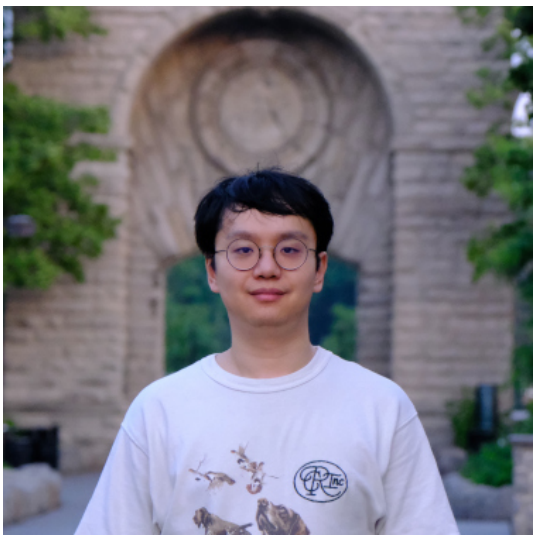

**Huan Zhang**
Postdoctoral Researcher
Department of Computer Science
Carnegie Mellon University (CMU)
Homepage: https://huan-zhang.com/

**Biography**: Huan Zhang is a postdoctoral researcher at Carnegie Mellon University, supervised by Prof. Zico Kolter. He received his Ph.D. degree in Computer Science at UCLA in 2020, advised by Prof. Cho-Jui Hsieh. Huan's research focuses on the robustness and trustworthiness of artificial intelligence (AI), especially on using formal verification and provable methods to evaluate and enhance the robustness of machine learning models, such as deep neural networks and tree ensembles. Huan systematically studied the robustness of many machine learning scenarios including reinforcement learning, natural language processing, and image generation. Huan led the development of α,β-CROWN, a toolbox for neural network robustness verification, which won the first prize in Verification of Neural Networks Competition (VNN-COMP'21). Huan Zhang was awarded an IBM Ph.D. fellowship during 2018 - 2020 and the 2021 AdvML Rising Star Award sponsored by MIT-IBM Watson AI Lab.

**Title**: How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers

**Abstract**: Neural networks have become a crucial element in modern artificial intelligence. However, they are often black-boxes and can behave unexpectedly and produce surprisingly wrong results under malicious inputs. When applying neural networks to mission-critical systems such as autonomous driving and aircraft control, it is often desirable to formally verify their trustworthiness such as safety and robustness. Unfortunately, the complexity of neural networks has made the task of formally verifying their properties very challenging. To tackle this challenge, I first propose an efficient verification algorithm based on linear relaxations of neural networks, which produces guaranteed output bounds given bounded input perturbations. The algorithm propagates linear inequalities through the network efficiently in a backward manner and can be applied to arbitrary network architectures. To reduce relaxation errors, I develop an efficient optimization procedure that can tighten verification bounds rapidly on machine learning accelerators such as GPUs. Lastly, I discuss

how to further empower the verifier with branch and bound by incorporating the additional branching constraints into the bound propagation procedure. The combination of these advanced neural network verification techniques leads to α,β-CROWN (alpha-beta-CROWN), a scalable, powerful and GPU-based neural network verifier that won the 2nd International Verification of Neural Networks Competition (VNN-COMP'21) with the highest total score.