PdF:
Exploiting LP Infeasibility Certificates for Efficient and Trustworthy NN Verification

Background

Neural networks are State-of-the-Art in numerous machine learning tasks ranging from image classification to beating world-class human players in the games of Chess and Go. These achievements have led to the implementation of neural networks in safety-critical domains like aircraft controllers [3] and autonomous driving. In these areas, failures can have dramatic consequences. Therefore, provable guarantees about the behavior of the corresponding neural networks are necessary. However, we can only provide guarantees for Neural Networks in practice, if the techniques providing these guarantees are, both, trustworthy and scalable.

Idea

Prior work has shown how neural network verification tools can be extended to provide proof certificates for their verification results [2]. On the one hand, proof production can be used to build formally verified proof checkers which can be used to increase the trust in the verification results provided by a solver [1]. On the other hand, proof production also opens the door to Conflict Driven Clause Learning [5]: During the exhaustive exploration of a neural network's input space, partial proof certificates are potentially reusable in different input regions which can speed up neural network verification. While there exists some initial work towards CDCL for NN verification [4], efficiently exploiting infeasibility certificates for NN verification turns out to be more challenging than its propositional counterpart. A major reason for this lies in the cost of computing small, ideally irreducible, infeasible subsets.

Project Plan

Following an in-depth literature research on pre-existing approaches, this project can be taken in two directions: We could either come up with a more efficient approach to use infeasibility certificates for Conflict Driven Clause Learning, or we could explore how infeasibility certificates can help to reduce the workload of verified proof-checkers to make verified proof-checking more computationally feasible.

Supervisors

Samuel Teuber, teuber@kit.edu, Room 203 (Geb. 50.34) Philipp Kern, philipp.kern@kit.edu, Room 203 (Geb. 50.34)

References

[1] Remi Desmartin et al. "A Certified Proof Checker for Deep Neural Network Verification". In: *arXiv preprint arXiv:2405.10611* (2024).

[2] Omri Isac et al. "Neural Network Verification with Proof Production". In: *2022 Formal Methods in Computer-Aided Design (FMCAD)* (2022), pp. 38–48.

[3] Kyle D. Julian et al. "Guaranteeing Safety for Neural Network-Based Aircraft Collision Avoidance Systems". In: *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. 2019, pp. 1–10. DOI: 10.1109/DASC43569.2019.9081748.

[4] Zongxin Liu et al. "DeepCDCL: A CDCL-based Neural Network Verification Framework". In: *Theoretical Aspects of Software Engineering - 18th International Symposium, TASE 2024, Guiyang, China, July 29 - August 1, 2024, Proceedings*. Ed. by Wei-Ngan Chin et al. Vol. 14777. Lecture Notes in Computer Science. Springer,

2024, pp. 343–355. DOI: 10.1007/978-3-031-64626-3_20. URL: https://doi.org/10.1007/978-3-031-64626-3%5C_20.

[5] João P. Marques Silva et al. “GRASP - a new search algorithm for satisfiability”. In: *Proceedings of the 1996 IEEE/ACM International Conference on Computer-Aided Design, ICCAD 1996, San Jose, CA, USA, November 10-14, 1996*. Ed. by Rob A. Rutenbar et al. IEEE Computer Society / ACM, 1996, pp. 220–227. DOI: 10.1109/ICCAD.1996.569607. URL: https://doi.org/10.1109/ICCAD.1996.569607.