

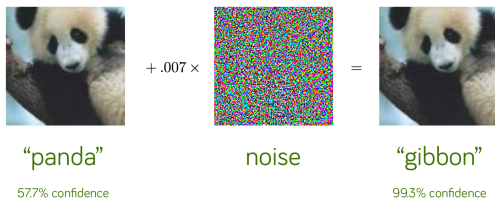
Formal Verification of Neural Networks

Tobias Ladner

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

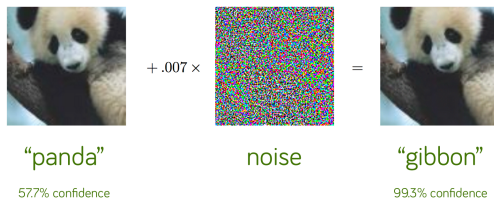
July 13th, 2023

Motivation



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

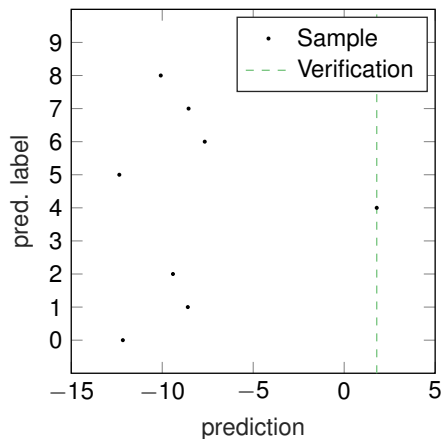
Motivation



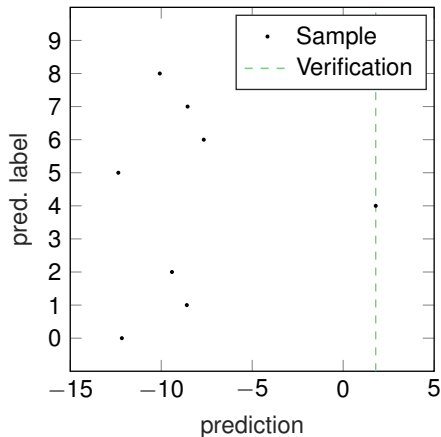
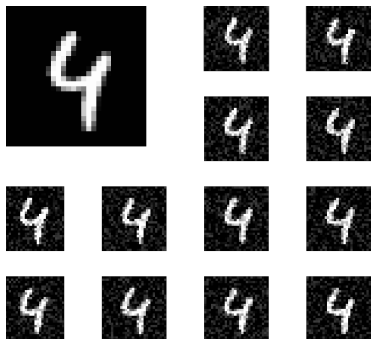
[1] Goodfellow, I. J., Shlens, J., Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.



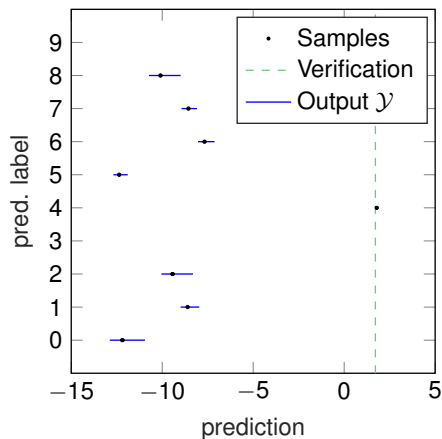
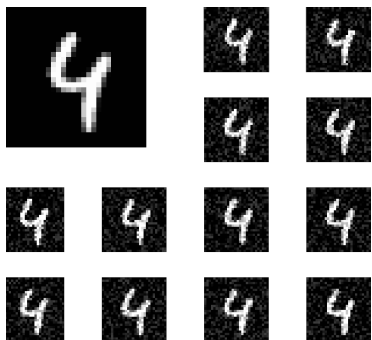
How to verify uncertainties



How to verify uncertainties



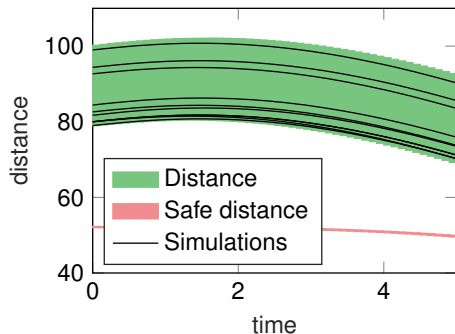
How to verify uncertainties



→ The network predicts "4" for all noisy images!

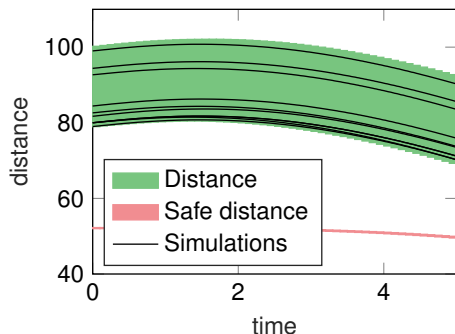
Topic: Training + Formal Verification of Reinforcement Learning Agents

Car Example

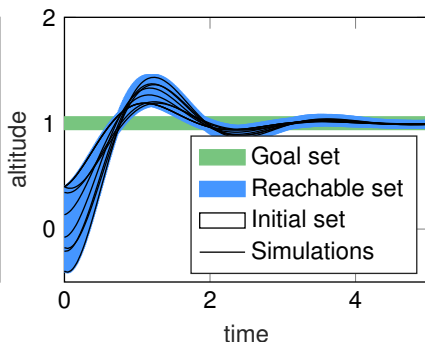


Topic: Training + Formal Verification of Reinforcement Learning Agents

Car Example



Drone Example



Interested? Contact me!

Tobias Ladner

tobias.ladner@tum.de