Formal Verification of Neural Networks

Tobias Ladner

Prof. Dr.-Ing. Matthias Althoff
Cyber-Physical Systems Group
Technische Universität München

February 5th, 2024
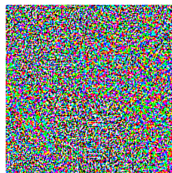
Motivation
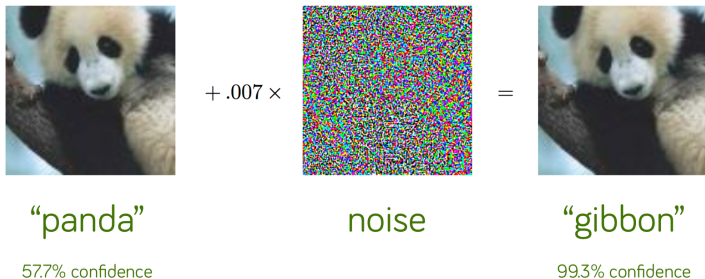

"panda"

57.7% confidence

noise

"gibbon"

99.3% confidence

---
[1] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples". In: *International Conference on Learning Representations*. 2015.
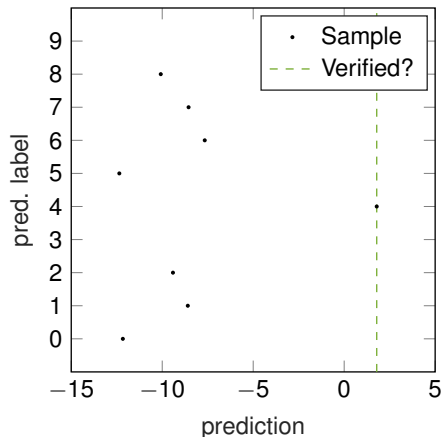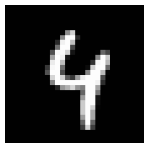
Motivation

$+\,.007\,\times$ =

"panda" noise "gibbon"

57.7% confidence 99.3% confidence

Adversarial examples[1] limit the applicability of neural networks in cyber-physical systems!

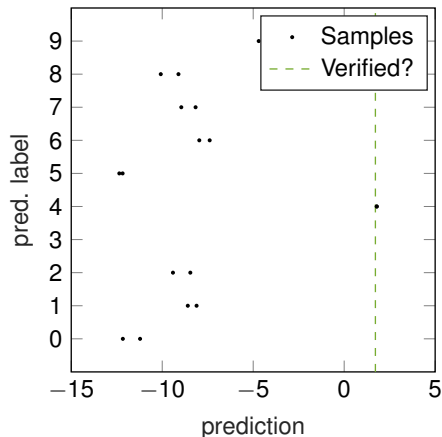[1] Goodfellow, Shlens, and Szegedy, "Explaining and harnessing adversarial examples".
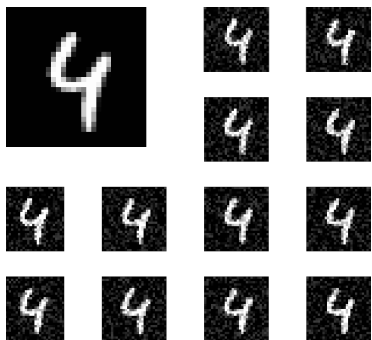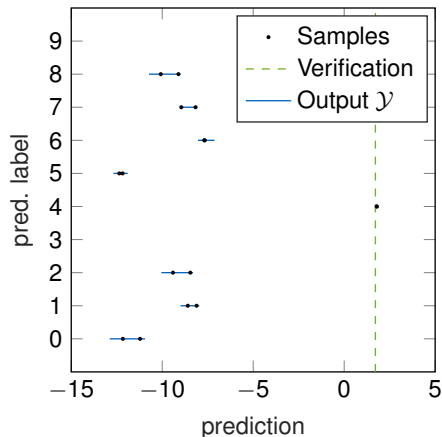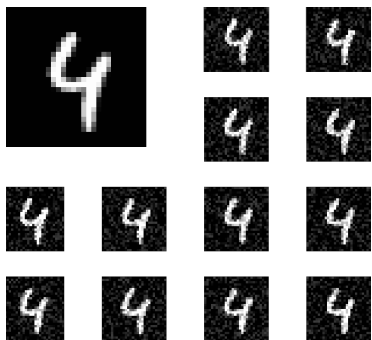
Motivation

Let us demonstrate the formal verification of neural networks by an example:

Motivation

Let us demonstrate the formal verification of neural networks by an example:

Motivation

Let us demonstrate the formal verification of neural networks by an example:

Your tasks:

- Literature review on different techniques to verify neural networks:
  - Set-based techniques
  - Optimization-based techniques

Your tasks:

- Literature review on different techniques to verify neural networks:
    - Set-based techniques
    - Optimization-based techniques

As an example, please check NNV in CORA[2]

Interested? Contact me!

Tobias Ladner
tobias.ladner@tum.de

---

[2]https://cora.in.tum.de/pages/neural-networks/