Karlsruher Institut für Technologie

Application-oriented Formal Verification Research Group
Institute of Information Security and Dependability (KASTEL)
Am Fasanengarten 5
76131 Karlsruhe
https://formal.kastel.kit.edu

Differential Verification for Computational Graphs

Bachelor Thesis

Neural networks (NNs) are frequently retrained when new data becomes available or compressed to improve execution speed. In such cases, it is important to ensure that the new NN behaves similarly to the old one, at least in certain regions of the input space. To achieve this, differential verification approaches analyze differences in the NNs' weight matrices and employ custom abstractions for differences of activation functions.
Prior work has resulted in the implementation of tools such as ReluDiff [2] and NeuroDiff [1], which have been developed completely from scratch. And are therefore missing out on many features that contribute to the success of established NN verifiers. However, reasoning over weight matrix differences can be reformulated using computational graphs, which can be automatically handled by powerful existing NN verifiers like AUTO_LIRPA [3]. By framing the problem in this way, we hope to leverage potential performance benefits from these well-established tools.

Task.

- Study existing literature on differential verification.
- Implement a transformation of weight matrix differences into a computational graph for differential verification in PyTorch
- Develop and implement an optimizable abstraction of $\mathrm{ReLU}(x) - \mathrm{ReLU}(y)$ within the AUTO_LIRPA framework.

Keywords: Neural Network Verification, Differential Verification, Equivalence Verification

References

[1] Brandon Paulsen et al. "NeuroDiff: Scalable Differential Verification of Neural Networks using Fine-Grained Approximation". In: *35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020, Melbourne, Australia, September 21-25, 2020.* IEEE, 2020, pp. 784–796. DOI: 10.1145/3324884.3416560.

[2] Brandon Paulsen et al. "ReluDiff: differential verification of deep neural networks". In: *ICSE '20: 42nd International Conference on Software Engineering, Seoul, South Korea, 27 June - 19 July, 2020.* Ed. by Gregg Rothermel et al. ACM, 2020, pp. 714–726. DOI: 10.1145/3377811.3380337.

[3] Kaidi Xu et al. "Automatic Perturbation Analysis for Scalable Certified Robustness and Beyond". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual.* Ed. by Hugo Larochelle et al. 2020.

Supervisors:
Philipp Kern, philipp.kern@kit.edu, Room 203 (Bldg. 50.34)