B.Sc./ M.Sc. Thesis

Untersuchung der statischen Stabilität neuronaler Netze

Neuronale Netze (NNs) spielen eine immer größere Rolle in unserer aller Leben. Auch in technischen Systemen kommen Sie zunehmend zum Einsatz. Dabei ist es von entscheidender Bedeutung für die Zertifizierbarkeit solcher auf künstlicher Intelligenz basierender Systeme, dass sie verifizierbar robust gegenüber Störungen sind. Kleinste, von Menschen nicht mal wahrnehmbare Änderungen der Eingangsgrößen, sollten nicht dazu führen, dass sich die Modellantwort ins Gegenteil verkehrt. Die statische Stabilität NNs beschreibt, wie sensitiv sie auf Eingangsstörungen reagieren, was u.a. durch die Lipschitz-Konstante (LK) des NNs charakterisiert werden kann. Deren exakte Bestimmung ist i.A. zwar nicht möglich, allerdings existieren unterschiedlich genaue Abschätzungen der tatsächlichen LK. Es wurde aber gezeigt, dass die Schärfe dieser Abschätzung zum Teil beträchtlich mit der Größe der untersuchten NNs variieren kann. Für die Zertifizierbarkeit ist eine verlässliche Abschätzung mit bekannter Unsicherheit allerdings unerlässlich. Im Rahmen dieser Arbeit sollen deshalb verschiedene Abschätzungen zunächst implementiert und anschließend auf ihre Schärfe hin untersucht werden.

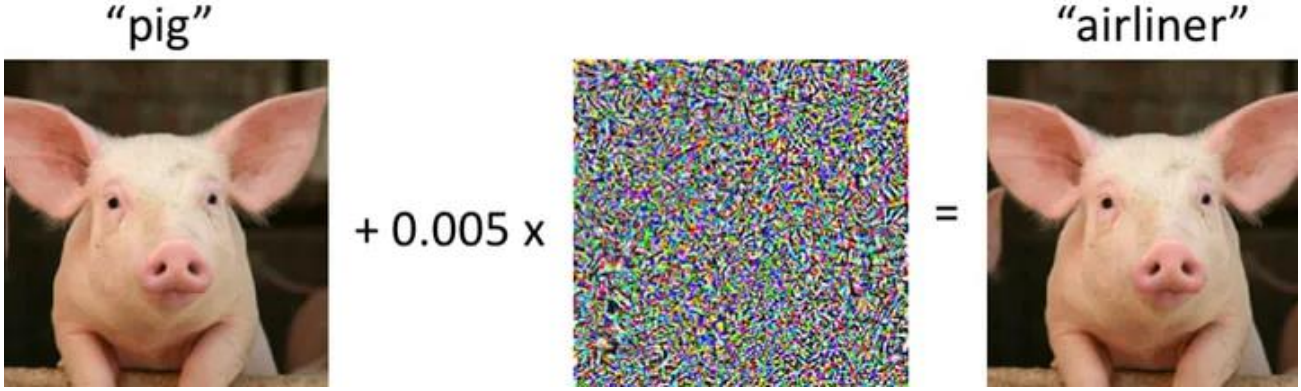

Abb. 1: Illustration der Sensitivität einer NN-Vorhersage ggü. Eingangsstörungen. Das NN soll erkennen, was für ein Objekt auf dem gegebenen Bild dargestellt ist. Für das Originalbild (links) wird korrekterweise ein Schwein erkannt. Durch Überlagerung eines für Menschen nicht wahrnehmbaren Rauschens (rechts) wird jedoch ein Flugzeug erkannt [übernommen von https://medium.com/swlh/gradient-based-adversarial-attacks-an-introduction-526238660dc9].

Aufgaben

- Literaturrecherche zur Bestimmung der LK von NNs
- Implementierung verschiedener Methoden zur Abschätzung der LK
- Training von unterschiedlich großen NNs mit bekannter LK
- Untersuchung der Schärfe der Approximationen in Abhängigkeit von der Tiefe der NNs

Folgende Vorkenntnisse sind wünschenswert:

- Programmiererfahrungen, vorzugsweise mit Python
- idealerweise erste Erfahrungen mit maschinellem Lernen, insbesondere neuronalen Netzen

Bearbeitungsbeginn und –dauer: ab sofort, 4-6 Monate (je nach PO)

Für weitere Informationen wenden Sie sich bitte an:

Hannes Mandler, M.Sc.
hannes.mandler@itlr.uni-stuttgart.de
+49 (0)711 685 62636

Adrian Grimm, M.Sc.
adrian.grimm@ifr.uni-stuttgart.de
+49 (0)711 685 69089