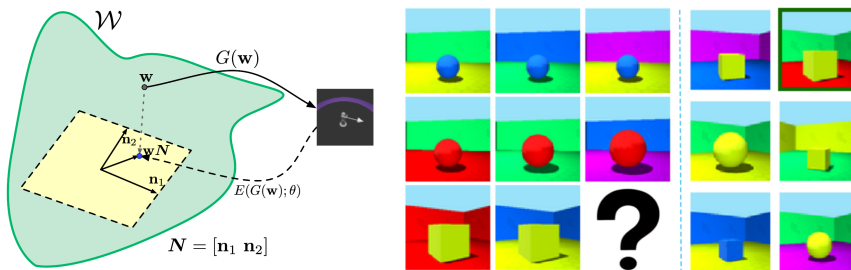


Feature Extraction aus trainierten Netzen durch nachträgliches Disentanglement

Deep Learning Explainable AI Feature Extraction Lineare Algebra

Tiefe Neuronale Netze spielen auch im autonomen Fahren eine immer größere Rolle und haben hier ganz neue Möglichkeiten im Bereich der Wahrnehmung und Planung eröffnet. Trotz aller Vorteile haben tiefe Netze jedoch gravierende Nachteile: Sie sind komplex, undurchsichtig, schwer zu verstehen und haben Millionen von Parametern. Dabei muss besonders bei sicherheitskritischen Anwendungen, wie dem autonomen Fahren, die Funktionsweise unbedingt nachvollziehbar und interpretierbar sein. Methoden und Algorithmen, welche für Menschen verständliche Erklärungen generieren können, werden unter dem Begriff Explainable AI zusammengefasst.



Nach dem Training eines tiefen Netzes ist es oft schwierig aussagekräftige Zwischenrepräsentationen zu finden, welche die Funktionsweise erklären. Ein neuer Ansatz [1] schafft hier Abhilfe, indem Zwischenergebnisse mit dem Verfahren der Singular Value Decomposition umgeformt werden. Deine Aufgabe ist es dieses Verfahren zu implementieren und insbesondere auf eine kürzere Laufzeit hin zu optimieren. Nachdem die Implementierung robust ist, wirst du das Verfahren mit verschiedenen Architekturen und Parametern testen. Auch sollst du untersuchen wie die extrahierten Features visualisiert werden können und ob diese als Grundlage für weitere Methoden benutzt werden können.

Du solltest Grundlagenwissen zu Deep Learning mitbringen und auch schon erste Programmiererfahrung mit Python gesammelt haben. Wenn das auf dich zutrifft und du Interesse hast, freue ich mich auf deine Bewerbung!

Die Arbeit besteht aus folgenden Teilen:

- + Literaturrecherche zu nachträglichem Disentanglement
- + Implementierung des Verfahrens
- + Laufzeit optimierung und Robustifizierung
- + Vergleich des Verfahrens

Gerne beantworte ich dir Fragen zur Thematik, Referenzliteratur oder sonstigen Themen. Frag mich einfach unverbindlich oder bewirb dich direkt!

[1] Khrulkov et al. (2021). "Disentangled Representations from Non-Disentangled Models"

Institut für Mess- und
Regelungstechnik (MRT)
Prof. Dr.-Ing. Christoph Stiller

Betreuer
Felix Hauser, M.Sc.

Programmiersprache¹
Python fortgeschritten

System, Frameworks
PyTorch, Git, Linux

Weitere Voraussetzungen:

- Grundlagen Deep Learning
- Selbständiges Arbeiten

Sprachen
Deutsch, Englisch

Melde dich bei Interesse oder Fragen einfach bei:

Felix Hauser
Raum: 033
Tel.: +49 721 608-42342
E-Mail: felix.hauser@kit.edu

Oder bewirb dich direkt mit einem aktuellen Notenauszug und unserem Fragebogen!



¹ **Sprachniveau**

grundlegend	< 500 Codezeilen (LOC)
fortgeschritten	500 – 5000 LOC
erfahren	> 5000 LOC