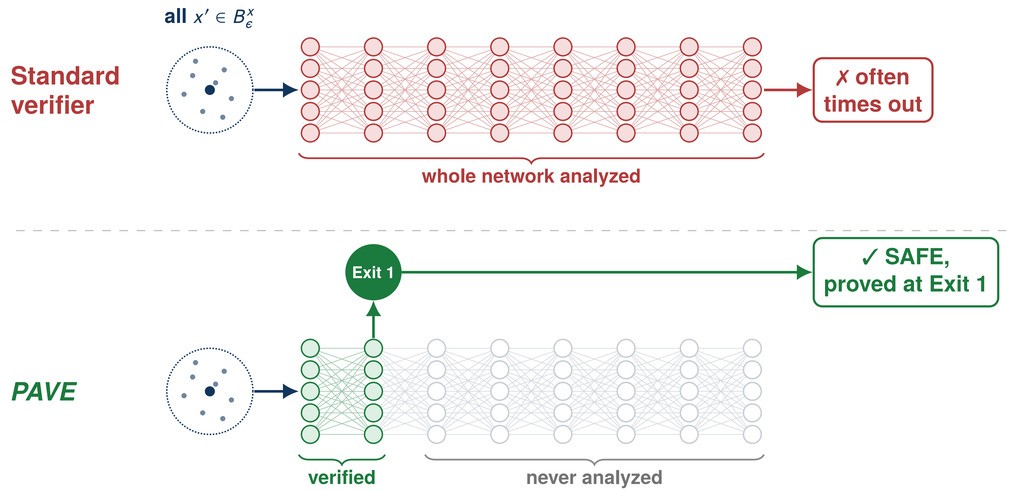
Pareto DNN Verification

Yizhak Y. **Elboher**[1,2,*] Avraham **Raviv**[3,*] Amihay **Elboher**[4]
Zhouxing **Shi**[5] Hillel **Kugler**[3] Guy **Katz**[1]

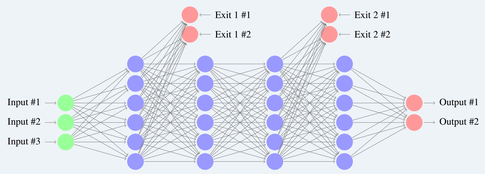
[1]Hebrew University of Jerusalem [2]Masaryk University [3]Bar-Ilan University [4]Ben-Gurion University [5]UC Riverside [*]Equal contribution

Add early exits. *Conditionally* verify exit by exit. Stop early.



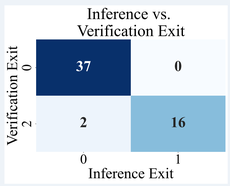
The problem, and the insight

- Proving a network robust is NP-hard. Verifiers often time out on ResNet-18 and VGG-16.
- Yet many inputs are classified confidently after only a few layers, and early-exit networks stop there.

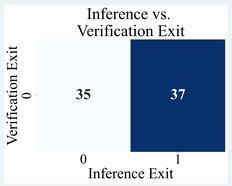


Can verification stop there too?

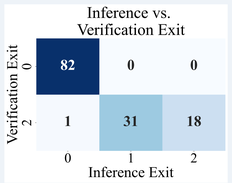
Why it works



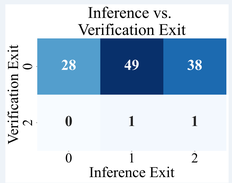
(a) ResNet-18 `SAFE`



(b) ResNet-18 `UNSAFE`



(c) LeNet-5 `SAFE`



(d) LeNet-5 `UNSAFE`

`SAFE` inputs that exit early in inference are also verified at that exit.

Provably: with a stable trace, exact verification is fixed-parameter tractable.

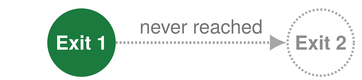
Exits keep accuracy

Thin, frozen classifiers added after training. The base network does not change.

Model	Base	+ Exits
FC-6 / MNIST	98.18	98.2
LeNet-5 / CIFAR-10	70.02	69.93
ResNet-18 / CIFAR-10	86.43	86.11
VGG-16 / CIFAR-10	93.45	93.14

Conditional queries

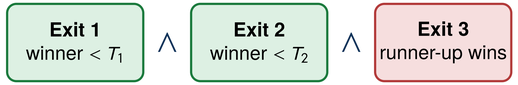
The runner-up wins at Exit 2.
Is that a counterexample?



Only if every earlier exit stayed silent.
Sound and complete whenever the verifier underneath is.

One query, many parts

One output head became many.
What does one check look like now?



Many simple queries become one parallelizable query with many parts, spanning several output heads at once.

≈80%
of queries solved in a fraction of the time

55 vs 24
VGG-16 queries solved vs. direct verification

10×
faster `SAFE` proofs at best, from two sound shortcuts

≈0.3 pt
largest accuracy drop from adding exits

NEURAL INFORMATION PROCESSING SYSTEMS **2026**

Paper

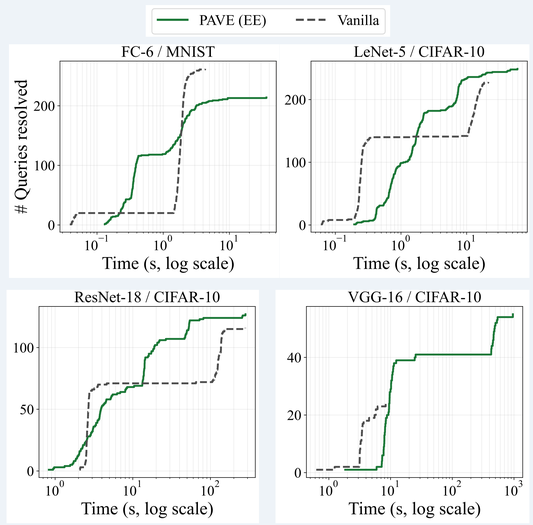
PAVE: verify exit by exit

Check each exit's sub-network in order, with α-β-CROWN underneath. Two sound shortcuts:

- **Break:** winner always above the threshold, and no earlier counterexample: `SAFE`, stop.
- **Continue:** winner always above $1 - T$: no runner-up can win here, skip this exit.

Mean `SAFE` time	**without**	**with**
FC-6	13.0 s	**1.1 s**
LeNet-5	56.9 s	**7.3 s**
ResNet-18	409.4 s	**42.6 s**
VGG-16	none proved	**1532.4 s**

More queries solved, and sooner



- Hard mid-range ϵ: LeNet-5 at ϵ=0.005, **39 vs. 2**.
- Verifies **ResNet-18 on CIFAR-100**: 14 samples proved `SAFE` only with exits.
- The gain grows with network size.

Supported in part by the EU (RobustifAI, 101212818; views are the authors' only), Horizon 2020 Bio4Comp (732482), the Israel Science Foundation (190/19) and the Data Science Institute at Bar-Ilan University.