REPAIR

Robust and ExPlainable AI for Radarsensors

Programm / Ausschreibung	Mobilität der Zukunft, Mobilität der Zukunft, MdZ - Konjunkturpaket (2021) FT	**Status**	abgeschlossen
Projektstart	01.10.2021	**Projektende**	31.03.2025
Zeitraum	2021 - 2025	**Projektlaufzeit**	42 Monate
Keywords	Robuste Radarsignalverarbeitung, Explainable AI, robuste Neuronale Netze, Bayessche Neuronale Netze,		

Projektbeschreibung

Eine Grundvoraussetzung für die breite Akzeptanz von autonomen Fahrzeugen ist die robuste und zuverlässige automatische Fahrzeugführung in allen denkbaren Situationen ohne notwendigen menschlichen Eingriff. Die Grundlage hierfür bietet die Sensorik zur Umweltwahrnehmung, wobei Radarsensoren wegen Wetter- und Beleuchtungsunabhängigkeit ein unverzichtbarer Bestandteil sind. In realen Verkehrssituationen können mannigfaltige Stör- und Umwelteinflüsse wie die Temperatur, verschiedene Verkehrssituationen, als auch eine gezielte Manipulation von Eingangsdaten (d.h. adversarial attacks) auftreten. Das Sensorsystem muss in der Lage sein, mit diesen Einflüssen umzugehen sowie zu erkennen wann die eigenen Vorhersagen unsicher sind. Wird die Vorhersage eines Sensors in einer bestimmten Situation als ‘unsicher’ eingestuft, so kann darauf reagiert werden indem andere Sensoren mit höherer Konfidenz berücksichtigt werden oder eine gesicherte Notfallsroutine aktiviert wird.

Zentrales Projektziel ist die Entwicklung von erklärbaren und robusten Machine Learning (ML) Modellen zur Objektdetektion mittels Radarsensoren. Für sicherheitskritische Anwendungsgebiete ist es essentiell, genau zu verstehen, wie ein ML Modell seine Vorhersage bestimmt. Mittels Explainable AI kann sichergestellt werden, welche Einflussfaktoren dafür entscheidend sind. Um die Robustheit der ML Modelle zu erhöhen wird an drei Bereichen geforscht; (i) es werden Hybridmodelle entwickelt - diese vereinen die Robustheit von klassischer Modellierung mit der Performancesteigerung der ML Methoden; (ii) durch Bayessche Neuronale Netze wird Unsicherheits-Information zu den Objektdetektionen bestimmt – davon profitiert die Weiterverarbeitung und sicherheitsrelevante Aspekte können garantiert werden; (iii) die Robustheit eines Neuronalen Netzes hängt umittelbar mit dem „Distribution Shift“ zwischen Trainings- und Testdaten zusammen. Um ein robustes Verhalten auf Distribution Shifts zu gewährleisten, wird eine nicht-parametrische Methode unter Verwendung der Wasserstein Distanz entwickelt und mit verschiedenen Normalisierungsmethoden verglichen.

Mit den erforschten, robusten und nachvollziehbaren ML Modellen soll gezeigt werden, dass ML Modelle auch für die Anwendung in sicherheitskritischen Bereichen sinnvoll und nützlich sind. Das Risiko einer fehlerhaften Entscheidung wird dadurch quantifizierbar. Die Evaluierung der ML Modelle anhand verschiedener Anwendungsfälle ermöglicht einen kritischen Vergleich in einem breiten Problemspektrum im Kontext von autonomen Fahrzeugen und der Umweltwahrnehmung mittels

Radarsensoren.

Abstract

A basic requirement for the widespread acceptance of autonomous vehicles is a robust and reliable automatic vehicle guidance in all possible situations without the need for human intervention. The foundation for that is provided by sensors for the environmental perception, with radar sensors being an indispensable technology due to their independence of weather and lighting conditions. In real world traffic situations a manifold of disturbances and environmental influences can occur, such as temperature effects, different traffic situations as well as a targeted manipulation of input data (i.e. adversarial attacks). The sensor system must be able to address these influences robustly and to recognize when its own predictions are uncertain. If the prediction of a sensor is classified as "uncertain" in a specific situation, other sensors with higher confidence can be given more credence or a secured emergency routine can be triggered.

The main project goal is the development of explainable and robust machine learning (ML) models for object detection using radar sensors. For safety-critical applications it is essential to understand exactly how an ML model determines its prediction. Explainable AI can be used to ensure which influencing factors are decisive. In order to increase the robustness of the ML models, research is focused in three areas; (i) Hybrid models are developed - these combine the robustness of classical modeling techniques with the superior performance of ML methods; (ii) Bayesian neural networks provide uncertainty information about the object detections – this information can be exploited in higher level processing steps and safety-relevant aspects can be guaranteed; (iii) the robustness of a neural network is directly related to the “distribution shift” between training and test data. To ensure robust behavior on distribution shifts, a non-parametric method using the Wasserstein distance is developed and compared with different normalization methods.

The aim of the developed robust and “explainable” ML models is to demonstrate their usefulness and advantage in safety-critical applications. Furthermore, these methods make the risk of a wrong decision quantifiable. The evaluation of the ML models on the basis of different use cases enables a critical comparison in a wide range of problems in the context of autonomous vehicles and environmental perception using radar sensors.

Endberichtkurzfassung

In "Robust and Explainable AI for Radar sensors" (REPAIR) wurde untersucht, wie neuronale Netze (NNs) für die Verarbeitung von Radarsignalen verbessert werden können – insbesondere in Hinblick auf Interferenzen. Ein zentrales Problem bisheriger Modelle war, dass sie sehr empfindlich auf Veränderungen der Störparameter reagierten: Wenn sich die Bedingungen im Testfall von denen im Training unterschieden, brach die Vorhersagequalität ein. Um diese Herausforderung zu meistern, entwickelten wir verbesserte NN-Architekturen, die auch bei variierenden Störbedingungen zuverlässig funktionieren – z.B. bei unterschiedlichen Einfallswinkeln der Interferenzen. Ein weiterer wichtiger Fortschritt war die Integration des Objektdetektors direkt in den Trainingsprozess des neuronalen Netzes. Der Detektor ist essenziell, da die Aufgabe eines Radarsystems letztlich darin besteht, Objekte zu erkennen. Frühere Netzwerke ignorierten diesen Aspekt, was zu unzuverlässigen Vorhersagen führte. Außerdem wurde das neue Netz so entworfen, dass es relevante physikalische Größen wie Distanz, Geschwindigkeit und Einfallswinkel der Objekte unabhängig voneinander verarbeitet, was es robuster, physikalisch plausibler aber auch recheneffizienter macht. Neben den verbesserten NNs entwickelten wir auch ein Hybridmodell, welches auf der sogenannten Fraktionellen Fouriertransformation (FrFT) basiert. Mit der FrFT lassen sich FMCW Interferenzsignale optimal erkennen und entfernen. Der Vorteil dieses Verfahrens liegt darin, dass es, verglichen mit

Black-Box NNs, mit viel weniger Parametern auskommt, und sich viel leichter analysiert lässt, da jeder Parameter eine klare Funktion hat. Tests mit öffentlich verfügbaren Radardaten bestätigten, dass die Methode auch unter realen Bedingungen zuverlässig funktioniert – und sogar mit reellwertigen Radarsensoren (welche der Industriestandard sind) gut verwendet werden kann.

Projektkoordinator

- Technische Universität Graz

Projektpartner

- Infineon Technologies Austria AG