Formal analysis of neural networks

David Hudák

Supervisor: doc. RNDr. Milan Češka, Ph.D.

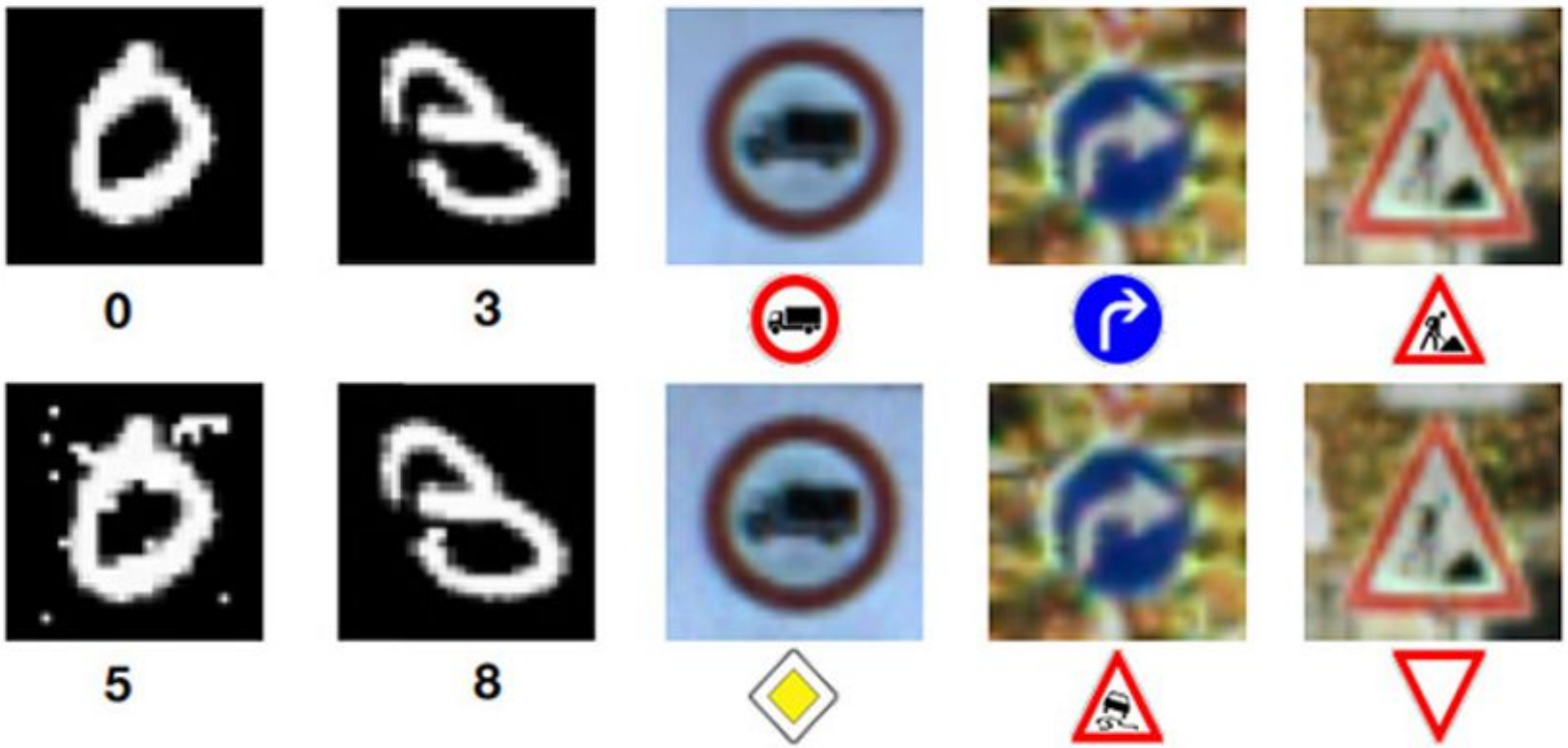
13th June 2022

Motivation

- Usage of neural networks grows
- Safety-critical areas $\rightarrow$ verification required
- Neural network = **black box**
- Existing tools cannot verify commonly used networks
- Robustness – endurance against input perturbations

Motivation

- Not locally robust network can change classification even for small perturbations



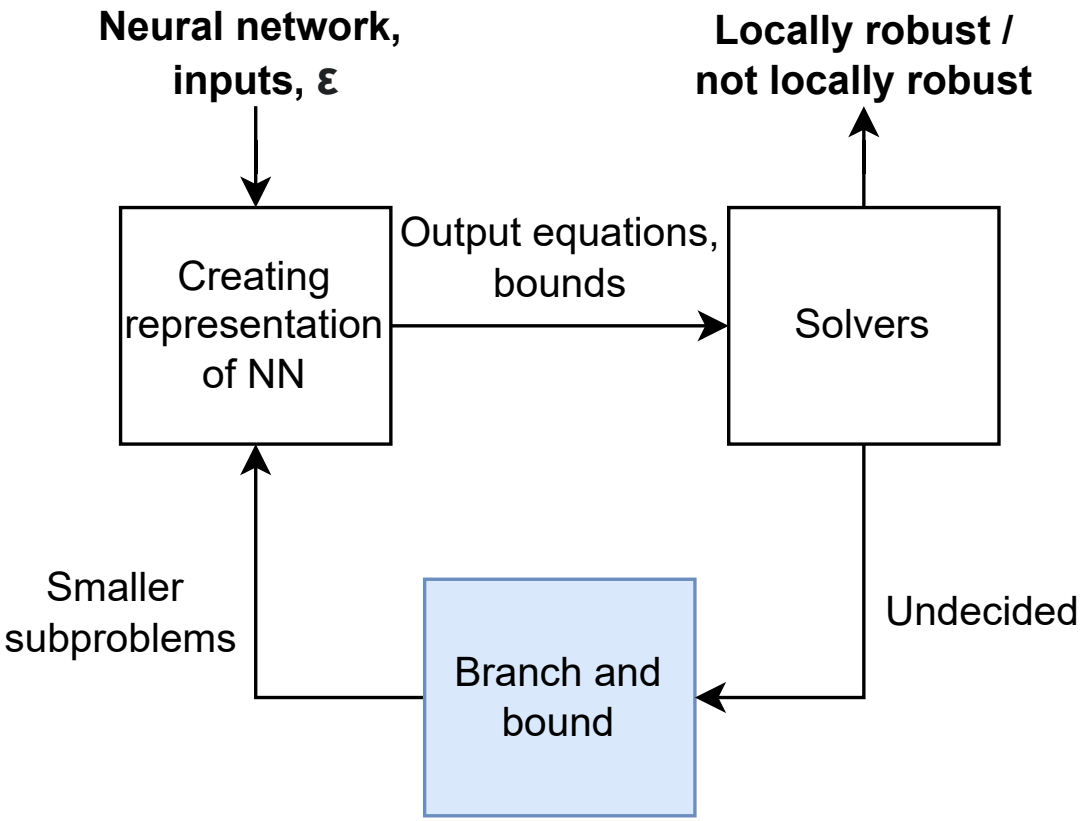
[1]

[1] https://spectrum.ieee.org/slight-street-sign-modifications-can-fool-machine-learning-algorithms

Objectives of the Thesis

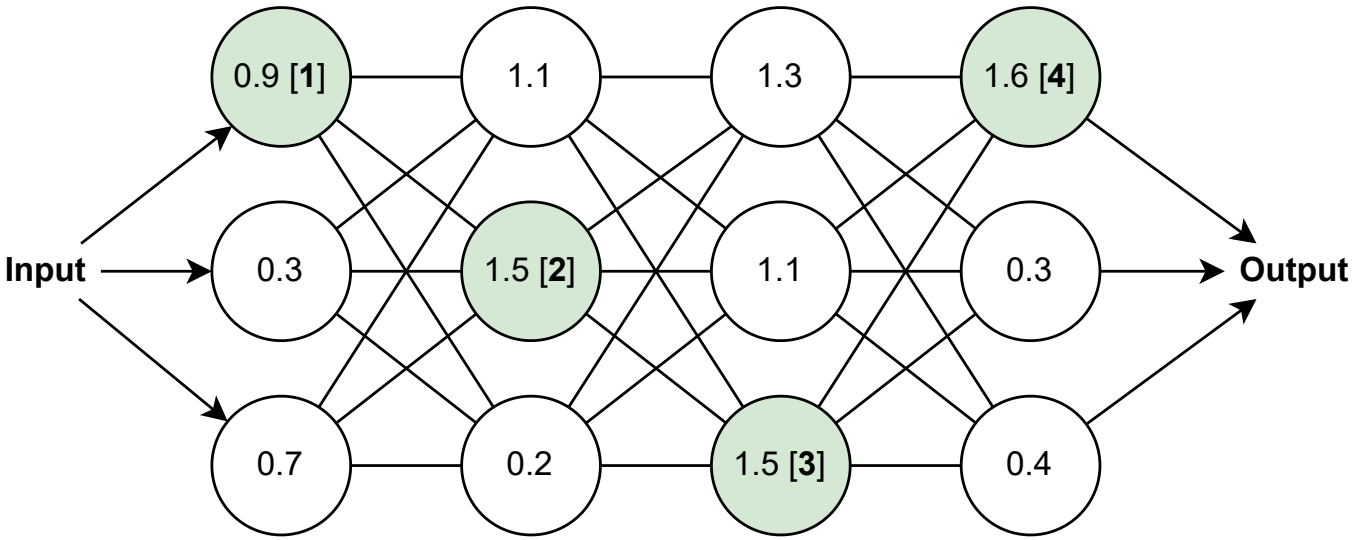
- Design and implementation of improvements to verification algorithms
- Used toolkit – **VeriNet**
  - Symbolic approach to local robustness verification
- Goals
  - Extend the set of verifiable cases in a given timeouts
  - Improve the solution of current verifiable cases
- Outputs
  - New universally applicable branch and bound strategies
  - Extension of the set of experimented networks within VeriNet
  - Analysis of the neural networks verification behavior
  - Faster solution of more complex cases

Main Verification Cycle



Splitting Strategies

- Splitting strategy = controller of a branch and bound algorithm
- Experiments with 5 new different strategies + 1 original



Experimental Results

- Various experiments with 9 different networks
  - Largest networks – sigmoid and tanh networks with 3000 hidden nodes and ReLU network with 1536 nodes
- Discovered unexpected behavior – **branch implosions**
  - For 10×20 ReLU network causes more than 22-fold acceleration of already resolved cases
- Semi-hierarchical strategy shows most significant improvement

Table: Results for three medium-sized networks with timeouts 15 and 30 minutes

Results	Original	Semi-hierarchical
Solved safe cases	582	598
Undecided cases	98	82
Solved unsafe cases	753	753

Thank You for Your Attention!

Network	Original time	Semi-hierarchical time
MNIST 20×40	1800.07 s (†)	0.88 s
MNIST 10×20	1800.07 s (†)	6.98 s
MNIST 10×20	1800.05 s (†)	7.89 s
MNIST 10×20	1800.10 s (†)	1.46 s