

Reproduction of dynaroars/neuralsat: VNN-COMP 2023 ACAS Xu benchmark

K-Veritas Team

Submitted by K-Veritas Team

kv:2610.00040v1 · Published 2026-10-13

Independent reproduction · Benchmark or leaderboard entry · Security and cryptography

Abstract

Ran NeuralSAT (src/main.py, default settings, as vnncomp_scripts/run_instance.sh calls it) at the pinned commit on all 186 instances of the VNN-COMP 2023 acasxu benchmark (45 ACAS Xu networks, properties 1 to 10, per-instance timeout from instances.csv, 116 s for most), one instance at a time, counting unsat, sat, timeout and error verdicts and the wall-clock time per instance. Benchmark files come from the ChristopherBrix/vnncomp2023_benchmarks repository; their SHA-256 hashes are printed at the start of the run. One instance (network 1_9, property 7) did not honour its 116 s timeout: the process ran for about 17.7 hours without writing a result and is counted as error, which accounts for most of the run's duration. The src/example folder (109 MB of sample networks) was moved out since the run does not use it. Newer packages than requirements.txt pins (PyTorch 2.11 with CUDA 12.8, gurobipy 13 with its restricted license). One NVIDIA RTX 5060 Ti.

About this record

Title, authors and abstract were provided by the authors and reviewed for completeness, not for scientific soundness. Every number in this record is copied from the signed data of the sealed reports [2], which remain authoritative. This cover page is not signed. Check it against the SHA-256 listed at <https://kveritas.org/records/2610.00040v1>, and verify each report with its code bundle: `kveritas verify <report.pdf> --bundle <bundle.zip>`.

1 Sealed reports

The experiments were sealed with K-Veritas [1], which binds each result to the code, hardware and time that produced it.

Report	Label	Sealed	Runs	Invocations	Signer
r1 [2]	NeuralSAT: VNN-COMP 202...	2026-10-11	1	1	Signed by the K-Veritas server

2 Results

r1: NeuralSAT: VNN-COMP 2023 ACAS Xu (186 instances)

Run	Metric	Final value	Step
1	instance_wall_s	15.9079	186

Run	Inline claim	Value	Phase
1	acasxu_unsat	138	
1	acasxu_sat	45	
1	acasxu_timeout	2	
1	acasxu_unknown	0	

Run	Inline claim	Value	Phase
1	acasxu_error	1	
1	acasxu_solved	183	
1	acasxu_instances	186	
1	acasxu_solved_mean_wall_s	15.45	

3 Setup

Report	Item	Value
r1	Operating system	linux / amd64
r1	CPU	Intel(R) Xeon(R) W-2145 CPU @ 3.70GHz, 16 cores
r1	GPU	NVIDIA GeForce RTX 5060 Ti, 16311 MiB
r1	Memory	31.0 GB
r1	Total run time	66823.6 s

4 Files

Each report and code bundle must hash to these values. A bundle belongs to the report whose signed data names its hash.

File	SHA-256	Size
r1.pdf	e418af4f44f19d2be78bb09cf00a9bbc044a7491fb72c25d039ac9bdd8ad169c	7.2 MB
r1.kvbundle.zip	5284d74da77ca2025ee1da99c4da131bcb4787f7acfd6f22ccae1fbf8e25a5ec	1.2 MB

5 Links and licence

Paper: <https://arxiv.org/abs/2307.10266>

Reproduction of: <https://github.com/dynaroars/neuralsat/tree/c5097698a32a6dc04d176dcf72b61a931b89cdac>

Record: CC-BY 4.0. Code: MIT, as declared by the authors. Tags: neural-network-verification, vnn-comp, acas-xu, formal-methods.

Listed, not endorsed. Review checks the artifact, never the science. Reports are never edited; changing them makes a new version. Other fields are corrected in place, and every correction is logged in metadata.yaml.

References

[1] M. K. Keita and C. Homan. Computer Science Conferences Should Require Nonrepudiable Experimental Results. NeurIPS 2026 Position Paper Track. arXiv:2605.08586.

[2] Sealed experiment report r1 (NeuralSAT: VNN-COMP 2023 ACAS Xu (186 instances)), kv:2610.00040v1/r1. Data hash 3b3ef9f60b7761126e7fdbbc5d74125b5c270e74bce9f27f4d886441f1916cb0.
<https://kveritas.org/records/2610.00040v1/r1.pdf>