Efficient Certified Reasoning for Binarized Neural Networks

Jiong Yang
Georgia Institute of Technology, United States

Yong Kiam Tan
Institute for Infocomm Research (I^2R), A*STAR, Singapore
Nanyang Technological University, Singapore

Mate Soos
University of Toronto, Canada

Magnus O. Myreen
Chalmers University of Technology, Sweden
University of Gothenburg, Sweden

Kuldeep S. Meel
Georgia Institute of Technology, United States
University of Toronto, Canada

Abstract

Neural networks have emerged as essential components in safety-critical applications—these use cases demand complex, yet trustworthy computations. Binarized Neural Networks (BNNs) are a type of neural network where each neuron is constrained to a Boolean value; they are particularly well-suited for safety-critical tasks because they retain much of the computational capacities of full-scale (floating-point or quantized) deep neural networks, but remain compatible with satisfiability solvers for qualitative verification and with model counters for quantitative reasoning. However, existing methods for BNN analysis suffer from either limited scalability or susceptibility to soundness errors, which hinders their applicability in real-world scenarios.

In this work, we present a scalable and trustworthy approach for both qualitative and quantitative verification of BNNs. Our approach introduces a native representation of BNN constraints in a custom-designed solver for qualitative reasoning, and in an approximate model counter for quantitative reasoning. We further develop specialized proof generation and checking pipelines with native support for BNN constraint reasoning, ensuring trustworthiness for all of our verification results. Empirical evaluations on a BNN robustness verification benchmark suite demonstrate that our certified solving approach achieves a 9× speedup over prior certified CNF and PB-based approaches, and our certified counting approach achieves a 218× speedup over the existing CNF-based baseline. In terms of coverage, our pipeline produces fully certified results for 99% and 86% of the qualitative and quantitative reasoning queries on BNNs, respectively. This is in sharp contrast to the best existing baselines which can fully certify only 62% and 4% of the queries, respectively.

2012 ACM Subject Classification Theory of computation → Logic and verification

Keywords and phrases Neural network verification, proof certification, SAT solving, model counting

The Boolean nature of BNNs allows both the networks and their specifications to be encoded as Boolean formulas which, in turn, enables the use of off-the-shelf SAT solvers and model counters to verify those specifications. Unfortunately, existing combinatorial solving methods for such analyses suffer from limited scalability. In contrast, custom methods for verifying input-output specifications scale to much larger (and more general) neural networks [6, 14, 16], but are susceptible to errors; in the annual neural network verification competitions, participating tools have been repeatedly shown to produce incorrect conclusions. *Certified reasoning* [9], where a tool generates both a conclusion and an independently-checkable proof of that conclusion, has long become a mainstay of trustworthiness for modern SAT solvers [15, 5, 7, 11] but it remains nascent for neural network verification tools [16].

In this work, we address the challenges of scalability and trustworthiness in BNN verification by developing an efficient and certified approach for both qualitative and quantitative reasoning on BNNs. Our contributions are as follows.

- For qualitative reasoning, we integrate a native representation of BNN constraints into a modern CNF-XOR SAT solver [10] to enhance its BNN solving efficiency.
- We then extend the solver's associated UNSAT proof format and verified proof checker with native support for BNN constraints, together enabling an efficient solving and certification pipeline for CNF-XOR-BNN formulas.
- Building upon this pipeline, we further develop a certified approximate model counter for quantitative reasoning over BNNs, leveraging native XOR and BNN representations for both effective reasoning and certification.

Empirically, our approach achieves state-of-the-art certified performance in both qualitative and quantitative reasoning for BNNs. Notably, we observe that the compactness of proof certificates with native BNN proof steps enables fast, verified proof checking.

- For qualitative reasoning, our end-to-end approach produced certified answers for 99% of the benchmark UNSAT queries, achieving a 9× speedup over alternative CNF- and PB-based approaches. Figure 1 (left) shows that our solver CryptoMiniSat-BNN, with the checker cake_xlrup-BNN, consistently provides improved solving and checking performance compared to RoundingSat [3] with VeriPB [2] and CaDiCaL [1] with cake_lpr [12].
- For quantitative reasoning, our certified counting approach answered 86% of the queries, with 218× speedup over the baseline which could only fully certify 4% of the queries. Figure 1 (right) presents that our counter ApproxMCCert-BNN offers superior counting runtime and its generated certificates were also readily checked by our checker CertCheck-BNN, compared to the baseline ApproxMCCert with CertCheck [13].

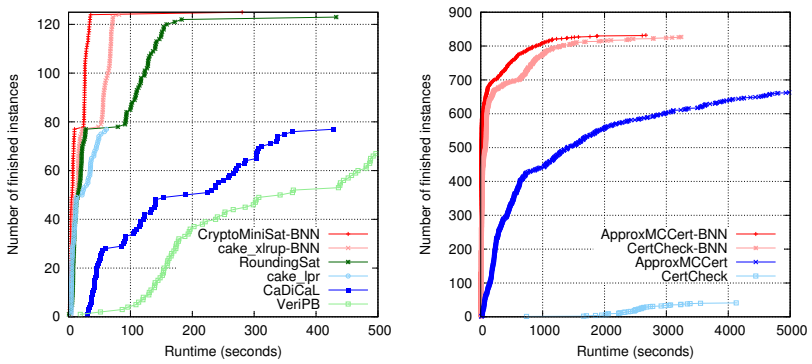

Figure 1 (Left) Runtime comparison of solvers and proof checkers. (Right) Runtime comparison of counters and certificate checkers. The cumulative number of finished instances for a given time limit, i.e., each point (x, y) indicates that the tool completed y instances within x seconds.

By developing solving and certification in tandem, our work offers a trustworthy approach to BNN verification with promising scalability. In fact, during the development of our certification pipeline, we identified and fixed a bug in our solver's implementation of the watching scheme for BNN constraints, which highlights the practical importance of certification.

Looking forward, this framework opens the door to certifying large-scale binarized vision and language models [4, 18], as well as extending certification to quantized neural networks for efficient on-device deployment [17, 8].

References

1. Armin Biere, Katalin Fazekas, Mathias Fleury, and Maximillian Heisinger. CaDiCaL, Kissat, Paracooba, Plingeling and Treengeling entering the SAT Competition 2020. In *Proc. of SAT Competition 2020 – Solver and Benchmark Descriptions*, 2020.
2. Bart Bogaerts, Stephan Gocht, Ciaran McCreesh, and Jakob Nordström. Certified dominance and symmetry breaking for combinatorial optimisation. *Journal of Artificial Intelligence Research*, 2023.
3. Jan Elffers and Jakob Nordström. Divide and conquer: Towards faster pseudo-boolean solving. In *Proc. of IJCAI*, 2018.
4. Yefei He, Zhenyu Lou, Luoming Zhang, Jing Liu, Weijia Wu, Hong Zhou, and Bohan Zhuang. BiViT: Extremely compressed binary vision transformers. In *Proc. of ICCV*, 2023.
5. Marijn Heule, Warren Hunt, Matt Kaufmann, and Nathan Wetzler. Efficient, verified checking of propositional proofs. In *Proc. of ITP*, 2017.
6. Guy Katz, Clark Barrett, David L. Dill, Kyle Julian, and Mykel J. Kochenderfer. Reluplex: An efficient smt solver for verifying deep neural networks. In *Proc. of CAV*, 2017.
7. Peter Lammich. Efficient verified (un)sat certificate checking. In *Proc. of CADE*, 2017.
8. Ji Lin, Jiaming Tang, Haotian Tang, Shang Yang, Wei-Ming Chen, Wei-Chen Wang, Guangxuan Xiao, Xingyu Dang, Chuang Gan, and Song Han. AWQ: Activation-aware weight quantization for llm compression and acceleration. In *Proc. of MLSys*, 2024.
9. R.M. McConnell, K. Mehlhorn, S. Näher, and P. Schweitzer. Certifying algorithms. *Computer Science Review*, 2011.
10. Mate Soos, Karsten Nohl, and Claude Castelluccia. Extending SAT solvers to cryptographic problems. In *Proc. of SAT*, 2009.
11. Yong Kiam Tan, Marijn J. H. Heule, and Magnus O. Myreen. cake_lpr: Verified propagation redundancy checking in cakeml. In *Proc. of TACAS*, 2021.
12. Yong Kiam Tan, Marijn J. H. Heule, and Magnus O. Myreen. Verified propagation redundancy and compositional UNSAT checking in CakeML. *Int. J. Softw. Tools Technol. Transf.*, 2023.
13. Yong Kiam Tan, Jiong Yang, Mate Soos, Magnus O. Myreen, and Kuldeep S. Meel. Formally certified approximate model counting. In *Proc. of CAV*, 2024.
14. Shiqi Wang, Huan Zhang, Kaidi Xu, Xue Lin, Suman Jana, Cho-Jui Hsieh, and J Zico Kolter. Beta-CROWN: Efficient bound propagation with per-neuron split constraints for complete and incomplete neural network verification. In *Proc. of NeurIPS*, 2021.
15. Nathan Wetzler, Marijn J. H. Heule, and Warren A. Hunt. DRAT-trim: Efficient checking and trimming using expressive clausal proofs. In *Proc. of SAT*, 2014.
16. Haoze Wu, Omri Isac, Aleksandar Zeljić, Teruhiro Tagomori, Matthew Daggitt, Wen Kokke, Idan Refaeli, Guy Amir, Kyle Julian, Shahaf Bassan, Pei Huang, Ori Lahav, Min Wu, Min Zhang, Ekaterina Komendantskaya, Guy Katz, and Clark Barrett. Marabou 2.0: A versatile formal analyzer of neural networks. In *Proc. of CAV*, 2024.
17. Guangxuan Xiao, Ji Lin, Mickael Seznec, Hao Wu, Julien Demouth, and Song Han. SmoothQuant: Accurate and efficient post-training quantization for large language models. In *Proc. of ICML*, 2023.
18. Yichi Zhang, Ankush Garg, Yuan Cao, Łukasz Lew, Behrooz Ghorbani, Zhiru Zhang, and Orhan Firat. Binarized neural machine translation. In *Proc. of NeurIPS*, 2023.