

Partnerspezifischer Schlussbericht

Berichtszeitraum: 01.04.2020 - 31.12.2023

FZI Forschungszentrum Informatik

Vorhabenbezeichnung	KI Data Tooling - Methoden und Werkzeuge für das Generieren und Veredeln von Trainings-, Validierungs- und Absicherungsdaten für KI-Funktionen autonomer Fahrzeuge
Laufzeit des Vorhabens	01.04.2020-31.12.2023
Förderkennzeichen	19A20001J
Fälligkeitsdatum	30.06.2024
Erstellungsdatum	18.12.2023
Version	1.0

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages



Dokumenteninformation

Autoren

Daniel Bogdoll, FZI Forschungszentrum Informatik

Jacob Langner, FZI Forschungszentrum Informatik

Rupert Polley, FZI Forschungszentrum Informatik

Philipp Rigoll, FZI Forschungszentrum Informatik

Kevin Rösch, FZI Forschungszentrum Informatik

Julian Truetsch, FZI Forschungszentrum Informatik

Kontakt

Jacob Langner

FZI Forschungszentrum Informatik

Haid-und-Neu-Straße 10-14

76131 Karlsruhe

Deutschland

Phone: +49 721 9654-154

Email: langner@fzi.de

Projektwebseite: <https://www.ki-datatooling.de/de/>



Inhaltsverzeichnis

1 Zusammenfassung / Summary	7
2 Übersicht	9
2.1 Aufgabenstellung	9
2.2 Voraussetzungen für das Vorhaben	9
2.2.1 BMWK-Fachprogramm “Neue Fahrzeug- und Systemtechnologien“	9
2.2.2 KI-Strategie der Bundesregierung	11
2.3 Planung und Ablauf	12
2.4 Stand der Wissenschaft und Technik	12
2.5 Zusammenarbeit mit anderen Stellen	13
3 Eingehende Darstellung	15
3.1 Erzielte Ergebnisse	15
3.1.1 AP1.1 Relevante Aspekte: Generierung und Qualitätsbeurteilung	15
3.1.2 AP1.2 Anpassung Toolchain und Bereitstellung synthetischer Radardaten	16
3.1.3 AP1.3 Validierung des Sensoreindrucks	17
3.1.4 AP2.1 Abstraktion von Sensorik	17
3.1.5 AP2.2 Optimierte Trainingsstrategie auf synthetischen, augmentierten und Real-Daten	20
3.1.6 AP2.3 Generierung und Selektion von Corner Cases	21
3.1.7 AP3.1 Methoden / Toolentwicklung für Autolabeling	32
3.1.8 AP3.2 Aufnahme von Realdaten	35
3.1.9 AP3.3 Aufbereitung der Trainingsdaten	36
Erarbeitung von Kriterien zur Bewertung der Datenqualität in der jeweiligen Verarbeitungsstufe	36
3.1.10 AP3.4 Augmentierung von Realdaten	40
3.1.11 AP3.5 Kontextidentifizierung	56
3.1.12 AP4.2 Schnittstellen und Beschreibungssprache	62
3.1.13 AP4.3 Methodenbaukasten und Toolchain	63
3.1.14 AP4.4 Evaluation und Demonstration	65
3.2 Zahlenmäßiger Nachweis	68
3.3 Notwendigkeit der Förderung	68
3.4 Veröffentlichungen	68
3.4.1 Öffentlichkeitsarbeit am FZI	68
3.4.2 Betreute Abschlussarbeiten	68
3.4.3 Erfolgte Veröffentlichungen	69



Abbildungsverzeichnis

Abbildung 1: Auszug diskutierter Radarsensoreffekte und -eigenschaften	16
Abbildung 2: Aufbau und Struktur des idealisierten, geometriebasierten Radarsensormodells	16
<i>Abbildung 3.3:</i> Veranschaulichung der Projektion des Kamerabildes in eine Punktwolke. Auf der rechten Seite mit einem Realbeispiel zur Veranschaulichung des Sichtfelds der Kamera	18
<i>Abbildung 3.4:</i> Darstellung der Methode in einem Datensatz mit 360° Abdeckung durch Kameras und LiDAR.	19
Abbildung 5: Flussdiagramm der Methode zur Beschreibung und Generierung von Corner Cases. Auf der Grundlage einer Corner Case Taxonomie und der Sprache OpenSCENARIO wurde eine Master-Ontologie entwickelt, die alle notwendigen Attribute zur Beschreibung komplexer Szenarien enthält. In einer 1:n Beziehung können Ontologien abgeleitet werden, die einzelne Szenarien beschreiben. Diese Szenarien werden dann automatisiert in das OpenSCENARIO-Format konvertiert, was die direkte Ausführung in Simulationsumgebungen ermöglicht.	22
Abbildung 6: Master Ontologie. Die Ontologie ist in der Lage, Szenarien zu beschreiben, die auf den sieben Abschnitten Szenario und Umgebung, Entitäten, Hauptelemente des Szenarios, Aktionen, Bedingungen, Wetter und Tageszeit sowie Corner Case-level basieren.	22
Abbildung 7: Szenario-Ontologie zur Beschreibung eines Fahrzeugs, das in ein nebliges Gebiet einfährt. Die Ontologie beschreibt das Szenario auf der Grundlage der sieben Abschnitte Szenario und Umgebung, Entitäten, Hauptelemente des Szenarios, Aktionen, Bedingungen, Wetter und Tageszeit sowie Corner Case Level.	23
Abbildung 8: Visualisierung der realisierten Corner Case-Szenarien	24
Abbildung 9: Erkennung von Corner Cases mit einem VAE	25
Abbildung 10: Überblick über die Pipeline zur Erstellung der Diskrepanzkarten. Das Originalbild wird durch ein semantisches Segmentierungsmodul geleitet, um eine semantische Maske zu erstellen. Daraus versucht ein GAN, das Originalbild zu rekonstruieren. Die Diskrepanzkarte wird durch den Vergleich des Originalbildes mit dem rekonstruierten Bild erstellt.	26
Abbildung 11: Datensätze für Training und Evaluation	26
Abbildung 12: Das linke Bild zeigt den latenten Raum einer VAE mit einer latenten Merkmalskarte der Größe $64 * 4 * 4$. Dieser latente Raum ist deutlich weniger geeignet, um Anomalien durch Clustering zu erkennen, als der Raum in der Mitte, der von einer VAE erzeugt wird, bei der der latente Raum die Größe $512 * 4 * 4$ hat. Das Bild auf der rechten Seite zeigt den latenten Raum in der Mitte nach der Clusterbildung.	27
Abbildung 13: Das linke Diagramm zeigt die Anzahl der Punkte pro Szenario. Die y-Skala ist logarithmiert. Das rechte Diagramm zeigt die Anzahl der gefundenen Cluster für die inkonsistenten Punkte der beiden Szenarien 3 und 4.	28
Abbildung 14: Beispiele für korrekt erkannte Anomalien aufgrund unerwarteter Verhaltensmuster. Anomalien sind dargestellt in rot und gelb.	29
Abbildung 15: Erforderliche Schritte für einen Instanz-weisen Vergleich von LiDAR- und Kameradaten	30
Abbildung 16: Zunächst wird eine quantitative Analyse mit Monte Carlo Dropout durchgeführt. Dann wird eine qualitative Analyse durchgeführt, bei der der geschätzten Punktwolken durchgeführt, um zu sehen, wie gut Corner Cases in 3D abgebildet wurden.	30
Abbildung 17: Vergleich von Pseudo-LiDAR mit Kameradaten. Es ist sichtbar, dass eine geometrische Punktwolke ähnlich zu LiDAR-Daten erkannt wird.	31
Abbildung 18: Vergleich von Pseudo-LiDAR mit Kameradaten. Es ist sichtbar, dass eine geometrische Punktwolke unähnlich zu LiDAR-Daten erkannt wird, was eine Anomalie darstellt.	32
Abbildung 19: Bewertung der Bewegung von Instanzen über die Zeit mit der Chamfer-Distanz.	32



Abbildung 20: Physischer Aufbau, zur Validierung geometrischer Eigenschaften des Radarsensormodells	36
Abbildung 21: Fahrzeug mit fehlerhaftem 3D-Bounding-Box-Label aus dem Datensatz (rot), präzidiertem Bounding Box (blau) und korrigiertem Bounding-Box-Label (grün)	37
Abbildung 22: Tracks verschiedener Objekte (Zeilen), die verifiziert werden sollen	39
Abbildung 23: Künstlich erzeugtes, realistisches Bildrauschen auf dem Cityscapes-Datensatz	40
Abbildung 24: Darstellung der Maskenqualität von Cityscapes	42
Abbildung 25: Resultierendes augmentiertes Bild, zusammengesetzt aus Vordergrundelementen und Hintergrundelementen	42
Abbildung 26: Ungenauigkeiten in der Tiefenkarte (rechts)	43
Abbildung 27: Verdeckung bei Augmentierung, von links nach rechts: a) einzufügende, ausgeschnittene Person b) mit Hilfe der Tiefenprädiktion bestimmte Maske c) nicht verdeckter Teil der Person wird eingefügt d) augmentiertes Bild mit Verdeckung	44
Abbildung 28: Sechs Batches, für die jeweils die in den Kontextbildausschnitt (jeweils oben links) einzufügenden Personen nach der gelernten Metrik sortiert dargestellt sind. Visuell passendere Personen haben einen kleineren Metrik-Abstand (weiter oben links).	45
Abbildung 29: Gelernte Anpassung der Histogramme der eingefügten Bildausschnitte	46
Abbildung 30: Das Netzwerk wird trainiert das Histogramm mit nur wenigen Parametern zu korrigieren, ähnlich zur Vorgehensweise bei der manuellen Bildbearbeitung	46
Abbildung 31: Durch Copy-Paste Augmentierung auf die Straße verschobene Objektpunktwolken	47
Abbildung 32: Neu generierte Beispiele eines mit dem Stanford Cars Datensatz trainierten DCGAN.	48
Abbildung 33: Neu generierte Beispiele eines mit dem A2D2 Datensatz trainierten DCGAN.	48
Abbildung 34: Fußgänger-Erzeugung mit Pix2Pix-Erweiterung	49
Abbildung 35: Einfügen eines realistisch aussehenden Verkehrsschilds in eine Verkehrsszene mit einem CycleGAN auf Grundlage eines Piktogramms	50
Abbildung 36: Vorgeschlagene Schildpositionen in einem Bild (rot); gefilterte Schildpositionen (grau)	50
Abbildung 37: Schematische Darstellung zur Bestimmung, ob ausreichen Platz für das neue Verkehrsschild und den zugehörigen Pfosten vorhanden ist	51
Abbildung 38: Prozess der Wetter-Augmentation	52
Abbildung 39: Beispiel der Wetter-Augmentation	53
Abbildung 40: Augmentationsbeispiel Tag (oben, real) → Nacht (unten, Augmentation)	54
Abbildung 41: Augmentationsbeispiel Nacht (oben, real) → Tag (unten, Augmentation)	55
Abbildung 42: Generierte Bilder (Augmentation: Nacht→ Tag) inkl. Rating des jeweiligen Bildes	55
Abbildung 43: Semantisch annotierte ins Kamerabild projizierte LiDAR-Punkte zur Abstandsmessung	57
Abbildung 44: Teilschritt der Methode bei der die relevanten Fahrzeuge (weiß) im Kamerabild identifiziert wurden	57
Abbildung 45: Tage des Jahres an denen aufgezeichnet wurde	58
Abbildung 46: Geographischer Überblick über die Datenaufzeichnungen	58
Abbildung 47: Relative Sonnenposition zum Fahrzeug	59
Abbildung 48: Schema zur Detektion von punktförmigen Objekten in geographischen Openstreetmap-Daten	60



Abbildung 49: Beispiele für sichtbare geographische Openstreetmap-Objekte im A2D2-Datensatz	61
Abbildung 50: Methode zur Suche nach Kontexten	62
Abbildung 51 Survey über Anomalie Datensätze	64
Abbildung 52 Schema des erzeugten Frustrums	64
Abbildung 53 Resultate der DBSCAN und Mean Shift Verfahren	65
Abbildung 54: Hardware des Demonstrators bestehend aus einem Messeaufbau mit einem integrierten berührungsempfindlichen Bildschirm	66
Abbildung 55: Demonstrator zur realistischen Positionierung von neuen Verkehrsschildern in existierenden Fahraufzeichnungen	66
Abbildung 56: Demonstrator zur Augmentation einzelner Fahrzeuge	67

Tabellenverzeichnis

Tabelle 1: Alle möglichen Szenarien, die beim Vergleich auftreten können. Die Spalte "Farbe" gibt an, welche Farbe zur Visualisierung der einzelnen Szenarien verwendet wurde.	28
--	----



1 Zusammenfassung / Summary

Zusammenfassung

Die Verwendung von Algorithmen, welche Methoden der künstlichen Intelligenz einsetzen, wird aktuell als einer der Schlüsselfaktoren gesehen, um automatisierte Fahrfunktionen effizient umzusetzen. Trainings-, Validierungs- und vor allem Testdaten sind für die Entwicklung und die Absicherung von solchen KI-Systemen im automobilen Umfeld essentiell. Da KI-Systeme umfassend trainiert und validiert werden müssen, kann das Vorliegen einer geeigneten Datenbasis erheblich dazu beitragen, Entwicklungszyklen zu verkürzen.

KI Data Tooling adressiert diese Notwendigkeit erstmals systematisch und umfassend. Hierbei funktioniert KI Data Tooling als ein Baustein der KI Projektfamilie innerhalb der VDA Leitinitiative autonomes und vernetztes Fahren (VDA-LI). Im Rahmen der VDA-LI wurden offene Fragestellungen bei der Umsetzung automatisierter Fahrfunktionen identifiziert und in Projekte und Projektfamilien gruppiert. Als Teil der KI Projektfamilie, welche Fragestellungen rund um den Einsatz von künstlicher Intelligenz für automatisierte Fahrfunktionen bearbeitet, wird KI Data Tooling im engen Austausch mit den anderen Projekten stehen, um Anforderungen, erhobene Daten und entwickelte Methoden auszutauschen.

KI Data Tooling entwickelt Methoden und Tools, die eine Beschreibung, Verarbeitung, Selektion, Generierung, Komprimierung und Bereitstellung oder Übertragung von Daten, wie beispielsweise Einzelbildern oder Sequenzen von unterschiedlichen Kamera-, Radar- oder Lidar-Sensoreindrücken von relevanten Szenarien, ermöglichen.

Ziel ist es, eine Datenbasis für das Training, die Validierung sowie für Test und Absicherung von KI-Systemen herzustellen. Dabei werden Lösungen erzeugt, die es ermöglichen, effizient und systematisch Daten unterschiedlicher Güte, unterschiedlichen Ursprungs (synthetisch erzeugt oder auf Basis von Realdaten), unterschiedlicher Sensorsetups oder -technologien zu nutzen und weiter zu verwerten.

Um die Methoden und Tools zu verifizieren, werden sie am Beispiel des Use Cases Fußgängererkennung verwendet und optimiert. Dieser Use Case ermöglicht die direkte Beurteilung der Methoden und Tools anhand konkreter Anforderungen an die entstehenden Daten und er ist zentral für den erfolgreichen Einsatz automatisierter Fahrfunktionen. Darüber hinaus profitiert das Training von Fußgängererkennungsalgorithmen von der Einbeziehung synthetischer Daten, da diese erlauben, auch gefährliche Grenzsituationen systematisch abzubilden. Schließlich ermöglicht der Use Case den engen Austausch mit dem Schwesterprojekt KI Absicherung der KI Projektfamilie, welches exemplarisch den gleichen Use Case betrachtet.

Für die Datengenerierung existieren unterschiedliche Ansätze, die in diesem Vorhaben erstmals umfassend und gemeinsam betrachtet und gegenübergestellt werden, um ihre Nutzung für das Training von KI-Methoden in der Praxis zu untersuchen: Verwendung aufbereiteter Realdaten, synthetisch generierte Daten sowie hybride Ansätze, welche Realdaten synthetisch augmentieren.

Durch die integrierte Betrachtung von Realdaten und synthetischen Daten sowie der Nutzung von Effizienzpotentialen bei deren Kombination wird in diesem Vorhaben erstmals eine „Daten-Factory“ als Komplettlösung für das Training und die Validierung von KI-basierten automatisierten Fahrfunktionen erzeugt.

Summary

The use of algorithms that employ artificial intelligence methods is currently seen as one of the key factors in the efficient implementation of automated driving functions. Training, validation and, above all, test data are essential for the development and validation of such AI systems in the automotive environment. As



AI systems have to be comprehensively trained and validated, the availability of a suitable database can make a significant contribution to shortening development cycles.

KI Data Tooling addresses this need systematically and comprehensively for the first time. KI Data Tooling functions as a component of the KI-Projektfamilie within the VDA-LI for Autonomous and Connected Driving. As part of the VDA-LI, open issues in the implementation of automated driving functions were identified and grouped into projects and project families. As part of the KI-Projektfamilie, which deals with issues relating to the use of artificial intelligence for automated driving functions, KI Data Tooling will be in close contact with the other projects in order to exchange requirements, collected data and developed methods.

KI Data Tooling develops methods and tools that enable the description, processing, selection, generation, compression and provision or transfer of data, such as individual images or sequences of different camera, radar or lidar sensor impressions of relevant scenarios.

The aim is to create a database for the training, validation, testing and safeguarding of AI systems. Solutions are created that make it possible to efficiently and systematically use and further utilize data of different quality, different origins (synthetically generated or based on real data), different sensor setups or technologies.

In order to verify the methods and tools, they are used and optimized using the example of the pedestrian detection use case. This use case enables the methods and tools to be assessed directly on the basis of specific requirements for the resulting data and is central to the successful use of automated driving functions. In addition, the training of pedestrian detection algorithms benefits from the inclusion of synthetic data, as this allows dangerous borderline situations to be systematically mapped. Finally, the use case enables a close exchange with the sister project AI Safeguarding of the AI project family, which uses the same use case as an example.

Different approaches exist for data generation, which are being examined and compared comprehensively and jointly for the first time in this project in order to investigate their use for training AI methods in practice: Use of processed real data, synthetically generated data as well as hybrid approaches that synthetically augment real data.

Through the integrated consideration of real data and synthetic data as well as the use of efficiency potentials in their combination, this project will for the first time create a "data factory" as a complete solution for the training and validation of AI-based automated driving functions.



2 Übersicht

2.1 Aufgabenstellung

Das Projekt KI Data Tooling entwickelt Methoden und Werkzeuge für eine effiziente Erweiterung der Datenbasis für das Training, die Validierung und die Absicherungen von KI-Funktionen für das automatisierte Fahren. Die neuen Methoden und Werkzeuge sollen insgesamt Daten besser beschreiben, verarbeiten, selektieren, generieren, veredeln, erweitern, komprimieren und bereitstellen oder übertragen können.

Dabei werden Lösungen erzeugt, die es ermöglichen, effizient und systematisch Daten unterschiedlicher Güte, unterschiedlichen Ursprungs (synthetisch erzeugt oder auf Basis von Realdaten), unterschiedlicher Sensorsetups oder -technologien zu nutzen und wiederzuverwerten.

Die Projektpartner des Verbundprojekts setzen damit gezielt auf eine Datenkomplettlösung im Bereich KI-basierter Funktionen für das automatisierte Fahren. Dabei sollten die Daten gesamtheitlich gedacht werden, von der Simulation über die Augmentierung bis hin zu aufgezeichneten Realdaten. Für jede Art der Daten wurden maßgeschneiderte Lösungen erdacht, die den Einsatz dieser Daten für die Entwicklung und Absicherung KI-basierter Funktionen im Fahrzeug ermöglichen bzw. verbessern. Dabei wurde auf Erkenntnissen des „Schwesterprojekts“ KI Absicherung aufgesetzt sowie ein intensiver Austausch mit allen weiteren Projekten der VDA LI KI-Projektfamilie gepflegt.

Um die Ergebnisse einer breiten Öffentlichkeit zugänglich zu machen, wurden diese in exemplarischen Workflows zusammengefasst. Die Workflows zeigen dabei die Anwendersicht der Methoden und Verfahren, die im Projekt entwickelt wurden. Darüber hinaus sollten zahlreiche Demonstratoren entwickelt werden, die der Projektergebnisse anschaulich visualisieren.

2.2 Voraussetzungen für das Vorhaben

Die Ziele und intendierten Wirkungen des Vorhabens KI Data Tooling stehen im Einklang mit dem BMWK-Fachprogramm „*Neue Fahrzeug- und Systemtechnologien*“, innerhalb dessen es gefördert werden soll. Darüber hinaus leistet KI Data Tooling einen Beitrag zu den in der von der Bundesregierung im November 2018 veröffentlichten „*Strategie Künstliche Intelligenz*“.

Die Bezüge des Vorhabens zu den im Fachprogramm bzw. der Strategie jeweils definierten Themen- bzw. Handlungsfeldern werden im Folgenden näher erläutert.

2.2.1 BMWK-Fachprogramm „Neue Fahrzeug- und Systemtechnologien“

Das BMWK-Fachprogramm formuliert als Hauptziel die „Stärkung der industriellen und wissenschaftlichen Innovationskraft Deutschlands im Bereich des automatisierten Fahrens“¹.

Das hier skizzierte Vorhaben leistet einen zentralen Beitrag zur Stärkung der Wettbewerbsfähigkeit des Wirtschafts- und Wissenschaftsstandorts Deutschland. Durch das Projekt werden die nationale Automobilindustrie und Forschungslandschaft befähigt, bestehende Kompetenzen² für eine Schlüsseltechnologie im Bereich des automatisierten und vernetzten Fahrens auf- und auszubauen. Die Aufrechterhaltung und der Ausbau der Innovationskraft der deutschen Automobilindustrie allgemein und insbesondere auch der Automobilzulieferer ist von erheblichem Bundesinteresse, da es langfristig zur Generierung von Wertschöpfung am Standort Deutschland beiträgt.

¹ Bundesministerium für Wirtschaft und Energie, „Fachprogramm Neue Fahrzeug- und Systemtechnologien,“ Mai 2015.

² Bundesministerium für Bildung und Forschung, „Lernende Systeme müssen der Gesellschaft dienen – nicht umgekehrt,“ 16 März 2018. [Online].



Das Vorhaben weist einen konkreten Bezug zur Programmsäule „*Automatisiertes Fahren*“ des BMWK-Fachprogramms³. In Summe adressiert KI Data Tooling drei der dort definierten sieben Themenfelder, wie in der folgenden Tabelle beschrieben.

Forschungsziel	Beitrag KI Data Tooling
Themenfeld „Innovative Sensorik und Aktoriksysteme“	
Weiterentwicklung Sensorik- und Aktoriksysteme und dazugehöriger Architektur	• Bereitstellung von Tools und Werkzeugen für die effiziente Erzeugung von Trainings-, Test- und Validierungsdaten für die Entwicklung und Absicherung KI-basierter Fahrfunktion • Übertragung synthetisch erzeugter Sensordaten auf die Realität
Themenfeld „Schnelle, sichere und zuverlässige Kooperation durch Kommunikation“	
Berücksichtigung der informations- und kommunikationstechnischen Herausforderungen wie z.B. Datenhandling	• Entwicklung bitrateneffizienter Ansätze zur Übertragung von Bilddaten mit Fokus auf relevante Bereiche • Entwicklung von Verfahren, welche die effiziente simultane Kompression mehrerer Sensordaten erlaubt • Besondere Berücksichtigung von seltenen und kritischen Situationen in realen und synthetischen Daten • Anforderungs- und Schnittstellendefinitionen sowie Demonstration von Werkzeugen zur Datenverarbeitung
Themenfeld „Neuartige Verfahren zur Datenfusion und -verarbeitung“	
Optimierung der Verfahren zur Sensordatenfusion und -auswertung	• Entwicklung von Verfahren, mit denen bereits während der Aufnahme von Sensordaten im Fahrzeug diese Daten intelligent und zielgerichtet prozessiert werden mit dem Ziel, • die Datenmengen an der Quelle zu reduzieren • nachgelagerte Verarbeitungsschritte effizienter zu gestalten und • Zeit und Kosten in der gesamten Kette von Datenaufnahme bis zur Verwendung einzusparen

³ Bundesministerium für Wirtschaft und Energie, „Fachprogramm Neue Fahrzeug- und Systemtechnologien,“ Mai 2015.



2.2.2 KI-Strategie der Bundesregierung

Die Bundesregierung hat zwölf prioritäre Handlungsfelder formuliert, „um die Zukunft der KI und deren Nutzung in Deutschland und Europa mitgestalten zu können“⁴. Als ein zentrales Vehikel für die Umsetzung dieser Handlungsfelder wird die Forschungsförderung benannt.

Das Vorhaben KI Data Tooling im engeren und die KI-Projektfamilie im weiteren Sinn adressieren unmittelbare zwei dieser Handlungsfelder:

- Forschung in Deutschland und Europa stärken, um Innovationstreiber zu sein
 - Erklärbarkeit und Transparenz von KI sind für uns der Schlüssel für das Vertrauen in die KI. Für eine Vielzahl von Anwendungen der KI stellt sich die Technologie jedoch noch als sogenannte „Blackbox“ dar: Es ist für Nutzer und Betroffene eines KI-Systems oft weder nachvollziehbar noch transparent, wie das System zu Entscheidungen oder Ergebnissen gekommen ist; dies gilt schon für algorithmenbasierte Systeme zur Entscheidungsvorbereitung und erst recht für autonom arbeitende und entscheidende Systeme. Entscheidungen müssen sich nachvollziehen lassen, damit KI-Systeme als „trusted AI“ akzeptiert werden können und rechtlichen Anforderungen genügen.

Die KI-getriebene Datenwissenschaft ermöglicht es, in großen Datenmengen („Big Data“) Muster, Untergruppen und Zusammenhänge zu erkennen. ... **Die Nutzung pseudonymisierter und anonymisierter Daten** kann ... dazu beitragen, ... das wirtschaftliche Potenzial von Big-Data-Anwendungen zu erreichen.

Die Bundesregierung wird die **Forschung an modernen Pseudonymisierungs- und Anonymisierungsverfahren** und für die **Erstellung synthetischer Trainingsdaten** vorantreiben und in die Praxis bringen.

Eine weitere Schlüsseltechnologie stellt das Autonome Fahren dar, bei dem KI dazu beitragen kann, unsere Mobilität sicherer, umweltfreundlicher, leistungsfähiger und preiswerter zu gestalten. Forschung und Entwicklung, insbesondere für fortschrittliche Elektroniklösungen, sind notwendig, damit ein Fahrzeug eigenständig und auch in komplexen Umgebungen Objekte im Straßenverkehr richtig erkennt und zuverlässige Fahrentscheidungen trifft. Dies erfordert die **Echtzeitverarbeitung sehr großer Datenmengen aus unterschiedlichen Quellen wie der Fahrzeugsensorik**, digitalen Karten und der mit dem Fahrzeug vernetzten Infrastruktur.

- Daten verfügbar machen und Nutzung erleichtern
 - Aufgrund der hohen Sicherheitsanforderungen im Automobilbereich kommt dort der Zuverlässigkeit von KI-Verfahren eine zentrale Rolle zu. Hierfür ist die Gewinnung ausreichender Trainingsdaten ... von besonderer Bedeutung. Zudem gilt es, auch mit Blick auf zukünftige Zulassungen neue Verfahren, Methoden und Gütekriterien zu entwickeln. Vor diesem Hintergrund ist eine **enge unternehmensübergreifende Zusammenarbeit der Automobilindustrie bei der Erzeugung/Aufnahme, effizienten Verwaltung und Auswertung von Fahr- und Sensordaten** gefragt. Nur durch eine solche Kooperation zur Realisierung und Absicherung von KI-Systemen für das automatisierte Fahren wird sich die internationale Wettbewerbsfähigkeit künftig sicherstellen lassen.

⁴ Bundesregierung, „Strategie Künstliche Intelligenz der Bundesregierung,“ November 2018. [Online].



2.3 Planung und Ablauf

Für die Ausplanung und Realisierung des Vorhabens wurde eine Projektstruktur mit einer Unterteilung in fünf Teilprojekte (TP) definiert. Die Hauptschwerpunkte und Zielsetzungen der einzelnen Teilprojekte lassen sich folgendermaßen zusammenfassen:

- In **TP1** werden die Anforderungen zur Generierung synthetischer Daten definiert, Methoden zu deren Generierung weiterentwickelt, synthetische Daten generiert und deren generierter synthetischer Sensoreindruck validiert.
- Das **TP2** betrachtet Effizienzaspekte synthetischer Daten wie Kompression, Corner Cases und Kontextualisierung und beantwortet konstruktiv die Frage, ob und wie die gemischte Datenbasis aus synthetischen und Realdaten effizient für das Training und die Validierung von KI-Funktionen zukünftig eingesetzt werden kann.
- Die Generierung und Aufbereitung von Realdaten ist in TP3 angesiedelt. Insbesondere werden im Rahmen des **TP3** Verfahren zum automatischen Labeling untersucht und es werden Ansätze zur Generierung von hybriden Daten untersucht.
- Im **TP4** werden Leitlinien erarbeitet und querschnittliche Aspekte betrachtet. Dies betrifft sowohl den Austausch mit anderen Projekten der KI-Projektfamilie, als auch die Untersuchung von Anforderungen des Use Cases sowie der Definition von Schnittstellen und Austauschformaten.
- **TP5** schafft die Voraussetzungen dafür, dass die Projektpartner in den definierten Arbeitspaketen effizient zusammenarbeiten können.

Das FZI Forschungszentrum Informatik beteiligte sich an allen Teilprojekten.

2.4 Stand der Wissenschaft und Technik

Trainings-, Validierungs- und Testdaten sind für die Absicherung von KI-Systemen essentiell. Dabei ist der Einsatz von realen, also mit Hilfe von Testfahrzeugen und entsprechenden Sensoren (Kameras, Lidar, Radar, usw.) unabdingbar. Die Erfassung und Aufbereitung von realen Daten sind jedoch sehr aufwendig und teuer. Eine zentrale Herausforderung ist, dass neben den eigentlichen Sensordaten des Fahrzeugs auch Referenzdaten, sogenannte Ground Truth oder Labels, erfasst werden müssen.

Eine Möglichkeit ist daher die Verwendung von Simulationen. Bei der Simulation der Kamera geht man in der Regel von einem Lochkamera-Modell aus und vernachlässigt Effekte wie Verzeichnung, Unschärfe, Vignettierung aber auch Bildsensor-Rauschen. Physikalische Kameramodelle existieren, sind aber stark vereinfacht und damit unrealistisch. Mittlerweile ist es jedoch unumstritten, dass die Sensoreffekte einen erheblichen Einfluss auf die Performance von Algorithmen haben und daher mit in der Toolkette berücksichtigt werden müssen. Zur Bewertung der optischen Daten wird zumeist nur der persönliche optische Eindruck oder auch die Übertragbarkeit durch Validierung auf anderen, realen Datensätzen herangezogen.

Ein essentieller Bestandteil in modernen Lernverfahren (z. B. neuronale Netze) ist die Bereitstellung entsprechender Trainingsdaten, die direkt in die Lernprozesse gespeist werden. In der Regel sind die Lernmethoden dabei überwacht, was zusätzlich zu den aufbereiteten Sensordaten semantische Label (z. B. Annotationen von Objekten auf Bildern) benötigt. Damit wird einem neuronalen Netzwerk ermöglicht, entsprechende Funktionalitäten zu erlernen. Der schwierigste Teil ist dadurch nicht notwendigerweise der Lernalgorithmus selbst, sondern die entsprechende Bereitstellung der Daten und ihrer Label. Dabei ist zum einen die Art und Weise der Bereitstellung wichtig, aber auch die Beschaffenheit der jeweiligen Label um in der gewünschten Zieldomäne erfolgreiche Anwendung zu finden.



Zur automatisierten Generierung und Veränderung von Bildern existieren unterschiedlichste Ansätze. Die Ansätze reichen von klassischen Bildverarbeitungsalgorithmen hin zu neuronalen Netzen. Dabei können komplett neue Situationen erstellt werden oder bereits existierende Aufnahmen nach einer bestimmten Vorgabe verändert werden. Der Vorteil bei der Verwendung dieser Methoden würde darin liegen, dass nicht nur die Labels direkt mit erzeugbar wären, sondern der reale Sensoreindruck explizit bzw. implizit mit beachtet wird. Ihre Anwendung im Kontext des automatisierten Fahrens verspricht eine robuste und verlässliche Quelle zur Datengenerierung und -Augmentierung.

Für die zukünftig anfallenden Mengen von Trainingsdaten ist es notwendig, geeignete Methoden zur Selektion relevanter Sequenzen, die gelabelt werden sollen oder auf die im Training besonderes Augenmerk gelegt werden soll, zu entwickeln. Dies wurde bereits im BMBF-geförderten Projekt „KI Plattform Konzept“ festgestellt und näher erläutert. Neben klassischen Selektionsverfahren, wie der manuellen Auswahl, gibt es bereits Vorarbeiten zu Methoden, die ungewöhnliche und damit wichtige Situationen, die häufig als Corner Cases bezeichnet werden, automatisiert detektieren. Außerdem ist die gezielte semantische Suche nach spezifischen Situationen notwendig.

Bei der Corner Case-Detektion werden typischerweise zwei Ansätze unterschieden. Zum einen gibt es die sogenannten semantischen Corner Cases, welche sich auf Daten beziehen, deren Inhalt etwas semantisch beschreibbar Ungewöhnliches darstellt. Als Beispiel kann hier eine mittels einer Kamera aufgenommene Szene genannt werden, bei der ein Fußgänger plötzlich auf die Straße rennt. Dieser ersten Art der Corner Cases stehen die datengetriebenen Corner Cases gegenüber. Hierbei bedingt nicht der Inhalt die Zugehörigkeit zur Corner-Case-Klasse, sondern die Verteilung der gesamten Messwerte. So können beispielsweise eine ungewöhnliche Pixelverteilung oder ein Sensorrauschen semantisch irrelevant sein, aber für die KI-Funktion große Schwierigkeiten bereiten, wenn es um korrekte Aussagen über das Gesehene geht.

Für den sinnvollen Umgang mit einem Datensatz werden in der Regelkontext-Daten verwendet. Diese enthalten in der Regel Informationen, wie den Aufnahmezeitpunkt und Aufnahmeort. Kontextdaten können dazu dienen, sich einen Überblick über die Zusammensetzung der aufgenommenen Daten zu verschaffen. Aber auch beim Auffinden spezifischer Daten spielen sie eine Rolle. Neben den Daten mit den dazugehörigen Kontext-Daten existieren in anderen Datenquellen, wie beispielsweise Wetterdatenbanken und bei Kartendiensten, weitere Informationen. Die einfache Verknüpfung dieser Kontexte sollte Teil einer effizienten Tool-Kette sein. Auch die automatisierte Erstellung von entsprechenden Übersichten und Statistiken würden den Umgang mit großen Datensätzen vereinfachen.

2.5 Zusammenarbeit mit anderen Stellen

KI Data Tooling hatte einen regen Austausch und nutzte Ergebnisse der Projekte KI-Absicherung, SetLevel4To5 sowie VVMethoden. Insbesondere in der ersten Projekthälfte, als in KI Data Tooling, noch keine Simulationsmodelle zur Verfügung standen, konnten exemplarische 3D-Assests der genannten Projekte genutzt werden. Des Weiteren wurden Erfahrungen zu simulationsbasierten Sensormodellen mit den Projekten diskutiert. Hierbei stellte VVMethoden, Erfahrungen mit Kamera- und Lidar-Sensorik und SetLevel4To5, mit Radar-Sensorik bereit. Für das Training und die Validierung von KI-Modellen konnte unter anderem auf die Erfahrung aus KI-Absicherung zurückgegriffen werden. Zudem gab es engen Austausch mit den parallellaufenden Projekten KI Delta Learning und KI-Wissen.

Für die Schwesterprojekte der KI-Projektfamilie der VDA-LI wurden mehrere Vernetzungstreffen im Rahmen von Zwischen- und Abschlussevents durchgeführt, an denen Vertreter*innen aller Projekte teilnahmen und so einen engen fachlichen Austausch sicherstellten.

Die Zusammenarbeit mit anderen Stellen wurde darüber hinaus im Allgemeinen meist von einzelnen Partnern vorangetrieben, die beispielsweise bereits in bestimmten Gremien vertreten waren. Das führt dazu, dass die Konzepte aus KI Data Tooling auch weiterhin in neue Projekte sowie in Standardisierungsvorhaben eingebracht werden und damit den langfristigen Impact des Projektes und seiner Ergebnisse stärken.





3 Eingehende Darstellung

Die Forschungsfortschritte im Vorhaben KI Data Tooling sind insbesondere basierend auf der inzwischen recht gut verstandenen Technologie der Convolutional Neural Networks, die insbesondere für die im Bereich des autonomen Fahrens anfallenden Sensordaten wie LiDAR- oder Kamerabilder. Parallel hierzu hat die sogenannte *Transformer*-Technologie, insbesondere aus der Anwendung für Sprach- bzw. Textmodelle heraus, viel Traktion aufgenommen und wurde im Zuge der Entwicklung auch als eine mögliche Alternative zu CNN-basierten Ansätzen zur Bilderkennung evaluiert. Die Transformer-Architektur hatte seit der ersten Veröffentlichung 2017 einen etwas gedämpften Start, mit einem Fokus insbesondere auf Sprachmodelle, bis Anfang 2021 mit DALL-E das erste große Diffusion-Modell zur Bilderzeugung veröffentlicht wurde, das auf der Transformer-Architektur aufsetzt und plötzlich das Tor zur Bildanwendung weit aufgerissen hat.

Wie der Einsatz dieser neuen Architekturparadigmen (Transformer) und den darauf aufbauenden Large Language Models (LLM) und den damit in Verbindung stehenden Foundation Models für das Anwendungsfeld des autonomen Fahrens bewerkstelligt werden kann, und wie die im Vorhaben KI Absicherung gewonnenen Erkenntnisse auf diese neuer Technologieschritte angewandt werden können, bleibt abzusehen. Diese Technologien bieten aber insbesondere aufgrund der deutlich geringeren Datenanforderung für die Spezialisierung (nicht für das ursprüngliche Training der Basismodelle, die extrem große Datenmengen erfordern) einen sehr interessanten weiteren Forschungsfokus für die Zukunft.

3.1 Erzielte Ergebnisse

3.1.1 AP1.1 Relevante Aspekte: Generierung und Qualitätsbeurteilung

Arbeitspaket AP1.1 erstellte Anforderungen an die spätere Generierung von Sensordate mittels Simulation, als auch für die realen Messungen, zur Validierung sowie Parametrisierung der Simulationsmodelle. Das FZI begleitete und gestaltetet die Diskussion aus Sicht der Radar-Sensorik. Hierbei wurde der für AP3.2 genutzte Radarsensor als Grundlage verwendet und technische Charakteristika wie, Trägerfrequenz, Field of View (FoV) oder Antennenkonfiguration eingebracht. Auf Basis dieser wurde mit den Partnern, hier insbesondere mit Robert Bosch und Ansys, grundlegende Effekte von Radarsensoren, wie Multipath Reflections, Mikro-Doppler, Materialdurchdringung oder Clutter-Effekte, diskutiert und relevante Szenarien für die simulationsbasierte Datengenerierung erstellt. Basierend auf diesen wurde in Szenarien-Katalog erstellt und die einzelnen Szenen bzgl. ihrer Relevanz priorisiert.

Modellierte Effekte:	Radarkanal
Rangebegrenzung (hart)	Interaktionen
FoV-Begrenzung (hart)	Reflexionen
Rauschen auf Messgrößen	Streuung
Winkel (Azimut und ggf. Elevation)	Beugung
Distanz	Brechung
Geschwindigkeit	Mikrodoppler (zur Klassifizierung von Fußgänger/Radfahrer)
Trennfähigkeit/Auflösung phänomenologisch	Dämpfung
False negatives (Aussetzer)	verdeckter Einbau
	Wetter (Regen, Nebel, ...)
	nahe Fahrbahn
	Einfluss weiterer Radarsensoren (eigene/fremde) - Interference
False positives (Geister) durch Mehrfachreflektionen, Straßenclutter, etc.	Sensor
	Hardwareeffekte
	Rauscheffekte/SNR
Verdeckung	Latenz
	DSV-Effekte
Detektion trotz Verdeckung	Modulation
	Auflösung
	Trennfähigkeit
RCS-Abschätzung	Verarbeitung der Locations
	Tracking
	Filtern
	Clustern
	Klassifizieren

Abbildung 1: Auszug diskutierter Radarsensoreffekte und -eigenschaften

3.1.2 AP1.2 Anpassung Toolchain und Bereitstellung synthetischer Radardaten

Das Arbeitspaket entwickelte ein idealisiertes Radarsensormodell, basierend auf Ansätzen geometrischer und physikalischer Optik.

In der ersten Projektphase wurde in Kooperation mit den Partnern ein einheitliche, d.h. unterschiedliche Sensormodalitäten umfassende, modularen Werkzeugkette skizziert. Augenmerk des FZI war die Berücksichtigung eines geometriebasierten Radarsensormodells (G-RSM), bzw. deren Verarbeitungskomponenten. Basierend auf dieser wurden potentielle Schnittstellen (Modelle können mehrere Module zusammenfassen, ohne die Schnittstellen offenzulegen) und Datenformate definiert.

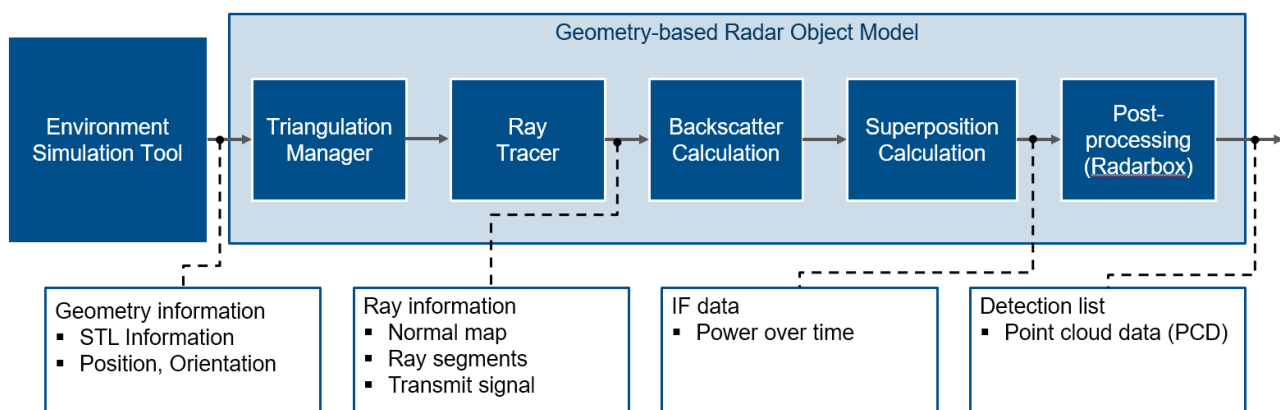


Abbildung 2: Aufbau und Struktur des idealisierten, geometriebasierten Radarsensormodells

Im Verlauf der Umsetzung kam es zu einer Neupriorisierung der Arbeiten des FZI. Arbeiten zur Berücksichtigung von Störeinflüssen (Boden- und Wetterclutter), soweit sie nicht durch geometriebasierte Pfadausbreitung modelliert werden, bspw. Rauigkeitseigenschaften, welche nicht als 3D-Mesh modelliert sind sowie Wettereinflüsse, durch nicht modellierte Partikel im Ausbreitungspfad, wurden nicht mehr berücksichtigt. Das entwickelte Radarmodell umfasst eine Triangulation Manager, welcher 3D-Szenen in mehrere Triangle-Meshes zerlegt. Diese bilden die Basis für ein rudimentäres Ray-Tracing (geometrische Optik), bei welchem alle rückstrahlenden Triangles detektiert werden. Hierbei werden Multipath-Reflektionen als auch Ausbreitungs- und Materialdämpfung berücksichtigt. Die Dämpfung bzw. Reflektion von



dielektrischen Materialien, die nur teilweise reflektieren, sind nicht berücksichtigt. Basierend auf den identifizierten Triangles, sowie die jeweilige Pfadanzahl und -länge findet eine Backscatter-Berechnung statt. Hierbei werden Berechnungen der physikalischen Optik (Kirchhoff-Helmholtz Integrals) verwendet, um das Rückstrahlverhalten der einzelnen Triangles zu bestimmen. Basierend auf den nachverfolgten einzelnen Strahlen pro Triangle, wird anschließend eine Superposition, basierend auf der räumlichen Ausdehnung berechnet. Hier lässt sich anschaulich darlegen wie die in AP1.1 diskutierten Parameter das Sensormodell beeinflussen. Das FoV sowie die Auflösung des Radarsensors, bestimmen die aggregierten Strahlen im Rahmen der Superposition. Die nachfolgende Verarbeitung zur Erzeugung von Detektionen, beruht auf der Signalverarbeitung des physischen Radarsensors.

3.1.3 AP1.3 Validierung des Sensoreindrucks

Für die Validierung der synthetischen Radardaten hat sich das FZI auf geometrische Eigenschaften beschränkt. Die Validierung basiert sowohl realer Messungen (vgl. AP3.2) sowie mathematischer Modelle für geometrische Objekte. Bei der Validierung stand die korrekte Modellierung der Radar-Cross-Section (RCS) sowie die korrekte Winkel- und Range-Modellierung im Fokus.

Für die realen Messungen wurde ein Radarsensor, basierend auf dem Entwicklungsboard TI AWR1843, in Betrieb genommen und mehrere Messungen in einer kontrollierten Umgebung (keine Laborumgebung), mit Retroreflektoren, aufgenommen (vgl. AP3.2). Hierbei stand die Variation unterschiedlicher geometrischer Szenen im Vordergrund. Target bildeten Eckreflektoren mit unterschiedlichen Kantenlängen. Zusätzlich wurde die RCS-Modellierung mit mathematischen Modellen wie einer planaren Platte validiert.

Des Weiteren hat das FZI die Materialvermessung bzgl. Radar und Lidar koordiniert und Anforderungen und Ergebnisse mit den Projektpartnern diskutiert und abgeglichen. Dafür wurden Anforderungen sowie eine Materialliste mit dem Konsortium abgestimmt, um eine bestmögliche Verwendung im Projekt zu gewährleisten. Anschließend wurde die pkTec GmbH für die Radarmessungen und das ILM der Universität Ulm für die Lidarmessungen beauftragt. Die Messergebnisse wurden mit den Partnern diskutiert und angepasste Anforderungen, in mehreren Zyklen, an die Unterauftragnehmer weitergegeben.

Wegen der in AP1.2 genannten Neupriorisierung der Themen begleitete das FZI die Abstimmungen zur Generierung der projektweiten synthetischen Radardaten, trat aber nicht als Anforderungsgeber bzw. Validator auf.

3.1.4 AP2.1 Abstraktion von Sensorik

Das Ziel dieses Arbeitspakets bestand darin, eine Methode zur Abstraktion von Sensordaten zu implementieren, insbesondere die Kombination von Kamerainformationen mit einer Punktwolke, um eine RGB-Punktwolke zu generieren. Dies dient dazu, das Verständnis und die Darstellung der von den Sensoren erfassten Umgebung zu verbessern.

Methodik: Die in diesem Projekt verwendete Methodik umfasst mehrere Schlüsselschritte. Als erstes wurde ein passender Datensatz ausgewählt. Eine Anforderung an den Datensatz ist eine Anwendung der entwickelten Methode auf ein 360° Sichtfeld. Außerdem muss die extrinsische Kalibrierung zwischen den Sensor gegeben sein. Die Abstraktion findet hier durch Projektion der Farbwerte des Bildes in die Punkte der LiDAR-Punktwolke statt. Die Informationsdichte sinkt insgesamt, da nicht alle Pixel des Kamerabildes zu einem Punkt in der LiDAR-Punktwolke korrespondieren. Aber die nachfolgende Datenverarbeitung kann mit einer angereicherten Punktwolke arbeiten. Die Grundprinzipien der Methode werden in *Abbildung 3.3* dargestellt.

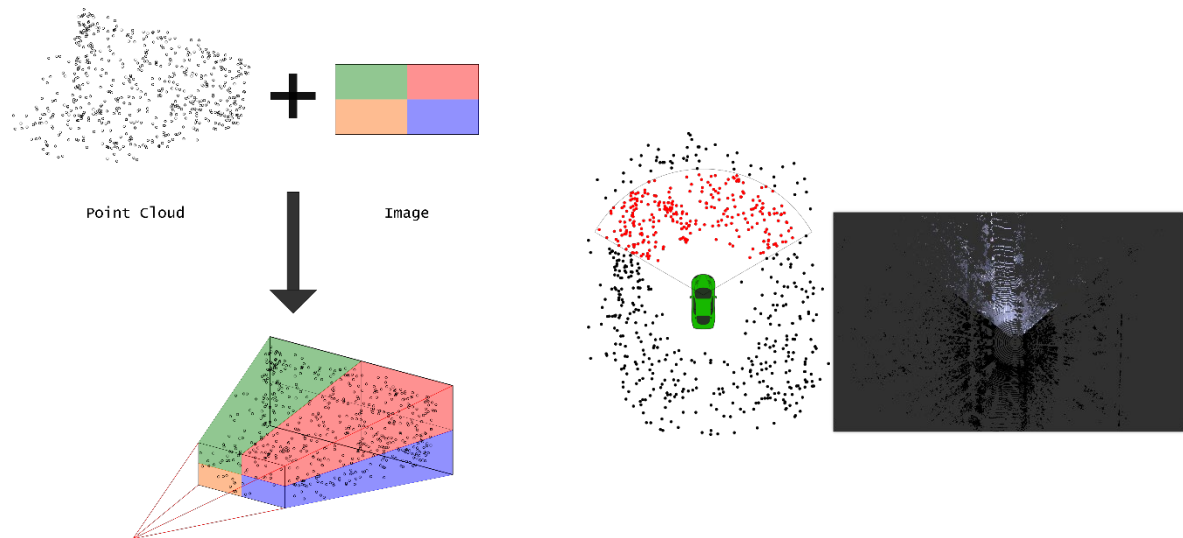


Abbildung 3.3: Veranschaulichung der Projektion des Kamerabildes in eine Punktwolke. Auf der rechten Seite mit einem Realbeispiel zur Veranschaulichung des Sichtfelds der Kamera

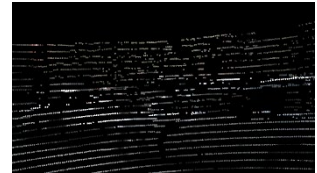
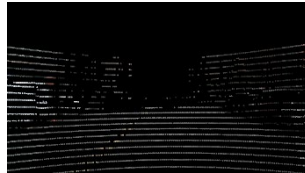
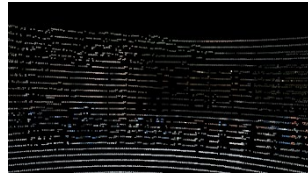
Die Methode wurde schließlich auf dem Lyft Level5 Perception Datensatz⁵ angewendet. Der Datensatz wurde ausgewählt da sich die Sichtfelder der Kameras und des LiDARs stark überlappen und die Kalibrierung der Sensoren vorliegt. *Abbildung 3.4* zeigt die Kamerabilder des Fahrzeugs als Referenz. Die Farbwerte der Kamera werden dann wie bereits beschrieben in die Punktwolke projiziert um eine Punktwolke mit Farbwerten in jeder Distanzmessung zu erhalten.

⁵ Houston, John, et al. "One thousand and one hours: Self-driving motion prediction dataset." Conference on Robot Learning. PMLR, 2021.

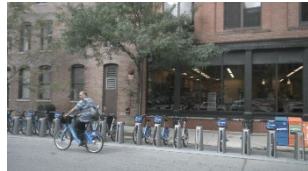
Kamera vorne



RGB Punktwolke
vorne



Kamera hinten



RGB Punktwolke
hinten

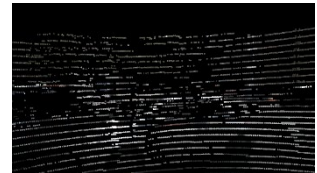
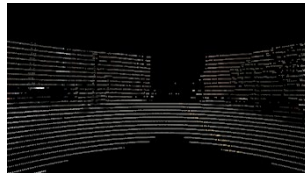
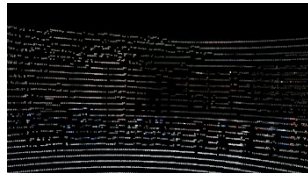


Abbildung 3.4: Darstellung der Methode in einem Datensatz mit 360° Abdeckung durch Kameras und LiDAR.



3.1.5 AP2.2 Optimierte Trainingsstrategie auf synthetischen, augmentierten und Real-Daten

Evaluation und Bewertungsmetriken für das Training mit synthetischen, augmentierten und realen Daten (E2.2.3)

Verfahren zur Bewertung der Repräsentativität von Datensätzen

Für E2.2.3 wurde am FZI Ansätze zur Bewertung der Repräsentativität von Datensätzen untersucht. Eine generische Bewertung der Repräsentativität bzgl. beliebiger Kriterien ist nur durch den Vergleich mit einem als repräsentativ angenommenen Referenzdatensatz möglich. Dieser Referenzdatensatz muss nicht gelabelt sein, sollte aber unter Vermeidung aller möglichen Bias und Verzerrungen aufgenommen werden. Im einfachsten Fall soll nur entschieden werden, ob der Referenzdatensatz und der Testdatensatz aus derselben statistischen Verteilung stammen, und damit ob der Testdatensatz den Referenzdatensatz repräsentativ abbildet. Es gibt eine ganze Reihe an statistischen Verfahren, die so genannten Two-Sample-Hypothesis-Tests, die geeignet sind zu entscheiden, ob zwei Stichproben (und somit auch Datensätze) aus derselben statistischen Verteilung gezogen worden sind. Da es sich bei den in autonomen Fahrzeugen eingesetzten Sensordaten jedoch um sehr hochdimensionale Daten handelt muss zunächst eine Dimensionsreduktion durchgeführt werden bevor die statistischen Tests angewandt werden können⁶. Es wurden hierzu Experimente mit den Dimensionsreduktionsverfahren Principal-Component-Analysis (PCA), Independent-Component-Analysis (ICA) und die Ausgabe eines ResNet-18⁷ verwendet. Als statistische Testverfahren wurde einerseits der Kolmogorov-Smirnov-Test (KS-Test) und andererseits ein Permutationstest mit der Maximum Mean Discrepancy (MMD)⁸ als Teststatistik untersucht. Der KS-Test mit Bonferroni-Korrektur kann dabei nur die Marginalverteilungen vergleichen, während der Permutationstest mit der MMD als Teststatistik ein echter mehrdimensionaler Test ist, der die gesamten Verteilungen vergleicht. Verglichen wurden in den Experimenten der Train-Split des Cityscapes-Datensatzes mit dem Val- und Test-Split desselben Datensatzes sowie mit dem Train-Split und dem Test-Split des KITTI-Datensatzes. Ausgabe aller Tests ist der p-Value für die Nullhypothese, dass die Verteilungen identisch sind. Für alle Verfahren war der p-Value für definitiv nicht-identische Datensätze (Cityscapes-Train verglichen mit KITTI), sehr viel kleiner als zu erwarten war. Jedoch konnten die Verfahren MMD-Permutationstest mit PCA als Dimensionsreduktion und MMD-Permutationstest mit ResNet als Dimensionsreduktion eine mit der Erwartung konsistente Anordnung der p-Values (kleiner für den Vergleich des Cityscapes-Train-Splits mit den anderen Cityscapes-Splits als im Vergleich mit den KITTI-Splits) produzieren.

Untersuchung von Trainingsansätzen zur Anreicherung unterrepräsentierter Datensätze durch künstlich augmentierte Realdaten innerhalb des Trainingsverlaufs

Das vom FZI in AP3.4 entwickelte Augmentierungsverfahren, mit denen Objekte oder Verkehrsteilnehmer aus einer Szene in eine andere Szene eingefügt werden können, wurde in E2.2.3 zu einem Online-Augmentierungsverfahren weiterentwickelt. Online-Augmentierungsverfahren bedeutet dabei, dass kein statischer augmentierter Datensatz generiert wird, sondern dass die Augmentierung stattdessen parallel zum Training eines neuronalen Netzwerkes durchgeführt wird. Der Vorteil dabei ist, dass einerseits kein zusätzlicher Speicherplatz für einen potentiell wesentlich größeren augmentierten Datensatz benötigt wird, und andererseits vor allem, dass durch die Online-Augmentierung eine höhere Variabilität in den augmentierten Bildern erreicht werden kann (die effektive Anzahl an unterschiedlich augmentierten Bildern

⁶ S. Rabanser, S. Günnemann, und Z. Lipton, „Failing Loudly: An Empirical Study of Methods for Detecting Dataset Shift“, in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2019.

⁷ Z. C. Lipton, Y.-X. Wang, und A. Smola, „Detecting and Correcting for Label Shift with Black Box Predictors“, Feb. 2018, doi: 10.48550/arXiv.1802.03916.

⁸ A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, und A. Smola, „A Kernel Two-Sample Test“, Journal of Machine Learning Research, Bd. 13, Nr. 25, S. 723-773, 2012.



über den Trainingsverlauf ist größer). Dies führt im Allgemeinen auch zu besseren Trainingsergebnissen. In E2.2.3 wurden dabei auch eine Ablation-Study durchgeführt um die Relevanz verschiedener Bestandteile des Augmentierungsverfahrens wie kontextuelle Platzierung, Skalierung und Blending zu untersuchen. Dabei konnte das Verfahren eine verbesserte Detektion von Fußgängern demonstrieren. Eine kontextuelle Platzierung war dabei für die Verbesserung der Average Precision (AP) für Detektion und Instanzsegmentierung mittels eines Mask-RCNN-Netzwerkes nicht wesentlich. Wichtig für eine Verbesserung der AP war jedoch eine Variation der Skalierung der eingefügten Personen (zufällig oder abhängig von der Einfügeposition). Dies weist darauf hin, dass für die Verbesserung der Detektionsergebnisse die Erhöhung der Varianz der Daten in gewissen Fällen wichtiger sein kann als die Realitätsnähe der generierten augmentierten Daten.

3.1.6 AP2.3 Generierung und Selektion von Corner Cases

Zu Beginn wurde in AP2.3.1 zusammen mit Partnern an der Definition von Corner Cases in unterschiedlichen Modalitäten gearbeitet⁹. Danach wurde der Fokus auf die Analyse und Entwicklung von Methoden zur Identifikation von Corner Cases im Bereich der Trajektorienplanung gelegt. In dem Paper "Space, Time, and Interaction: A Taxonomy of Corner Cases in Trajectory Datasets for Automated Driving"¹⁰ wurde zusammen mit anderen Partnern eine mehrdimensionale Einordnung und Beschreibung von Corner Cases in Trajektoriendaten erarbeitet. Diese soll als Grundlage für die Entwicklung von Methoden und Algorithmen in der Anwendung dienen. Eine Kernerkenntnis der Arbeit war, dass Corner Cases in Trajektoriendaten sehr abhängig von äußeren Einflüssen/Kontexten sein können. Basierend auf diesem Ergebnis wurde eine ausführliche Literaturrecherche im Bereich der Modellierung und Generierung von Trajektoriendaten durchgeführt.

Aus der Literatur abgeleitet wurde an einem Verfahren basierend auf Normalizing Flows gearbeitet, welches sich eignet um, gegeben einem Datensatz, die Verteilung der Daten zu lernen. Dies kann genutzt werden um per Sampling zu überprüfen ob neue Daten innerhalb der gelernten Verteilung liegen oder auch um neue Daten zu generieren. Die Eigenschaften von Architekturen die auf Normalizing Flows aufbauen, benötigen eine sehr gut definierte Eingangsdaten. Da die Länge und Menge von Trajektorien pro Szene jedoch stark variieren können, wurde als vorgeschaltetes Netzwerk ein Autoencoder entwickelt. mit dessen Hilfe Trajektorien beliebiger Länge in ein geeignetes Format gebracht werden sollen. Hierbei wurden jeweils unterschiedliche Autoencoderarchitekturen getestet und implementiert. Um die komplexe Rekonstruktionsaufgabe für das Pretraining der Autoencoder besser zu definieren, wurden unterschiedliche zusätzliche Hilfsaufgaben definiert die auf dem Prinzip des physikalisch informierten Trainings basieren. In der Abschlussphase hat sich herausgestellt, dass vor allem durch den Erfolg von Diffusionsnetzwerken neue Möglichkeiten für generative Modelle aufkamen. Die Methode wurde so flexibel konzipiert, dass der Austausch des generativen Elements am Ende der Verarbeitungskette möglich war.

Die Arbeiten zu AP2.3.2 wurden bereits in den Zwischenberichten Nr. 4 und 5 detailliert beschrieben. Hier folgt eine aktualisierte und zusammenfassende Version dieser Darstellungen:

1/6: Ontologie für verschieden granulare Corner Cases

In der Veröffentlichung Bogdoll *et al.*, „One Ontology to Rule Them All: Corner Case Scenarios for Autonomous Driving“, ECCV-W, 2022 beschreiben wir die bisher einzige Ontologie, die Corner-Case

⁹ Florian Heidecker, Jasmin Breitenstein, Kevin Rösch, Jonas Löhdefink, Maarten Bieshaar, Christoph Stiller, Tim Fingscheidt, Bernhard Sick, "An Application-Driven Conceptualization of Corner Cases for Perception in Highly Automated Driving", in Proc. of IV, 2021, Nagoya, Japan, p. 644-651

¹⁰ K. Rösch, F. Heidecker, J. Truetsch, K. Kowol, C. Schick Tanz, M. Bieshaar, B. Sick, and C. Stiller, "Space, Time, and Interaction: A Taxonomy of Corner Cases in Trajectory Datasets for Automated Driving", in Proc. of SSCI, IEEE CIVTS, 2022, Singapore, p. 1-8

Beschreibungen innerhalb der Ontologie aufnimmt und Szenarien aller Corner Case Level abbildet. Die Beschreibung der Szenarien basiert auf der vorherigen Publikation Bogdoll et al.; “Description of Corner Cases in Automated Driving: Goals and Challenges”; ICCV Workshop „Embedded and Real-World Computer Vision in Autonomous Driving“.

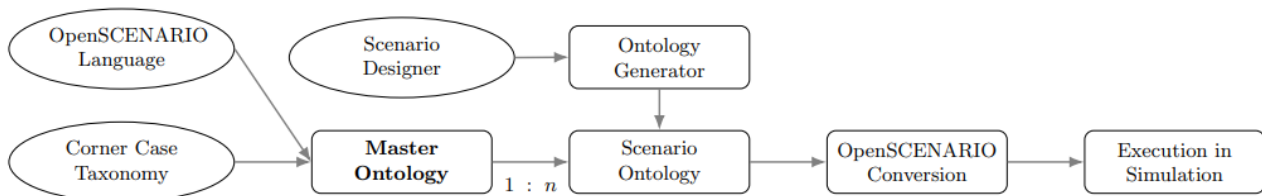


Abbildung 5: Flussdiagramm der Methode zur Beschreibung und Generierung von Corner Cases. Auf der Grundlage einer Corner Case Taxonomie und der Sprache OpenSCENARIO wurde eine Master-Ontologie entwickelt, die alle notwendigen Attribute zur Beschreibung komplexer Szenarien enthält. In einer 1:n Beziehung können Ontologien abgeleitet werden, die einzelne Szenarien beschreiben. Diese Szenarien werden dann automatisiert in das OpenSCENARIO-Format konvertiert, was die direkte Ausführung in Simulationsumgebungen ermöglicht.

Aufbauend auf dem Konzept, welches in Abbildung 5 gezeigt wird, können auf Basis der Master-Ontologie (Abbildung 6) spezifische Szenario-Ontologien erstellt und simuliert werden. Mehr Details zur Ontologie finden sich in der Veröffentlichung.

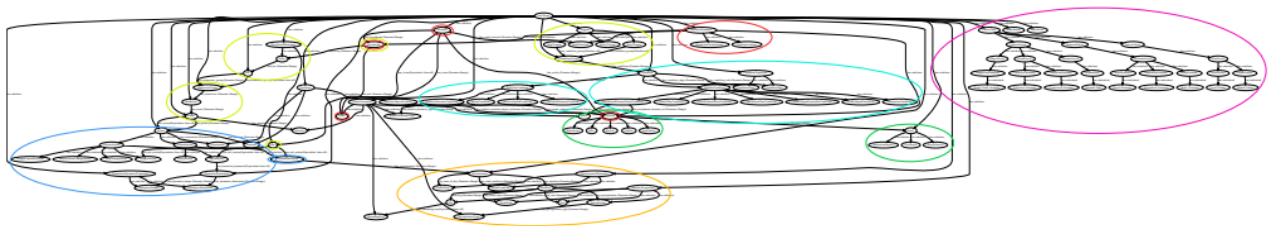


Abbildung 6: Master Ontologie. Die Ontologie ist in der Lage, Szenarien zu beschreiben, die auf den sieben Abschnitten Szenario und Umgebung, Entitäten, Hauptelemente des Szenarios, Aktionen, Bedingungen, Wetter und Tageszeit sowie Corner Case-level basieren.

2/6: Szenarienkatalog aufbauend auf der erstellten Ontologie

Auf Basis der Master-Ontologie wurde ein Szenarienkatalog spezifischer Szenario-Ontologien erstellt. Auf Basis des aufsteigenden Schwierigkeitsgrads der Corner Case Szenarien nach Breitenstein (Breitenstein, 2020) wurden somit zehn Szenarien verschiedener Kritikalität erstellt. Die Skalierbarkeit des Frameworks ermöglicht die Erstellung weiterer Szenarien. Eine Szenario-Ontologie wird in Abbildung 7 gezeigt.

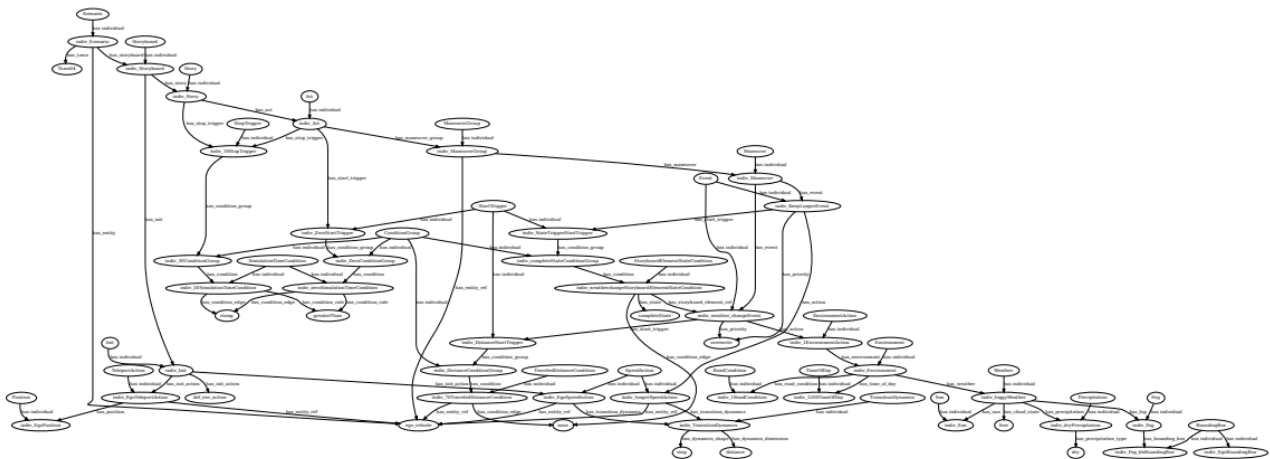


Abbildung 7: Szenario-Ontologie zur Beschreibung eines Fahrzeugs, das in ein nebliges Gebiet einfährt. Die Ontologie beschreibt das Szenario auf der Grundlage der sieben Abschnitte Szenario und Umgebung, Entitäten, Hauptelemente des Szenarios, Aktionen, Bedingungen, Wetter und Tageszeit sowie Corner Case Level.

Aufbauend auf solchen Szenario-Ontologien können vollständig temporale Szenarien simuliert werden, wie in Abbildung 8 dargestellt.

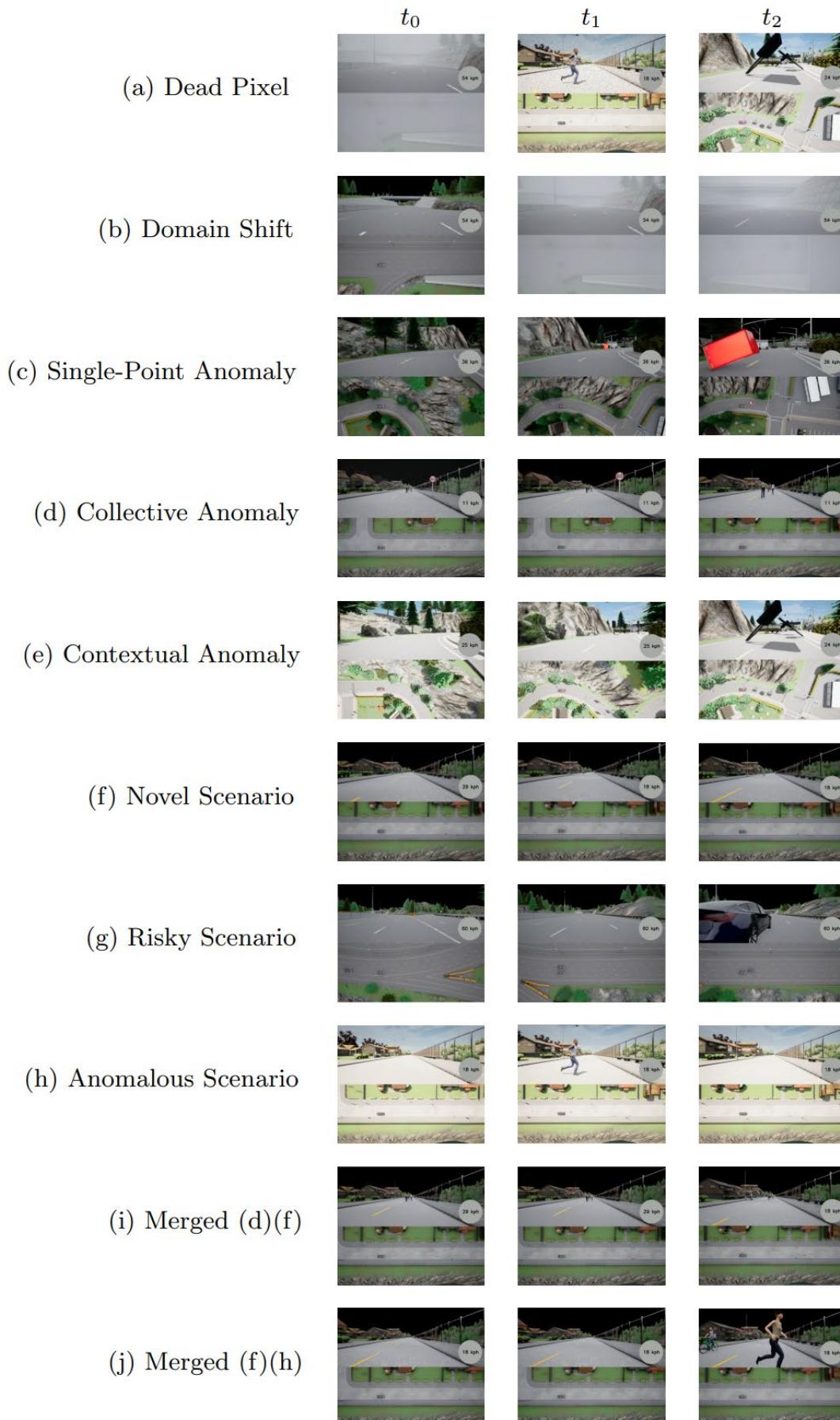


Abbildung 8: Visualisierung der realisierten Corner Case-Szenarien

3/6: Darstellung von Verkehrssituationen in einer Repräsentation im latenten Raum

Der gewählte Ansatz optimiert gleichzeitig auf die Repräsentation im latenten Raum sowie die Erkennung von Corner Cases. Eine schematische Darstellung ist in Abbildung 9 zu sehen:

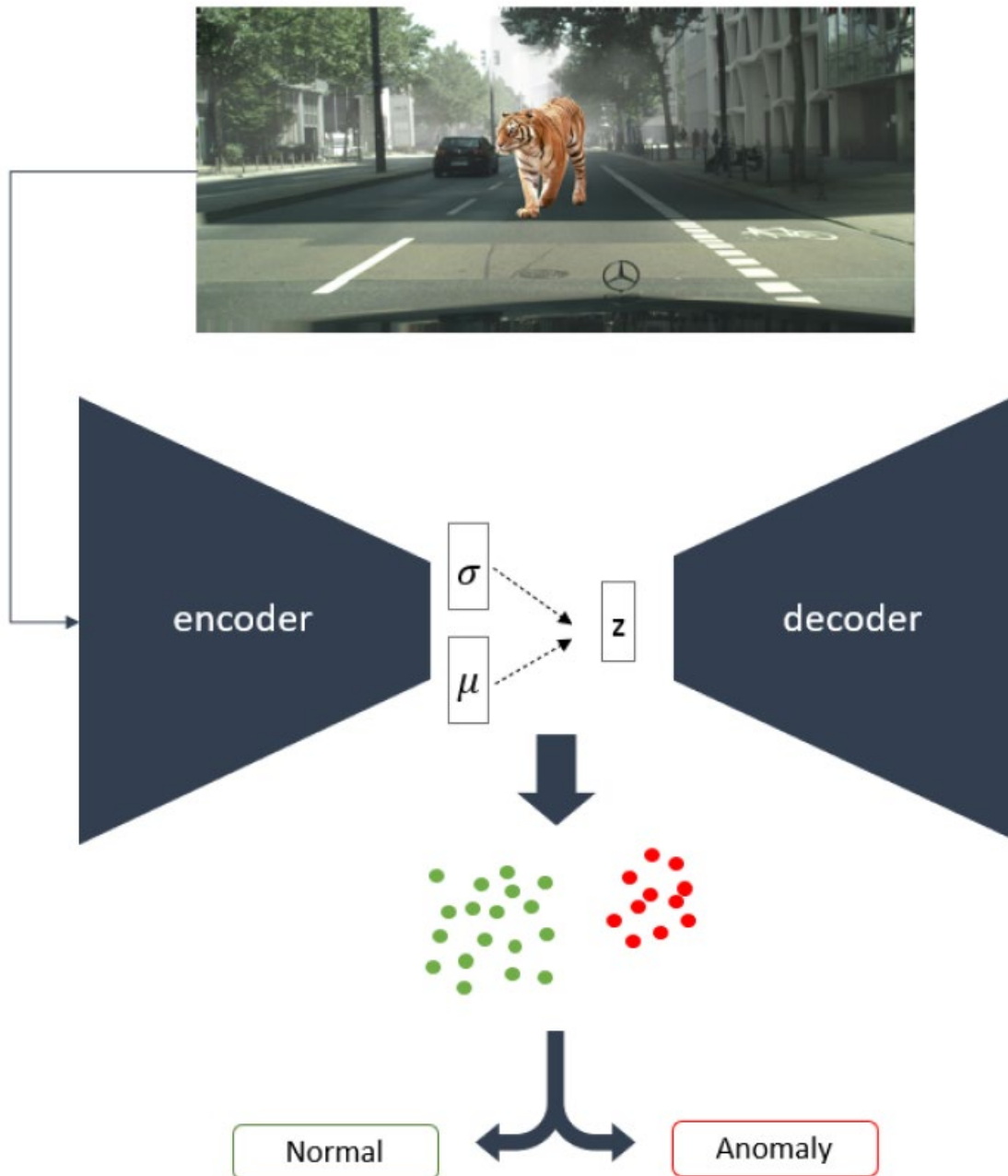


Abbildung 9: Erkennung von Corner Cases mit einem VAE

Um dem Netzwerk genauere Informationen über Anomalien zukommen zu lassen, wurde eine sogenannte Diskrepanzkarte generiert und dem VAE als weiterer Input zur Verfügung gestellt, wie in Abbildung 10 zu sehen ist.

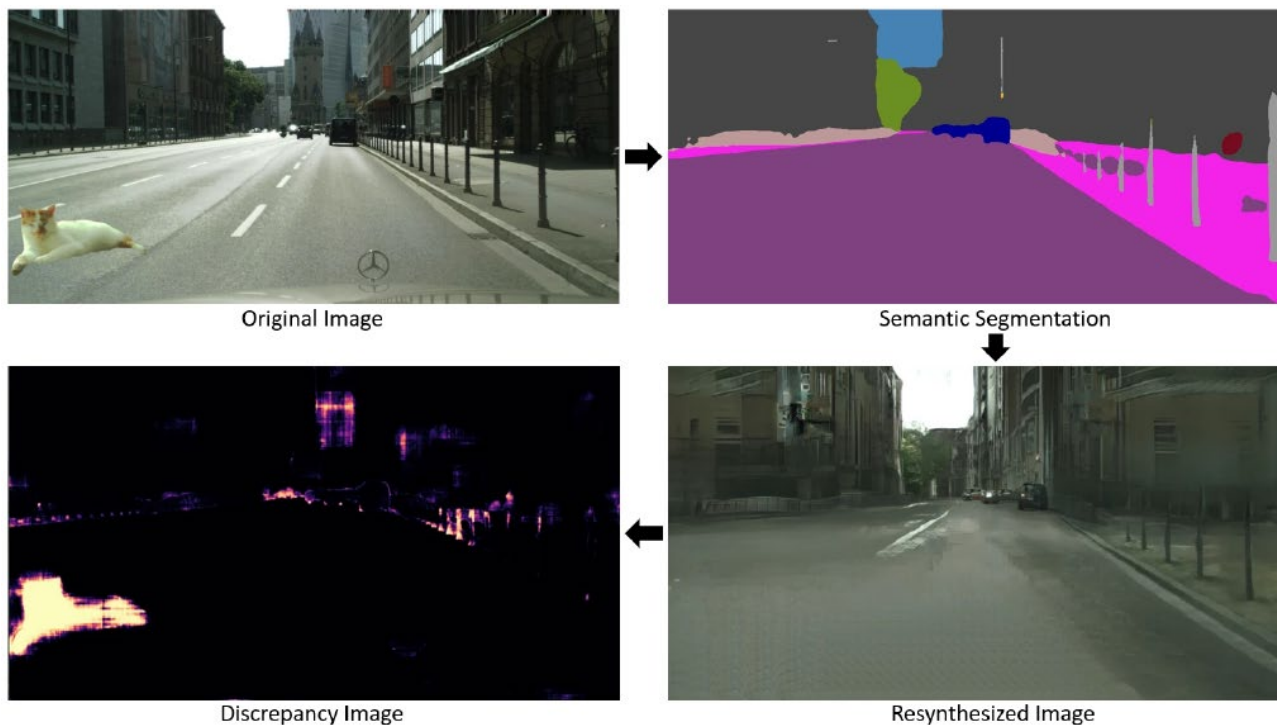


Abbildung 10: Überblick über die Pipeline zur Erstellung der Diskrepanzkarten. Das Originalbild wird durch ein semantisches Segmentierungsmodul geleitet, um eine semantische Maske zu erstellen. Daraus versucht ein GAN, das Originalbild zu rekonstruieren. Die Diskrepanzkarte wird durch den Vergleich des Originalbildes mit dem rekonstruierten Bild erstellt.

Auf Basis von verschiedenen Datensätzen (Abbildung 11) haben wir dieses Konzept ausführlich trainiert und evaluiert. Domain Shifts wurden durch ähnliche Datensätze vermieden, wo möglich.



Abbildung 11: Datensätze für Training und Evaluation

Es zeigt sich, dass insb. durch kleine Anomalien, die Unterscheidungen zwischen Normalität und Corner Case herausfordernd ist. Dies ist insbesondere in den Clustern des VAE latenten Raum sichtbar, wie Abbildung 12 zeigt.

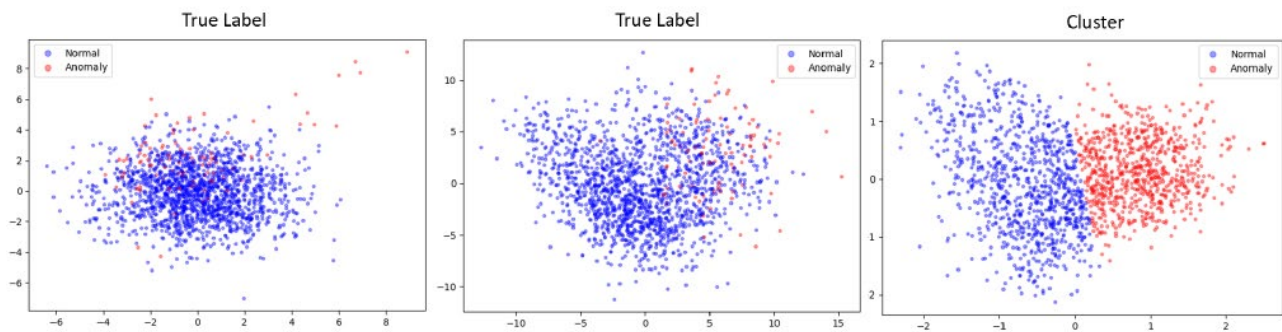


Abbildung 12: Das linke Bild zeigt den latenten Raum einer VAE mit einer latenten Merkmalskarte der Größe $64 * 4 * 4$. Dieser latente Raum ist deutlich weniger geeignet, um Anomalien durch Clustering zu erkennen, als der Raum in der Mitte, der von einer VAE erzeugt wird, bei der der latente Raum die Größe $512 * 4 * 4$ hat. Das Bild auf der rechten Seite zeigt den latenten Raum in der Mitte nach der Clusterbildung.

Im Vergleich zu SotA Anomalie-Detektoren erhalten wir vergleichbare False-Positiv Raten, stellen jedoch zusätzlich eine vollständige kompakte Repräsentation im latenten Raum bereit, welche als kompakte Repräsentation deutlich effizienter in der Handhabung ist und sich für weitere Tasks, z.B. prädiktive Ansätze, eignet. Weitere Details finden sich in der Veröffentlichung D. Bogdoll *et al.*, „Conditioning Latent-Space Clusters for Real-World Anomaly Classification“, IEEE Symposium Series on Computational Intelligence (SSCI), 2023

4/6: Untersuchung von Ausgangsformaten bzgl. der Eignung für automatisiertes Clustern

In Bezug auf die Eignung verschiedener Ausgangsformate und deren Eignung für spezielle Methoden zur Erkennung und Clustering von Corner Cases wurde eine umfangreiche Literaturstudie durchgeführt, in welcher über 110 Quellen untersucht und abschließend 50 in einer Survey verarbeitet worden sind. Das Gesamtergebnis, als Ausschnitt aus dem veröffentlichten Paper, zeigt sich in Figure 1.

	Camera	Lidar	Radar	Multimodal	Object level
Confidence score	9	3	0	0	2
Reconstructive	6	1	1	0	1
Generative	6	0	0	0	0
Feature extraction	4	0	9	5	1
Prediction	0	0	0	0	4

Figure 1: Overview of anomaly detection approaches based on camera, lidar, radar, multimodal, and abstract object level data.

Insbesondere Methoden aus den Kategorien „Reconstructive“ und „Generative“ verwenden einen latenten Raum im Prozess der Anomaliedetektion. Die Arbeiten wurden im April 2022 angeschlossen und beim CVPR Workshop on Autonomous Driving veröffentlicht (Bogdoll, 2022). Zusätzlich wurde eine kritische Analyse verfügbarer Daten im Bereich der Corner Case Detektion durchgeführt, welche insbesondere den Fokus auf Kameradaten bestätigt hat. Details finden sich in D. Bogdoll *et al.*, „Perception Datasets for Anomaly Detection in Autonomous Driving: A Survey“, IEEE Intelligent Vehicles Symposium (IV), 2023.

5/6: Untersuchung der Eignung erkannter Grenzfälle für Autolabeling

Zur Analyse wurde untersucht, inwiefern erkannte Anomalien, sogenannte Proposals, zum Autolabeling geeignet sind. Hierzu wurden erkannte Anomalien im KITTI-360 Datensatz untersucht. Es wurden statische



und dynamische Anomalien untersucht. Die Ergebnisse von zwei Bewegungsprädiktoren zur Anomalieerkennung (Supervised (SV) und Self-Supervised (SSV)) wurden verglichen und wie folgt kategorisiert und farblich codiert, wie Tabelle 1 zeigt:

Tabelle 1: Alle möglichen Szenarien, die beim Vergleich auftreten können. Die Spalte "Farbe" gibt an, welche Farbe zur Visualisierung der einzelnen Szenarien verwendet wurde.

	SV Part	SSV Part	Consistent	Color
Scenario 1	static	static	✓	green
Scenario 2	dynamic	dynamic	✓	blue
Scenario 3	static	dynamic	✗	red
Scenario 4	dynamic	static	✗	yellow

Unter Betrachtung von über 20.000 Frames hat sich folgendes Bild ergeben:

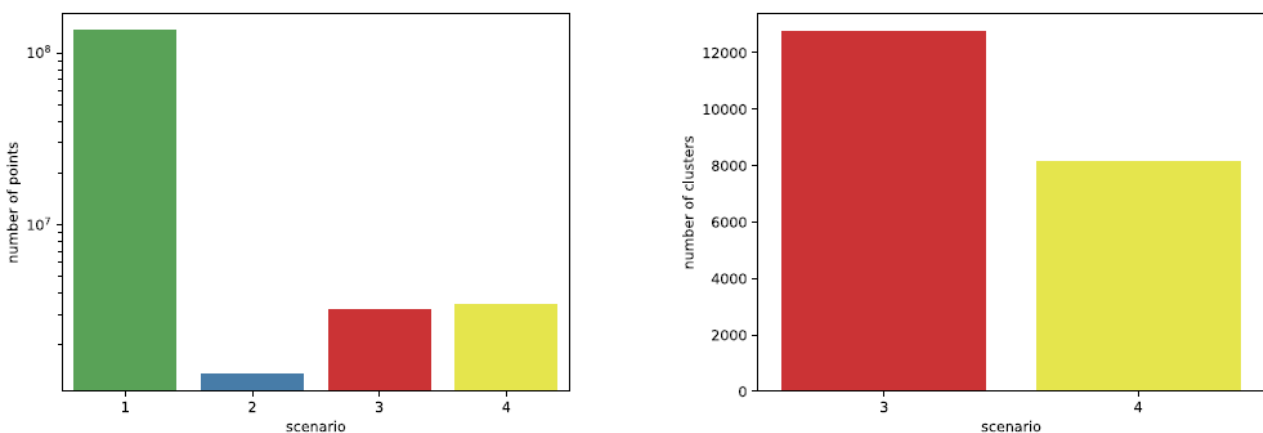


Abbildung 13: Das linke Diagramm zeigt die Anzahl der Punkte pro Szenario. Die y-Skala ist logarithmiert. Das rechte Diagramm zeigt die Anzahl der gefundenen Cluster für die inkonsistenten Punkte der beiden Szenarien 3 und 4.

Abbildung 13 zeigt die Anzahl der Punkte pro Szenario und die Anzahl der Cluster für die beiden inkonsistenten Szenarien. Aus dem Diagramm auf der linken Seite ist zu erkennen, dass die Anzahl der Punkte, die von beiden Teilen als statisch eingestuft wurden, am größten ist. Dies ist nicht überraschend, da im Allgemeinen die meisten Punkte in KITTI statisch sind. Die Anzahl der Punkte für Szenario 3 und 4, bei denen die Vorhersagen der beiden Anomalie-Detektoren nicht übereinstimmen, ist deutlich höher als die Anzahl der Punkte, bei denen beide Teile einen Punkt als dynamisch einstufen. Während der Analyse wurde Vielzahl von Fällen beobachtet, die sich in drei Hauptkategorien einteilen lassen:

- Korrekt konsistente Anomalie-Erkennung (TP)
- Korrekt konsistente Normalitäts-Erkennung (TN)
- Fehlerhafte inkonsistente Anomalie-Erkennung (FP)

Korrekt inkonsistente Anomalie-Erkennung (FN)

Hierbei hat sich erneut gezeigt, dass zahlreiche False-Positive Erkennungen, wie typisch in der Anomalie-Detektion. Diese können als Proposals für das Autolabeling verwendet werden, müssen jedoch durch menschliche Orakel bestätigt werden.

Beispiele zur Anomalie Erkennung finden sich in Abbildung 14:



Abbildung 14: Beispiele für korrekt erkannte Anomalien aufgrund unerwarteter Verhaltensmuster. Anomalien sind dargestellt in rot und gelb.

6/6: Untersuchung des Einflusses unterschiedlicher Klassen von Corner Cases innerhalb der aufgebauten Ontologie auf die Eignung von sensorübergreifendem Autolabeling

Verschiedene Typen von Corner Cases können Auswirkungen auf Single-, aber auch Multisensorkonzepte haben. Für diesen Teilbereich haben wir eine Architektur untersucht, die auf Kamera- und LiDARdaten basiert. Dieser wird in Abbildung 15 dargestellt. Um Punktwolken instanziell zu vergleichen, muss bestimmt werden, welche Punkte zu jeder Instanz in den LiDAR- und Kamera-Daten gehören. Um herauszufinden, wie stark ein Pseudo-LiDAR aus Kameradaten vom LiDAR für eine Instanz abweicht, wird ermittelt, welche Punkte in LiDAR und Pseudo-LiDAR zu einer Instanz gehören. Da bekannt ist, wie viele Punkte im Pseudo-LiDAR zu einer Instanz gehören, können die Punkte betrachtet werden, die zu der Instanz im LiDAR gehören. Sobald die entsprechenden Punkte im Pseudo-LiDAR und LiDAR für jede Instanz gefunden wurden, wird die Chamfer-Distanz als Abstandsmaß verwendet, um die Leistung des Pseudo-LiDARs Instanz für Instanz zu bewerten. Ein großer Abstand bedeutet, dass das Pseudo-LiDAR im Vergleich zu den LiDAR-Daten schlechter abschneidet, was eine Anomalie bedeuten kann.

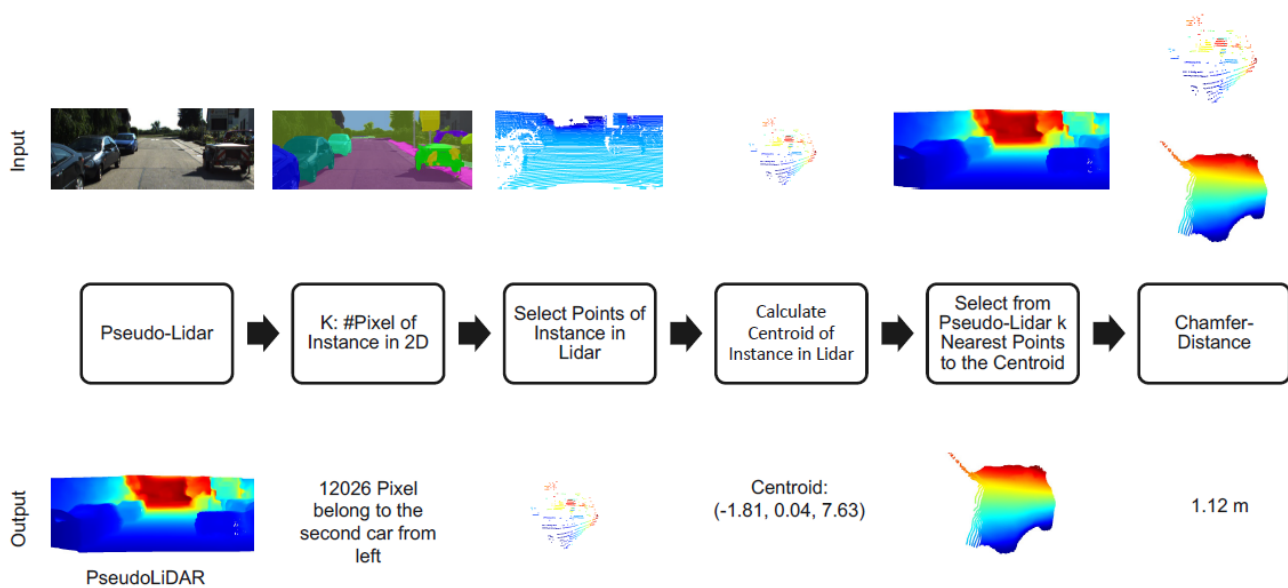


Abbildung 15: Erforderliche Schritte für einen Instanz-weisen Vergleich von LiDAR- und Kameradaten

Auf Basis derselben Daten wie im vorherigen Schritt wurden unterschiedliche Corner Cases und deren Einfluss auf die Methodik analysiert. Als Basis für den Prozess werden systematische Unsicherheiten in der semantischen Segmentierung genutzt. Auf Basis zahlreicher Datensätze wurde eine quantitative und qualitative Evaluation durchgeführt, wie Abbildung 16 zeigt.

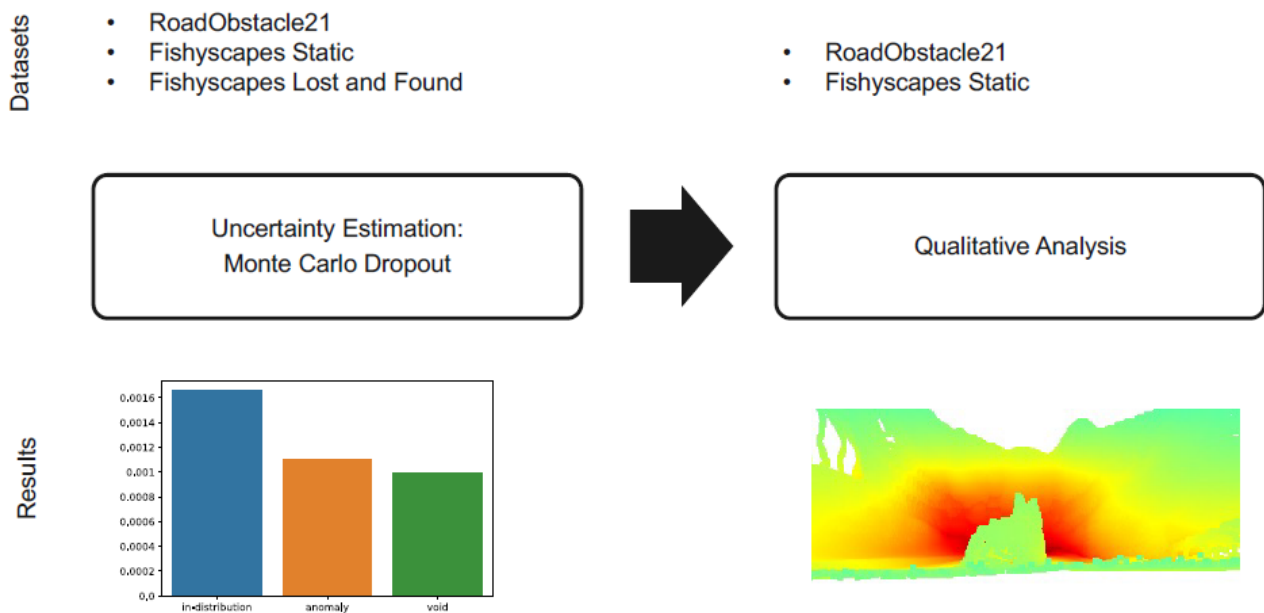


Abbildung 16: Zunächst wird eine quantitative Analyse mit Monte Carlo Dropout durchgeführt. Dann wird eine qualitative Analyse durchgeführt, bei der der geschätzten Punktwolken durchgeführt, um zu sehen, wie gut Corner Cases in 3D abgebildet wurden.

Es zeigt sich insbesondere, dass Chamfer-Distanzen und Unsicherheiten zwischen normalen Daten und Corner Cases nah beieinander liegen. Insbesondere bei Klassen, die teils als Corner Cases betrachtet werden (z.B. Hunde), zeigt die self-supervised Methodik hier ihre Stärken. Während klassische Methoden einen falschen Corner Case und somit falsche Label (FP) liefern würden, eignet sich der sensorübergreifende Ansatz hervorragend aufgrund erhöhter Robustheit, wie Abbildung 17 exemplarisch zeigt.

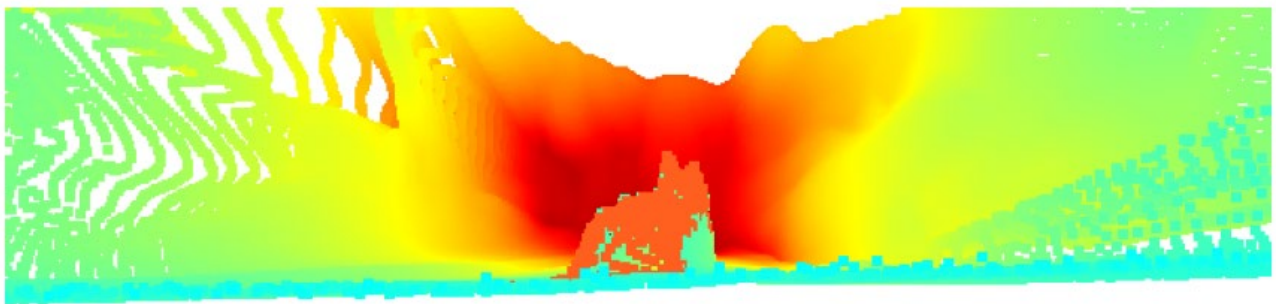


Abbildung 17: Vergleich von Pseudo-LiDAR mit Kameradaten. Es ist sichtbar, dass eine geometrische Punktwolke ähnlich zu LiDAR-Daten erkannt wird.

Bei echten Anomalien, z.B. augmentierten transparenten (unrealistischen) Tieren auf der Fahrbahn, schlägt die Methodik an und zeigt starke Unterschiede zu geometrischen LiDAR-Daten; Anomalien im Bildbereich werden nicht in den geometrischen Raum des Pseudo-LiDAR gemappt, was eine Unterscheidung zulässt, wie Abbildung 18 zeigt:



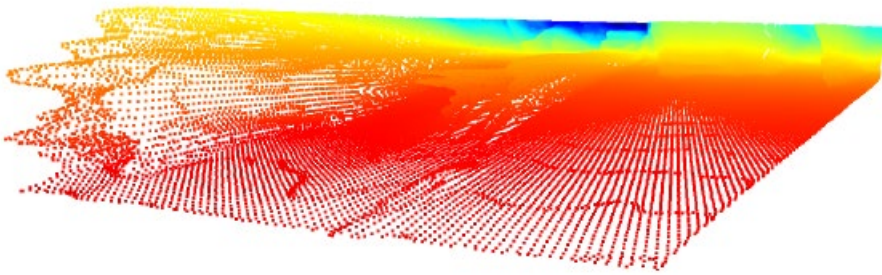


Abbildung 18: Vergleich von Pseudo-LiDAR mit Kameradaten. Es ist sichtbar, dass eine geometrische Punktwolke unähnlich zu LiDAR-Daten erkannt wird, was eine Anomalie darstellt.

Auf der anderen Seite zeigt sich, dass die Methodik wenig geeignet ist, um dynamische Anomalien zu erkennen, da das Rauschen der Prädiktion einen tatsächlichen Effekt überdeckt; statische Objekte können nicht von anomal dynamischen unterschieden werden, wie Abbildung 19 zeigt.

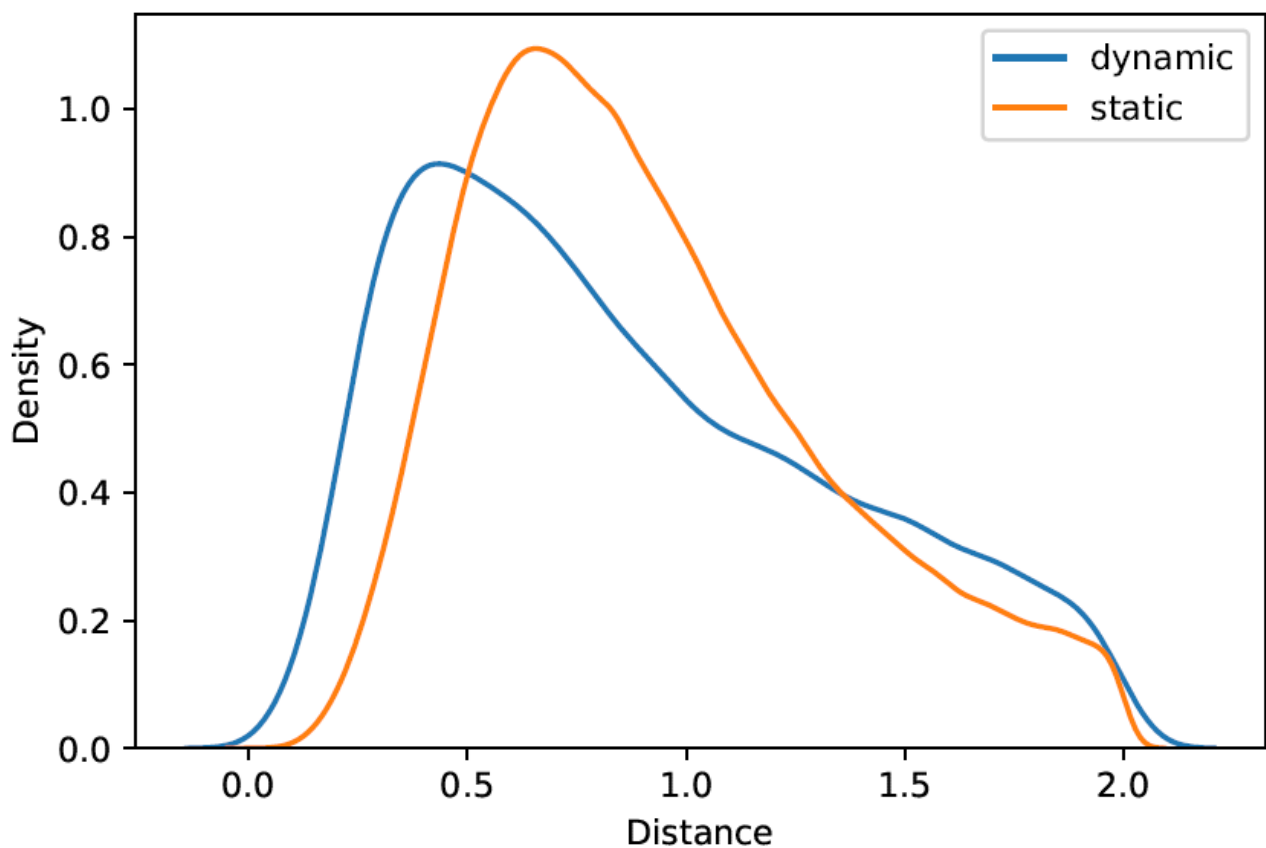


Abbildung 19: Bewertung der Bewegung von Instanzen über die Zeit mit der Chamfer-Distanz.

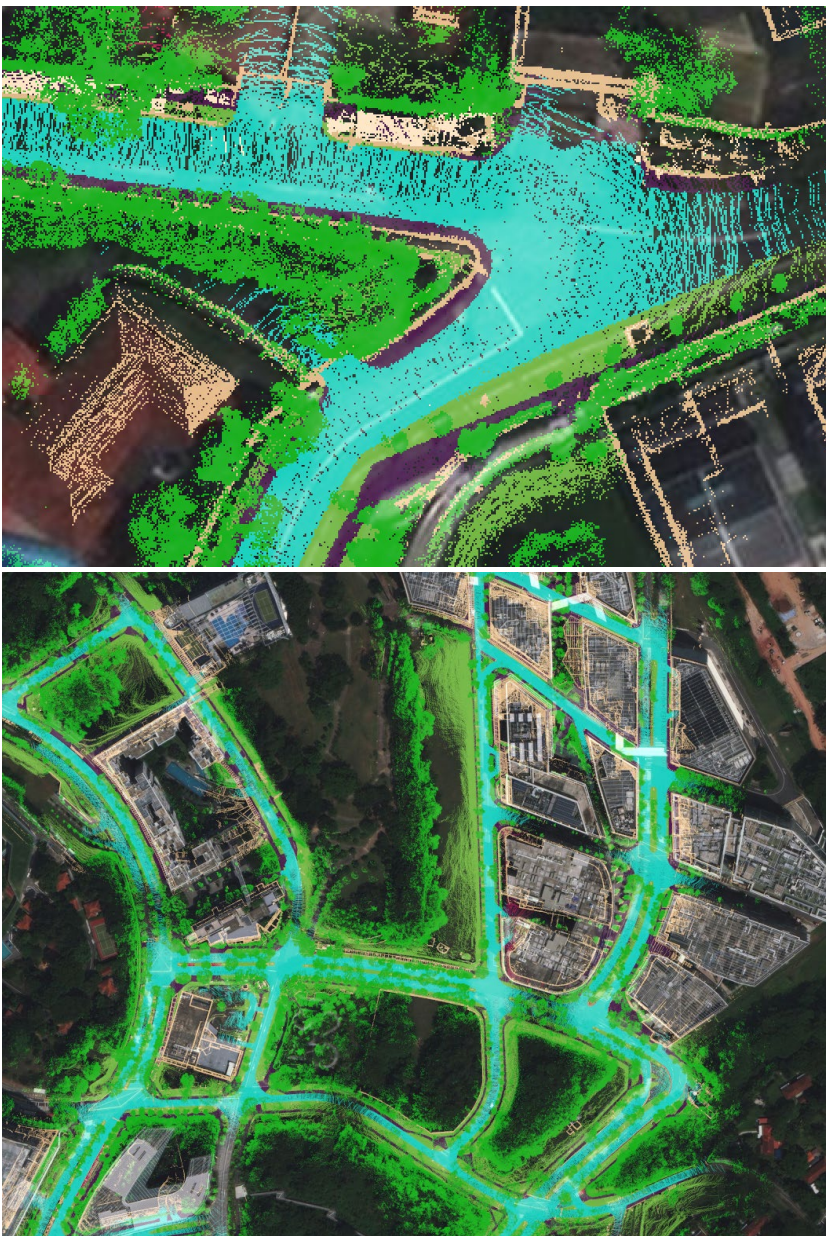
Zusammenfassend lässt sich sagen, dass statische Corner Case Klassen durch sensorübergreifendes Labeling mit höherer Erfolgsquote als dynamische Corner Cases gelabelt werden können.

3.1.7 AP3.1 Methoden / Toolentwicklung für Autolabeling

In AP3.1 entwickelte das FZI mehrere Fusionsansätze, mit denen Informationen zwischen unterschiedlichen Sensormodalitäten fusioniert und ergänzt werden können. Dieser Informationsgewinn wird daraufhin verwendet um automatisierte Pre-Labeling Prozesse zu optimieren.

Das FZI beschäftigte sich in AP3.1 hauptsächlich mit der Generierung von Kartenmaterial mithilfe von multimodaler Sensorik und unterschiedlichen Perspektiven. Während beispielsweise ein LiDAR und ein Kamerasensor unterschiedliche Sensormodalitäten besitzen und so unterschiedliche Informationen liefern, sind sie bei einer Anbringung auf einem Versuchsträger doch in Ihrer wahrnehmbaren Umgebung limitiert. Das Hinzuziehen von Kamerabildern aus einer Top-Down Perspektive, wie beispielsweise von Satelliten, Flugzeugen oder Drohnen erweitert die wahrnehmbare Umgebung durch die Ausrichtung der Perspektive. Durch die Verwendung von Luftbildern ist eine Erhöhung der Genauigkeit in der Aufgabe der Segmentierung und Generation von Karteninformationen zu erwarten.

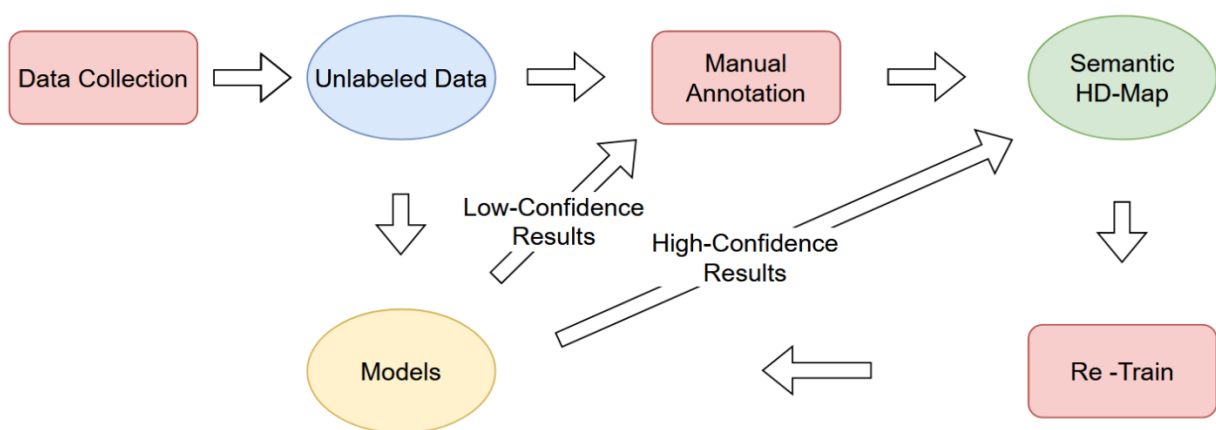
Datensätze mit gelabelten Luftbildern sind oft nur klein und sind deswegen ein Bottleneck für das Training von Neuronalen Netzen. Um dies entgegenzuwirken wurde ein Tool entwickelt, mit dem gelabelte LiDAR Punkte auf ein Luftbild projiziert werden. Da viele LiDAR Segmentierungs Datensätze vorhanden sind, und LiDAR Segmentierung auch mit weiterer Kamerasegmentierung erweitert werden kann, ist dies ein geeigneter Weg um Labels auf Luftbilder zu übertragen. Anhand eines Beispiels der Regionen Singapur und Boston, welche im NuScenes Datensatz enthalten sind, wurden gelabelte LiDAR Punkte verwendet um gelabelte Luftbilder der Regionen zu erhalten.



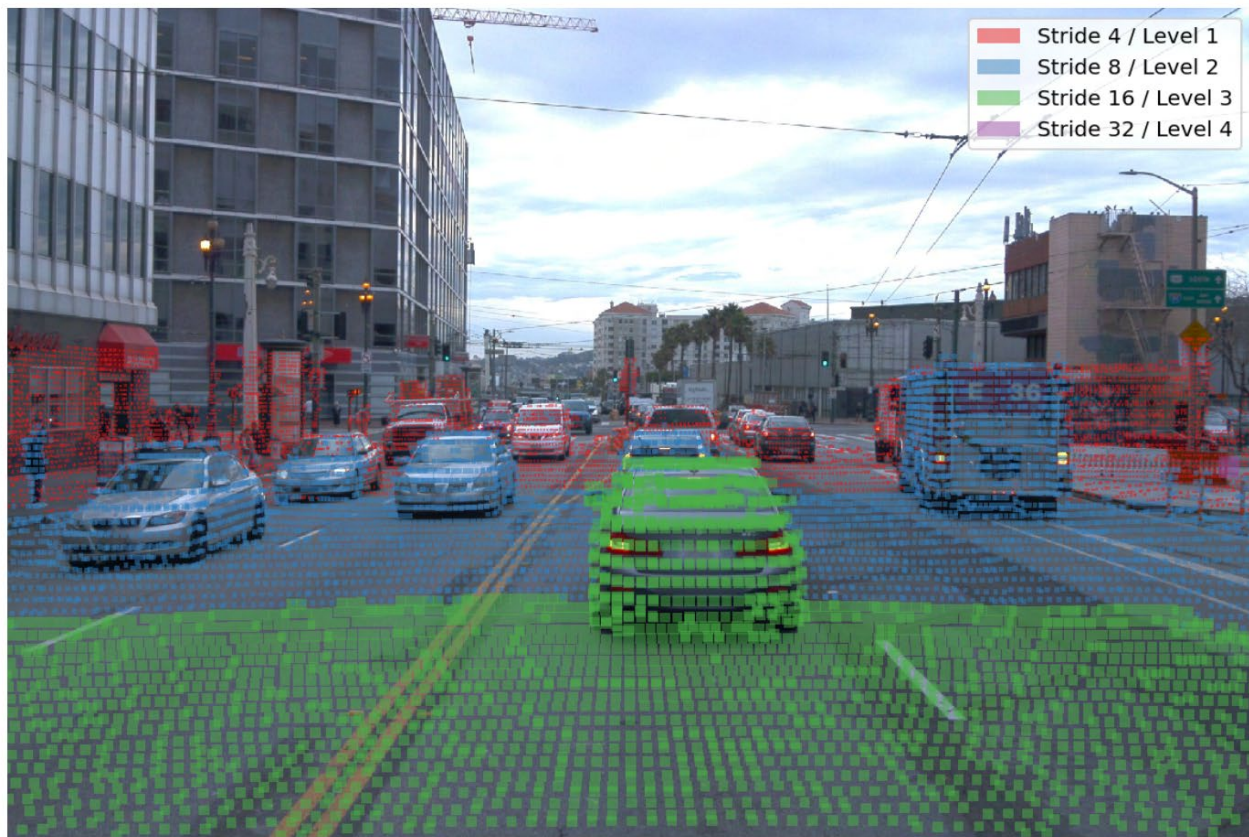


Als Übersicht ist hier der Bereich zu sehen, der mit temporal fusionierten, lokalisierten und segmentierten LiDAR Punktwolken im Luftbild gelabelt wurde. Aufgrund der Auflösung des gelabelten Luftbildes von 400 Megapixeln ist hier nur eine niedrig aufgelöste Version dargestellt.

Mit diesen gewonnenen Labels und anderen segmentierten Luftbild Datensätzen wurden neuronale Netze trainiert, mit denen weitere Regionen gelabelt werden sollen, für die keine LiDAR Punktwolken zur Verfügung stehen. Für das Labeln weiterer Regionen wurde Human-In-The-Loop Labeling Verfahren eingesetzt und dazugehöriges Tooling dafür erstellt. Mithilfe des Programms CVAT und der trainierten neuronalen Netze sowie dem Foundation Modell SegmentAnything wurden Resultate mit hoher Konfidenz direkt als Label gespeichert während Resultate mit niedriger Konfidenz manuell verifiziert oder falsifiziert werden. Die Luftbildsegmentierungsmodelle wurden mit den neuen gelabelten Daten erneut trainiert um die Performance des weiteren Pre-Labelings zu erhöhen.



Schlussendlich entstand im Rahmen des AP3.1 eine Masterarbeit mit dem Titel "Machine Learning basierte Fusion von Kamera- und LiDAR-Daten für das teilautomatisierte Pre-Labeling" in der untersucht wurde ob durch eine ML-basierte Fusion der Kamera- und LiDAR-Daten eine Verbesserung des Pre-Labelings von 3D Bounding Boxen erreicht werden kann. Für die Fusion wurde ein Modell konzipiert und trainiert, welches die komplementären Informationen der beiden Sensormodalitäten fusioniert und mit Hilfe der erlernten gemeinsamen Repräsentation die gewünschten 3D Bounding Boxen ausgibt. Genauer wurde hauptsächlich untersucht wie am besten Bildfeatures, die aus einem CNN gewonnen werden mit Voxelfeatures einer voxelisierten Punktwolke fusioniert werden können. Die Voxel im 3D Raum werden in das 2D Kamerabild projiziert um passende Bildfeauters mit Voxelfeatures zu fusionieren. Durch die Kameraprojektion erscheinen weit entfernte Voxel kleiner im Bild als nahe Voxel. Diese verschiedenen Skalierungen stellen ein Problem für die Fusion dar. Wird ein Voxel, dessen räumliche Ausdehnung sehr klein ist, auf eine Featuremap mit grober Auflösung projiziert, fehlen die räumlichen Details in den Bildfeatures. Dies ist bei weit entfernten Voxeln der Fall. Wird dahingegen ein Voxel nahe der Kamera ins Bild projiziert, möchte man möglichst high-level Features der tieferen Layers zur Fusion nutzen. Aufgrund der großen abgedeckten Fläche im Bild ist dabei eine gröbere Auflösung ausreichend. Um dieses Problem zu lösen wurden die Bildfeautures mit einer Feature Pyramid Network aufbereitet und eine Funktion entwickelt, die jedem Voxel ein Level in der Feature Pyramide zuordnet.



Zusätzlich wurden Layer trainiert, die einen Offset für jedes Voxelzentrum bestimmen, um die Voxelfeatures besser mit Bildfeatures zu fusionieren. Auf dem Benchmark des Datensatzes Waymo wurde durch diese Methode eine deutlich erhöhte Detektion von Vulnerable Road Users erreicht, da besonders hier die LiDAR Punktwolke oft nicht genug Punkte enthält um diese zuverlässig zu erkennen.

3.1.8 AP3.2 Aufnahme von Realdaten

Das Arbeitspaket lieferte Daten zu Validierung des Radarsensormodells (vgl. AP1.2). Hierzu wurde basierend auf dem TI AWR 1843 und dem DCA1000EVM ein HW/SW-System umgesetzt, welches die Aufzeichnung des Radarsignals direkt an der Intermediate Frequency (IF) Schnittstelle ermöglicht. Basierend auf den aufgezeichneten Daten, wird mittels Matlab die Detektionen sowie Range-Azimuth berechnet. Dieses Vorgehen wurde gewählt, um mit aufgenommenen Daten unterschiedliche Postprocessing-Konfigurationen zu evaluieren.

Basierend auf diesem System wurden unterschiedliche Messungen zur Validierung des Systems und darauf aufbauend zur Validierung des geometriebasierten Sensormodells (vgl. AP1.3).

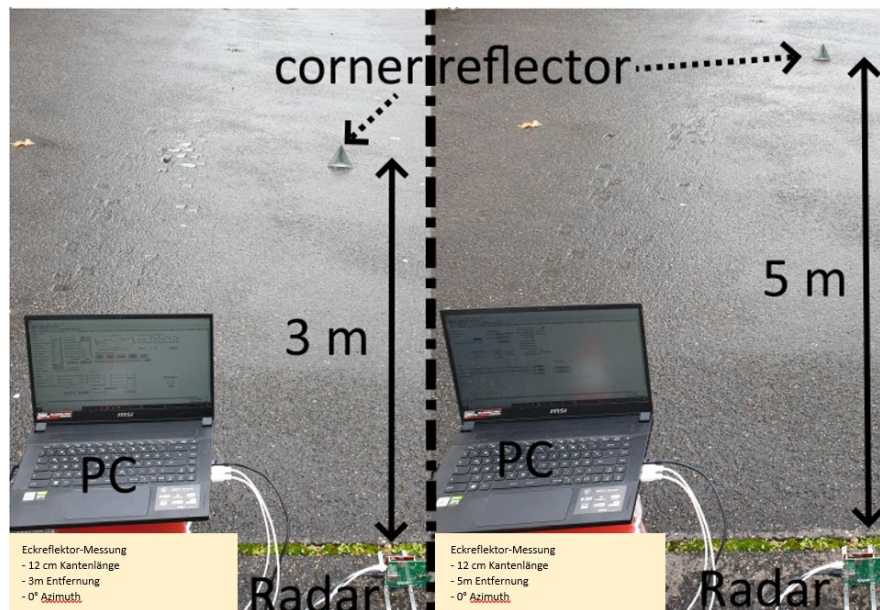


Abbildung 20: Physischer Aufbau, zur Validierung geometrischer Eigenschaften des Radarsensormodells

3.1.9 AP3.3 Aufbereitung der Trainingsdaten

Erarbeitung von Kriterien zur Bewertung der Datenqualität in der jeweiligen Verarbeitungsstufe

Am FZI wurde für das Ergebnis E3.3.5 an Verfahren zur Bewertung der Datenqualität der in verschiedenen Sensoren (Kamera und LiDAR) koregistrierten Daten, insbesondere mit Fokus auf die Bewertung und Sicherstellung der Konsistenz von Annotations-Identifikatoren in den verschiedenen Sensor-Modalitäten des Datensatzes, gearbeitet.

Eine Literaturrecherche hat ergeben, dass es bereits eine Vielzahl an Verfahren zum Training von tiefen neuronalen Netzen auf Datensätzen mit verrauschten (d. h. teilweise inkorrekten) Annotationen (Label) gibt¹¹. Für die Zielsetzung des KI Data Tooling Projekts sind dabei vor allem Ansätze interessant, welche inkorrekte Label in den aufgezeichneten Datensätzen detektieren (um nachgelagert eine manuelle Korrektur zu ermöglichen) oder selbstständig korrigieren (z. B. ¹²). Andere Ansätze zielen stattdessen darauf ab, beim Training auf Datensätzen mit verrauschten Labeln bessere Ergebnisse zu erzielen, ohne die Label explizit zu korrigieren. Solche Verfahren sind für die Entwicklung hochautomatisierter Fahrzeuge weniger geeignet, da nur eine Korrektur der Label im ursprünglichen Datensatz die Weiterverwendung dieses verbesserten Datensatzes für andere Entwicklungsschritte erlaubt. Weiterhin erleichtert die Korrektur der Label die Nachvollziehbarkeit und Validierung sowohl des Datensatzes als auch der damit entwickelten Systemkomponenten.

¹¹ H. Song, M. Kim, D. Park, Y. Shin, und J.-G. Lee, „Learning From Noisy Labels With Deep Neural Networks: A Survey“, *IEEE Transactions on Neural Networks and Learning Systems*, S. 1-19, 2022, doi: [10.1109/TNNLS.2022.3152527](https://doi.org/10.1109/TNNLS.2022.3152527).

¹² C. G. Northcutt, L. Jiang, und I. L. Chuang, „Confident Learning: Estimating Uncertainty in Dataset Labels“, *arXiv:1911.00068 [cs, stat]*, Apr. 2021, Zugriffen: 2. Mai 2022. [Online]. Verfügbar unter: <http://arxiv.org/abs/1911.00068>

Aufbauend auf dem Stand der Technik wurden am FZI dazu die bekannten Verfahren zur Verwendung im Entwicklungsprozess hochautomatisierter Fahrzeuge adaptiert bzw. weiterentwickelt:

1. Es wurde ein Verfahren entwickelt, das gleichzeitig die Klassenlabel und die Lokalisations-Label (d. h. die 3D Bounding Box) korrigieren kann. Dazu wurde ein auf einem Ensemble von Detektoren basierender Ansatz aus der Literatur ¹³ zur Korrektur von fehlerhaften 2D-Bounding-Box-Annotationen weiterentwickelt. Das Verfahren erlaubt die Korrektur dreidimensionaler Bounding-Boxen allein aus monokularen Kameradaten ohne auf z.B. LiDAR-Messdaten zurückgreifen zu müssen.
2. Die Literaturrecherche hat keine Verfahren zur Korrektur der im Datensatz abgelegten Instanz-IDs hervorgebracht. Aufbauend auf Ansätzen zum Objekttracking mit tiefen neuronalen Netzen sollen metrische Lernverfahren angewendet werden, um falsche Instanz-IDs im Datensatz zu identifizieren.

Identifikation falscher Klassenlabel und 3D-Bounding-Boxen

Bei der manuellen oder automatischen Annotation von Datensätzen können Objekte vereinzelt falsche Klassen-Labeln zugeordnet werden (z.B. ein Motorrad als Fahrrad gelabelt werden) oder die Objekt-Position, Orientierung oder Dimension ungenau angegeben werden. Diese falschen Label können die Ergebnisse von auf diesen Datensätzen trainierten Neuronalen Netzwerken beeinträchtigen. Für E3.3.5 wurde ein Verfahren entwickelt, das gleichzeitig die Klassenlabel und die Lokalisations-Label (d. h. die 3D Bounding Box) korrigieren kann. Dazu wurde ein auf einem Ensemble von Detektoren basierender Ansatz aus der Literatur (Wang et al.) zur Korrektur von fehlerhaften 2D-Bounding-Box-Annotationen weiterentwickelt. Das Verfahren erlaubt die Korrektur dreidimensionaler Bounding-Boxen allein aus monokularen Kameradaten ohne auf z.B. LiDAR-Messdaten zurückgreifen zu müssen. Dies kann vorteilhaft sein, da in vielen Labeltools die Annotation der 3D-Bounding-Boxen anhand der Punktwolken der LiDAR-Messdaten geschieht. Durch die so über verschiedene Verarbeitungsstufen geschehende Koregistrierung der Annotation der Objekte in verschiedenen Sensor-Modalitäten wird die Abhängigkeit der Korrektheit der Annotationen von einer Sensordomäne (in diesem Fall den Kameradaten), reduziert.

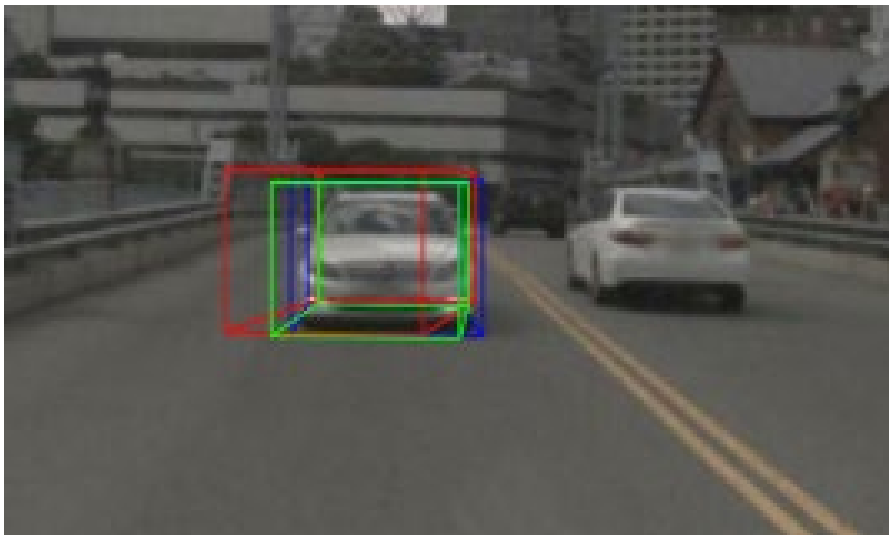


Abbildung 21: Fahrzeug mit fehlerhaftem 3D-Bounding-Box-Label aus dem Datensatz (rot), prädizierter Bounding Box (blau) und korrigiertem Bounding-Box-Label (grün)

¹³ S. Wang, J. Gao, B. Li, und W. Hu, „Narrowing the Gap: Improved Detector Training with Noisy Location Annotations“. arXiv, 12. Juni 2022. doi: [10.48550/arXiv.2206.05708](https://doi.org/10.48550/arXiv.2206.05708).



Die meisten 3D-Objektdetektoren präzisieren für jedes Objekt zunächst mehrere Bounding-Boxen und Klassenscores, die in einem weiteren Verarbeitungsschritt abhängig von der Überschneidung der Bounding-Boxen und den Klassenscores mittels Non-Maximum-Suppression (NMS) zu einer einzelnen Prädiktion pro Objekt reduziert werden. Für die Korrektur der 3D-Bounding-Boxen wird das monokulare 3D-Detektionsnetz FCOS3D¹⁴ zur Inferenz auf dem zu korrigierenden Datensatz angewandt und die einzelnen Prädiktionen vor der Non-Maximum-Suppression mit der Bounding-Box-Annotation in dem Datensatz fusioniert. Dabei werden die Koordinaten der annotierten Bounding-Box und derjenigen prädierten Bounding-Boxen, deren Überschneidung (genauer die 3D-IoU) mit der Bounding-Box über einem festgelegten Schwellwert liegen, gemittelt. Die Koordinaten der prädierten Bounding-Boxen werden dabei mit dem jeweiligen prädierten Klassenscore gewichtet. Die korrigierte Klasse kann wie üblich bei Ensemble-Methoden über ein Voting der prädierten Klassenscores bestimmt und mit dem Klassenlabel aus dem Datensatz abgeglichen werden. Anstelle der 3D-IoU kann für das Matching der prädierten Bounding-Boxen mit der annotierten Bounding-Box auch der Abstand des Mittelpunktes der Bounding-Boxen verwendet werden. In der Evaluation erwies sich die 3D-IoU vorteilhaft für die Erkennung von falschen Klassenlabel und die Distanz der Mittelpunkte für die Detektion von ungenauen Bounding-Box-Annotationen.

Das Verfahren kann prinzipiell zwar auch zur vollautomatischen Korrektur der Annotationen angewandt werden, in der Praxis sinnvoller ist es aber wohl das Verfahren nur zur Suche nach potenziell falschen Annotationen einzusetzen, die dann nachgelagert manuell überprüft und ggf. korrigiert werden können. So ist die maximal zu erwartende Qualität der korrigierten Label nicht beschränkt auf die Detektionsgenauigkeit des eingesetzten monokularen 3D-Detektionsnetzes. In einem Beispielszenario, in dem angenommen wird, dass im NuScenes-Datensatz 10% aller Bounding-Box-Annotationen fehlerhaft bzw. ungenau seien, konnten über 95% aller fehlerhaften Annotation erkannt werden. Das Verfahren verringert im Vergleich zu einer rein manuellen Kontrolle der Label den Anteil der erneut zu überprüfenden Label auf ein Sechstel.

Identifikation falscher Instanz-IDs

Die Literaturrecherche hat keine Verfahren zur Korrektur der im Datensatz abgelegten Instanz-IDs hervorgebracht. Diese Instanz-IDs erlauben die Identifikation des selben Verkehrsteilnehmers über mehrere aufeinanderfolgende Bilder in dem Datensatz. Insbesondere wenn das Labeling primär in der für den Menschen wenig intuitiven LiDAR-Sensordomäne geschieht, passieren leicht Labelingfehler, bei dem einer Objektinstanz in einem Bild eine andere Instanz-ID als in dem letzten vorherigen Bild, in dem das Objekt zu sehen war, zugewiesen wird. Die Gefahr dafür ist insbesondere dann groß, wenn ein Objekt temporär aus dem Sichtbereich der Kamera verschwindet oder mehrere Verkehrsteilnehmer sich eng aneinander vorbeibewegen. Um die Konsistenz der Annotationsidentifikatoren über die LiDAR-Sensorgrenze hinweg sicherzustellen, wurden aufbauend auf Ansätzen zum Objekttracking mit tiefen neuronalen Netzen metrische Lernverfahren angewendet, um falsche Instanz-IDs im Datensatz zu identifizieren. Dazu wird ein siamesisches Netzwerk durch die Minimierung des Triplet-Loss trainiert, ein Embedding für die Bildausschnitte in der 2D-Bounding-Box um die Objektinstanzen zu produzieren, für das die Kosinusdistanz zwischen den Embeddings zweier Bildausschnitte der selben Instanz klein und die zweier unterschiedlicher Instanzen groß ist. Wird für jede Bounding-Box in dem Datensatz ein solches Embedding berechnet, kann z.B. mittels Clustering-Verfahren bestimmt werden, welche Bounding-Boxes zur selben Objektinstanz

¹⁴ T. Wang, X. Zhu, J. Pang, und D. Lin, „FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection“. arXiv, 24. September 2021. doi: [10.48550/arXiv.2104.10956](https://doi.org/10.48550/arXiv.2104.10956).

gehören. Instanz-ID-Label in dem Datensatz, die inkonsistent mit diesem Clustering sind, sind vermutlich falsch und sollten nochmals überprüft werden.

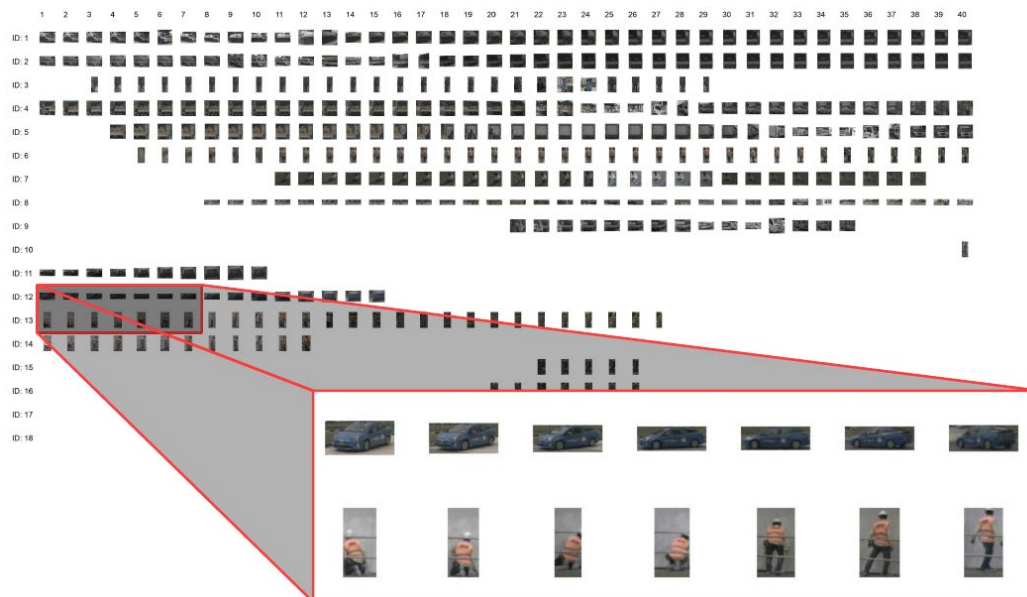


Abbildung 22: Tracks verschiedener Objekte (Zeilen), die verifiziert werden sollen

In der Evaluation auf dem Nulmages-Datensatz konnte die prinzipielle Tauglichkeit des Ansatzes gezeigt werden. Die Verwendung zusätzlicher Konsistenzbedingungen unter Einbeziehung der zur Verfügung stehenden Metadaten wie annotierter Klasse, Position der Objektinstanzen in globalen Koordinaten und ID des Bildes in dem die Objektinstanz vorkommt konnte die Robustheit des Ansatzes deutlich verbessern.

Zur gezielten Erzeugung von Störmustern geringer Leistung deren Hinzufügen für KI-Verfahren maximale Leistungseinbußen generiert, wurde ein vortrainiertes Netzwerk zur Erzeugung von realistischen, an die Kamera angepassten Bildrauschen verwendet¹⁵. Das durch dieses Netzwerk generierte Bildrauschen berücksichtigt sowohl die spezifischen Eigenschaften der aufnehmenden Kamera und des aufgenommenen Bildes bzw. der Szenerie und kann daher das Kamerarauschen wesentlich exakter abbilden als klassische Verfahren dafür (z.B. einfaches Gaußsches Rauschen). Mit diesem Netzwerk können Bilder mit einer beliebigen Anzahl realistischer Rauschmuster augmentiert werden, was die gezielte Untersuchung des Einflusses dieser realistischen Rauschmuster auf z.B. die Detektionsgenauigkeit eines Detektionsnetzwerkes erlaubt. Somit können auch Szenarien, in denen Kamerarauschen besonders problematisch sein kann, identifiziert werden. Da es sich bei dem das Bildrauschen generierende Netzwerk um ein Normalizing-Flow-Netzwerk mit nur geringer Eingangsgröße handelt, welches zunächst nur für einzelne Bildpatches mit geringer Auflösung Bildrauschen generieren kann, musste eine Rekonstruktion des verrauschten Bildes aus den einzeln augmentierten Bildpatches implementiert werden. Abbildung 1 zeigt ein solches rekonstruiertes verrauschtes Bild für den Cityscapes-Datensatz.

¹⁵ S. Kousha, A. Maleky, M. S. Brown, und M. A. Brubaker, „Modeling sRGB Camera Noise with Normalizing Flows“. arXiv, 1. Juni 2022. doi: [10.48550/arXiv.2206.00812](https://doi.org/10.48550/arXiv.2206.00812).



Abbildung 23: Künstlich erzeugtes, realistisches Bildrauschen auf dem Cityscapes-Datensatz

3.1.10 AP3.4 Augmentierung von Realdaten

In AP3.4 wurde untersucht, ob und wie sich klassische Methoden und Methoden des maschinellen Lernens zur Augmentierung von Sensordaten nutzen lassen.

Klassische Augmentierungsverfahren für Kameradaten (E3.4.1)

Copy-Paste-Augmentierung

In der Welt der Computer Vision und des maschinellen Lernens ist die Erkennung und Verarbeitung von Objekten in verschiedenen Szenen eine entscheidende Aufgabe. Die Integration virtueller Elemente, wie beispielsweise Fußgänger, in Szenen ist ein wichtiger Schritt, um Modelle auf die Bewältigung realer Szenarien vorzubereiten. Die Copy-Paste-Augmentierung ermöglicht es, synthetische Daten zu generieren, die realen Szenen ähneln und gleichzeitig die Kontrolle über wichtige Faktoren wie Größenverhältnisse und Okklusion bieten.

Die Genauigkeit von Tiefenkarten und Segmentierungsbildern in realen Datensätzen kann variieren und kleine Ungenauigkeiten können zu Verzerrungen führen, die die Vorhersagen von Computer Vision-Modellen beeinflussen. Durch die Anpassung der Tiefenkarten und die Hervorhebung wichtiger Merkmale wie Fahrzeugkanten können wir die Genauigkeit und Zuverlässigkeit unserer Modelle verbessern.

Darüber hinaus sind die Copy-Paste-Augmentierungsmethoden äußerst nützlich, um das Training von Modellen mit begrenzten Datenressourcen zu unterstützen. Durch die Erstellung von vielfältigen, synthetischen Datensätzen können Modelle besser generalisieren und sind besser auf die Herausforderungen der echten Welt vorbereitet, insbesondere wenn es um die Interaktion zwischen Menschen und ihrer Umgebung geht.

Zusammengefasst ermöglichen die beschriebenen Copy-Paste-Augmentierungsmethoden die Verbesserung der Leistung von Computer Vision-Modellen, indem sie die Realitätsnähe der Trainingsdaten erhöhen und die Vorhersagegenauigkeit in komplexen Szenarien steigern. Dies trägt dazu bei, dass solche Modelle besser in der Lage sind, reale Szenen und Situationen zu verstehen und entsprechend zu reagieren.

Die hierfür erstellte Methode wählt zunächst ein zufälliges Bild und die dazugehörige Maske als Vordergrundelement sowie ein separates Bild und dessen Maske als Hintergrundelement aus.

Im nächsten Schritt werden Augmentierungsfunktionen angewendet. Dabei erfolgt eine zufällige horizontale oder vertikale Spiegelung des ausgewählten Vordergrund- und Hintergrundelements samt zugehöriger Masken.

Im Vordergrundbild werden die Objekte der Klasse "Person" erkannt, und es wird eine zufällige Person ausgewählt. Auf dem Hintergrundbild wird eine zufällige Position innerhalb der Kategorien "Straße", "Boden" und "Gehweg" gefunden, an der die ausgewählte Person platziert wird. Vor der Platzierung wird die Person entsprechend der Positionierung vergrößert oder verkleinert.

Auswahl Datensatz

Es wird ein Datensatz mit (Instanz-) Segmentierungsmasken für die zu augmentierende Klasse benötigt. Ein kontextuelles Einfügen setzt zudem semantische Segmentierungsmasken für Straße und/oder Fußweg voraus. Für die korrekte Skalierung der eingefügten Objekte muss zudem die Kamerakalibrierung in dem Datensatz hinterlegt sein. Vorteilhaft ist eine möglichst große Varianz der Daten, d.h. die Bilder sollten nicht in zu kurzen Zeitabstand aufgenommen sein damit nicht z.B. die selben Personen in zu vielen Bildern zu sehen sind. Diese Kriterien werden zum Beispiel vom Cityscapes-Datensatz erfüllt, welcher auch für die meisten durchgeführten Experimente verwendet worden ist.



Da nicht jedes Bild im Cityscapes-Datensatz perfekt annotiert ist, wird die Matting-Funktion verwendet, um die entsprechenden Masken eigenständig zu generieren. Die Matting-Funktion ermöglicht die Erstellung präziserer Masken und verhindert beim Einsetzen des Objekts (Vordergrundbild) in den Hintergrund scharfe Kanten zwischen dem Objekt und dem Hintergrund.

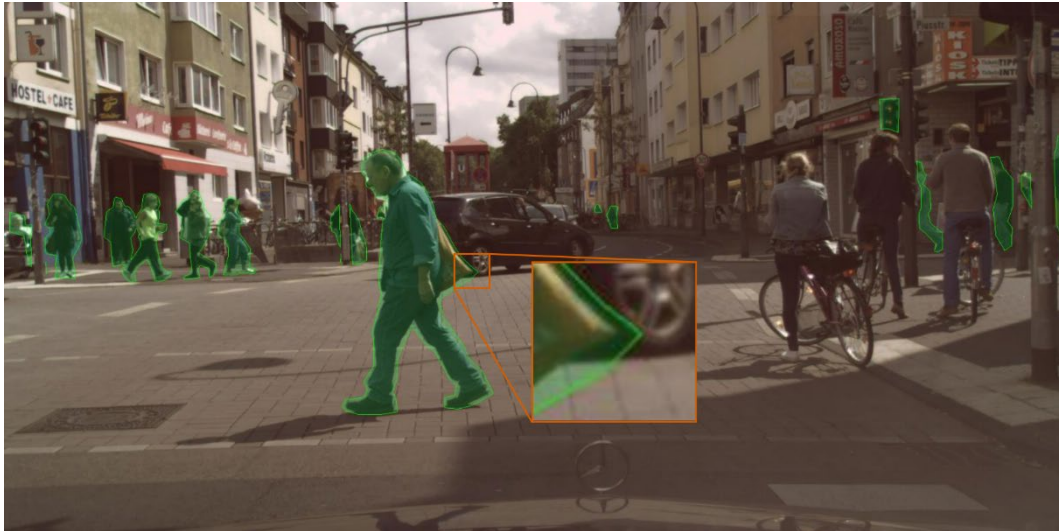


Abbildung 24: Darstellung der Maskenqualität von Cityscapes



Abbildung 25: Resultierendes augmentiertes Bild, zusammengesetzt aus Vordergrundelementen und Hintergrundelementen

Verdeckung bei Augmentierung

Das Ziel dieser Arbeit lässt sich kurz zusammenfassen als die Entwicklung eines Algorithmus, der virtuelle Fußgängerbilder basierend auf ihrer Platzierung in der Szene unter Berücksichtigung von Maßstab und Okklusionsbeziehung mithilfe des Cityscapes-Datensatzes, einer Tiefenkarte und eines Segmentierungsbildes in die Szene einfügen kann. Darüber hinaus sollen die Bilder mit eingefügten Fußgängern in großer Anzahl gleichzeitig erzeugt werden können. So können durch Verdeckung bedingte (z.B. hinter geparktem Fahrzeug hervortretender Fußgänger) Corner-Cases gezielt durch Insertion von Verkehrsteilnehmern aus Datensequenzen einer Situation in die Sequenzen einer anderen erzeugt werden. Es wurden verschiedene Techniken zur Generierung von Tiefenkarten und Segmentierungsbildern entwickelt, wobei wir uns entschieden haben, den Algorithmus von MiDaS für die Tiefenkarte und den Algorithmus von ERFnet für das Segmentierungsbild zu verwenden.

Während des Testlaufs stellten wir jedoch fest, dass aufgrund der leichten Ungenauigkeiten in der Tiefenkarte die Unschärfe der Grenzen in der Tiefenkarte zwischen verschiedenen Objekten im Hintergrund eine nach innen gewölbte Form aufweisen kann und zu mehr Unsicherheit bei der Okklusionsvorhersage führen kann.

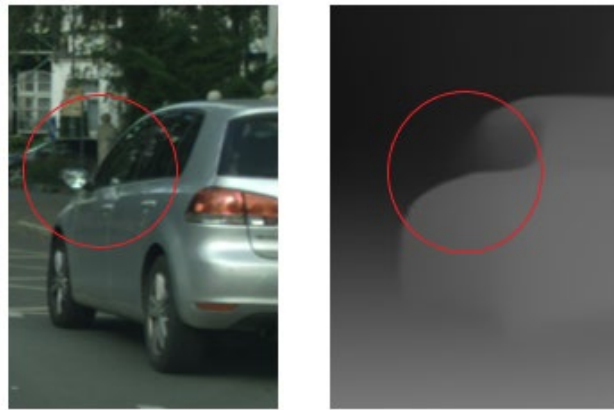


Abbildung 26: Ungenauigkeiten in der Tiefenkarte (rechts)

Um die Unsicherheit auszugleichen und die Tiefenkarte für die Berechnung der Okklusionstauglichkeit nutzbar zu machen, muss sie vor der Verwendung überarbeitet werden. Die häufigste Unsicherheit betrifft die Kanten der Fahrzeuge, die eine falsche Interferenz mit dem eingefügten Fußgänger verursachen können. Da sie nach Fußgängern die häufigsten Objekte sind, müssen die Kanten der Fahrzeuge in den Tiefenkarten deutlich erkennbar sein. Um die Genauigkeit zu gewährleisten, müssen die Kanten der Fahrzeuge betont werden, um sie deutlicher hervorzuheben.

Um dies zu erreichen, wählen wir zuerst die Masken der Fahrzeuge in der Hintergrundmaske aus und löschen die übrigen Masken im Bild. Da sie in einer bestimmten Graustufe vorliegen, sollte dieser Schritt schnell verarbeitet werden können. Nach diesem Prozess werden auch extrem kleine Konturen gelöscht, da sie möglicherweise irrtümlicherweise in die Maske aufgenommen wurden und nur minimalen Einfluss auf das Ergebnis haben. Der minimale Tiefenwert dm jedes Fahrzeugs wird durch den Vergleich der Fläche innerhalb der Kontur und derselben Fläche in der Tiefenkarte bestimmt. Hierbei wird nur der minimale gültige Tiefenwert aufgezeichnet (d.h. der kleinste Tiefenwert innerhalb der Kontur, ohne die nach innen gewölbten Bereiche). Nach Erhalt eines geeigneten minimalen Tiefenwerts dm werden die nach innen gewölbten Bereiche in der Tiefenkarte mit dem minimalen Tiefenwert dm gefüllt, um die Kante eines Fahrzeugs vollständiger und glatter zu machen.

Das Vorgehen funktioniert wie folgt:

Nachdem wir die Informationen zur Höhe und Breite erhalten haben, verwenden wir MiDaS, um die Tiefenkarte des Hintergrunds vorherzusagen. Mithilfe der im Datensatz verfügbaren Masken können wir die Disparität der Fußgänger berechnen. Anschließend verwenden wir einen Algorithmus, der auf ERFNet basiert, um Pixel, die zu Straßen, Gehwegen, Fahrzeugen, Verkehrsschildern usw. gehören, zu segmentieren und ihre Disparitäten aus der Tiefenkarte zu sammeln. Das Okklusionsverhältnis wird durch den Vergleich der Disparität des Fußgängers und der umgebenden Objekte aus der Tiefenkarte gelöst. Pixel mit größerer Disparität gehören zum Vordergrund und verdecken das Objekt.



Abbildung 27: Verdeckung bei Augmentierung, von links nach rechts: a) einzufügende, ausgeschnittene Person b) mit Hilfe der Tiefenprädiktion bestimmte Maske c) nicht verdeckter Teil der Person wird eingefügt d) augmentiertes Bild mit Verdeckung

Automatische Auswahl von passenden Augmentierungsobjekten

Bei der Augmentierung von Kamerabildern mit klassischen Augmentierungsverfahren stellt sich die Frage, wie passende Kombinationen realer Sensordaten und realer Datenobjekte identifiziert werden können, die bei der Erstellung eines Corner-Cases zu möglichst realistischen Resultaten führen. Als einzufügendes Datenobjekt wurden auch hier Fußgänger ausgewählt, der Ansatz lässt sich aber auch ohne größere Anpassungen auf andere Objektklassen übertragen. Bei der in AP3.4 vom FZI entwickelten Augmentierungspipeline, welche Bildausschnitte aus einem Bild abhängig von der Einfügeposition korrekt skaliert in einem anderen Bild einfügt, ist für einen überzeugenden visuellen Eindruck des augmentierten Bildes vor allem die korrekte Beleuchtung, das heißt Helligkeit, Kontrast, Farbe, Beleuchtungswinkel etc. des eingefügten Objekts entscheidend. So genannte Image-Harmonization-Ansätze, die für diesen Zweck ebenfalls in AP3.4 untersucht worden sind, können diese Bildeigenschaften für den eingefügten Bildausschnitt anpassen, so dass dieser besser zum Hintergrundbild passt. Allerdings funktionieren diese Verfahren nicht ohne Artefakte und können nicht jeden eingefügten Bildausschnitt korrekt an die Beleuchtungsverhältnisse im Hintergrundbild anpassen. Außerdem können auch noch andere Faktoren mitbestimmen, ob der eingefügte Bildausschnitt plausibel an der entsprechenden Stelle in das Hintergrundbild eingefügt werden kann. So ist zum Beispiel vom Bildkontext abhängig welche Gelenkposen bei Personen plausibel sind (z.B. auf einer Bank sitzend, über einen Bordstein steigend etc.), Kleidung ist von den Wetterverhältnissen abhängig, etc. Anstatt zu versuchen, jeden möglichen Einflussfaktor für sich genommen zu modellieren und zu korrigieren, wurde am FZI für Ergebnis E3.4.1 ein Verfahren entwickelt, mit dem automatisiert und ohne manuelle Spezifikation der Auswahlkriterien oder der zu beachtenden Einflussfaktoren abhängig vom Hintergrundbild und der ausgewählten Einfügeposition ein passender Bildausschnitt ausgewählt werden kann. Es handelt sich dabei um ein siamesisches Convolutional-Neural-Network, welches den lokalen Kontext um eine mögliche Einfügeposition (spezifiziert durch einen rechteckigen Bildausschnitt um die Einfügeposition) mit einem Einfügekandidaten vergleicht, und nach durchgeführtem Training eine Metrik ausgibt, die quantifiziert wie gut der einzufügende Bildausschnitt zum Hintergrundkontext passt. Dieses siamesische Netzwerk wird selbstüberwacht trainiert, das heißt ohne dass zusätzliche Annotationen (über die im Datensatz bereits vorhandenen Annotationen für Segmentationsmasken hinaus) verfügbar sein müssen. Da es sich um ein metrisches Lernverfahren handelt, werden positive Beispielpaare - in diesem Fall ein Bildausschnitt und ein Hintergrundbild, die gut zueinander passen - und negative Beispielpaare - in diesem Fall ein Bildausschnitt und ein Hintergrundbild, die schlecht

zueinander passen, benötigt. Die positiven Beispielpaare werden während dem Trainings durch eine Person (ausgeschnitten anhand der Segmentierungsannotation, mit dem Hintergrund ausmaskiert) und ihrem lokalen Kontext in ihrem Ursprungsbild gebildet. Da dies ein reales Beispiel für eine Person und ihrem tatsächlichen Hintergrund ist, passt die Person offensichtlich gut zu diesem Hintergrund. Damit das Netz nicht einfach nur die Person in beiden Bildausschnitten miteinander vergleicht, wird im Kontextbild die Person durch das einfarbige Ausfüllen der Bounding-Box der Person verdeckt. Negative Beispiele sind dann einfach zufällig vertauschte Personen und Hintergründe.

Der Ansatz wurde auf dem Cityscapes-Datensatz evaluiert und es konnte demonstriert werden, dass mit diesem Verfahren tatsächlich zuverlässig passende Kombinationen von einzufügenden Personen und Hintergrundbildern gefunden werden können (auch an Stellen im Hintergrundbild, an denen sich im Originalbild keine Person befand). Der Ansatz kann sowohl für eine online-Augmentierung parallel zum Training eines Detektionsnetzwerkes passende Kombinationen von Personen und Hintergrundbildern innerhalb eines Batch finden, als auch über einen gesamten Datensatz die am besten passendsten Personen gegeben eines Hintergrundbildes und einer Einfügeposition.

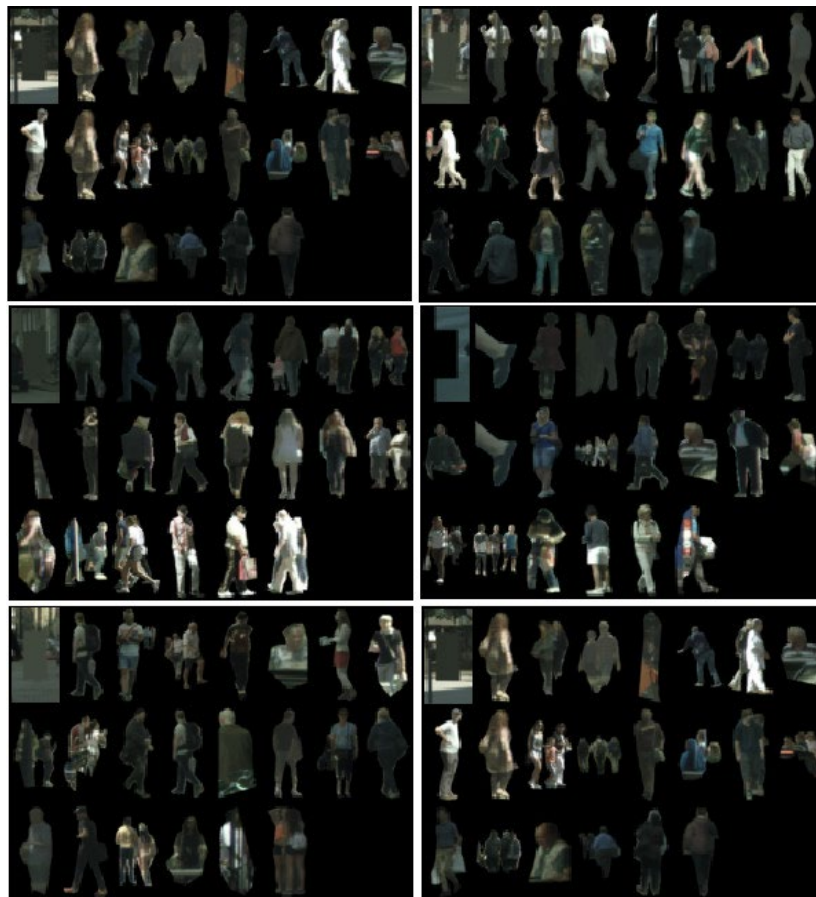


Abbildung 28: Sechs Batches, für die jeweils die in den Kontextbildausschnitt (jeweils oben links) einzufügenden Personen nach der gelernten Metrik sortiert dargestellt sind. Visuell passendere Personen haben einen kleineren Metrik-Abstand (weiter oben links).

Image Harmonization für Augmentierung mit klassischen Verfahren

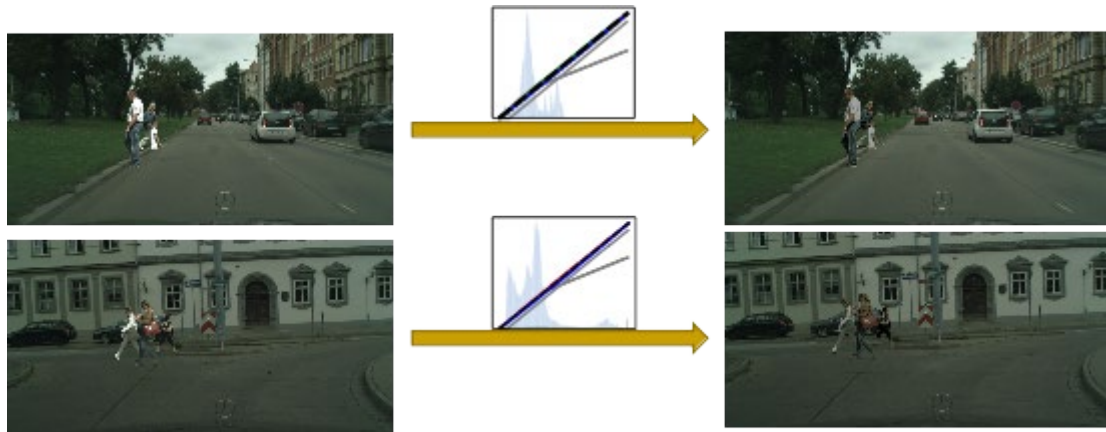


Abbildung 29: Gelernte Anpassung der Histogramme der eingefügten Bildausschnitte

Im Ergebnis E3.4.1 sind Augmentierungsverfahren entwickelt worden, die Objekte aus Bildern in einem Datensatz ausschneiden und in einem anderen Datensatz einfügen, wobei die Bildausschnitte aus dem Ursprungsdatensatz an die Beleuchtungsverhältnisse im Zieldatensatz angepasst werden. Dabei wird die Farbe und Helligkeit der eingefügten Bildausschnitte so angepasst, dass das resultierende augmentierte Bild einen realistischen Eindruck macht und die Augmentierung möglichst nicht an der nicht zum Bildhintergrund passenden Beleuchtung des eingefügten Objektes zu erkennen ist. Dazu wurde ein tiefes neuronales Netz trainiert das Farb-Histogramm mit nur wenigen Parametern für z.B. Helligkeit, Kontrast, Sättigung und Glanzlichter automatisch anzupassen, ähnlich zur manuellen Vorgehensweise mit Bildbearbeitungsprogrammen. Das Training findet dabei selbstüberwacht statt, das heißt es müssen keine zusätzlichen Daten gelabelt werden. Dazu werden während des Trainings die zu augmentierenden Bildausschnitte mit zufälligen Farb- und Helligkeitsanpassungen angepasst, jedoch nicht aus ihrem ursprünglichen Bild ausgeschnitten. Somit ist die optimale Farb- und Helligkeitsanpassung für diese Bildausschnitte bekannt - diejenige, die das Originalbild wiederherstellt.

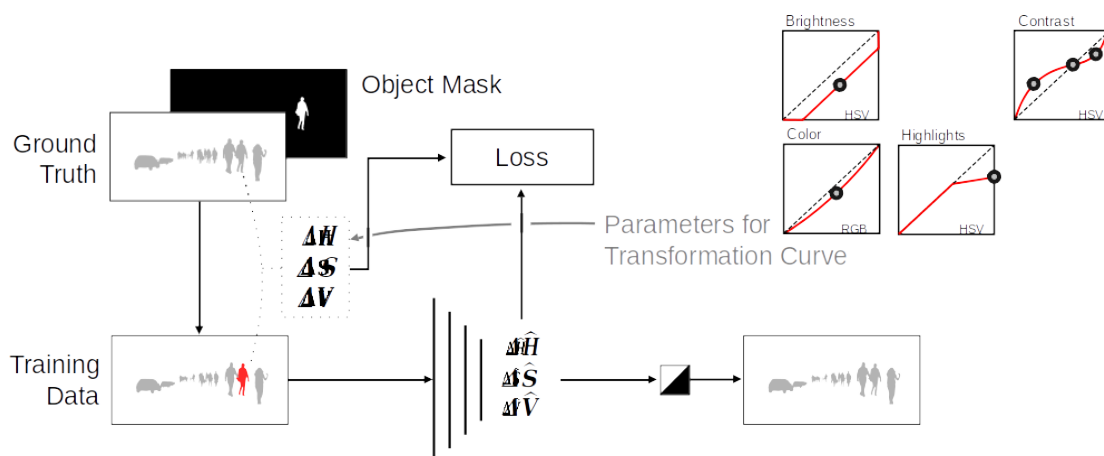


Abbildung 30: Das Netzwerk wird trainiert das Histogramm mit nur wenigen Parametern zu korrigieren, ähnlich zur Vorgehensweise bei der manuellen Bildbearbeitung

Klassische Augmentierungsverfahren für LiDAR-Daten (E3.4.2)

Am FZI wurden in E3.4.2 basierend auf den in E3.4.1 für Kameradaten vom FZI entwickelten Augmentierungsverfahren zum Einfügen von Straßenverkehrsteilnehmern aus Kamerabildern einer Szene in Kamerabilder einer anderen Szene ein Augmentierungsverfahren zum Einfügen von

Straßenverkehrsteilnehmern aus der LiDAR-Punktwolke einer Szene in die LiDAR-Punktwolke einer anderen Szene entwickelt. Mit beiden Verfahren zusammen lassen sich somit augmentierte Sensordaten für die beiden wichtigsten Sensordomänen autonomer Fahrzeuge (Kamera und LiDAR) generieren, und somit z.B. auch neue Corner-Cases in beiden Sensordomänen darstellen. Mit dem Verfahren, welches äquivalent zum kameradatenbasierten Ansatz aus E3.4.1 im Wesentlichen einen lokalen Ausschnitt der Sensordaten um das gewünschte Augmentierungsobjekt herum (z.B. einen Fußgänger), in diesem Fall der Punktwolke des LiDARs, ausschneidet und in einer anderen Punktwolke einer anderen Szene einfügt, konnten realistisch wirkende LiDAR-Sensordaten generiert werden.

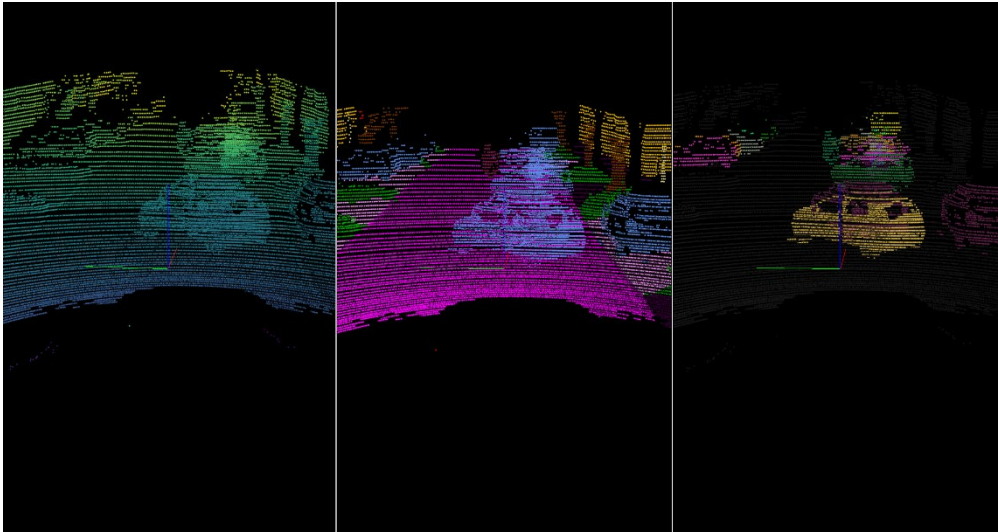


Abbildung 31: Durch Copy-Paste Augmentierung auf die Straße verschobene Objektpunktwolken

GAN-basierte Verfahren zur Kameradaten-Augmentierung (E3.4.3)

In diesem Ergebnis der Einsatz von Generative Adversarial Networks (GANs) geprüft. GANs lernen eine Approximation der Verteilung, welche einem Datensatz zugrunde liegt, und sind in der Lage anschließend neue Beispiele zu erzeugen, die aus dieser Verteilung stammen könnten.

GAN-Architekturen sind Gegenstand intensiver Forschung, weshalb es vielfältige Ansätze gibt und ihre Anzahl stetig steigt. Zunächst wurden einige existierenden Ansätze auf ihre Einsatzfähigkeit zur Augmentierung im Automotive-Kontext untersucht. Dabei wurden Augmentierungsverfahren betrachtet, die das gesamte Bild umfassen, aber auch Verfahren, die einzelne Objekte verändern oder hinzufügen. Hierbei musste für die existierenden GAN-Architekturen, jeweils ein Nutzungskonzept entwickelt werden und die Umsetzbarkeit bewertet werden. Anschließend wurde für eine Auswahl der als geeignet eingestufte Verfahren ein entsprechender prototypischer Prozess aufgebaut.

Zu Beginn wurden erste Experimente mit einfachen GAN-Architekturen durchgeführt, um die zu erwartenden Ergebnisse einschätzen zu können. Hierfür wurde zunächst die DCGAN-Architektur¹⁶ verwendet. Diese Architektur besteht aus Convolutional-Schichten, welche sich bei der Generierung von Bildern bewährt haben. Für das Training wurde der Stanford Cars Datensatz¹⁷, der Bilder von Autos beinhaltet, verwendet.

¹⁶ A. Radford, L. Metz, und S. Chintala, „Unsupervised representation learning with deep convolutional generative adversarial networks“, *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, S. 1-16, 2016.

¹⁷ J. Krause, M. Stark, J. Deng, und L. Fei-Fei, „3D object representations for fine-grained categorization“, *Proc. IEEE Int. Conf. Comput. Vis.*, S. 554-561, 2013, doi: 10.1109/ICCVW.2013.77.

Neu generierte Beispiel-Bilder aus der vom DCGAN gelernten Verteilung des Datensatzes sind in **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellt.



Abbildung 32: Neu generierte Beispiele eines mit dem Stanford Cars Datensatz trainierten DCGAN.

Die Bilder im Stanford Cars Datensatz zeigen alle die gleiche Objektklasse (Autos), wodurch es sich häufig wiederholenden Strukturen, z. B. die Scheinwerfer, gibt. Daher wurde im nächsten Schritt das Verhalten des Netzes bei heterogeneren Bildern untersucht. Hierzu wurde das DCGAN mit den Kamera-Bildern des A2D2-Datensatzes¹⁸ trainiert. Die Bilder des A2D2-Datensatzes sind deutlich diverser, da hier keine Objektklasse festgelegt ist, sondern die Bilder von der Fahrt mit einem Versuchsträger stammen. Die vom DCGAN generierten neuen Beispiele des A2D2-Datensatzes sind in **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellt.



Abbildung 33: Neu generierte Beispiele eines mit dem A2D2 Datensatz trainierten DCGAN.

¹⁸ J. Geyer u. a., „A2D2: Audi Autonomous Driving Dataset“, 2020.

Im nächsten Schritt wurde untersucht, ob sich einzelne Objekte (am Beispiel von Autos) innerhalb einer Szene generieren lassen. Hierzu wurde die Pix2Pix-Architektur¹⁹ verwendet und für die Generierung einzelner Objekte einer Szene angepasst²⁰. Für das Training mussten zunächst die Objekte, die erzeugt werden sollen, in den Trainingsbildern lokalisiert werden. Zur Lokalisierung wurde YOLO v3²¹ verwendet. Für das Training wurden die Objekte mit Rauschen Abbildung 32 bzw. mit den Konturen des ursprünglichen Objekts Abbildung 33 überzeichnet.

In einem weiteren Experiment war das Ziel einen Bildausschnitt markieren zu können, der anschließend von dem GAN mit einem generierten Fußgänger gefüllt wird. Im Experiment wurde der Cityscapes Datensatz²² genutzt, in dem die vorhandenen Fußgänger bereits gelabelt sind. Die Ergebnisse sahen nicht überzeugend aus. Ein Beispiel kann in Abbildung 34 betrachtet werden.



Abbildung 34: Fußgänger-Erzeugung mit Pix2Pix-Erweiterung

Der gleiche Ansatz wurde außerdem für die Generierung von Verkehrsschildern betrachtet. Hierzu wurde der GTSDb Datensatz²³ verwendet. In diesem sind die Verkehrsschilder inkl. des Schildtyps gelabelt. Durch die einfachere Struktur von Verkehrsschildern im Vergleich zu Fußgängern, bestand die Annahme, dass die Erzeugung einfacher sei. Statt Rauschen wurden ein Piktogramm eines Verkehrsschildes als Ausgangsbild genutzt (siehe Abbildung 35 Fehler! Verweisquelle konnte nicht gefunden werden.). Zur objektiven Evaluation wurde erneut auf YOLO v3 und die IoU-Metrik zurückgegriffen. Zur Bewertung wurde YOLO darauf trainiert Verkehrsschilder zu erkennen. Die Trainingsdaten bestanden dabei aus den Realdaten, augmentierten Daten oder einer Mischung aus beiden Datensätzen. Anschließend wurden die trainierten YOLO-Netze auf einem realen Test-Datensatz bewertet. Es konnte hierbei keine Verbesserung der Erkennungsraten durch die Augmentierung beobachtet werden.

¹⁹ P. Isola, J. Y. Zhu, T. Zhou, und A. A. Efros, „Image-to-image translation with conditional adversarial networks“, *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, Bd. 2017-Janua, S. 5967-5976, 2017, doi: 10.1109/CVPR.2017.632.

²⁰ X. Ouyang, Y. Cheng, Y. Jiang, C.-L. Li, und P. Zhou, „Pedestrian-Synthesis-GAN: Generating Pedestrian Data in Real Scene and Beyond“, 2018.

²¹ J. Redmon und A. Farhadi, „YOLOv3: An Incremental Improvement“, *Tech Rep.*, S. 1-6, 2018.

²² M. Cordts u. a., „The Cityscapes Dataset for Semantic Urban Scene Understanding“, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Juni 2016, S. 3213-3223. doi: 10.1109/CVPR.2016.350.

²³ S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, und C. Igel, „Detection of traffic signs in real-world images: The German traffic sign detection benchmark“, in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, Aug. 2013, S. 1-8. doi: 10.1109/IJCNN.2013.6706807.



Abbildung 35: Einfügen eines realistisch aussehenden Verkehrsschilds in eine Verkehrsszene mit einem CycleGAN auf Grundlage eines Piktogramms

Entscheidend ist die Position, an der ein neues Verkehrsschild eingefügt wird. Zur Positionsbestimmung werden das Kamerabild und die zugehörigen LiDAR-Daten kombiniert. Mit Hilfe von rechtlichen Regelungen über die Positionierung von Verkehrsschildern ist es, ausgehend von der dreidimensionalen Position der Fahrbahnkante, möglich realistische Verkehrsschildpositionen zu erzeugen (siehe Abbildung 36 **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die Vorschläge werden anschließend noch weiter gefiltert. So wird mit Hilfe der LiDAR-Daten ermittelt, ob in der Szene auch ausreichend Platz für den Pfosten eines Verkehrsschilds vorhanden ist (siehe Abbildung 37). Durch den Austausch der rechtlichen Regelungen, können mit dieser Methode realistische Verkehrsschildpositionen beliebiger Staaten generiert werden.



Abbildung 36: Vorgeschlagene Schildpositionen in einem Bild (rot); gefilterte Schildpositionen (grau) ²⁴

²⁴ P. Rigoll, P. Petersen, J. Langner, und E. Sax, „Parameterizable LiDAR-Assisted Traffic Sign Placement for the Augmentation of Driving Situations with CycleGAN“, in Advances in Systems Engineering, Bd. 364, L. Borzemski, H. Selvaraj, und J. Świątek, Hrsg., in Lecture Notes in Networks and Systems, vol. 364. , Cham: Springer International Publishing, 2022, S. 403-417. doi: 10.1007/978-3-030-92604-5_36.

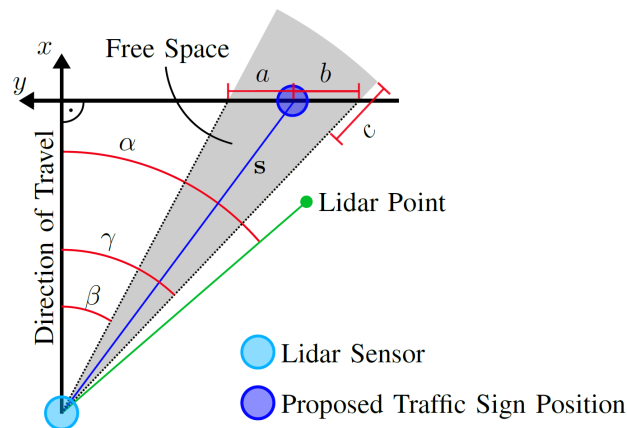


Abbildung 37: Schematische Darstellung zur Bestimmung, ob ausreichend Platz für das neue Verkehrsschild und den zugehörigen Pfosten vorhanden ist²⁴

Außerdem wurde eine Methode zur automatischen Datensatzerweiterung hinsichtlich unterschiedlicher Wettersituationen untersucht²⁵. Das Wetter ist bei der Datenaufnahme während einer Realfahrt nicht beeinflussbar. Bestimmte Wettersituationen können in einem Datensatz daher unterrepräsentiert sein. Das Ziel dieser Methode ist, die Abdeckung eines Datensatzes hinsichtlich der beobachteten Wettersituationen zu vergrößern. Hierzu werden Aufnahmen während häufig vorkommende Wettersituationen so verändert, als wären sie bei seltenen Wettersituationen aufgenommen worden.

Das grundlegende Vorgehen bei dieser Methode ist in Abbildung 38 dargestellt. Ist nicht der gesamte Datensatz annotiert, kann ein weiteres neuronales Netz zur Annotation verwendet werden. Im Falle von Wettersituationen, kann das WeatherNet²⁶ Verwendung finden. Für die Experimente haben wir den Oxford-Datensatz²⁷ verwendet, bei dem die Wettersituationen vollständig annotiert sind.

²⁵ Rigoll, P., P. Petersen, L. Ries, J. Langner, und E. Sax. „Augmentation von Kameradaten mit Generative Adversarial Networks (GANs) zur Absicherung automatisierter Fahrfunktionen“. In Fahrerassistenzsysteme und automatisiertes Fahren, herausgegeben von VDI Wissensforum GmbH, 41-48. VDI Verlag, 2022. <https://doi.org/10.51202/9783181023945-41>.

²⁶ Ibrahim, M., Haworth, J., und Cheng, T. (2019): WeatherNet: Recognising Weather and Visual Conditions from Street-Level Images Using Deep Residual Learning, IJGI, Bd. 8, Nr. 12, S. 549.

²⁷ Maddern, W., Pascoe, G., Linegar, C. und Newman, P. (2017): 1 year, 1000 km: The Oxford RobotCar dataset, The International Journal of Robotics Research, Bd. 36, Nr. 1, S. 3-15.

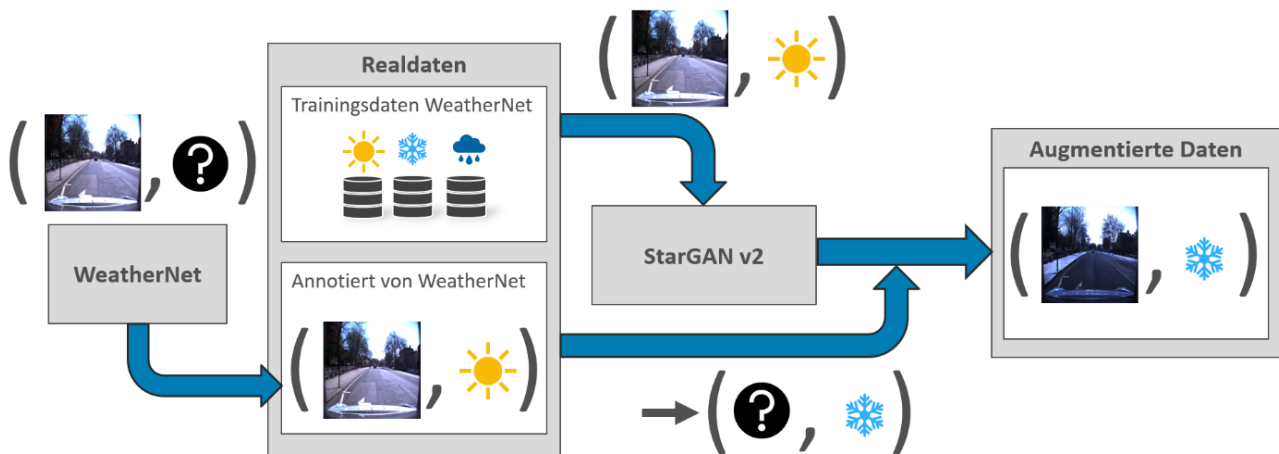


Abbildung 38: Prozess der Wetter-Augmentation²⁵

Zur Augmentation wird ein Generative Adversarial Network (GAN)²⁸ verwendet. Für diese Methode ist ein Standard-GAN jedoch nicht ausreichend, da dieses nur grundlegend neue Bilder generiert, statt bestehende Bilder abzuändern. Daher wird eine Abwandlung des Standard-GAN verwendet. Das StarGAN v2²⁹ ist in der Lage Bilder zwischen mehreren Domänen, hier Wettersituationen, zu transformieren. Abbildung 39 zeigt einige Augmentationsbeispiele.

²⁸ I. Goodfellow et al., "Generative Adversarial Nets," p. 9, 2014.

²⁹ Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," arXiv:1912.01865 [cs], Apr. 2020, Accessed: Feb. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1912.01865>

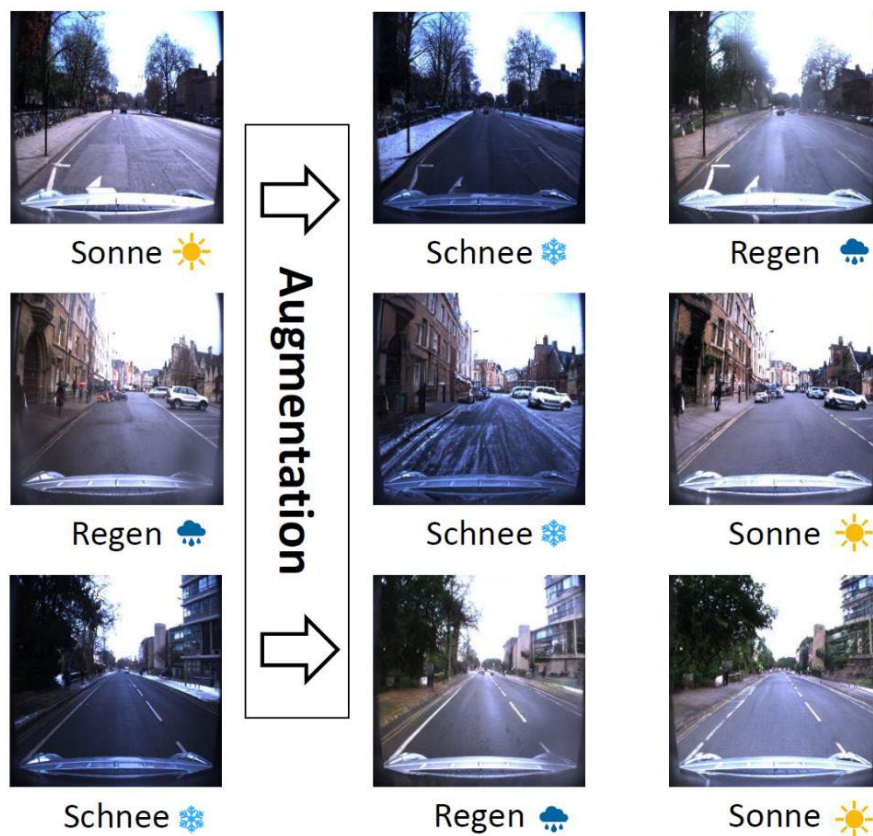


Abbildung 39: Beispiel der Wetter-Augmentation²⁵

Kriterien zur Bewertung der Güte der augmentierten Daten (E3.4.6)

Bilder die mit künstlichen neuronalen Netzen erzeugt werden, müssen vor ihrer Nutzung eingehend auf ihre Qualität und realitätstreue überprüft werden. Dabei ist es nicht ausreichend, die Bilder nur mit dem menschlichen Blick zu bewerten. Beispielsweise besitzen generierende Netze oft Dekonvolution-Schichten, welche sich in regelmäßigen Artefakten in den Bildern manifestieren können. Diese Artefakte sind mit dem menschlichen Auge oft nicht direkt erkennbar. Das macht deutlich, dass weitere Methoden für die technische Bewertung notwendig sind.

Ein Bild lässt sich als Superposition vieler Wellenfunktionen in den einzelnen Farbkanälen beschreiben. Um diese Effekte zu erkennen, nutzen wir das aus und betrachten die Frequenzen, welche im Bild vorkommen. Dafür zerlegen wir das Bild in einzelne Frequenzen und betrachten den Beitrag jeder Frequenz zum Gesamtbild. Hierzu wird die Diskrete-Kosinus-Transformation³⁰ verwendet. Die daraus erhaltene Verteilung wird über mehrere Bilder gemittelt. Dieses Vorgehen wird für die augmentierten Bilder wiederholt und anschließend die beiden entstandenen Durchschnittsverteilungen voneinander abgezogen. In der Differenz der gemittelten Spektren sind die Artefakte in den augmentierten Bildern deutlich durch überrepräsentierte Frequenzen (hellere Punkte) sichtbar. Auch wenn die Artefakte für den Menschen schwer wahrnehmbar sind,

³⁰ J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, "Leveraging Frequency Analysis for Deep Fake Image Recognition".

können diese das Training eines neuronalen Netzes beeinflussen. Es existieren aber auch schon weiterentwickelte Dekonvolution-Schichten, um diese Artefakte zu verhindern/minimieren³¹.

Eine weitere Methode zur automatisierten Bewertung von Augmentationen ist die Fréchet Inception Distance (FID)³². Hier werden mit Hilfe eines bereits trainierten Netzes die den generierten und realen Daten zugrundeliegenden Verteilungen verglichen.

Eine Bewertung der Farbanpassung bei einem Domänen-Transfer wird im HSV-Farbraum (Farbwert, Farbsättigung, Hellwert) durchgeführt. Hier werden die Farb-Histogramme über mehrere Bilder gemittelt und mit der Jensen-Shannon-Divergenz (JSD)³³ verglichen.

Außerdem eine Methode³⁴ zur Erkennung von augmentierten Bildern untersucht. Die Methode verwendet einen vortrainierten Diskriminator, welcher ein Bild als Eingabe erhält und ein Rating-Wert zwischen 0 und 1 ausgibt, wobei ein Wert von 1 bedeutet, dass ein generiertes Bild als Eingabe vermutet wird. Für die Untersuchung wurden mit der StarGANv2-Architektur³⁵ augmentierte Bilder erzeugt. Als Ausgangsdatensatz wurde der BDD100k-Datensatz³⁶ verwendet. Durch die entsprechenden Labels konnte der Datensatz in Tag- und Nacht-Bilder aufgeteilt werden. Anschließend wurde die StarGANv2-Architektur dann mit diesen beiden Klassen trainiert und die Augmentation in die jeweils andere Klasse ausgeführt (siehe Abbildung 40 und Abbildung 41).

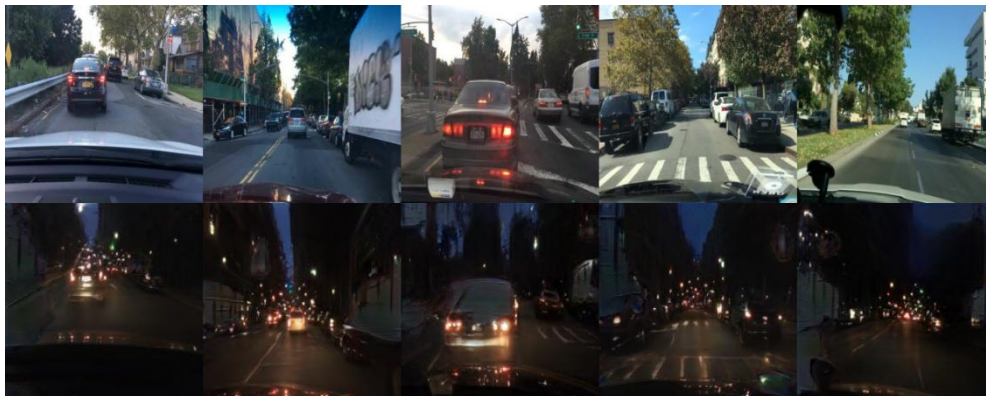


Abbildung 40: Augmentationsbeispiel Tag (oben, real) → Nacht (unten, Augmentation)

³¹ A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, "Checkerboard artifact free sub-pixel convolution," p. 16.

³² M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," arXiv:1706.08500 [cs, stat], Jan. 2018, Accessed: Feb. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1706.08500>

³³ J. Lin, "Divergence measures based on the Shannon entropy," IEEE Transactions on Information Theory, vol. 37, no. 1, pp. 145-151, Jan. 1991, doi: 10.1109/18.61115.

³⁴ Wang, Sheng-Yu, Oliver Wang, Richard Zhang, Andrew Owens, und Alexei A. Efros. „CNN-Generated Images Are Surprisingly Easy to Spot... for Now“. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8692-8701. Seattle, WA, USA: IEEE, 2020. <https://doi.org/10.1109/CVPR42600.2020.00872>.

³⁵ Choi, Yunjey, Youngjung Uh, Jaejun Yoo, und Jung-Woo Ha. „StarGAN v2: Diverse Image Synthesis for Multiple Domains“. ArXiv:1912.01865 [Cs], 26. April 2020

³⁶ Yu, Fisher, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, und Trevor Darrell. „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“. ArXiv:1805.04687 [Cs], 8. April 2020. <http://arxiv.org/abs/1805.04687>.



Abbildung 41: Augmentationsbeispiel Nacht (oben, real) → Tag (unten, Augmentation)

Als zusätzlicher Vergleichsdatensatz wurden mit der Simulationsumgebung Carla³⁷ weitere Bilder erzeugt. Bei der Untersuchung wurden die realen Bilder, bis auf wenige Ausreißer, zuverlässig als nicht generiert erkannt. Die generierten Bilder werden vom Netz in den meisten Fällen auch richtig bewertet. Das funktioniert im Falle der Nacht-Bilder besser als bei den Tag-Bildern.

Als weiterer Vergleich wurde das Rating von simulierten Bildern untersucht. Interessanterweise klassifiziert das Netz die simulierten Bilder weder eindeutig als reale, noch als generierte Bilder. Eine Eignung zum Filtern von visuell besseren Augmentationen konnte nicht festgestellt werden. Hierfür wurden einige generierte Beispiele inkl. ihres Ratingwerts betrachtet (siehe Abbildung 42). Fehler! Verweisquelle konnte nicht gefunden werden.).



Abbildung 42: Generierte Bilder (Augmentation: Nacht→ Tag) inkl. Rating des jeweiligen Bildes

Die untersuchte Methode zur Erkennung von augmentierten Bildern ist in der Lage, mit dem StarGAN2 generierte Beispiele zu erkennen und von realen Beispielen zu unterscheiden. Eine Verallgemeinerung auf alle synthetischen Bilder ist aber nicht möglich, das zeigt das Experiment mit simulierten Beispielen.

Die Ergebnisse aus diesem Arbeitspaket wurden in folgenden Publikationen vorgestellt:

- Rigoll et al. „Parameterizable LiDAR-Assisted Traffic Sign Placement for the Augmentation of Driving Situations with CycleGAN“, International Conference On Systems Engineering, Dezember 2021.
- Rigoll, P., P. Petersen, L. Ries, J. Langner, und E. Sax. „Augmentation von Kameradaten mit Generative Adversarial Networks (GANs) zur Absicherung automatisierter Fahrfunktionen“. In

³⁷ Dosovitskiy, Alexey. „CARLA: An Open Urban Driving Simulator“, o. J., 16.



Fahrerassistenzsysteme und automatisiertes Fahren, herausgegeben von VDI Wissensforum GmbH, 41-48. VDI Verlag, 2022. <https://doi.org/10.51202/9783181023945-41>.

3.1.11 AP3.5 Kontextidentifizierung

E3.5.1 Definition von Kontextinformationen und deren formale Datenstrukturen (E3.5.1, E3.5.2)

In AP3.5 wurde in Kooperation mit den anderen Partnern in diesem Arbeitspaket eine Liste von Kontexten definiert, die im Gesamtprojekt von Interesse sind. Dabei wurde besonders darauf geachtet, dass die Kontexte für den weiteren Projektverlauf relevant sind und es auch eine realistische Aussicht darauf gibt, dass die jeweiligen Kontexte auch von einem der Partner identifiziert bzw. bereitgestellt werden können.

In einem weiteren Ergebnis wurde, wiederum in Kooperation mit den Partnern, spezifiziert wie die Kontexte konkret abgespeichert werden können. Hierbei wurde sichergestellt, dass die Kontexte möglichst verständlich, aber dennoch einfach maschinenlesbar und damit durchsuchbar abgelegt werden. Außerdem wurde in diesem Zuge auch abgestimmt, welcher Partner, welche Kontexte identifiziert bzw. liefert.

Spezifikation, Generierung, Analyse und Auswertung von Kontextinformationen (E3.5.3, E3.5.4, E3.5.5)

Es wurden LiDAR- und Kamera-Daten kombiniert, um die Bebauung zu klassifizieren. Hierzu war es zunächst nötig, State-of-the-Art neuronale Netze für die Semantische Segmentierung zu identifizieren. Durch Projektion der LiDAR-Daten in die Kamerabildebene, war es anschließend möglich einzelnen LiDAR-Punkten eine Objektklasse zuzuordnen. Zunächst werden durch eine semantische Segmentierung die Objektklassen der sichtbaren Objekte im Kamerabild bestimmt. Anschließend kann der Abstand zu den erkannten Objekten durch die Projektion der LiDAR-Punkte ins Kamerabild berechnet werden. Durch geometrische Berechnungen ist es anschließend möglich, der Pixelhöhe eines Objekts auf dem Bild eine Höhe in der Realität zuzuschreiben. Mit dieser Methode lässt sich die Höhe von Gebäuden aber zum Beispiel auch von Vegetation bestimmen.

Außerdem wurde eine Methode zur Abstandsbestimmung aus Kamera- und LiDAR-Daten konzeptioniert und prototypisch implementiert. Die Methode erlaubt die Abstandsbestimmung des Ego-Fahrzeugs von einem anderen Objekt. Konkret wird der Abstand zu einem Fahrzeug gemessen, das sich in Fahrtrichtung vor dem Ego-Fahrzeug befindet. Für das Verfahren ist eine semantische Segmentierung des Kamerabilds notwendig. Da die notwendigen Daten und die semantische Segmentierung im Projekt nicht zur Verfügung stehen, wurde mit dem öffentlich verfügbaren A2D2-Datensatz **Fehler! Textmarke nicht definiert.** gearbeitet. Für die semantische Segmentierung kann auch ein neuronales Netzwerk, wie beispielsweise das DeepLabv3³⁸ trainiert werden. Durch die Kombination von semantisch segmentierten Kamerabildern, LiDAR-Daten und einer Reihe von Filtern wird nach einem vorausfahrenden Fahrzeug gesucht (siehe Abbildung 43 und Abbildung 44). Die LiDAR-Daten liefern die notwendige dreidimensionale Information. Aus der Kombination dieser Daten wird der Abstand zum Vorderfahrzeug berechnet.

³⁸ L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," arXiv:1706.05587 [cs], Dec. 2017, Accessed: Jan. 12, 2022.

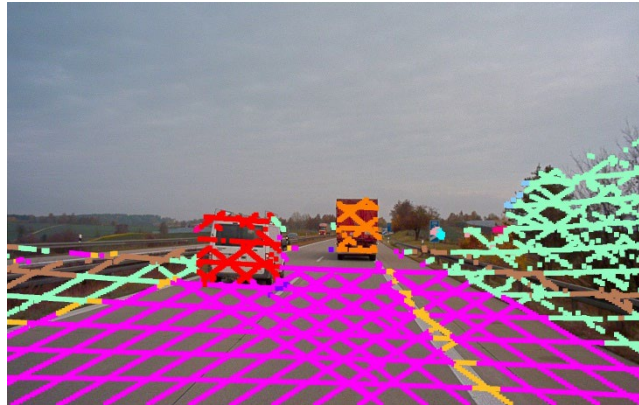


Abbildung 43: Semantisch annotierte ins Kamerabild projizierte LiDAR-Punkte zur Abstandsmessung

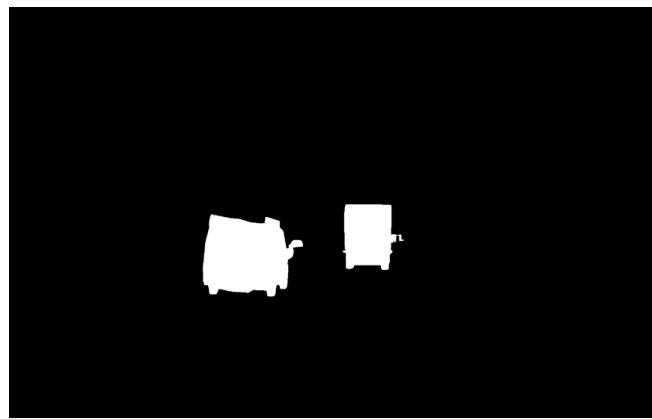


Abbildung 44: Teilschritt der Methode bei der die relevanten Fahrzeuge (weiß) im Kamerabild identifiziert wurden

Um einen Überblick über die gesammelten Daten zu bekommen und gegebenenfalls Lücken in den Daten zu identifizieren, werden die Daten auf Basis der Metadaten/Kontexten analysiert. Da im Projekt zum Berichtszeitraum erst eine sehr kleine Zahl an Daten zur Verfügung stand, wird das Vorgehen im Folgenden an den A2D2-Daten³⁹ demonstriert.

Die Methode basiert dabei auf einer sehr kleinen initiale Metadaten-Basis: Zeitstempel, GPS-Position und Fahrtrichtung. Darauf setzt dann die Methode zur Sonnenpositionsbestimmung auf, die auch im Rahmen dieses Arbeitspakets entstanden ist.

Im Folgenden, wird nur ein kleiner Einblick in mögliche Analysen gegeben. Als Maß für die Abdeckung wurde hier die gefahrene Strecke bzw. die Anzahl der Datenpunkte betrachtet. Zunächst wird die Abdeckung hinsichtlich der Jahreszeiten untersucht (siehe Abbildung 45).

³⁹ Geyer et al., “A2D2: Audi Autonomous Driving Dataset,” arXiv:2004.06320 [cs, eess], Apr. 2020.

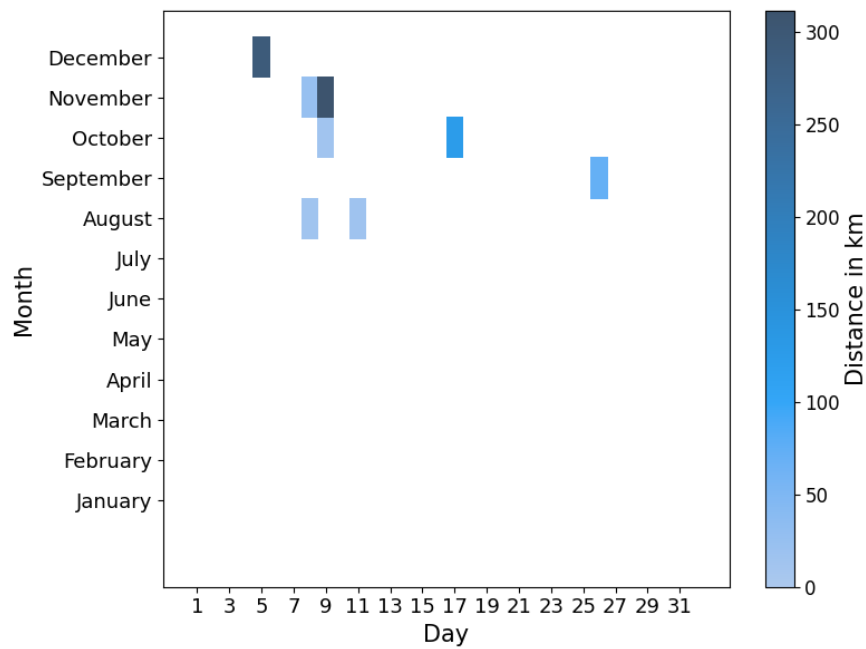


Abbildung 45: Tage des Jahres an denen aufgezeichnet wurde

Aber auch die Abdeckung hinsichtlich der geographischen Abdeckung ist möglich (siehe Abbildung 46).

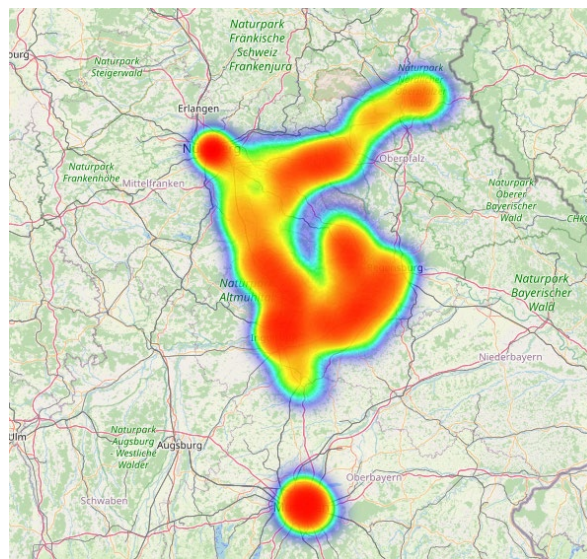


Abbildung 46: Geographischer Überblick über die Datenaufzeichnungen

Werden diese Basis-Metadaten mit den entwickelten Methoden erweitert, werden auch komplexere Analysen umsetzbar. Durch die Analyse der relativen Sonnenposition können Lücken in den Daten gefunden werden (siehe Abbildung 47).

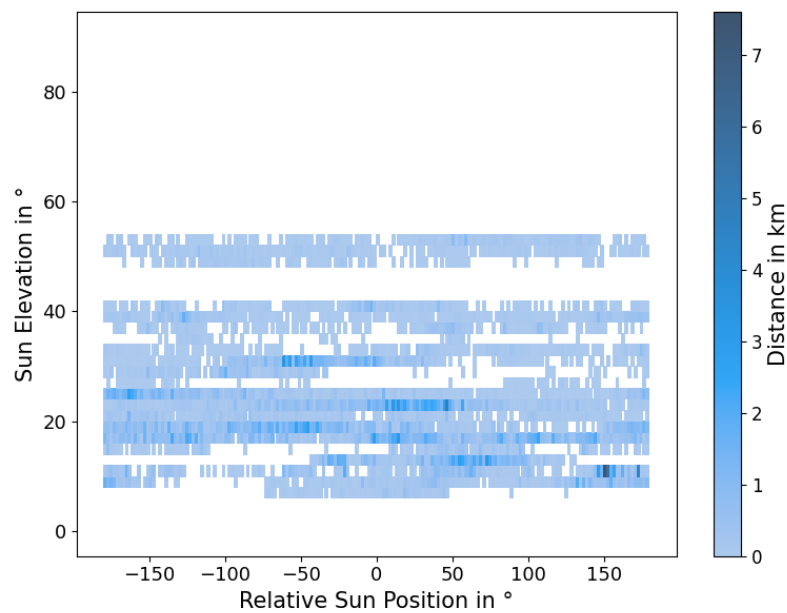


Abbildung 47: Relative Sonnenposition zum Fahrzeug

Anknüpfend an die allgemeine Analyse von Kontexten und die Detektion von besonderen Sonnenpositionen wurde die weitere Verknüpfung mit geographischen Informationen untersucht. Die dabei erarbeiteten Methoden wurden veröffentlicht⁴⁰. Ziel war es, aus geographischen Daten zu ermitteln, welche Objekte in einem Bild sichtbar sind. Hierfür wurden die Fahrzeugposition und die Fahrtrichtung als Ausgangsdaten verwendet. Basierend hierauf wurden Methoden entwickelt, um festzustellen, ob ein geographisches Objekt sichtbar ist. Neben annähernd punktförmigen Objekten, wie beispielsweise Ampeln (siehe Abbildung 48), wurden auch ausgedehnte Objekte, wie Brücken, betrachtet.

⁴⁰ P. Rigoll, L. Ries, und E. Sax, „Scalable Data Set Distillation for the Development of Automated Driving Functions“, in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, Okt. 2022, S. 3139-3145. doi: 10.1109/ITSC55140.2022.9921868.

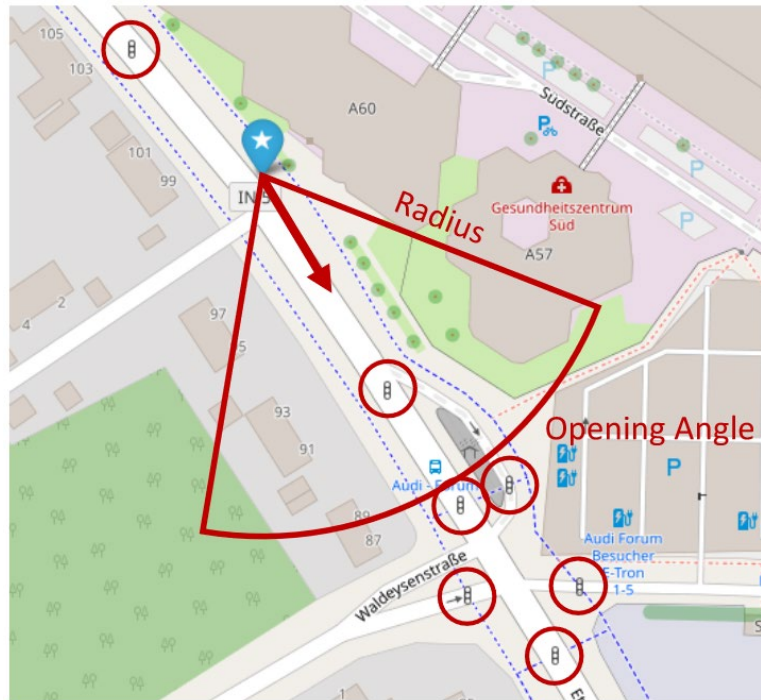


Abbildung 48: Schema zur Detektion von punktförmigen Objekten in geographischen Openstreetmap-Daten⁴⁰

Für die Untersuchung wurde der A2D2-Datensatz und Openstreetmap-Daten verwendet. In Abbildung 49 sind einige Beispiele dargestellt.



Abbildung 49: Beispiele für sichtbare geographische Openstreetmap-Objekte im A2D2-Datensatz⁴⁰

Statt einer regelbasierten Kontext-Analyse wurde zusätzlich ein generischerer Ansatz verfolgt. Die Information wird hierbei nicht aus anderen Kontexten extrahiert, sondern stammt aus den Kamerabildern. Hierzu wird ein State-of-the-Art neuronales Netz (CLIP⁴¹) verwendet, das Bilder und Texte als Vektoren repräsentiert. Die Vektoren teilen sich hierbei einen gemeinsamen Vektorraum. Ähnlichen Bildern und Texten werden ähnliche Vektoren zugeordnet.

Das erlaubt die Generierung bzw. die Suche nach diversesten Kontexten. In Abbildung 50 ist die Gesamtmethode skizziert. Abhängig davon was man über den entsprechenden Kontext weiß, kann man die Suche unterschiedlich gestalten. Wir haben das für eine textuelle Beschreibung, eine Beispielbild und einen Bildausschnitt untersucht⁴².

⁴¹ A. Radford u. a., „Learning Transferable Visual Models From Natural Language Supervision“, S. 16, 2021.

⁴² Preprint: P. Rigoll, P. Petersen, H. Stage, L. Ries, und E. Sax, „Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain“. arXiv, 21. April 2023. [Online]. Verfügbar unter: <http://arxiv.org/abs/2304.10247>

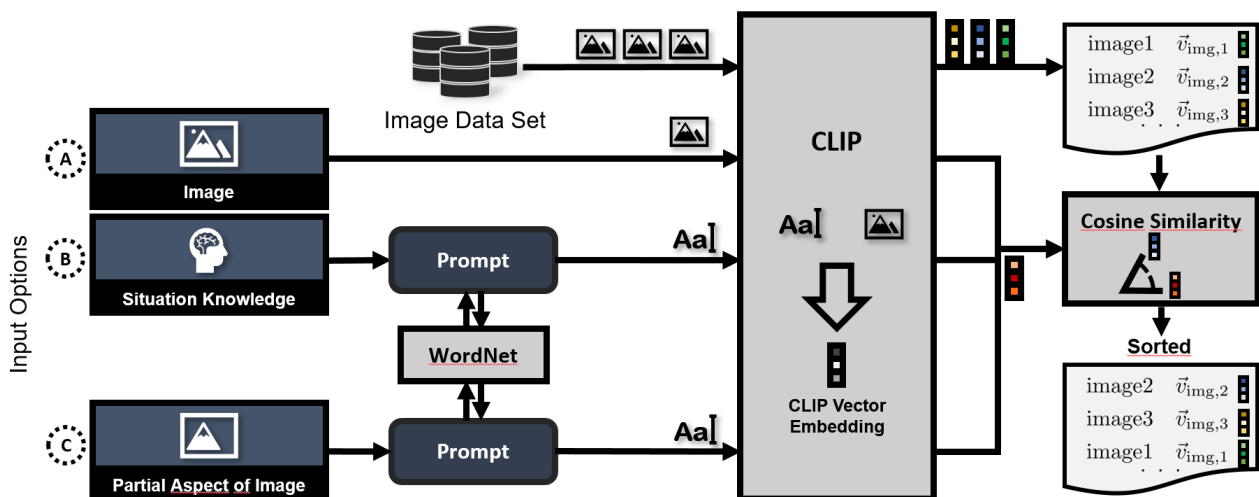


Abbildung 50: Methode zur Suche nach Kontexten⁴²

Mit der Methode kann zwar weder eine Vollständigkeit erreicht, noch False-Positives ausgeschlossen werden, dennoch ist es ein generisches Tool, das ohne manuellen Aufwand weitere Kontexte zugänglich macht. Ein Beispiel aus unseren Experimenten fanden wir überbelichtete Bilder, die wir mit dieser Methode im BDD100k-Datensatz⁴³ ausfindig gemacht haben.

Die Analyse der Kontextdateien aus dem Projekt, erlaubt eine Beurteilung der Abdeckung der Projektdaten. Betrachtet werden die Kontextdaten, die durch die Arbeiten in AP3.5 angereichert wurden. Das Ergebnis der Analyse ist ein zusammenfassender Bericht mit verschiedenen Visualisierungen der Kontextdaten. Außerdem werden auch Übersichten über die gemachten Kameraaufnahmen bei verschiedenen Kontexten generiert.

Die Ergebnisse aus diesem Arbeitspaket wurden in folgenden Publikationen vorgestellt:

- Rigoll, Philipp, Lennart Ries, und Eric Sax. „Scalable Data Set Distillation for the Development of Automated Driving Functions“. In 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 3139-45. Macau, China: IEEE, 2022. <https://doi.org/10.1109/ITSC55140.2022.9921868>.
- P. Rigoll, P. Petersen, H. Stage, L. Ries, und E. Sax, „Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain“. arXiv, 21. April 2023.

3.1.12 AP4.2 Schnittstellen und Beschreibungssprache

In Arbeitspaket 4.2 wurden gemeinsame Schnittstellen und Beschreibungssprachen mit dem Konsortium diskutiert und für das Projekt festgelegt. Die gemeinsamen Arbeitsergebnisse sind im gemeinsamen Schlussbericht aufgeführt.

⁴³ Yu, Fisher, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, und Trevor Darrell. „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“. ArXiv:1805.04687 [Cs], 8. April 2020. <http://arxiv.org/abs/1805.04687>.



3.1.13 AP4.3 Methodenbaukasten und Toolchain

Tools zur Unterstützung der Sensorkalibrierung und Übertragung von Labels (E4.3.2)

Um Labels zwischen Intermodaler und Multimodaler Sensorik übertragen werden können müssen alle Sensoren sowohl intrinsisch als auch extrinsisch zueinander und dem Fahrzeugkoordinatensystem kalibriert werden. Dies ist oft ein manueller Prozess, welcher viel Zeit in Anspruch nimmt. Daher wurde in AP4.3 ein Toolkit entwickelt, welches viele Schritte des Kalibrierungsprozesses optimiert und die benötigte Menschliche Interaktion minimiert. Der Kalibrierungsprozess wird in mehreren Nodes aufgeteilt. Das Kalibrierungstoolkit bedingt, dass das Framework ROS verwendet wurde um LiDAR und Kamera Rohdaten aufzuzeichnen. Die Kalibrierungs Pipeline erwartet, dass ein LiDAR Sensorkoordinatensystem als Fahrzeugkoordinatensystem als Basiskoordinatensystem definiert wird auf welches alle anderen Sensoren kalibriert werden. Zuerst werden mit einer Node LiDAR Sensoren zu dem Basiskoordinatensystem kalibriert. Daraufhin werden Kameras zu den Punktwolken kalibriert.

LiDAR Kalibrierung: In dieser Node, werden LiDAR Sensoren zum Fahrzeugkoordinatensystem kalibriert. Dafür werden Trainingsdaten benötigt, welche gematchte LiDAR Punkte in Sensor und Fahrzeugkoordinatensystem darstellen. Diese Trainingsdaten können mit einer GUI schnell assoziiert werden und durch die Robustheit der Methode müssen nur wenige Punkte manuell assoziiert werden. Mit drei verschiedenen Lossfunktionen lernt ein Adam Optimierer die Transformation der Sensoren zum Fahrzeugkoordinatensystem. Die Learningrate während des Trainings wird wie typisch für Adam decayed. Eine intrinsische Kalibrierung der LiDAR Sensoren ist nicht notwendig. Die erlernten Transformationen werden automatisch als URDF Datei maschinenleserlich gespeichert.

Kamerakalibrierung: In dieser Node werden Kameras zu einer (fusionierten) LiDAR Punktwolke kalibriert. Hierfür wird eine grobe initiale Kamerakalibrierung benötigt, wobei die intrinsische Kalibrierung vom Hersteller angegeben wird und die extrinsische aus dem 3D Modell des Fahrzeugs auslesbar ist. Es werden 3 Verzerrungsmodelle unterstützt: Fish-Eye, Plumb Bob und Lochkamera. Zwei verschiedene Lossfunktionen (Mean Average Error & Huber) werden verwendet um mit dem KI-Optimierungsverfahren (RMSProp) die Focal-Length, das Kamerazentrum, die Distortion Parameter und die Transformation bestehend aus Rotation und Translation aus den Trainingsdaten lernt. Dieses iterative Verfahren verwendet eine dynamische Learning Rate und erzeugt auch mit nur wenigen Trainingsdaten eine sehr gute Kalibrierung. Ein zusätzliches Tool im Toolkit ermöglicht eine schnelle Erzeugung der Trainingsdaten durch eine responsive GUI und die Unterstützung verschiedener Hotkeys und Color-Codes.

Weitere Tools im Toolkit erleichtern den Umgang mit aufgezeichneten Daten und synchronisieren intelligent die aufgezeichneten Daten der multimodalen Sensorik.

Das Toolkit wurde vom FZI verwendet um die Sensorik des Versuchsträgers CoCar NextGen zu kalibrieren. In diesem Versuchsträger sind 6x statische 4D LiDAR Sensoren, 6x 360° LiDAR Sensoren und 9x Kameras verbaut, wobei davon 2 Kameras ein Fischaugen Objektiv haben. Trotz der hohen Anzahl an verbauter Sensorik ist die Kalibrierung aller Sensoren mit nur minimalen Trainingsdatenerzeugung möglich und führt zu sehr guten Ergebnissen.

Im Kontext von AP4.3 und E4.3.2 befasst sich auch das FZI - neben dem Fokus auf Kalibrierung - mit der Erkennung und Generierung von Sondersituationen in vorhandenen Daten. Wir haben in der Publikation Perception Datasets for Anomaly Detection in Autonomous Driving: A Survey Datensätze mit Anomalien aufgelistet, und nur die CODA Datensatz-Familie verfügt über multimodale Daten von Kamera und Lidar Sensoren:

Dataset	Year	Sensors	Size (Test/Val)	Resolution	Anomaly Source	Temporal	#OOD Classes	Groundtruth
Fishyscapes [17], [31], [32]								
FS Lost and Found	2019	Camera	275 / 100	2048 × 1024	Recording	✗	1	Semantic Mask
FS Static	2019	Camera	1,000 / 30	2048 × 1024	Data Augmentation	✗	1	Semantic Mask
CAOS [33], [34]								
StreetHazards	2019	Camera	1,500	1280 × 720	Simulation	✓	1	Semantic Mask
BDD-Anomaly	2019	Camera	810	1280 × 720	Class Exclusion	✗	3	Semantic Mask
SegmentMelfYouCan [35], [36]								
RoadAnomaly21	2021	Camera	100 / 10	2048 × 1024	Web Sourcing	✗	1	Semantic Mask
RoadObstacle21	2021	Camera	327 (+55) / 30	1920 × 1080	Recording	✓	1	Semantic Mask
CODA [37], [38]								
CODA-KITTI	2022	Camera, Lidar	309	1242 × 1376	Void Classes	✗	6	Bounding Boxes
CODA-nuScenes	2022	Camera, Lidar	134	1600 × 900	Void Classes	✗	17	Bounding Boxes
CODA-ONCE	2022	Camera, Lidar	1,057	1920 × 1020	Automated OOD Proposal	✗	32	Bounding Boxes
CODA2022-ONCE	2022	Camera, Lidar	717	1355 × 720	Automated OOD Proposal	✗	29	Bounding Boxes
CODA2022-SODA10M	2022	Camera	4,167	1280 × 720	Automated OOD Proposal	✗	29	Bounding Boxes
Wuppertal OOD Tracking [22], [39], [40]								
Street Obstacle Sequences (SOS)	2022	Camera, Depth	1,129	1920 × 1080	Recording	✓	13	Instance Mask
CARLA-WildLife (CWL)	2022	Camera, Depth	1,210	1920 × 1080	Simulation	✓	18	Instance Mask
Misc								
Lost and Found [41], [42]	2016	Stereo Cameras	2,104	2048 × 1024	Recording	✓	42	Semantic Mask
WD-Pascal [43], [44]	2019	Camera	70	1920 × 1080	Data Augmentation	✗	1	Semantic Mask
Vistas-NP [45], [46]	2020	Camera	11,167	Varying	Class Exclusion	✗	4	Semantic Mask

Abbildung 51 Survey über Anomalie Datensätze

Der CODA Datensatz beinhaltet Groundtruth zu Anomalien in der Form von 2D Bounding Boxen für Kamera Bilder. Die Datensätze liefern teilweise dazugehörige Lidar Punktwolken, allerdings keine Groundtruth in diesen, weder in der Form von 3D BoundingBoxen, noch punktweise. Auch die 2D Bounding Boxen weisen nicht die höchste Genauigkeit auf, insbesondere in Hinsicht auf andere Datensätze, welche gängigerweise semantische Segmentierung für die Groundtruth nutzen.

Da der Fokus des AP auf der Übertragung von Labels zwischen verschiedenen Modalitäten liegt, wollen wir diesen Fokus fortführen und nutzen Label Transfer Techniken, um Sondersituationen in Daten zu detektieren. Anschließend können wir zu Validierungs- und Evaluationszwecken aus diesen Daten einen neuen Datensatz erzeugen, welcher neben den single-frame Anomalien auch den temporalen Kontext vor und nach der Detektion beinhaltet.

Das Ziel ist es also, 2D Labels der Anomalien auf 3D Labels in Punktwolken zu übertragen und gleichzeitig die Genauigkeit der Daten zu verbessern. Wir streben also punktweise Labels statt 3D Bounding Boxen an. Um zunächst die Sub-Punktwolke zu identifizieren, die zu der Anomalie innerhalb der 2D Bounding Box assoziiert ist, wenden wir ein Frustum an:

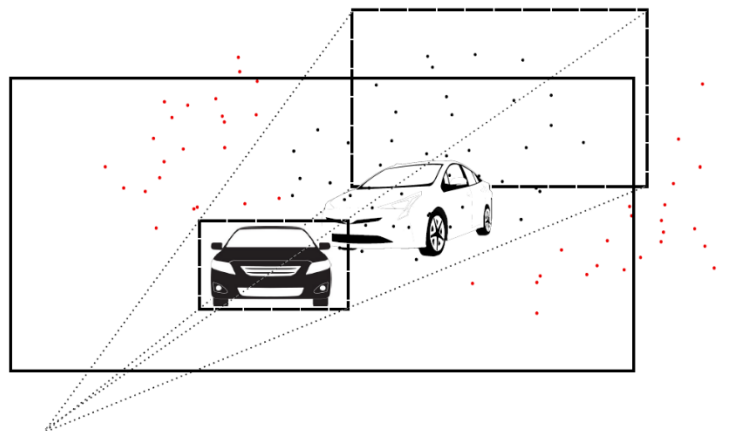


Abbildung 52 Schema des erzeugten Frustums

Um diese Punktwolke kann nun nicht einfach eine 3D Bounding Box gesetzt werden, da die fehlenden Tiefeninformationen aus dem Kamerabild dazu führen, dass nicht alle Punkte innerhalb des Frustums

relevant sind. Entsprechend nutzen wir das Zentrum der 2D Bounding Box, um DBSCAN und Mean Shift Clustering Algorithmen im 3D Raum auszuführen:

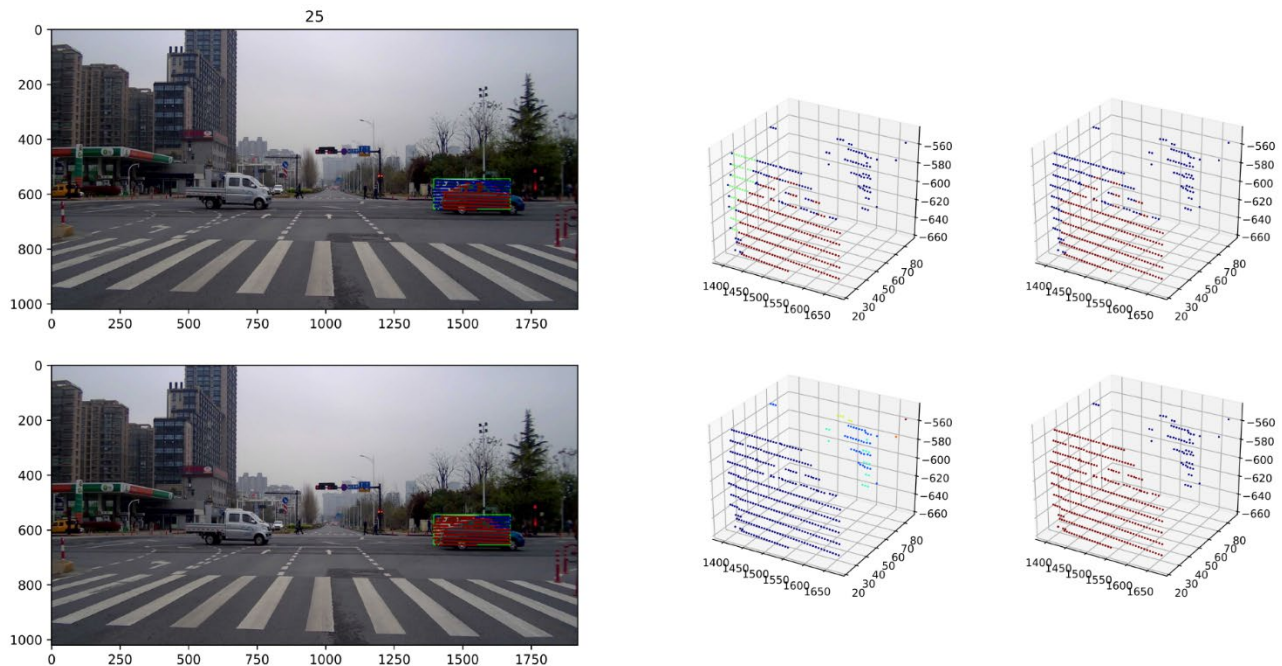


Abbildung 53 Resultate der DBSCAN und Mean Shift Verfahren

Diese Ansicht zeigt oben DBSCAN Resultate und unten Mean Shift Resultate. Die mittlere Grafik zeigt alle erkannten Cluster, während die linke und rechte die auf Basis des 2D Center Punkts ausgewählten Cluster zeigen. Dieses Tool ermöglicht es, Cluster-Resultate gegeneinander auszuspielen. Stimmen die Ergebnisse überein, liegt keine Sondersituation vor. Sind die ausgewählten Cluster - wie im Beispiel - jedoch stark unterschiedlich, wurde eine Sondersituation detektiert. Nun kann das umliegende Szenario aus dem ursprünglichen Datensatz extrahiert werden, um eine Datenbank von Sondersituationen mit temporalem Kontext zu erzeugen.

3.1.14 AP4.4 Evaluation und Demonstration

Evaluation von Verfahren und Methoden in der Gesamttoolkette zur Erzeugung von aufbereiteten Datensätzen (E4.4.1)

In AP4.4 werden die Methoden und Verfahren, welche im Rahmen des Projekts entwickelt werden, gesammelt und evaluiert werden. Da die Methoden und Verfahren im gesamten Projekt sehr divers sind und jeweils mit eigenen Metriken evaluiert werden, war es nötig, zum Einsammeln der Ergebnisse einen entsprechenden Rahmen zu schaffen. Hierfür wurde ein Evaluationsbogen entwickelt. Dieses Formular wurde so generisch gehalten, dass sich jedes Arbeitspaket darin wiederfindet. Gleichzeitig wurde darauf geachtet, dass eine Vergleichbarkeit der Ergebnisse, vor allem innerhalb eines Arbeitspakets, gewährleistet bleibt. Um sicherzustellen, dass diese Bedingungen auch tatsächlich erfüllt sind, wurde der Bogen mit den Arbeitspaketen des gesamten Projekts abgestimmt. Technisch wurde der Evaluationsbogen im projektweiten Online-Portal realisiert. Hier ist es durch Vorlagen und Stichwörter möglich, die von den Arbeitspaketen ausgefüllten Evaluationsbögen gesammelt darzustellen und es wurde am Ende ein schneller und sinnvoller Überblick gewährleistet.

Für die WG Model Zoo wurde ein Konzept für die Abgabe von Maschinenlern-Modellen entwickelt. Das Konzept umfasst eine Checkliste für die Abgabe und ein Beispiel GIT-Repository, das als Boilerplate verwendet werden kann.

Evaluation von Verfahren und Methoden in der Gesamtoolkette zur Erzeugung von aufbereiteten Datensätzen (E4.4.3)

Es wurde ein Demonstrator entwickelt, der zwei Methoden aus AP3.4 zeigt. Die Hardware des Demonstrators besteht aus einem Messeaufbau mit integriertem, berührungsempfindlichem Bildschirm (siehe Abbildung 54). Eine Methode ermöglicht die realistische Positionierung von neuen Verkehrsschildern in existierenden Fahraufzeichnungen (siehe Abbildung 55). Dieser Prozess wurde interaktiv aufbereitet, sodass ein Nutzer in Echtzeit neue Verkehrsschilder in einer Szene platzieren kann. Dabei lassen sich diverse Parameter für die Verkehrsschildpositionierung von einem Nutzer wählen. Als zweiter Showcase wurde aus AP3.4 die Augmentation von einzelnen Fahrzeugen identifiziert (siehe Abbildung 56). Der Nutzer kann in einer real aufgezeichneten Szene Fahrzeuge auswählen und mit einem generativen neuronalen Netz verändern. Das Ergebnis wird in annähernder Echtzeit visualisiert.



Abbildung 54: Hardware des Demonstrators bestehend aus einem Messeaufbau mit einem integrierten berührungsempfindlichen Bildschirm



Abbildung 55: Demonstrator zur realistischen Positionierung von neuen Verkehrsschildern in existierenden Fahraufzeichnungen

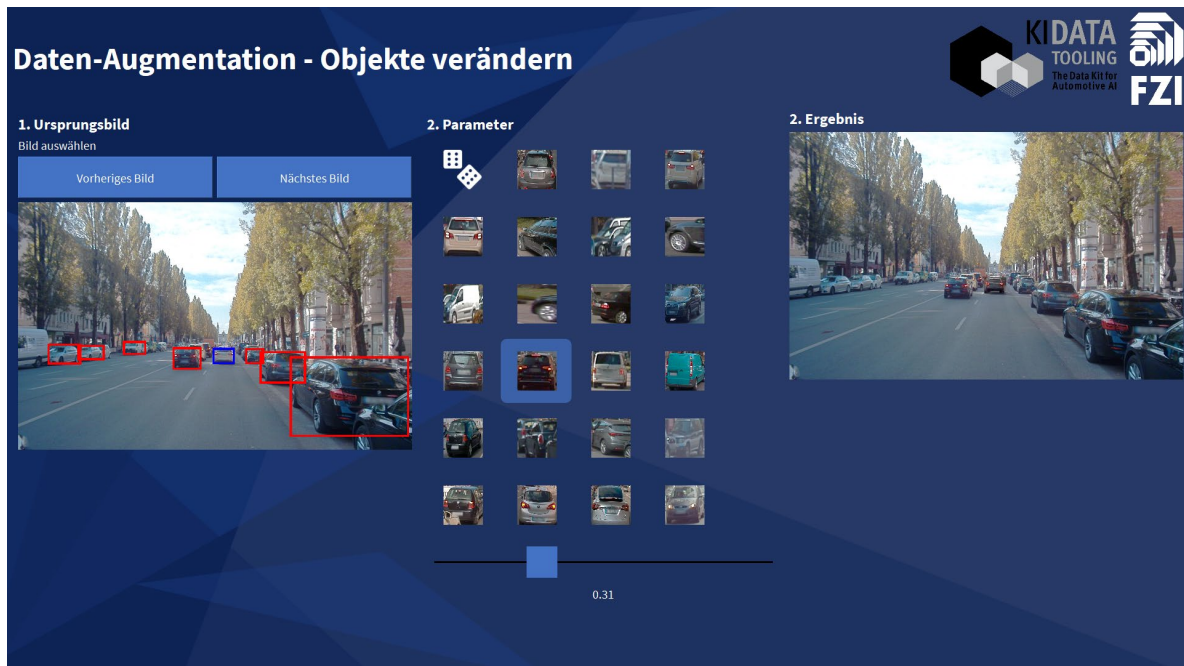


Abbildung 56: Demonstrator zur Augmentation einzelner Fahrzeuge



3.2 Zahlenmäßiger Nachweis

Der zahlenmäßige Nachweis wurde separat bereitgestellt.

3.3 Notwendigkeit der Förderung

Das Vorhaben KI Data Tooling war als Verbundprojekt konzipiert, welches durch Bereitstellung umfangreicher Eigenmittel der beteiligten Industriepartner sowie durch Mittel der öffentlichen Forschungsförderung des Bundes realisiert werden sollte.

Die Notwendigkeit der Zuwendung durch das BMWK ergab sich aus den ehrgeizigen und risikobehafteten angestrebten technischen und wissenschaftlichen Zielen. Nur im Verbund konnten die großen technologischen Herausforderungen angegangen und der notwendige technologische Hub erarbeitet werden. Das damit einhergegangene Risiko war von keinem einzelnen Partner zu stemmen. Nur in Kooperation von Wissenschaft und Industrie konnten die komplexen Fragestellungen zielführend erforscht werden.

Das Forschungsvorhaben wäre ohne finanzielle Förderung durch das BMWK nicht durchführbar gewesen.

3.4 Veröffentlichungen

3.4.1 Öffentlichkeitsarbeit am FZI

- Dauerhafte Ausstellung des entstandenen Videos und Poster innerhalb des Projekts KI Absicherung in den Laboren des FZI, insbesondere im Rahmen des FZI House of Living Labs
- Ausstellung und Demonstration der Projektergebnisse im Rahmen des FZI Open House 2023 sowie 2024

3.4.2 Betreute Abschlussarbeiten

- Querfurth, Lennart: Validierung von augmentierten Fußgängern im Bereich automatisiertes Fahren mit Methoden des Deep Learning, Bachelorarbeit, 2021
- Zhou, Junze: Bildbasierte Sonnenpositionsbestimmung mit ResNet, Bachelorarbeit, 2021
- Sun, Haotong: Methoden zur realistischen Verdeckung von künstlichen Fußgängern mit Hilfe von tiefen neuronalen Netzen, Bachelorarbeit, 2021
- Tillert, Wiebke H.: Konzeptionierung einer Pipeline zur Augmentation von Bildern aus dem Straßenverkehr mittels Maschinellen Lernens, Karlsruher Institut für Technologie, Bachelorarbeit, 2021
- Leonhardt, Artur: Merkmalsuntersuchung von mit Generative Adversarial Networks augmentierten Automotive-Bildern zur Beurteilung der Verwendung als Beispieldaten, Karlsruher Institut für Technologie, Bachelorarbeit, 2021
- Schober, Jan: Evaluation der Einsatzmöglichkeit von Augmentation mittels Generative Adversarial Networks beim Closed-Loop Integrationstest von hochautomatisierten Fahrfunktionen, Karlsruher Institut für Technologie, Masterarbeit, 2021
- Golks, Niklas: Objektbezogene Untersuchung eines Latent Space zur Gewinnung neuer Kontexte, Karlsruher Institut für Technologie, Masterarbeit, 2022
- Wilhelm, Marian: Automatisierte Farb- und Lichanpassung für Bildmontagen von Verkehrsszenen mit Hilfe Neuronaler Netze, Masterarbeit, 2022
- Dernst, Artur: Metrisches Lernen zur Detektion falscher Instanz-IDs in einem Datensatz für das autonome Fahren, Karlsruher Institut für Technologie, Bachelorarbeit, 2023



- Mai, Jingming: Towards High-Quality Datasets: A Monocular Approach for Detecting Mistakes in Labels and 3D Bounding Box Annotations, Karlsruher Institut für Technologie, Masterarbeit, 2023
- Skrobiranda, Lars: Selbstüberwachtes metrisches Lernen zur Erzeugung realitätsnaher zusammengesetzter Bilder von Straßenszenen, Karlsruher Institut für Technologie, Bachelorarbeit, 2023/2024
- Klatte, Marvin: Copy-Paste-Augmentation im Kontext des autonomen Fahrens, Karlsruher Institut für Technologie, Masterarbeit, 2023/2024
- Guneshka, Stefani: Ontology-based corner case scenario simulation for autonomous driving, Karlsruher Institut für Technologie, Bachelorarbeit, 2022
- Klaus, Simon: Anomaly Detection in the Latent Space of VAEs, Karlsruher Institut für Technologie, Bachelorarbeit, 2023
- Schilling, Marcus: Anomaly Detection in 3D Space for Autonomous Driving, Karlsruher Institut für Technologie, Masterarbeit, 2022

3.4.3 Erfolgte Veröffentlichungen

- Bogdoll et al.; “Description of Corner Cases in Automated Driving: Goals and Challenges”; ICCV Workshop „Embedded and Real-World Computer Vision in Autonomous Driving“
- Rigoll et al. „Parameterizable LiDAR-Assisted Traffic Sign Placement for the Augmentation of Driving Situations with CycleGAN“, International Conference On Systems Engineering, Dezember 2021.
- Rigoll, P., P. Petersen, L. Ries, J. Langner, und E. Sax. „Augmentation von Kameradaten mit Generative Adversarial Networks (GANs) zur Absicherung automatisierter Fahrfunktionen“. In *Fahrerassistenzsysteme und automatisiertes Fahren*, herausgegeben von VDI Wissensforum GmbH, 41-48. VDI Verlag, 2022. <https://doi.org/10.51202/9783181023945-41>.
- Rigoll, Philipp, Lennart Ries, und Eric Sax. „Scalable Data Set Distillation for the Development of Automated Driving Functions“. In 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 3139-45. Macau, China: IEEE, 2022. <https://doi.org/10.1109/ITSC55140.2022.9921868>.
- K. Rösch *et al.*, "Space, Time, and Interaction: A Taxonomy of Corner Cases in Trajectory Datasets for Automated Driving," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, Singapore, 2022, pp. 86-93, doi: 10.1109/SSCI51031.2022.10022241.
- Klemp, M., Rösch, K., Wagner, R., Quehl, J., & Lauer, M. (2023). LDFA: Latent Diffusion Face Anonymization for Self-driving Applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3198-3204).
- Reichert, H., Lang, L., Rösch, K., Bogdoll, D., Doll, K., Sick, B., & Zöllner, J. M. (2021, September). Towards Sensor Data Abstraction of Autonomous Vehicle Perception Systems. In *2021 IEEE International Smart Cities Conference (ISC2)* (pp. 1-4). IEEE
- Heidecker, F., Breitenstein, J., Rösch, K., Löhdefink, J., Bieshaar, M., Stiller, C. & Sick, B. (2021, July). An application-driven conceptualization of corner cases for perception in highly automated driving. In *2021 IEEE Intelligent Vehicles Symposium (IV)* (pp. 644-651). IEEE.
- D. Bogdoll *et al.*, “Anomaly Detection in Autonomous Driving: A Survey”, *Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, New Orleans, USA, 2022



- D. Bogdoll *et al.*, “One Ontology to Rule Them All: Corner Case Scenarios for Autonomous Driving”, European Conference on Computer Vision (ECCV) Workshop, Tel Aviv, Israel, 2022
- P. Rigoll, P. Petersen, H. Stage, L. Ries, und E. Sax, „Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain“. arXiv, 21. April 2023.
- D. Bogdoll *et al.*, „Conditioning Latent-Space Clusters for Real-World Anomaly Classification“, IEEE Symposium Series on Computational Intelligence (SSCI), 2023
- D. Bogdoll *et al.*, “Perception Datasets for Anomaly Detection in Autonomous Driving: A Survey”, IEEE Intelligent Vehicles Symposium (IV), 2023