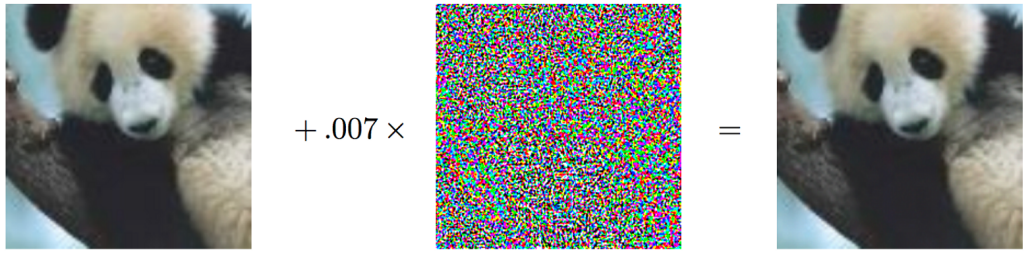
Verification of Neural Networks

Tobias Ladner

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

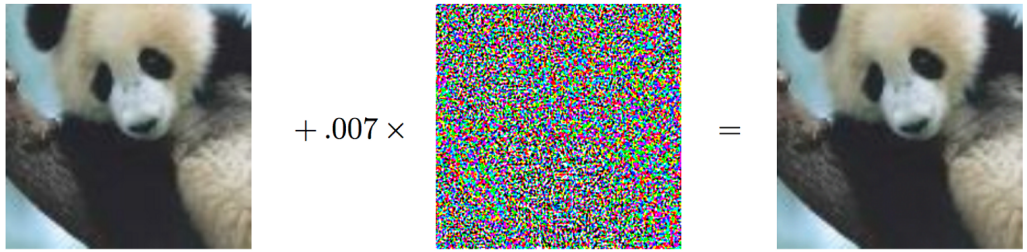
January 31st, 2023

Motivation

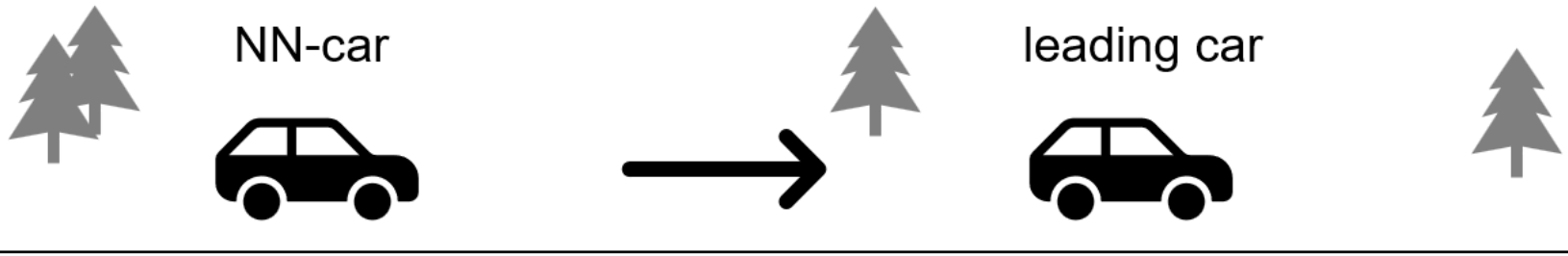


[1] Goodfellow, I. J., Shlens, J., Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

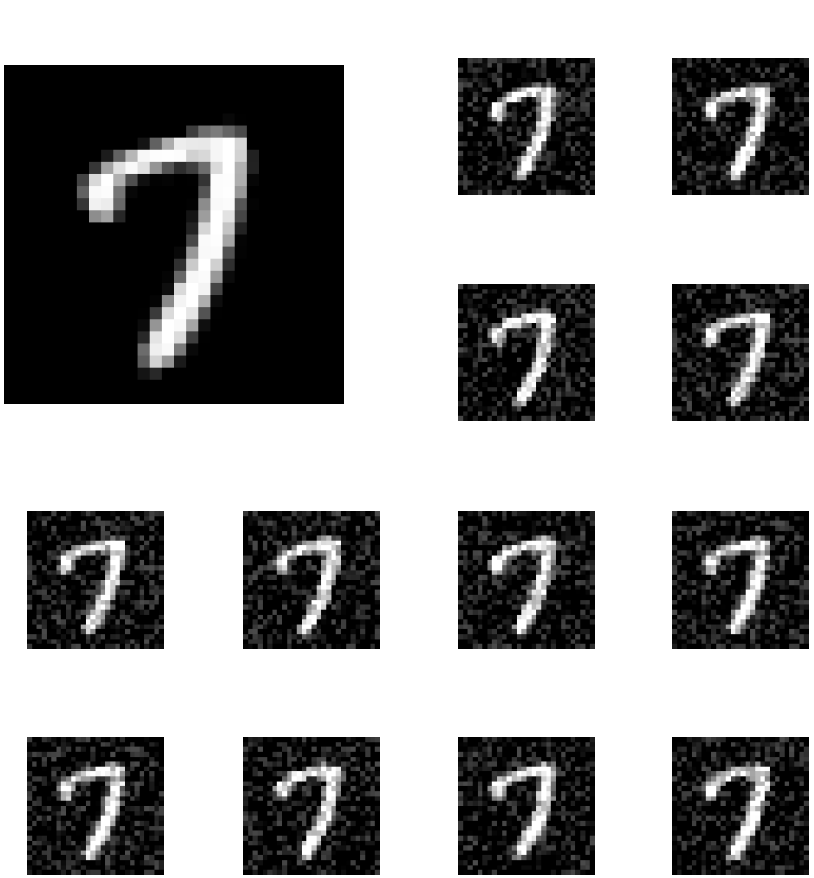
Motivation

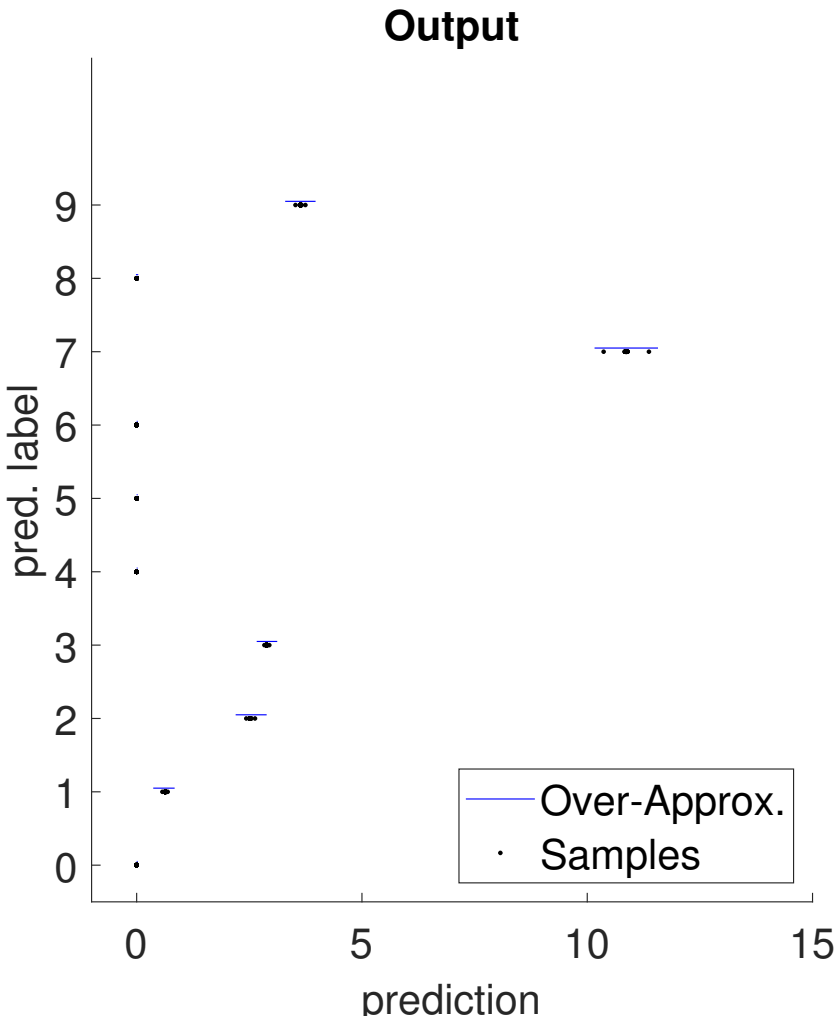


[1] Goodfellow, I. J., Shlens, J., Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.



How to verify noise





→ The networks predicts ”7” for all noisy images!

Topics

Multiple topics that are related to this idea, including

- Implementation of competition benchmark
- Extensions to graph neural networks
- Algorithms to tighten over-approximation
- ...

Interested? Contact me!

Tobias Ladner
tobias.ladner@tum.de