

Fortschrittszentrum LERNENDE SYSTEME

EIN KI-QUICK-CHECK DES KI-FORTSCHRITTSZENTRUMS



KONTAKT



Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Nina Schaaf

nina.schaaf@ipa.fraunhofer.de

Omar de Mitri

omar.de.mitri@ipa.fraunhofer.de

Hang Beom Kim

hang.beom.kim@ipa.fraunhofer.de

IN ZUSAMMENARBEIT MIT



IDS Imaging Development Systems GmbH

Alexander Balz

a.balz@ids-imaging.de

Dr. Robert-Alexander Windberger

r.windberger@ids-imaging.de

VISUALISIERUNG DER ENTSCHEIDUNGSFINDUNG NEURONALER NETZE

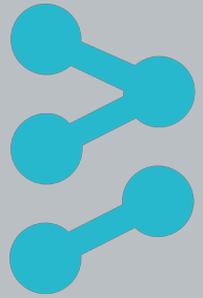
Ausgangssituation

IDS hat eine Trainingsplattform entwickelt, die es Fachexperten ermöglicht, durch Hochladen eigener Bilddaten Neuronale Netze zu trainieren. Trotz hoher Genauigkeiten der eingesetzten Neuronalen Netze gestaltet sich eine Diagnose im Fehlerfall oftmals schwierig und hemmt so die Verbreitung der Algorithmen. Eine von IDS entwickelte intuitive Visualisierung in Form sogenannter Attention Maps (Heat Maps) soll ermöglichen, kritische Entscheidungen leichter nachzuvollziehen, um damit die Akzeptanz Neuronaler Netze im industriellen Umfeld zu erhöhen. Um die Leistung des Heat Map-Verfahrens bewerten zu können, fehlt ein verlässliches Standardverfahren, welches die Ansprüche einer industriellen Umgebung repräsentiert.

Lösungsidee

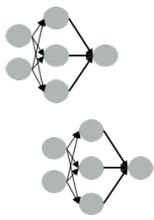
Ziel ist es, zu überprüfen, ob durch den Einsatz von Attention Maps Adversarial Attacks sowie verzerrte (engl. biased) Vorhersagen des Neuronalen Netzes erkannt werden können. Zu diesem Zweck werden verschiedene Methoden zur Erzeugung von Attention Maps anhand zuvor definierter Metriken miteinander verglichen. Um die Prüfung durchführen zu können, müssen zudem Techniken zur Erzeugung von Adversarial Attacks und zur Erstellung Neuronaler Netze mit künstlichen Fehlern (Bias) eingesetzt werden. Vielversprechende Techniken und Metriken können zu einem Testprozess zusammengeführt werden, welcher die Güte von Heat Map-Verfahren nachvollziehbar quantifiziert.

VISUALISIERUNG DER ENTSCHEIDUNGSFINDUNG NEURONALER NETZE MITTELS SMART-KAMERA

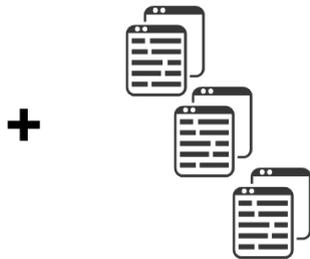


EIN KI-QUICK-CHECK DES KI-FORTSCHRITTSZENTRUMS

Neuronale Netze



Visualisierungstechniken



Attention Maps



© Monopoly919 - stock.adobe.com

Metriken



Neuronaler Netze im industriellen Umfeld führen wird.

Nutzen

Durch die von IDS bereitgestellte Trainingsplattform kann ein Fachexperte sein Wissen über die Bildinhalte in einen Algorithmus einfließen lassen, um diesen stetig anzupassen und zu verbessern. Eine intuitiv zu verstehende Visualisierung der Aufmerksamkeit des Neuronalen Netzes ermöglicht es Nutzern, schnell Fehler nachzuvollziehen und Schwachstellen im Datensatz zu finden. Die evaluierten Heatmap-Techniken visualisieren die Bildbereiche, welche für die Entscheidung durch das Neuronale Netz besonders relevant sind. Diese Visualisierungen ermöglichen es einem Laien, zu erkennen, wie die trainierten Netze funktionieren, was langfristig zu einer erhöhten Akzeptanz

Umsetzung der KI-Applikation

Die Umsetzung wurde mit dem öffentlich zugänglichen ImageNet-Datensatz durchgeführt. Um Bias und Adversarial Attacks zu erkennen, wurde ein verzerrtes Netz trainiert und mit Rauschmustern überlagerte Bilder generiert. Des Weiteren wurden Bewertungsmetriken zur Messung der Qualität der erzeugten Attention Maps entwickelt. Zur Evaluation der beiden Use Cases wurden drei Methoden zur Erzeugung von Attention Maps auf verschiedene Netzarchitekturen angewandt und die erzeugten Heat Maps mithilfe der Metriken bewertet. Die Ergebnisse zeigen, dass Attention Maps verzerrte Vorhersagen kenntlich machen und Hinweise auf Adversarial Attacks liefern können.

Die erarbeiteten Methoden helfen IDS dabei, eigens entwickelte Heat Map-Verfahren zu evaluieren und optimieren.

Fortschrittszentrum LERNENDE SYSTEME

EIN KI-QUICK-CHECK DES KI-FORTSCHRITTSZENTRUMS



Fraunhofer-Institut für Arbeitswirtschaft
und Organisation IAO



Fraunhofer-Institut für Produktions-
technik und Automatisierung IPA

Kooperationspartner:



Gefördert durch:



Baden-Württemberg

MINISTERIUM FÜR WIRTSCHAFT, ARBEIT UND WOHNUNGSBAU

Ansprechpartner:

Dr. Matthias Peissner

Telefon +49 711 970-2311

matthias.peissner@iao.fraunhofer.de

Prof. Dr. Marco Huber

Telefon +49 711 970-1960

marco.huber@ipa.fraunhofer.de

www.ki-fortschrittszentrum.de

ÜBER DAS KI-FORTSCHRITTSZENTRUM »LERNENDE SYSTEME«

Das KI-Fortschrittszentrum »Lernende Systeme« unterstützt Firmen dabei, die wirtschaftlichen Chancen der Künstlichen Intelligenz und insbesondere des Maschinellen Lernens für sich zu nutzen. In anwendungsnahen Forschungsprojekten und in direkter Kooperation mit Industrieunternehmen arbeiten die Stuttgarter Fraunhofer-Institute für Arbeitswirtschaft und Organisation IAO sowie für Produktionstechnik und Automatisierung IPA daran, Technologien aus der KI-Spitzenforschung in die breite Anwendung der produzierenden Industrie und der Dienstleistungswirtschaft zu bringen. Finanzielle Förderung erhält das Zentrum vom Ministerium für Wirtschaft, Arbeit und Wohnungsbau Baden-Württemberg.

Europas größte Forschungskooperation auf dem Gebiet der KI

Das KI-Forschungszentrum ist Forschungspartner des Cyber Valley, einem Konsortium

aus den renommierten Universitäten Tübingen und Stuttgart, dem Max-Planck-Institut für intelligente Systeme und einigen führenden Industrieunternehmen. In gemeinsamen Forschungslabors werden Grundlagenforschung und anwendungsorientierte Entwicklung zu aktuellen wie auch zukünftigen Bedarfen behandelt und vorangetrieben.

Menschzentrierte KI

Alle Aktivitäten des Zentrums verfolgen das Ziel, eine menschenzentrierte KI zu entwickeln, der die Menschen vertrauen und die sie akzeptieren. Nur wenn Menschen mit neuen Technologien intuitiv interagieren und vertrauensvoll zusammenarbeiten, kann ihr Potenzial optimal ausgeschöpft werden. Daher konzentrieren sich die Forschungsaktivitäten unter anderem auf die Themen Erklärbarkeit, Datenschutz, Sicherheit und Robustheit von KI-Technologien.