S. Dhillon, Luca Daniel., *International Conference on Machine Learning (ICML)*, 2018.

32. Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach. Tsui-Wei Weng*, **Huan Zhang***, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, Luca Daniel, *International Conference on Learning Representations (ICLR)*, 2018.

33. Attacking Visual Language Grounding with Adversarial Examples: A Case Study on Neural Image Captioning. Hongge Chen*, **Huan Zhang***, Pin-Yu Chen, Jinfeng Yi, Cho-Jui Hsieh, *56th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018.

34. Is Robustness the Cost of Accuracy? Lessons Learned from 18 Deep Image Classifiers. Dong Su*, **Huan Zhang***, Hongge Chen, Jinfeng Yi, Pin-Yu Chen, Yupeng Gao., *European Conference on Computer Vision (ECCV)*, 2018.

35. Towards Robust Neural Networks via Random Self-ensemble. Xuanqing Liu, Minhao Cheng, **Huan Zhang**, Cho-Jui Hsieh, *European Conference on Computer Vision (ECCV)*, 2018.

36. EAD: Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples. Pin-Yu Chen, Yash Sharma, **Huan Zhang**, Jinfeng Yi and Cho-Jui Hsieh, *In AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

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

38. Gradient Boosted Decision Trees for High Dimensional Sparse Output. Si Si, **Huan Zhang**, Sathiya Keerthi, Dhruv Mahajan, Inderjit Dhillon and Cho-Jui Hsieh, *34th International Conference on Machine Learning (ICML)*, 2017.

39. Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent. Xiangru Lian, Ce Zhang, **Huan Zhang**, Cho-Jui Hsieh, Wei Zhang and Ji Liu, *Advances in Neural Information Processing Systems (NIPS)*, 2017.

40. HogWild++: A New Mechanism for Decentralized Asynchronous Stochastic Gradient Descent. **Huan Zhang**, Cho-Jui Hsieh, Venkatesh Akella, *IEEE International Conference on Data Mining (ICDM)*, 2016.

41. Fixing the Convergence Problems in Parallel Asynchronous Dual Coordinate Descent. **Huan Zhang**, Cho-Jui Hsieh, *IEEE International Conference on Data Mining (ICDM)*, 2016.

42. Sublinear Time Orthogonal Tensor Decomposition. Zhao Song, David P. Woodruff, **Huan Zhang**, *Advances*

in Neural Information Processing Systems (NIPS), 2016.

43. A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order. Xiangru Lian, **Huan Zhang**, Cho-Jui Hsieh, Yijun Huang and Ji Liu, *Advances in Neural Information Processing Systems (NIPS)*, 2016.

- *Workshop papers* (* indicates co-first authors)

44. Enhancing Certifiable Robustness via a Deep Model Ensemble. **Huan Zhang**, Minhao Cheng and Cho-Jui Hsieh, *ICLR 2019 Safe Machine Learning Workshop*, 2019.

45. Realtime Query Completion via Deep Language Models. Po-Wei Wang, **Huan Zhang**, Vijai Mohan, Inderjit S. Dhillon and J. Zico Kolter, *SIGIR Workshop On eCommerce*, 2018.

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

47. Burst Mode Processing: An Architectural Framework for Improving Performance in Future Chip Microprocessors. **Huan Zhang**, Rajeevan Amirtharajah, Christopher Nitta, Matthew Farrens and Venkatesh Akella, *Workshop on Managing Overprovisioned Systems, Co-located with ASPLOS-19*, April 2014.

48. HySIM: Towards a Scalable, Accurate and Fast Simulator for Manycore Processors. Kramer Straube, **Huan Zhang**, Christopher Nitta, Matthew Farrens and Venkatesh Akella, *3rd Workshop on the Intersections of Computer Architecture and Reconfigurable Logic, Co-located with MICRO-46*, December 2013.

Patents

2010 **Blind Guide Device Based on the Smart Phone**, China Patent, ZL.2010 2 0516516.9, Yang Yang, Huan Zhang, Ding Zhao et al.

Open Source Projects

α,β-CROWN: A Neural Network Verification Toolbox (2021-) `http://abCROWN.org`.

I lead the development of α,β-CROWN, an efficient and scalable neural network verification toolbox that won the highest total score in 2nd and 3rd International Verification of Neural Network Competition (VNN-COMP 2021 and 2022).

AutoLiRPA: A Neural Network Perturbation Analysis Library (2020-) `http://PaperCode.cc/AutoLiRPA`.

I lead the development of AutoLiRPA, an easy-to-use library capable of automatically giving provable bounds under input or weight perturbations for complex neural networks and other general computational functions.

LightGBM on GPU (2016-2017) `https://github.com/huanzhang12/lightgbm-gpu`.

LightGBM is a popular tree boosting package with high efficiency on large-scale datasets. I accelerated its decision tree construction process on GPUs with 7 to 8 times speedup. My code reaches production quality and has been merged into the LightGBM official repository.

List of Awards

Top Highest Score Award 3nd International Verification of Neural Networks Competition (VNN-COMP 2022).

Top Highest Score Award 2nd International Verification of Neural Networks Competition (VNN-COMP 2021). I led a multi-institutional team (CMU, Columbia University, UCLA, UIUC and Drexel University) and we developed the α,β-CROWN verification toolbox which won VNN-COMP 2021 and 2022 with the highest total score. VNN-COMP is an important event in the field of formal verification of neural networks, attracting 10+ teams each year from elite universities doing related research across the globe. More details can be found at : `https://cacm.acm.org/careers/255662` and `https://github.com/stanleybak/vnncomp2022`

Adversarial Machine Learning (AdvML) Rising Star Award Sponsored by MIT-IBM Watson AI Lab, 2021, See award details at `https://sites.google.com/view/advml/advml-rising-star-award`

IBM PhD Fellowship 2018-2020, Details at `https://research.ibm.com/university/awards/fellowships.html`

Student Travel Award NIPS 2016, 2017, 2018; ICML 2018; ICDM 2016; ICLR 2018; ACM CCS 2017.

National Merit Scholarship Ministry of Education, China, 2011, awarded to top 2% students.

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

National Merit Scholarship Ministry of Education, China, 2009, awarded to top 2% students.

First Prize China Undergraduate Mathematical Contest in Modeling, 2009.

Second Prize East China Undergraduate Mathematical Contest in Modeling, 2009.

Recent Professional Activities

- *Invited Talks*

2022 **Johns Hopkins University**, Institute for Assured Autonomy Seminar Series, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2022 **Carnegie Mellon University (CMU)**, AI Seminar, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2021 **University of California Santa Barbara (UCSB)**, Computer Science Colloquium, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2021 **Northeastern University**, Security Seminar, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2021 **University of Illinois at Urbana-Champaign (UIUC)**, Computer Science Speakers Series, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2021 **University of Southern California (USC)**, AI Seminar, title "How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers".

2021 **Lorentz Center Workshop on Robust Artificial Intelligence**, title "Robust Reinforcement Learning Against Adversarial Perturbations on State Observations".

2021 **Bosch Center for Artificial Intelligence (BCAI)**, title "Complete and Incomplete Neural Network Verification with Efficient Bound Propagations".

2020 **3rd Workshop on Formal Methods for ML-Enabled Autonomous Systems**, title "Robustness Verification for Ensemble Stumps and Trees".

- *Workshop/Tutorial Organizer*

2022 **Workshop (lead organizer)**, 1st Workshop on Formal Verification of Machine Learning, co-located with ICML 2022.

2022 **Workshop (lead organizer)**, Queer in AI Workshop, ICML 2022.

2022 **Workshop**, Workshop on Socially Responsible Machine Learning, co-located with ICLR 2022.

2022 **Tutorial (lead organzier)**, Formal Verification of Deep Neural Networks: Theory and Practice, AAAI 2022.

2021 **Workshop**, Workshop on Security and Reliability of Machine Learning, co-located with ATVA 2021.

2021 **Tutorial**, Lorentz Center Workshop on Robust Artificial Intelligence, title "auto_LiRPA: An Automatic Neural Network Verification Library".

- *Guest Lectures*

2020 **University of Nebraska Lincoln**, title "CROWN: A Linear Relaxation Framework for Neural Network Verification".

2020 **Stony Brook University**, title "Complete and Incomplete Neural Network Verification with Efficient Bound Propagations".

- *Reviewers, Editors*

Conference Paper Reviewer/Program Committee, NIPS 2016, 2018, 2019, 2020, 2021, 2022; ICML 2019, 2020, 2021; ICLR 2019, 2020, 2021, 2022; AAAI 2020, 2021; UAI 2020, 2021; AISTATS 2021, 2022; CVPR 2020, 2021. USENIX 2020..

Senior Program Committee/Area Chair, AAAI 2022.

Journal Reviewer, Journal of Machine Learning Research (JMLR), IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), Springer Journal of Machine Learning.

Journal Editor, Special Issue "Black-Box Algorithms and Their Applications", MDPI Algorithms, 2021.

Journal Editor, Trustworthy Machine Learning Research Topic, Frontiers in Big Data, 2021.