KI ABSICHERUNG
Safe AI for Automated Driving

Perspectives on Safety: Estimating and Proving

Tom Thielo, Christian Brunner,
Kai Fabi, Jonas Schneider

DNN specific Safety Concern: Unreliable Confidence / Brittleness

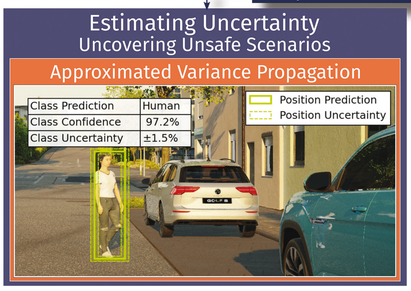
Epistemic Uncertainty Estimation

For safety critical applications knowing the limitations of an AI model is crucial. This knowledge on what a neural network does not know can be obtained by estimating uncertainties, which quantify what level of trust should be given to its predictions.
In the Bayesian modelling approach one can distinguish two types of uncertainty. Aleatoric uncertainty accounts for statistical noise inherent to the observations, e.g., sensor noise. Epistemic uncertainty captures the ignorance about which model generated the observed data. The latter is important as it is required to identify samples outside the domain of collected training data.

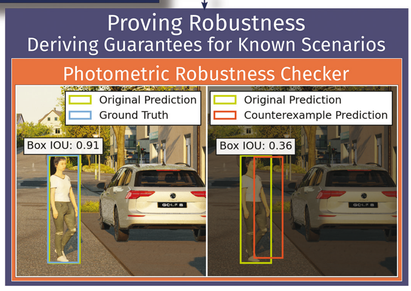
Proving Photometric Robustness

Formal verification of neural networks aims to give guarantees about their intended behavior in specified scenarios. For this purpose, the intended behavior has to be described by a set of testable properties. These properties can be used to describe robustness, i.e., how should a limited and meaningful change of the input data maximally influence the predictions of the neural network. Properties based on input data are referred to as local robustness properties. Approaches that utilize local robustness properties require a suitable test data set per property to enable a meaningful verification. That is, transformed inputs must still be part of the ODD.

Deep Neural Network Verification



The neural network is certain in its prediction...



...but not robust under transformations.

Approximated Variance Propagation

Popular techniques for epistemic uncertainty estimation, e.g., Monte Carlo Dropout require drawing samples at inference time. As the computational cost grows linearly with the number of samples, it becomes prohibitive for real-time applications like autonomous driving. Our mechanism implements a sampling-free epistemic uncertainty estimation using variance propagation. By injecting variances on the neuron activations and propagating them through the network, we derive an uncertainty estimate in a single shot. Thus, we achieve comparable quality while being real-time capable.

Safety Hypothesis:
The method addresses the safety concern Unreliable Confidence Information. Quantifying epistemic uncertainty helps to identify inputs the network is not familiar with or only has sparse information on.

Photometric Robustness Checker (PRC)

Our PRC encodes the neural network and the definition of counterexamples into a mixed integer linear program (MILP). The search space is based on an input sample and photometric transformations, e.g., contrast change. Solutions to this MILP are counterexamples for the robustness properties under test and consequently refute them. If no counterexample exists, we have proven that the network is robust against a transformation. While our approach yields proofs on robustness, it is currently computationally infeasible for large neural networks.

Safety Hypothesis:
The method addresses the safety concern Brittleness of DNNs. The intended behavior of neural networks is proven for specific properties, i.e. robustness against photometric transformations.

EFS

For more information contact:
jonas.schneider@efs-auto.com

KI Absicherung is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

KI FAMILIE

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

https://www.ki-absicherung-projekt.de @KI_Familie KI Familie

Image sources: EFS & Mackevision Medien Design GmbH