Exploiting Monotonicity for Training and Verification of Neural Networks

Background

Neural networks are great at solving many complex tasks, e.g., object detection [10], natural language processing [11], or chess [8]. To ensure the safety of neural networks in safety-critical environments, such as autonomous driving [12], there is a large field of research around the formal verification of neural networks [1]. However, the formal verification of neural networks is computationally hard [5]. Generally, there are two approaches to formally verify neural networks: The first approach encodes the specification and the neural networks as an optimization problem and applies an off-the-shelf solver, e.g., (mixed-integer) linear programming [3] or a satisfiability modulo theories solver [5]; The second approach computes a tight enclosure of the output set of a neural network, e.g., using interval bound propagation [4] or by propagating more expressive set representations like zonotopes [9, 7]. Recent research exploit monotonicity in neural networks to speed-up and improve the formal verification [6].

Description

For monotone functions, exact output bounds can be computed by propagating input bounds. Neural networks can be made monotone by small modification of their architecture; recent research shows promising results by making entire neural networks monotone [6]. However, such monotonic neural networks can only approximate monotone functions. In this thesis, we want to compose multiple monotonic neural networks to approximate arbitrary non-monotonic functions. The composition can be viewed as a mixed-monotone system [2]. Moreover, we want to investigate the performance of compositional monotonic neural networks and find efficient methods to exploit the mixed-monotonic behavior for their formal verification.

Tasks:

1. Literature research on formal verification of neural networks.
2. Implementation of monotonic neural network architecture and different methods for their composition.
3. Training different compositional monotonic neural networks and comparing the performance on monotonic and non-monotonic tasks.
4. Development of efficient algorithms for the formal verification of compositional monotonic neural networks.
5. Extensive evaluation and comparison with existing approaches for formal verification of neural networks.

References

[1] Christopher Brix, Mark Niklas Müller, Stanley Bak, Taylor T. Johnson, and Changliu Liu. First three years of the international verification of neural networks competition (VNN-COMP). *Int. Journal on Software Tools for Technology Transfer*, 25(3):329–339, 2023.

[2] Samuel Coogan and Murat Arcak. Efficient finite abstraction of mixed monotone systems. In *Proc. of the Int, Conf. on Hybrid Systems: Computation and Control (HSCC)*, pages 58–67, 2015.

[3] Claudio Ferrari, Mark Niklas Müller, Nikola Jovanović, and Martin Vechev. Complete verification via multi-neuron relaxation guided branch-and-bound. In *Proc. of the Int. Conf. on Learning Representations (ICLR)*, 2022.

[4] Sven Gowal, Krishnamurthy Dvijotham, Robert Stanforth, Rudy Bunel, Chongli Qin, Jonathan Uesato, Relja Arandjelovic, Timothy Arthur Mann, and Pushmeet Kohli. Scalable verified training for provably robust image classification. In *Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, pages 4841–4850, 2019.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Lukas Koller, M.Sc.

Research project:
DFG - SPP 2422

Type:
BT

Research area:
Formal Verification of Neural Networks

Programming language:
MATLAB

Required skills:
Machine Learning, Formal Methods

Language:
English

Date of submission:
16. Dezember 2024

For more information please contact us:

Phone: +49 (89) 289 - 18140

E-Mail: lukas.koller@tum.de

Website: ce.cit.tum.de/cps/

[5] Guy Katz, Clark Barrett, David L. Dill, Kyle Julian, and Mykel J. Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In *Int. Conf. on Computer Aided Verification (CAV)*, pages 97–117, 2017.

[6] Ouail Kitouni, Niklas Nolte, and Mike Williams. Robust and provably monotonic networks. *Machine Learning: Science and Technology*, 4(3), 2023.

[7] Niklas Kochdumper, Christian Schilling, Matthias Althoff, and Stanley Bak. Open- and closed-loop neural network verification using polynomial zonotopes. In *NASA Formal Methods*, pages 16–36, 2023.

[8] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[9] Gagandeep Singh, Timon Gehr, Matthew Mirman, Markus Püschel, and Martin Vechev. Fast and effective robustness certification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[10] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 7464–7475, 2023.

[11] Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136, 2023.

[12] Cunliang Ye, Yongfu Wang, Yunlong Wang, and Ming Tie. Steering angle prediction yolov5-based end-to-end adaptive neural network control for autonomous vehicles. *Proc. of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 236(9):1991–2011, 2022.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems