



KIDELTA
LEARNING

Scalable AI for Automated Driving

Zusammenfassender Schlussbericht KI Delta Learning

Entwicklung von Methoden und Werkzeugen zur effizienten Erweiterung und Transformation vorhandener KI-Module autonomer Fahrzeuge auf die Herausforderungen neuer Domänen oder komplexerer Szenarien.

Version	1.0
Editor	Dr.-Ing. Amin Hosseini
Projektkoordination	Mercedes-Benz AG und ZF Friedrichshafen AG

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Dokumenteninformation

Autoren

Franz Andert | DLR
Marius Bachhofer | ZF Friedrichshafen AG
Viviane Benzin | Mercedes-Benz AG
Saqib Bukhari | ZF Friedrichshafen AG
Thies de Graaff | DLR
Hanno Gottschalk | University of Wuppertal
Domenik Helms | DLR
Amin Hosseini | Mercedes-Benz AG
Tobias Kalb | Porsche Engineering Group GmbH
Stefan Matthes | BMW Group
Jens Mehnert | Robert Bosch GmbH
Roshan Muthaiya | CMORE Automotive GmbH
Jörg P. Schäfer | DLR
Florian Piewak | Mercedes-Benz AG
Manuel Schwonberg | CARIAD SE
Martin Simon | Valeo Schalter und Sensoren GmbH
Antonia van Betteray | University of Wuppertal
Tobias Wagner | Valeo Schalter und Sensoren GmbH
Julian Wiederer | Mercedes-Benz AG
Sebastian Wirkert | BMW Group
Christian Witt | Valeo Schalter und Sensoren GmbH

Reviewer

Amin Hosseini | Mercedes-Benz AG
Marius Bachhofer | ZF Friedrichshafen AG
Hanno Gottschalk | University of Wuppertal
Domenik Helms | DLR
Jens Mehnert | Robert Bosch GmbH
Christian Witt | Valeo Schalter und Sensoren GmbH

Kontakt

Dr.-Ing. Amin Hosseini

Function und Software Autonomous Driving - RD/ASF

Mercedes-Benz AG, Research & Development

Kolumbusstr. 19+21

71063 Sindelfingen

Germany

Email: projects@eict.de

Projektwebsite: www.ki-deltalearning.de

Revisionslog

Version	Datum	Kommentar	Autor	Partner
0.1	bis 05.07.2023	Input auf Confluence	s.o.	s.o.
0.2	06.-07.07.2023	Übertrag in Word und Layout	Nikos Papamichail	EICT
0.3	08.-12.07.2023	Finales Review	Amin Hosseini	Mercedes-Benz
1.0	27.07.2023	Einarbeitung Review und Finalisierung	Nikos Papamichail	EICT

Inhaltsverzeichnis

1 Einleitung	10
2 Kurzdarstellung	11
2.1 Gesamtzielsetzung	11
2.2 Die Ziele im Einzelnen	12
2.2.1 Trainingsdaten	13
2.2.2 Referenzmodelle	13
2.2.3 Methoden Delta Learning	13
2.2.4 Modelle Delta Learning	14
2.2.5 Bewertungen und Demonstratoren	14
2.3 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde	14
2.3.1 Bezug des Vorhabens zu den förderpolitischen Zielen	14
2.4 Planung und Ablauf des Vorhabens	16
3 Eingehende Darstellung	20
3.1 TP1: Datengewinnung	20
3.1.1 Ausgangssituation und Zielsetzung	20
3.1.2 Strukturierung	20
3.1.3 Zielerreichung	33
3.2 TP2: Transfer Learning	34
3.2.1 Ausgangssituation und Zielsetzung	34
3.2.2 Strukturierung	34
3.2.3 Zielerreichung	55
3.3 TP3: Didaktik	56
3.3.1 Ausgangssituation und Zielsetzung	56
3.3.2 Strukturierung	57
3.3.3 Zielerreichung	88
3.3.4 Referenzen	89
3.4 TP4: Automotive Tauglichkeit	91
3.4.1 Ausgangssituation und Zielsetzung	91
3.4.2 Strukturierung	92
3.4.3 Zielerreichung	110
3.4.4 Referenzen	111
3.5 TP5: Evaluation	112
3.5.1 Ausgangssituation und Zielsetzung	112
3.5.2 3.5.2 Strukturierung	112

3.5.3 Zielerreichung	130
3.6 TP6: Projektorganisation	130
3.6.1 Ausgangssituation und Zielsetzung	130
3.6.2 Strukturierung	131
3.6.3 Zielerreichung	144
4 Zusammenfassung	145
4.1 TP1 - Datengewinnung	145
4.2 TP2 - Transfer Learning	145
4.3 TP3 - Didaktik	146
4.4 TP4 - Automotive Tauglichkeit	148
5 Veröffentlichungen	150
5.1 Veröffentlichungen im Rahmen von Konferenzen	150
5.2 Veröffentlichungen in Journals, Buchprojekt	157

Abbildungsverzeichnis

Abbildung 1: Untersuchte Deltas im Projekt Delta-Learning	11
Abbildung 2: Zusammenhang Kernarbeitsziele im Projekt	13
Abbildung 3: Realisierter Projektplan inklusive der Änderungen	18
Abbildung 4: Web-Ansicht der Plattform für integriertes Datenmanagement und -verarbeitung	22
Abbildung 5: Ansicht eines Datensatzes mit Hervorhebung der Reiter für Metadaten	23
Abbildung 6: Am Fahrzeug angebrachte Sensoren.	25
Abbildung 7: Zwei Aufnahmefahrzeuge von KI Delta Learning und vom DLR am Testfeld Niedersachsen	26
Abbildung 8: Beispielhafte Semantische Segmentierung der Referenz-Kamera	27
Abbildung 9: Semantisch annotierte Punktwolke als Resultat des entwickelten Lidar-Modells	29
Abbildung 10: Erzeugte Ansicht aus der 3D-Rekonstruktion einer synthetischen Szene	30
Abbildung 11: Beispiele für erzeugte synthetische Daten mit schwierigen Wetterbedingungen	31
Abbildung 12: Verschieden Darstellungen eines Motion Capture Bildes	32
Abbildung 13: Layer Stitching mit und ohne Stitching Layer zum Messen von Änderungen im Neuronalen Netz	36
Abbildung 14: Im Projekt erarbeitete Klassifikation von Domänenadaptionsverfahren.	37
Abbildung 15: Illustration einiger Beispielbilder	39
Abbildung 16: Übersicht über die Rare-Event-Simulation-basierte Fehleridentifizierung	42
Abbildung 17: Übersicht über die Verwendung eines Knowledge Graphen zur Optimierung einer Verkehrszeichenerkennung.	51
Abbildung 18: Übersicht über die Verwendung eines Gradient Reversal Layers zur Optimierung des Performance bei Schlechtwettersituationen.	52
Abbildung 19: Übersicht über einer verbesserte Version von Adaptive Instance Normalization	53
Abbildung 20: Übersicht über Verschiedene Methoden der Domain Anpassung	54
Abbildung 21: Übersicht über das Konzept zur Domain Gap Quantifizierung.	55
Abbildung 22: Überblick über die Themenbereiche und Fragestellungen des Teilprojekt 3: Didaktik!	58
Abbildung 23: Der iterative Trainingsprozess von EasyAdapt zur unüberwachten Domänenanpassung	59
Abbildung 24: Überblick über das Framework zum unüberwachten Einlernen unbekannter Objekte für die semantische Segmentierung	60
Abbildung 25: Single-Frame unüberwachte 3D Posenschätzung	61
Abbildung 26: NN Bausteine	64
Abbildung 27: Original Kitti Bild (links) und beispielhafte Diffusionsstufen	65
Abbildung 28: Grobe Architektur und Funktionsweise des PointPillar Netzwerkes	66

Abbildung 29: Darstellung der verschiedenen Augmentierungsmethoden, die für die Experimente verwendet wurden.	67
Abbildung 30: Darstellung der verschiedenen Funktionen, mit denen der Schwierigkeits-Hyperparameter um Laufe des Trainings moduliert wurde.	67
Abbildung 31: Interaktive Datenaufzeichnung	68
Abbildung 32: Live-Anzeige von Corner Cases während der Aufzeichnung von Motion Capture Daten	69
Abbildung 33: Human Pose Estimation (rechts: falsche Erkennung bei Anomalien)	69
Abbildung 34: Implementierung des Monte-Carlo-Dropout in DeepLab	72
Abbildung 35: Aufeinanderfolgende Bilder aus den Cityscapes- und A2D2-Datensätzen	73
Abbildung 36: Konzept der Domain-Adaptation-Methode: Komponenten und Trainingsaufbau.	74
Abbildung 37: Konzept der Domain-Adaptation-Methode: Zyklus des aktiven Lernens für den geleiteten Generator	74
Abbildung 38: Visueller Vergleich der gesteigerten Performance	75
Abbildung 39: Schema der Generierung unserer semi-synthetischen Datensätze	75
Abbildung 40: Übersicht über die Architektur des Cycle-Consistent-World-Models	78
Abbildung 41: Beispiel der visuellen Randomisierung	79
Abbildung 42: Human Pose Estimation	80
Abbildung 43: Überlagerung der Segmentierung	81
Abbildung 44: Pipeline für das Entfernen von Duplikaten eines Single-Shot-Detektors	82
Abbildung 45: Globale Cross-Sensor-Attention mit Spatial Awareness	83
Abbildung 46: Überblick über CRAT-Pred.	84
Abbildung 47: Blockdiagramm des Aufbaus des Multitask-CenterNet	86
Abbildung 48: Erzielte Ergebnisse für Netze mit Detektions- und/oder Segmentierungsprädiktionen.	87
Abbildung 49: Multitask-Netzwerk trainiert mit Pseudolabels verschiedener Spezialisten Netzwerke.	87
Abbildung 50: Iterative Erstellung größerer Graphen bis zur vollständigen Segmentierung	88
Abbildung 51: Taxonomie von Corner Cases aus der Perspektive von Interpretationsproblemen eines neuronalen Netzes	93
Abbildung 52: Inferenz durch den Graph-basierten Auto-Encoder und die Kernel Density Estimation	95
Abbildung 53: Ergebnisse der Anomalie-Erkennung	96
Abbildung 54: Architektur des entwickelten "Robust MVD Baseline Model" Verfahrens	96
Abbildung 55: Übersicht des entwickelten Verfahrens "SF2SE3"	98
Abbildung 56: Beispielhafte Architektur eines Capsule Networks für Hierarchical Novelty Detection	99

Abbildung 57: Auswertung der Hardwareanforderungen	100
Abbildung 58: Kategorien der Übersichtstabelle geeigneter Zielhardwareplattformen.	100
Abbildung 59: Schematische Übersicht der automatischen Hardware-Charakterisierung	101
Abbildung 60: Vorhersage der zu erwartenden Hardware-Ausführungszeiten	101
Abbildung 61: Ergebnis der Benchmarks ausgeführt auf dem Neural Compute Stick 2	102
Abbildung 62: Ergebnis der Benchmarks ausgeführt auf der integrierten Intel UHD 630 GPU	102
Abbildung 63: Hyperparameter Exploration für ein GA basiertes Channel Pruning auf DeepLabV3+	104
Abbildung 64: HW-bewusste Kompression von DeepLabV3+ mithilfe von GA basiertem Channel Pruning	104
Abbildung 65: Performanz Vergleich des Single-Shot-Detektor Netzwerks	105
Abbildung 66: Ergebnisse des Interpretable Pruning nach der Optimierung im Vergleich zu HRank Algorithmus.	106
Abbildung 67: Accuracy auf den Validierungsdaten über das Training hinweg	107
Abbildung 68: Darstellung der Tucker-Zerlegung	108
Abbildung 69: Analyse des generischen KI-Benchmarks	109
Abbildung 70: Dieselbe Visualisierung wie in Abbildung 69, aber für den AlexNet-Benchmark.	109
Abbildung 71: Beispielhafte Heatmap für AP1.3 für das zweite Projektinkrement	115
Abbildung 72: Gruppenbild vom dritten Review and Refinement Treffen	116
Abbildung 73: Logos der KI Familie und der vier Einzelprojekte	136
Abbildung 74: Projektlogos mit Hintergrundbildern	136
Abbildung 75: Startansicht der KI Delta Learning-Webseite	138
Abbildung 76: Gruppenfoto vom Kick Off 22/23.01.2020	139
Abbildung 77: Zwei Screenshots aus der virtuellen Zwischenpräsentation	140
Abbildung 78: Bilder von der Abschlussveranstaltung	143

Tabellenverzeichnis

Tabelle 1: Beiträge des Vorhabens KI-Delta Learning zu den Forschungszielen des BMWK-Fachprogramms.....	15
Tabelle 2: Übersicht über die Arbeiten, die z.B. in Projektinkrement I behandelt wurden	48
Tabelle 3: Gegenüberstellung der geplanten Ergebnisse zur Zielerreichung für AP 2.4 und AP 2.5.	55
Tabelle 4: Ergebnisse der Analyse bzgl. Robustheit von aktuellen Tiefenschätzungsverfahren ...	97
Tabelle 5: Ausführungszeiten auf verschiedenen Hardwareplattformen vor und nach der Kompression.....	110

1 Einleitung

Ziel des Vorhabens KI Delta Learning war die Entwicklung von Methoden und Werkzeugen zur effizienten Erweiterung und Transformation vorhandener KI-Module autonomer Fahrzeuge auf die Herausforderungen neuer Domänen oder komplexerer Szenarien. KI-Module sind der Kern der kognitiven Intelligenz automatisierter Fahrzeuge und somit eine Schlüsseltechnologie für immer höhere Automationsgrade von Assistenzsystemen bis hin zum autonomen Fahren. Daher sind KI-Module ein zentraler Schlüssel für die künftige Wertschöpfung der deutschen Automobilindustrie. Die Markteinführungsstrategie der deutschen Automobilindustrie für diese Assistenzsysteme erfolgt schrittweise hin zu immer höheren Automationsgraden und größeren Einsatzbereichen der Automation. Im Mittelpunkt des Projekts standen die schrittweisen Erweiterungen (nachfolgend „Deltas“ genannt) des Einsatzbereichs von Assistenzsystemen und den zugehörigen KI-Modulen, welche parallel in sechs Richtungen entsprechend unterschiedlicher Use-Case-Szenarien erfolgt sind.

So wurden Deltas durch verbesserte Sensorik, den verschiedenen Verkehrsräumen wie Autobahnen oder Baustellen, Fahrten durch verschiedene Länder, in unterschiedlichen Jahren mit neuen Verkehrsformen und Verkehrsteilnehmern wie E-Scootern, in wechselnden Umweltbedingungen wie Tag, Nacht, Sonne oder Regen sowie Deltas durch weiterentwickelte Netze betrachtet. Diese schrittweise, strukturierte Erweiterung von KI-Modulen in Richtung der sechs genannten Deltas ist das für das Projekt namensgebende „Delta Learning“. Dabei müssen nicht notwendigerweise alle sechs Erweiterungen gleichzeitig durchgeführt werden. Durch effizientes Delta Learning wird sich die Automation von Assistenzsystemen viel mehr schrittweise erhöhen. Delta Learning hatte zum Ziel, bereits für begrenzte Einsatzbereiche und -orte trainierte KI-Module inkrementell zu erweitern ohne den sonst üblichen Trainings- und Optimierungsvorgang mit sehr hohem Aufwand komplett neu durchzuführen. Bisher existierten keine hinreichend effizienten und stabilen Methoden und Werkzeuge für ein solches Delta-Learning.

Die grundlegende Erweiterung der aktuellen Generationen von KI-Algorithmen ermöglicht einen entscheidenden Sprung in Richtung der großflächigen Realisierung autonomer Fahrzeuge. Damit stellt KI Delta Learning einen wichtigen Innovationsbaustein für die Wettbewerbsfähigkeit der deutschen Automobil- und Zuliefererindustrie in einer immer kompetitiveren Ökonomie dar.

2 Kurzdarstellung

2.1 Gesamtzielsetzung

Die verfügbaren Funktionen der „künstlichen Intelligenz“ (KI) können die Herausforderungen eines autonomen Fahrzeuges nur in einem sehr eingeschränkten Zielraum und anhand ineffizienter Trainingsstrukturen bedienen. Trotz der anhaltenden Fortschritte im maschinellen Lernen, tendieren die verfügbaren Verfahren dazu, vorhandene Datensätze auswendig zu lernen und schon in leicht unterschiedlichen Szenarien unvorhersehbar zu reagieren.

Ziel des Vorhabens KI-Delta Learning war die Entwicklung von Methoden und Werkzeugen zur effizienten Erweiterung und Transformation vorhandener KI-Module autonomer Fahrzeuge auf die Herausforderungen neuer Domänen oder komplexerer Szenarien. Es wurden Verfahren entwickelt, die es erlauben, das bereits erworbene Wissen optimal zu nutzen und nur die zusätzlichen Anforderungen - das Delta - mit minimalem Aufwand neu zu lernen.

Dabei wurde eine doppelte Strategie verfolgt: Zum einen sollte auf neue Szenarien robuster reagiert werden können, zum anderen sollte durch effizientere Datengenerierung der Einsatzbereich der Fahrzeuge erweitert werden. Um solcherart dedizierte Lösungen für künstliche Intelligenz in uneingeschränktem Einsatz autonomer Fahrzeuge zu entwickeln, mussten Experten aus der Automobilindustrie mit anerkannten Experten aus der Wissenschaft zusammengebracht werden. Für das Vorhaben KI-Delta Learning konnte ein schlagkräftiges Konsortium mit starker technischer Expertise für dieses Feld zusammengestellt werden.

Das Projekt konzentrierte sich auf Deltas, die die Nutzbarkeit der aktuellen Generation von KI-Algorithmen stark einschränken: Für die Nutzung im Automobil ist es notwendig, die Deltas in der Robustheit der Algorithmen zu reduzieren und den Laufzeitanforderungen in eingebetteten Systemen gerecht zu werden. Weitere Deltas die überbrückt werden sollten, waren der Umgang mit verschiedenen Sensoriken, verschiedenen Bedingungen zwischen Ländern wie z.B. Verkehrsregeln oder dem Straßenbild, verschiedenen Umweltbedingungen als auch der noch nicht ausreichende Realitätsgrad synthetischer Daten Deltas. Auch das sehr ressourcenaufwändige Training und Labeling von neuronalen Netzen wurde als signifikante Einschränkung heutiger Algorithmen identifiziert.

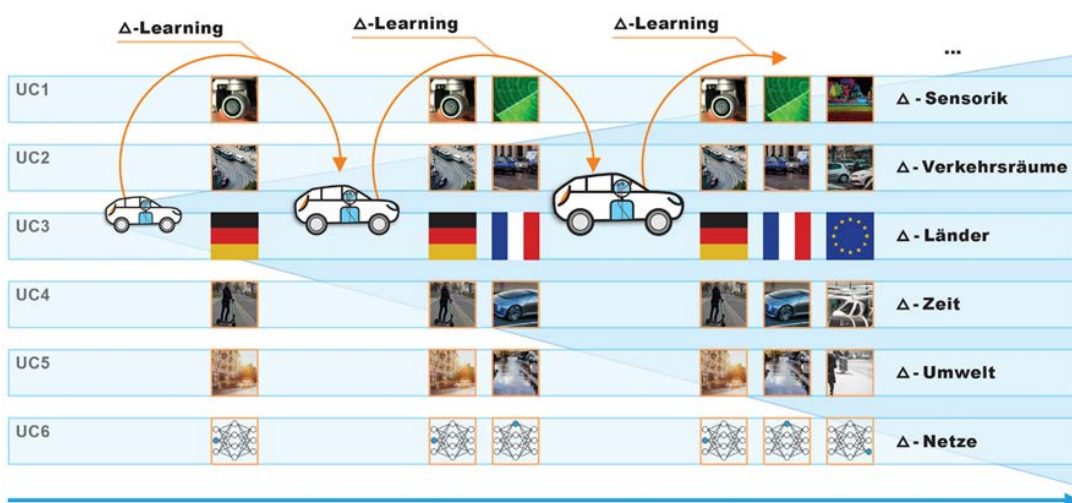


Abbildung 1: Untersuchte Deltas im Projekt Delta-Learning

Abbildung 2 macht diese Deltas und das Delta-Learning greifbarer. Die erste Zeile (Delta-Sensorik) symbolisiert, wie ein bestehendes KI-System mittels Delta-Learning durch zusätzliche Sensorik erweitert werden kann. Analog symbolisieren die Zeilen Delta-Verkehrsräume/Delta-Länder/Delta-Umwelt, wie ein KI-System, das für ein einen speziellen Verkehrsraum/Land/Umweltbedingungen konzipiert wurde, durch das Delta-Learning auf andere Verkehrsräume/Länder/Umweltbedingungen erweitert werden kann. Unter Delta-Zeit sind Phänomene, zu verstehen, die zur Zeit der Konzeption des Systems noch nicht vorhanden waren (wie z.B. Flugtaxi). Durch Delta-Learning soll ein bestehendes KI-System um solche neuen Phänomene erweitert werden können. Delta-Netze befasste sich mit der Zielsetzung bereits in Neuronalen Netzen modelliertes Wissen auf effizientere Neuronale Netze abzubilden.

Das Vorhaben KI-Delta Learning sollte den Stand der Technik in all diesen Bereichen so weit vorantreiben, dass die nächste Generation der KI-Algorithmen alle Deltas in einer Weise beherrscht, die sie für einen uneingeschränkten Einsatz im autonomen Fahrzeug qualifiziert. Die entstehenden Methoden und Prozesse können, wie in Abbildung 2 gezeigt, als iterative Werkzeuge verstanden werden, sodass KI-Algorithmen auf die jeweiligen Delta-Herausforderungen erweitert werden können. Über den entstehenden Werkzeugkasten hinaus zielte das Vorhaben auf einen Standard für die Bewertung der Wirksamkeit verschiedener Methoden der Delta-Überbrückung.

Eine große Herausforderung bestand darin, dass der aktuelle publizierte Stand der Wissenschaft sehr wahrscheinlich nicht ausgereicht hat, um dieses Ziel zu erreichen. Es war nicht ausreichend nur den Reifegrad bekannter Verfahren zu erhöhen, sondern es waren neue Techniken zu entwickeln, die erst die entsprechenden Möglichkeiten schaffen werden. Dies war insbesondere Aufgabe der Wissenschaftspartner, die parallel mehrere verschiedene innovative Konzepte entwickelt, eruiert und gegeneinander vergleichen haben, so dass am Ende des Projekts stärkere Methoden zur Verfügung stehen, welche die Konkurrenzfähigkeit der Projektpartner mit internationalen Wettbewerbern, die in der grundlegenden KI-Forschung sehr stark von Synergien aus anderen Geschäftsmodellen profitieren können, zumindest erhält.

Die grundlegende Erweiterung der aktuellen Generationen von KI-Algorithmen ermöglicht einen entscheidenden Sprung in Richtung der großflächigen Realisierung autonomer Fahrzeuge. Damit stellt KI-Delta Learning einen wichtigen Innovationsbaustein für die Wettbewerbsfähigkeit der deutschen Automobil- und Zuliefererindustrie in einer immer kompetitiveren Ökonomie dar.

2.2 Die Ziele im Einzelnen

Die formulierte Gesamtzielsetzung des Vorhabens KI-Delta Learning lässt sich in fünf wissenschaftliche und technische Kernarbeitsziele herunterbrechen.

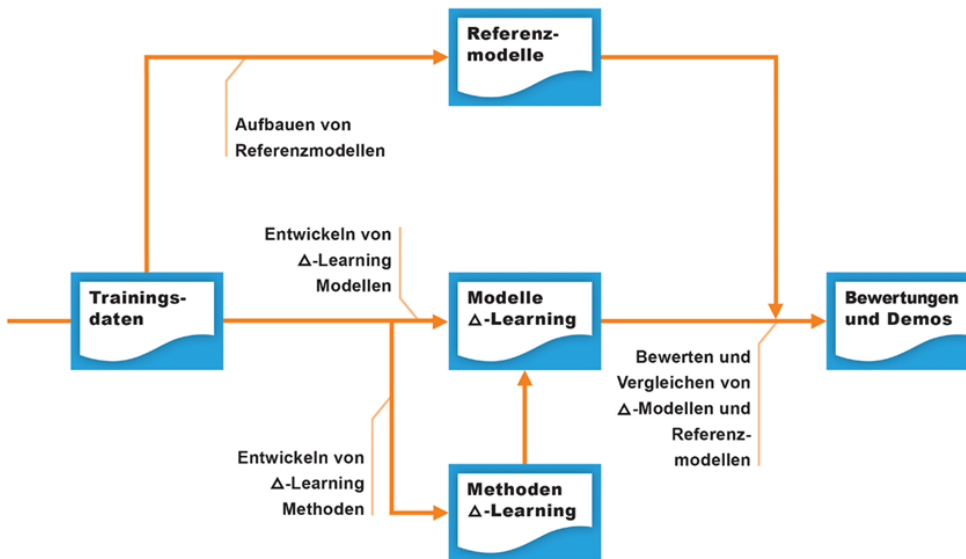


Abbildung 2: Zusammenhang Kernarbeitsziele im Projekt

2.2.1 Trainingsdaten

Das erste grundlegende Arbeitsziel von KI-Delta Learning war der Aufbau eines Referenzdatensatzes. Dieser Referenzdatensatz sollte explizit so konstruiert werden, dass definierte Deltas und zugehörige Lösungsstrategien bewertet werden können und somit auch im Anschluss an das Vorhaben als Benchmark für Delta Learning-Probleme gelten. Die Referenzdaten haben sich aus zwei Quellen gespeist. Erstens wurden bestehende Datensätze wie KITTI oder Cityscapes bewertet und der Nutzbarkeit nach Datenschutzgrundverordnung (DSGVO) entsprechend in die Referenzdaten integriert. Zweitens wurde innerhalb des Projektes ein Datensatz erstellt, anhand dessen alle adressierten Deltas des Projektes analysiert werden können. Dieser Datensatz sollte dabei reale und synthetische Daten vereinen. Die Erstellung des Datensatzes sollte es ermöglichen, dass aus dem Vorhaben eine Norm zur Bewertung von Methoden für Delta Learning hervorgeht, die über die Projektlaufzeit hinaus Gültigkeit und Relevanz hat.

2.2.2 Referenzmodelle

Zur Schaffung einer Referenz und einer Basis für die Bewertung des Erfolgs verschiedener Methoden wurden Referenzmodelle implementiert, die den gegenwärtigen Stand der Technik repräsentieren. Die Modelle orientierten sich an automotive-typischen Aufgabenfällen und haben je nach Delta unterschiedliche Aufgaben adressiert. Dieses Arbeitsziel war entscheidend, um den Erfolg der zu entwickelten Delta Learning-Methoden und -Modelle zu bewerten, aber auch, um die Methoden auf Basis der Referenzmodelle zu entwickeln. Auch haben die Referenzmodelle dazu gedient, den Einfluss der verschiedenen Deltas auf das avisierte Performanzkriterium zu quantifizieren.

2.2.3 Methoden Delta Learning

Im Arbeitsziel Methoden Delta Learning wurden spezifische Methoden zur Überbrückung der identifizierten Deltas definiert. Für jedes Delta wurden im Projekt verschiedene Methodiken untersucht, evaluiert und bewertet. Darüber hinaus beinhaltetete das Arbeitsziel eine Konsolidierung und gesamtheitliche Bewertung der zugrundeliegenden Methoden anhand der Deltas.

2.2.4 Modelle Delta Learning

Abgeleitet aus den Methoden wurde das Arbeitsziel Modelle Delta Learning definiert. Hier war das Ziel, mithilfe der identifizierten oder neu entwickelten Methoden funktionstüchtige Modelle zu entwickeln, welche eine Lösung eines spezifischen Deltas darstellen. Während die Methoden das Vorgehen beschreiben, sind die Modelle Resultate der Methoden und sind direkt anwendbare Ergebnisse des Projektes.

2.2.5 Bewertungen und Demonstratoren

Das Arbeitsziel Bewertungen und Demonstratoren war die finale Kommunikation der Ergebnisse des Vorhabens KI-Delta Learning. Um die Ergebnisse bewerten zu können, wurden bereits bewährte als auch in diesem Projekt neu zu entwickelnde Bewertungskriterien herangezogen. So konnte der Fortschritt der neu entwickelten Methoden und Modelle im Vergleich zu dem Stand der Technik messbar gemacht werden. Darüber hinaus sollten die Ergebnisse auch in visuelle Demonstratoren einfließen, sodass ein greifbarer Eindruck über die Erfolge des Projekts vermittelt werden kann.

Insgesamt hat sich das Projekt KI-Delta Learning innerhalb von Unsicherheiten und durch von schnellen Verbesserungszyklen geprägte Forschungsgebiete bewegt. Im Projekt wurde daher das Konzept der rollierenden Planung genutzt und die iterativen Schleifen führen zu einem inkrementellen Entwicklungsprozess. Außerdem wurden bei der Methodenentwicklung mehrere aussichtsreiche Konzepte von verschiedenen Partnern und deren jeweiligen Kompetenzen parallel und jeweils mit minimalem Aufwand eruiert. Dies hat die schrittweise Verfeinerung von Zielen unter Einbeziehung aktueller interner und externer Entwicklungen ermöglicht, indem der Entwicklungsstand an Meilensteinen und Quality Gates bewertet wurde und die Anforderungen daraufhin überprüft und verfeinert wurden. Durch regelmäßige Synchronisation zwischen TPs und APs wurde sichergestellt, dass vielversprechende Methodiken AP-übergreifend kommuniziert wurden und für den Fortschritt in den einzelnen parallel arbeitenden Teilen des Projekts genutzt werden konnten.

2.3 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

2.3.1 Bezug des Vorhabens zu den förderpolitischen Zielen

2.3.1.1 BMWi-Fachprogramm

Das BMWi-Fachprogramm, im Rahmen dessen das vorliegende Vorhaben gefördert wurde, formuliert als Hauptziel die „Stärkung der industriellen und wissenschaftlichen Innovationskraft Deutschlands im Bereich des automatisierten Fahrens“ [1].

Der Zeitdruck ist hoch und eine systematische Förderung der Erforschung von KI im Allgemeinen [2] und der Entwicklung von KI-basierten automatischen Fahrfunktionen war und ist weiterhin daher zwingend geboten. So empfiehlt die Expertenkommission Forschung und Innovation (EFI) der Bundesregierung in ihrem neuen Gutachten, ihre Aktivitäten bei autonomen Systemen und KI weiterhin zu verstärken – ein Impuls, der von den Ressorts BMBF und BMWi positiv aufgegriffen wurde [3] [4].

Das hier dokumentierte Vorhaben hat einen zentralen Beitrag zur Stärkung der Wettbewerbsfähigkeit des Wirtschafts- und Wissenschaftsstandorts Deutschland geleistet. Durch das Projekt wurden die nationale Automobilindustrie und Forschungslandschaft befähigt, bestehende Kompetenzen für eine Schlüsseltechnologie im Bereich des automatisierten und

vernetzten Fahren auf- und auszubauen [5]. 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.

Das Vorhaben weist einen konkreten Bezug zur Programmsäule „Automatisiertes Fahren“ des BMWi-Fachprogramms auf [1]. In Summe adressiert KI-Delta Learning drei der dort definierten sieben Themenfelder:

Themenfeld	Forschungsziel	Beitrag KI Delta Learning
Innovative Sensorik und Aktoriksysteme	Weiterentwicklung Sensorik- und Aktoriksysteme und dazugehörige Architektur	<ul style="list-style-type: none"> Steigerung der Nutzbarkeit bestehender Sensorik-Architekturen bei sich ändernden Sensor-Setups Weiterverwendung von Daten von einer Sensorgeneration zur nächsten bzw. die Überbrückung von Deltas zwischen unterschiedlichen Sensortypen (z.B. zwischen Video, Lidar und Radar)
Neuartige Verfahren zur Datenfusion und -verarbeitung	Optimierung der Verfahren zur Sensordatenfusion und -auswertung	<ul style="list-style-type: none"> Herausforderung aus Sicht der KI besteht nicht primär durch sich ändernde Automatisierungsgrade, sondern z.B. durch sich ändernde Einsatzorte Befähigung zur Weiterverwendung bestehender Verfahren in anderen Umgebungen bzw. Einsatzorten bzw. Use Cases – mit dem Fokus auf verschiedene Wahrnehmungsaufgaben
Angepasste Testverfahren und Validierung	Entwicklung neuartiger Indikatoren, Testverfahren und -methoden für Test und der Absicherung der automatisierten Fahrfunktionen	<ul style="list-style-type: none"> Entwicklung von Ansätzen für den robusten Einsatz von KI-Modulen, die u.a. Widerstandsfähigkeit gegen unbekannte, seltene oder kritische Szenen ermöglichen Messbarmachung von Robustheit durch Definition von „Kritikalitätsfunktionen“ messbar gemacht Definition von Robustheit als Lernziel für KI-Module

Tabelle 1: Beiträge des Vorhabens KI-Delta Learning zu den Forschungszielen des BMWi-Fachprogramms

2.3.1.2 KI-Strategie der Bundesregierung

Die Bundesregierung hat im Rahmen der Strategie Künstliche Intelligenz [6] das Ziel ausgegeben, Deutschland zu einem führenden KI-Standort zu machen und die Entwicklung der KI gemeinwohlorientiert und zum Nutzen des Menschen voranzutreiben. Hierzu wurden zwölf Handlungsfelder definiert, in denen die Technologie sowie deren Förderung strukturell verbessert werden sollte. Im Rahmen dieses Vorhabens sind hiervon drei hervorzuheben: die Förderung der Forschungslandschaft, der Transfer der neuesten Methoden in die Wirtschaft und Industrie sowie die Verfügbarmachung von Daten.

Im Strategie-Papier wird weiter der Aspekt der Mobilität speziell herausgehoben: „Wir wollen, dass KI ihr Potenzial voll entfaltet, um unsere Mobilität insgesamt sicherer, umweltfreundlicher, effizienter und preiswerter zu gestalten.“ Um diesem Anspruch gerecht zu werden muss KI entwickelbar und modular sein, sodass neue Aspekte und Erkenntnisse (Deltas) sicher und effizient in bestehende Algorithmen integriert werden können. Hierfür hat dieses Vorhaben einen wesentlichen Beitrag geleistet, indem es:

- *Förderung der Forschungslandschaft*: Forschungsunternehmen und Wissenschaftspartner in ein Konsortium aus Industrie und Politik eingebunden hat, um so die Innovationskraft der deutschen Forschungslandschaft zu stärken.

- *Verfügbarmachung von Daten*: Daten erhoben, katalogisiert, strukturiert und vor allem für ein breites Forum aus Partnern verfügbar gemacht hat, um gemeinsame Nutzung und Verarbeitung im Bereich der deutschen Automobilindustrie zu ermöglichen.
- *Transfer der neuesten Methoden in die Wirtschaft und Industrie*: den unmittelbaren Transfer von Forschungsergebnissen in die Wirtschaft gefördert hat. Durch die Einbindung von Unternehmen aus allen Bereichen der automobilen Wertschöpfungskette (OEM, Tier1 und Zulieferer) und in den Kontext der VDA Leitinitiative wurde sichergestellt, dass Ergebnisse schnell und standardisiert bereits in frühen Stadien des Projekts genutzt werden können.

2.4 Planung und Ablauf des Vorhabens

Die nötigen Arbeiten um die Ziele von KI Delta Learning zu erreichen wurden in 6 Teilprojekten (TPs) organisiert: Drei TPs in denen die eigentliche Forschungs- und Entwicklungsarbeit zum Lernen von Deltas verortet war:

- **TP2: Transfer Learning**
- **TP3: Didaktik**
- **TP4: Automotive Tauglichkeit**

Der projekteigene Datensatz wurde in **TP1: Datengewinnung** erzeugt, während die begleitende Prüfung der Zwischenergebnisse und gegebenenfalls die Aktualisierung der Projektroadmap in **TP5: Evaluation** geleistet wurde. Die in Projekten dieser Größenordnung nötige Koordination und die Ergebnisverbreitung waren in **TP6: Projektorganisation** angesiedelt. Die untenstehende Grafik zeigt den Projektstrukturplan mit diesen TPs und den dazugehörigen Arbeitspaketen (APs).

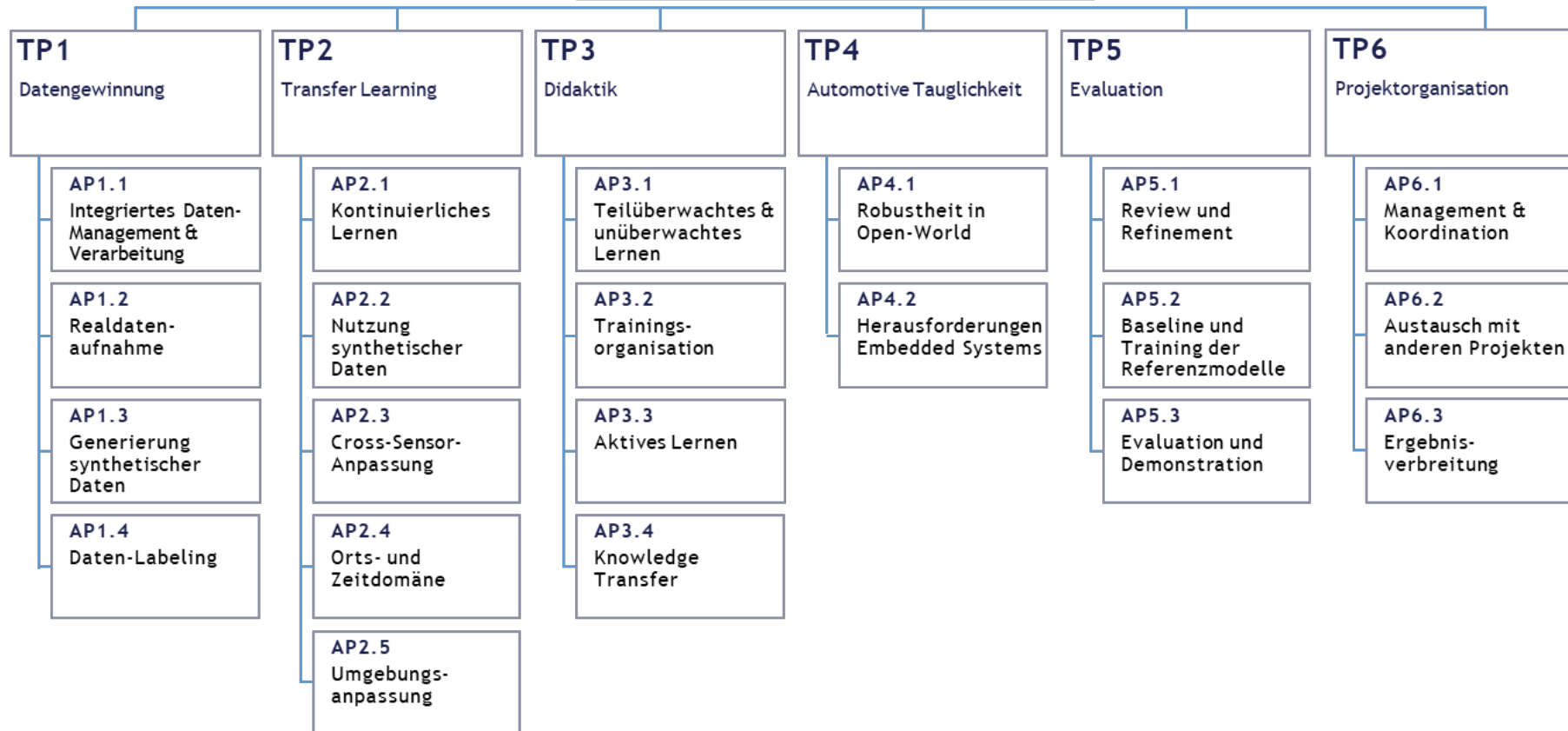


Abbildung: Projektstrukturplan von KI Delta Learning

Die oben erwähnte rollierende Planung wurde konkret über die Etablierung von vier Projektinkrementen (PIs) realisiert, in denen die Ergebnisse immer weiter verfeinert wurden:

- Projektinkrement I: Grundlegende Arbeiten
- Projektinkrement II: Entwicklung
- Projektinkrement III: Optimierung
- Projektinkrement IV: Evaluation

Der initial vorgesehene Zeitplan musste wegen der Verzögerungen durch die Covid-19-Pandemie und die langwierige Abstimmung der Datenschutzvereinbarung etwas gestreckt werden, weswegen auch die Projektverlängerung von vier Monaten vom Konsortium beantragt und vom Projektträger genehmigt wurde. Der so entstandene Zeitplan der PIs ist in der folgenden Grafik wiedergegeben.

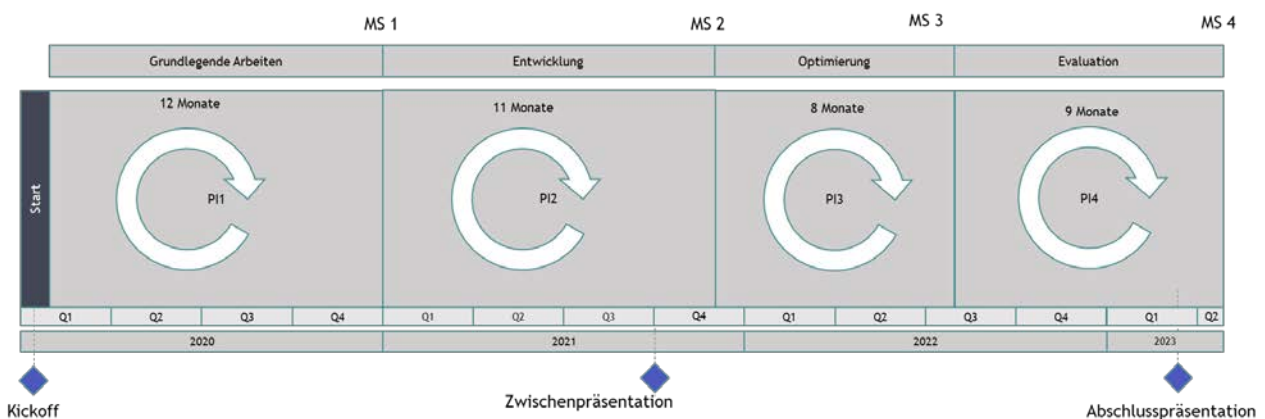


Abbildung 3: Realisierter Projektplan inklusive der Änderungen

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

[2] B. Schölkopf, „Kybernetische Revolution,“ [sueddeutsche.de](http://www.sueddeutsche.de), 15 März 2018. [Online]. Available: <http://www.sueddeutsche.de/politik/aussenansicht-kybernetische-revolution-1.3907249>. [Zugriff am 24 April 2018].

[3] Bundesministerium für Bildung und Forschung, „Verstärkt in Forschung und Innovation investieren,“ 28 Februar 2018. [Online]. Available: <https://www.bmbf.de/de/verstaerkt-in-forschung-und-innovation-investieren-5739.html>. [Zugriff am 24 April 2018].

[4] Bundesministerium für Wirtschaft und Energie, „Machnig: G7 intensivieren Austausch und Kooperation zu Künstlicher Intelligenz,“ 27 März 2018. [Online]. Available: <https://www.bmwi.de/Redaktion/DE/Pressemitteilungen/2018/20180327-machnig-g7-intensivieren-austausch-und-kooperation-zu-kuenstlicher-intelligenz.html>. [Zugriff am 24 April 2018].

[5] Bundesministerium für Bildung und Forschung, „Lernende Systeme müssen der Gesellschaft dienen - nicht umgekehrt,“ 16 März 2018. [Online]. Available:

<https://www.bmbf.de/de/lernende-systeme-muessen-der-gesellschaft-dienen---nicht-umgekehrt-5842.html>. [Zugriff am 24 April 2018].

[6] Bundesregierung, „Strategie Künstliche Intelligenz der Bundesregierung,“ 11 2018. [Online]. Available: https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie.pdf. [Zugriff am 05 12 2018].

3 Eingehende Darstellung

3.1 TP1: Datengewinnung

3.1.1 Ausgangssituation und Zielsetzung

Das Ziel des Teilprojekts Datengewinnung war die Entwicklung von Methoden und Werkzeugen, um eine effizientere Datengenerierung zu ermöglichen, mit dem Ziel vorhandene KI-Module von autonomen Fahrzeugen auf neue Domänen und komplexe Szenarien zu erweitern.

Die Generierung von Daten ist sehr aufwändig und kostspielig, da insbesondere die Sensorintegration und Kalibration als auch die Annotationen viele manuelle Schritte erfordern. Das Teilprojekt Datengewinnung soll diese Prozesse durch automatisierte Annotation vereinfachen und optimieren. Das erste grundlegende Arbeitsziel von KI-Delta Learning war daher der Aufbau eines Referenzdatensatzes, der explizit so konstruiert wurde, dass definierte Deltas und zugehörige Lösungsstrategien bewertet werden können und somit als Benchmark für Delta Learning-Probleme etabliert werden können. Die Datensätze sollten reale und synthetische Daten vereinen und auf die Anforderungen des Delta Learnings angepasst sein.

Das Teilprojekt bestand aus vier Arbeitspaketen: AP1.1, das ein integriertes Daten-Management und -Verarbeitungssystem entwickelt hat, AP1.2 und AP1.3, die die Erzeugung eines Referenzdatensatzes mit Real- und synthetischen Daten und dessen Bewertung umfasst haben und AP1.4, das die Entwicklung und Implementierung von automatischen Annotationsverfahren der gesammelten Daten umfasst hat, um den manuellen Aufwand zu minimieren.

Das Teilprojekt 1 "Datengewinnung" hat auf die Entwicklung neuer Verfahren zur effizienten, gezielten, datenschutzkonformen Erzeugung und Annotation von Trainingsdaten für Delta Learning gezielt, sowie auf die Bewertung der Güte der annotierten Trainingsdaten und Technologien für deren Datenmanagement und Processing ab. Das Projektziel war die Schaffung eines Referenzdatensatzes, der als Benchmark für Delta Learning-Probleme verwendet werden kann und somit den Einsatz künstlicher Intelligenz in autonomen Fahrzeugen ermöglichen soll.

3.1.2 Strukturierung

Das Teilprojekt TP1 war in AP1.1, AP1.2, AP1.3 und AP1.4 unterteilt.

AP1.1 hat sich mit dem integrierten Datenmanagement und der Datenverarbeitung befasst. Die Herausforderungen der globalen Aufstellung und firmenübergreifenden Kooperationen erfordern Lösungen, die die bestehende Diversität aller Stakeholder beherrschbar bleiben lassen und ein effizientes Zusammenarbeiten ermöglichen. Die Abstraktion und Automatisierung sind Schlüssel dazu. Die Abstraktion transformiert reale und im Allgemeinen unterschiedliche Entitäten auf eine höhere Ebene und schafft eine einheitliche und übergreifende Beschreibung und transparente Semantik. Die Automatisierung ermöglicht die Standardisierung und Automatisierung von Abläufen und Workflows. In AP1.1 wurde ein System zum integrierten Daten-Management und -Processing implementiert, das Daten, Prozesse und Tools miteinander vernetzt und automatisiert, wie sie für die Entwicklungen zum Delta Learning in allen TPs dieses Projekts notwendig waren.

AP1.2 hat sich mit der Realdatenaufzeichnung befasst. Das Ziel war, geeignete Daten für das Delta Learning zur Verfügung zu stellen. Das Projekt hat die Bereitstellung von Rohdaten, die Akquisition von Daten und deren Aufbereitung sowie die Unterstützung bei der

Datenanreicherung enthalten. Insbesondere wurden hier Methoden entwickelt, um die zu labelnden Key-Frames auszuwählen und dadurch die Varietät des Datensatzes zu maximieren. Eine hohe Diversität an Daten ist wichtig, um eine entsprechende Generalisierung des Modells zu ermöglichen.

AP1.3 hat sich auf die Generierung synthetischer Daten konzentriert. In diesem Teilprojekt ging es darum, eine geeignete Methode zu entwickeln, um synthetische Daten für das Delta Learning zu generieren. Hierfür mussten geeignete Modelle für die Generierung von Daten entwickelt werden. Dies umfasste zum einen die Modellierung der verwendeten Sensoren, der Modellierung der 3D-Umgebung und die Entwicklung von Annotierungsstrategien für Simulationsumgebungen.

AP1.4 war für die Datenaufbereitung und -analyse zuständig. Die Datenaufbereitung sollte dabei helfen, die Qualität der Daten zu verbessern, während die Datenanalyse einen Einblick in die Daten liefert und Erkenntnisse für die weitere Entwicklung von autonomen und KI-basierten Fahrfunktionen bieten sollte. In diesem Teilprojekt wurden auch Modelle entwickelt, um die Qualität der Daten zu bewerten und die Ergebnisse zu visualisieren.

3.1.2.1 AP1.1 Integriertes Datenmanagement und -Verarbeitung

3.1.2.1.1 Zielsetzung

Ziel des Arbeitspaketes 1.1 war die Erstellung eines Systems zum integrierten Datenmanagement und der Verarbeitung dieser Daten. Dieses System soll die Daten, die bei Versuchsfahrten (AP 1.2) aufgenommen werden, speichern und - soweit möglich - automatisch verarbeiten. Den Nutzern soll dabei eine Oberfläche angeboten werden die einen guten Überblick über die Daten liefert und es erlaubt, die Daten soweit möglich einzusehen und zu evaluieren. Dabei sind die wesentlichen Anwendungsfälle:

- Betrachten der Datensätze zu einer Begutachtung
- Automatische Verarbeitung der Datensätze mittels mehrerer Prozessschritte
- Manuelle Freigabe von "Key-Frames" (besonders relevanten Momenten in einem Datensatz)

Weiterhin sollten Aspekte wie Nutzerrechte-Management oder Anonymisierung der Daten in Hinblick auf persönliche Daten in den Kamerabildern, wie beispielsweise Gesichtern oder Fahrzeugkennzeichen, integriert werden.

Die Verarbeitung der Daten soll soweit möglich automatisiert erfolgen, und einen Austausch von einzelnen Prozessschritten ermöglichen, um einfach an unterschiedliche Workflows anpassbar zu sein.

3.1.2.1.2 Erzielte Ergebnisse

Im Rahmen des AP 1.1 wurde eine Plattform für integriertes Datenmanagement und -verarbeitung aufgebaut, die im Wesentlichen aus zwei Teilen besteht; Die *Benutzeroberfläche* einerseits, die es erlaubt Informationen über die Daten zu erhalten und diese zu editieren, sowie die zugrundeliegende *Plattform* andererseits, die größtenteils automatisiert die einzelnen Datenverarbeitungsschritte ausführt.

Die Benutzeroberfläche bietet zum einen die Übersicht über alle verfügbaren Datensätze, inklusive wichtiger Metadaten wie beispielsweise Länge des Datensatzes, Datum der Erstellung, sowie Status in der automatischen und manuellen Verarbeitung der Daten. Darüber hinaus ist es ebenfalls möglich, nach Datensätzen zu filtern oder zu suchen.

Suche nach Aufnahmen

Manuelle Freigabe der "Key-frames"

Automatische Datenverarbeitung

Labeling

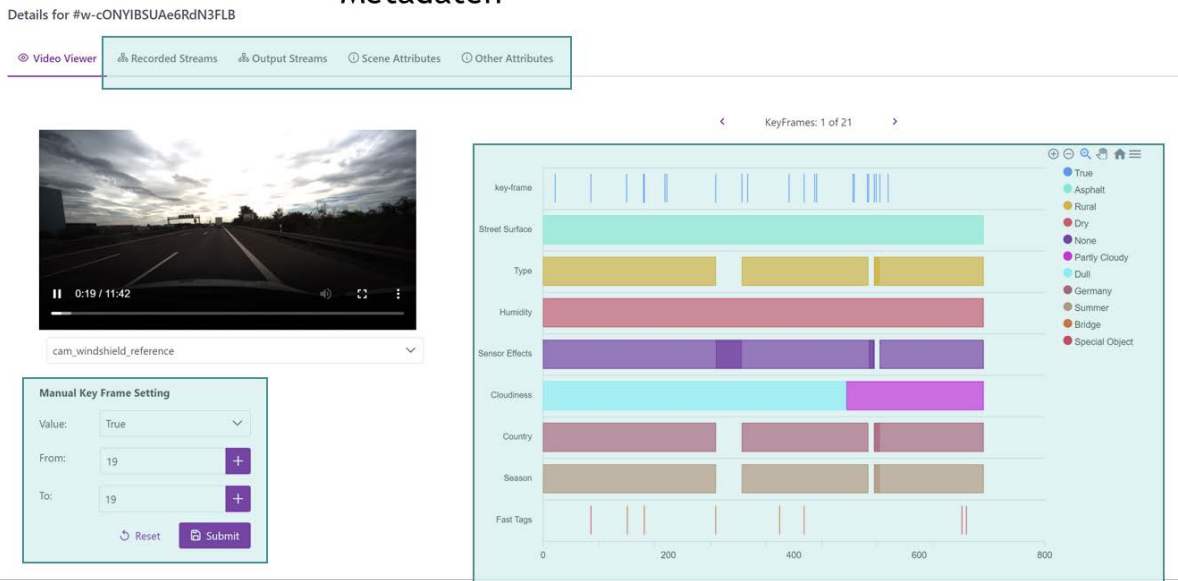
Details	ID	Path	Duration	Recording Status	Key Frame Status	Annotation Status	Created	Last modified
	X4t9NIBczspO8vZq4I	BB_XY_6226_2022-07-14-08-42-05/	1414	key-frame-anonymized	approved	ready-for-delivery	2022-07-25T10:53:35.059541Z	
	UwdkNYIBSUA6R0N2IAH	BB_XY_6226_2022-07-13-15-05-21/	1039	key-frame-anonymized	approved	ready-for-delivery	2022-07-25T12:45:58.859275Z	
	w-c0NYIBSUA6R0N3FLB	BB_XY_6226_2022-07-13-12-13-21/	702	key-frame-anonymized	approved	ready-for-delivery	2022-07-25T11:12:22.671390Z	
	wn55_iEBczspO8vZq4I	BB_XY_6226_2022-06-29-09-04-40/	179	key-frame-anonymized	approved	ready-for-delivery	2022-07-14T11:30:26.059683Z	
	P6svYMBKaktNw_X3v0B	BB_XY_6226_2022-08-22-15-53-34/	7380	key-frame-anonymized	approved	session-created	2022-09-10T02:19:28.170079Z	
	Xq8m0iMBKaktNw_XzpzP	BB_XY_6226_2022-08-24-16-28-36/	5885	key-frame-anonymized	approved	session-created	2022-09-13T18:40:10.472545Z	
	H6x7joMBKaktNw_XEybX	BB_XY_6226_2022-08-22-14-43-40/	2921	key-frame-anonymized	approved	session-created	2022-09-10T08:19:03.550274Z	
	EkyzYMBKaktNw_X2gpp	BB_XY_6226_2022-08-22-13-59-30/	2647	key-frame-anonymized	approved	session-created	2022-09-10T04:16:19.905439Z	
	IADBC4E8FPFqLmZjgID	BB_XY_6226_2022-03-15-13-34-03/	581	key-frame-anonymized	approved	session-created	2022-05-28T18:44:43.279037Z	
	H6vLi8BczspO8vZq4I	BB_XY_6226_2022-07-12-07-05-30/	13628	key-frame-anonymized	approved	session-created	2022-07-23T21:06:48.468586Z	

Abbildung 4: Web-Ansicht der Plattform für integriertes Datenmanagement und -verarbeitung mit Hervorhebung wichtiger Metadaten sowie der Funktion zum Suchen und Filtern.

Für einzelne Datensätze wiederum bietet die Benutzeroberfläche eine Übersicht über die Videodaten, eine Darstellung des zeitlichen Verlaufs einzelner Attribute in dem Datensatz, sowie eine Möglichkeit zur manuellen Freigabe von "Key-Frames".

Aus den freigegebenen Key-Frames wird wiederum eine Liste von Zeitstempeln generiert, die an ein Team weitergegeben wird, deren Aufgabe es ist, die Daten (Video und Lidar) zu diesem Zeitstempel zu labeln, d.h. die einzelnen Bereiche in den Bilddaten Objektklassen zuzuordnen. Dieses Labeling (auch Annotation genannt) bildet später die Grundlage (Ground Truth) für das Training der Algorithmen.

Metadaten



“Key Frame” Freigabe

Attribute

Abbildung 5: Ansicht eines Datensatzes mit Hervorhebung der Reiter für Metadaten, der Darstellung der Attribute im Zeitverlauf sowie der Funktionen für die "Key Frame" Freigabe.

Auf der Plattform für die Datenverarbeitung laufen eine Reihe von Prozessen sequentiell ab, um die Rohdaten aufzubereiten. Diese Schritte sind im Einzelnen:

- Validieren und Importieren der Daten
- Erstellung von Videos aus den Rohdaten
- Automatische Vorauswahl der Key-Frames
- Extraktion aller Sensordaten für die Key-Frames (nach manueller Freigabe), incl. Anonymisierung.
- Erstellung einer Session in C.LABEL für das Labeling-Team

Diese einzelnen Prozessschritte sind als sog. Container implementiert, also eigenständig lauffähige Prozesse. Dies erlaubt eine hohe Flexibilität bei der Umsetzung, z.B. für das Ändern oder Ergänzen einzelner Prozessschritte.

3.1.2.1.3 Gegenüberstellung der vorgegebenen Ziele

Im Vergleich zu den Zielen lässt sich sagen, dass im AP 1.1 die Plattform für die Datenverarbeitung erfolgreich aufgebaut werden konnte, und alle angestrebten Einsatzzwecke erfüllt werden konnten. Die Datensätze können schnell und einfach überblickt und durchsucht werden, Datenverarbeitung ist größtenteils automatisiert und die manuelle Nutzung ist durch eine übersichtliche Oberfläche vereinfacht. Zuletzt sind Datenschutzaspekte in Hinblick auf Anonymisierung von Gesichtern und Kennzeichen berücksichtigt. Die Implementierung als Container-basierte Cloud-Lösung ermöglicht einen hohen Grad an Flexibilität und Automatisierung der Datenverarbeitung.

3.1.2.2 AP1.2 Realdatenaufnahme

3.1.2.2.1 Zielsetzung

Die Zielsetzung von AP1.2 war die Erstellung eines sorgfältig ausgewählten Trainingsdatensatzes, der mithilfe definierten Anforderungen bezüglich möglicher Domänenänderungen koordiniert wird. Im Besonderen sollten die im Projekt definierten Deltas Sensoren, Verkehrsräume, Länder, Umwelt und Zeit abgedeckt werden.

Der Datensatz sollte iterativ erstellt werden, um neue Anforderungen und Erkenntnisse zu berücksichtigen. Es sollten außerdem intelligente Systeme entwickelt werden, um die Bewertung der Güte und Diversität des Datensatzes zu unterstützen. Zusätzlich sollte die Datenaufnahme durch den Einsatz von Verkehrsinfrastrukturanlagen ermöglicht und bei Bedarf durch ergänzende Daten erweitert werden.

3.1.2.2.2 Erzielte Ergebnisse

AP1.2 umfasste die Themen

- Definition des Sensor Setup,
- die zu fahrenden Routen,
- die Auswahl und Spezifikation der zu annotierenden Daten, und die
- Evaluierung und Validierung der Daten.

3.1.2.2.2.1 Aufnahme-Fahrzeug

Für die Spezifikation des Aufnahmefahrzeugs wurde frühzeitig die *Working-Group Sensor Specifications* gegründet. Dadurch konnte sichergestellt werden, dass die Anforderungen aller Partner berücksichtigt werden konnten. Außerdem wurde eine enge Zusammenarbeit zwischen der Fahrzeug-Integration und der Sensorhersteller für die Seriensensorik gewährleistet. Die endgültige Spezifikation umfasste

- 1x 8.9MP Referenz-Kamera mit 120° Abdeckung
- 1x 2.3MP Fischaugen-Kamera mit 130° Abdeckung
- 9x 2.3MP Protokoll-Kamera mit 90° Abdeckung
- 1x 2MP Serienkamera Bosch MPC3 mit 100° Abdeckung
- 1x Referenz-LiDAR Velodyne 128
- 2x Protokoll-LiDAR Velodyne 32
- 1x Serien-LiDAR Scala Mobility Kit 16
- 1x Serien-LiDAR Innoviz One
- 2x Serienradar InnoSenT 5G3 20fps
- 1x IMU Oxford RT3000 (hochgenaues GNSS).

Zu jedem Sensor wurde eine entsprechende Verbauposition und -ausrichtung festgelegt. Neben der Sensorik, wurden hier auch weitere Themen besprochen, wie etwa die Systemarchitektur, die Strategien zur Synchronisierung und Kalibration und der Datenformate.

Durch die verzögerte GUA3 Beauftragung, wurde erst im ersten Halbjahr 2021 mit dem Fahrzeugaufbau begonnen. Die Integration und Kalibration wurde Großteils von CMORE übernommen. Die softwareseitige Integration wurde mit den Partnern Bosch, InnoSenT und Valeo abgestimmt, die zum Teil ein entsprechendes Software Development Kit entwickelt haben, um die Aufnahme im entsprechenden ROS Message Typ zu ermöglichen.

Durch die späte Einigung und Unterzeichnung der Datenschutz-Verträgen, konnte die Qualitätsprüfung und damit der Fahrzeugaufbau erst im ersten Halbjahr 2022 fertiggestellt wurde. Fokus der Qualitätsprüfung war insbesondere die Evaluation der Sensorpositionen, Synchronisation und Kalibration im Aufnahmemodus. Dafür wurden beispielsweise mehrere Sensoren ineinander fusioniert, um eine entsprechende hohe Güte der Kalibration zu gewährleisten.



Abbildung 6: Am Fahrzeug angebrachte Sensoren.

3.1.2.2.2 Routenplanung

Auch die Routendefinition wurde in einer gleichnamigen Working Group bearbeitet.

Zunächst wurden die allgemeine Anforderungen der Partnern gesammelt. Anschließend wurden konkrete Vorschläge für die zu fahrenden Routen erarbeitet und Feedback eingearbeitet. Ziel war es, die im Projekt definierten Deltas abzudecken. Die entstandenen Routen waren:

- der "Local Drive", eine kurze Strecke (80km) zwischen Sindelfingen, Stuttgart und Leonberg, die etwa 40 mal zu verschiedenen Jahreszeiten, Wettersituationen und Tageszeiten gefahren wurde,
- der "Day Drive", eine Tagesfahrt (430km) die mehrere unterschiedliche Regionen und Städte zwischen Stuttgart und Friedrichshafen umfasst und 4 Mal gefahren wurde und
- der "Multi-Day Drive", eine Strecke die über weite Teile von Deutschland geht (1500km) und Nürnberg, Leipzig, Berlin, Braunschweig und Würzburg umfasst. Auf dem Testfeld Niedersachsen in Braunschweig wurden die Aufnahmen mit einer Verkehrsinfrastrukturanlag unterstützt.

Nach einer Zwischenauswertung wurde außerdem ein zweiter Local Drive durchgeführt, der im Allgäu, also geographisch höher liegt, um fehlende Deltas, wie unterschiedliche Straßenbeläge oder Schnee abzudecken. Diese Route verlief auf 175 km durch Lindau, Oberstaufen, Füssen, dem Riedbergpass und Isny.

Darüber hinaus wurden Routenplanungen für verschiedene Länderoptionen erzeugt. Nach der Einigung des Konsortiums wurde die Auslandsfahrt in Italien durchgeführt. Sie führte 2880km von Lindau über Bologna und Pescara und auf vielen Landstraßen an der Ostküste bis nach Neapel und anschließend auf der Autobahn an der Westküste zurück.

Insgesamt sind in Deutschland und Italien 240 Recordings mit insgesamt 193 Stunden Aufnahmen entstanden.



Abbildung 7: Zwei Aufnahmefahrzeuge von KI Delta Learning und vom DLR am Testfeld Niedersachsen

3.1.2.2.2.3 Labeling

In der *Working Group Labeling* wurden die Annotationen spezifiziert und das Labeling Format definiert. Als Ziel wurde festgelegt, die Referenz-Kamera durch semantische Segmentierung im Sichtfeld der Kamera und den Referenz-LiDAR durch 3D Bounding Boxen im vollen 360° Umfang zu annotieren.

In der Working Group fand auch eine Qualitäts-Validierung der Annotationen statt, die zeigte, dass die annotierten Daten von CMORE eine insgesamt gute Qualität haben.

3.1.2.2.2.4 Keyframe Selection

Für eine einfachere Verarbeitung der Daten, wurde festgelegt, dass während der Fahrt der Beifahrer besondere Momente mit einem Tool festhält und durch Tags beschreibt, wie beispielsweise Sonnenreflektionen oder außergewöhnlich gekleidete Fußgänger. Dies ist notwendig um eine automatisierte Auswahl von Aufnahmedaten zu ermöglichen.

Zur weiteren Anreicherung der aufgenommenen Daten, wurden verschiedene Algorithmen implementiert, die den Sonnenstand berechnen oder Informationen aus der Open-Source-Karte "OpenStreetMap" extrahieren.

Es wurde ein Algorithmus entwickelt, der anhand solcher Tags eine automatisierte Auswahl an Frames erlaubt. Dieser Algorithmus versucht solche Frames auszuwählen, die im bestehenden Datensatz unterrepräsentiert sind und erlaubt dadurch eine iterative Verbesserung des Datensatzes. Diese Tags und Metadaten dienen gleichzeitig als Quelle, um eine tiefere Auswertung bezüglich der Diversität durchzuführen.

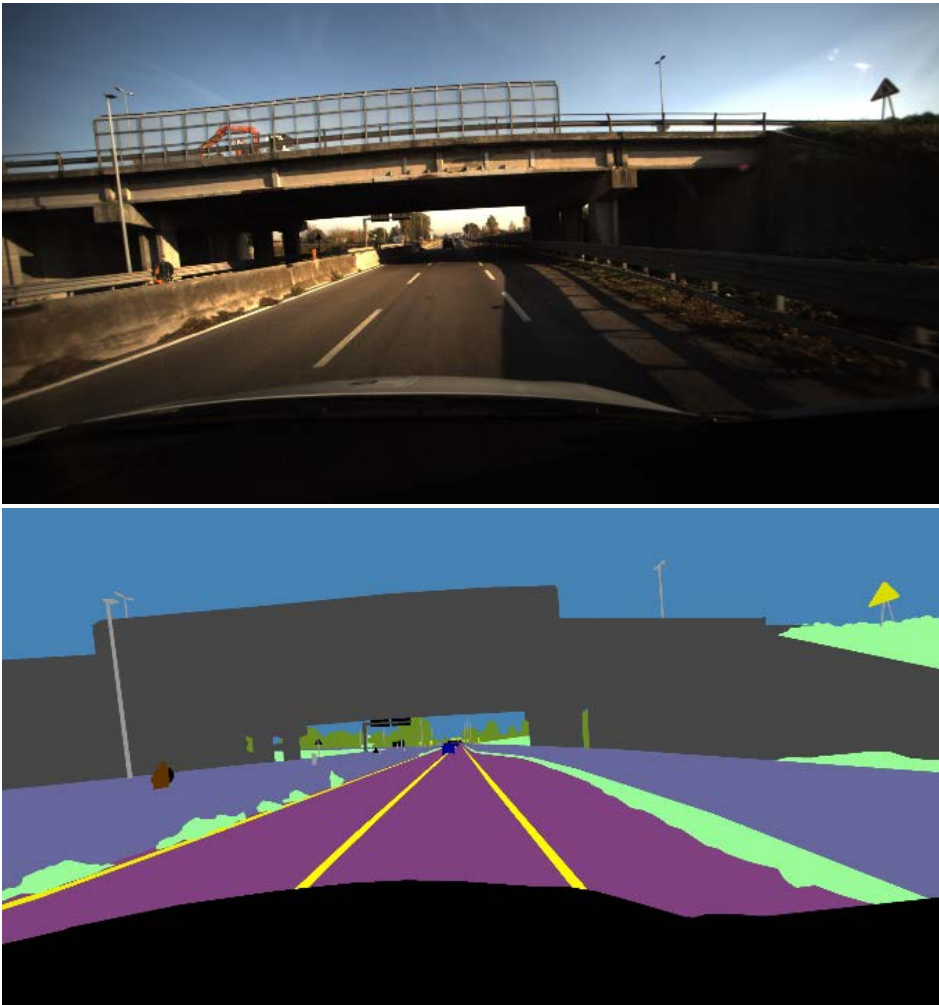


Abbildung 8: Beispielhafte Semantische Segmentierung der Referenz-Kamera

3.1.2.2.2.5 Bewertung des Datensatzes

Um die Verteilung der Labels zu überprüfen und menschliche Vorurteile zu vermeiden, wurden verschiedene Bewertungen durchgeführt.

Dazu wurden sowohl manuelle Qualitätsprüfungen durchgeführt, als auch Tools entwickelt, um die Routen und Key-Frames automatisiert auszuwerten. Durch die Auswertung der Metadaten und Online-Tags, wurden alle Routen analysiert und vielversprechende Szenarien besonders bei der Keyframe-Auswahl berücksichtigt. Insgesamt lässt sich sagen, dass ein Großteil der Deltas erfolgreich durch die eingefahrenen Routen abgedeckt wird. Durch die intensive Planung sind genügend Aufnahmen zu verschiedenen Uhrzeiten und an verschiedenen Lokalitäten entstanden.

Durch eine weitere Evaluation der Key-Frames konnte, im Rahmen der Routenwahl, eine hohe Vielfalt innerhalb des Datensatzes nachgewiesen werden.

3.1.2.2.3 Gegenüberstellung der vorgegebenen Ziele

Insgesamt wurden alle Ziele in AP1.2 erreicht. Durch Vorarbeiten bei der Spezifikation von Sensoren, Routen und Labels, wurde sichergestellt, dass alle einzufahrenden Deltas bestmöglich aufgenommen wurden. Durch die parallel entwickelten Auswerte-Algorithmen, wurden Zwischen-Evaluationen durchgeführt und die Datenaufnahme bzw. die Auswahl der Key-Frames bei Bedarf angepasst. Dadurch ist ein Datensatz entstanden, der insbesondere die Anforderungen an Güte und Diversität erfüllt und ein System mit dem solche Auswertungen ermöglicht werden.

3.1.2.3 AP1.3 Generierung synthetischer Daten

3.1.2.3.1 Zielsetzung

Für die Realisierung von KI-Modulen eines Fahrzeugs (bspw. für die Perzeption) ist eine sehr große Menge von Trainingsdaten notwendig, um die große Vielfalt im realen Straßenverkehr näherungsweise abdecken zu können. Diese Vielfalt setzt sich aus verschiedenen Faktoren zusammen, wozu insbesondere die unterschiedlichsten Erscheinungsformen von Objekten, die verschiedensten Wettersituationen und auch die große Menge an möglichen Szenarien zählen. Um die hohen Zeit- und Kostenaufwände zu reduzieren, die mit der realen Datenerhebung und dem Labeling der Rohdaten verbunden sind, war es das Ziel von AP1.3 das KI-Training und Testen durch synthetische Daten zu unterstützen. Hierfür hatte sich das Arbeitspaket insbesondere als Ziel gesetzt den Realismus synthetischer Daten zu erhöhen, als auch relevante Daten gezielter zu erzeugen. Hierbei wurde überwiegend die Simulationssoftware CARLA eingesetzt.

3.1.2.3.2 Erzielte Ergebnisse

3.1.2.3.2.1 Gesteigerter Realismus synthetischer Daten

Der Open-Source-Simulator CARLA bietet direkt einige Standardmodelle für virtuelle Sensoren an. Da insbesondere der bereitgestellte Lidar-Sensor sehr simplifiziert aufgebaut ist, ist der Realismus der resultierenden Punktwolken einer Lidar-Messung stark eingeschränkt. Um dieses Problem zu lösen wurde ein neues Lidar-Modell für CARLA entwickelt, welches auf einem realen Lidar-Sensor basiert. Da das in CARLA bzw. der zugrundeliegenden Unreal Engine integrierte Raytracing-Feature schwer von außen zugreifbar ist, wurde das Lidar-Modell auf Basis von Bilddaten und Tiefeninformationen entwickelt. Durch die Kombination beider Informationen kann eine dreidimensionale Punktwolke bestimmt werden, die in anschließenden Schritten weiterverarbeitet wird. Hierbei kann aus den Farbinformationen eine entsprechende Intensität abgeleitet und anschließend zu einem optischen Signal umgerechnet werden. Dieses kann dann wie in einem realen Lidar-Sensor weiterverarbeitet werden. Letztendlich finden noch weitere Verarbeitungsschritte statt, um reale Effekte wie mehrfache Echos, sogenannte "Ghost Points" oder aber einen gekippten Spiegel (wie er in realen Sensoren zum Einsatz kommen kann) zu simulieren. Die resultierende Punktwolke kann dann durch die von CARLA bereitgestellten semantischen Informationen automatisch gelabelt werden, sodass resultierende Daten für Supervised Learning eingesetzt werden können.

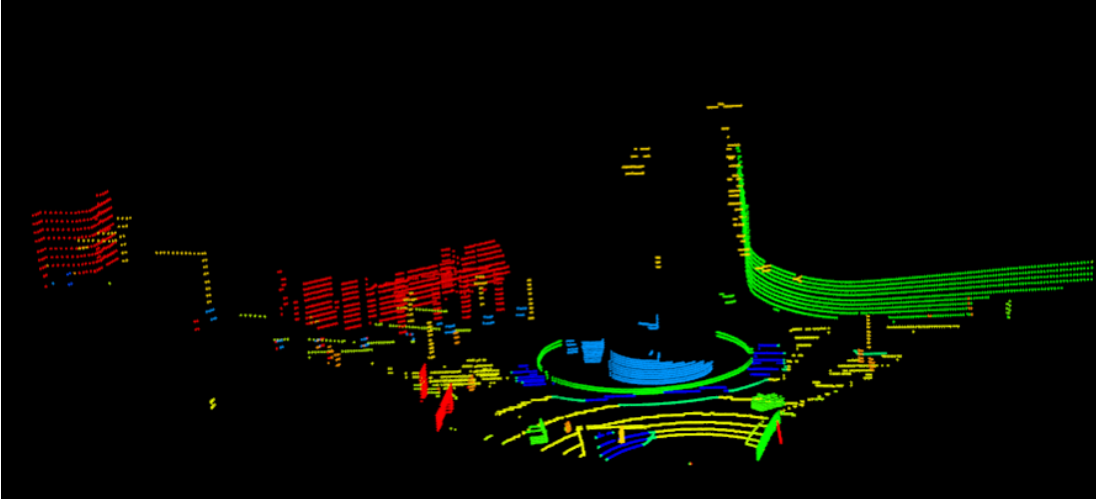


Abbildung 9: Semantisch annotierte Punktwolke als Resultat des entwickelten Lidar-Modells

Neben Bild- und Lidar-Daten werden auch Radar-Sensoren im automotive Umfeld häufig eingesetzt. Deren Simulation gestaltet sich aufgrund der eingesetzten Wellenlänge aber als sehr schwierig, da hierbei insbesondere Materialeigenschaften und Streueffekte von Strahlen die resultierenden Messungen stark beeinflussen. Aus diesem Grund wurde für ein realistisches Radar-Modell ein spezifisches Simulationsframework eingesetzt und eine Prozessierung entwickelt, die einem realen Radar-Sensor nachempfunden ist. Hierbei werden insbesondere die Antennenkonfiguration, verwendete Übertragungsmedien von Radiofrequenzen sowie Rauschquellen modelliert. Das resultierende Modell wurde abschließend in einer kontrollierten Umgebung gegen einen realen Radar-Sensor evaluiert.

Neben den eingesetzten Sensoren bietet CARLA auch eine begrenzte Menge von virtuellen Welten, in denen Simulationen abgespielt und Daten erzeugt werden können. Dadurch ist die Vielfalt der resultierenden Daten naturbedingt stark eingeschränkt, sodass der Bedarf an zusätzlichen realistischen Karten besteht. Hierfür wurde ein Ansatz auf der Basis von Neural Radiance Fields (NeRF) entwickelt, durch den anhand von Kamera- und Lidar-Daten einer realen Szene ein virtuelles Abbild konstruiert werden kann. Da NeRFs in ihrer ursprünglichen Variante nur für kleinere, statische Szenen angewendet werden können, wurden diverse Ansätze aus der Literatur kombiniert und weiterentwickelt, um die Anwendbarkeit auch in hoch-dynamischen Verkehrssituationen zu realisieren. Das Resultat des Trainings eines NeRFs auf einer spezifischen Szene ist ein neuronales Netz, welches im eigentlichen Sinne zur Erzeugung neuer Ansichten (d.h. neuer Kamerapositionen) aus der Szene verwendet wird. Es kann aber auch die in dem neuronalen Netz kodierte 3D-Repräsentation via Sampling zu einer expliziten Grafikrepräsentation umgerechnet werden, welche anschließend weiter aufbereitet und dann in einen Simulator integriert werden kann.



Abbildung 10: Erzeugte Ansicht aus der 3D-Rekonstruktion einer synthetischen Szene (links Farbbild, rechts Tiefe). Es wurden synthetische Szenen zur Evaluierung gegen die bekannte Ground Truth verwendet.

3.1.2.3.2.2 Generierung synthetischer Daten

Für das Training einer Perzeptions-KI werden in der Regel nicht-sequentielle Daten eingesetzt, da sonst zu viel Redundanz in den Daten besteht. Bei der Erstellung realer Datensätze werden hierbei von vielen Stunden Videomaterial einzelne Key-Frames selektiert und anschließend gelabelt. Dieses Prinzip sollte in einem synthetischen Datensatz zwar ähnlich abgebildet werden, würde aber bei der gleichen Vorgehensweise das Simulationsbudget ineffizient ausnutzen. Daher wurden verschiedene Ansätze entwickelt um gezielt Key-Frames zu generieren. Während eine einfachere Strategie zufällig das Ego-Fahrzeug sowie andere Verkehrsteilnehmer in der Welt platziert um anschließend ein einzelnes Frame zu erzeugen, ermöglicht eine weitere Strategie das Berechnen eines optimalen Pfades durch die virtuelle Welt, um die gesamte Vielfalt einer Karte effizient abzudecken. Im Gegensatz zu der zufälligen Platzierung können so auch spannende Verkehrssituationen durch die Interaktion der Verkehrsteilnehmer entstehen. Kombiniert wurden die Strategien mit verschiedensten Wetterbedingungen, welche insbesondere in realen Datensätzen häufig fehlen.

Um synthetische Daten ideal mit realen Daten kombinieren zu können, sollten die Modalitäten der realen und virtuellen Sensoren ähnlich sein. Hierfür wurden die realen Sensorsetups des öffentlichen nuScenes-Datensatzes, sowie auch von projektinternen Versuchsfahrzeugen in

CARLA nachgebildet. Zusätzlich sollten auch die Labelspezifikationen der Daten übereinstimmen, wobei CARLA in der zur Projektlaufzeit verfügbaren Version noch einige Lücken aufwies: Ein Fahrzeug wurde nur generell als "Fahrzeug" gelabelt, nicht aber als eine feinere Klasse wie z.B. "Auto" oder "LKW". Hierfür wurden auch Lösungen entwickelt, die ein genaueres Labeling ermöglichten.

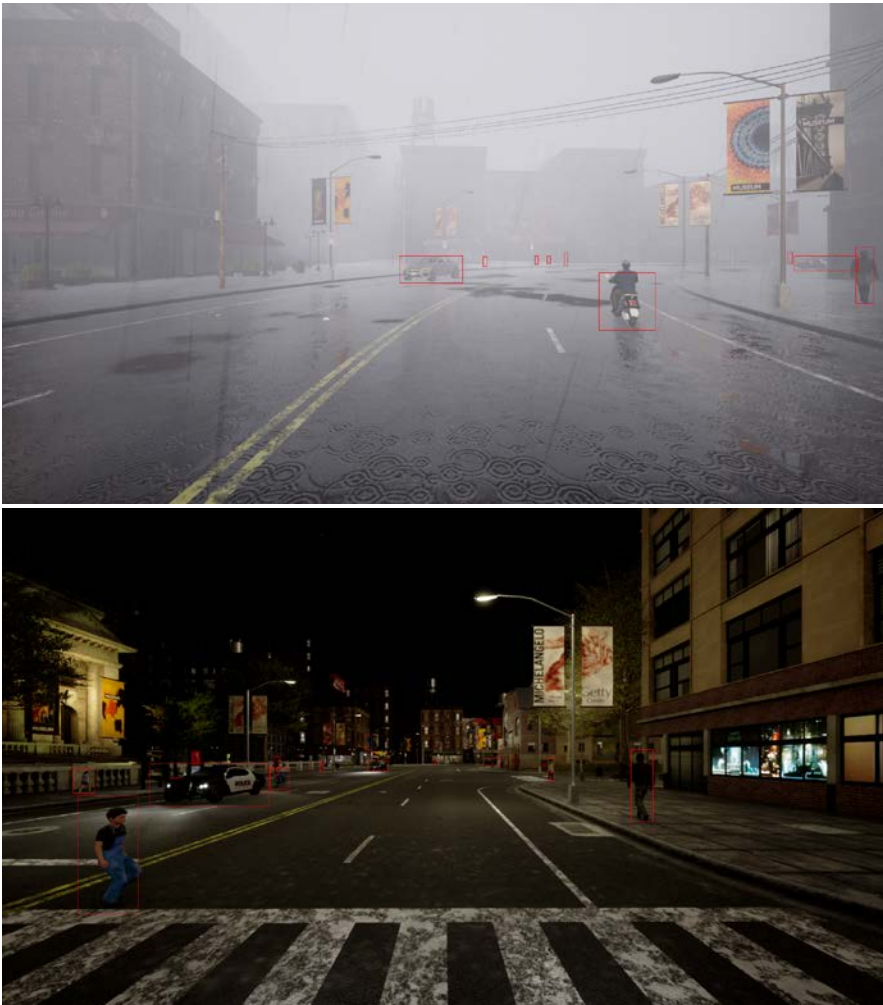


Abbildung 11: Beispiele für erzeugte synthetische Daten mit schwierigen Wetterbedingungen : links mit starkem Regen und Nebel; rechts bei Nacht.

Neben der Simulation verkehrlicher Daten in CARLA wurden auch in einem Motion Capture Labor reale Daten mit E-Scootern und Hoverboards aufgezeichnet, welche anschließend synthetisch repliziert wurden. So entstehen zu den realen Daten synchrone synthetische Daten, welche Untersuchungen bzgl. des Domain Shifts von Simulation zur Realität ermöglichen. Hierbei können auch synthetischen Daten variiert werden, indem bspw. die Avatare der Personen ausgetauscht werden. Schlussendlich lassen sich aus der Simulation mit wenig Aufwand die entsprechenden Datenannotationen berechnen, wie z.B. Tiefe, semantische Segmentierung, Bounding Boxes und Posen, welche für weiterführende Aufgaben verwendet werden können.

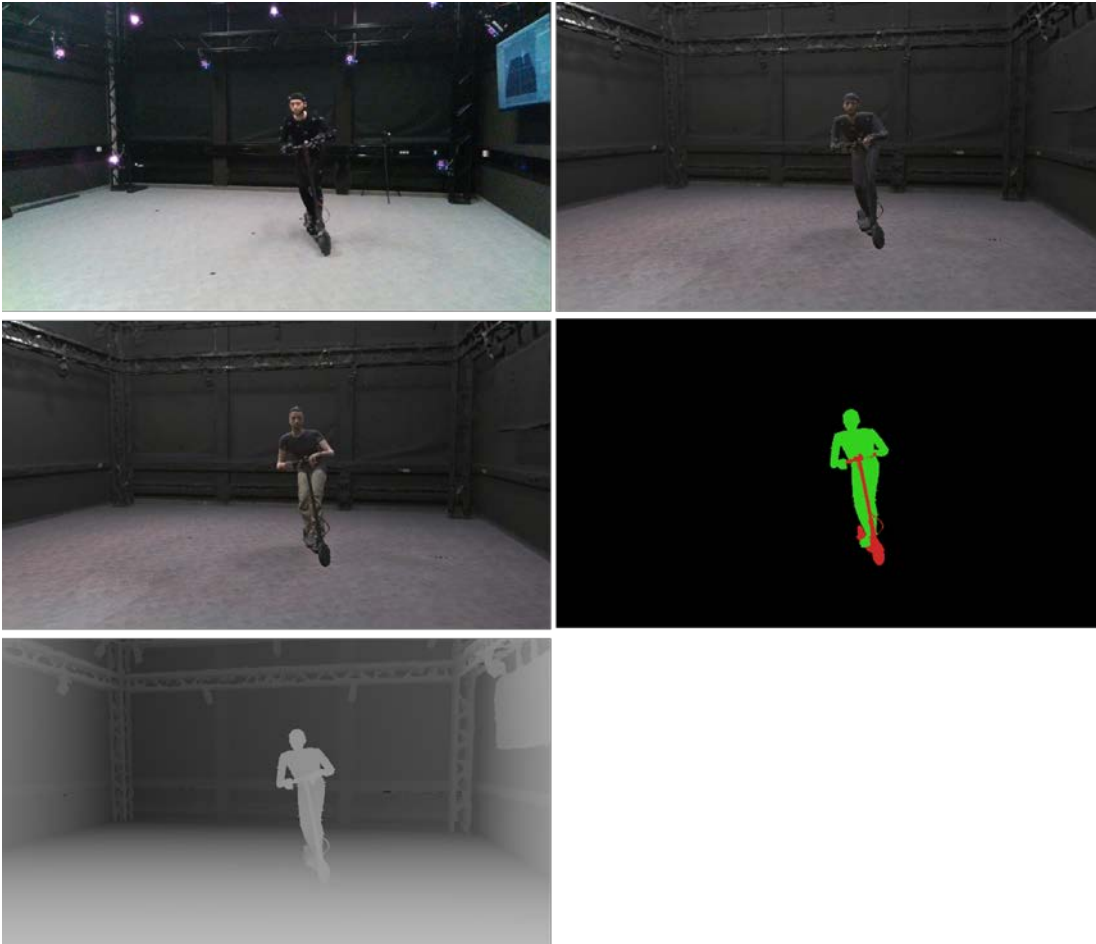


Abbildung 12: Verschieden Darstellungen eines Motion Capture Bildes : Echtes RGB-Bild (oben links), synchrones synthetischen RGB-Bild mit Motion Capture Anzug (oben rechts), synchrones synthetischen RGB-Bild mit gewöhnlicher Kleidung (Mitte links), semantische Segmentierung aus der Simulation (Mitte rechts), Tiefenbild aus der Simulation (unten links). Für eine klarere Darstellung werden Annotationen wie Bounding Boxes oder Posen nicht dargestellt.

3.1.2.3.2.3 Gegenüberstellung der vorgegebenen Ziele

Die erzielten Ergebnisse demonstrieren, dass das AP1.3 die gesteckten Ziele zur Verfeinerung synthetischer Daten sowie deren Erzeugung erreichen konnte. Die Realisierung verbesserter Sensormodelle steigert die Übertragbarkeit synthetischer Daten auf die Realität. Gleichzeitig erhöhen die Methoden zur Integration neuer virtueller Welten die Datenvielfalt, die synthetisch erzeugt werden kann. Schlussendlich wurden so generierte synthetische Daten in anderen Arbeitspaketen zur Entwicklung weiterführender Ansätze verwendet.

3.1.2.4 AP1.4 Daten-Labeling

3.1.2.4.1 Zielsetzung

Das AP1.4 hatte sich zum Ziel gesetzt, die im Projekt KI-Delta Learning aufgenommenen Daten mit Hilfe der in TP2, TP3 und TP4 entstandenen Annotationsstrategien zu labeln und die Menge der Labels gleichzeitig zu nutzen, um die zur Verfügung stehenden Annotationsstrategien zu evaluieren. Die Evaluation sollte einerseits gegen die manuell definierten Label aus GUA3 durchgeführt werden. Andererseits sollten implizite Qualitätsmerkmale durch Ensemblestrategien abgeleitet werden. Darüber hinaus sollte durch die Evaluation der in TP2, TP3 und TP4 entwickelten Annotationsstrategien geeignete Ensembles für die jeweiligen

Domänen (z.B. Wettersituation oder Landschaftstyp) gefunden werden, um so die Qualität der automatisch erzeugten Label mit Hilfe der gegebenen Mittel maximieren zu können.

3.1.2.4.2 Erzielte Ergebnisse

Zu Beginn von AP1.4 wurden die zu evaluierenden Tasks aus der Menge der Interessens- und Aktivitätsgebiete der Partner gewählt. Zu diesen Tasks wurden in Abstimmung mit den Partnern Metriken zur Evaluation der Modelle gewählt. Auf Basis der technischen und rechtlichen Einschränkungen, die durch den Austausch von Software unter den Partnern entstehen, wurde in AP1.4 eine Strategie entwickelt, die beiden Problembereichen aus dem Weg geht. Konkret sollten die Partner ihre Annotationsstrategien in ihren Habitaten auf den im Projekt eingefahrenen Daten anwenden, sodass nur die inferierten Labels von den Partnern an AP1.4 weitergegeben werden hätten müssen, statt den Code weiterzugeben. Rückmeldungen seitens der Partner waren grundlegend positiv, jedoch fehlten am Ende die Ressourcen für dieses Unterfangen, da die im Projekt eingefahrenen Daten zu spät zur Verfügung standen.

Als Ersatz für die in AP1.4 anvisierten Ziele hat Valeo an einer Delta-Labeling Strategie für Bounding Boxes gearbeitet. Konkret interpoliert das implementierte Verfahren Bounding Boxes zwischen Key-Frames, um somit die Boxen in höherer Frequenz zu erhalten.

Porsche Eng. implementierte ein Verfahren zur Anonymisierung von Kameradaten, sodass das Training von Deep Learning Modellen weiterhin funktioniert, indem sensible Daten nicht verpixelt sondern mittels eines GAN durch alternative Daten des gleichen Typs (d.h. Gesichter bzw. Nummernschilder) ausgetauscht werden.

Das DLR hat einen "Greedy" Algorithmus entwickelt, der bei korrekter Bereitstellung von Metadaten (Straßenbelag, Wetterbedingungen, ...) während der Aufnahme bereits eine geeignete Teilmenge der sonst zu großen Menge an Aufnahmen auswählt, um diese für das manuelle Annotieren vorzuschlagen. Die so gewählte Teilmenge suggeriert nach den Erkenntnissen in TP3 bereits deutlich höhere Qualität von darauf trainierten Deep Learning Modellen. Der Algorithmus ist auf Basis der Erkenntnisse des DLR in AP3.3 entstanden.

3.1.2.4.3 Gegenüberstellung der vorgegebenen Ziele

Da die im Projekt aufgenommenen Daten deutlich zu spät kamen und dem AP1.4 zu wenig Annotationsstrategien aus TP2, TP3 und TP4 zu Verfügung standen, hat das AP1.4 die eigentlichen Ziele leider nicht erfüllen können. Dennoch wurden durch die Zielstrebigkeit von AP1.4 Zwischenergebnisse erreicht, die anderweitig wiederverwendet werden können. Dazu gehört u.a. die Evaluierung mehrerer Annotationsstrategien auf Daten, ohne eine Ground-Truth auf diesen Daten zur Verfügung stehen zu haben, sowie eine Vorauswahl von Daten auf Basis von Metadaten als Eingabe in ein anschließendes Active Learning Verfahren, das bereits von einer Verminderung der Menge der Daten bzgl. der Ausführungszeit profitiert.

3.1.3 Zielerreichung

Die Ziele in TP1 umfassten auf der einen Seite die Entwicklung neuer Verfahren zur effizienten, gezielten, datenschutzkonformen Erzeugung, Annotation und Auswertung von Trainingsdaten. Auf der anderen Seite sollten sowohl Real-, als auch Synthetische Datensätze anhand dieser Methodik erzeugt werden. Beide Ziele wurden erfolgreich erreicht.

Durch die Entwicklung des integrierten Datenmanagements in AP1.1 wurden eine datenschutzkonforme für alle Firmenpartner Plattform nutzbare geschaffen. Durch Abstraktion

und Automatisierung von Abläufen wurde ein effizienter Arbeitsprozess geschaffen, der viele Prozesse miteinander verknüpft.

Dieses wurde erfolgreich in AP1.2 bei der Realdatenaufnahme für die Datenauswahl und den Labelingprozess genutzt. Dort wurden außerdem eine Vielzahl von Prozessen geschaffen, die einen umfangreichen, hochqualitativen Datensatz ermöglichen. Durch ausführliche Planung konnten nahezu alle Deltas eingefahren werden und ausgeglichen in einem Datensatz repräsentiert werden. Darüber hinaus wurden Tools entwickelt, um eine einfache Bewertung jederzeit zu erlauben und ein iteratives Vorgehen zu ermöglichen.

AP1.3 hat geeignete Methodiken und Modelle entwickelt, um reale und virtuelle Daten miteinander zu kombinieren und daraus realitätsnahe synthetische Daten zu erzeugen. Die entsprechenden Lösungen erlauben ein hochgenaues Labeling und beinhalteten eine realitätsnahe virtuelle Sensorik. Dadurch lassen sich Realdatenaufnahmen sinnvoll und effizient ergänzen.

AP1.4 hat seine Ziele aus der VHB grundsätzlich erreicht. Durch die fehlenden Beiträge an Annotationsstrategien der Projektpartner, konnte das entwickelte Ensembleverfahren jedoch nicht angewendet werden. Stattdessen wurden die Inhalte dahingehend umdefiniert, dass eigene Annotationsstrategien entwickelt wurden, um so eine sinnvolle Alternative zu erreichen. Insgesamt konnte so trotzdem eine Bewertung von Labeling-Strategien entwickelt werden.

3.2 TP2: Transfer Learning

3.2.1 Ausgangssituation und Zielsetzung

Auf dem aktuellen Forschungsstand basierende Funktionen der „künstlichen Intelligenz“ (KI) können die Herausforderungen eines autonomen Fahrzeuges nur in einem sehr eingeschränkten Zielraum und anhand ineffizienter Trainingsstrukturen bedienen. Trotz der jüngsten Fortschritte im maschinellen Lernen, tendieren die verfügbaren Verfahren dazu, vorhandene Datensätze auswendig zu lernen und schon in leicht unterschiedlichen Szenarien unvorhersehbar zu reagieren.

Ziel des Vorhabens KI-Delta Learning war die Entwicklung von Methoden und Werkzeugen zur effizienten Erweiterung und Transformation vorhandener KI-Module autonomer Fahrzeuge auf die Herausforderungen neuer Domänen oder komplexerer Szenarien. Im Teilprojekt 2 wurden Verfahren entwickelt, die es erlauben, das bereits erworbene Wissen optimal zu nutzen und nur die zusätzlichen Anforderungen - das Delta - mit minimalem Aufwand neu zu lernen.

3.2.2 Strukturierung

Aus TP2 soll eine Verfahrenssammlung resultieren, die die Wiederverwendung von Daten und Labels ermöglicht. Zudem soll der Wert der vorhandenen Daten gesteigert werden, indem das Nutzungsspektrum deutlich erweitert wird. Damit wird auch die Notwendigkeit neuer annotierter Daten auf ein Minimum reduziert oder es kann bei den neuen Daten auf Labels verzichtet werden.

Konkret war AP2.1 in erster Linie eine Sammlung von Methoden, die für anwendungsgetriebenen Arbeitspakete AP2.2 bis AP2.5 benötigt werden, um die Domänenanpassung möglich zu machen.

In AP2.2 wurden synthetische Daten genutzt, um einerseits Experimente mit nicht vorhandenen Sensoren möglich zu machen und andererseits zu evaluieren, in wieweit es möglich ist, mit

synthetischen Daten Realdaten zu ersetzen. Das soll Kosten für Daten und Annotationsdaten sparen.

Der Fokus von AP2.3 war die Domänenadaption wie z.B. von Kamera 1 zu Kamera 2 und Lidar 1 zu Lidar 2 und spiegelt den im typischen Serienentwicklungsprozess üblichen Übergang von alter zu neuer Sensorik.

AP2.4 sollte Verfahren liefern, die mit größeren örtlichen und zeitlichen Domänenlücken zurechtkommen. Dieser Anwendungsfall ist besonders interessant, da die Verfahren beispielsweise mit Fahrzeugen zurechtkommen müssen, die es zum Zeitpunkt der Datenaufnahme noch nicht gab. Außerdem sollten sie die Erschließung neuer Ortsdomänen mit minimalem Aufwand für die Generierung und Annotierung von Daten ermöglichen.

Für AP2.5 wurden Verfahren geliefert, die unabhängig von Änderungen in der Umgebung funktionieren. Beispielsweise soll ein Verfahren, das auf während des Tages aufgenommenen Daten basiert, auch in der Nacht funktionieren. Gleiches gilt für das Funktionieren bei noch weiteren hinzukommenden Umgebungsparametern wie z.B. Wetter und Stau.

3.2.2.1 AP2.1 Kontinuierliches Lernen

3.2.2.1.1 Zielsetzung

Im Teilgebiet der KI namens Kontinuierliches Lernen werden Methoden entwickelt, um bereits trainierte KI-Module zu aktualisieren oder zu erweitern, ohne den Gesamtdatensatz neu zu trainieren. Dadurch kann effizient auf neue Datendomänen erweitert werden, auch wenn nur wenige Beispiele vorliegen. Die drei grundlegenden Fragestellungen des kontinuierlichen Lernens sind *Katastrophales Vergessen*, *Zwischenrepräsentationen* und *Kontinuierliches Training*. Katastrophales Vergessen bezieht sich auf die Situation, in der neuronale Netze vergessen, was sie gelernt haben, wenn sie über einen längeren Zeitraum mit einer begrenzten Anzahl an Trainingsbeispielen trainiert werden und diese Trainingsbeispiele wenig Ähnlichkeit mit den bisher gelernten Konzepten haben. Zwischenrepräsentationen ermöglichen es neuronalen Netzen, grundlegende Konzepte zu lernen und somit eine Vielzahl von (potenziell neuen) Aufgaben effizienter zu lösen. Kontinuierliches Training bezieht sich auf die besonderen Herausforderungen, die beim kontinuierlichen Lernen auftreten können, wie zum Beispiel Änderungen in Labeling-Spezifikationen über die Zeit.

3.2.2.1.2 Erzielte Ergebnisse im Einzelnen

Im Folgenden wird auf die erzielten Ergebnisse in den oben genannten thematischen Clustern eingegangen.

3.2.2.1.2.1 Katastrophales Vergessen

Ziel war es, kontinuierliche Lernmethoden hinsichtlich katastrophalem Vergessen für den Kontext von spezifischen Länder- und Domänenwechsel sowie für das Klasseninkrementelle Lernen zu quantifizieren und Verfahren zu entwickeln dieses zu mindern. Es wurden verschiedene Regularisierungsmethoden, Knowledge Distillation Verfahren und Replay-Verfahren evaluiert. Die Ergebnisse dieser Untersuchung legten offen, dass katastrophales Vergessen im Klassen-inkrementellen Lernen sich deutlich stärker ausprägt als im Domänen-inkrementellen Lernen und dass in den untersuchten Fällen auf Knowledge-Distillation basierende Verfahren die besten Ergebnisse erzielen. Bei den Regularisierungsbasierten Methoden stellte sich das als "Modelling the Background for Incremental Learning in Semantic Segmentation" vorgestellte

Verfahren bei mehreren Partnern als besonders vielversprechend dar. Weiterhin zeigten die Ergebnisse auch, dass kontinuierliche Lernmethoden sich in ihrer Wirkweise in Klassifikationsaufgaben und Semantische Segmentierung unterschieden. Um zu messen wieso und an welchen Stellen ein neuronales Netz von katastrophalem Vergessen betroffen ist, wurde eine Sammlung von Tools verwendet, die es ermöglichen, die Änderungen des Netzes im Lernprozess zu messen. Unter anderem wurden dafür Centered Kernel Alignment, Linear Mode Connectivity und eine angepasste Form von Layer Stitching verwendet, vgl. Abbildung 13.

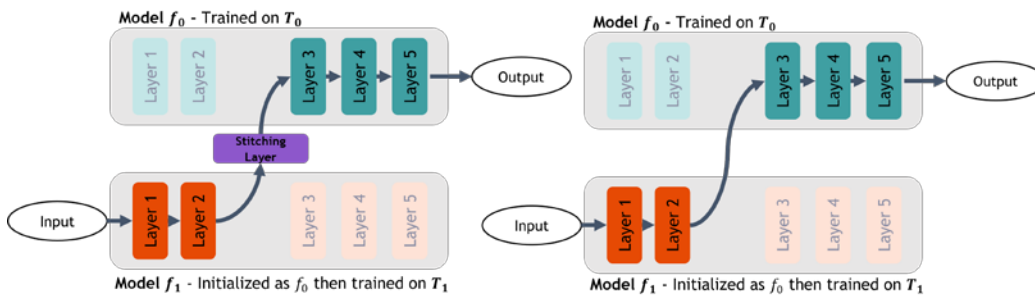


Abbildung 13: Layer Stitching mit und ohne Stitching Layer zum Messen von Änderungen im Neuronalen Netz

Mithilfe dieser Methoden wurde anschließend katastrophales Vergessen im Klassen-inkrementellen Lernen (AP3.4) und im Domänen-inkrementellen Lernen (AP2.1) untersucht. Die Ergebnisse für das Domänen-inkrementelle Lernen zeigten, dass katastrophales Vergessen sich besonders in dem ersten Convolutional Layer auswirkt, das nachhaltig auch die folgende Batch Normalization Layer beeinflusst. Folgeexperimente zeigten, dass der Effekt durch besondere Augmentierungen und/oder Vortraining auf ImageNet behoben wurde. Eine wahrscheinliche Erklärung ist, dass diese Netze allgemeinere Merkmale in den ersten Schichten aufweisen, die im kontinuierlichen Lernen dann nicht weiter überschrieben werden. Ein Vergleich unserer Trainingsmethode und der Baseline-Methode zeigt, dass unser angepasstes Training die Repräsentation im gesamten Netz stabilisiert und damit katastrophales Vergessen vermeidet.

Zusätzlich zu arbeiten in der Perzeption wurde katastrophales Vergessen auch im Kontext Reinforcement Learning zur direkten Fahrzeugsteuerung untersucht. Das entwickelte Verfahren führt zu robusteren Agenten und reduziert somit weiter das Vergessen.

3.2.2.1.2.2 Zwischenrepräsentationen

Auf maschinellem Lernen basierte Verfahren haben in der Regel eine drastische Leistungseinbuße, wenn sie mit Daten aus einer anderen Verteilung als während des Trainings konfrontiert werden. Wahrnehmung in einer sich ständig verändernden Umgebung wie der realen Welt ist daher immer noch eine große Herausforderung. Ein Beispiel für eine ständige Veränderung ist die Beleuchtung einer Szene, aber es kann auch Veränderungen in der aktuellen Wettersituation geben, wo die Umgebung innerhalb weniger Minuten komplett von Schnee oder Nebel bedeckt sein kann. Obwohl die Leistungseinbuße vermieden werden kann, indem man neue Daten aus den unbekanntem Domänen sammelt und manuell kennzeichnet, erfordert dies eine enorme Menge an Datensammlung und -kennzeichnung.

Um jedoch das manuelle Labeling zu umgehen und das Modell trotzdem an neue Domänen anzupassen, gibt es verschiedene Verfahren, für die die Partner folgende Taxonomie erstellten:



Different Task Definitions

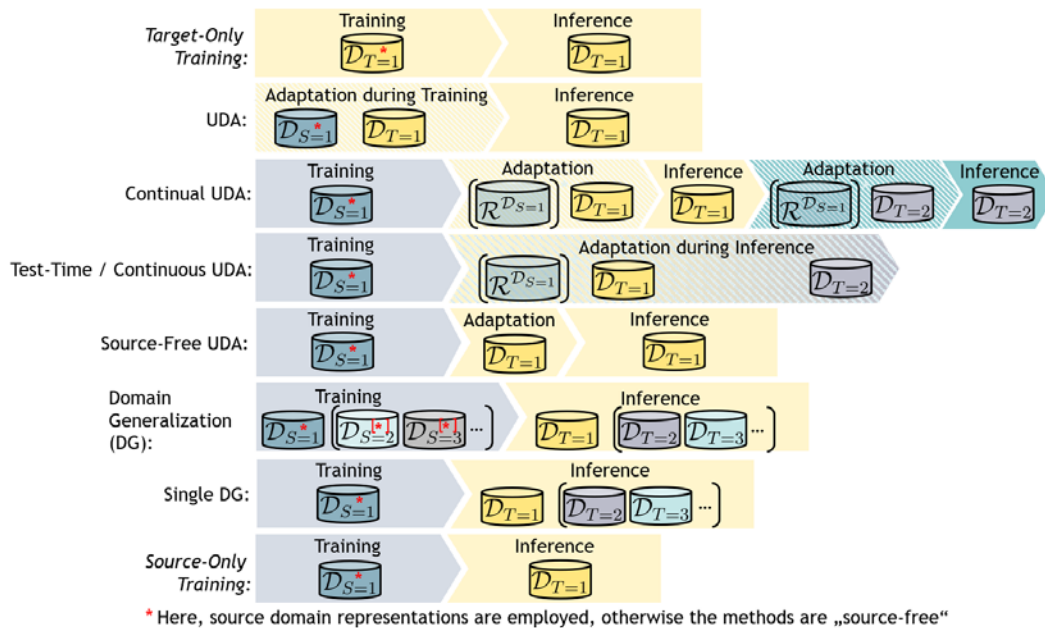


Abbildung 14: Im Projekt erarbeitete Klassifikation von Domänenadaptionsverfahren.

Ein Beispiel ist die unüberwachte Domänenanpassung (UDA). UDA-Methoden versuchen, Wissen von einer gelabelten Quell-Domäne auf eine ungelabelte Ziel-Domäne zu übertragen. Dies wird durch die Minderung der Diskrepanz zwischen der Quell- und Zielverteilung im Eingaberaum, dem Merkmalsraum, dem Ausgaberaum oder sogar in mehreren Räumen parallel erreicht. Das Ziel von UDA ist jedoch nicht ausreichend, da es die Leistung auf der Quell-Domäne vernachlässigt und sich nur auf die Ergebnisse der Ziel-Domäne konzentriert. Daher ist UDA besonders nützlich für das gut untersuchte Szenario von Synthetisch-zu-Realem, bei dem das Wahrnehmungssystem die synthetische Domäne nie wieder trifft. Im Kontext des automatisierten Fahrens kann jedoch die Umgebung auf die das Modell gerade angepasst wurde, zu einer zuvor gesehenen Ziel-Domäne oder zur Quell-Domäne zurückkehren. Deshalb untersuchten die Partner ebenfalls Verfahren zur test-time Domänenadaptation und Domänengeneralisierung. Erstere adaptieren das System kontinuierlich auf graduelle Veränderungen, Letztere lernen eine gegen Domänenänderung möglichst robuste Repräsentation. Die im Projekt entwickelten Verfahren führen zu besseren Adaptionsergebnissen, allerdings bleibt der Goldstandard weiterhin das Trainieren auf Daten der Ziel-Domäne.

3.2.2.1.2.3 Kontinuierliches Training

Beim Training über die Zeit treten verschiedene Probleme auf.

1. Einerseits kann sich die Aufgabe über die Zeit ändern, z.B. durch neu auftretende Unterklassen.
2. Andererseits kann sich auch die Datendomäne ändern, z.B. durch neue Mode auf den Straßen.

Im Folgenden werden untersuchte Ansätze zu beiden Problemstellungen vorgestellt.

3.2.2.1.2.3.1 Veränderung der Aufgabe über die Zeit

Bei der Entwicklung von Modellen für maschinelles Lernen besteht ein häufiges Problem darin, dass sich die Anforderungen an die zu erkennenden Kategorien ändern können. Entweder will man zusätzliche Klassen zu den bereits definierten Klassen oder deren Unterklassen erkennen. Im Projekt wurde der letztere Fall im Kontext der Bildsegmentierung untersucht. Da bei der Bildsegmentierung jedem Pixel eine Klasse zugeordnet werden muss, ist dies ein wahrscheinliches Szenario. Zum Beispiel möchte man die Klasse Mensch je nach Kontext entweder durch Person oder Fahrer ersetzen. Dies würde zusätzliche Beschriftungskosten verursachen, da alle Vorkommen von Mensch neu beschriftet werden müssten. Es wurde untersucht, wie eine gute Leistung für die neue Klassifikation erreicht werden kann, während so wenige Bilder wie möglich neu beschriftet werden müssen.

Der erfolgreichste untersuchte Ansatz verwendet einen modifizierten Kreuzentropieverlust, bei dem für Pixel, deren Unterklassen unbekannt sind, das Label der Superklasse eingeht, wobei deren Wahrscheinlichkeit durch die Summe der prädizierten Wahrscheinlichkeiten der Unterklassen berechnet wird. Zentrales Ergebnis war, dass bei der Verfeinerung der Labels ein kleiner Teil der neu gelabelten Bilder ausreicht, um eine fast identische Performance wie bei der Neubelabelung aller Bilder zu erzielen. Konkret reichte das Neulabeling von 50% aller Bilder in allen Fällen aus, um quasi identische Ergebnisse zu erhalten. Für viele untersuchte Fälle reichte es allerdings auch aus, 10% der Bilder neu zu labeln. Die Ergebnisse der Partner deuten darauf hin, dass je eindeutiger die Klassen sind (z. B. sind LKW und Bus leichter zu unterscheiden als Person und Fahrer), desto weniger Stichproben erforderlich sind um die neuen Klassen zu lernen.

3.2.2.1.2.3.2 Veränderung der Daten über die Zeit

Neben möglichen Änderungen in den Anforderungen an die zu erkennenden Kategorien, ist ein weiteres Problem der heutigen Netzwerkarchitekturen, dass sie anfällig sind für veränderte Umgebungsbedingungen. Dies äußert sich typischerweise in einem rapiden Abfall der Ergebnisse in der neuen Zieldomäne. In Anwendungen wie dem autonomen Fahren, ist es nun oftmals der Fall, dass sich die Zieldomäne nicht nur einmal ändert, sondern sich fortlaufend ändern kann. Da manche der Domänenänderungen zudem zyklisch sind, ist eine kontinuierliche Integration an neuem Domänenwissen zu dem bereits bestehenden Wissen von Vorteil.

Um zusätzlich den Annotationsaufwand möglichst gering zu halten, wurden für die Aufgabe der semantischen Segmentierung Methoden aus dem Bereich Continual Unsupervised Domain Adaptation untersucht. Diese Methoden zielen nicht nur auf gute Ergebnisse in der aktuellen Zieldomäne ab, sondern betrachten auch die Ergebnisse auf den zuvor gelernten Zieldomänen. In den Anfangsanalysen stellte sich heraus, dass Ansätze die auf simplem „Adaptive Instance Normalization“ (AdaIN) aufbauen, für reale Daten kaum besser funktionieren als gängige Datenerweiterungstechniken. Der Grund hierfür ist, dass durch den AdaIN basierten Stiltransfer nicht nur Artefakte erzeugt werden, sondern auch die klassenspezifischen Unterschiede nicht richtig eingefangen werden und lediglich der dominierende Stil in einem Zielbild kopiert wird. Durch das Beheben dieser Probleme, konnte schließlich eine Methodik entwickelt werden, welche für synthetische und reale Daten deutliche Vorteile zeigt. Ein paar Beispielbilder, die einmal mit AdaIN und einmal mit unserem Ansatz transferiert wurden, sind in Abbildung *Abb-STGT-1* zu sehen.

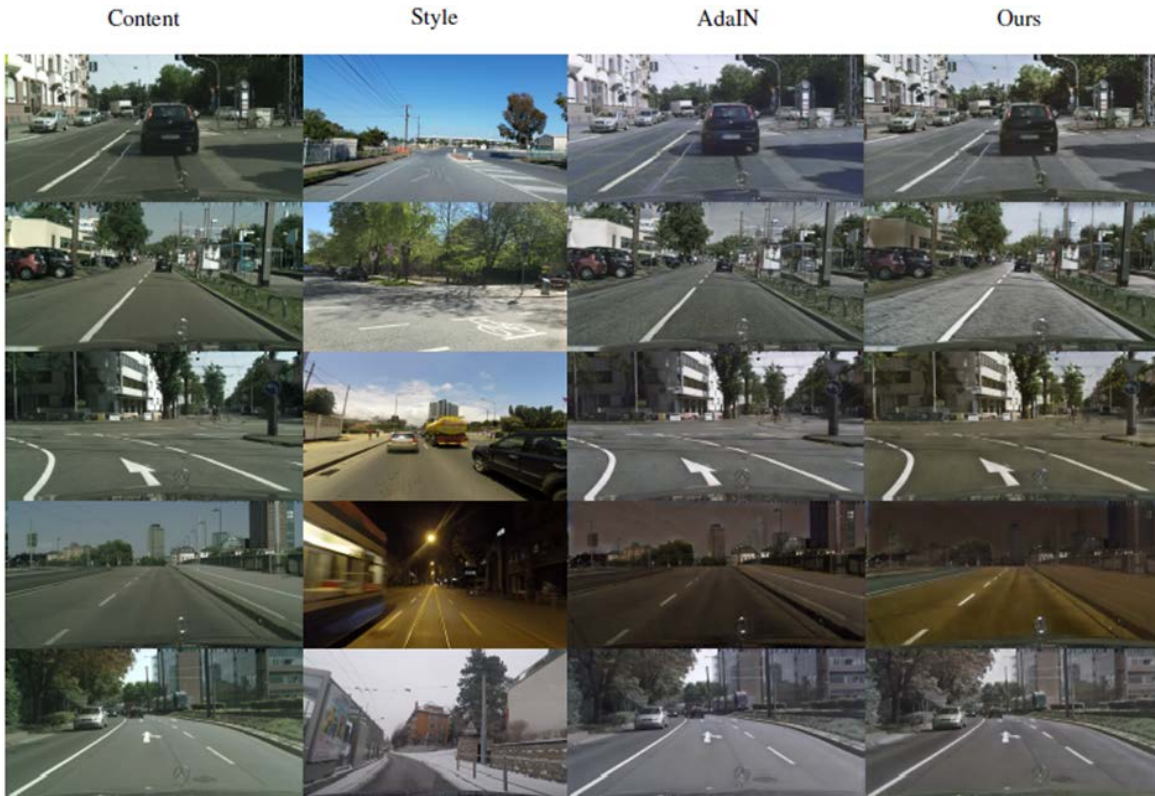


Abbildung 15: Illustration einiger Beispielbilder , die einmal mit AdaIN und einmal mit dem Ansatz der Universität Stuttgart transferiert wurden.

In weiteren Analysen stellte sich zudem heraus, dass die entwickelte Methodik ebenfalls deutlich besser funktioniert als adversarial learning basierte Ansätze.

Während die zuvor beschriebenen Techniken die Adaptierung des Modells an die neuen Umgebungsbedingungen offline und mit einem fixen Zieldatensatz durchführen, kann das Modell auch online direkt während der Testzeit angepasst werden. Damit wird sichergestellt, dass das Modell an die Domänenänderung angepasst wird, die es während der Testzeit tatsächlich erfährt. Um ein Modell direkt während der Testzeit anzupassen, wurden zwei Ansätze entwickelt. Ersterer ist eine Erweiterung der zuvor beschriebenen Methodik und nutzt Stiltransfer für die Adaptierung während der Testzeit. Auch in diesem Fall führte der entwickelte Ansatz zu signifikant besseren Ergebnissen als die Vergleichsmethoden. Der zweite Ansatz beruht auf einem robusten Mean Teacher. In Experimenten für die Aufgabe der Klassifikation zeigte sich, dass der entwickelte Mean Teacher Ansatz nicht nur besonders gut für sich graduell entwickelnde Domänen geeignet ist, sondern auch für sukzessiv schwerer werdende Domänenänderungen.

3.2.2.1.2.4 Gegenüberstellung der vorgegebenen Ziele

Alle Partner haben ihre Ziele über die Unterarbeitspakete in diesem Projekt abgearbeitet. Kein Partner hat sich Ziele gesteckt, die er nicht erreichen konnte. Wie in einer PI-Planung üblich haben die Partner die Möglichkeit genutzt Unterarbeitspakete, die aufgrund der vorangegangenen Experimenten uninteressant wurden, neu zu planen, um ein besseres Gesamtergebnis zu erzielen. Darüber hinaus haben Partner durch den regen Austausch weitere Erkenntnisse gewonnen.

3.2.2.2 AP2.2 Nutzung synthetischer Daten

3.2.2.2.1 Zielsetzung

Synthetische Daten werden gerne genutzt, um a) in der Realität schwer oder unmöglich zu erzeugende Daten zu ersetzen, wie z.B. potenzielle Unfallsituationen, oder b) Ground Truth Label zu generieren, die sonst nur mit hohem Aufwand zu gewinnen sind. Jedoch existiert ein Delta zwischen synthetischen Daten und Realdaten, wodurch die Leistungsfähigkeit von Modellen, die auf oder mit synthetischen Daten trainiert wurden, im Allgemeinen nicht übertragbar ist. In diesem Arbeitspaket sollen Methoden zur Überbrückung dieses Deltas entwickelt werden. Ziel ist es, dass synthetische Daten verwendet werden können, um die Leistungsfähigkeit von Modellen auf Realdaten zu verbessern. Aus diesem Grund werden in diesem AP die Methoden entwickelt, welche die Anwendung der gezielt erzeugten synthetischen Daten unter Berücksichtigung des Domainshift ermöglichen.

Zudem lassen sich anhand simulierter Daten Untersuchungen bezüglich des domain gaps unter kontrollierten und reproduzierbaren Bedingungen durchführen. So kann beispielsweise der Einfluss verschiedener (simulierter) Sensoren, Wetterbedingungen usw. auf der gleichen Szene untersucht werden. Zudem soll der domain gap zwischen den synthetischen und den realen Daten mit dem besonderen Fokus daraufhin untersucht werden, wo in den neuronalen Netzwerken domänenspezifisches Wissen gespeichert ist.

3.2.2.2.2 Erzielte Ergebnisse im Einzelnen

3.2.2.2.2.1 Generierung von sicherheitskritischen Szenarien durch differenzierbare Simulation

Wie bei Angriffen auf den Bildraum beobachtet wurde, hat die gradientenbasierte Optimierung das Potenzial, schneller und erfolgreicher zu sein als BBO. Darüber hinaus gibt es einen Trend zur durchgängigen Differenzierbarkeit, sowohl in der Simulation als auch bei den Antriebsagenten. Die Verwendung differenzierbarer Komponenten ermöglicht die gradientenbasierte Generierung von adversen Verkehrsszenarien. In diesem Beitrag beantworten wir eine wichtige Frage: Muss die gesamte Simulationspipeline differenzierbar sein, um nützliche Gradienten für die Optimierung von Verkehrsszenarien zu liefern? Wir präsentieren KING, einen einfachen und effektiven Ansatz für die Generierung sicherheitskritischer Szenarien. Unsere Kernidee besteht darin, ein kinematisches Fahrradmodell als Stellvertreter für die tatsächliche Dynamik eines Fahrsimulators zu verwenden und sicherheitskritische Störungen von unkritischen Ausgangsszenarien über Backpropagation zu lösen. Des Weiteren zeigen wir, dass KING anspruchsvolle, aber lösbare Testfälle für Fahrsysteme generiert, die sowohl (1) einen Planer, der auf einem Raster aus der Vogelperspektive (BEV) agiert, als auch (2) einen Kamera- und LiDAR-basierten Fahrgenten verwenden. Schließlich zeigen wir, dass die von KING generierten Szenarien die ursprüngliche Trainingsverteilung, die nur eine begrenzte Vielfalt aufweist, ergänzen können. Dies führt zu einer verbesserten Kollisionsvermeidung.

3.2.2.2.2.2 Vorhersage menschlicher Bewegungen: Von der Simulation zur realen Welt

Im Vergleich zu öffentlich verfügbaren Datensätzen haben wir uns darauf konzentriert, einen Datensatz zu erstellen, der speziell auf die Anwendungsfälle im Zusammenhang mit autonomen Fahraufgaben zugeschnitten ist. Obwohl CARLA ein umfangreiches Sensorenset für die Umgebungswahrnehmung bietet, ist die Interaktion zwischen verschiedenen Fußgängern und ihre Fähigkeit, realistische Aktionen auszuführen, im Vergleich zu modernen Videospielen wie GTA-V ziemlich eingeschränkt. Einfache Interaktionen mit der Umwelt, wie z.B. das Hinsetzen auf eine Bank an einer Bushaltestelle, sind in CARLA nicht mit überschaubarem Aufwand zu animieren

und konnten daher nicht modelliert werden. Darüber hinaus haben wir festgestellt, dass die Gehstile der Fußgänger weniger vielfältig sind als in GTA-IM. Nicht zuletzt ist unser aktueller Datensatz etwa zwei Größenordnungen kleiner als JTA. Dieses Problem hängt mit der Wiederholungsfunktion in CARLA zusammen, bei der wir während der Datenerstellung auf mehrere Probleme gestoßen sind.

3.2.2.2.3 Anpassung von synthetischen an reale Bereiche

Für beide Ansätze gibt es noch viele offene Fragen. Das Self-Training mit Metaseg scheint noch ein gewisses Leistungspotenzial zu haben, da die Qualität der Pseudo-Etiketten sehr gut ist. Allerdings ist das Trainieren von quellenfreien Pseudo-Labels eine große Herausforderung und wir entwickeln neue Methoden, um die Leistung noch weiter zu steigern.

Auch die Zufallsauswahl muss weiter erforscht werden, da es bemerkenswert ist, dass eine so kleine Anzahl von Bildern eine so signifikante Leistungssteigerung bewirken kann, so dass eine tiefergehende Analyse sehr nützlich wäre. Interessant wäre auch die Frage, ob nur die eigentlichen Zieldaten einen solchen Zuwachs bewirken oder auch andere reale Daten aus einem anderen Bereich wie z.B. A2D2.

3.2.2.2.4 Simulation-Real-Anpassung

Um unseren Ansatz mit RegDA zu vergleichen, haben wir ihn in die Transfer-Learning-Bibliothek integriert (unsere Integration ist noch nicht öffentlich zugänglich). Wir folgen dem Trainingsverfahren von RegDA, das davon ausgeht, dass eine Epoche nicht eine Iteration über den gesamten Datensatz ist, sondern aus 500 Batches besteht. Für das Training von RegDA wird eine Batchgröße von 32 und für unseren AuxCycleGAN-Ansatz eine Batchgröße von 8 verwendet. Bei der Domänenanpassung unseres Datensatzes der Hochschule Reutlingen (siehe vorherige Ergebnisse und AP1.3 für Details) von synthetischen Daten mit einem Schauspieler in Freizeitkleidung zu synthetischen Daten mit einem Schauspieler in einem Motion-Capture-Anzug ist unser AuxCycleGAN-Ansatz in der Lage, RegDA zu übertreffen und liegt nahe an der Leistung der Quelldomäne. Wir haben festgestellt, dass durch den Wechsel von einer linear abfallenden zu einer zyklischen Lernrate, die zwischen einer oberen und einer unteren Schranke oszilliert, die Domänenanpassungsleistung sowie die Trainingszeit verbessert werden können. Während CycleGAN eine linear abfallende Lernrate verwendet, zeigt die Ablationsstudie von AuxCycleGAN ohne Hilfsaufgabe, die im Grunde CycleGAN mit einer zyklischen Lernrate ist, eine bessere Leistung. AuxCycleGAN schneidet sogar noch besser ab und unterstützt unseren auf Hilfsaufgaben basierenden Ansatz für den LSP-Datensatz. Unser Ansatz liegt in diesem Szenario hinter RegDA zurück, schneidet aber immer noch besser ab als CycleGAN. Wir trainieren oder modifizieren das Aufgabenmodell für die menschliche Posenschätzung nicht neu, sondern führen nur eine Sensoranpassung in die durch das Aufgabenmodell bekannte Domäne durch. Die Posekonfigurationen der Zieldomäne sind möglicherweise nicht durch die Posen der Quelldomäne abgedeckt und werden daher nicht vom Aufgabenmodell gelernt. Im Gegensatz dazu passt RegDA das Aufgabenmodell an die Zieldomäne an. Für dieses Szenario sind weitere Analysen erforderlich.

3.2.2.2.5 Rare-Event-Simulation-basierte Fehleridentifizierung

Das DLR hat sich in AP 2.2 mit dem simulativen Testen von KI-Systemen beschäftigt. Hierbei wurde eine Methode zur automatisierten Identifizierung kritischer Simulationsparameter entwickelt, bei denen die KI besonders fehleranfällig ist. Die Grundidee ist hierbei, dass Optimierungsalgorithmen verwendet werden, um kritische (und oftmals seltene) Ereignisse

innerhalb der Simulation häufiger sichtbar zu machen. Das grundlegende Verfahren der Methodik ist in Abbildung 16 dargestellt.

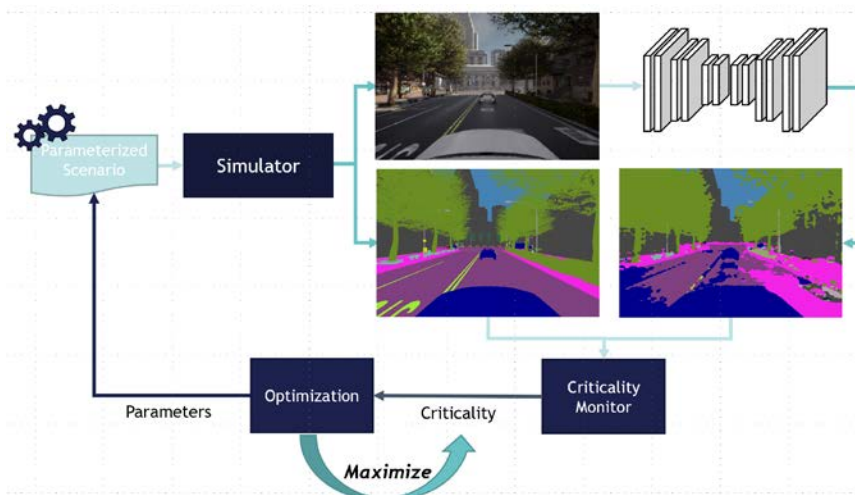


Abbildung 16: Übersicht über die Rare-Event-Simulation-basierte Fehleridentifizierung

Bei diesem Verfahren wird innerhalb einer Schleife ein parametrisiertes Szenario wiederholt abgespielt und die Kritikalität des jeweiligen Simulationslaufes gemessen. Basierend auf dieser Kritikalität erzeugen die Optimierungsalgorithmen ein neues Set an Input-Parametern für das Szenario, wodurch ein neuer Simulationslauf erzeugt wird. Mit jedem Lauf dieser Schleife erhalten die Optimierungsalgorithmen mehr Informationen über das System und können mit höherer Wahrscheinlichkeit einen kritischen Simulationslauf erzeugen, bei dem die KI viele Fehler macht.

Diese kritischen Testläufe können dann genutzt werden, um zu identifizieren, wo die KI noch fehleranfällig ist, und so zur Verbesserung der Robustheit beitragen.

Die Methode wurde in AP 2.2 entwickelt und in AP 2.4 auf dem Anwendungsfall Wetter evaluiert. Dort wurde die Methode dann genutzt, um bestimmte Wettersituationen zu finden, die zu einer hohen Anzahl von Fehlklassifikationen führen. Die Ergebnisse aus dieser Studie können in Deliverable E4.2.2.4 gefunden werden.

3.2.2.2.6 Verstärkte Instanzsegmentierung mit synthetischen Daten

Zusammenfassend lässt sich sagen, dass synthetische Daten vielversprechende Eigenschaften aufweisen, um die Leistung von DNNs in der realen Welt der Bildinferenz zu verbessern. Sie sind realistisch, vielfältig und können gelernt werden. Obwohl synthetische Daten noch nicht in der Lage sind, reale Daten für das Training vollständig zu ersetzen, können sie einen Mangel an realen Daten leicht überwinden und zur Verbesserung der Leistung neuronaler Netze verwendet werden. In dieser Arbeit heben wir die Vorteile der Vorverarbeitung synthetischer Bilder vor ihrer Verwendung hervor. Durch die Entfernung der nicht verwertbaren Instanzen aus den synthetischen Daten wird der Rechenaufwand verringert und die Genauigkeit sowohl bei synthetischen als auch bei realen Bildern verbessert. Wir stellen hauptsächlich zwei Strategien vor, die synthetische und reale Bilder kombinieren. Ein erster Ansatz besteht im Vortraining auf synthetischen Daten und der Feinabstimmung auf realen Daten. Diese Methode überwindet die derzeitige Leistung des PointRend-Moduls bei Stadtansichten, wobei die beiden Datensätze getrennt bleiben. Die Steuerung der Feinabstimmungsphase durch konsistente Hyperparameter

sorgt für eine noch bessere Leistung. Der zweite Ansatz besteht darin, einen gemischten Datensatz mit synthetischen und realen Bildern zusammenzustellen. Um die beste Leistung zu erzielen, müssen die realen Bilder vollständig einbezogen werden, während die Menge der zusätzlichen synthetischen Bilder an die zulässige Anzahl der Iterationen angepasst werden sollte. Die Methode der Datensatzmischung übertrifft auch die derzeitige Leistung des PointRend-Moduls und übertrifft sogar den Ansatz der Feinabstimmung. Ein Teil dieser Ergebnisse dürfte von der Wahl des Netzwerks, der durchgeführten Aufgabe oder der beiden Datensätze und ihrer Verschiebung abhängen. Viele Elemente dieser Studie, wie die verschiedenen vorgestellten Strategien und die Gesamtmethodik, könnten jedoch auch für andere Anwendungsfälle mit synthetischen Daten nützlich sein.

3.2.2.2.7 Auswirkung der Quelldatensätze auf die Leistung der Bereichsanpassung

In diesem AP haben wir untersucht, inwieweit die Menge der synthetischen Trainingsdaten maximiert werden sollte. Bei leistungsfähigen Simulationsumgebungen ist es oft naheliegend, eine riesige Menge von Proben zu erzeugen, um die bestmögliche Leistung zu erzielen. Experimentelle Ergebnisse aus diesem AP und aus der Literatur zeigen jedoch, dass nicht nur die Menge, sondern auch die Übereinstimmung mit der erwarteten Zieldomäne von großer Bedeutung ist. Dies bestätigt erneut den Bedarf an Metriken, die es erlauben, den Abstand zwischen zwei Domänen zu messen, um geeignete Bilder für das Training zu erzeugen oder auszuwählen.

3.2.2.2.8 VoxGRAF: Schnelle 3D-bewusste Bildsynthese mit spärlichen Voxel-Gittern

In dieser Arbeit untersuchen wir spärliche Voxel-Gitter als Repräsentation für 3D-bewusste Bildsynthese. Wir stellen fest, dass der Schlüssel zur Erzeugung spärlicher Voxel-Gitter in der Kombination von progressivem Wachstum, Beschneidung und Regularisierung liegt, um eine scharfe Oberfläche zu erzeugen, die effizient gerendert werden kann. Unser Ansatz übertrifft alle Methoden, die keinen neuronalen Renderer verwenden. Anstatt das neuronale Rendering ganz zu verwerfen, finden wir es vorteilhaft, ein flaches CNN für die Verfeinerung zu verwenden. Dieses CNN kann die Verzerrung von Datensätzen modellieren, ist aber deutlich schwächer als Standardansätze für neuronales Rendering, die Feature Maps mit niedriger Auflösung hochrechnen. Unser Ansatz kann den Abstand zu Modellen, die stark auf neuronalem Rendering aufbauen, verringern, doch bleibt ein Kompromiss zwischen 3D-Konsistenz und Bildtreue bestehen. Die Frage, ob ein gewisses Maß an neuronalem Rendering erforderlich ist, um die beste Leistung zu erzielen, ist ein wichtiger Aspekt für die zukünftige Forschung. Schließlich hängt die Geschwindigkeit unserer Methode von der Spärlichkeit der modellierten Szene ab. Daher werden sich die Rendering-Zeiten bei komplexeren Datensätzen als den in der Literatur üblichen wahrscheinlich erhöhen.

3.2.2.2.9 Übertragung von Bereichen auf einen benutzerdefinierten Datensatz unter Verwendung semantischer Konsistenz

Wie in unseren Experimenten zu sehen ist, erhöht die semantische Konsistenz die Modellgenauigkeit, ist aber im Vergleich zu den ursprünglichen CARLA-Frames und Cityscapes immer noch unterdurchschnittlich. Unser Ziel für die Zukunft ist es, die Original-CARLA-Frames zu übertreffen und näher an die Leistung von Cityscapes heranzukommen. Zu diesem Zweck versuchen wir, kleine Bildausschnitte aus den Quell- und Zielbildern zu extrahieren und diese, eingeschränkt auf die L2-Norm, miteinander zu kombinieren. Dies führt zu ähnlichen Bildausschnitten, die als Input für ein GAN verwendet werden können, was die Unterscheidungsaufgabe erheblich erschwert und somit zu realistischeren Bildern führt.

3.2.2.2.10 Wavelet-basierte unüberwachte Label-to-Image-Übersetzung

In dieser AP wurde ein Rahmen für die semantische Bildsynthese in einer ungepaarten Umgebung vorgeschlagen (USIS). Dabei wird ein neuartiger WaveletSPADE-Generator zusammen mit einem UNet und einem unbedingten, auf Wavelet basierenden Ganzbild-Diskriminator eingesetzt. Das UNet fördert die Klassentrennbarkeit und den Erhalt des Inhalts, während der Diskriminator die Farb- und Texturverteilung realer Bilder berücksichtigt. Die Wirksamkeit des vorgeschlagenen Rahmens für die semantische Bildsynthese wurde anhand von 3 anspruchsvollen Datensätzen gezeigt: Stadtansichten, ADE20K und Cocosuff. USIS übertraf frühere ungepaarte GANs und näherte sich der Leistung von überwachten Systemen an.

3.2.2.2.3 Gegenüberstellung der vorgegebenen Ziele

Alle Partner haben ihre Ziele über die Unterarbeitspakete in diesem Projekt selbst vorgegeben und entsprechend abgearbeitet. Aufgrund kluger Planung hat kein Partner sich Ziele gesteckt, die er nicht erreichen konnte. Wie in einer PI-Planung üblich haben die Partner die Möglichkeit genutzt Unterarbeitspakete, die aufgrund der vorangegangenen Experimenten uninteressant wurden, neu zu planen, um ein besseres Gesamtergebnis zu erzielen. Darüber hinaus haben Partner durch den regen Austausch weitere Erkenntnisse gewonnen und damit die vorgegebenen Ziele übertroffen.

3.2.2.3 AP2.3 Cross-Sensor-Anpassung

3.2.2.3.1 Zielsetzung

Die kontinuierliche Entwicklung autonomer Fahrzeuge bedingt, dass sich Sensorgenerationen von einer zur nächsten in Eigenschaften wie zum Beispiel Rauschverhalten, Öffnungswinkel, Auflösung stark unterscheiden können. Die Änderungen dieser Eigenschaften haben großen Einfluss auf die Ergebnisse der KI-Module. Das führt dazu, dass man aufgenommene Daten und Annotationsdaten, welche für einen Sensor erstellt wurden, nicht mehr für die Folgegeneration des Sensors nutzen kann. Daher werden in diesem AP Verfahren des Transfer Learnings erforscht, welche die Überbrückung dieses Deltas unterstützen, was auch ein Delta zwischen unterschiedlichen Sensortypen (wie z.B. zwischen Video und Lidar) sein kann.

Um dies zu bewerkstelligen, werden Methoden aus dem Bereich des kontinuierlichen Lernens verwendet und spezialisiert, beziehungsweise wurden Methoden genutzt, die mit unterschiedlichen Domänen gleichzeitig zurechtkommen. Wesentlich dabei ist, dass das relevante Wissen aus der Quelldomäne über eine passende Zwischenrepräsentation in die Zieldomäne überführt wird. Die Zwischenrepräsentation kann auch eine Sammlung von Wissensaspekten aus der Quelldomäne sein, die in der Zieldomäne rekombiniert werden. Das Vorgehen kann in einem Schritt durchgeführt (End-To-End), aber auch in mehrere Unterschritte aufgeteilt werden.

Der wichtigste Arbeitsanteil ist aber das Verständnis der Deltas, die als Gütemaße, Metriken und Verlustfunktionen in den Verfahren vorkommen. Nach dem Verständnis kommt die mathematische Beschreibung. Kann man ein Delta beispielsweise in eine Verlustfunktion für ein Neuronales Netz fassen, so kann dieses im Lernprozess minimiert werden.

3.2.2.3.2 Erzielte Ergebnisse im Einzelnen

Die Partner haben sich in diesem Arbeitspaket auf Video→Video, Lidar→Lidar, Radar→Radar und multimodale Transfer-Lernmethoden und der Analyse dieser Methoden konzentriert. Die Partner haben in vielfältiger Weise zu den Ergebnissen von APs beigetragen.

Aufgaben waren Objekterkennung, semantische Segmentierung, Tiefenschätzung und Klassifikation. Domänenlücken sind unterschiedliche Sensortypen, Auflösung, Öffnungswinkel und Sensor-Einbauposition. Zur Reduzierung von Domänenlücken werden GANs, physikalische Transformationen, Datenerweiterung, Wissensgraphen und spezielle Verlustfunktionen verwendet bzw. kombiniert.

Eine genaue Übersicht über die Arbeiten, die z.B. in Projektinkrement I behandelt wurden, mit der Methodik, den Gütemaßen, den genutzten Datensätzen, den behandelten Deltas wie auch den Anwendungsfällen, die zu Ergebnis E1.2.2.3 beigetragen haben zeigt ganz konkret die folgende Tabelle.

Delta-Use-Case(s) / Setting	Application Use-Case(s)	Dataset(s)	Metric(s)	Loss(es)	Training / Setting	Used (baseline) model(s)	Delta reduction method(s)
Camera1 (old) → Camera 2 (new)	Semantic Segmentation	Two internal cityscapes like data sets	(Delta-)IoU	Cross entropy	Supervised / Unsupervised Domain Adaption (UDA)	Small internal network / Change to more standard network is planned, if KI-DL- Dataset is available	Data augmentation, initial rescale and crop
Lidar 1 (lowres) → Lidar 2 (highres)	3D Object Detection	Internal multimodal dataset with 4 Lidar sensors	spatial analysis of recall, qualitative	Focal Loss, smooth L1 loss	Supervised	PIXOR-like single-stage 3D object detector	tbd, e.g., gradient reversal, aux loss
Camera1 (old) → Camera 2 (new)	Object Recognition	CIFAR-10 → STL-10	Accuracy	Contrastive Loss	Supervised	ResNet-50 with cross-entropy	Joint Embedding Space
Lidar 1 (lowres) → Lidar 2 (highres)	Point cloud upsampling Semantic Segmentation	SemanticKITTI	Chamfer Distance, EMD	Chamfer Distance, EMD	Supervised	Caccia et al.; Rangenet++; PointNet-alike	
Lidar 1 (lowres) → Lidar 2 (highres)	Semantic Segmentation (SemSeg)	nuScenes-lidarseg, SemanticKITTI	<ul style="list-style-type: none"> Domain Adaptation: qualitative SemSeg: IoU 	<ul style="list-style-type: none"> Domain Adaptation: adversarial loss + cycle Consistency loss SemSeg: multiclass softmax cross-entropy 	<ul style="list-style-type: none"> Domain Adaptation: GAN approach SemSeg: supervised 	<i>not applicable to this eval</i> in future we will use: LiLaNet, (RangeNet)	appearance adaptation based on CycleGAN

Delta-Use-Case(s) / Setting	Application Use-Case(s)	Dataset(s)	Metric(s)	Loss(es)	Training / Setting	Used (baseline) model(s)	Delta reduction method(s)
Lidar domain shift analysis with different mounting position especially vertically	Point cloud transformation analysis between sensors in different mounting height	Lyft Lvl 5	not applicable for this eval. (density analysis gmm likelihood)	not applicable for this eval.	not applicable for this eval.	not applicable for this eval.	not applicable for this eval.
Camera □ Camera, LiDAR □ LiDAR	Semantic Segmentation, Point cloud transformation	GTA5, SYNTHIA, Cityscapes, CARLA, KITTI, KITTI360	IoU Earth Mover Distance (EMD) Chamfer Distance (CD)	Contrastive Loss, CE CD-EMD	Unsupervised	DeepLabv2 with ResNet-101 PU-NET PU-GAN CycleGAN	Contrastive method conditional GANs
Camera □ LiDAR	End-to-End Driving	CARLA	Driving score, route completion (CARLA Leaderboard metrics)	L1 loss	Supervised	CILRS, LBC, Late Fusion, Geometric Fusion	Attention for integrating representations from different modalities
Camera One (low mounting position) → Camera Two (high mounting position)	Object Detection	Internal ZF Truck Front Facing Camera Dataset, implicitly KITTI Dataset for baseline model training	Qualitative	not applicable for this evaluation yet.	Supervised	faster_rcnn_resnet101_kitti	not applicable for this approach yet.

Delta-Use-Case(s) / Setting	Application Use-Case(s)	Dataset(s)	Metric(s)	Loss(es)	Training / Setting	Used (baseline) model(s)	Delta reduction method(s)
Camera 1 – Camera 2 (RGB, depth to be decided)	Human Pose Estimation (top-down-approach, therefore Object Detection will be investigated later)	Custom motion capture dataset	PCKh, Average Precision	MSE, MAE	(currently) Unsupervised DA	Simple Baselines for Human Pose Estimation (and Tracking) https://github.com/Microsoft/human-pose-estimation.pytorch	CycleGAN-based

Tabelle 2: Übersicht über die Arbeiten, die z.B. in Projektinkrement I behandelt wurden

Zudem gab es rege Teilnahme an diversen Arbeitsgruppen, wie WG Labeling, WG Sensor Specification, WG Requirements Analysis for Data Management und WG Route Definition. Diese Arbeitsgruppen waren rein von den Interessen der Partner getrieben und zeigen den guten Austausch der Partner bzgl. den gegebenen Themen.

Im weiteren Projektverlauf wurden Wissensgraphen verwendet um KI-Verfahren auch mit Zusatzdaten verbessern zu können, die nicht direkt in Bild und Annotationsdaten enthalten sind. Es wurde eine Ablationsstudie gemacht um den Einfluss von zusätzlichen Zieldomänenannotationsdaten auf den Domänentransfer abschätzen zu können. Für den Auflösungsdomänentransfer wurde ein Lidar-Verfahren entwickelt, welches zwischen niedriger und hoher Auflösung transferiert. Im Digitallabor wurde ein Detektionsalgorithmus entwickelt, der mit Hilfe von synthetischen Daten trainiert wurde, aber auch auf realen Daten funktioniert. Durch die Nutzung moderner Netzstrukturen wurde mit Hilfe einer multimodalen Fusion ein Verfahren entwickelt, welches in einem Netz Kamera und Lidar-Daten fusioniert. Aufgabe war es Objekte wie Ampeln und Fahrzeuge zu erkennen. Um auf-akkumulierte Lidarpunktwolken zu verstehen wurden 2D und 3D-Darstellungen entwickelt, die die Analyse solcher Daten möglich machen. Verfahren für die Semantische Segmentierung wurden verbessert, indem im Training kontrastives Lernen verwendet wurde. Es wurden eine Domänenlückenanalyse zwischen unterschiedlichen Kamerapositionen durchgeführt. Um alte Lidar-Daten weiter nutzen zu können wurden Cycle-GANs verwendet. Aus den Ergebnissen dieser Experimente ergaben sich neue Zielsetzungen für die folgenden Unterarbeitspakete.

Es wurden Wissensgraphen im Falle des Sensorwechsels untersucht, wozu auch Code beigesteuert wurde. Für 3D-Box-Detection wurde eine Sensor-Lückenanalyse durchgeführt. Das Hochskalieren von Lidar-Daten wurde in synthetischer Weise untersucht, durch das Weglassen von Scanlinien. Die Domänenadaption für die menschliche Posenanalyse wurde verbessert. Die Lidar zu Lidar Domänenadaption wurde verbessert. Es wurde analysiert, wie performant die Kanalinterpolation von Radardaten ist. Es gab eine Publikation über die Frequenzverzerrung generativer Modelle. Es wurde die Bereichsverschiebung zwischen Lidar-Sensoren in verschiedenen Montagepositionen untersucht, insbesondere die Veränderungen, die diese Bereichsverschiebung bei den Eingabedaten für neuronale Netze hervorruft. Untersucht wurde auch das Vorhandensein und die Abschwächung der Domänenlücke, die entsteht, wenn Kameras in unterschiedlichen Positionen und Ausrichtungen montiert werden. Genauer gesagt wurde die Leistungslücke untersucht, die bei der Erkennung von Objekten aus verschiedenen Kamerapositionen entsteht.

Aufbauen auf die vorangegangenen Untersuchungen fand eine Untersuchung der Domänenlücke in Lidar-Objektdetektionsnetzwerken statt, woraus auch eine Publikation und ein Verfahren entstanden ist. Es wurden sensorübergreifende Verfahren zum Simulation-zu-Real-Transfer entwickelt und evaluiert. Weiter hat ein Partner eine Bewertung von Objektvorschlägen an ungesehenen Objekten durchgeführt, was einen tieferen Einblick bzgl. der Generalisierung von KI-Verfahren möglich macht. Eines der größten Probleme im KI-Bereich ist die multimodale Fusion, die von einem Partner mit Hilfe modernster Netzwerkstrukturen und auf Netzwerkebene angegangen wurde. Das Ziel eines Partners war es den Einfluss der Domänen-Verschiebung zwischen Verkehrsteilnehmer auf der Ebene der 3D-Punktwolke zu quantifizieren. Für diese Aufgabe konzentriert dieser sich hauptsächlich auf Autos als die Zielart von Interesse. Außerdem zielte dieser darauf, auf diesen Daten ein Modell zu trainieren, das es ermöglicht neue Punktwolken mit dem verfügbaren Modell zu vergleichen und festzustellen, von welcher

Sensoreinrichtung sie mit größerer Wahrscheinlichkeit stammen. Der Großteil der Kameradaten, die in den Kfz-Datensätzen erfasst werden, wird mit einem Pkw-ähnlichen Fahrzeug gesammelt. Dies kann jedoch zu einem Leistungsabfall führen, wenn Modelle, die mit solchen Daten trainiert wurden, auf größere Fahrzeugklassen wie Lkw, Busse usw. mit einer anderen Einbauposition angewandt werden. Die Bestimmung des Grades des Einflusses der Kameramontageposition auf die Leistung von Objekterkennungsnetzen und die Strategien, die erforderlich sind, um diese Lücke zu schließen, ist eine wichtige Frage, die ein Partner angegangen hat.

In der letzten Phase des Projektes haben die meisten Partner im Wesentlichen die angefangenen Arbeiten weitergeführt und/oder verfeinert. Ein Beispiel dafür ist der Transfer von synthetischen Daten bei der Nutzung von CycleGAN mit der Aufgabe eines Transfer des Posenmodells auf reale Daten. Auch die Aufgabe der LIDAR-Auflösungsvergrößerung wurde weitergeführt durch Zusatzinformationen wie Tiefendaten. Das Trajektorienplanungsthema wurde ebenfalls fortgeführt mit Zusatzinformationen von verfolgten Objekten. Gleiches gilt für die LIDAR-Lückenanalyse, bei der es um die Untersuchung des Deltas unterschiedlichen Verbaupositionen ging, was auch ein fortgeführtes Thema ist. Zudem würde auch die LIDAR-Lückenanalyse zwischen LIDARen mit unterschiedlicher Scanlinienanzahl fortgeführt. Eine weitere Lückenanalyse hat sich mit dem Delta zwischen unterschiedlicher Videoauflösungen bei der Objektdetektion beschäftigt. Die Überbrückung der Wetterdomänenlücke bei LIDARen wurde mit der Methodik der Gradienteninvertierung angegangen. Eine letzte Lückenanalyse fand bzgl. des Domänenwechsels auf Feature-Ebene statt.

3.2.2.3.3 Gegenüberstellung der vorgegebenen Ziele

Alle Partner haben ihre Ziele über die Unterarbeitspakete in diesem Projekt selbst vorgegeben und entsprechend abgearbeitet. Aufgrund kluger Planung hat kein Partner sich Ziele gesteckt, die er nicht erreichen konnte. Wie in einer PI-Planung üblich haben die Partner die Möglichkeit genutzt Unterarbeitspakete, die aufgrund der vorangegangenen Experimenten uninteressant wurden, neu zu planen, um ein besseres Gesamtergebnis zu erzielen. Darüber hinaus haben Partner durch den regen Austausch weitere Erkenntnisse gewonnen und damit die vorgegebenen Ziele übertroffen.

3.2.2.4 AP2.4 und AP2.5 Orts- und Zeitdomäne und Umgebungsanpassung

3.2.2.4.1 Zielsetzung

Das Kernziel von AP2.4 behandelt den effizienten Umgang mit Orts- und (Lang-) Zeitänderungen, wie zum Beispiel dem Wechsel von Jahreszeiten oder des Erscheinungsbilds von unterschiedlichen Ländern. Das AP2.5 behandelt weiterhin den effizienten Umgang mit Umgebungsänderungen, zum Beispiel bedingt durch Wettereinflüsse. Im Zuge vergangener Einarbeitung und der Literaturrecherche wurden Querschnittsaspekte sowie methodologische Ähnlichkeiten beider Arbeitspakete erkannt, da verschiedene Umgebungsänderungen ebenfalls durch Orts- und (Lang-) Zeitänderungen beschrieben werden können (z.B. Schnee-Erkennung bei plötzlichem Niederschlag als Umgebungsänderung, im Winter als (Lang-) Zeitänderungen oder in Schweden als Ortsänderung). Aus diesem Grund findet die Organisation und somit die Berichterstattung der Arbeitspakete AP2.4 und AP2.5 gemeinschaftlich statt. Die dadurch entstandenen Synergien zwischen den Partnern, sowie der intensivere Austausch unterstützten die Zielerreichung der geplanten Themen.

Die Arbeitspakete 2.4 und 2.5 wurden im Projekt als Use-Case-APs aufgesetzt, was dazu führt, dass mehrere Partner ihre Methoden und Ansätze in anderen APs entwickeln und in diesen APs

auf die beschriebenen Use Cases anwenden. Dadurch existiert eine enge Verzahnung der jeweiligen APs. Gleichzeitig ist die methodische Diversität im AP 2.4 und 2.5 sehr hoch und gruppiert sich grob in:

- Knowledge Engineering
- Lidar-Wahrnehmung
- Inkrementelles Lernen
- Rare-Event Simulation
- Domain Generalisierung
- Domain Transfer Learning
- Domain Gap Quantifizierung

3.2.2.4.2 Erzielte Ergebnisse im Einzelnen

Im Folgenden wird exemplarisch auf einzelne erzielten Ergebnisse der oben genannten thematischen Cluster eingegangen.

3.2.2.4.2.1 Knowledge Engineering

Die Nutzung von Knowledge Graphs ist eine weitere Möglichkeit dem Modell zusätzliches Wissen zuzuführen, um eine verbesserte Generalisierung zu erhalten. In diesen APs wurden Knowledge Graphs dazu eingesetzt die Erkennung von Verkehrszeichen zu verbessern. In den Graphen sind hierbei a-priori Informationen zu Konventionen, Taxonomien und Relationen von Verkehrszeichen enthalten. Eine Übersicht der Zusammenhänge kann in der Abbildung 17 betrachtet werden. Es konnte gezeigt, dass die Nutzung solcher zusätzlicher Informationen die Generalisierung für andere Datensätze deutlich verbessern kann.

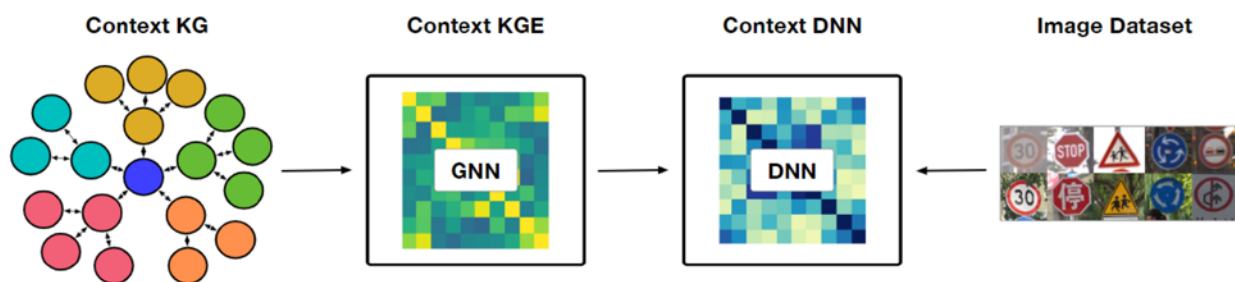


Abbildung 17: Übersicht über die Verwendung eines Knowledge Graphen zur Optimierung einer Verkehrszeichenerkennung.

3.2.2.4.2.2 LiDAR-Wahrnehmung

Auch die Umfeldwahrnehmung durch LiDAR-Sensoren kann durch Wetter- oder Umgebungsänderungen unter einer deutlichen Performance-Reduktion leiden, daher waren auch Untersuchungen im Bereich LiDAR Domain-Änderungen Teil dieser APs. In den Untersuchungen konnte gezeigt werden, dass eine Gradient Reversal Layer den Domainänderungs-bedingten Performanceverlust, der durch unterschiedliche Wettersituationen ausgelöst wird, reduzieren

kann (siehe Abbildung 18). Dabei konnte auch die Performance in der Ursprungs-Domäne erhöht werden. Weitere Untersuchungen haben sich mit dem Thema Unsicherheitsabschätzung für LiDAR-Sensoren unter verändernden Wettersituation beschäftigt und herausgefunden, dass diese bei semantischer Segmentierung für widrige Wettersituationen helfen kann die Segmentierung zu verbessern, aber zugleich die punktweise-Erkennung wetter-bezogenen Rauschens an seine Grenzen stößt.

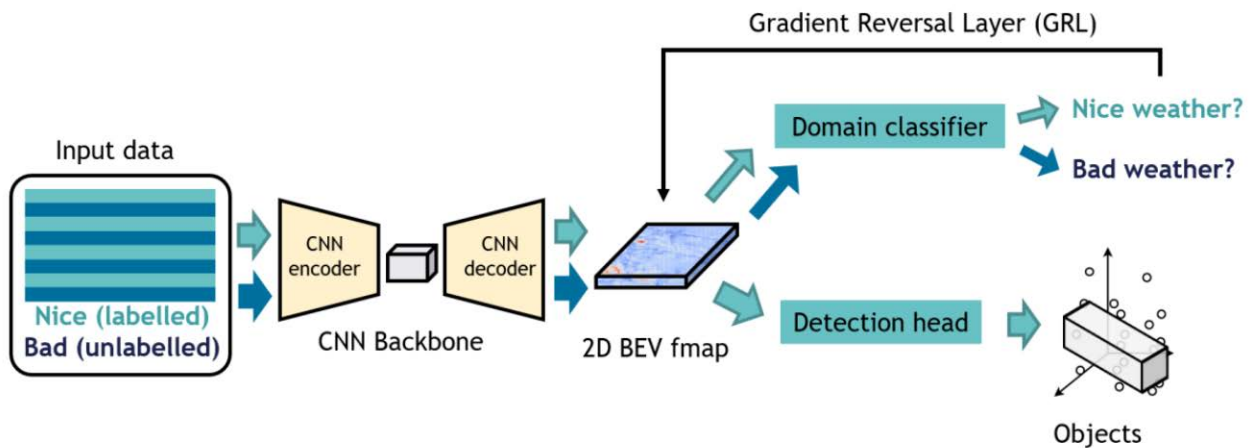


Abbildung 18: Übersicht über die Verwendung eines Gradient Reversal Layers zur Optimierung der Performance bei Schlechtwettersituationen.

3.2.2.4.2.3 Inkrementelles/Kontinuierliches Lernen

Inkrementelles kontinuierliches Lernen bezeichnet eine Lernmethodik, bei der unterschiedliche Target-Daten sequentiell bereitgestellt werden, sodass eine oder mehrere neue Domänen nicht wie sonst üblich in einem Trainingsprozess, sondern nacheinander gelernt werden müssen. Zusätzlich muss die Domain Gap berücksichtigt werden. In den Untersuchungen in diesen APs wurde eine Adaptierungsmethode im Eingabebereich, also auf Pixel-Level, evaluiert. Dabei wurde die populäre Methode Adaptive Instance Normalization (AdaIN) genutzt, um die visuellen Merkmale der Source und der eingehenden Target-Domänen anzugleichen. Es wurde eine verbesserte AdaIN-Version entwickelt, welche Skip-Connections und einen klassenweisen Style Transfer einführt. Wie in der Abbildung 19 zu sehen kommen die Klassen-Information in der Target-Domäne von Pseudo-Labels. Mit dieser Technik konnte auf dem herausfordernden real-real Domänen Shift zwischen Cityscapes und ACDC eine signifikante Verbesserung erzielt werden.

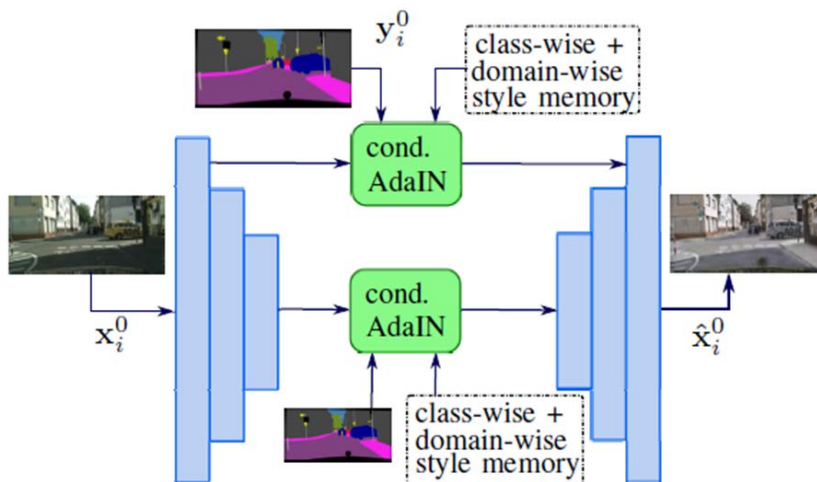


Abbildung 19: Übersicht über einer verbesserte Version von Adaptive Instance Normalization (AdaIN) für den Umgang mit Inkrementellen/Kontinuierlichen Lernen.

3.2.2.4.2.4 Domain Generalisierung

In diesem Cluster wurden Methoden zur Domaingeneralisierung entwickelt und evaluiert. Diese unterscheiden sich von anderen Methoden, da keine Daten aus der Zieldomäne vorliegen und somit auf eine oder mehrere unbekannte Zieldomänen hin generalisiert werden muss (siehe Abbildung 20). Dies orientiert sich eng an den Zielen von AP2.4 und AP2.5, da die neuen Zeit- und Umgebungsdomänen in der Realität häufig unbekannt sein können. Konkret wurden in diesem Bereich sowohl Ansätze der Domaingeneralisierung als auch der Domainadaptierung für semantische Segmentierung untersucht und es konnte festgestellt werden, dass beide Gruppen von Ansätzen für eine teilweise signifikant verbesserte Domaingeneralisierung sorgen. Auch die Domainadaptierungsansätze erhöhen, obwohl sie auf eine bestimmte, bekannte Zieldomäne trainiert werden, die Generalisierung für unbekannte Target-Domänen. Hervorzuheben sind der Datensatz Mapillary Vistas, der besonders gute Generalisierungseigenschaft gezeigt hat, und die vision transformer Architekturen, mit deren Hilfe es gelungen ist die Domain Gap stark zu reduzieren. Weitere Arbeiten brachten das Ergebnis, dass eine einfache Fusionierung unterschiedlicher Netzgewichte die Generalisierung deutlich verbessern kann.

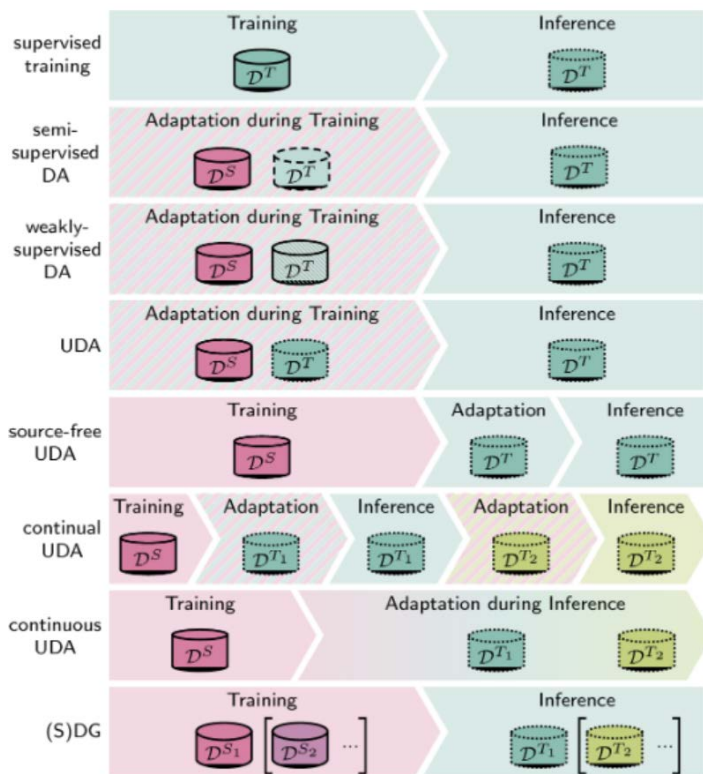


Abbildung 20: Übersicht über Verschiedene Methoden der Domain Anpassung . Dabei werden Unsupervised Domain Adaption (UDA) und Domain Generalization (DG) gegenübergestellt indem die verwendeten Daten einer Ursprungs-Domäne (\mathcal{D}^S) und einer Ziel-Domäne (\mathcal{D}^T) dem Training oder der Inferenz zugeordnet werden.

3.2.2.4.2.5 Domain Gap Quantifizierung

In diesem Themenbereich ging es darum Methoden zur Quantifizierung von Domain Gaps zur Überbrückung von Ortsdomänen zu evaluieren (siehe Abbildung 21). Für die kamera-basierte semantische Segmentierung haben diese Untersuchungen neue Erkenntnisse über das Verhalten neuronaler Netze gebracht, die einen Domain Gap verarbeiten müssen. Es wurde festgestellt, dass der Domain Gap layer-abhängig ist, d.h. unterschiedliche Teile des Netzes sind unterschiedlich stark von dieser Domainänderung betroffen. Es wurden verschiedene Metriken evaluiert und verglichen, aber auch bestimmte Methoden wie die Dimensionalitätsreduktion mittels Singulärwertzerlegung (SVD) verworfen aufgrund von mathematisch inkonsistenten Ergebnissen. Insgesamt sind die entwickelten Metriken vielfältig auch für weitere Bereiche wie Active Learning oder Out-of-Distribution Detection einsetzbar.

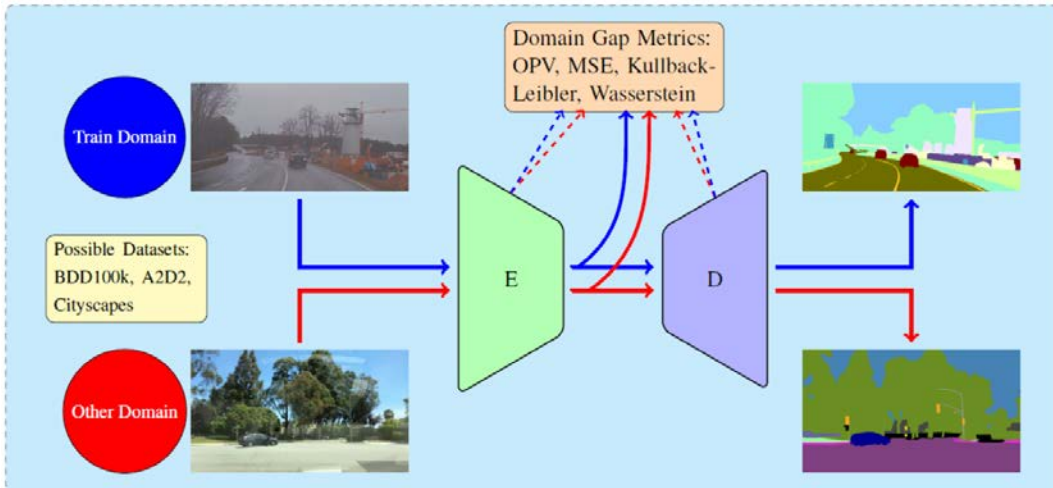


Abbildung 21: Übersicht über das Konzept zur Domain Gap Quantifizierung.

3.2.2.4.3 Gegenüberstellung der vorgegebenen Ziele

Nach tieferen Einblick in die erzielten Ergebnisse im Einzelnen, kann die Zielerreichung an den folgenden ursprünglich geplanten Ergebnissen gemessen werden:

ID	Ergebnis	Zielerreichung
E1.2.2.4 & E1.2.2.5	Methoden, welche die Orts-Zeit-Domänen-Anpassung und Umweltanpassung unterstützen bzw. möglich machen	✓
E2.2.2.4 & E2.2.2.5	Erste Implementierung von Methoden aus E1.2.2.4 & E1.2.2.5.	✓
E3.2.2.4 & E3.2.2.5	Ausgewählte Methoden von E2.2.2.4 & E2.2.2.5 sind optimiert und Alternativen wurden ausprobiert.	✓
E4.2.2.4 & E4.2.2.5	Finaler Ansatz für die ausgewählten Methoden aus E3.2.2.4 & E3.2.2.5 ist vorhanden	✓

Tabelle 3: Gegenüberstellung der geplanten Ergebnisse zur Zielerreichung für AP 2.4 und AP 2.5.

Dies bedeutet, dass für die oben genannten Themenbereiche im Zusammenhang mit dem effizienten Umgang mit Orts- und (Lang-) Zeitänderungen sowie mit Umgebungsänderungen, verschiedene Methoden recherchiert, implementiert, optimiert und evaluiert wurden. Es entsteht demzufolge eine umfangreiche Sammlung dieser Methoden, welche für zukünftige Ansätze als Referenz herangezogen werden können und die zukünftige Forschung und Entwicklung in diesen Bereichen unterstützt. Für eine detaillierte Aufstellung der Methodensammlung wird auf die jeweiligen Meilensteinberichte, Deliverables und vor Allem auf die entstandenen Publikationen, wie zum Beispiel die auf internationalen Konferenzen publizierten Paper verwiesen.

3.2.3 Zielerreichung

Im Rahmen von TP2 hatten wir uns zum Ziel gesetzt, die Fähigkeit von neuronalen Netzwerken zur Verarbeitung von Daten verschiedener Sensor- und Ortsdomänen zu verbessern. Dazu wurden

vier Arbeitspakete definiert: Continuous Learning, Synthetic Data, Cross-Sensor-Adaption und Orts- und Zeitdomäne.

Wir freuen uns, mitteilen zu können, dass alle Ziele aller Arbeitspakete erfolgreich erreicht wurden. Durch die Implementierung von Continuous-Learning-Verfahren in **Arbeitspaket 2.1** konnten Probleme wie "catastrophic forgetting" gemildert werden und die Lerneffizienz dieser Netze gesteigert werden.

Durch den Einsatz von synthetischen Daten in **Arbeitspaket 2.2** konnten die Netze auf eine größere Varietät von Daten vorbereitet werden, was zu einer Verbesserung bezüglich nicht gesehener Daten führte.

Die Cross-Sensor-Adaption in **Arbeitspaket 2.3** ermöglichte eine Übertragung von Wissen aus einem Trainingsdatensatz auf einen anderen, was die Effizienz des Trainingsprozesses weiter erhöhte.

Die Implementierung der Orts- und Zeitdomänen-Features aus **Arbeitspaket 2.4** und **Arbeitspaket 2.5** schließlich ermöglichte es den Netzen, Daten aus verschiedenen Orten und Zeiten zu integrieren, was zu einer verbesserten Generalisierung und einer höheren Fähigkeit zur Vorhersage führte.

Im Hinblick auf unsere Projektziele können wir somit feststellen, dass das Transfer Learning-Teilprojekt ein bedeutender Erfolg war und wesentlich dazu beitragen wird, die Effektivität neuronaler Netzwerke zu erhöhen und deren Einsatzmöglichkeiten in verschiedenen Bereichen zu erweitern.

3.3 TP3: Didaktik

3.3.1 Ausgangssituation und Zielsetzung

Der kommerzielle Erfolg von Produkten basierend auf maschinellem Lernen beruht überwiegend auf überwachtem Lernen (supervised learning) anhand eines zur Aufgabenstellung zugehörigen Trainingsdatensatzes. Diese Art des Trainings neuronaler Netze bedarf einer immensen Anzahl an Labels für die verwendeten Trainingsdaten sowie eines beträchtlichen Hardware-Einsatzes für die vielfache Wiederholung der Trainingschritte. Das Delta, das Teilprojekt 3 adressiert, ist also die effiziente Nutzung der vorhandenen Informationen aus Datensätzen.

Überwachtes Lernen impliziert die Definition eines Trainingsdatensatzes und einer anschließend zufälligen und vielfachen Auswahl und Präsentation der Datenpunkte während Trainingszeit eines lernenden Algorithmus. Diese Art des Trainings neuronaler Netze hat die wissenschaftliche Community und Tech-Industrie zu signifikanten Erfolgen im Bereich der künstlichen Intelligenz gebracht.

Im Anwendungsbereich der automobilen Perzeption ist es notwendig, eine große Anzahl an gelabelten und hochaufgelösten Trainingsdatenpunkten zur Verfügung zu stellen, um den Ansprüchen an Leistungsfähigkeit der Algorithmen innerhalb einer hochkomplexen Umwelt gerecht zu werden. Für diesen Anwendungsfall bringt überwachtes Lernen verschiedene Probleme mit sich, die in zwei Problembereichen zusammengefasst werden können:

1. Zeit- und Ressourcenaufwand für Training: Für eine zuverlässige KI-basierte Perzeption im automatisierten Fahren wird eine immense Anzahl an Trainingsdaten benötigt, welche

den realen Datenraum bestmöglich abdeckt. Folglich rechnet die Automobilindustrie mit einer enormen, sukzessive ansteigenden Menge an Trainingsdaten zur Abdeckung des Einsatzbereiches der KI-Funktionen in der Perzeption. Zur Verarbeitung dieser Datenmengen sind beträchtliche Hardwareressourcen notwendig. Bereits zum heutigen Zeitpunkt werden in der automobilen Perzeption Algorithmen trainiert, welche mehrere Wochen GPU-Zeit in Anspruch nehmen. Neben dem hohen zeitlichen Aufwand, welchen das Training neuronaler Netzwerke mit sich bringt, ist mit dem Training auch ein hoher Energieaufwand verbunden.

2. Zeit- und Ressourcenaufwand für Datenlabeling: Labeling von Trainingsdaten für neuronale Netze wird zum heutigen Zeitpunkt größtenteils in akribischer Handarbeit vorgenommen. Daraus ergibt sich ein hoher zeitlicher sowie monetärer Aufwand, um die Daten für überwachtes Lernen zur Verfügung zu stellen.

Zusammengefasst ist sowohl aus wirtschaftlicher, als auch ökologischer Sicht eine Optimierung des Trainingsprozesses – also einer zielgerichteten Didaktik – sinnvoll.

3.3.2 Strukturierung

TP3 Didaktik war in vier Arbeitspakete (AP) eingeteilt. Für teilüberwachtes und unüberwachtes Lernen (AP3.1) kann eine Analogie im Lernprozess eines Kleinkindes gefunden werden. So lernt es ohne externes Feedback bzw. Überwachung durch Eltern, dass z. B. Hindernisse nicht durchschreitbar sind. Ein Beispiel für teilüberwachtes Lernen ist, wenn der unüberwachte Lernprozess durch selektives Feedback unterstützt wird. Im vorliegenden Beispiel, die Information durch die Eltern, dass Türen geöffnet werden können. Trainingsorganisation (AP3.2) findet eine Analogie im gesteuerten Lernprozess eines Kleinkindes. Hier wird durch eine externe Instanz, wie beispielsweise die Eltern, eine Lernstrategie für das Kleinkind vorgegeben. Ein Beispiel ist das Erlernen von mathematischen Methoden. Hier muss das Kleinkind zuvorderst die Addition lernen, bevor komplexere Aufgaben, wie eine Multiplikation erfolgreich und effizient erlernt werden können.

Eine ähnliche Analogie findet sich für das Aufgabenpaket Aktives Lernen (AP3.3). Hier definiert, im Gegensatz zur Trainingsorganisation, das Kleinkind selbst, für welche Aufgabe weiterer Input zum Verständnis notwendig ist. Ein Beispiel ist das fehlende Verständnis einer bestimmten Aufgabenstellung während des Erlernens der Multiplikation, woraufhin das Kind bei der unterstützenden Instanz, zum Beispiel die Eltern, aktiv nachfragt. Auch für Knowledge Transfer (AP3.4) findet sich eine Analogie im schulischen Lernen. Wenn ein*e Lehrer*in einem Schulkind einen Sachverhalt erklärt, macht er oder sie höchstwahrscheinlich kleinere Fehler in seiner oder ihrer Erklärung. Nichtsdestotrotz ist der Lerneffekt für das Kind höher über die teilweise fehlerhafte Erklärung des Lehrers, als über autodidaktisches Lernen aus einem Buch. Dieser Effekt findet sich auch im Lernprozess eines neuronalen Netzes wieder. Ein erhöhter Lerneffekt ergibt sich insbesondere bei mehreren Lehrern zu ähnlichem Schulstoff (multi-teacher learning), sowie Erarbeitung eines Themas über mehrere Aufgaben (multi-task learning).

Didaktik

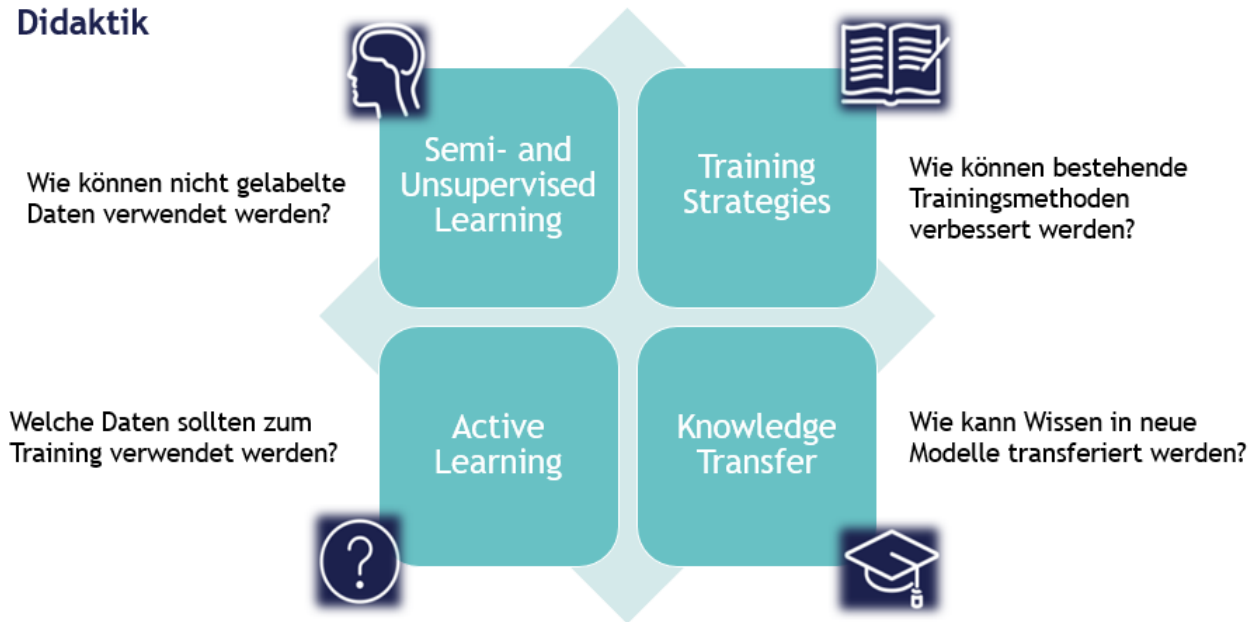


Abbildung 22: Überblick über die Themenbereiche und Fragestellungen des Teilprojekt 3: Didaktik!

3.3.2.1 3.3.2.1 AP3.1 Teilüberwachtes und unüberwachtes Lernen

3.3.2.1.1 Zielsetzung

Das Ziel von AP3.1 bestand darin, Methoden für unüberwachtes Lernen und teil-überwachtes Lernen im automobilen Umfeld zu untersuchen, um den sonst benötigten manuellen Labelingaufwand möglichst zu reduzieren. Bei dem klassischen überwachten Lernen ist ein gelabelter Datensatz erforderlich, um ein Modell zu trainieren, dessen Erstellung zeitaufwendig und teuer sein kann, insbesondere bei komplexen Wahrnehmungsaufgaben wie der Objekterkennung oder Segmentierung. Im Gegensatz dazu können Algorithmen für unüberwachtes Lernen aus nicht gekennzeichneten Daten lernen, die viel zahlreicher und einfacher zu sammeln sind. Dadurch kann das System aus einem breiteren Spektrum von Datenquellen lernen und möglicherweise neue Muster oder Merkmale aufdecken, die beim überwachten Lernen nicht erkannt worden wären. Innerhalb von AP3.1 wurden dazu folgende Arten von unüberwachten Lernmethoden untersucht:

- Unüberwachtes Lernen: betrachtet Algorithmen, welche ohne die Zunahme von Labels lernen können.
- Teilüberwachtes Lernen: Algorithmen, welche, zusätzlich zu gelabelten Daten, ungelabelte Daten mit in den Trainingsprozess einbeziehen.
- Selbst-überwachtes Lernen: Lernverfahren, welche in den Daten vorhandene Trainingssignale verwenden, um einen Proxy Task zu lösen, mit der das Verfahren relevante Informationen für die Zielaufgabe lernt.
- Schwach-überwachtes Lernen: betrachtet Lernverfahren, welche relevante, aber ungenaue oder verrauschte Trainingssignale für den Trainingsprozess verwenden, um den Labelingaufwand zu verringern.

Mithilfe dieser Methoden wurde es möglich den Labelingaufwand in verschiedenen Bereichen zu reduzieren unter anderem: für die Domänenanpassung zwischen Simulation und Realen Trainingsdaten für Kamera und Real-Daten, 3D Human Pose Estimation, Instanz Segmentierung sowie der Tiefenschätzung für Kamerabilder. Diese Methoden wurden auf öffentlich verfügbaren Datensätzen entwickelt und später auf dem projektinternen Datensatz evaluiert.

3.3.2.1.2 Erzielte Ergebnisse

Der folgende Abschnitt beschreibt die wichtigsten Ergebnisse, die in den unterschiedlichen Themenfeldern des AP3.1 erzielt worden sind. Ein großer Teil der Arbeit beschäftigte sich mit effizienten Lernmethoden für die semantische Segmentierung, da der Labelingaufwand dort besonders hoch ist weil jeder Pixel eines Bildes gelabelt werden muss. Hierzu wurde ein zweistufiges Verfahren zur unüberwachten Domänenanpassung entwickelt, das auf die Methoden des Self-Learning sowie Domain Adaptation durch Angleichung der Verteilung im Feature-Raum durch ein Selbstüberwachtes Clustering basiert. Dieses zweistufige Verfahren wurde zu einem iterativen Verfahren erweitert, wie es in der Abbildung 23 zu sehen ist. Der iterative Prozess besteht dabei aus einer Kombination von Self Training und des Selbstüberwachten Clusterings. Hierbei werden die vom Modell selbst generierten Label iterativ verbessert. Mit diesem System gelang es die Güte des Systems auf der Target Domäne im Vergleich zur vorherigen Variante um 10% von 46% mean Intersection over Union (mIoU) auf 56% mIoU zu steigern. Weiterhin wurde auch AP-übergreifend in Kooperation mit AP2.4/2.5 ein ausführliches Survey-Paper zum Thema der unüberwachten Domänenadaption geschrieben.

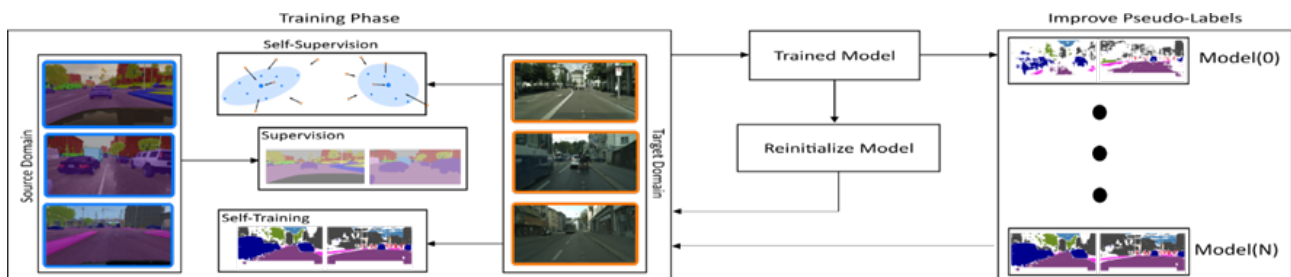


Abbildung 23: Der iterative Trainingsprozess von EasyAdapt zur unüberwachten Domänenanpassung

Neben der unüberwachten Domänenanpassung ist es auch notwendig neue Klassen einzulernen. Dazu wurden zwei Methoden für das unüberwachte Auffinden und Nachlernen neuer Klassen entwickelt. Das Verfahren ist in der Abbildung 24 näher dargestellt. Dabei werden möglicherweise unbekannte Objekte durch eine segmentweise Unsicherheitsquantifizierung identifiziert. Die Bilder werden auf die gefundenen Objekte zugeschnitten und mit einem Bildklassifikationsnetz werden Merkmale extrahiert. Dieser Merkmalsvektoren werden mithilfe von DBSCAN geclustert. Durch Kombination der Unsicherheitsquantifizierung und des Clustering werden Pseudo-Labels erstellt, mit welchen das Netz Klassen-inkrementell nachtrainiert wird. In der zweiten Methode werden die Merkmalsvektoren nicht geclustert, sondern es wird eine paarweise Distanzmatrix berechnet. Das DNN wird dann um eine beliebige Anzahl an Klassen erweitert und mit einer modifizierten Loss Funktion nachtrainiert. Diese Loss Funktion soll die unbekannt Objekte auf die neuen Klassen verteilen und clustert diese gleichzeitig auf Basis der Distanzmatrix.

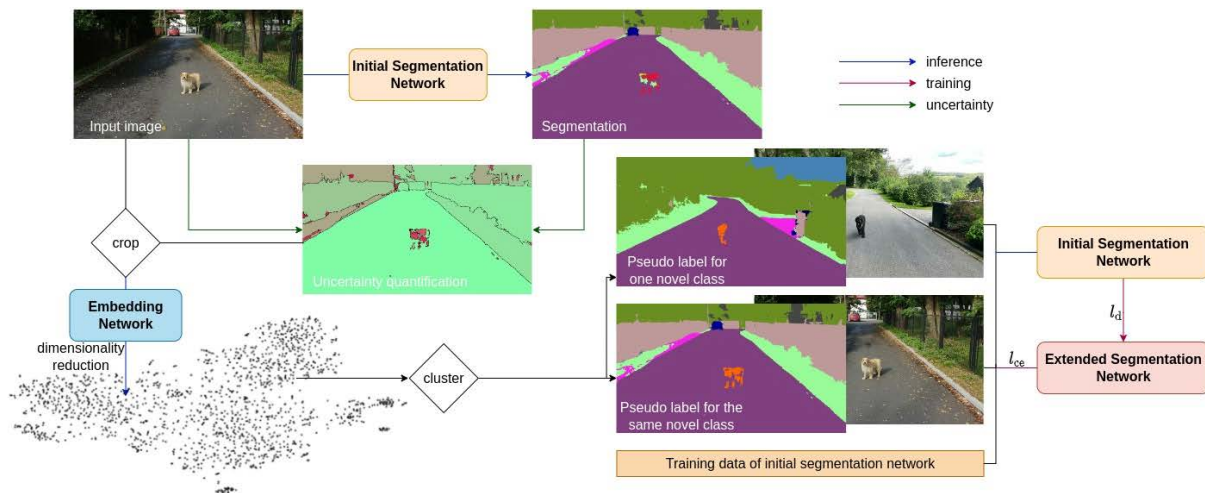


Abbildung 24: Überblick über das Framework zum unüberwachten Einlernen unbekannter Objekte für die semantische Segmentierung

Eine Limitierung beider Methoden ist, dass sie sehr sensibel bezüglich der Anomalie Detektion, des Clustering und der gleichzeitig wechselnden Domäne sind. Deshalb wurde weiterhin auch untersucht wie gut sich Methoden für das Klassen-inkrementelle Lernen und der unbewachten Domänenanpassung kombinieren können. Die Grundidee ist, dass in dem synthetischen Datensatz nicht alle Klassen und Objekte vorhanden sind die in der Realität existieren. Daher ist ein Modell welches auf dem synthetischen Datensatz trainiert worden ist, nicht in der Lage, Objekte die nicht in der Simulation existieren, zu detektieren. Dieses Modell muss daher so angepasst werden, dass es Klassen und Objekte, die ausschließlich durch die synthetischen Daten erlernt wurden, auch in der Realität erkennt und zusätzlich zuvor ungesehene Daten hinzulernen kann. Dabei wurden Verfahren des Class-Balanced Self-Training mit Multi-Scale Fourier Domain Adaption (M-FDA) und einer speziellen Augmentierungsstrategie kombiniert um klassen-inkrementelle Lernen auch auf ungesehen Domänen zu ermöglichen.

Eine weitere Möglichkeit Labelingaufwand zu reduzieren ist es, Labels mit geringeren Detailgrad zu verwenden beispielsweise Bounding Boxes anstelle von pixel-genauen Segmentierungen. Eben dazu wurde ein Konzept zum schwach überwachten Lernen von Instanzsegmentierungen mithilfe der Nutzung von Stereotiefenkarten entwickelt, implementiert, experimentell ausgewertet, die Schwächen analysiert und eine abschließende Bewertung vorgenommen. Ziel des Konzepts war, kostengünstige Objektannotationen von Bounding Boxes halbautomatisch in aufwendige Objektsegmentierungen umzuwandeln, indem die in Tiefenbildern vorhandenen Objektkanten genutzt werden. Als Grundlage sollte der Mask-R-CNN Ansatz dienen, der so umgebaut werden sollte, dass die zusätzlichen Bounding Boxes und die Tiefenkarten effektiv einfließen können. Der Ansatz erzielte einen AP-Wert von 36,9, reagierte jedoch sensibel auf Hyperparameteränderungen und funktionierte nicht gut mit den derzeitigen Tiefenkarten. Die Studie kommt zu dem Schluss, dass der Schwerpunkt auf der Verbesserung der Tiefenschätzung liegen sollte und dass schwach überwacht Lernen mit Bounding Boxes nicht konkurrenzfähig mit anderen veröffentlichten Ansätzen war.

Ebenso aufwendig wie das Labeling von Semantischer Segmentierung ist das Labeling von menschlichen Posen in 3D, da jeder sogenannte Keypoint in 3D Koordinaten gelabelt werden muss. Deshalb war ein Ziel die Entwicklung einer Methode zur unüberwachten 3D Posenschätzung aus Multi-View Bildern.

Die erste Implementierung des vorgestellten Verfahrens zeigte vielversprechende Ergebnisse, allerdings wurde keine zeitliche Information einbezogen, was die Qualität der prädizierten 3D Posen stark beeinflusst hat. Daraufhin wurde das Konzept erweitert unter Berücksichtigung der zeitlichen Information, Während des Trainings hatten wir Zugriff auf Multiview Bilder. Trainiert wurde ein temporales Convolutional Neural Network, um die 3D Posen zu schätzen. Zuerst wurden mittels einer vortrainierten 2D-Körperposenschätzung 2D Keypoints extrahiert. Die 3D Pose wurde zunächst mit Hilfe der Epipolar Geometrie (Triangulation) berechnet. Wir nennen die durch die Triangulation berechneten 3D Posen „3D Pseudo Ground-Truth“. Die „3D Pseudo Ground-Truth“ wurde für das Training des Netzes verwendet. Während des Trainings wurde die vom Netz geschätzte 3D Pose mit Hilfe der „Camera Consistency“ in die jeweils andere „View“ transformiert, mit dem Ziel, den 3D-Suchraum weiter einzuschränken. Eine Limitierung beider Methoden ist, dass sie die Multiview Bilder nutzten, die allerdings im Context des autonomen Fahrens nicht so oft verfügbar sind.

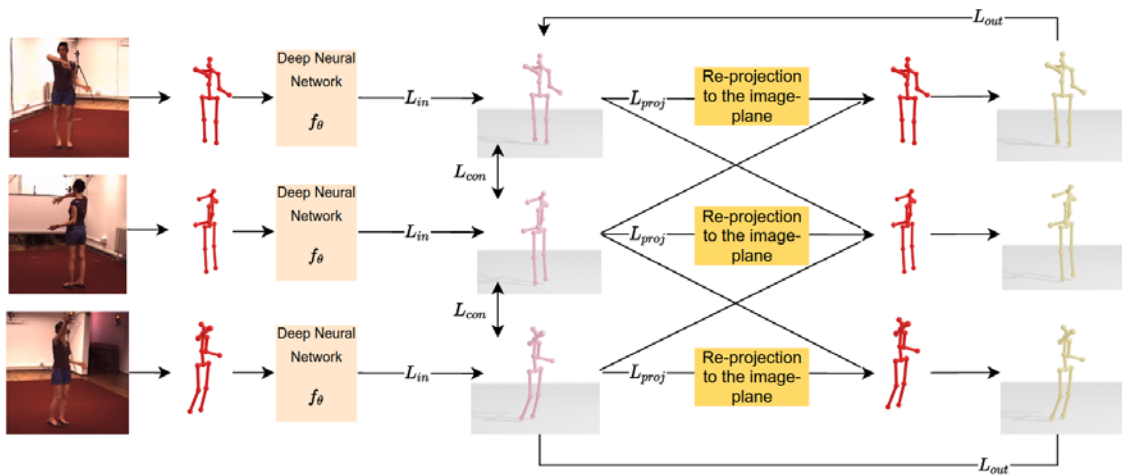


Abbildung 25: Single-Frame unüberwachte 3D Posenschätzung

Weiterhin wurden Methoden für die unüberwachte monokulare Tiefenschätzung entwickelt, die die Generierung von Tiefendaten aus monokularen Bildern ermöglichen soll, ohne dabei an Genauigkeit und Robustheit im Vergleich zu teuren Lidar-Sensoren zu verlieren. Zu diesem Zweck wurde zunächst die Auswirkungen des Einsatzes verschiedener Aufmerksamkeitsmechanismen auf ein aktuelles Baseline-DNN untersucht. Es wurde festgestellt, dass die Verwendung von kanalbasierter Aufmerksamkeit bei der Tiefenschätzung zu einer besseren Leistung als andere Techniken führte. Anschließend untersuchten wir verschiedene Strategien zur weiteren Leistungsverbesserung, wie z. B. die Einbeziehung zusätzlicher Merkmalskarten und die Verwendung von partiellen Tiefenkarten als Input für das Posenschätzungsnetzwerk während des Trainings. Dadurch konnte die Fähigkeit mit Randbereichen und texturlosen Regionen in den Bildern besser umzugehen verbessert werden. Aufgrund der sicherheitskritischen Natur des autonomen Fahrens wurden auch die Modellierung der Unsicherheit für Tiefenschätzungsnetzwerke erforscht. Es stellte sich heraus, dass es nicht immer eine lineare Beziehung zwischen dem Grad der Unsicherheit und den Vorhersagefehlern zu geben scheint. Insgesamt waren wir in der Lage, durch die Einbeziehung der verschiedenen oben beschriebenen Strategien die Baseline in den realen KITTI- und DDAD-Datensätzen um bis zu 2 % bzw. 29 % zu verbessern. In Bezug auf die Unsicherheitsmetriken wurde festgestellt, dass sich verschiedene Methoden tatsächlich auf verschiedene Aspekte der Bilder konzentrieren, die unsicherer sind, jedoch diese Regionen sich nicht immer

überschneiden. Zu den Einschränkungen des vorgeschlagenen Ansatzes gehört die Unfähigkeit, mit ungünstigen Bedingungen wie Regen, reflektierenden Oberflächen, Bewegungsunschärfe usw. umzugehen.

Ein anderer untersuchter Ansatz befasste sich auch mit der Generierung von Texturen für 3D-Modelle. Wenn ein 3D-Designer ein neues 3D-Modell entwickelt, so wird in einem ersten Schritt die 3D-Struktur erstellt (d.h. die Oberfläche des Objekts), welches anschließend texturiert wird. Die Texturierung stellt hierbei einen wesentlichen zeitlichen Aufwand dar. Durch die Generierung einer Textur kann so indirekt der Labelingaufwand reduziert werden, indem mit weniger Aufwand Assets für eine Simulation erzeugt werden und schlussendlich vermehrt vielfältige synthetische Daten erzeugt werden können. Der entwickelte Ansatz nimmt ein untexturiertes 3D-Modell und fügt es in eine bestehende Szene ein. Anschließend werden verschiedene Ansichten gerendert, wobei gleichzeitig eine Maske für die untexturierten Stellen erzeugt wird. Ein gerendertes Bild und die Maske werden anschließend von einem trainierten Inpainting-Modell (z.B. Stable Diffusion v2) verarbeitet, sodass die untexturierten Stellen ausgemalt werden. Dieses Resultat wird anschließend wieder auf das 3D-Modell projiziert um so die Textur zu erzeugen. Auf diese Weise konnten in beispielhaften Szenen qualitativ gute Ergebnisse erzielt werden. Eine Schwierigkeit stellt hierbei noch die nötige hohe Auflösung von Texturen dar, die aufgrund der limitierten Auflösung von Inpainting-Modellen schwer zu erreichen ist.

Die Arbeit zu Unüberwachten Domänenanpassung von LiDAR Daten wurde in der ersten Hälfte des Projektes in AP3.1 bearbeitet und später in AP2.3 fortgesetzt. Die in der VHB genannten Ziele wurden bis auf die genannten Änderungen und der Evaluierung auf dem projekt-internen Datensatz alle erreicht.

3.3.2.1.2.1 Gegenüberstellung der vorgegebenen Ziele

Die in der Vorhabensbeschreibungen definierten Ziele wurde nach Abschluss der verschiedenen Projektinkremente reevaluiert und so im Laufe des Projekts Aufwände angepasst, aufgrund von thematische Überschneidung mit andere Arbeitspakete oder da die erzielten Ergebnisse eines erarbeiteten Konzepts für die selbst-überwachte Trajektorienplanung nicht zielführend erschien. Bis auf diese genannten Änderungen wurden die in der Vorhabensbeschreibung genannten Ziele erreicht und die geplanten Ergebnisse auf internationalen Konferenzen sowie in den Deliverables publiziert. Lediglich die Evaluierung der Methoden auf dem projekt-internen Datensatzes wird noch im Anschluss des Projektes nachgeholt, aufgrund der späten Verfügbarkeit der Daten.

3.3.2.2 AP3.2 Trainingsorganisation

Ziel von AP3.2 war die Untersuchung von hierarchischen Datenmodelle zum Training Neuronaler Netze, die Erarbeitung von Ansätzen und Dokumentation für die Erweiterung der Performanz von Neuronalen Netzen (NN) über Trainingsorganisation. Dazu gehören die Auswahl und Aufbereitung von Daten, die Entwicklung adäquater Modelle und passenden Hyperparametern. Die Arbeiten von AP3.2 können in die folgenden drei Themen eingeordnet werden:

- Hierarchische Datenmodelle zum Training Neuronaler Netze
- Datenaugmentierung für 3D-Daten
- Co-Simulation und live Corner Case Erkennung

Im Folgenden werden die Arbeiten der Partner kurz vorgestellt, die im Rahmen des Arbeitspakets Trainingsorganisation geplant und durchgeführt wurden. Für eine detaillierte Beschreibung der jeweiligen Studien und Ergebnissen sei hiermit auf die partnerspezifischen Schlussberichte bzw. Deliverables verwiesen.

3.3.2.2.1 Hierarchische Datenmodelle zum Training Neuronaler Netze

3.3.2.2.1.1 Zielsetzung

Der Einsatz von hierarchischen Datenmodellen im Kontext von Neuronalen Netzen verfolgte in AP3.2 zwei Ziele:

1. Wiederverwendbarkeit von Modulen Neuronaler Netze und Reduktion der Modell Komplexität
2. Reduktion von Textur-Bias und Steigerung der Robustheit unter Domain-Verschiebung

3.3.2.2.1.2 Erzielte Ergebnisse

3.3.2.2.1.2.1 Wiederverwendbarkeit von NN Modulen und Reduktion der Modell Komplexität

Um die Überparameterisierung, also eine ineffiziente Nutzung der Gewichte, von neuronalen Netzen (NN) zu reduzieren, wurden die Analogien von hierarchischen Iterationsverfahren und residual Netzen (ResNet) [1] ausgenutzt. Der Fokus bei den untersuchten Iterationsverfahren lag bei Mehrgitter (MG) Methoden [2], welche hierarchische Verfahren sind um große (nicht-) lineare Gleichungssysteme effizient zu lösen.

Eine wichtige Komponente in MG sind die Vergrößerungen der Auflösungsstufen, sogenannte Restriktionen. Die Vergrößerung von Auflösungen ist auch für Faltungsnetze charakteristisch und wird z.B. durch Pooling-Operationen umgesetzt. Insbesondere die ResNet-Architekturen sind analog zur Struktur von MG und ermöglichen eine MG Perspektive in der Auflösungsdimension (MgNet) [2]. Diese Analogie von MG und ResNet motiviert die Möglichkeit Gewichtstensoren wieder zu verwenden (vgl. Abb. 3). Das Wiederverwenden von Gewichtstensoren reduziert die Anzahl der Gesamtgewichte des Netzes signifikant, ohne dass es einen negativen Effekt auf die Klassifikationsgenauigkeit hat.

In Faltungsnetzen ist die Kanaldimension in der Regel voll besetzt. Um hier die Anzahl der Gewichte auch einzusparen, wurde eine hierarchische MG Struktur für der Channel Dimension, ähnlich MGIC [4], entwickelt, die weitere MG Komponenten beinhaltet. Die resultierende Architektur verbindet MG in der Auflösungsdimension und in der Channeldimension und heißt daher Multigrid in all Dimensions (MgiaD) [5]. Diese Modelle haben weniger Parameter und keinen Verlust in der Klassifikationsgenauigkeit und reduzieren daher erfolgreich die Überparameterisierung von Neuronalen Netzen.

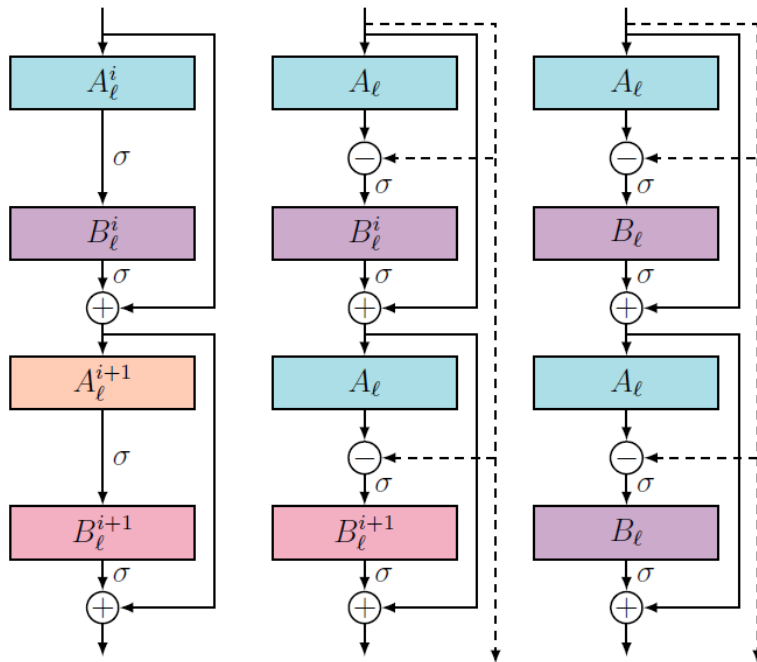


Abbildung 26: NN Bausteine ; Links: standard. ResNet-Block, kein Teilen von Gewichten, Mitte: MgNet, Teilen eines Gewichts-Layers (A), MgNet, Teilen von zwei Gewichts-Layers (A und B) auf einer Auflösungsstufe I.

Die aus diesem Ansatz entstandenen Architekturen sollten ursprünglich durch eine Blackbox-Architektursuche mit MorphNet weiter optimiert werden. Dies wurde durch eine Optimierung mit Hilfe von lokaler Fourier-Analyse (LFA) ersetzt. Durch LFA können Singulärwerte der Faltungsmatrizen effizient berechnet werden. Mit den größten und kleinsten Singulärwerten kann eine Richardsen-Iteration mit Optimalen Hyperparametern implementiert werden und ermöglicht so den Einsatz von weiteren MG Analogien.

3.3.2.2.1.2.2 Reduktion von Textur-Bias und Steigerung der Robustheit unter Domain-Verschiebung

Für das zweite Ziel, die Reduktion von Textur-Bias, wurde anisotrope Diffusion [7,8,9] auf Bilder für Objekt Detektion angewendet. Anisotrope Diffusion ist ein PDE basiertes Verfahren aus der klassischen Bildkompression, das Farbinformation vornehmlich orthogonal zu den Gradienten weitergibt und damit Änderungen in der Textur bewirkt, während Formen und Kanten weitgehend erhalten bleiben. Neben sowohl qualitativer Optimierung als auch Verschnellerung der Implementation, wurde anisotrope Diffusion in AP3.2 als Augmentierungsmethode eingesetzt, um eine stärkere Domänengeneralisierung zu erreichen. Dazu wurden verschiedene Datensätze (z.B. Kitty [10] und BDD100k daytime clear [12]) mit unterschiedlich stark diffundierten Bildern erweitert, Yolo [11] Objektdetektionssätze auf diesen trainiert und ihre Performanz auf Datensätzen wie IDD [13] und VKitti [14] entweder nachträglich evaluiert oder schon während der Trainings gemessen, um schwankungsbedingten fehlerhaften Ergebnissen vorzubeugen.



Abbildung 27: Original Kitti Bild (links) und beispielhafte Diffusionsstufen nach 200 (mittig) und 600 (rechts) Zeitschritten.

Erste Ergebnisse, die eine Verbesserung der Domänengeneralisierung versprechen ließen (Training auf Kitti, Evaluation auf VKitti, BDD, IDD), konnten im zweiten und aufwendigeren Versuchsaufbau nicht bestätigt werden (Training auf BDD "daytime clear", in das Training integrierte Evaluation auf Datensätzen mit Domänenverschiebung wie VKitti, Kitti, BDD100k "daytime bad weather", IDD). Da auch die verwendete Baseline "Gaußsche Diffusion" in unserem Versuch keine verbesserte Domänengeneralisierung brachte, halten wir weitere Untersuchungen in Settings für nötig, die von Gaußscher Diffusion profitieren.

3.3.2.2.2 Datenaugmentierung für 3D-Daten

3.3.2.2.2.1 Zielsetzung laut VHB

Ansätze und Dokumentation für die Erweiterung der Performanz von neuronalen Netzen über Trainingsorganisation:

1. Auswahl einer Baseline und Erstellung von Ergebnissen für die 3D Objekterkennung auf Lidardaten mithilfe der in AP5.2 erstellten Referenzmodelle
2. Entwicklung von Augmentationsverfahren für Lidardaten mit einstellbarem „Schwierigkeitsgrad“
3. Untersuchung der Auswirkung der Augmentationsverfahren während des Trainings

4. Scheduling der Schwierigkeitsgrade der Augmentierung und Untersuchung deren Auswirkungen auf Training und letztendlich Leistungsfähigkeit des Modells

3.3.2.2.2 Erzielte Ergebnisse

1. Bei der 3D Objekterkennung auf Lidardaten sollen zuvor ausgewählte Klassen von Objekten in Lidar-Punktwolken mit einer möglichst eng anliegenden Bounding Box versehen werden. Diese werden üblicherweise mit einem Mittelpunkt, der Ausdehnung in die drei Dimensionen, sowie dem Orientierungswinkel des Objekts beschrieben. Das PointPillars [15] Netzwerk wurde als Baseline zur 3D Objekterkennung für dieses Projekt ausgewählt, implementiert und die original Ergebnisse reproduziert. PointPillars ist ein Single Stage Detektor, der einen Voxelisierungsansatz verfolgt. Hierbei ist die Höhe eines Voxels unendlich, sodass sich die namensgebenden Pillars ergeben. Dadurch ist es möglich 2D Convolutions anzuwenden. Somit hat Pointpillars sowohl eine kürzere Laufzeit als auch einen geringeren Speicherbedarf als andere Modelle zur 3D Objekterkennung, was es am geeignetsten für die Anwendung im Autonomen fahren macht.

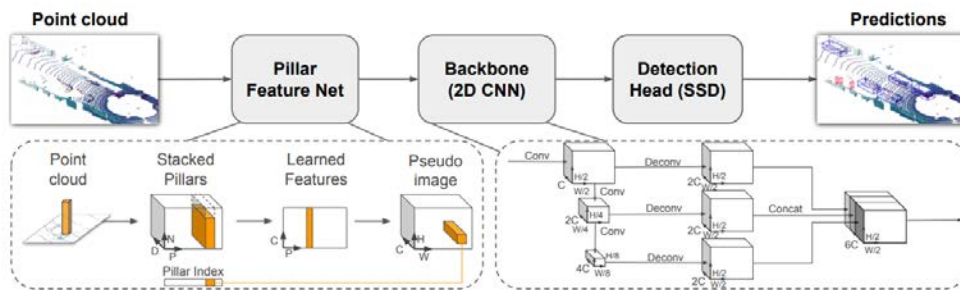


Abbildung 28: Grobe Architektur und Funktionsweise des PointPillar Netzwerkes . Genommen aus der original Publikation [15].

2. Für das Training eines 3D Objektdetektors sind, wie bei Deep Learning-Verfahren üblich, eine große Anzahl annotierter Daten notwendig. Da das Aufnehmen und Annotieren von Daten kosten- und zeitintensiv ist, wird Augmentierung verwendet. Augmentierung beschreibt eine Methode, um einen gegebenen Datensatz künstlich mit weiteren Daten anzureichern, um die Menge und Diversität der Daten zu erhöhen. Augmentierung wird in der Regel (als Abgrenzung zur Simulation) während des Trainingsprozesses "on the fly" angewandt. Aus dem State of the Art wurden die meist verwendeten Augmentierungsverfahren für 3D Objekterkennung zusammengetragen. Diese reichen von einfachen Transformationen wie Rotation oder Skalierung bis zum Einfügen von Objekten in die aktuelle Punktwolke. Zum Anwenden einer dieser Verfahren werden zufällig aus einem vorgegebenen Intervall Parameter gezogen und die Punktwolke entsprechend verändert, zum Beispiel um einen zufälligen Winkel zwischen -90 und 90 Grad rotiert. Der Schwierigkeitsgrad ergibt sich nun aus den erwähnten Intervallen. Je größer diese Intervalle, desto häufiger treten Augmentierungen auf, die von der ursprünglichen Verteilung der Daten abweichen. Alle ausgewählten Augmentierungsverfahren wurden implementiert und auch als Pseudocode zur Verfügung gestellt.

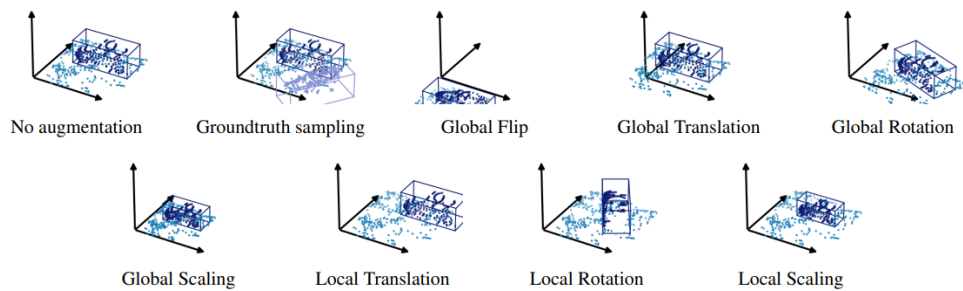


Abbildung 29: Darstellung der verschiedenen Augmentierungsmethoden, die für die Experimente verwendet wurden.

3. Die gesammelten Augmentierungsverfahren wurden auf dem öffentlich zugänglichen Kitti [16] Datensatz ausgewertet. Für eine höhere Vergleichbarkeit wurden zusätzlich zu PointPillars auch drei andere 3D Objektdetektoren getestet. Außerdem wurden einige Experimente auf einem weiteren öffentlichen Datensatz wiederholt. Hierbei stellte sich heraus, dass die Wirkung der Augmentierungsmethoden sowohl vom verwendeten Netzwerkmodell als auch vom Datensatz abhängig ist und somit keine allgemeine optimale Augmentierungsstrategie gefunden werden kann. Stattdessen müssen für jeden vorliegenden Fall individuell die passenden Augmentierungsmethoden und Parameter ermittelt werden.
4. Um die Auswirkungen von verschiedenen Schwierigkeitsgraden der Augmentierung während des Trainings zu untersuchen, wurde ein Hyperparameter eingeführt, der die Modulation der Schwierigkeit (wie in 2. beschrieben) für alle Augmentierungsmethoden gleichermaßen ermöglicht. Dieser Hyperparameter wurde in einer Serie von Experimenten mit dem PointPillars Netzwerk und auf dem Kitti Datensatz auf unterschiedliche Weise während des Trainings moduliert. Dabei konnten keine regelmäßigen positiven Effekte durch die Veränderung des Schwierigkeitsgrads während des Trainings beobachtet werden.

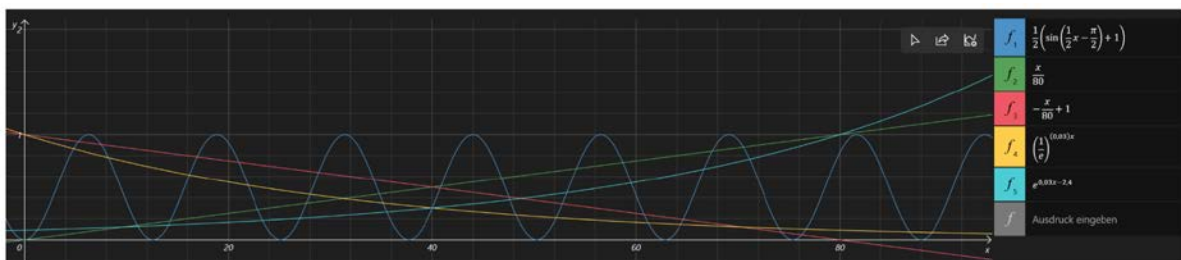


Abbildung 30: Darstellung der verschiedenen Funktionen, mit denen der Schwierigkeits-Hyperparameter um Laufe des Trainings moduliert wurde.

3.3.2.2.2.3 Co-Simulation und live Corner Case Erkennung

3.3.2.2.2.3.1 Zielsetzung

Heutige Simulationen können nicht verwendet werden, um Modelle für menschliches Verhalten zu entwickeln, da Animationen repetitiv und ohne Kontext abgespielt werden.

Daher war das Ziel die Szenen-relevante Datenaufzeichnung zu ermöglichen, indem realistische Interaktionen durch eine gemeinsame Simulation von Mensch und Sensorplattform durchgeführt werden. Weiterhin sollte eine Methode zur Erkennung von Corner Cases entwickelt werden, die

es erlaubt schon während der Aufzeichnung seltene Vorkommnisse auszugleichen, um die Struktur der resultierenden Trainingsdaten zu verbessern.

3.3.2.2.2.3.2 Erzielte Ergebnisse

Es wurde eine Simulation entwickelt mit der komplexe Animationen im virtuellen Kontext aufgezeichnet werden können. Unter Verwendung von Motion Capture und Virtual Reality sind sowohl die Umgebung als auch der eigene Körper virtuell sichtbar und immersiv erlebbar. Weiterhin können verschiedene Ground Truth Arten (RGB, Depth, Semantic/Instance Segmentation, Bounding Box, Human Joints 2D/3D, Meta Daten) generiert werden.



Abbildung 31: Interaktive Datenaufzeichnung (links: Sicht eines Fahrzeugs, rechts: Sicht des Probanden)

Zudem wurde ein Ansatz entwickelt mit dem seltene Posen während der Aufzeichnung erkannt werden können. Seltene Posen können so zeitnah wiederholt werden bis sie nicht mehr als selten erkannt werden, was den Ausgleich von Corner Cases während der Aufzeichnung ermöglicht.

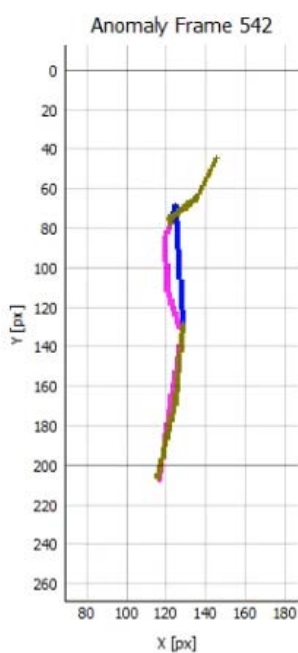


Abbildung 32: Live-Anzeige von Corner Cases während der Aufzeichnung von Motion Capture Daten

Mit Hilfe der Aufzeichnung von realistischen Motion Capture Animationen wurden eigene Datensätze erzeugt. Zur Evaluation des Ansatzes wurde Human Pose Estimation verwendet um nicht-ausgeglichene und ausgeglichene Datensätze zu vergleichen.

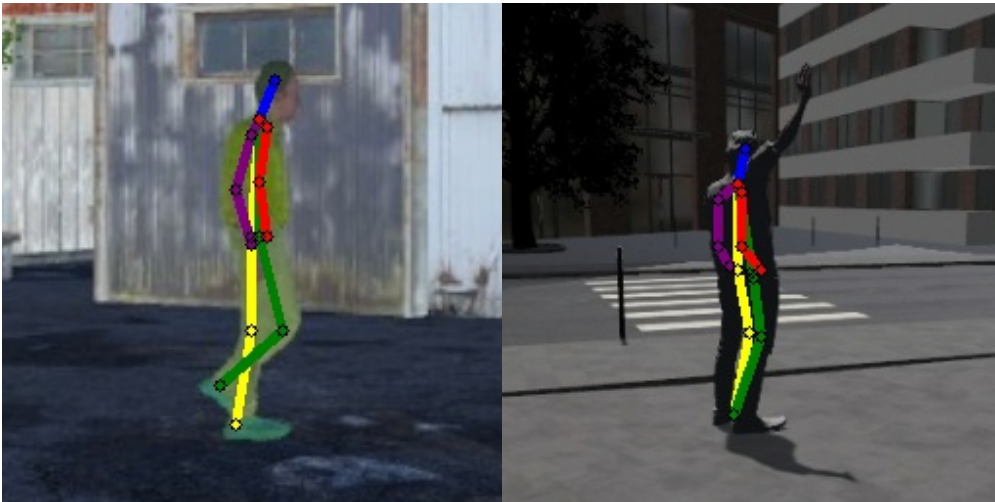


Abbildung 33: Human Pose Estimation (rechts: falsche Erkennung bei Anomalien)

Ist in einem Datensatz eine Abweichung vom durchschnittlichen Inhalt nur sehr kurz enthalten, kann ein Modell diese Daten kaum lernen. Durch diese seltenen Daten sind große Datenmengen für das Training notwendig, was das Training verlangsamt. Durch besser balancierte Datensätze kann diese benötigte Datenmenge reduziert und so das Training beschleunigt werden.

Die verwendete PCKh-Metrik wurde im Detail untersucht, in dem Anomalie-Sektionen in den aufgezeichneten Daten manuell in die Klassen „Anomalie“ und „nicht-Anomalie“ gelabelt wurden. So konnte die geringe Auswirkung von Corner Cases auf die Metrik nachgewiesen werden, was bedeutet, dass die Metrik gute Werte liefert obwohl die enthaltenen seltenen Posen deutlich schlechter erkannt wurden. Diese Labels werden ausschließlich zur Evaluation verwendet und sind nicht für den Ausgleich von Datensätzen notwendig.

3.3.2.2.2.3.3 Gegenüberstellung der vorgegebenen Ziele

Alle angestrebten Ziele wurden erreicht. Die Evaluierung für die Erkennung und den Ausgleich von Corner Cases wurde, bereits während der Aufzeichnung, erfolgreich durchgeführt. Die Evaluierung der Szenen-relevanten Daten steht noch aus.

Beide Ansätze dienen der Reduzierung von Deltas zwischen realen und simulierten Bewegungen.

Zudem wurde an Hand der PCKh-Metrik aufgezeigt, dass Modelle trotz guter Metrik-Werte bei Corner Cases fehlschlagen können, obwohl diese in den Trainingsdaten (kurz) enthalten waren. Es wurde auch gezeigt, dass dieses Verhalten durch die entwickelte Methode ausgeglichen werden kann.

3.3.2.3 AP3.3 Aktives Lernen

In diesem AP wurden Ansätze erörtert, die es einem Algorithmus ermöglichen Trainingsdaten während des Trainings effizient selbst auszuwählen, wodurch sowohl der Labeling-Aufwand reduziert als auch die Trainingszeit verkürzt werden sollen. Die untersuchten Ansätze lassen sich

grob in zwei Kategorien einteilen: Solche, die auf der Grundlage von unterschiedlichen Metriken Situationen herausfiltern, die nicht mit der bisherigen Verteilung der Trainingsdaten übereinstimmen und Solche, die speziell die aktuell falsch klassifizierten Daten berücksichtigen. Es folgt eine kurze Darstellung der Arbeit der Partner, die im Rahmen des Arbeitspakets durchgeführt wurde.

3.3.2.3.1 Erforschung aktiver Lernmethoden auf der Grundlage von Konsistenz und Unsicherheit

3.3.2.3.1.1 Zielsetzung

Hierbei sollte untersucht werden, ob die Inkonsistenz der Vorhersagen diverser ML-Modelle, die für unterschiedliche, aber vergleichbare Aufgaben trainiert wurden (z.B. Objekterkennung und semantische Segmentierung), als Auswahlkriterium für das aktive Lernen (AL) verwendet werden kann. Weiterhin sollte überprüft werden, ob die selbstüberwachten Methoden aus AP3.1 und die in KI-Absicherung identifizierten Unsicherheitsmetriken für die Auswahl zu labelnder Daten geeignet sind.

3.3.2.3.1.2 Erzielte Ergebnisse & Gegenüberstellung

Es wurden verschiedene AL-Verfahren für die semantische Segmentierung und Objekterkennung entwickelt, deren Aquisefunktionen auf Inkonsistenzen und Unsicherheiten in den Prädiktionen beruhen. Die Konsistenz zwischen semantischer Segmentierung und detektierten 2D-Boxen wurde durch zwei Metriken implementiert, welche auf einer gewichteten IoU basieren. Ein weiterer Ansatz verwendet die Konsistenz zwischen prädizierten Segmentierungen für unterschiedlich augmentierte Bilder als Auswahlkriterium. Zusätzlich zu den auf Konsistenz basierenden AL-Verfahren wurden diverse Unsicherheitsmetriken für das aktive Lernen analysiert, u.a. Entropie, Divergenz und Best-vs-Second-Best. Aufgrund der Verzögerung bei der Erstellung des KI-DL Datensatzes wurden alle Ansätze auf dem Cityscapes Datensatz evaluiert.

Dabei hat sich gezeigt, dass konsistenzbasierte AL-Methoden eine effiziente Lösung darstellen und Ergebnisse liefern, die gleichwertig oder besser sind als alternative Ansätze aus der Literatur, wie z.B. unsicherheitsbasierte AL-Verfahren. Jedoch schnitt die Kombination aus semantischer Segmentierung und Objekterkennung für das aktive Lernen nicht besser ab als die ausschließliche Verwendung von Segmentierungen. Dies ist darauf zurückzuführen, dass sich ein effektiver Vergleich unterschiedlicher Modalitäten in der Praxis als schwierig erwiesen hat.

Abgesehen von der Untersuchung der selbstüberwachten Methoden aus AP3.1 wurden alle ursprünglichen Ziele erreicht. Da die Ergebnisse der selbstüberwachten Methoden nicht erfolgversprechend waren, wurden sie hier nicht weiter verfolgt. Stattdessen wurden die bisher genannten Ansätze zusätzlich für regionenbasiertes AL analysiert, d.h. für die automatisierte Selektion zu beschriftender Regionen (anstelle ganzer Bilder). Die Resultate zeigen, dass dieses AL-Regime deutlich effizienter ist.

3.3.2.3.2 Untersuchung aktiver Lernmethoden basierend auf Vorhersagekonfidenz und Szenenanalyse

3.3.2.3.2.1 Zielsetzung

Im Vordergrund dieser Arbeit stand die Erforschung von aktiven Lernmethoden basierend auf Vorhersagekonfidenz und Szenenanalyse. Dabei sollte untersucht werden, welchen Einfluss Pseudolabel auf Basis bereits trainierter Netzarchitekturen auf das weiterführende Training haben und wie Szenenbeschreibungen in AL-Verfahren integriert werden können.

3.3.2.3.2 Erzielte Ergebnisse & Gegenüberstellung

Zunächst wurden die aktiven Lernmethoden *Zufallsauswahl* (Random Sampling), *Unsicherheitsabschätzung*, *Dropout-Unsicherheit* und *Lernverlust* für aktives Lernen (LLAL) implementiert und mit der Baseline des nicht-aktiven Lernalers für das Bildklassifikationsproblem auf dem CIFAR10-Datensatz verglichen. Als weiterer Aspekt wurde untersucht, ob Pseudo-Labeling zu einer Verbesserung der Gesamtleistung führt. Bei dieser Methode des teilüberwachten Lernens werden nicht-annotierte Daten verwendet und nach den ersten Trainingsepochen Pseudoannotationen (Pseudo-Label) mit dem zu trainierenden Netzwerk erzeugt. Die Ergebnisse dieser Studie zeigten, dass die Methode der zufälligen Auswahl (d. h. eine naive aktive Lernmethode) eine ähnliche Leistung erbrachte wie andere moderne aktive Lernmethoden (z. B. Unsicherheitsabschätzung, Dropout-Unsicherheit und Lernverlust für aktives Lernen (LLAL)).

Nach der Bildklassifizierung haben wir unsere Untersuchung auf das Problem der semantischen Segmentierung ausgeweitet. Zu diesem Zweck wurden die Unsicherheitsschätzungsmethode Monte-Carlo-Dropout und die Approximation der Unsicherheit durch die Softmax-Wahrscheinlichkeiten als aktive Lernauswahlmethoden verwendet. Auf der Grundlage der geschätzten Unsicherheit werden unmarkierte Stichproben bewertet und mit Hilfe eines Stichprobenverfahrens ausgewählt. Die Ergebnisse dieser Studie zeigten auch, dass die Ergebnisse der auf Monte-Carlo-Dropout basierenden aktiven Lernmethode denen der auf Zufallsauswahl basierenden aktiven Lernmethode entsprachen.

Schließlich untersuchten wir auch aktives Lernen auf der Grundlage einer Taxonomie zur Szenenbeschreibung. Zu diesem Zweck haben wir eine Taxonomie für die Beschreibung einer Automobilszene abgeleitet und Domäneninformationen für den Sampling-Prozess genutzt. Dies ermöglicht es dem aktiven Lernprozess, die Leistung des Netzwerks für bestimmte Szenenbeschreibungen zu bewerten und anspruchsvolle Szenarien zu erfassen. Darüber hinaus führen wir einen Gewichtungungsverlust ein, der auch Domänenverteilungen einbezieht. Unsere Experimente haben gezeigt, dass die Auswahl allein auf der Grundlage von Domäneninformationen, selbst mit dem modifizierten Verlust, keinen Vorteil gegenüber der Zufallsstichproben-Basislinie bietet.

3.3.2.3.3 Aktives Lernen für die semantische Segmentierung in realistischen Fahrscenarien

3.3.2.3.3.1 Zielsetzung

In diesem Teil war die ursprüngliche Zielsetzung die Kombination von Fahrzeugdaten und Infrastrukturdaten zur automatisierten Detektion spezieller (d.h. seltener, kritischer) Ereignisse und Lernen derselben. Aufgrund der starken Verzögerung in der Erstellung des KI DL Datensatzes mussten Teile des Plans geändert werden. Stattdessen wurde der Fokus auf den Umgang mit realistischen Datenverteilungen im Bereich des autonomen Fahrens gelegt, insbesondere im Hinblick auf Redundanzen.

3.3.2.3.3.2 Erzielte Ergebnisse & Gegenüberstellung

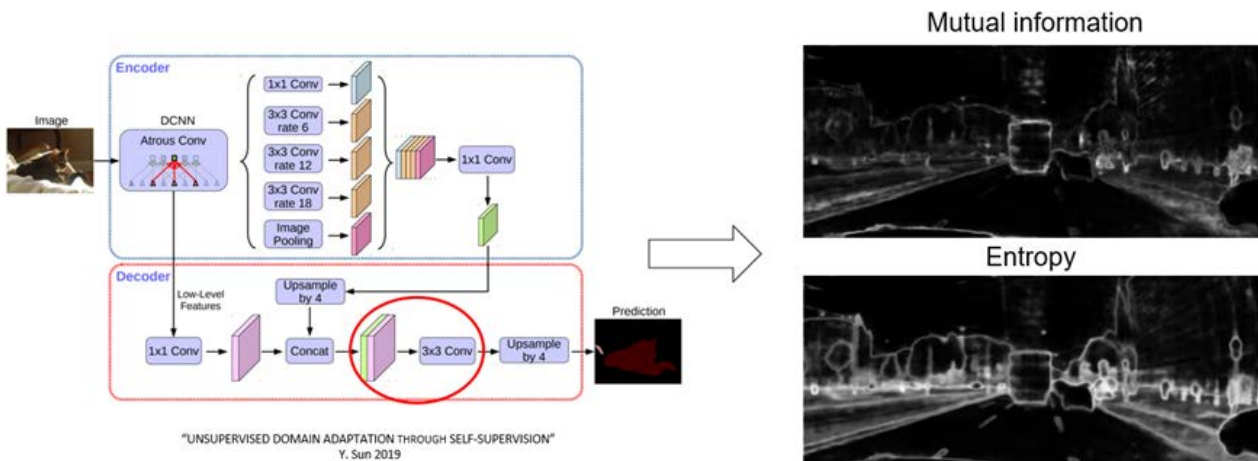


Abbildung 34: Implementierung des Monte-Carlo-Dropout in DeepLab : Links ist die Implementierung des Monte-Carlo-Dropout in das DeepLab-v3-Framework zu sehen. Die Dropout-Layer werden in die rot umrandeten Schichten eingefügt. Nach Sampling von 5 Dropout-Masken ergeben sich die Entropie und Transinformation (engl. Mutual Information) auf der rechten Seite. Die Transinformation ist ein besseres Maß für die Akquisition von Bildern.

Dazu wurden die aktiven Lernverfahren BALD und BatchBALD auf die Aufgabe der semantischen Segmentierung übertragen. Das BatchBALD-Verfahren adressiert neben der Auswahl einer Teilmenge zur manuellen Annotation auch die Herausforderung, dass diese Teilmenge möglichst redundanzfrei bzgl. ihres Informationsgehaltes sein muss. Akquisitionsfunktionen, welche zu akquirierende Datenpunkte (Bilder) unabhängig voneinander bewerten (z.B. auf Basis der Unsicherheit des gegebenen Bildes) neigen dazu, viele Redundanzen zu generieren, wenn Batches an Daten für die Annotation selektiert werden müssen. Das BatchBALD Verfahren hingegen bewertet den Informationsgehalt aller Bilder eines Batches gemeinsam. Das entwickelte Verfahren wurde in verschiedenen Szenarien getestet. Die Annahme war dabei, dass die Performance eines aktiven Lernverfahrens stark von dem Datensatz abhängig ist, von dem neue Trainingsdaten akquiriert werden.

Zunächst wurde das Verhalten auf dem Cityscapes Datensatz untersucht, bei dem es sich um einen sehr gut vorselektierten Datensatz handelt. Das führt dazu, dass auch ohne eine intelligente Akquisitionsfunktion, welche Redundanzen verhindert, bei einer Auswahl mit einfachen Akquisitionsfunktionen Batches entstehen, welche einen positiven Effekt auf das Training haben. Liegt jedoch ein Datensatz vor, welcher viele Redundanzen enthält, wie z.B. der A2D2 Datensatz, sollte eine Akquisitionsfunktion, welche Redundanzen in der Selektion nicht verhindert, weniger gute Ergebnisse liefern und eine Methode wie BatchBALD einen positiven Effekt haben. Das letztere Szenario ist dabei der realistischere Fall im automotiven Kontext.



Abbildung 35: Aufeinanderfolgende Bilder aus den Cityscapes- und A2D2-Datensätzen . Dies zeigt, dass selbst die aufeinanderfolgenden Bilder im Cityscapes-Datensatz unterschiedlich und vielfältig sind, während aufeinanderfolgende Bilder im A2D2-Datensatz sehr ähnlich sind und redundante Informationen enthalten.

Diese Analyse wurde auf einer Vielzahl von Datensätzen ausgeweitet, die sich im Bezug auf ihre Redundanz unterscheiden: Pascal VOC, Cityscapes, A2D2 und aus A2D2 generierte Datenpools mit variierender Redundanz. In diesem Zusammenhang wurde zudem die Wechselwirkung von Methoden des halbüberwachten Lernens mit dem aktiven Lernen betrachtet. Eine weitere Frage, die hierbei untersucht wurde ist, ob die Auswahl von Bildern mit Hilfe von Akquisitionsfunktionen des aktiven Lernens einen positiven Einfluss auf das unüberwachte Lernen auf dem nicht gelabelten Teil der Daten hat und welchen Einfluss der Datenpool auf diese Methoden hat. Die Ergebnisse wurden in dem Paper „Best Practices in Active Learning for Semantic Segmentation“ [17] zusammengefasst. Insbesondere zeigte sich, dass je nach Datenredundanz und Zahl der gelabelten Beispiele ganz andere Active-Learning Verfahren verwendet werden sollten. Bei hoher Redundanz haben sich besonders Core-Set-Methoden als geeignet herausgestellt.

3.3.2.3.4 Aktives Transferlernen mit Domain-adaptierten Daten

3.3.2.3.4.1 Zielsetzung

Hierbei war das Ziel eine Domänenanpassungsmethode zu entwickeln, welche mit möglichst wenigen Labels die Domänenlücke zwischen realen und synthetischen Daten verringert. Dabei sollte anschließend mit Hilfe des aktiven Lernens überprüft werden, ob durch die Wahl von informativen Datenpunkten die Methode noch verbessert werden kann.

3.3.2.3.4.2 Erzielte Ergebnisse & Gegenüberstellung

Im Rahmen des Projektes wurde eine modulare halb-überwachte Domain-Adaptation-Methode mit CycleGAN entwickelt, welche mit Hilfe einer ausgewählten Menge an annotierten Realdaten zur Zielaufgabe (z.B. semantische Segmentierung) geleitet wird. Eine Übersicht über die entwickelte Methode ist in

Abbildung 36 dargestellt. Details unserer Methode sind in [18] veröffentlicht.

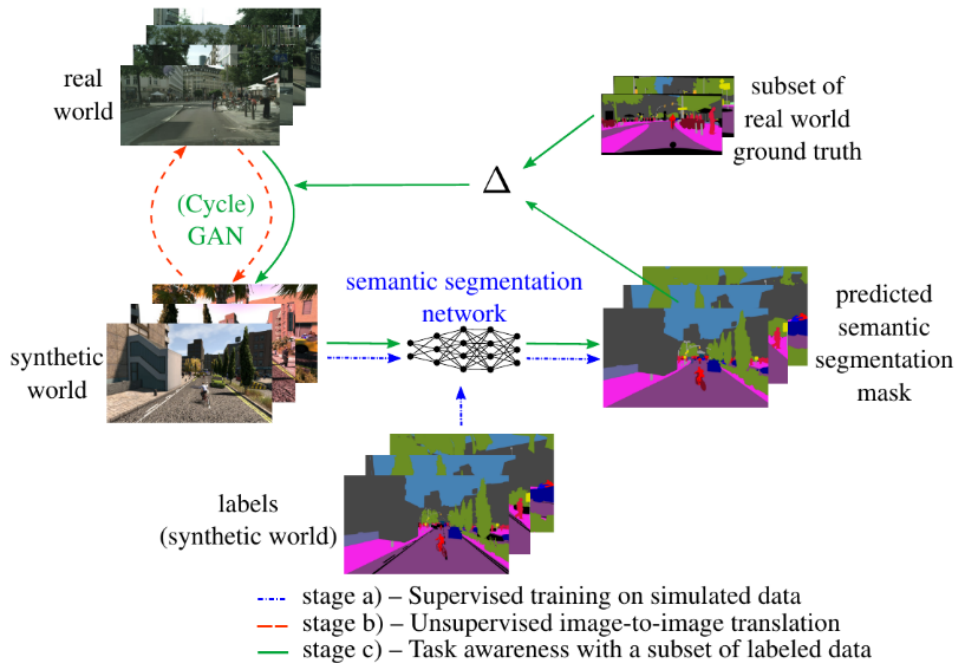


Abbildung 36: Konzept der Domain-Adaptation-Methode: Komponenten und Trainingsaufbau.

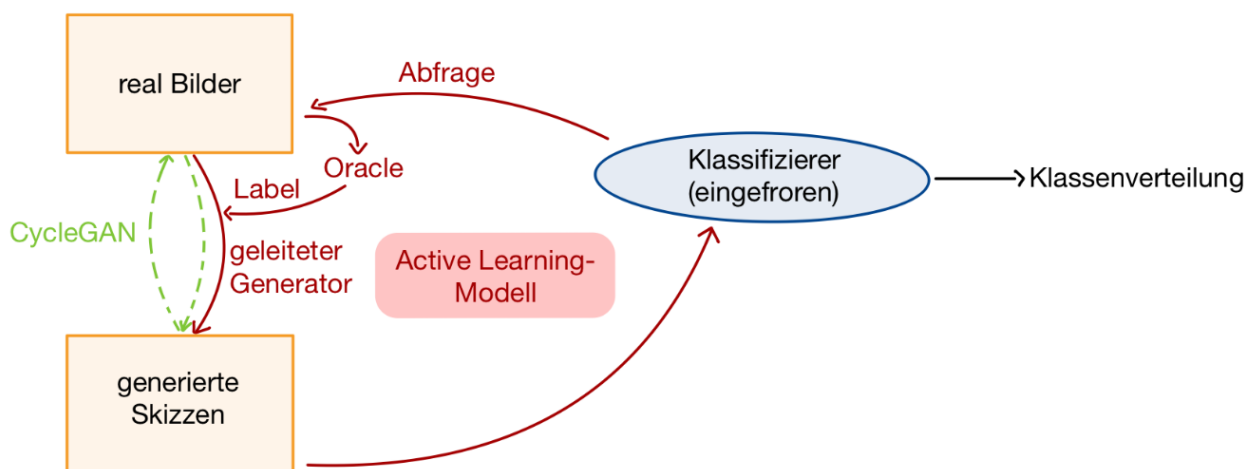


Abbildung 37: Konzept der Domain-Adaptation-Methode: Zyklus des aktiven Lernens für den geleiteten Generator . Bildquelle: [18]

Für die Auswahl der zu annotierenden Datenpunkte, welche für den dritten Schritt des Trainings genutzt werden, haben wir unsere Methode in das AL-Framework (<https://github.com/ac121/deep-active-learning-pytorch>) eingebunden. Der AL Zyklus ist im rechten Teil der Abbildung BUW-1 visualisiert. Dabei wird nur der Generator mit Hilfe der Daten adaptiert, nicht jedoch das auf synthetischen Daten trainierte Netz. Experimente mit dem in AP1.3 erzeugten Carla-Datensatz haben gezeigt, dass schon mit wenigen (max. 10%) annotierten zufällig gewählten Daten eine verbesserte Vorhersagegenauigkeit (gemessen in mIoU) erzielt wird (vgl. Abbildung 38).

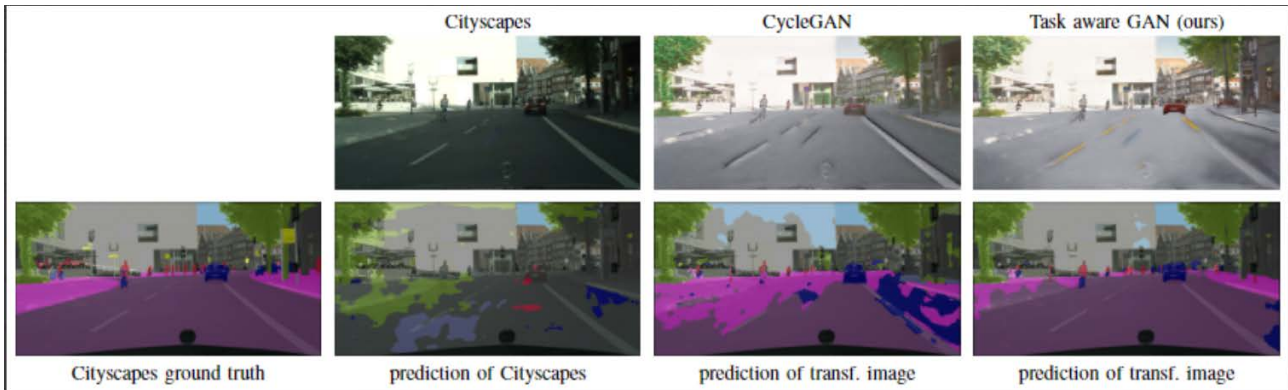


Abbildung 38: Visueller Vergleich der gesteigerten Performance durch unsere halb-überwachte zielaufgabenbedachte Domain-Adaptation-Methode. Bildquelle: [18]

Um Einflüsse von Domänenunterschieden auf aktive Lernverfahren zu untersuchen, wurden zwei semi-synthetische Datensätze erstellt (siehe Abbildung 39). Diese ermöglichten die Implementierung einer Prototypisierungsumgebung für AL von Objektdetektoren. Vergleiche von AL-Experimenten auf den semi-synthetischen Datensätzen mit solchen auf Realdaten wie BDD100k zeigte ähnliches Verhalten, was statistisch ausgewertet wurde. Da die beste Performanz auf beiden Arten von Daten durch primitive Selektionsalgorithmen wie der Softmax-Entropie erzielt wurde, wurde von der Entwicklung spezieller Unsicherheitsmetriken für Domain-adaptierte Daten abgesehen. Die geschaffene Umgebung erlaubte schnelle, automatisierte und detaillierte Untersuchungen konzeptionell verschiedener Architekturen (YOLOv3, Faster R-CNN und RetinaNet) auf insgesamt vier Datensätzen (darunter die beiden Realdatensätze Pascal VOC und BDD100k).

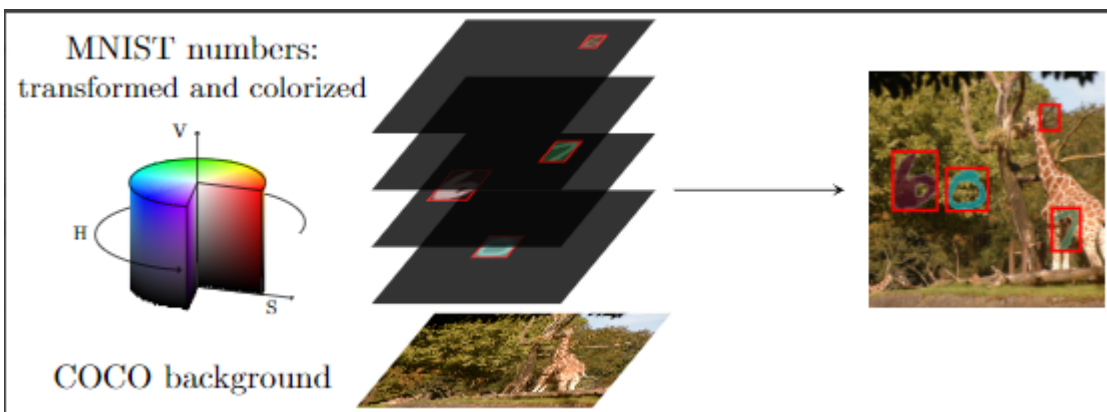


Abbildung 39: Schema der Generierung unserer semi-synthetischen Datensätze basierend auf Hintergründen aus MS COCO und Vordergrundinstanzen aus MNIST und EMNIST. Bildquelle: [19]

Alle Ergebnisse wurden in ihrer Grundidee erreicht. Für das aktive Lernen wurde auf weniger rechenintensive Aufgabenstellungen (Klassifikation) bzw. Datensätze mit geringerer Komplexität (Prototypisierungsumgebung für Objektdetektoren) zurückgegriffen, um schneller Experimente durchführen zu können. Die Aussagekraft der Methode wird dadurch nicht beeinträchtigt, da die Reduzierung der Netzaufgaben keine Reduzierung der Komplexität der Domaintransferaufgabe darstellt.

3.3.2.3.5 Aktives Lernen durch Ausnutzung von Modellunsicherheiten für 3D Objektdetektion & semantische Segmentierung

3.3.2.3.5.1 Zielsetzung aus VHB

1. Verwendung und Entwicklung von Methoden zur Nutzung aleatorischer (Daten-) Unsicherheiten zur Filterung von Trainingsdatensätzen für Objekterkennung (Lidar) oder semantischer Segmentierung, um Datenunsicherheit zu reduzieren und somit die Modellgüte zu verbessern
2. Verwendung und Entwicklung von Methoden zur Nutzung epistemischer (Modell-) Unsicherheiten zur Auswahl von Samples während des Trainings für Objekterkennung (Lidar) oder semantische Segmentierung
3. Anpassungen von Fehlerfunktionen basierend auf Unsicherheiten

3.3.2.3.5.2 Erzielte Ergebnisse & Gegenüberstellung

Bei der Bearbeitung wurde sich auf Punkt 2 der ursprünglich in der VHB genannten Punkte konzentriert.

Zunächst wurde die Anwendung von Aktiven Lernen für die 3D Objekterkennung auf LiDAR Daten untersucht. Das PointPillars [20] Netzwerk wurde als Baseline zur 3D Objekterkennung für dieses Projekt ausgewählt, implementiert und die originalen Ergebnisse reproduziert. Pointpillars weist hierbei sowohl eine kürzere Laufzeit als auch einen geringeren Speicherbedarf als andere Modelle zur 3D Objekterkennung, was es am geeignetsten für die Anwendung im Autonomen Fahren macht. Zudem ist auch der Zeitbedarf beim Training geringer, was für das Aktive Lernen Schema eine notwendige Erleichterung für die Experimente darstellt.

Anschließend an die Implementierung der Baseline wurde die gängigen Aktive Lernen Schleife implementiert. Hierbei wurde die Shannon Entropie als Unsicherheitsmaß verwendet. Monte-Carlo Dropout mit 5 Forward-Passes wurde verwendet, um die Unsicherheiten zu schätzen. Dafür wurden Dropout Layer nach jedem Convolutional Layer im 2D Backbone von PointPillars eingefügt.

Die Experimente wurden auf dem öffentlich verfügbaren Kitti [21] und nuScenes [22] Datensatz durchgeführt. Beide Trainings-Datensätze wurden dem Aktiven Lernen Schema entsprechend in einen bekannten und unbekanntem Teil separiert. Beide Experimente zeigen, dass eine gewöhnliche aktive Lernschleife mit einem Entropie-Auswahlkriterium, wie sie in vielen konventionellen maschinellen Lernansätzen verwendet wird, für diese Aufgabe der 3D-Objekterkennung auf Punktwolken nicht geeignet ist. Ein Grund dafür ist die anspruchsvolle Aufgabe der Objekterkennung. Das größte Problem, das bei den Experimenten auftrat, war die hohe Berechnungszeit von den Experimenten, wie sie hier vorgestellt wurden, da für jede Iteration der aktiven Lernschleife das Netz neu trainiert werden muss.

Anschließend wurden die Auswirkungen von Aktivem Lernen bei der Fusion von Gewichten zur semantischen Segmentierung von Bildern untersucht. Die Fusion von Gewichten bietet eine Möglichkeit zur Annäherung an ein tiefes Ensemble ohne Einbußen bei der Laufzeiteffizienz. Die Gewichte können durch Wiederholung des Trainings mit denselben Daten und Hyperparametern erzeugt werden, einen Unterschied zwischen den Trainingsläufen stellt nur die Initialisierung dar.

Für die Experimente wurde der BDD100k-Datensatz [23] und das DeepLabV3+-Modell [24] verwendet.

Zwei unterschiedliche Auswahlkriterien wurden getestet. Beide basieren auf der mean intersection over union (mIoU) der einzelnen DNNs sowie der mIoU des Orakeltests. Beim Orakeltest werden die True-Positives der Vorhersagen entsprechend den Ground-Truth-Daten fusioniert. Je höher die mIoU nach der Fusion der Vorhersagen, d.h. nach dem Orakeltest, desto mehr unterscheiden sich die Schwachstellen der einzelnen DNNs voneinander. Je unterschiedlicher die Schwachstellen sind, desto höher ist das Potenzial für die Vorhersageleistung bei der Fusion. Die Unterschiede der mIoU-Werte nach dem Orakeltest mit den einzelnen DNNs liefern also gute Auswahlkriterien für die Auswahl der Trainingsdaten. Wir unterscheiden zwischen der Differenz zum niedrigen und höheren mIoU-Wert (diff_low, diff_high).

Aus den durchgeführten Experimenten lässt sich schließen, dass die angewandten Kriterien für die Auswahl der Trainingsdaten auch mit einer Verringerung des mIoU verbunden sind. Es ist zu beobachten, dass im direkten Vergleich mit zufällig ausgewählten Bildern beide Kriterien (diff_low und diff_high) besser abschneiden, aber eine stetige Reduzierung des mIoU zu beobachten ist. Wir schließen daraus, dass die Auswahl der Trainingsdaten nicht zu einem verbesserten mIoU führt. Der höchste mIoU-Wert wird durch die Verwendung des gesamten Validierungsdatensatzes erreicht.

3.3.2.3.6 Erforschung aktiver und Domänen-adaptierter Lernmethoden zur direkten Fahrzeugsteuerung

3.3.2.3.6.1 Zielsetzung

Ziel war es Methoden im Gebiet des aktiven Lernens zu erforschen, um die Notwendigkeit aktueller Deep-Learning-Modelle nach großen Datenmengen zu reduzieren und weiterhin ein Verfahren zur Sim2Real-Adaption zu entwickeln, um gelerntes Wissen aus Simulationsumgebungen auf die Realität zu transferieren. Beides sollte im Kontext der direkten Fahrzeugsteuerung erfolgen.

3.3.2.3.6.2 Erzielte Ergebnisse & Gegenüberstellung

Zur Identifikation von Datenpunkten mit Potenzial zur Performancesteigerung eines Algorithmus wurde ein Ensemble-Model entwickelt, aus welchem die aleatorische und epistemische Unsicherheit einer Trajektorie approximativ extrahiert werden kann. Auf Basis der epistemischen Unsicherheitsschätzung wurde eine Metrik definiert, welche die Neuheit/den Informationsgehalt einer Trajektorie für das Lernen des Umweltmodells bewertet und es erlaubt vor allem bisher schlecht gelernte Daten in das Training zu integrieren. Da die Identifikation von interessanten Daten innerhalb einer Datenbasis einen hohen Rechenaufwand bedarf, wurde ein Prioritized-Experience-Replay-Buffer (PER) für Trajektorien-Daten entwickelt (im Gegensatz zu bereits existierenden PER für einzelne Transitionen). Da das entwickelte Model sowohl als Input Sensordaten entgegen nimmt, als auch Prädikationen im Sensordaten-Raum macht, wird bei der Prädikation die ganze Szene, also sowohl statische Infrastruktur als auch dynamische Verkehrsteilnehmer, berücksichtigt.

Zur Generierung neuer Datenpunkte für das weiterführende Training neuronaler Netze und zur Trajektorienerzeugung mit Ausgangspunkt synthetischer Daten mit Ziel reale Daten wurden *Cycle-Consistent-World-Models for Domain Independent Latent Imagination* (siehe Abbildung 40)

entwickelt. Der Ansatz ermöglicht es, trotz nicht-angeglichener simulierter und realer Daten, ein Prädikationsmodell zu lernen, ohne dass die realen Daten alle Aspekte der simulierten Daten enthalten müssen (z.B. keine Information über die Güte einer Trajektorie in den realen Daten). Der erforschte Ansatz wurde sowohl in CARLA evaluiert, generalisiert aber auch über die Domäne des autonomen Fahrens hinaus.

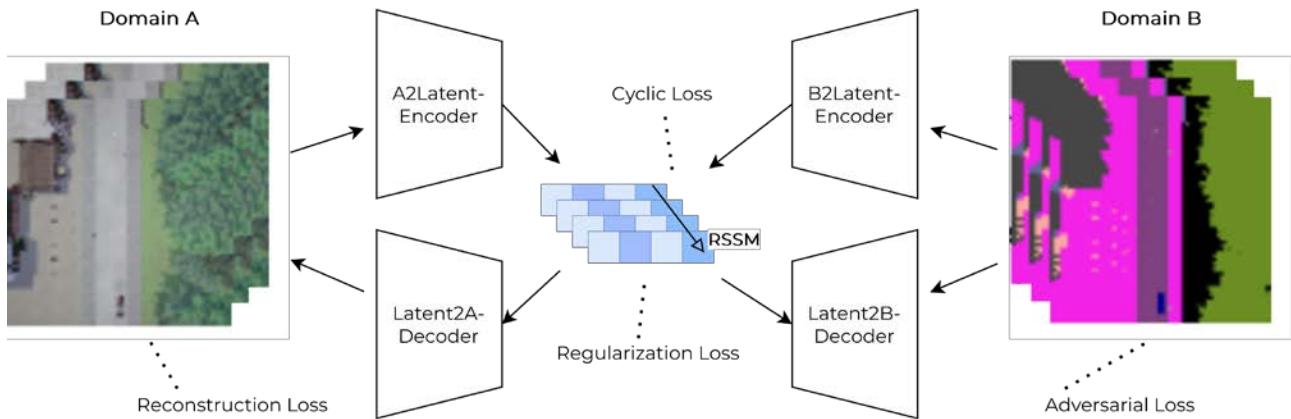


Abbildung 40: Übersicht über die Architektur des Cycle-Consistent-World-Models. Links zu sehen sind die Realdaten, ganz rechts die synthetischen Daten aus der Simulation und zentral die temporale Komponente, welche Vorhersagen in die Zukunft in beiden Modalitäten ermöglicht.

Alle Ziele wurden grundsätzlich erreicht und entsprechende Ergebnisse erzeugt. Um effektiv und effizient durch Experimente iterieren zu können, wurde der CARLA-Simulator anstelle von Realdaten für Training und Evaluation genutzt. Aus rechentechnischen Limitierungen wurden niedriger-dimensionale Sensordaten genutzt als sie typischerweise in realen autonomen System zur Anwendung kommen. Da die entwickelten Ansätze aber methodisch agnostisch gegenüber diesen praktischen Einschränkungen sind, sind die Ergebnisse dennoch übertragbar auf reale Systeme und Szenarien.

3.3.2.3.7 Automatische Erkennung von Corner Cases und gezielte Nachgenerierung von Trainingsdaten

3.3.2.3.7.1 Zielsetzung

Ist in einem Datensatz eine Abweichung vom durchschnittlichen Inhalt nur sehr kurz enthalten, kann ein Modell diese Daten kaum lernen. Zudem wird die Metrik bei solchem kurzen aber schlechten Detektionsverhalten kaum beeinflusst, sofern der Testdatensatz nicht speziell dafür entwickelt wurde.

Daher war es das Ziel eine Methode zu entwickeln, mit der Corner Cases automatisch erkannt und in einer Simulation nachgeneriert werden können.

3.3.2.3.7.2 Erzielte Ergebnisse & Gegenüberstellung

Zur Evaluierung und Lösung dieses Problems wurde eine Simulation entwickelt, mit der kontrollierbare und parametrisierbare visuelle Randomisierungen durchgeführt werden können. Darauf folgt die Generierung verschiedener Ground Truth Arten (RGB, Depth, Semantic/Instance Segmentation, Bounding Box, Human Joints 2D/3D, Meta Daten). So können vielfältige Daten generiert werden, die immer noch inhaltlich korrekt sind aber visuell divergieren.



Abbildung 41: Beispiel der visuellen Randomisierung

Zudem wurde eine Posen-basierte Corner Case Erkennung entwickelt, die Frames mit seltenen Posen identifiziert. Durch eine Verbindung der Corner Case Erkennung mit der Simulation können die relevanten Frames in der Simulation mit einer visuellen Randomisierung nachgeneriert werden. Aus den ursprünglichen und den neuen Frames entsteht ein neuer, ausgeglichener Datensatz.

Mit Hilfe der Aufzeichnung von realistischen Motion Capture Animationen wurde ein eigener Datensatz erzeugt, auf dem die Kombination von Corner Case Erkennung und Simulation angewandt wurde. Die Unterschiede zwischen ursprünglichem und neuem Datensatz wurden mit dem Training und der Auswertung eines Human Pose Estimation Modells evaluiert.



Abbildung 42: Human Pose Estimation (rechts: falsche Erkennung bei Anomalien)

Die verwendete PCKh-Metrik wurde im Detail untersucht, indem Anomalie-Sektionen in den aufgezeichneten Daten manuell in die Klassen „Anomalie“ und „nicht-Anomalie“ gelabelt wurden. So konnte die geringe Auswirkung von Corner Cases auf die Metrik nachgewiesen werden, was bedeutet, dass die Metrik gute Werte liefert obwohl die enthaltenen seltenen Posen deutlich schlechter erkannt wurden. Diese Labels werden ausschließlich zur Evaluation verwendet und sind nicht für den Ausgleich von Datensätzen notwendig.

Alle angestrebten Ziele wurden erreicht. Die Evaluierung wurde für die Erkennung und den Ausgleich von Corner Cases durch Nachgenerierung wurde erfolgreich durchgeführt. Zudem wurde an Hand der PCKh Metrik aufgezeigt, dass Modelle trotz guter Metrik-Werte bei Corner Cases fehlschlagen können, obwohl diese in den Trainingsdaten (kurz) enthalten waren. Sowie, dass dieses Verhalten durch die entwickelte Methode ausgeglichen werden kann.

3.3.2.4 Zielsetzung

Der Fokus von AP3.4 lag auf der Übertragung von in Algorithmen enthaltenem Expertenwissen hin zu Algorithmen, die für den produktiven Einsatz geeignet sind, indem sie durch Knowledge Transfer beispielsweise mehrere Aufgaben in einem Netz vereinen, an Performance oder Robustheit gewinnen oder ihre Rechenkomplexität reduziert wird. Knowledge Transfer ist nicht allein auf die Übertragung von einem neuronalen Netz auf ein anderes beschränkt. Von besonderem Interesse ist auch die Nachbildung des Verhaltens eines klassischen Funktionsmoduls durch ein äquivalentes neuronales Netz, um es mit anderen neuronalen Netzen verketteten zu können. Dies erschließt die Möglichkeit, ein bestehendes System an geänderte Anforderungen, z.B. an die unterschiedlichen Deltas, die in den Gesamtprojektzielen definiert sind, mittels End-to-End-Optimierung anzupassen.

Knowledge Transfer bezeichnet kein spezifisches Verfahren, sondern viel mehr eine Gruppe diverser Methodiken, welche für die Repräsentation von Wissen und die dessen Weitergabe zwischen Algorithmen eingesetzt werden. Beides sollten in den Forschungsarbeiten im Rahmen von AP 3.4 gleichermaßen berücksichtigt werden. Zwischen verschiedenen Anwendungsfällen gibt es große Unterschiede in der Art der Wissensrepräsentation und entsprechend auch den Verfahren, die zum Einsatz kommen können, um Wissen zu übertragen. Bei neuronalen Netzen beispielsweise liegt das Wissen des Algorithmus in der Netzarchitektur und den gelernten Gewichten und kann durch die Generierung von sogenannten Pseudolabels für das Training eines anderen Netzes extrahiert werden. Diese weit verbreitete Knowledge-Transfer-Methode wird allgemein hin auch als Teacher-Student-Verfahren bezeichnet. Bei anderen Anwendungsfällen, z.B. Sensor Fusion, ist Knowledge Transfer komplexer und erfordert eine vollständige

Nachbildung, sogenannte Surrogate Networks, von etablierten, modellbasierten Ansätzen durch lernbare Gegenstücke.

Ziel des Arbeitspakets war die Implementierung, Bewertung und Dokumentation bestehender Knowledge-Transfer-Methoden, deren Optimierung und Analyse, sowie die Neu-Modellierung von klassischen Methoden mittels neuronaler Netze. Übergeordnete Motivation war eine umfangreiche Methodensammlung für breit gefächerte Anwendungsfälle (Objektdetektion, Segmentierung, Sensorfusion, Trajektorienprädiktion).

3.3.2.4.1 Ausnutzung zeitlicher Redundanz zur Reduktion von Rechenkomplexität

3.3.2.4.1.1 Zielsetzung

Das ursprüngliche Ziel in AP3.4 war die Ausnutzung zeitlicher Redundanz zur Reduktion von Rechenkomplexität. Dabei war zunächst ein System zu entwickeln, bei dem komplexe Architekturen nur auf Key-Frames ausgeführt werden und zwischen diesen z.B. durch die Berechnung von optischen Flüssen die Prädiktionen der CNNs weiter propagiert werden. Um dieses Ziel zu erreichen, wurden verschiedene Verfahren zur Berechnung von optischen Flüssen evaluiert. Dabei zeigte sich, dass Ansätze zur Berechnung von dichten optischen Flüssen eine ähnliche Laufzeit haben, wie die Ausführung der komplexen CNN Architekturen und mithin keinen Vorteil bieten, da man in diesem Fall auch direkt die CNN Ansätze ausführen kann.

3.3.2.4.1.2 Erzielte Ergebnisse

Schwerpunkt in diesem AP lag mithin alternativ auf dem Thema Transfer Learning zwischen der Kamera und LiDAR Domäne. Dazu wurde eine Methodik entwickelt, um Ansätze des Self-Trainings, die im Detail in AP3.1 untersucht wurden, mittels dieses Transfers zu verbessern.

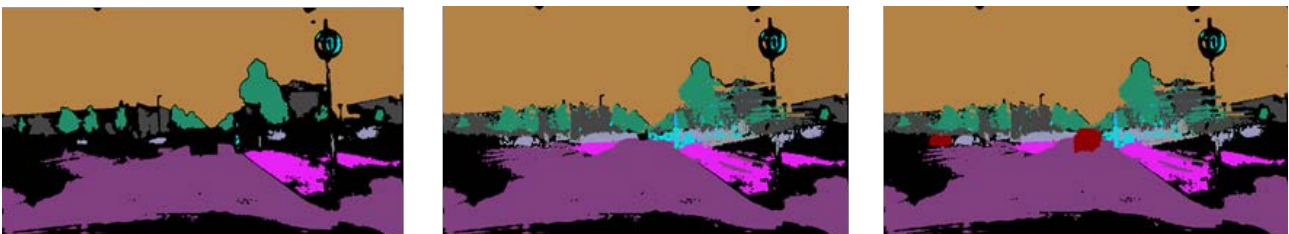


Abbildung 43: Überlagerung der Segmentierung aus dem ersten Bild der Sequenz (ohne dynamische Objekte, links), mit der Rückprojektion aus der aggregierten Punktwolke (ohne dynamische Objekte, Mitte) und der dynamischen Objekte aus dem ersten Bild der Sequenz (rechts).

Die Qualität der durch dieses Verfahren gewonnenen Daten hängt maßgeblich von der Kalibrierung der Daten ab. Da zu Beginn der Arbeiten zu diesem Thema noch keine Projektdaten vorlagen, wurden eigene Daten verwendet. Die in TP1 gewonnenen Erfahrungen im Bereich der Kalibrierung von LiDAR- und Kamerasensoren wurden genutzt, um die Kalibrierung der eigenen Fahrzeuge weiter zu optimieren.

3.3.2.4.2 Modellierung von Verarbeitung variabler Ein- und Ausgangsformate und -strukturen

3.3.2.4.2.1 Zielsetzung

Ziel der Arbeiten im Arbeitspaket 3.4 war die Entwicklung von Methoden für die Übertragung von Wissen aus klassischen Algorithmen hin zu neuronalen Netzen. Insbesondere sollten gelernte Alternativen für unterschiedliche klassische Verfahren untersucht werden, welche oft zur

Nachverarbeitung der Ergebnisse neuronaler Netze eingesetzt werden. Diverse Komponenten eines klassischen Systems für Autonomes Fahren, insbesondere Objektdetektion und -tracking sowie Trajektorienprädiktion, sollten im Rahmen der Forschungsziele anpassungsfähig und optimierbar gemacht werden und dadurch zentrale Erkenntnisse für einen nachhaltigen praktischen Einsatz gewonnen werden.

3.3.2.4.2 Erzielte Ergebnisse

Ergebnis der Arbeiten ist ein Graph Convolutional Network basierter Algorithmus, welcher eine maschinell lernbare Non-Maximum-Suppression durch Message Passing zwischen Detektionskandidaten modelliert. Non-Maximum Suppression ist ein weit verbreiteter heuristischer Algorithmus, der eingesetzt wird, um Duplikate in bild-basierter Objektdetektion zu reduzieren. Als Kriterium für die Identifizierung und Entfernung von Duplikaten dient in der Regel lediglich die Überlappung mehrerer Detektionen unabhängig von dem zugrundeliegenden Bildinhalt. Das entwickelte Netzwerk erreicht bei reduzierter Inferenzzeit eine bessere Performance als klassische Non-Maximum Suppression und realisiert damit den letzten erforderlichen Schritt zu einer vollständig maschinell gelernten Pipeline für Objektdetektion.

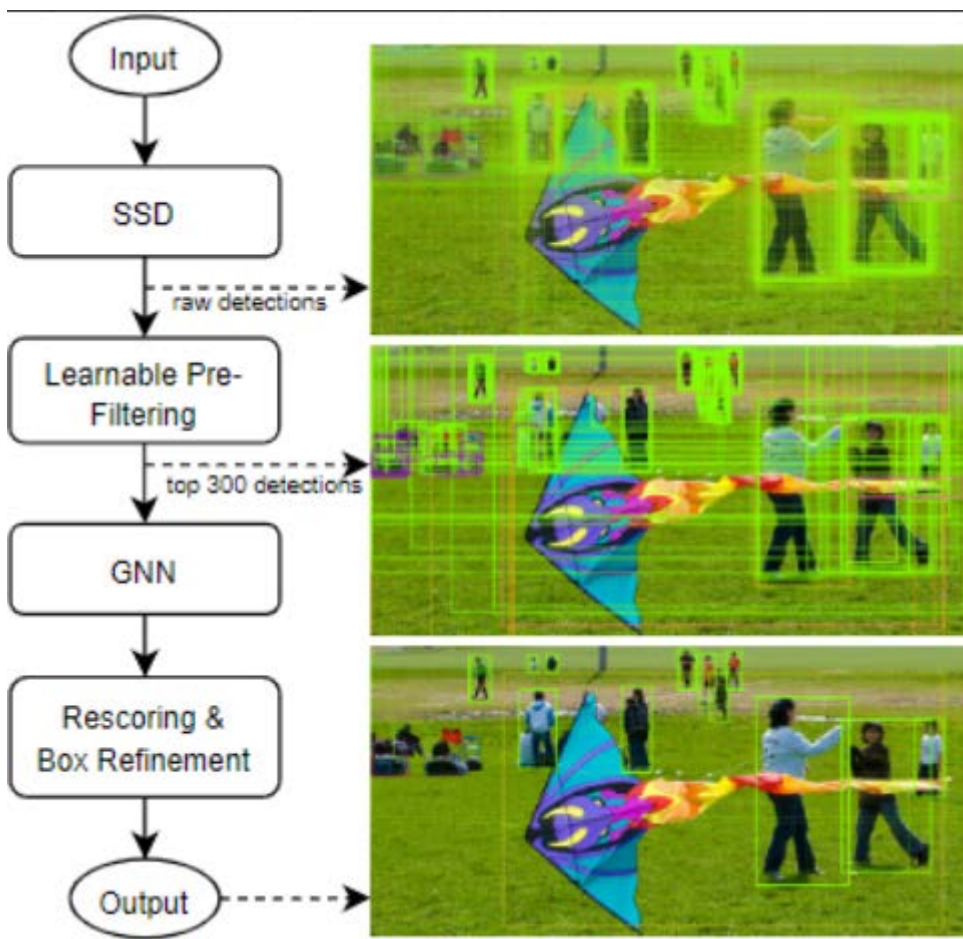


Abbildung 44: Pipeline für das Entfernen von Duplikaten eines Single-Shot-Detektors mithilfe Graph Neural Network basierter Non-Maximum Suppression.

Ferner resultierte die Forschung in einen neuartigen Transformer-basierten Ansatz für Multi-Camera-Fusion, welcher die Informationen mehrerer Sensoren bereits während der Bildung von initialen Objekthypothesen konsolidiert und nicht erst durch zeitliche Integration wie für probabilistische Filter üblich. Wesentlicher Bestandteil von SpatialDETR ist ein neuronales

Attention-Modul mit sogenannter Spatial Awareness, das während des Trainings eine global einheitliche und von extrinsischen Parametern unabhängige Wissensrepräsentation lernt, die vom Ziel-Sensor-Setup unabhängig ist. Insbesondere für nahe, große Objekte, welche im Sichtbereich mehrerer Kameras liegen, bewies die Methode großes Potential als Alternative oder Ergänzung zu etablierten physikalisch-fundierte Single-Sensor-Unsicherheitsmodellen, welche nicht selten ein aufwendiges Post-Processing erfordern.

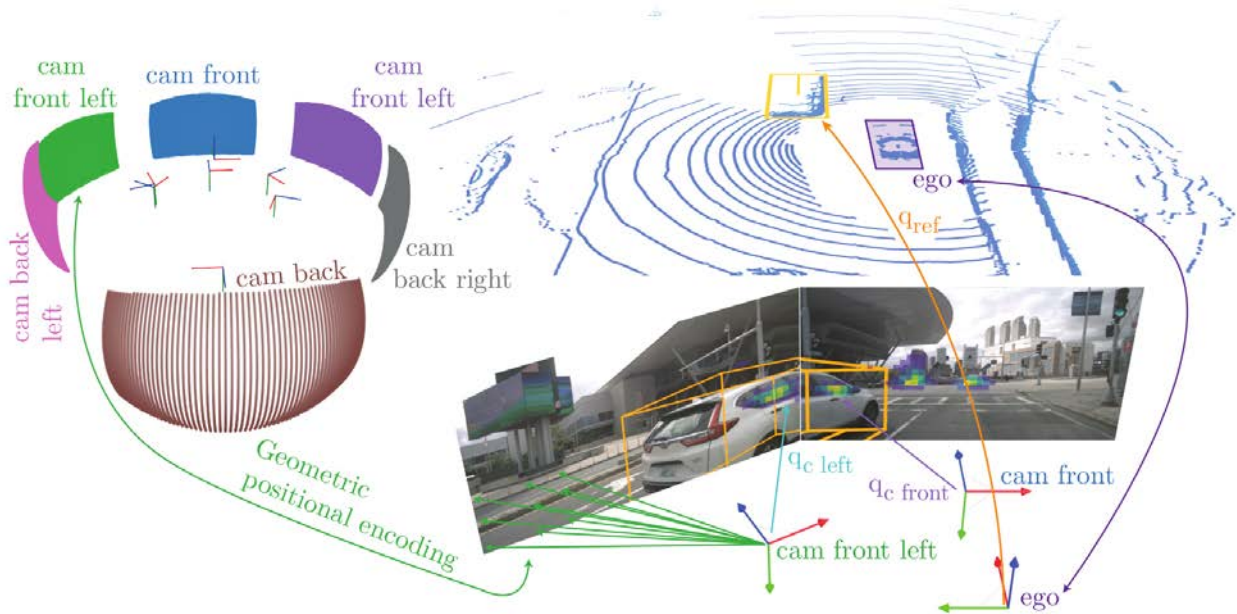


Abbildung 45: Globale Cross-Sensor-Attention mit Spatial Awareness . Die Fusion von 3D-Information individueller Bilder nutzt den globalen Richtungsvektors jedes Pixels.

Eine weitere Arbeit aus dem Bereich Trajektorienprädiktion brachte zukunftsweisende Erkenntnisse im Bereich kartenloser Bewegungsvorhersage und setzte neue Maßstäbe. Ergebnis ist ein CRAT-Pred, ein Expertenalgorithmus auf Basis eines Crystal Graph Convolutional Neural Networks mit Multi-Head Self-Attention, welches den Einfluss zwischen Verkehrsteilnehmern im Hinblick auf ihre zukünftigen Trajektorien schätzt. Eingesetzt als Kriterium für das Pruning von Trainingsdaten durch Entfernen von Agenten mit niedriger Relevanz, erreichte ein Prädiktionsnetz State-Of-The-Art-Performance trotz einer deutlich geringeren Anzahl an Parametern als vergleichbare, ähnlich performante Methoden.

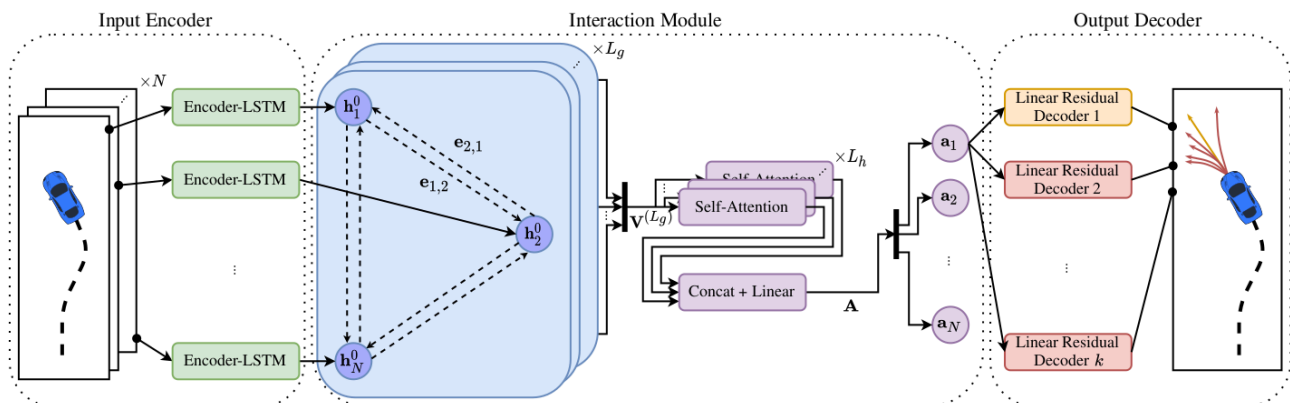


Abbildung 46: Überblick über CRAT-Pred. Das Interaction Module modelliert soziale Interaktionen zwischen mehreren Agenten durch ein Graph Neural Networks und Multi-Head Self-Attention.

3.3.2.4.2.3 Gegenüberstellung

Alle angestrebten Ziele wurden grundsätzlich erreicht. Es wurden diverse Surrogate-Netze für klassische Methoden entworfen, implementiert und evaluiert und dokumentiert. Die Arbeiten zu zeitlicher Sensorfusion folgten dem Trend zum breiten Einsatz von Transformer-Architekturen und beschränkten sich dafür auf synchronisierte Bilddaten, welche keine zeitliche Integration, wie ursprünglich formuliert, erforderten.

3.3.2.4.3 Knowledge Distillation

3.3.2.4.3.1 Zielsetzung

Ziel der Arbeit war es zu untersuchen weshalb *Knowledge Distillation* Verfahren in der kontinuierlichen Semantischen Segmentierung so viel effektiver sind wie in der Klassifikation und wie Vergessen in kontinuierlichen trainierten Netzwerken auftritt.

Die Arbeiten in AP2.1 legten die Vermutung nahe, dass sich Vergessen in neuronalen Netzen für die semantische Segmentierung und der Klassifikation womöglich unterschiedlich äußert. Insbesondere da Verfahren, die auf „Knowledge Distillation“ beruhen, deutlich unterschiedliche Ergebnisse im Klassen-inkrementellen Lernen erzielen. Aus diesem Grund wurde untersucht, wie sich das Vergessen in neuronalen Netzen für die semantische Segmentierung äußert und wie dies womöglich unterschiedlich zu der Klassifikation ist.

3.3.2.4.3.2 Erzielte Ergebnisse

In den Untersuchungen wurde besonderes Augenmerk auf die „Hintergrund-Klasse“ gelegt, welche für irrelevante Objekte im Bild verwendet wird. Die Ergebnisse zeigten, dass der semantische Wechsel der Bedeutung der Hintergrund-Klasse einer der Hauptgründe für Vergessen im Klassen-Inkrementellen Lernen ist. Dies äußert sich, indem die ursprünglichen Klassen nun als Hintergrund erkannt werden, während ausschließlich die zuletzt gelernten Klassen korrekt klassifiziert werden. Ein näherer Blick in die gelernten Repräsentationen des neuronalen Netzes legte jedoch nahe, dass dabei die ursprünglich gelernten Merkmale für diese Objekte nicht vergessen werden, sondern diese Merkmale fälschlicherweise nun als Indiz für den Hintergrund verwendet werden und nicht als Indiz für die zuvor gelernten Klassen. Diese Hypothese konnte mittels verschiedener Visualisierungsmethoden bestätigt werden.

Einige der zuvor gelernten Klassen wurden allerdings nicht als Hintergrund, sondern als eine der neuen Klassen erkannt. Eine nähere Untersuchung der Klassen zeigte, dass dies ausschließlich Klassen betraf, die visuell verwandte Klassen in den beiden Klassen-Inkrementen hatten. So wurde beispielsweise die Klasse „Bus“, der Teil der zuerst gelernten Klassen war, mit der Klasse „Zug“ verwechselt, die als neue Klasse dazu gelernt worden ist. Diese Verwechslung wird dadurch verursacht, dass die Klassen nicht gleichzeitig gelernt werden, weshalb das Modell keine Merkmale lernt um die beiden Klassen voneinander zu unterscheiden. Dieses Phänomen ist als „Inter-Task-Confusion“ bekannt und könnte eine mögliche Beschränkung von kontinuierlichen Lernverfahren sein, die nicht auf Replay basieren.

Zuletzt wurde noch belegt, dass Knowledge Distillation in der Klassen-inkrementellen semantischen Segmentierung aufgrund des Hintergrunds eine ähnliche Funktionsweise wie

Replay-basierte Verfahren in der Klassifikation haben. Dadurch lässt sich die ursprüngliche Diskrepanz der Methoden erklären.

3.3.2.4.3.3 Gegenüberstellung

Die angestrebten Ziele wurden erreicht. Es wurden allerdings anstelle eines domänen-inkrementelles Setups, zwischen verschiedenen Ländern, ein klassen-inkrementelles Setup für die Untersuchung herangezogen, da dort die Effekte des katastrophalen Vergessens ausgeprägter sind.

3.3.2.4.4 Teacher-Student-Verfahren für Multi-Task Learning und Self Supervised Learning

3.3.2.4.4.1 Zielsetzung

In diesem Aufgabenpaket wurden zwei Problemstellungen verfolgt. Zum einen wurde die Zusammenlegung verschiedener Netze zu einem Netz im Bereich der Computer Vision durch Multitasking Netze erprobt und evaluiert. Zweitens wurde mit Self Supervised Learning gearbeitet, um ein existierendes Multi-Task-Netz für Bildrekonstruktion und Klassifizierung um die Aufgabe der Instanzsegmentierung zu erweitern bzw. das gelernte Wissen zu transferieren.

3.3.2.4.4.2 Erzielte Ergebnisse

Im Bereich der Multitask Netze wurde ein Algorithmus zur gleichzeitigen Prädiktion von Semantischer Segmentierung, Bounding Box Detektion und Posenschätzung erstellt [25]. Es wird ein gemeinsames Backbone verwendet, sodass der Großteil der Berechnungsoperationen gegenüber einer parallelen Verwendung von Netzen eingespart werden kann. Lediglich die Köpfe des Netzes laufen nebeneinander. Dadurch lässt sich die Geschwindigkeit der Inferenz um ein Vielfaches beschleunigen.

Er wurde auf dem MS COCO Datensatz trainiert und evaluiert. Hierbei wurde CenterNet [26] für die entsprechenden Tasks erweitert.

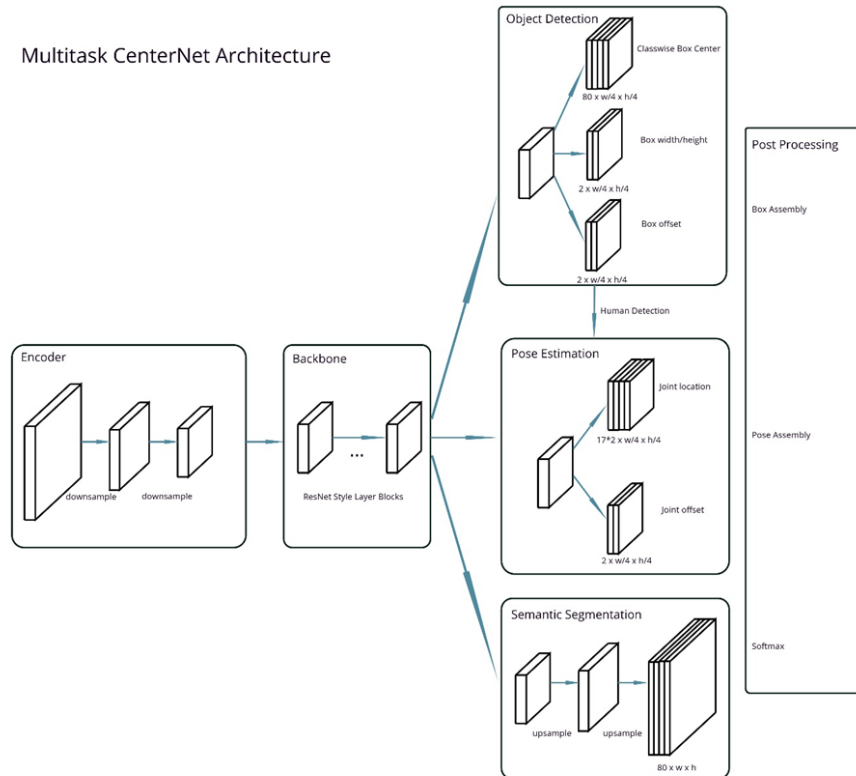


Abbildung 47: Blockdiagramm des Aufbaus des Multitask-CenterNet für den COCO Datensatz.

Die Resultate lassen erkennen, dass trotz der Einsparung der Berechnungsoperationen nahezu gleichwertige Detektionsperformanz erzielt werden kann, und zum Teil auch marginal übertroffen wird. Die gilt für eine eingeschränkte Auswahl an Task-Kombinationen.

Network	Segmentation mIoU	Detection mAP
Only Seg	49,0%	NA
Only Det	NA	36.3%
Seg + Det	48,9%	36.7%

Abbildung 48: Erzielte Ergebnisse für Netze mit Detektions- und/oder Segmentierungsprädiktionen.

Darüber hinaus lässt sich das Netz auch mit Daten aus verschiedenen Datensätzen trainieren. Es wird hierfür ein Spezialistennetzwerk auf jedem Datensatz trainiert, welcher Inferenzen auf dem Ziel-Datensatz durchführt. Die Zusammenlegung der Inferenzen aller Spezialistennetzwerke lässt sich nun verwenden, um ein Multitask Netzwerk für alle Tasks zu trainieren.



Abbildung 49: Multitask-Netzwerk trainiert mit Pseudolabels verschiedener Spezialisten Netzwerke.

Im Bereich des selbstüberwachten Lernens wird Masked-Autoencoder (MAE) [27] zur Instanzsegmentierung verwendet. Hierbei wird ein zu 75% Maskierter Autoencoder darauf trainiert, ein Bild zu rekonstruieren. Wenn der Autoencoder trainiert ist, ist er in der Lage neben Bildrekonstruktionen auch andere Downstream Tasks wie Bildklassifikation zu betreiben. Die Aufgabe in AP3.4 besteht darin, ihn für Instanzsegmentierung (Vordergrund/Hintergrund-Trennung) zu verwenden. Hierfür wird der Fehler der Bildrekonstruktion eines Trainierten Netzes berechnet. Dieser Fehler kann verwendet werden, um die Zugehörigkeit der sichtbaren Patches zu den Unsichtbaren (maskierten) zu schätzen. Ein Graph wird erstellt, welcher Patches als Knoten und Fehler als Kanten hat. Schließlich kann durch den Community Detection Algorithmus

[28] die Zugehörigkeit von Patches zueinander zu Clustern führen. Diese Cluster werden wieder in den Autoencoder gefüttert, sodass sich größere semantische Gruppen bilden.

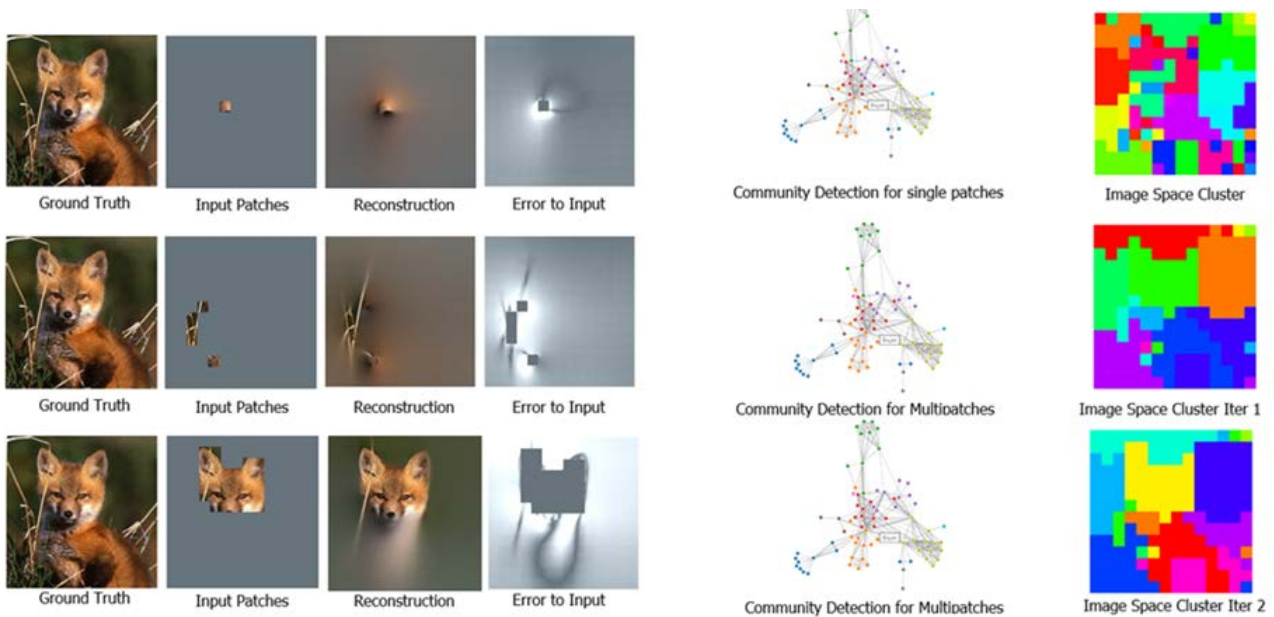


Abbildung 50: Iterative Erstellung größerer Graphen bis zur vollständigen Segmentierung

3.3.2.4.4.3 Gegenüberstellung der vorgegebenen Ziele

Die geplanten Ziele wurden erreicht. In Multitask Centernet konnte die Zusammenlegung mehrerer Netze und Verringerung der Latenz bei gleichbleibender Performanz gezeigt werden. In Masked Autoencoder wurde ein Ansatz gezeigt, Bildrekonstruktion neben Klassifikation auch für Instanzsegmentierung zu verwenden.

3.3.3 Zielerreichung

Im Teilprojekt Didaktik wurden Entwicklungsfortschritte erreicht und wissenschaftliche Erkenntnisse gewonnen, welche einen effizienten Umgang mit vorhandenen Daten sowie die Nutzung von Daten jenseits der vollständig annotierten Daten ermöglichen. Weiterhin wurden Methoden zur Verbesserung und effizienten Gestaltung des Trainingsprozesses vorgelegt.

Im Teilüberwachten Lernen (AP 3.1) wurden Arbeiten im Bereich der Semantischen Segmentierung und der Posenschätzung durchgeführt und zu einem erfolgreichen Abschluss gebracht. Hier konnte durch die effiziente Nutzung von nicht gelabelten Daten die Präzision der KI-basierten Perzeption signifikant gesteigert werden. Die Domänenanpassung, welche notwendig wird, wenn sich Trainingsdomäne (Daten) statistisch von der Anwendungsdomäne (Fahrertrieb) unterscheiden, bildet einen weiteren Schwerpunkt des AP 3.1, welche insbesondere mit einem Übersichtsartikel zum State of the Art eingehend bearbeitet wurde. Die Forschung zu Self-Labeling, also der Verwendung der Vorhersagen eines Segmentierungsnetzes anstelle 'echter' gelabelter Trainingsdaten erfüllte die in diese Technik gesetzten Hoffnungen nicht.

Im Bereich der Trainingsorganisation konnte die Anzahl der Parameter tiefer neuronaler Netze, bei gleichbleibender Leistungsfähigkeit, erheblich reduziert und das Training, auch durch Texturreduktion, robustifiziert werden. Ebenfalls wurde die gezielte Produktion von Daten zur

unterstützung des Trainingsprozesses mit Hilfe von Motion Capturing in der Posenerkennung erfolgreich durchgeführt. Die gesetzten Ziele des AP wurden somit erreicht.

Im Arbeitspaket 3.3 **Aktives Lernen** wurden für eine breite Auswahl von Perzeptions-Modellen Akquisestrategien erfolgreich entwickelt, welche den Lernerfolg mit besonders effizienten Daten gezielt unterstützen. Weiterhin konnte die Nutzung von Daten aus anderen Domänen, wie z.B. synthetischen Daten, wie geplant durchgeführt werden. Nicht zuletzt wurden Methoden zur Erkennung von außergewöhnlichen Ereignissen, so genannten Corner Cases, vorgestellt.

Die im Arbeitspaket 3.4 **Knowledge Transfer** durchgeführten Untersuchungen betreffen zum einen die Übertragung von Wissen in der Zeitlichen Dimension der Szenenerkennung sowie den Übertrag von bereits erlerntem Wissen eines Neuronalen Netzwerkes oder eines physikbasierten Bilderkennungsalgorithmus auf ein anderes Neuronales Netz. Die hier durchgeführten Arbeiten erreichen somit die Ziele der Vorhabenbeschreibung ebenfalls.

Mit geringfügigen Einschränkungen, da die Self-Labeling Methode sich in der Durchführung als wenig effizient erwies und daher durch wirkungsvollere Ansätze ersetzt wurde, erreichen die vorgelegten Ergebnisse die in der Vorhabenbeschreibung für das Teilprojekt Didaktik gesetzten Ziele. Im Teilprojekt konnten vielversprechende Entwicklungsfortschritte erzielt werden und es wurden zahlreiche wissenschaftliche Publikationen vorgelegt.

3.3.4 Referenzen

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
- [2] U. Trottenberg, C. W. Oosterlee,, and Anton Schüller. Multigrid. Academic Press, San Diego, 2001.
- [3] Juncai He and Jinchao Xu. MgNet: A unified framework of multigrid and convolutional neural network. *Science China Mathematics*, 62(7):1331-1354, July 2019
- [4] Moshe Eliasof, Jonathan Ephrath, Lars Ruthotto, and Eran Treister. "MGIC: Multigrid-in-Channels Neural Network Architectures", *ArXiv abs/2011.09128* , November 2020
- [5] Antonia van Betteray, Matthias Rottmann and Karsten Kahl. "MGiaD: Multigrid in all dimensions. Efficiency and robustness by coarsening in resolution and channel dimensions." *ArXiv abs/2211.05525* (2022)
- [6] Kahl, K., Kintscher, N. Automated local Fourier analysis (aLFA). *Bit Numer Math* 60, 651-686 (2020). <https://doi.org/10.1007/s10543-019-00797-w>
- [7] C. Schmaltz, J. Weickert und A. Bruhn, "Beating the Quality of JPEG 2000 with Anisotropic Diffusion" 2009.
- [8] J. Weickert, M. Welk, und M. Wickert, "L2-Stable Nonstandard Finite Differences for Anisotropic Diffusion," in *Scale Space and Variational Methods in Computer Vision*, 2013, pp. 380-391.

- [9] J. Weickert und M. Welk, "Tensor Field Interpolation with PDEs," in Visualization and Processing of Tensor Fields, J. Weickert and H. Hagen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 315-325. doi: 10.1007/3-540-31272-2_19.
- [10] A. Geiger, P. Lenz, C. Stiller, und R. Urtasun, "Vision meets robotics: the KITTI dataset," The International Journal of Robotics Research, vol. 32, pp. 1231-1237, Sep. 2013, doi: 10.1177/0278364913491297.
- [11] C.-Y. Wang, I.-H. Yeh und H.-Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks". arXiv, 2021. doi: 10.48550/ARXIV.2105.04206.
- [12] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan und Trevor Darrell, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning". arXiv, 2020, doi: 10.48550/1805.04687.
- [13] G. Varma, A. Subramanian, A. Namboodiri, M. Chandraker und C V Jawahar, "IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments". arXiv, 2018, doi: 1811.10200.
- [14] A. Gaidon, Q. Wang, Y. Cabon, und E. Vig, "Virtual Worlds as Proxy for Multi-Object Tracking Analysis" 2016, pp. 4340-4349. [Online]. Available: <https://elib.dlr.de/105154/>.
- [15] Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [16] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012.
- [17] "Best Practices in Active Learning for Semantic Segmentation" Mittal and Niemeijer et al. arXiv 2023
- [18] Annika Mütze, Matthias Rottmann, and Hanno Gottschalk. Semi-Supervised Domain Adaptation with CycleGAN Guided by Downstream Task Awareness. In Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP, pages 80-90. INSTICC, SciTePress, 2023.
- [19] Tobias Riedlinger, Marius Schubert, Karsten Kahl, Hanno Gottschalk, and Matthias Rottmann. Towards Rapid Prototyping and Comparability in Active Learning for Deep Object Detection. arXiv preprint, 2022.
- [20] Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [21] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012.

- [22] Caesar, Holger, et al. "nusenes: A multimodal dataset for autonomous driving." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [23] Yu, Fisher, et al. "Bdd100k: A diverse driving dataset for heterogeneous multitask learning." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [24] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
- [25] F. Heuer, S. Mantowsky, S. Bukhari und G. Schneider, MultiTask-CenterNet (MCN): Efficient and Diverse Multitask Learning using an Anchor Free Approach, ERCVAD Workshop at ICCV, 2021 (under review)., 2021.
- [26] X. Zhou, D. Wang und P. Krähenbühl, „Objects as Points,“ <http://arxiv.org/abs/1904.07850>, 2019.
- [27] K. He, X. Chen, S. Xie, Y. Li, P. Dollár und R. Girshick, „Masked Autoencoders Are Scalable Vision Learners,“ in *preprint arXiv:2111.06377*, 2021.
- [28] V. D. Blondel, J.-L. Guillaume, R. Lambiotte und E. Lefebvre, „Fast Unfolding of Communities in Large Networks,“ in *Journal of Statistical Mechanics: Theory and Experiment*, 2008.

3.4 TP4: Automotive Tauglichkeit

3.4.1 Ausgangssituation und Zielsetzung

Zu Projektbeginn wurden zwei konkrete Bereiche identifiziert, in denen aktuelle KI-Verfahren Schwächen aufweisen, die aber kritisch für den Erfolg hochautomatisierter Fahrzeuge sind:

Robustheit in Open World: Da die Umwelt, der ein hochautomatisiertes Fahrzeug ausgesetzt sein wird, hochkomplex und vielfältig ist, ergibt sich eine Anzahl möglicher Situationen im Straßenverkehr, die nicht durch einen endlichen Datensatz repräsentiert werden können (Open-World-Problem). Es wird in der realen Welt immer Situationen geben, die in den Trainingsdaten nicht antizipiert wurden. Die Perzeption eines automatisierten Fahrzeugs muss aber mit allen möglichen Situationen zurechtkommen, da sonst die Sicherheit der Verkehrsteilnehmer beeinträchtigt wäre. Insbesondere muss die KI-basierte Perzeption robust gegenüber Störungen, wechselnden Umweltbedingungen sowie unbekanntem Objekten sein.

Es bedurfte daher Methoden, welche die Robustheit eines KI-Systems quantifizierbar machten. Diese sollten dann genutzt werden, um Ansätze zur Steigerung der Robustheit, etwa durch die Kombination von unterschiedlichen Sensordaten oder die aktive Erkennung von unbekanntem Objekten oder Situationen zu entwickeln.

Herausforderungen Embedded Systems: Um entsprechend schnell auf Änderungen in der Umgebung reagieren zu können, muss die Perzeption des Fahrzeugs in Echtzeit anwendbar sein. Aus kommerziellen Gründen unterliegt die im Fahrzeug eingesetzte Hardware aber strengeren Grenzen in Bezug auf Rechenleistung und Speicher, als es für andere Anwendungsgebiete von KI-Technologien der Fall ist. Aktuelle KI-Verfahren setzen aber recht hohe Anforderungen an die benötigte Hardware.

Daher sollen für den automobilen Kontext relevante Hardwareplattformen in Kombination mit auszuführenden KI-Modellen im Hinblick auf gewisse Randbedingungen wie bspw. der Echtzeitfähigkeit und dem Speicherverbrauch untersucht werden. Darauf aufbauend sollten Methoden zur Performanzoptimierung der KI-Modelle entwickelt werden, um die Integration in die Zielhardware zu ermöglichen.

3.4.2 Strukturierung

Um die oben aufgeführten Probleme lösen zu können, wurde für beide Probleme ein dediziertes Arbeitspaket definiert.

In AP4.1 sollte die Robustheit eines KI-Systems quantifizierbar gemacht werden, sodass anschließend Ansätze zur Steigerung der Robustheit entwickelt werden können. Dies sollten bspw. durch die Kombination von unterschiedlichen Sensordaten oder die aktive Erkennung von unbekanntem Objekten oder Situationen realisiert werden.

In AP4.2 sollten für den automobilen Kontext relevante Hardwareplattformen in Kombination mit auszuführenden KI-Modellen im Hinblick auf gewisse Randbedingungen wie bspw. der Echtzeitfähigkeit und dem Speicherverbrauch untersucht werden. Darauf aufbauend wurden Methoden zur Performanzoptimierung der KI-Modelle entwickelt, sodass die Integration in die Zielhardware möglich wird.

3.4.2.1 Robustheit in Open World

3.4.2.1.1 Zielsetzung

Im Gegensatz zu wissenschaftlichen Benchmarks, die auf begrenzten Datensätzen arbeiten, müssen automatisierte Fahrzeuge mit allen Situationen in der Open-World zurechtkommen. Dies gilt entlang der gesamten Wirkkette des autonomen Fahrzeugs, von der Perzeption über die Verhaltens-Vorhersage bis hin zur Bewegungs-Planung. In allen Stufen der Wirkkette werden häufig gelernte Modelle eingesetzt. Diese Modelle müssen somit robust gegenüber Störungen, zum Beispiel wechselnde Umweltbedingungen, unbekannte Objekten oder abnormale Verhaltensweisen, sein. Für die Optimierung eines des gelernten Modells ergibt sich so eine neue Anforderung: Die Systeme müssen nicht nur hohe Performanz aufweisen, sondern auch robust sein, z.B. müssen sie eigene Fehler erkennen, oder sehr gut generalisieren können, z.B. unempfindlich gegenüber wechselnden Wetterbedingungen sein.

Diese Anforderung lässt sich in zwei Ziele formulieren: Die Erkennung von neuen Situationen sowie von Prädiktions-Fehlern, und die Steigerung der Performance in diesen Situationen. Zu ersterem Ziel wurden Ansätze im Bereich der Unsicherheitsschätzung und Anomalie-Erkennung entwickelt. Das zweite Ziel wird insbesondere mit dem Erlernen von robusten 3D Szenen-Repräsentation adressiert und die Nutzung von Label-Hierarchien. Der folgende Abschnitt geht auf die wichtigsten Ziele sowie die erzielten Ergebnisse hinsichtlich der Robustheit in der Open-World ein.

3.4.2.1.2 Erzielte Ergebnisse

Die Analyse der Robustheit neuronaler Netze in sogenannten "Corner Cases" bedarf einer grundlegenden Auseinandersetzung mit der Ausprägung solcher Corner Cases und deren Einfluss auf die Prädiktions-Fähigkeit der neuronalen Netze. Hierzu wurde in diesem AP die Definitionen und die Eigenschaften von Corner Cases evaluiert. Im Gegensatz zu vorherigen Arbeiten wurden die Ursachen von „Corner Cases“ aus der Perspektive der Interpretationsprobleme der

neuronalen Netze betrachtet. Die Leistung und Robustheit des Netzes wurde in Bezug auf die Verfügbarkeit, der bei der Entwicklung und Validierung verwendeten Daten analysiert und systematisiert, vgl. Abbildung 51.

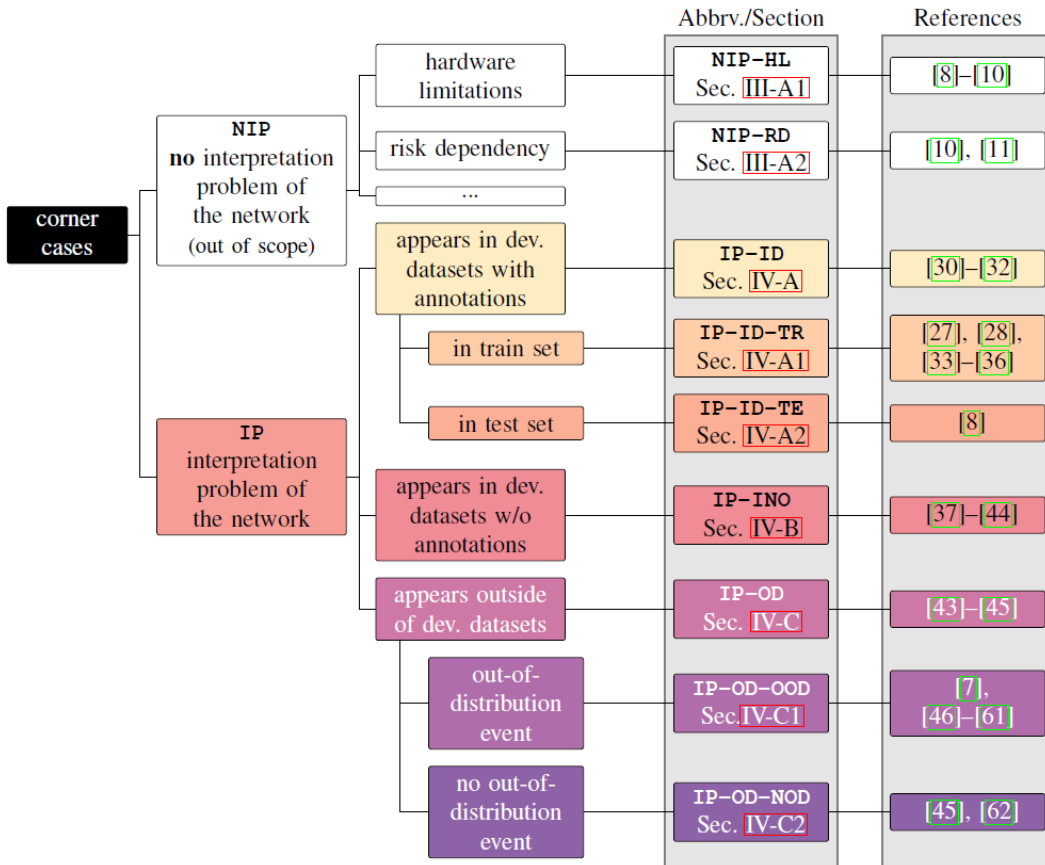


Abbildung 51: Taxonomie von Corner Cases aus der Perspektive von Interpretationsproblemen eines neuronalen Netzes

Für die ersten Implementierungen wurde eine vereinfachte Problemdarstellung betrachtet, nämlich Fahrzeuge mit Fahrradträgern zu erkennen. In den Experimenten wurde der Datensatz „Berkeley DeepDrive“ verwendet, um den Bildabschnitt eines Fahrzeugs (im Datensatz als "Car" bezeichnet) zu generieren. Während der Vorverarbeitung wird die Ground Truth der Objektlokalisierung vom Datensatz verwendet, um sicherzustellen, dass der für das Training verwendete Datensatz normalisiert werden könnte. Da der Berkeley-Datensatz nur sehr wenige Objekte mit Fahrradträgern enthält, wurden auch einige Bilder aus der Google-Bildersuche für das Training des neuronalen Netzes verwendet. Um sicherzustellen, dass das neuronale Netz auch die Fahrzeuge mit und ohne Fahrradträger von den normalen Fahrzeugen unterscheiden kann, wird ein zusätzlicher Testdatensatz mit Fahrradträgern verwendet. Mit Hilfe von Heatmaps werden ungewöhnliche Bereiche im Bild als Anomalien klassifiziert. In Ergebnissen wurde gezeigt, dass verschiedene Arten von Fahrzeugen mit Fahrradträgern korrekt als abnormale Objekte klassifiziert werden. Anschließend wurden mehrere Methoden untersucht, die in der Lage sind, zwischen den detektierten Objekten aus den bereits bekannten und den noch unbekannt Klassen zu unterscheiden. Dazu wurden keine Objekte der noch unbekannt Klassen während des Trainings der Modelle benötigt, was die hier vorgestellten Methoden von bisherigen Lösungsansätzen unterscheidet. Hier wurden mehrere Methoden zur hierarchischen Klassifikation betrachtet, welche die detektierten Objekte in eine Klassen-Hierarchie zuordnen

sollen. Zusätzlich dazu wurde ein neues Evaluierungsschema für die semantische Segmentierung vorgestellt, mit dem man Segmentierungsfehler innerhalb sowie außerhalb der Objektkategorie separat betrachten kann. Eine Reihe von Experimenten wurden durchgeführt, um das Verhalten verschiedener Netzwerkarchitekturen in unterschiedlichen Situationen, z.B. Umgebungsänderung oder das Auftauchen einer unbekannt Klasse, zu bewerten. Darüber hinaus wurden die Auswirkungen von verschiedener Klassentaxonomien für die Bewertung von Fehleranalysen außerhalb der Kategorie evaluiert.

Ein weiteres Mittel zur Analyse der Robustheit ist die Unsicherheitsschätzung. Ein Ziel war es eine punktweise Unsicherheitsschätzung für die semantische Segmentierung von Lidar-Punktwolken zu entwickeln, um darüber unterrepräsentierte oder sogar neue Klassen zu identifizieren. Überwachte Lernmethoden verlassen sich auf annotierte Daten und lernen die Repräsentation der bekannten Klassen. Durch eine Unsicherheitsschätzung erhält das Perzeptions-KNN die Möglichkeit, das Vorkommen von schlecht oder gar nicht gelernten Klassen zu signalisieren, sodass nachgelagerte Algorithmen entsprechend reagieren können. Mithilfe einer speziell entwickelten Metrik wurde die Kalibrierung verschiedener Modelle für semantische Segmentierung untersucht, besonders von unterrepräsentierten Klassen. Diese Metrik berücksichtigt Klassen-Imbalancen in den Daten und wertet zuverlässig die Kalibrierung eines Modells aus. Die Arbeit wurde auf einem Workshop auf der IROS 2022 publiziert. Um den Umgang mit den als problematisch identifizierten Klassen zu verbessern, wurde ein hierarchisches Multi-Label Modell entwickelt, das die strukturelle Verwandtschaft einiger Klassen ausnutzt, um Unsicherheit über die vorliegenden Daten auszudrücken. Hier wird dem Modell durch eine spezielle Lernfunktion implizit ermöglicht, auf Meta-Klassen zurückzugreifen. So wird sichergestellt, dass das Modell in schwer interpretierbaren Szenen gleichzeitig konfidente und zumindest teilweise korrekte Prädiktionen ausgibt, anstelle von unsicheren oder falschen Vorhersagen. Beispielsweise kann die Klassifikation für einen Fahrradfahrer auch "Mensch" sein, so wird ein Teil der Information in den Daten erhalten und geht nicht durch ein hohes Unsicherheitssignal verloren. Angewandt, implementiert und getestet wurden die Methoden für die semantische Segmentierung von LiDAR Punktwolken.

Ein weiterer Aspekt der Robustheitsanalyse ist die sogenannte Novelty-Erkennung, im Folgenden Anomalie-Erkennung genannt. Die Erkennung von Anomalien trägt unmittelbar zur Steigerung der Robustheit der eingesetzten KNNs bei, da deren Prädiktion häufig in unbekannt Szenarien ungenau ist. Zur Untersuchung der Qualität der Novelty-Detektion sind insbesondere Datensätze notwendig, welche diese sogenannten 'Known Unknowns' beinhalten.

Neben Anomalien in der Erscheinung von Objekten, können insbesondere abnormale Fahrverhalten katastrophale Folgen haben, z.B. Geisterfahrer oder aggressive Fahrer. Diese Bewegungs-Anomalien müssen unbedingt erkannt werden, um Unfälle zu vermeiden. Hierzu wurde ein neuartiger Ansatz zur Anomalie-Erkennung konzipiert. Da Fahranomalien selten vorkommen und in bestehenden Datensätzen nicht oder nicht in ausreichender Form annotiert sind, wird die Aufgaben der Anomalie-Erkennung als unüberwachte Lernaufgabe formuliert, die während des Trainings ohne Labels auskommt. Die entwickelte Methode nutzt die Fähigkeit von sogenannten Auto-Encodern eine Repräsentation der zugrundeliegenden Daten in einem latenten Raum des kNN zu lernen. Dieser Auto-Encoder nutzt als Eingangssignal die Szenen-Information. Diese besteht aus der Karteninformation und der vergangenen Trajektorien aller Verkehrsteilnehmer. Als Encoder wurde hierzu ein sogenannten Graph-basiertes neuronales Netz verwendet, welches neben den individuellen Bewegungs-Trajektorien auch die Interaktion

zwischen den Agenten einbezieht. In einem zweiten Schritt, wird eine Verteilung über alle latenten Merkmalsvektoren im Trainingsdatensatz geschätzt. Während der Inferenz können abnormale Verhaltensmuster in Regionen des latenten Raums mit geringer Wahrscheinlichkeit gefunden werden. Die Methode ist in Abbildung 52 dargestellt. Zur Untersuchung der Methode, wurden zwei Datensätze erstellt, ein interaktiver Highway-Datensatz mit zwei Agenten sowie ein Datensatz mit urbanen und ländlichen Szenarien. Dabei wurde für die Erzeugung der abnormalen Szenarien auf Simulation gesetzt. Beide Datensätze beinhalten annotierte, abnormale Fahrscenarien. Die Datensätze, die beschriebene Methode und weitere Baseline-Methoden wurden in den Publikationen "Anomaly Detection in Multi-Agent Trajectories for Automated Driving" (CoRL 2021) und "A Benchmark for Unsupervised Anomaly Detection in Multi-Agent Trajectories" (ITSC 2022) veröffentlicht. Qualitative Ergebnisse sind in Abbildung 53 dargestellt. Zur Evaluation wurden vorhandene Metriken zur Anomalie-Erkennung aus der Literatur übernommen und erweitert.

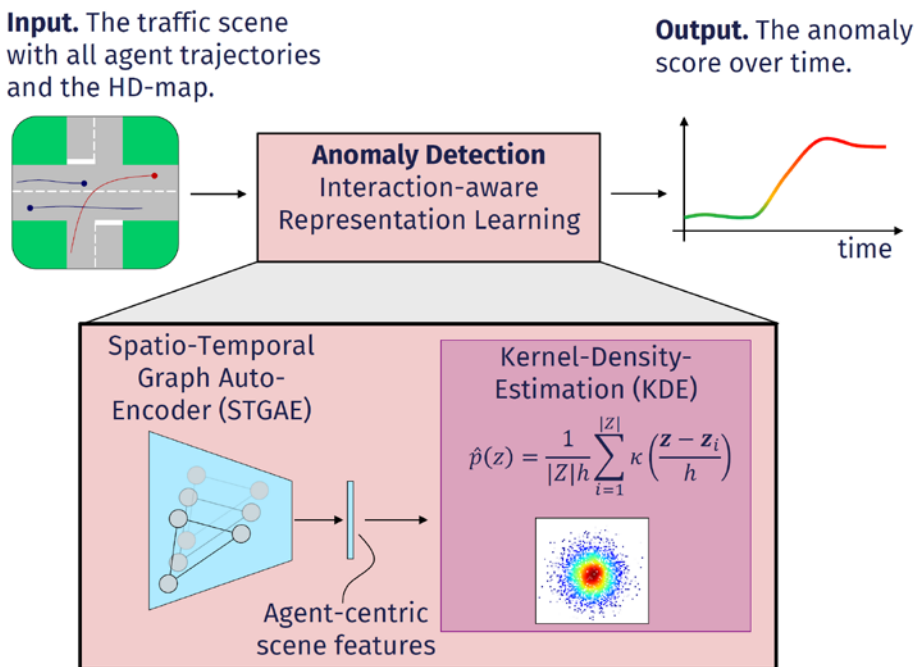


Abbildung 52: Inferenz durch den Graph-basierten Auto-Encoder und die Kernel Density Estimation zur Schätzung einer Wahrscheinlichkeitsdichte über die Merkmalsvektoren. Das Modell gibt den Anomalie-Score aus.

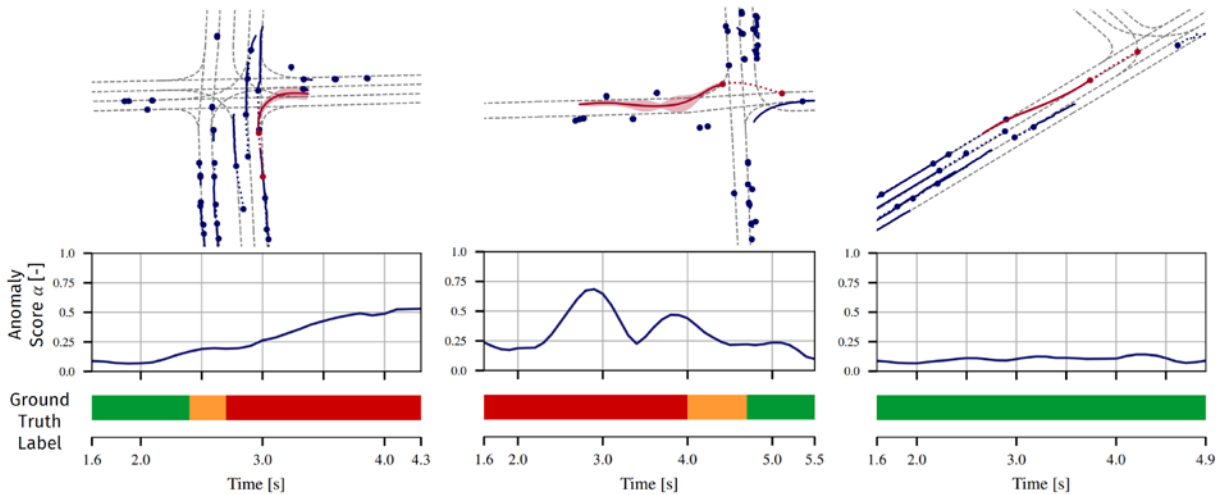


Abbildung 53: Ergebnisse der Anomalie-Erkennung : (links) Abnormales Abbiegeverhalten. Der Agent biegt in die Spur des ankommenden Verkehrs ein. (mitte) Der Agent bricht den Abbiegevorgang abrupt ab. (rechts) Normales Verhalten. Der Agent wechselt die Spur und der Anomalie-Score bleibt niedrig.

Ein wichtiger Beitrag zur Steigerung der Robustheit von kNNs sind robuste Repräsentation der erfassten Umwelt. Hierzu dienen besonders 3D Szenenrepräsentationen. Das Ziel war es einen lernbasierten Bottom-Up-Ansatz zur Verbesserung der Generalisierungseigenschaften zu entwickeln, indem eine 3D Szenenrepräsentation inkl. von 3D-Bewegung aus den Rohdaten berechnet wird.

Insbesondere sollte das Verfahren DeepTAM analysiert und weiterentwickelt werden. Im Rahmen dieses Projektes wurde DeepTAM insbesondere bezüglich Robustheit gegenüber unvorhergesehenen Situationen analysiert, da dies im automobilen Kontext entscheidend ist. Dabei wurde festgestellt, dass diese nur unzureichend gegeben ist (genauso wie auch bei anderen lernbasierten Tiefenschätzverfahren). Aufbauend darauf, wurde ein robustes Tiefenschätzverfahren mit dem Namen "Robust MVD Baseline Model" entwickelt (siehe Abbildung 54). Die Robustheit wurde hierbei vorallem durch die Wahl der Trainingsdaten und eine neu entwickelte Datenaugmentierung erzielt. Es wurde gezeigt, dass dieses Verfahren robuster ist als bisherige Verfahren inklusive DeepTAM (siehe Tabelle 1). Die Analyse und das entwickelte Verfahren wurden in der Publikation "A Benchmark and a Baseline for Robust Multi-view Depth Estimation" bei der 3DV 2022 Konferenz veröffentlicht.

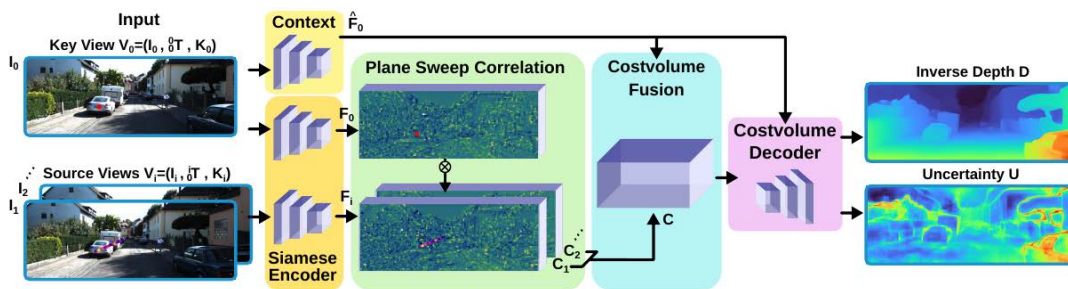


Abbildung 54: Architektur des entwickelten "Robust MVD Baseline Model" Verfahrens . Das Verfahren bekommt mehrere Bilder mit bekannter Kamerakalibrierung als Eingabe und berechnet die (inverse) Tiefenkarte für eines der Eingabebilder und die zugehörige Unsicherheit.

Approach	GT	GT	Align	KITTI		ScanNet		ETH3D		DTU		T&T		Average		
	Poses	Range		rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	time [s] ↓
a)																
COLMAP [16, 15]	✓	✗	✗	12.0	58.2	14.6	34.2	16.4	55.1	0.7	96.5	2.7	95.0	9.3	67.8	≈ 3 min
COLMAP Dense [16, 15]	✓	✗	✗	26.9	52.7	38.0	22.5	89.8	23.2	20.8	69.3	25.7	76.4	40.2	48.8	≈ 3 min
b)																
DeMoN [22]	✗	✗	t	15.5	15.2	12.0	21.0	17.4	15.4	21.8	16.6	13.0	23.2	16.0	18.3	0.08
DeepV2D KITTI [19]	✗	✗	med	(3.1)	(74.9)	23.7	11.1	27.1	10.1	24.8	8.1	34.1	9.1	22.6	22.7	2.07
DeepV2D ScanNet [19]	✗	✗	med	10.0	36.2	(4.4)	(54.8)	11.8	29.3	7.7	33.0	8.9	46.4	8.6	39.9	3.57
c)																
MVSNet [27]	✓	✓	✗	22.7	36.1	24.6	20.4	35.4	31.4	(1.8)	(86.0)	8.3	73.0	18.6	49.4	0.07
MVSNet Inv. Depth [27]	✓	✓	✗	18.6	30.7	22.7	20.9	21.6	35.6	(1.8)	(86.7)	6.5	74.6	14.2	49.7	0.32
CVP-MVSNet [25]	✓	✓	✗	156.7	2.2	137.1	15.9	156.4	13.6	(4.0)	(68.4)	24.7	52.9	95.8	30.6	0.49
Vis-MVSNet [30]	✓	✓	✗	9.5	55.4	8.9	33.5	10.8	43.3	(1.8)	(87.4)	4.1	87.2	7.0	61.4	0.70
PatchmatchNet [23]	✓	✓	✗	10.8	45.8	8.5	35.3	19.1	34.8	(2.1)	(82.8)	4.8	82.9	9.1	56.3	0.28
Fast-MVSNet [29]	✓	✓	✗	14.4	37.1	17.0	24.6	25.2	32.0	(2.5)	(81.8)	8.3	68.6	13.5	48.8	0.30
MVS2D ScanNet [26]	✓	✓	✗	21.2	8.7	(27.2)	(5.3)	27.4	4.8	17.2	9.8	29.2	4.4	24.4	6.6	0.04
MVS2D DTU [26]	✓	✓	✗	226.6	0.7	32.3	11.1	99.0	11.6	(3.6)	(64.2)	25.8	28.0	77.5	23.1	0.05
d)																
DeMoN [22]	✓	✗	✗	16.7	13.4	75.0	0.0	19.0	16.2	23.7	11.5	17.6	18.3	30.4	11.9	0.08
DeepTAM [31]	✓	✗	✗	68.7	0.4	(6.7)	(39.7)	20.4	19.8	58.0	9.1	40.0	12.9	38.8	16.4	0.85
DeepV2D KITTI [19]	✓	✗	✗	(20.4)	(16.3)	25.8	8.1	30.1	9.4	24.6	8.2	38.5	9.6	27.9	10.3	1.43
DeepV2D ScanNet [19]	✓	✗	✗	61.9	5.2	(3.8)	(60.2)	18.7	28.7	9.2	27.4	33.5	38.0	25.4	31.9	2.15
MVSNet [27]	✓	✗	✗	14.0	35.8	1568.0	5.7	507.7	8.3	(4429.1)	(0.1)	118.2	50.7	1327.4	20.1	0.15
MVSNet Inv. Depth [27]	✓	✗	✗	29.6	8.1	65.2	28.5	60.3	5.8	(28.7)	(48.9)	51.4	14.6	47.0	21.2	0.28
CVP-MVSNet [25]	✓	✗	✗	158.2	1.2	2289.0	0.1	1735.3	1.2	(8314.0)	(0.0)	415.9	9.5	2582.5	2.4	0.50
Vis-MVSNet [30]	✓	✗	✗	10.3	54.4	84.9	15.6	51.5	17.4	(374.2)	(1.7)	21.1	65.6	108.4	31.0	0.82
PatchmatchNet [23]	✓	✗	✗	29.0	16.3	70.1	16.7	99.4	3.5	(82.6)	(5.6)	39.4	19.3	64.1	12.3	0.18
Fast-MVSNet [29]	✓	✗	✗	12.1	37.4	287.1	9.4	131.2	9.6	(540.4)	(1.9)	33.9	47.2	200.9	21.1	0.35
MVS2D ScanNet [26]	✓	✗	✗	73.4	0.0	(4.5)	(54.1)	30.7	14.4	5.0	57.9	56.4	11.1	34.0	27.5	0.05
MVS2D DTU [26]	✓	✗	✗	93.3	0.0	51.5	1.6	78.0	0.0	(1.6)	(92.3)	87.5	0.0	62.4	18.8	0.06
Robust MVD Baseline	✓	✗	✗	7.1	41.9	7.4	38.4	9.0	42.6	2.7	82.0	5.0	75.1	6.3	56.0	0.06

Tabelle 4: Ergebnisse der Analyse bzgl. Robustheit von aktuellen Tiefenschätzungsverfahren. Evaluiert wurden auf den Datensätzen KITTI, ScanNet, ETH3D, DTU und Tanks And Temples. Als Metriken wurden der relative Fehler (rel) und der Anteil an korrekt geschätzten Pixeln (τ) verwendet. Ein Verfahren kann als robust gesehen werden, wenn es gut über Datensätze hinweg funktioniert.

Darüber hinaus sollte die Kompatibilität der Stixel-Repräsentation mit lernbasierten Verfahren geprüft werden. Dies wurde umgesetzt indem eine Netzwerkarchitektur entwickelt wurde, die direkt von Eingabe-Bildern und Tiefenkarten zu einer Stixel-Repräsentation abbilden kann. Es musste allerdings festgestellt werden, dass dies nicht einfach ist, da die Anzahl der Stixel pro Bild-Spalte variabel ist, aber gleichzeitig die Spalte vollständig abdecken muss. Es war schwer diese Variabilität in einer Netzwerkarchitektur zu ermöglichen und zu optimieren. Es muss daher festgestellt werden, dass Stixel-Repräsentationen und lernbasierte Verfahren nicht trivial kompatibel sind. Mit heutigen Technologien, z.B. Transformer Architekturen, könnte sich dies eventuell anders verhalten.

Zuletzt sollten 3D Bewegungen von beobachteten Objekten berechnet werden. Dies wurde mit dem entwickelten Verfahren "SF2SE3" umgesetzt, welches in Abbildung 55 dargestellt ist. In diesem Verfahren, wird angenommen, dass Bilder und Tiefenkarten gegeben sind. Die Tiefenkarten können von zusätzlichen Sensoren oder von unserem "Robust MVD Baseline Model" geschätzt werden. Zunächst wird ein bestehendes Modell zur Schätzung des optischen Flusses eingesetzt. Aus den somit bekannten optischen Fluss und den Tiefenkarten, werden dann einzelne Objekte segmentiert und die 3D Bewegung berechnet, indem Punkte mit der selben Bewegung zusammengefasst und gefiltert werden. Das Ergebnis ist eine Segmentierung bewegter Objekte mit den zugehörigen 3D Bewegungen. Das Verfahren wurde in der Publikation "SF2SE3: Clustering Scene Flow into SE(3)-Motions via Proposal and Selection" bei der GCPR 2022 veröffentlicht.

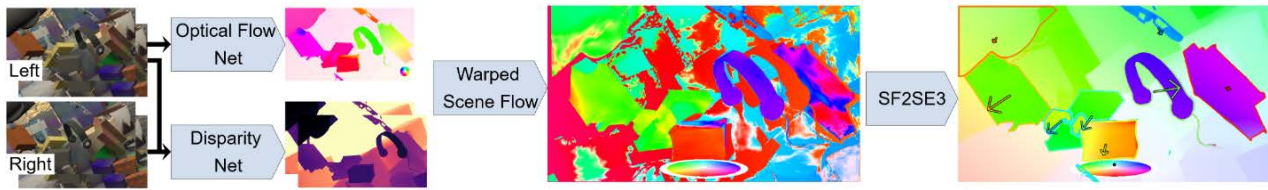


Abbildung 55: Übersicht des entwickelten Verfahrens "SF2SE3". Das Verfahren bekommt optischen Fluss und Tiefenkarten (bzw. Disparität) als Eingabe und berechnet daraus eine Segmentierung bewegter Objekte (im Bild rechts markiert mit farbigen Konturen) mit zugehörigen 3D Bewegungen (im Bild rechts markiert mit Pfeilen).

In einem weiteren Projekt wurden sogenannte hierarchische Label-Strukturen genutzt, um die Robustheit der kNNs zu verbessern. Hierfür wurde konkret der Anwendungsfall der Objekterkennung herangezogen, wobei ein dominierender Faktor der Robustheit die Verwechslungsgefahr zweier Klassen als auch insbesondere die Resilienz gegenüber unbekanntem Objekten ist. Ein Modell, welches nicht auf das letztere Problem konditioniert ist, würde ein unbekanntes Objekt schlicht mit einer der zur Trainingszeit bekannten Klassen verwechseln. Gleichzeitig kann für eine Fahrfunktion die Verwechslung zweier Klassen unterschiedlich kritisch sein (Verwechslung von Mensch und Auto ist kritischer als Auto und Bus). Zur Quantifikation der Robustheit wurde daher eine Misklassifikationsmatrix vorgeschlagen, deren Einträge die Schwere paarweiser Verwechslungen bewertet.

Um anschließend die Robustheit insbesondere gegenüber unbekanntem Objekten zu steigern, wurde ein Hierarchical Novelty Detection (HND) Verfahren auf der Basis von Capsule Networks entwickelt. Hierbei finden Klassifikationen von Objekten entlang einer Klassenhierarchie statt, sodass unbekannte Objekte durch das Modell in eine abstraktere Klasse in der Hierarchie eingeordnet werden sollen. Durch die Verwendung von Capsule Networks, welche eine intrinsische hierarchische Struktur aufweisen, wurde eine verbesserte Güte dieses Ansatzes gegenüber klassischer HND-Verfahren erwartet. Eine exemplarische Architektur ist in Abbildung 56 zu sehen. Zur Evaluation wurde das entwickelte Verfahren auf dem FashionMNIST-Datensatz angewendet, wobei eigens eine kleine Klassenhierarchie aufgebaut wurde. Als Baseline wurden klassische Top-Down oder Big-Bang Verfahren eingesetzt, wobei jedes Verfahren über eine für die Problemstellung definierte Misklassifikationsmatrix bewertet wurde. Die Ergebnisse zeigen, dass das entwickelte Verfahren robuster gegenüber unbekanntem Objekten im Vergleich zu den Baseline-Modellen ist. Eine umfassende Beschreibung der Methodik und der Ergebnisse wurden auf der IV22 veröffentlicht: <https://ieeexplore.ieee.org/document/9827249>.

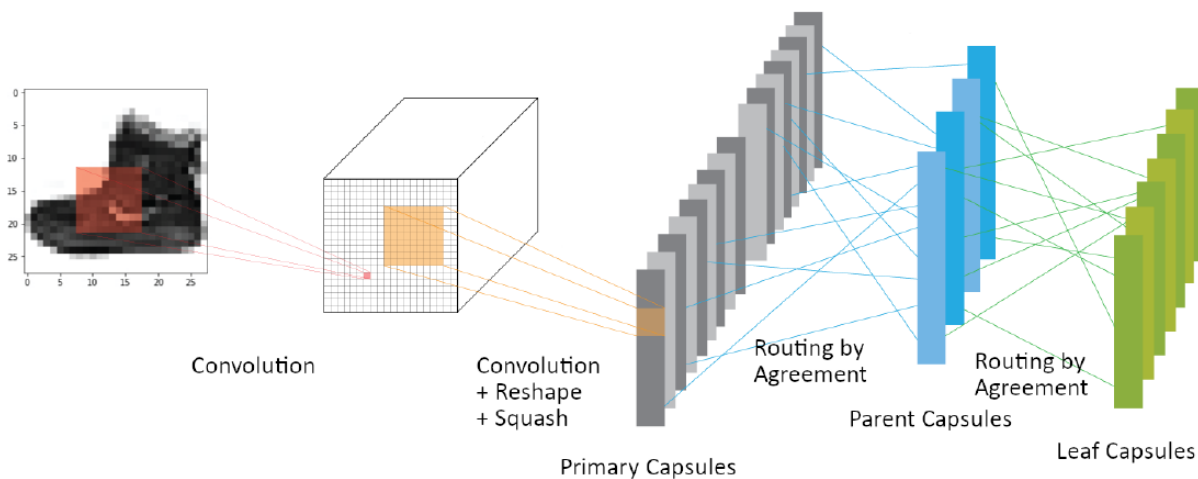


Abbildung 56: Beispielhafte Architektur eines Capsule Networks für Hierarchical Novelty Detection

3.4.2.1.3 Gegenüberstellung der vorgegebenen Ziele

Die in der VHB definierten Ziele wurden erreicht. Unter anderem wurde im Bereich der Schätzung von 3D Tiefeninformation durch die gezielte Auswahl an Trainingsdaten sowie durch Daten-Augmentierung die Robustheit gegenüber neuen Szenen gesteigert, was vor allem für die Anwendung im Bereich des autonomen Fahrens notwendig ist, da es in der Open World immer zu neuen Situationen kommen kann. Zudem wurde aus der Kombination eines graph-basierten neuronalen Netzes und einer Kernel Dichte-Schätzung eine Methode entwickelt, welche Anomalien in Fahrsituationen erkennt. Damit ist es möglich auf abnormale Fahrmanöver zu reagieren und somit selbst in kritischen Fahrsituationen sicher zu agieren.

3.4.2.2 AP4.2 Herausforderungen Embedded Systems

3.4.2.2.1 Zielsetzung

Das Ziel von AP4.2 war es, Methoden zu erstellen und Informationen zusammenzutragen, um die Übersetzung von KI Systemen vom Laborrechner auf die automobilen eingebetteten Hardwareplattform zu erleichtern. Diese geschieht im Wesentlichen in drei Phasen:

1. die Auswahl der geeigneten Hardwareplattform und/oder der Bestimmung der Vorgaben (Zahl der Rechenoperationen, Speichermenge, Höchsttemperatur, etc.) die aus der Wahl der Hardwareplattform folgen
2. der Anpassung der Topologie der KI (Metaparameter des neuronalen Netzwerks) an diese Anforderungen
3. die automatische Synthese der KI für die Zielplattform

Für alle drei Phasen sollten Methoden zur Verbesserung/Erleichterung der Durchführung erstellt werden. Für Phase 1 beinhaltet dies die Erstellung einer Methode um die Wahl der geeigneten Hardware zu erleichtern sowie eine Übersicht der wichtigsten Vorgaben hinsichtlich relevanter Betriebsparameter, für Phase 2 sollten Methoden erarbeitet werden, die es erlauben, zielgerichtet Anpassungen an der Netzwerktopologie der KI vorzunehmen so dass diese sich besser für die Umsetzung in eingebettete Hardware eignet. Dies sollte insbesondere durch eine Vorhersagemethode der zu erwartenden KI-KPIs erzielt werden. Für Phase 3 sollten verschiedene

Methoden entwickelt werden, die sich in den Syntheseflow integrieren lassen und so die Zielparameter Ausführungszeit, Speicherbedarf und Energieumsatz automatisch und hardwarebewusst optimieren.

3.4.2.2.2 Erzielte Ergebnisse und Gegenüberstellung der vorgegebenen Ziele

3.4.2.2.2.1 KI Hardware Marktübersicht

Entsprechend der Zielsetzung wurde eine Aufstellung der Hardwareanforderungen automobiler, sicherheitskritischer KI und eine Übersicht geeigneter Zielhardwareplattformen erstellt. Dafür wurde ein Fragebogen abgestimmt und erstellt, um die individuellen industriellen Anforderungen der Projektpartner und externer Teilnehmer zu bestimmen. Die Ergebnisse wurden eingesammelt, ausgewertet und zu einer grundlegenden, tabellarischen Übersicht zusammengestellt. Fehlende Einträge in den Übersichten wurden dann anhand von Datenblättern für jeden einzelnen Eintrag ergänzt.

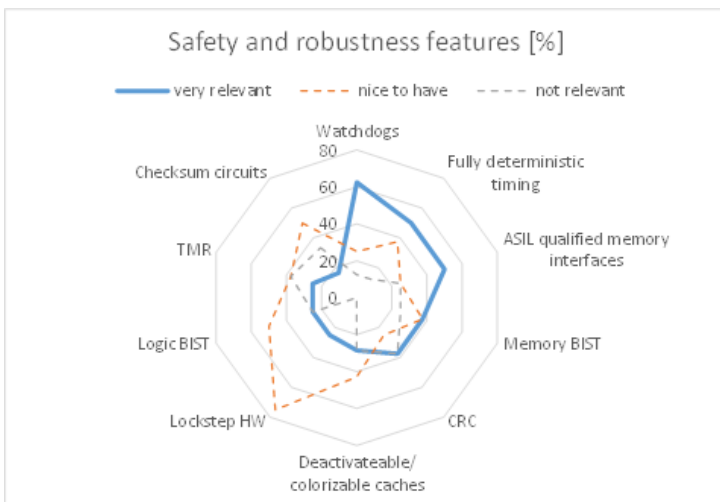


Abbildung 57: Auswertung der Hardwareanforderungen - hier Safety und Robustheitseigenschaften.

Datapath performance				Memory performance			Device features					
OP/s	Cores	Chipset	HW Features	Datatypes	mem config.	mem bandwidth	Design	Power	I/O connections	I/O bandwidth		
Reliability features				Certificates			Robustness and safety features			Frameworks supported		
Thermal	Humidity	IP prol. class	FIT	ASIL	AEC-Q100	Other	Watchdog	Lockstep	Other	(adaptive) AUTOSAR	AI	Frameworks supported

Abbildung 58: Kategorien der Übersichtstabelle geeigneter Zielhardwareplattformen.

In Abbildung 57 ist beispielhaft eines der Auswertungsergebnisse aufgetragen. Basierend auf der Aussage, dass über 60% der Befragten die Watchdog Technologie als sehr relevant angaben und fast 80% die Lockstep Technik verwenden würden wenn vorhanden, wurden diese beiden Features in der Hardware Übersicht (siehe Abbildung 58) gesondert aufgeführt und recherchiert. Andere Techniken wie etwa Prüfsummenschalungen wurden bei „Other“ erwähnt, aber nicht durchgehend für jede Hardware recherchiert. Die vollständige Auswertung und die über 40 Einträge der Hardwareübersicht wurden in E1.2.4.2a berichtet und liegen mittlerweile in Version 3 vor.

3.4.2.2.2 Hardwareabhängige Vorhersage der Laufzeit von KI Systemen

Um es einem KI-Entwickler, der möglicherweise keine genauen Hardwarekenntnisse besitzt dennoch zu ermöglichen, seine neuronalen Netze so zu wählen oder zu modifizieren, dass sie sich später auf der gewählten Hardwareplattform innerhalb der Vorgaben läuft, wurde eine Methode entwickelt, welche die zu erwartenden Eigenschaften wie etwa die Ausführungszeit schnell, leicht und für einen KI Entwickler leicht nutzbar zur Verfügung stellt. Konkret wurde dafür zunächst eine Hardwareanalysemethode entwickelt, welche zwar über Stunden oder ggf. Tage die Eigenschaften einer neuen Hardwareplattform vermisst, dies aber vollständig autonom tut, sodass hier keine hohen Arbeitsaufwände entstehen. Diese Untersuchung muss für jede neuartige Hardwareplattform auch nur einmal gemacht werden und ist dann für alle möglichen Anwendungen wiederverwendbar.



Abbildung 59: Schematische Übersicht der automatischen Hardware-Charakterisierung als Vorbereitung für eine hardwarebewusste Vorhersage der Ausführungszeiten von neuronalen Netzen.

Dabei hat es sich als unverzichtbar herausgestellt (siehe Abbildung 59), nicht nur die Hardware selbst, sondern auch den genutzten Syntheseflow mit zu charakterisieren, da dieser maßgeblichen Einfluss auf die Ressourcennutzung und somit Effizienz der Hardware hat.

Liegen die Charakterisierungsdaten einer Syntheseflow-Hardware Kombination vor kann der KI Entwickler das von uns entwickelte Vorhersagemodell direkt in seiner Entwicklungsumgebung (z.B. Tensorflow) nutzen. Wichtig war es hier, dass das Modell nur auf Eingangsgrößen angewiesen ist, die bereits vor dem Training der KI vorhanden sind (z.B. alle Metaparameter). Auf diese Weise kann das Training von neuronalen Netzen, welche am Ende niemals auf der Hardware innerhalb der Vorgaben umsetzbar wären vermieden und so viel Entwicklungszeit gespart werden.

```
t=24.8±0.533ms          1 HW layers
layer type: Convolution
exec type: MyriadXhwOp
layer name: conv5_block16_2_conv_3/Conv2D
```

Abbildung 60: Vorhersage der zu erwartenden Hardware-Ausführungszeiten . Hier aufgeschlüsselt pro Layer.

Wie in Abbildung 60 gezeigt, kann der KI Entwickler eine zusätzliche Python Bibliothek integrieren, und mit einem Satz Charakterisierungsdaten initiieren. Daraufhin können Methoden genutzt werden, um pro Layer eines neuronalen Netzes die zu erwartenden Ausführungszeiten und den Speicherbedarf dargestellt zu bekommen. Daraufhin kann entweder ein gänzlich neues neuronales Netz gewählt, oder das vorhandene manuell pro Sicht angepasst werden. Ebenso kann die Vorhersage Teil der Zielfunktion einer automatischen Architektursuche (NAS) sein.

Diese Methoden wurden sowohl auf einem dedizierten KI-Beschleuniger (einem Intel Neural Compute Stick 2) als auch auf einem Graphikbeschleuniger (der Intel UHD 630 integrierten Graphikeinheit) getestet.

Abbildung 61 zeigt das Ergebnis auf dem dedizierten KI-Beschleuniger. Hier kann man den gemittelten absoluten Fehler (in blau) in Prozent für alle gemodelten Schichten sehen. Die Gesamtabweichung (nicht absolut) von der eigentlichen Ausführungszeit ist in rot gekennzeichnet. Mit einigen Ausnahmen bleibt die Vorhersage beim KI-Beschleuniger unter 20%.

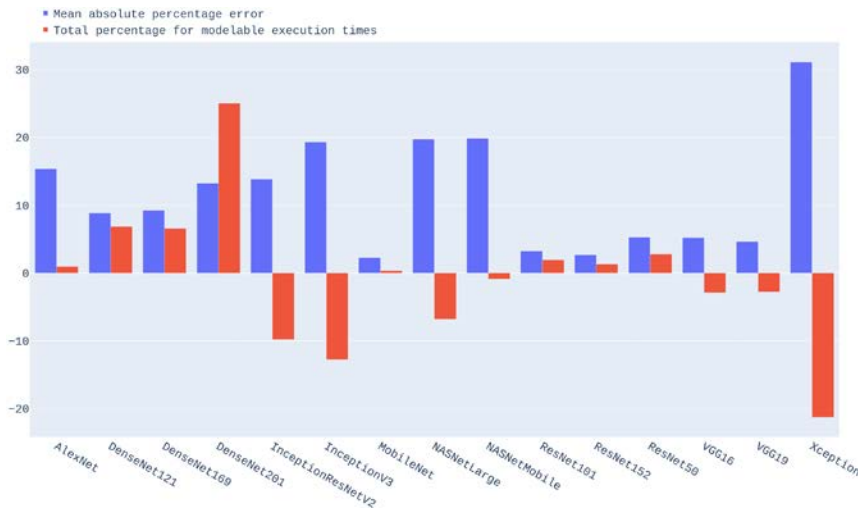


Abbildung 61: Ergebnis der Benchmarks ausgeführt auf dem Neural Compute Stick 2

Abbildung 62 zeigt die gleichen Fehler an, für fast die gleichen Netzwerke. Drei Netzwerke (NASNetLarge, NASNetMobile und ResNet50) konnten für diese Hardware nicht synthetisiert werden.

Die Fehler sind hier deutlich höher (meist unter 50%) als beim KI-Beschleuniger, da es alleine bei den gemessenen Ausführungszeiten deutlich höhere Varianzen gab, als beim KI-Beschleuniger.

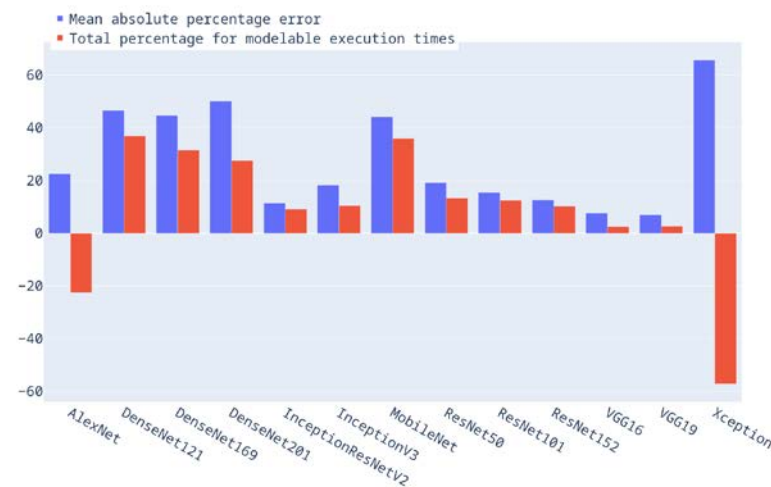


Abbildung 62: Ergebnis der Benchmarks ausgeführt auf der integrierten Intel UHD 630 GPU

3.4.2.2.3 HIL Integration eines KI Systems und Entwicklung eines heuristischen Pruningverfahrens

Es sollten im Projekt Algorithmen entwickelt, implementiert und evaluiert werden um das grobgranulare (Filter) Pruning optimal für eine vorgegebene Hardware zu konfigurieren. Wir konzentrierten uns daher auf den Einsatz eines hochmodernen, auf semantischer Segmentierung basierenden CNN, nämlich des DeepLabV3+ Modells, auf rekonfigurierbarer Hardware (FPGA). Die Hauptmotivation war die Entwicklung eines effizienten Beschleunigers unter Verwendung von High-Level-Synthese (HLS), um die hohen Rechen- und Speicheranforderungen eines so komplexen Modells wie DeepLabV3+ zu erfüllen. Zur weiteren Erleichterung der effizienten Ausführung auf der HW implementierten wir einen HW-bewussten genetischen Algorithmus (GA) und sucht nach Pareto-optimalen beschnittenen CNNs. Um HW-freundliche CNN-Architekturen zu entwerfen, wurde ein HW-Modell des entwickelten semantischen Segmentierungsbeschleunigers in die Pruning-Suche integriert. Daraufhin konnte die Pruning-Konfigurationen dadurch weiter verfeinert werden, dass HW-Heuristiken abgeleitet wurden, welche den Suchalgorithmus bei der Komprimierung des Netzwerks leiten.

Zur Erstellung der Genome potenzieller Lösungen wird eine explizite, bijektive Kodierung verwendet. Ein einzelnes Genom repräsentiert eine mögliche CNN-Beschneidungsstrategie und hat so viele genetische Loci wie Schichten im CNN. Jeder genetische Locus kapselt eine Beschneidungsrate auf der entsprechenden Schicht. Die Kodierung von Sparsamkeit zu Schichten kann intuitiv in sequenziellen Genomen erfasst werden, die eine Eins-zu-Eins-Anwendung von GA-Operatoren auf die CNN-Schichtenfolge ermöglichen. Benachbarte CNN-Schichten weisen eine höhere Merkmalskorrelation auf als entfernte Schichten. Daher können beschnittene Schichtbeziehungen, die in benachbarten genetischen Loci kodiert sind, in einer Population überleben und durch Ein-Punkt-Kreuzung wiederverwendet werden, um effizientere Nachkommen zu erzeugen. Je fitter die Eltern im Laufe der Generationen werden, desto bessere genetische Loci werden sie haben, was zu besseren Nachkommen führen kann. Die Mutation ermöglicht es den Nachkommen außerdem, den lokalen Minima ihrer Eltern zu entkommen.

Die GA-Suche beginnt zunächst mit einer Populationsgröße P . Jedes Individuum in der Population P ist ein CNN mit zufällig ausgewählten Beschneidungsraten r seiner Schichten. Diese Netze werden bewertet, um ihre mIOU- und HW-Schätzungen zu erhalten. Die mIOU-Werte während der Suche werden durch Bewertung des beschnittenen Modells anhand von Zufallsstichproben aus der Trainingsmenge ermittelt, um eine unvoreingenommene Population in Bezug auf die Validierungsmenge zu erhalten. Das HW-bezogene Optimierungskriterium kann entweder die Anzahl der Operationen im CNN oder die Latenzwerte sein. Die Individuen in der Population durchlaufen dann Crossover und Mutation, um Nachkommen zu erzeugen. Wir bewerten die mIOU- und HW-Schätzwerte der einzelnen Nachkommen. Auf der Grundlage ihrer Pareto-Optimalität überleben die ausgewählten Eltern und Nachkommen in der nächsten Generation. Dieser Prozess wird für g Generationen fortgesetzt.

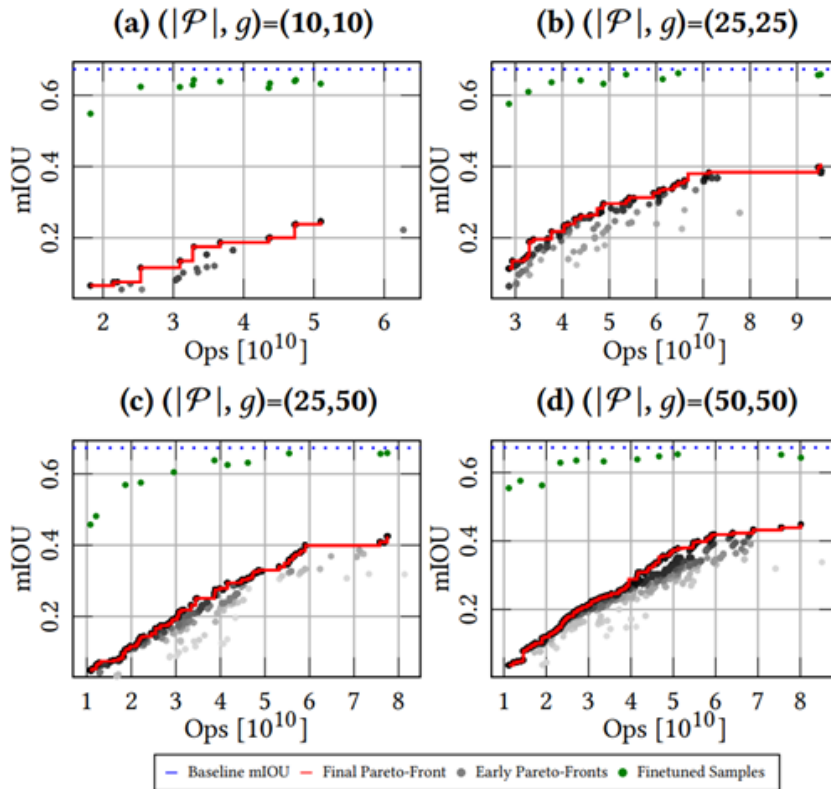


Abbildung 63: Hyperparameter Exploration für ein GA basiertes Channel Pruning auf DeepLabV3+

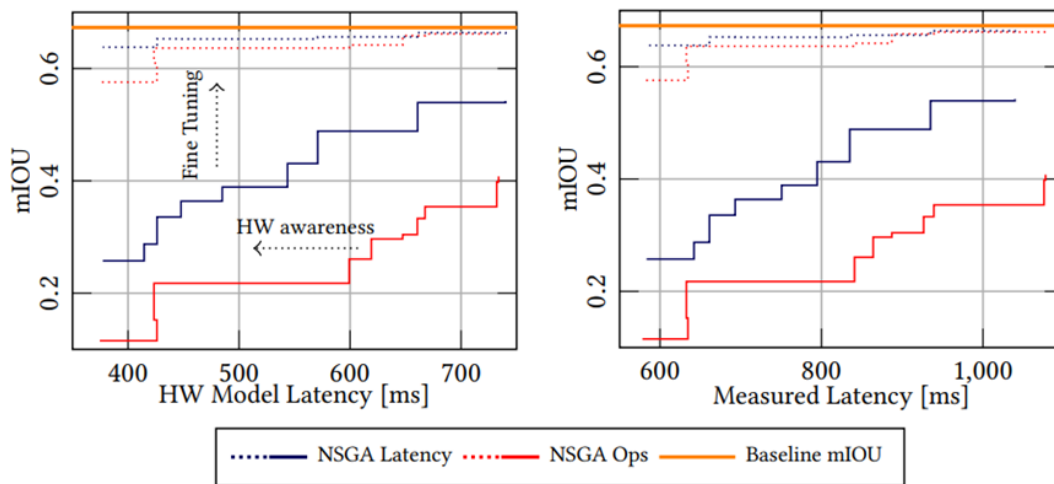


Abbildung 64: HW-bewusste Kompression von DeepLabV3+ mithilfe von GA basiertem Channel Pruning

Die Methode wurde erfolgreich erforscht und vollständig entwickelt. Wie in Abbildung 63 Abbildung 64 zu sehen ist, bietet sie je nach Heuristik-Parametern eine große Auswahl Pareto-optimaler Lösungen, die zwischen der Genauigkeit (gemessen in mIOU) und der Anzahl von Rechenoperationen abwägen hilft. Zusätzlich zeigt, wie wichtig der Fokus eben nicht nur auf die abstrakten Metriken wie die Zahl der Operationen ist. Wird stattdessen die wirkliche Ausführungszeit als Optimierungskriterium gewählt, zeigt sich beim Pruning eine etwa doppelt so hohe Genauigkeit bei gleicher Latenz. Nach dem Finetuning ist dieser Effekt deutlich geringer, aber dennoch noch vorhanden.

Wir haben auch überprüft, ob das auf genetischen Algorithmen basierende Pruning auf der Nvidia 1080 TI-GPU funktioniert. Wir stellen die Ergebnisse für ops-basierte und HW-fähige Optimierungsbeschränkungen bereit. Wir beobachten eine Reduzierung der Latenz um das 1,37-fache, wenn sie direkt für die Latenz optimiert wird (dritte Reihe), wobei mIoU nur minimal beeinträchtigt wird.

Model	Compute Complexity (GFLOPs)	Latency per batch (ms)	mIoU (%)
Baseline	276	114	71.07
Ops based	171	98	69.99
HW-aware	141	83	69.99

3.4.2.2.2.4 Entwicklung eines Pruningverfahrens, basierend auf Datenanalyse um das korrekte Verhalten in seltenen Fällen zu gewährleisten

Die Popularität von Deep Neural Networks und ihrer Anwendung auf eingebetteten Systemen und Edge-Geräten nimmt rasant zu. Die meisten eingebetteten Systeme sind in ihren Rechenkapazitäten und ihrem Speicher meist stark begrenzt. Um diese Einschränkungen gerecht zu werden, müssen die immer größer werdenden Modelle komprimiert werden, während ihre Genauigkeit beibehalten werden sollte. Dies ist durch verschiedene Methoden zum Beschneiden des Netzwerks möglich, auch Pruning genannt. Das Beschneiden der Modelle ist meist durch einen komplexen Algorithmus realisiert und es ist nur sehr schwierig nachvollziehbar, wie die Entscheidungen zum Entfernen verschiedener Filter / Neurone getroffen wird.

Zu Beginn des Projekts hat ZF einige grundlegende Methoden zum Pruning untersucht, ebenso wie bereits schnelle und optimierte Netzwerke. Zudem wurde zusammen mit dem Partner OFFIS eine Umfrage zu Anforderungen an KI auf Ressourcen beschränkter Hardware durchgeführt (s. KI Hardware Marktübersicht).

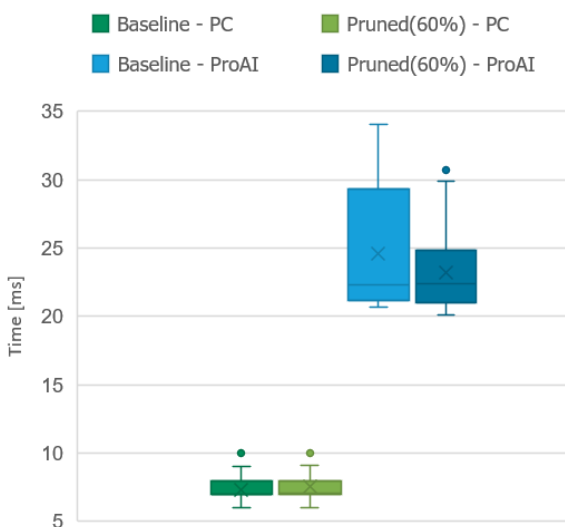


Abbildung 65: Performanz Vergleich des Single-Shot-Detektor Netzwerks zwischen automotive Computer und lokalem Workstation PC (mit NVIDIA P2000 GPU).

Im weiteren Verlauf des Projektes wurde die Methode Hrank [1] für Klassifizierung hinsichtlich der Verbesserung von Laufzeit und Speicherbedarf evaluiert und für auf den Bereich Objektdetektion erweitert. Weiterhin wurde ein Performanzvergleich zwischen automotive Computer und Workstation PC durchgeführt. In Abbildung 65 ist ein solcher Vergleich gezeigt, bei dem Baseline und das dazugehörige geprunte Modell analysiert wurde. Die Inferenzzeit zwischen Baseline und Komprimiertem Modell ist sowohl auf dem PC als auch auf dem automotive Computer sehr ähnlich. Der Unterschied wird bei der Anwendung auf der CPU statt der GPU deutlicher.

Basierend auf dieser Methode wurde eine Pruningmethode mit dem Namen „Interpretable Pruning“ entwickelt, die zusätzlich die Komponente der Erklärbarkeit beinhaltet. Durch die Verwendung von Heatmaps, die für jedes Inputbild der Pruningmethode generiert werden, kann eine visuelle Nachvollziehbarkeit des Prunings erreicht werden. Wir haben diesen Algorithmus in der letzten Projektphase noch weiter optimiert und konnten eine Kompressionsrate von 40% ohne große Verluste der Genauigkeit erreichen (siehe Abbildung 66).

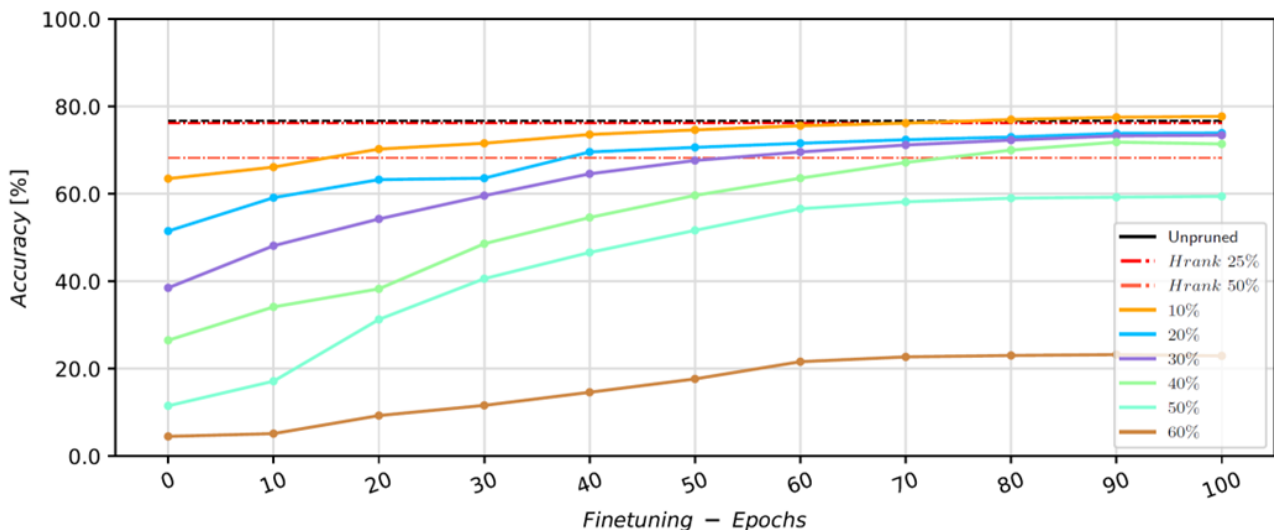


Abbildung 66: Ergebnisse des Interpretable Pruning nach der Optimierung im Vergleich zu HRank Algorithmus.

Die ersten definierten Ziele der VHB wurden bereits zu Beginn in Projektinkrement 1 und 2 erfüllt. Mit der Untersuchung und Validierung verschiedener Methoden und Netzwerke wurde eine grundlegende Basis geschaffen für die weitere Entwicklung neuer Methoden im Bereich des Prunings. Mit der Adaption von Hrank und der Entwicklung von Interpretable Pruning hat ZF zwei unterschiedliche Methoden hinsichtlich Performanz und Speichernutzung untersucht. Die HRank Methode wurde zusätzlich noch auf der ProAI hinsichtlich Performanz evaluiert.

3.4.2.2.2.5 Entwicklung einer quantisiert funktionierenden Batch Normalization Schicht um die De- und Requantisierung zu vermeiden

Batchnormalisierung findet sich in den meisten aktuellen Netzwerken. Vor allem in der beliebten ResNet Architektur ist es ein essentieller Bestandteil des Modelles. Bei der Quantisierung von Netzwerken wird oft die Batchnormalisierung ignoriert, was später zu Umwandlungen von dem quantisierten Wert wie z.B. int8 zu float32 führt. Das verlängert die Laufzeit und man kann weniger die positiven Effekte der Quantisierung ausnutzen oder es verhindert sogar sie Ausführung auf bestimmter Hardware, die keine floating point Operationen unterstützt. Vor dem

Hintergrund hat Valeo eine Reihe von verschiedenen Ansätzen ausprobiert, die die Batchnormalisierung ebenfalls quantisieren. Dabei wurde immer ein weiterer Pfad eingeführt mit einen "folded" Conv+BatchNorm Block, der im Training mit dem Standard Conv+BatchNorm Block kombiniert wird. Die Gewichte des "folded" Blocks werden quantisiert (Versionen 1, 2). Die Untersuchungen haben sich damit beschäftigt, wie man diesen "folded" Block einfügt und im Training behandelt. Dabei hat sich eine Methode mit einem "switch" als am besten erwiesen, bei der im Training an einem bestimmten Punkt der "switch" dafür sorgt, dass nur noch der quantisierte "folded" Block verwendet wird (Versionen 3, 4, 5). Die beste Version (5) hat fast die Leistung der Baseline erreicht, bei der die BatchNorm Layer nicht quantisiert sind.

Version	Base	1	2	3	4	5
Test accuracy	93.11	89.58	90.00	91.66	92.80	92.93

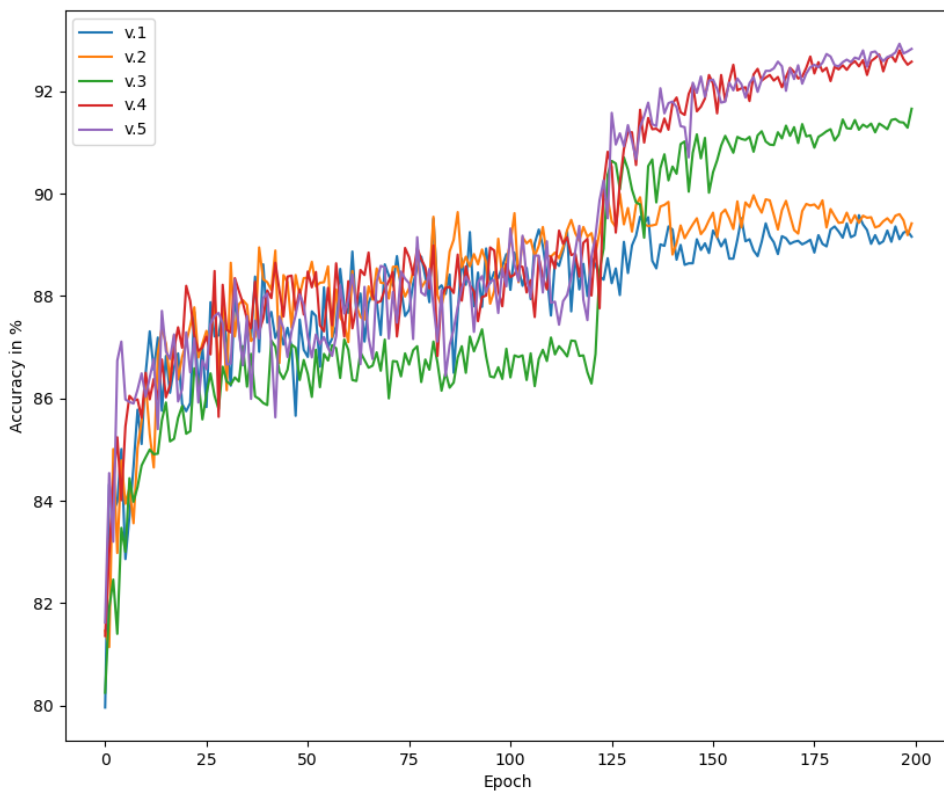


Abbildung 67: Accuracy auf den Validierungsdaten über das Training hinweg . Gut zu sehen ist der "switch" der um Epoche 125 umgelegt wird und zu einer großen Verbesserung führt.

3.4.2.2.2.6 Entwicklung einer Methode zur Reduktion der Netzwerkgrößen mithilfe von Tensor Compression

Tensorkompression ist vor allem eine Technik zur Reduktion des Speicherbedarfs, aber auch der Ausführungszeiten von neuronalen Netzen. Tensor Compression kann pro Schicht angewendet werden, was zur Erstellung von je drei Unterschichten führt, die sich ähnlich wie die ursprüngliche Schicht verhalten, aber in der Summe kleiner sind als die ursprüngliche Schicht. Eine der vielversprechendsten Komprimierungsmethoden ist die Tucker-Zerlegung, die das Tensor-Äquivalent zur Singulärwert-Zerlegung für Matrizen ist.

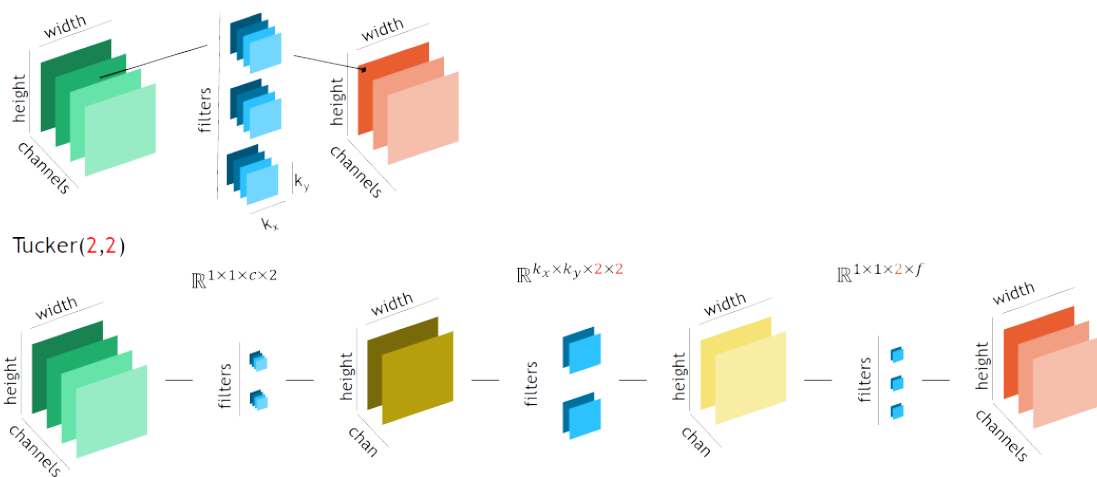


Abbildung 68: Darstellung der Tucker-Zerlegung ; Ein einzelner Layer der Größe k_x, k_y, c, f wird in drei Layer der Größen $1, 1, c, r_c$; k_x, k_y, r_c, r_f und $1, 1, r_f, f$ zerlegt, abhängig von den Matrix Rängen r_c und r_f . Abhängig von den Werten der Ränge können die drei entstehenden Layer signifikant kleiner sein als der Originallayer, aber sich dabei sehr ähnlich verhalten.

Abbildung 68 zeigt ein Beispiel für eine solche Dekomposition. In der oberen Zeile ist die zu komprimierende Originalschicht mit ihrem 3D-Eingangstensor (grün) mit 4 Kanälen, 3D-Ausgangstensor (rot) mit 3 Kanälen und ihrem 4D-Gewichtungstensor (blau) mit 4 Kanälen und 3 Filtern dargestellt. Die untere Zeile zeigt die Dekomposition mit beiden Rangwerten gleich zwei. Echte Schichten, auf die die Tucker-Zerlegung sinnvoll angewandt werden kann, müssen eine weitaus größere Anzahl von Kanälen und Filtern haben, und auch die Tucker-Ränge sind in der Regel wesentlich größer als zwei. Der erste Tensor aus der Zerlegung hat eine Größe von $1, 1, c, 2$ (da $r_c=2$) und liest somit die 4 Kanäle des Eingabetensors pixelweise ein und berechnet einen ersten Zwischenausgangstensor (braun) mit nur zwei Kanälen. Die innere Schicht hat die gleiche Kernelgröße wie das Original, benötigt aber nur zwei Kanäle und zwei Filter (da $r_c=r_f=2$) und erzeugt den zweiten Zwischentensor (gelb) mit nur zwei Kanälen. Die letzte Schicht liest dann die beiden Kanäle pixelweise ein und konvertiert sie zurück in die ursprünglichen 3 Kanäle des Ausgangstensors (rot). Die letzte Schicht fügt anschließend die f -Bias-Werte aus der ursprünglichen Schicht hinzu und fügt auch die ursprüngliche Aktivierungsfunktion hinzu.

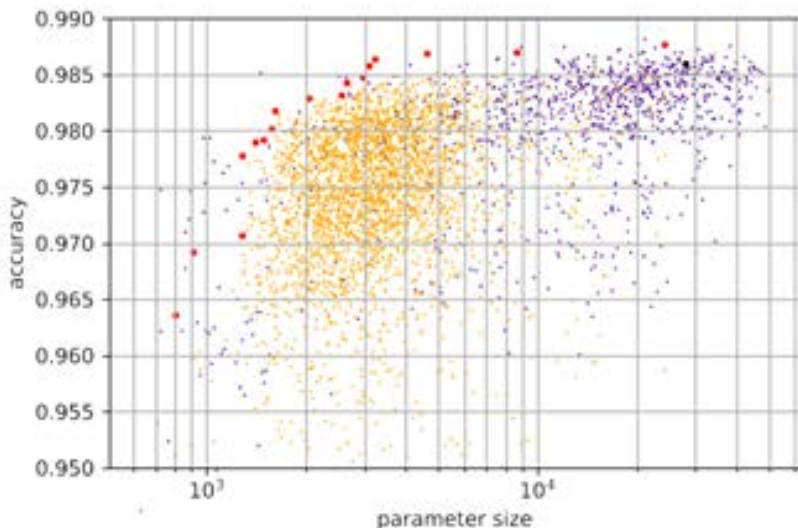


Abbildung 69: Analyse des generischen KI-Benchmarks . Die Anzahl der Parameter der Faltungsschichten auf der horizontalen Achse ist logarithmisch gegen die Top-1-Genauigkeit auf der vertikalen Achse aufgetragen. Schwarz: ursprüngliche AI-Topologie, orange und rot: Lösungen für NAS-gesteuerte Tensorkompression, rot: Pareto-Punkte, blau: vollständige NAS (auf Filtergrößen) zum Vergleich.

Wie in Abbildung 69 und Abbildung 70 dargestellt, verbessert die NAS-gesteuerte Tensorkompression eindeutig den Stand der Technik bei der Tensorkompression, was zu dreimal kleineren KI-Topologien bei gleichem Genauigkeitsverlust führt.

Die gefundenen Lösungen liegen sehr nahe an der Pareto-Front. Dies liegt daran, dass die Suchheuristik vielversprechende Konfigurationen leicht identifizieren kann und nur in den relevantesten Unterräumen sucht. Zum Vergleich hat OFFIS auch einen vollständigen NAS für die Filtergrößen durchgeführt. Da nach einer Änderung der Filtergrößen die KI komplett neu trainiert werden muss, ist dies extrem langsam und dauert selbst bei den leistungsstärksten GPUs über eine Stunde pro Lösung. Da die Tensorkomprimierung nur eine Feinabstimmung erfordert, rechnet sie 3-4 mal schneller, aber beide Methoden sind um Größenordnungen langsamer als der Stand der Technik bei der Tensorkomprimierung und dem Filter Pruning. Da die Optimierung jedoch nur einmal auf einem Server ausgeführt wird und dann potenziell Inferenzzeit für Millionen von eingebetteten Systemen einsparen kann, ist die Optimierungslaufzeit nicht das primäre Ziel von OFFIS, solange sie innerhalb vernünftiger Grenzen bleibt.

Die ersten Analysen zeigen, dass NAS mit Kompression sogar eine vollständige NAS-Suche im mittleren Bereich übertrifft (zwischen 5- und 10-fache Größenreduktion). Darunter und darüber wird die Pareto-Front von der vollständigen NAS-Suche dominiert. Nichtsdestotrotz führt die Kompression immer zu Lösungen, die sehr nahe an den besten Full-NAS-Lösungen liegen, und rechnet erheblich schneller.

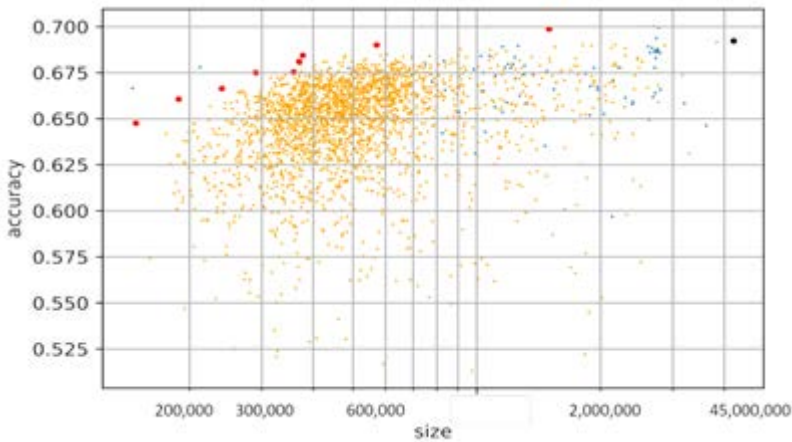


Abbildung 70: Dieselbe Visualisierung wie in Abbildung 69, aber für den AlexNet-Benchmark.

Der Vorteil für die benötigte Speichergröße (für den diese Technik entwickelt wurde) ist unmittelbar offensichtlich, es wurde dann im letzten Projektabschnitt noch untersucht, wie sehr sich diese Technik auch auf die Ausführungszeit auswirkt. Die Tensor Kompression kann ganz eindeutig die Zahl der benötigten Rechenoperationen reduzieren - je nach Hardware-Ressource muss das aber nicht notwendigerweise zu einer Reduktion der Ausführungszeiten führen.

	VPU		CPU	
	vorher	nachher	vorher	nachher
Generic	0.737s	0.867s	0.407s	0.381s
AlexNet	1.409s	0.989s	0.847s	0.349s

Tabelle 5: Ausführungszeiten auf verschiedenen Hardwareplattformen vor und nach der Kompression.

Wie in Tabelle 5 dargestellt ist, profitieren insbesondere heterogene Hardwarekomponenten wie CPUs sehr von der Reduktion der Zahl der Berechnungen und überwiegen den Overhead durch die zusätzlichen Schichten deutlich. Für massiv parallele Hardwarearchitekturen wie die VPU geht ein großer Teil des Vorteiles durch die Reduktion der Zahl der Rechenoperationen wieder verloren, da die entstandenen neuronalen Netze weniger regulär sind und sich daher weniger effizient berechnen lassen. Je nach Beispiel kann sich dies sogar in einer Erhöhung der Ausführungszeit auswirken. Die Reduktion des Speicherbedarfs bleibt allerdings auch hier erhalten.

3.4.2.2.3 Gegenüberstellung der vorgegebenen Ziele

Die in der VHB definierten Ziele wurden im ersten Projektinkrement diskutiert und es wurden Synergien identifiziert, wie etwa, dass nahezu alle beteiligten Partner sich in der Quantisierung engagieren möchten. Statt dessen wurde konsolidiert und es wurden neue, in der Zeit zwischen Antragstellung und P11 neu aufgekommene Ideen hinzugefügt. Auf diese Weise wurde neben der geplanten KPI Vorhersage, dem Pruning und der Quantisierung auch noch die Tensor Compression als Technik untersucht. In allen Bereichen wurden die geplanten Arbeiten durchgeführt, die geplanten Ergebnisse erzielt und dann publiziert.

3.4.3 Zielerreichung

Im Bereich der Robustheit in Open-World wurden die Ziele in zwei Kategorien eingeteilt: Quantifizierung der Robustheit von KI-Systemen und Methoden zur Steigerung der Robustheit von KI-Systemen.

Quantifizierung der Robustheit: Entsprechen der definierten Ziele wurden zwei Datensätze zur Untersuchung der Robustheit mit Hilfe von Novelty-Detektion spezifiziert, generiert und annotiert. Außerdem wurden Ansätze basierend auf tiefen neuronalen Netzen entwickelt, welche die Novelty in Verkehrssituationen evaluiert. Weiterhin wurde die Unsicherheit in der Tiefenschätzung untersucht sowohl basierend auf Lidar-Punktwolken als auch auf der Tiefenschätzung aus Kamerabildern. Außerdem wurde eine Methode entwickelt, welche die Unsicherheit in der geschätzten Pose von Verkehrsteilnehmern evaluiert. Die entwickelten Modelle helfen dabei die Robustheit der im Fahrzeug eingesetzten KI-Systeme in neuen und ungewohnten Situationen zu bewerten und somit auch in unsicheren Situationen dennoch sichere Fahrmanöver zu planen.

Steigerung der Robustheit: Zur Steigerung der Robustheit wurden Tiefenschätzungsverfahren entwickelt, welche auch unter neuen Bedingungen, z.B. Wetterbedingungen, robuste Vorhersagen treffen. Dieses so genannte zero-shot Verfahren setzt vor allem auf eine neue Trainingsstrategie unter Nutzung von Daten-Augmentierung um auch in unvorhergesehenen Situationen die Regressions-Leistung zu erhalten. Des Weiteren wurde eine hierarchische

Klassifikationsmethode untersucht und optimiert, welche dem KI-Modell die Möglichkeit gibt auf Klassen, welche in einer Label-Hierarchie weiter oben liegen, z.B. Fahrzeug statt PKW, zurückzufallen, falls in der gegebenen Situation die Klassifikation unsicher ist.

Es wurden im Bereich der Embedded Systems zu Projektbeginn Ziele in drei Kategorien definiert: Bewertung von KI Hardware, Vorhersage von Metriken und Optimierung von Metriken.

Bewertung von KI Hardware: Es wurde entsprechend der definierten Ziele eine Befragung industrieller und akademischer Anwender innerhalb und außerhalb des Projektes durchgeführt. Aus deren Auswertung ergab sich ein Ziel für eine Marktrecherche um die Daten der Befragung zu ergänzen. Daraufhin konnte eine Erhebung der Hardware-Zielplattformanforderungen und automobilspezifischer Kriterien durchgeführt und dokumentiert werden. Dieser Teil soll dem Embedded AI Entwickler eine Hilfestellung geben, welche am Markt verfügbare Hardwareplattform für die jeweilige Anwendung optimal ist.

Vorhersage von Metriken: Da die Kennzahlen der oben genannten Bewertung zwar einen guten Einblick in die generelle Eignung von KI Hardware gewähren, aber diese je nach Anwendung natürlich noch variieren, wurde eine Vorhersagemethodik entwickelt, die es KI-Entwicklern ermöglicht möglichst relevante Aspekte wie die genaue hardwareabhängig zu erwartende Latenz schon früh im Entwurfsprozess akkurat berücksichtigen zu können. Dies wurde durch eine Hardware-in-the-Loop (HIL) basierte Methode ergänzt, um Hardware bezogene Leistungskennzahlen bei der Inferenz von Neuronalen Netzen zu messen. Dadurch ist die Werteerstellung (Charakterisierung) sowie Überprüfung der Vorhersagemethode möglich. Nach der optimalen Wahl der Hardware in der ersten Kategorie ist hier die Wahl des optimalen neuronalen Netzes für diese Hardware möglich.

Optimierung von Metriken: In dieser Kategorie sollen nun die gewählten neuronalen Netze für die gewählte Hardware weiter verfeinert werden, so dass sich die Eigenschaften weiter verbessern. Hier wurden drei prinzipielle Methoden gewählt. Das Entfernen von nicht relevanten Netzwerkteilen (Pruning) wurde mithilfe eines Reinforcement Learning Agenten unter Zuhilfenahme der oben beschriebenen Vorhersagemethoden durchgeführt, um den Zusammenhang zwischen Komprimierungstechnik und KI-Beschleuniger zu lernen und optimal zu nutzen. Es wurde hier eine Pruning-Methode entwickelt, welche das Verlernen wichtiger Aspekte durch das Pruning vermeiden kann. Die zweite entwickelte Methode ist die Quantisierung zur Verringerung des Speicherbedarfs und des Rechenaufwandes. Hier wurden die vorhandenen Methoden derart erweitert, das nach der Quantisierung der Convolution und Activation Layer (ist heute Stand der Technik) auch die Batch Normalisierung Layer quantisiert werden können. Dies vermeidet eine aufwändige Dequantisierung vor und Requantisierung nach jeder Batch Normalisierungsschicht. Schließlich wurde eine Optimierungstechnik entwickelt, die auf Tensor Compression beruht und es ermöglicht, große Convolution Schichten durch wesentlich kleinere und daher weniger speicherintensive Schichten zu ersetzen. Dies ist für Speicher- und Energiebedarf immer vorteilhaft, für die Laufzeit allerdings nur für bestimmte, wenig parallelisierte Hardwaresysteme.

3.4.4 Referenzen

[1] M. Lin, R. Ji, Y. Wang, Y. Zhang, B. Zhang, Y. Tian und L. Shao, „HRank: Filter Pruning using High-Rank Feature Map,“ *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

3.5 TP5: Evaluation

3.5.1 Ausgangssituation und Zielsetzung

Um den Hub einer Entwicklung beurteilen zu können ist eine Evaluation der Ergebnisse am Ende des Projekts gegenüber dem Stand vor dem Projekt nötig. Teilprojekt 5 hat die verschiedenen Aspekte der Evaluation des Projekts zusammengefasst. Der Beitrag der technischen Teilprojekte (TP1 bis TP4) für den Einsatz Künstlicher Intelligenz im Hinblick auf Robustheit gegenüber Deltas und Methoden zur Integration von Deltas im Automotive-Umfeld wurden aufgezeigt.

Im Teilprojekt wurde als Baseline nicht nur der Stand der Technik zu Beginn des Projekts erhoben, sondern dieser wurde kontinuierlich aktualisiert um den Nutzen des Projekts nicht nur gegenüber dem Stand vor dem Projekt, sondern auch gegen mögliche Entwicklungen außerhalb des Projekts bewerten zu können. Um Vergleichbarkeit herzustellen war es zudem nötig, die identifizierten Modelle der Baseline jeweils mit den gleichen Daten zu trainieren, die auch für das Training der entwickelten Modelle genutzt wurden. Dieser Schritt wurde in jedem Projektinkrement durchgeführt. Da zu Beginn des Projektes noch keine eigenen Daten zur Verfügung stehen konnten, wurden für die erste Projektinkrement-Phase nicht nur für die Entwicklung, sondern auch für die Evaluation verfügbare Daten verwendet. Aufbauend auf der Evaluation der Modelle und Methoden in einem Projektschritt wurden in TP5 auch die Anforderungen für alle technischen Arbeitspakete im nächsten Projektschritt definiert. Dies galt nicht nur für den Projektstart zu dem die Anforderungen für Entwicklung, Use Cases und Datenaufnahme definiert werden mussten, sondern auch für den Start jedes Projekt-Inkrementes.

Zum Ende des Projekts in der letzten Projektinkrement-Phase hat TP5 dann die Gesamtevaluation des Projekts basierend auf Daten die nicht zum Training der Modelle genutzt wurden durchgeführt und die Performance mit der Performance der Referenzmodelle auf Basis von KPIs verglichen, die zu Beginn des Projektes im TP definiert wurden.

3.5.2 3.5.2 Strukturierung

Aufgrund der hohen Dynamik in der Entwicklung neuer Ansätze im Bereich KI war ein iterativer Zeitplan erforderlich, der eine Einbeziehung neuer Entwicklungen aus der Forschung ermöglichte. Außerdem war durch mehrere Projekt-Inkmente und iterative Prozesse eine kontinuierliche Verbesserung und Optimierung der zu entwickelnden Ansätze im Projekt KI-Delta Learning möglich. Aus diesen Gründen wurden die TP5-Arbeitspakete als Teil des Workflows der Projekt-Inkmente gesehen, die sich mehrfach im Projekt wiederholen. Wie bereits beschrieben bestand das Projekt aus einer Startphase und vier Projektinkrementen.

Alle Partner waren in TP5 beteiligt und haben zu den Ergebnissen beigetragen. TP5 bildet den Rahmen für Evaluation der Ergebnisse, Austausch der Ergebnisse und Aktualisierung der Anforderungen, in dem alle Partner zusammenkommen. Dieser Austausch wird im Rahmen geplanter Projekttreffen und Workshops realisiert.

3.5.2.1 AP5.1 Review und Refinement

Die bereits in der VHB beschriebenen Anforderungen, Spezifikationen und Use Cases wurden in der Startphase zum Beginn des Projekts verfeinert und auf System- und Methodenanforderungen heruntergebrochen. Insbesondere lag der Fokus auf der Definition der Anforderungen für die zu entwickelnde Tool-Kette, der Spezifikation der zu generierenden Daten und Beschreibung der Use Cases. Am Anfang des Projektes wurde eine zentrale Risikoanalyse als Teil der Datenschutzerklärung durchgeführt. Dabei werden alle Datenbearbeitungsschritte wie z.B.

Recording, Data-Offloading, Labeling, Training etc. berücksichtigt. Die Maßnahmen für die Risikoreduzierung wurden geplant und die Verantwortlichkeiten für die Datenschutzregeln werden definiert. Mit Hilfe der Risikoanalyse wurden die Datenschutz-Themen innerhalb des Projektes geklärt.

Außerdem wurden zu Beginn jedes neuen Projektinkrements die Ergebnisse des vergangenen Projektinkrements in Form von Evaluationen überprüft und diskutiert. Daraufhin wurden für die neue Projektinkrement-Phase die definierten Anforderungen und Spezifikationen anhand der neu gewonnenen Kenntnisse aktualisiert und angepasst. Die Entwicklung der Tool-Kette, der Daten-Generierung und der algorithmischen Entwicklung erfolgte dann mit aktualisierten Anforderungen. Dieses Vorgehen hat sich iterativ einmal pro Projekt-Inkrement bis zum Ende des Projekts wiederholt.

3.5.2.2 AP5.2 Baseline und Training der Referenzmodelle

Um auszuwerten, inwieweit die Überbrückung der in den technischen TPs behandelten Deltas den aktuellen Stand der Technik vorantreibt, musste dieser zunächst recherchiert und analysiert werden. Zum Beginn jedes Projekt-Inkrements und direkt nach Review und Refinement wurde der Stand-der-Technik neu untersucht um festzustellen, ob die verwendeten Referenzmodelle aktuell sind. Als Ergebnis wurden ggf. neue geeignete Referenzmodelle für die individuellen Use Cases als Baseline mit einbezogen. Außerdem wurden in diesem AP die ausgewählten Referenzmodelle mit vorhandenen Daten aus öffentlichen und zugänglichen Datensätzen sowie mit im Projekt aufgenommenen und gelabelten Daten trainiert, um eine Vergleichbarkeit mit den im Projekt erarbeiteten Modellen herzustellen.

3.5.2.3 AP5.3 Evaluation und Demonstration

In Abwesenheit von üblichen physikalischen Prototypen in Form von Versuchsträger, wurden die Ergebnisse dieses Forschungsprojektes in Form der Evaluationen demonstriert. Dabei wurden die Ergebnisse der technischen Teilprojekte sowohl unter qualitativen als auch unter quantitativen Gesichtspunkten bewertet. Es wurden verschiedene KPIs für die individuellen Use Cases definiert und automatisiert für die entwickelten Ansätze ausgewertet. Der Grundgedanke dabei war die Anforderung (Daten und Metrik) vorzugeben, jedoch den Lösungsansatz frei in der Hand der Projektpartner zu belassen. Das jeweilige Datenset und Metriken wurden iterativ am Anfang jedes Projektinkrements festgelegt. Insofern stellte jedes Projektinkrement eine Challenge - ähnlich zu Kaggle - an die Partner.

Ein Vergleich mit den entsprechenden KPIs der Baseline-Methoden aus AP5.2 lieferte eine quantitative Evaluation. Zudem wurde eine qualitative Evaluation und visuelle Demonstration der Ergebnisse auf Prototypen und durch Simulation auf aufgezeichneten Realdaten realisiert. Analog zu anderen APs dieses Teilprojektes wurde die Evaluation bis Ende des Projektes iterativ einmal pro Projekt-Inkrement durchgeführt. Die Ergebnisse aus AP5.3 wurden dann in AP5.1 Review und Refinement diskutiert und in Form von Dokumenten und Publikationen zur Verfügung gestellt.

3.5.2.4 AP5.1 Review und Refinement

3.5.2.4.1 Zielsetzung

Die in der VHB beschriebenen Anforderungen, Spezifikationen und Use Cases sollten in der Startphase zum Beginn des Projekts verfeinert und auf System- und Methodenanforderungen heruntergebrochen werden. Der Fokus sollte insbesondere auf der Definition der Anforderungen

für die zu entwickelnde Tool-Kette, der Spezifikation der zu generierenden Daten und Beschreibung der Use Cases liegen. Am Anfang des Projektes sollte eine zentrale Risikoanalyse als Teil der Datenschutzerklärung durchgeführt werden. Dabei sollten alle Datenbearbeitungsschritte wie z.B. Recording, Data-Offloading, Labeling, Training etc. berücksichtigt werden. Die Maßnahmen für die Risikoreduzierung sollen geplant und die Verantwortlichkeiten für die Datenschutzregeln sollen definiert werden. Mit Hilfe der Risikoanalyse sollten die Datenschutz-Themen innerhalb des Projektes geklärt werden.

Außerdem war vorgesehen zu Beginn jedes neuen Projekt-Inkrementes die Ergebnisse des vergangenen Projekt-Inkrementes in Form von Evaluationen zu überprüfen und zu diskutieren. Auf dieser Basis sollten die definierten Anforderungen und Spezifikationen für die neue Projekt-Inkrement-Phase anhand der neu gewonnenen Kenntnisse aktualisiert und angepasst werden. Die Entwicklung der Tool-Kette, der Daten-Generierung und der algorithmischen Entwicklung soll dann mit aktualisierten Anforderungen erfolgen. Laut Projektplan sollte dieses Vorgehen iterativ einmal pro Projekt-Inkrement bis zum Ende des Projekts erfolgen.

3.5.2.4.2 Ergebnisse

Die Überprüfung der Ergebnisse eines Projektinkrementes und die darauf aufbauende Nachjustierung der Ziele des kommenden Projektinkrementes wurde in zweitägigen "Review and Refinement"-Treffen durchgeführt. Am ersten Tag wurden jeweils die Ergebnisse der APs vorgestellt und diskutiert. Am zweiten Tag wurde in AP-Gruppenarbeit die "AP Capacity Heatmap" für das kommende Projektinkrement abgestimmt. Auf diesen von der Projektkoordination für jedes AP vorbereiteten Confluenceseiten wird festgehalten wann die beteiligten Partner mit welcher Personalstärke im kommenden Projektinkrement arbeiten. Außerdem wurden die konkret erwarteten Ergebnisse diskutiert und die dazu beitragenden Partner wurden festgelegt. Neben diesen Heatmaps als formelles Ergebnis stellen die bei der Erstellung geführten vielen Detaildiskussionen und der fruchtbare Austausch ein weiteres Highlight dieser Aktivitäten dar.

R&R1 Worksheet AP1.3 (SB)

Erstellt von Niko Papamichail vor Kurzem

Capacity Heat Map AP1.3 - PI2

- Please use the following color coding to fill out the heat map
- Each partner involved in the result shall indicate their planned capacity for the next PI
- The current capacities are based on the last entries by all partners and should be updated for the next PI during this workshop

Status	Color code
not available	
< 1 PM	
= 1 PM	
> 1 PM	

		2021											
		Q1			Q2			Q3			Q4		
		PI2 Development						PI3 Optimization					
		1	2	3	4	5	6	7	8	9	10	11	12
TP1 Datengewinnung	Last update	Total PMs	name of developer										
AP 1.3 Generierung synthetischer Daten													
	VALEO	27.05.2020	12										
	PEG	26.05.2020	2										
	FZI												
	OFFIS	25.05.2020	18										
	BUW	21.01.2021	3										
	HSRT	03.06.2020	6										
	InnoSent	08.12.2020	36										

Result Planning AP1.3 - PI2

- Please use the link in the table below to be forwarded to the respective result
- Please use the section "Tasks for this Result" to discuss what is expected to complete the result in PI2: i.e. what contributions are needed from each partner, when are these contributions due etc.
- Overall the idea is to plan the next PI and make sure all involved partners are on the same page what is expected from them to complete the result

ID	Result Title	Involved Partners	Link to result for PI2 planning
E2.2.1.3	Tools zur Generierung von Daten und synthetischer Datensatz. (Simulatorumgebung, GANs)	InnoSent, OFFIS, Valeo	Link

Gefällt mir Sei der Erste, dem dies gefällt.

Keine Stichwörter

Abbildung 71: Beispielhafte Heatmap für AP1.3 für das zweite Projektinkrement

Die Terminierung der Projektinkremente hat die Zeitpunkte der Review and Refinement Treffen bestimmt:

- Zwischen erstem und zweiten Projektinkrement: 01. und 02.12.2020
- Zwischen zweitem und drittem Projektinkrement: 21. und 22.09.2021
- Zwischen drittem und vierten Projektinkrement: 06. und 07.09.2022

Beim dritten Treffen wurde der Termin etwas nach hinten geschoben um den Start des Projektinkrements nicht vor der ferienbedingten Abwesenheit vieler Mitarbeiter durchzuführen, sondern ihn als gemeinsamen Auftakt zum letzten Projektinkrement bestmöglich zu nutzen. Es wurde auch deswegen Wert auf eine größtmögliche Teilnahme gelegt, weil es das erste physische Review and Refinement Treffen war. Die ersten beiden mussten aufgrund der Covid-19-Pandemie als virtuelle Veranstaltung durchgeführt werden. Im Konsortium bestand der klare Wunsch wieder ein Vor-Ort-Treffen zu veranstalten und auch die Projektleitung hat sich für ein physisches Treffen ausgesprochen um nach den personellen Wechseln im Managementteam die Projektteilnehmer persönlich kennen zu lernen. Die hohe Teilnehmerzahl von 60 zeigt das starke Interesse.



Abbildung 72: Gruppenbild vom dritten Review and Refinement Treffen

3.5.2.4.3 Gegenüberstellung der vorgegebenen Ziele

KI ist ein aktuelles Forschungsgebiet, das sich rasant weiterentwickelt und in dem sich während der Laufzeit eines Förderprojektes sehr wahrscheinlich neue Ideen, andere Ansätze oder unvorhergesehene Lösungswege entwickeln. Die iterative Projektstruktur mit den Review and Refinement Treffen hat dafür gesorgt, dass die untersuchten und entwickelten, beziehungsweise weiterentwickelten Methoden und Maßnahmen auf Höhe des State of the Art waren und geblieben sind. Ein deutliches Zeichen dafür sind die über 100 wissenschaftlichen Veröffentlichungen, die größtenteils auf sehr renommierten Konferenzen vorgestellt wurden und die somit aktiv das Forschungsfeld mit prägen. Auch die Tatsache, dass der nun etablierte Workshop on Autonomy@Scale, der aus KI Delta Learning heraus als Teil des IEEE Intelligent Vehicles Symposium initiiert wurde, auch nach Projektende weitergeführt wird, zeigt, dass die Aktivitäten des Projektes stets aktiver Teil der aktuellen, internationalen Forschungslandschaft waren.

3.5.2.5 AP5.2 Baseline und Training der Referenzmodelle

3.5.2.5.1 Zielsetzung

Die Entwicklung von KI und auch der Einsatz von KI für das autonome Fahren sind sich rasch weiterentwickelnde Forschungsfelder, die fortlaufend neue Ansätze und Erkenntnisse liefern. Forschungsrichtungen die heute noch als vielversprechend gelten, können durch neue Ergebnisse überholt oder zumindest stark beeinflusst werden, sodass sich die erfolgversprechendsten Wege

ein Ziel zu erreichen, in einem Zeitrahmen ändern können, der deutlich unter der Laufzeit dieses Projektes liegt. Um also die eingesetzten Ressourcen optimal einzusetzen ist es nötig das Forschungsfeld während der Projektlaufzeit kontinuierlich zu beobachten und neue Tendenzen in das Konsortium zu tragen und die Projektarbeit gegebenenfalls nach zu justieren. Das ist die Aufgabe von AP5.2. Dort wurde vor und nach jedem Projektinkrement der Stand-der-Technik neu untersucht um festzustellen, ob die verwendeten Referenzmodelle aktuell sind. Außerdem wurden in diesem AP die ausgewählten Referenzmodelle mit vorhandenen Daten aus öffentlichen und zugänglichen Datensätzen sowie mit im Projekt aufgenommenen und gelabelten Daten trainiert um eine Vergleichbarkeit mit den im Projekt erarbeiteten Modellen herzustellen.

3.5.2.5.2 Erzielte Ergebnisse im Einzelnen

Zu Beginn des Projektes wurde eine umfangreiche State-of-the-Art-Recherche durchgeführt und als Deliverable 03a dem Konsortium zur Verfügung gestellt. Auf über 200 Seiten wurde für alle technischen Bereiche des Projektes die Ausgangslage zusammengefasst. Während der Projektlaufzeit wurden dann zum Ende jedes Projektinkrements in verschiedenen Bereichen Defizitanalysen für den aktuell erarbeiteten Stand durchgeführt. Darauf aufbauend wurden Pläne für die nächste Optimierung im kommenden Projektinkrement abgestimmt. Dies wurde für die TPs 2-4 durchgeführt, also für die Themen Transfer Learning, Didaktik und Automotive-Tauglichkeit. Die Ergebnisse wurden als Teil der Deliverables 04, 07, 11 und 15 dokumentiert. Es war geplant, diese Defizitanalysen mit dem KI Delta Learning Datensatz durchzuführen, der aber erst deutlich verzögert zur Verfügung stand. Aus diesem Grund wurden dafür öffentlich zugängliche Datensätze verwendet.

Zu Projektende wurde nochmals ein eigenes Deliverable in AP5.2 erstellt, Deliverable 03b, in dem für die drei Forschungsfelder jeweils eine Übersicht zum Stand nach Ende der vier Projektinkremente zusammengestellt wurde, bestehend aus einer Methodensammlung, einem Katalog und einer Übersicht zum SOTA.

3.5.2.5.3 Gegenüberstellung der vorgegebenen Ziele

Alle vorgesehenen Ergebnisse wurden erbracht und über die Projektplattform den Partnern nach Abschluss der jeweiligen Projektinkremente zur Verfügung gestellt. Aufgrund der Verzögerungen wurden die Enden der Projektinkremente im Laufe des Projektes nach hinten geschoben und damit auch die Lieferzeitpunkte der Ergebnisse aus AP5.2. Abgesehen von Deliverable 03a wurden deswegen alle Ergebnisse später produziert als in der VHB vorgesehen, aber wegen der generellen Verzögerung des Projektes trotzdem zum vorgesehenen Zeitpunkt bezüglich des internen Workflow im Projekt.

3.5.2.6 AP5.3 Evaluation und Demonstration

3.5.2.6.1 Zielsetzung

Zur Gesamtevaluation des Projekts fasst dieser Bericht die Ergebnisse aus den Teilbereichen *Automotive-Tauglichkeit*, *Transfer Learning* und *Didaktik* zusammen und bereitet sie dahingehend auf, dass eine teilprojektübergreifende, thematische Übersicht entsteht. Die Gliederung erfolgt nach dem jeweiligen Zweck der erforschten Methoden und ermöglicht so einen schnellen Gesamtüberblick über die Themen. Teilweise sind die bearbeiteten Themen mehreren Bereichen zuzuordnen, und bei der Gliederung nach Methoden wurde versucht, die Inhalte möglichst redundanzarm wiederzugeben. Um die nachfolgenden Abschnitte zu erstellen, wurden die Ergebnisse der Partner zunächst tabellarisch erfasst und sind hier unabhängig von der Projektstruktur neu wiedergegeben.

3.5.2.6.2 Erzielte Ergebnisse im Einzelnen

3.5.2.6.2.1 Perzeptionsmethoden

Dieser Abschnitt umfasst die Entwicklungen von Methoden im Bereich der Perzeption bzw. Mustererkennung in Sensordaten. Allgemeines Ziel ist die (semantische) Interpretation der Sensordateninhalte (insb. Kamera und LiDAR). Angewendet auf Sensordaten die aus dem Fahrzeug heraus aufgenommen wurden, ermöglichen die Methoden damit die Wahrnehmung von Objekten und Hindernissen im unmittelbaren Fahrzeugumfeld.

3.5.2.6.2.1.1 Klassifikation

Ziel ist die Bestimmung eines semantischen Typs aus den gegebenen Eingabedaten. Typischerweise bedeutet dies, die Bedeutung des wesentlichen Kerninhalts zu klassifizieren, beispielsweise zu genau einem Eingabebild das wesentliche im Bild dargestellte Objekt zu klassifizieren. Stand der Technik ist die Klassifikation von mehreren hundert bis mehreren tausend verschiedener Klassen, so werden derartige Verfahren beispielsweise bei der Erkennung von Tier- und Pflanzenarten, Texterkennung, Verkehrszeichen und vielem mehr eingesetzt. Die Aufgabe wird oft als nachgelagerter Schritt einer gröberen Objekterkennung betrachtet, so dass in diesem Fall vor allem kleinere Bildregionen (Größenordnung bis kPixel) zu klassifizieren sind, und nur ein einziges zu klassifizierendes Objekt in den Eingabedaten wesentlich ist. Aufgrund der Generalisierbarkeit erfolgt die Forschung in diesem Bereich oft anwendungsübergreifend - so lässt sich ein Verfahren je nach Art der Trainingsdaten für verschiedene Zwecke verwenden und trainieren (z.B. überwacht oder teilüberwacht).

Im Rahmen von KI Delta Learning hat die **BUW** hierzu Forschungsarbeiten durchgeführt. Ziel ist die signifikante Reduktion der neuronalen Netze (bzw. der zu trainierenden Parameter innerhalb der Netze) bei möglichst nur unwesentlicher Reduktion der Klassifikationsqualität und -genauigkeit - in diesem Fall durch strukturierte Verringerung der Gewichte durch das Ausnutzen von Ähnlichkeiten von Mehrgitter Methoden, Teilen von Faltungsnetzen und Reduktion von Hierarchien innerhalb der Netze. Angewendet auf verschiedene Datensätze (Cifar10, Cifar100, FashionMNIST, TinyImageNet), ergab das Verfahren eine Reduktion von 50% - 90% der Parameter bei nahezu gleichbleibender, teilweise sogar leicht besserer, Genauigkeit der Klassifikation. Insbesondere für Probleme, die weniger Kapazität brauchen, lässt sich die Genauigkeit bei drastischer Gewichtsverringering erhalten. Die damit verbundene die Laufzeit- und Speicherplatzreduktion (vgl. Kompression / Pruning im Abschnitt Optimierungsmethoden) ist somit für den Einsatz in Echtzeitsystemen und bei eingeschränkter Hardware von Bedeutung. Offen ist hier die Anwendung auf automotive-spezifische Datensätze und komplexere Problemstellungen.

Verwandt dazu sind die Arbeiten von **ZF**, hier wurde am Beispiel der Datensätze Cifar100 und Pascal VOC, untersucht, wie Klassifikationsaufgaben für eingebettete System, d.h. unter reduzierten Systemanforderungen, optimiert werden können. Hierzu wurde zunächst ein HRank-Algorithmus für diverse Backbone-Architekturen entwickelt, und nach Erfolg auch auf komplexere Aufgaben wie Objektdetektion angewendet. Potenzial wird auch bei Segmentierungsaufgaben gesehen.

Die **Universität Stuttgart** hat in diesem Kontext Methoden zur Online Test-time Adaptation (TTA) entwickelt, das Thema adressiert die abfallende Leistungsfähigkeit von tiefen neuronalen Netzen, wenn sich die Umgebungsbedingungen ändern. Dabei wird das Modell direkt während der Testzeit mit den zur Verfügung stehenden Testdaten angepasst. Die entwickelte Methode

eignet sich besonders gut für Anwendungen, in denen sich die Umgebung kontinuierlich weiterentwickelt, kann allerdings auch mit schlagartigen Umgebungsänderungen sehr gut umgehen. Die entwickelte Methode wurde mit vergleichbaren generischen Daten wie oben (CIFAR10-C, CIFAR100-C, ImageNet-C, ImageNet-R, DomainNet-126) untersucht und auf einem breiten Spektrum unterschiedlicher Domänen-Änderungen getestet. Die Fehlerraten konnten je nach Datensatz / Modell um 30% - 60% reduziert werden. Potenziell eignet sich TTA auch für andere Aufgaben wie bspw. semantische Segmentierung. Die Methode kann potentiell ebenfalls an ihre Grenzen stoßen, wenn sehr ungünstige Testszenarien erfahren werden (stark korrelierte Daten).

Um robustere Maschinenmodelle in Bezug auf Domänenverschiebungen zu erlernen, hat **Bosch** neurosymbolische Ansätze untersucht und entwickelt, die eine Überwachung des Trainings durch bilddateninvariantes Hilfwissen ermöglichen. Das Hilfwissen wird zunächst in einem Wissensgraphen mit entsprechenden Konzepten und deren Beziehungen kodiert, der dann über ein Einbettungsverfahren in eine dichte Vektordarstellung transformiert wird. Mit Hilfe einer kontrastiven Verlustfunktion können die entwickelten Verfahren wie Knowledge Graph Neural Network (KG-NN) lernen, ihren visuellen Einbettungsraum und damit ihre Gewichte entsprechend dem bilddateninvarianten Wissensgraphen-Einbettungsraum anzupassen. Wir evaluieren KG-NN an visuellen Transfer-Lernaufgaben zur Klassifikation unter Verwendung des mini-ImageNet-Datensatzes und seiner Derivate sowie von Daten zur Verkehrszeichenerkennung (z.B. GTSRB und CTSD) und aus Deutschland und China. Die Ergebnisse zeigen, dass ein visuelles Modell, das mit einem Wissensgraphen als Trainer trainiert wurde, ein mit Cross-Entropie trainiertes Modell in allen Experimenten übertrifft, wenn die Domänenlücke zunimmt. Neben besserer Leistung und größerer Robustheit gegenüber Domänenverschiebungen passen sich diese KG-NN an mehrere Datensätze und Klassen an, ohne stark unter katastrophalem Vergessen zu leiden.

3.5.2.6.2.1.2 Semantische Segmentierung

Ziel ist die feingranulare Aufteilung der Eingabedaten in semantische Regionen, bei Bilddaten bedeutet dies idR. die pixelweise Klassifikation in (im Vergleich zur Klassifikation oben) wenige Klassen. In diesem Fall ist die Ausgabe zu einem gegebenen Eingabebild wieder ein Bild, in dem die Klassenzuordnung und optional die Konfidenz der Zuordnung in den Pixelwerten kodiert ist. Bezogen auf die Fahrzeug-Umfeldwahrnehmung im Straßenverkehr ist bspw. die im Cityscapes-Benchmark verwendete Unterteilung in 30 Klassen weit verbreitet, andere Benchmarks aus dem gleichen Anwendungsbereich verwenden hier oft ähnliche Klassen. Auch wenn das Verfahren zumeist auf Bilddaten angewendet wird, lassen sich prinzipiell auch andere Eingabedaten (bspw. LiDAR-Punktwolken) semantisch segmentieren. Hier existierte bereits zu Projektbeginn eine Vielzahl von realen und synthetischen Benchmark-Datensätzen aus dem Anwendungskontext Fahrzeug-Umfeldwahrnehmung (u.a. BDD, Cityscapes, A2D2, GTA5, Cynthia), und so wurde das Thema der semantischen Segmentierung im Projekt von vergleichsweise vielen Partnern untersucht.

Mit dem Ziel, eine exakte pixelweise Segmentierung durchzuführen, ist das Training von datengetriebenen Verfahren eine Herausforderung. Annotationen sind im Vergleich zur Annotation von Gesamtbild-Klassifikationen (z.B. für die Klassifikationsaufgabe) komplex, somit ist u.a. die manuelle Annotation sehr ressourcenintensiv. Teil- und unüberwachte Lernverfahren bieten damit einen Kosten- und Effizienzvorteil, zudem ermöglichen Verfahren zur Einbindung von annotierten synthetischen Daten das Training von Perzeptionssituationen, die in der Realität nur selten vorkommen oder nur mit hohem Aufwand bzw. Risiko erzeugbar sind. Im

verkehrlichen Kontext sind dies bspw. seltene, atypische und kritische Verkehrereignisse sowie die Erzeugung einer hohen Anzahl von Kombinationen von Sensor-, Verkehrs-, Umgebungs- oder Wettersituationen. Somit war es attraktiv, die gesamte Bandbreite an didaktischen Verfahren auf die Aufgabe der semantischen Segmentierung anzuwenden. Fortschritte wurden dabei in folgenden Bereichen erzielt:

- Teil- und unüberwachtes Lernen durch Domänenadaption (**DLR, CARIAD, TU Braunschweig, Universität Stuttgart**). Kernidee ist, den Trainingsprozess von einer annotierten Quelldomäne (Realdaten, synthetische Daten) auf eine nicht annotierte Zieldomäne (idR. Realdaten) zu übertragen, und Ähnlichkeiten zwischen beiden Domänen zur Automatisierung des Lernprozesses zu verwenden. Je nach Quell- und Zieldatensatz konnten dabei Genauigkeitsverbesserungen von ca. 10 bis 20 Prozentpunkten (Metrik mIoU) in der Zieldomäne erreicht werden. Theoretisch kann die Leistung in der Zieldomäne erreicht werden, die auch bei überwachtem Training erreichbar ist, dies erfordert jedoch entsprechende Überlappungen zwischen Quell- und Zieldomäne. Neben der Generalisierung für eine Adaptierbarkeit auf möglichst vielfältige Zieldomänen ist es somit ein Anknüpfungspunkt, optimale synthetische Daten für diese Aufgabe zu generieren (vgl. teilüberwachtes Lernen).
- In Ergänzung zu diesen Verfahren wurden auch Strategien zur Verbesserung des überwachten Lernens (**Mercedes-Benz**) sowie zum kontinuierlichen Lernen (**Porsche Engineering, Uni Stuttgart, ZF**) weiterentwickelt. Überwachtes Lernen kann als Grundlage für weitere aufsetzende Prozesse dienen (z.B. initiale Annotation neuer Quelldomänen), angegangen wurden Multi-Labeling Strategien um verrauschte und mehrdeutige Daten zu annotieren, und um Informationsverluste für nachfolgende Schritte zu vermeiden. Kontinuierliches Lernen zielte hier darauf ab, vergleichsweise kleine Änderungen (u.a. Wetter, Erweiterung einzelner Segmentierungsklassen) zu berücksichtigen und um das "Vergessen" von Inhalten zu verhindern bzw. gezielt zu steuern.
- Neben teilüberwachtem Lernen wurden Verfahren zum teilüberwachten Domänentransfer untersucht (**BUW**). Hierbei wurde ein Expertennetzwerk auf synthetischen Daten trainiert, um die darin enthaltene semantische Informationen abzugreifen. Mit Hilfe von neuronalen Netzen zur Bildanpassung werden speziell auf die Eingabedaten des Experten angepasste Bilder erlernt und erzeugt. Die Methode führt zu einer Steigerung der Leistung, es ist jedoch eine Herausforderung, bestehende Verzerrungen zwischen den Domänen (real zu synthetisch und umgekehrt) zu reduzieren.
- Im Bereich des aktiven Lernens wurde der Einfluss der Trainingsdatenauswahl auf das Trainingsergebnis untersucht und daraus Verfahren abgeleitet, wie die Bildauswahl möglichst optimiert werden kann, z.B. als Stichprobe für eine manuelle Annotation oder für automatisierte Verfahren, in denen häufige Ereignisse nicht überrepräsentiert werden sollen. Das **DLR** und die **Universität Freiburg** untersuchten Bewertungs- und Auswahlmethoden für typische Messkampagnendaten mit vielen redundanten Bildern, u.a. durch die Übertragung des BatchBALD Algorithmus auf die semantische Segmentierung. **BMW** verfolgte hier den Ansatz des superpixelbasierten aktiven Lernens, als einfache Vorsegmentierung der Bilddaten für nachfolgende Schritte. Hierdurch wird u.a. das Labeling vereinfacht, und es konnte mit am Beispiel des Cityscapes-Datensatzes

gezeigt werden, das eine um ca. 50% reduzierte Anzahl gelabelter Pixel (8% Superpixel vs. 20% reguläre Pixel) ein Modell mit gleicher Segmentierungsleistung erzeugen kann. Neben Bilddaten hat die **TU München** die Segmentierung von LiDAR-Punktwolken (nuScenes Datensatz) untersucht, insbesondere im Hinblick auf eine konsistente Segmentierung über die Zeit. Herausfordernd stellen sich hier dynamische Szenen dar, insbesondere wird hier ein robustes Matching von Punkten bewegter Objekte über die Zeit benötigt.

- Das Lernen bei veränderter Annotationsspezifikation wurde von **BMW** untersucht, am Beispiel eines Klassensplits zur Unterteilung u.a. der Klasse "Großes Fahrzeug" in "Bus" und "Truck" im Cityscapes Datensatz. Für alle untersuchten Split Szenarien wurde mit 50% neu gelabelten Daten eine ähnliche Performanz erreicht wie mit 100% neu gelabelten Daten. Das entwickelte Verfahren ist einfach und führt zu ähnlichen Ergebnissen wie das neue Labeln aller Bilder, bei entsprechend reduziertem Annotationsaufwand.
- Neben Bilddaten sind LiDAR-Daten und die Kombination beider Sensoren von Bedeutung. Für den Wissenstransfer von Kamerabildern zu LiDAR-Punktwolken wurden vom **DLR** Strategien zur Generierung von Pseudolabels entwickelt, die sich dafür eignen, komplexe semantische 3D-Objekte aus der Sensorkombination zu erfassen. Erzeugt wurden die Pseudolabels auf den Realdaten der Messkampagne.

Eine Weiterführung der semantischen Segmentierung sind Verfahren zur Instanzsegmentierung (siehe Abschnitt unten).

3.5.2.6.2.1.3 Tiefensegmentierung

Tiefensegmentierung ist insofern mit der semantischen Segmentierung vergleichbar, als dass eine pixelweise Zuweisung zu Bildinformationen erfolgt. Bei dieser Aufgabe werden hingegen die Bilder nach Tiefe segmentiert, und die Objektentfernung wird in den Pixelwerten des Ausgabebildes kodiert. Diese Ausgabe ist z.B. vergleichbar mit den Ausgaben einer Stereo-Tiefenschätzung (*depth image* oder *disparity map*), hier basierend auf maschinellen Lernverfahren, bei denen die Objektentfernung trainiert wurde. Vorteil gegenüber bspw. stereoskopischen Methoden ist, dass die Tiefensegmentierung auch bei Monokamerabildern funktioniert bzw. speziell für diese Art von Eingabedaten gedacht ist. Der Hardware- und Kalibrierungsaufwand ist im Vergleich zu stereoskopischen Methoden somit reduziert, ebenso reduzieren sich typische Fehlerquellen bspw. Auswirkungen auf das Kamerasystem aufgrund von Vibrationen. Es stellt sich somit die Frage, inwieweit die Leistungen von Tiefensegmentierungsverfahren mit den Leistungen anderer Verfahren vergleichbar sind und in welchen Fällen eine geringere Performanz zu erwarten ist. Eine mögliche Referenz für KI-basierte Tiefensegmentierung ist die von der **Universität Freiburg** weiterentwickelte Multi-View Depth Estimation mit einer eigenen Zusammenstellung von Benchmarkdaten (Robust MVD Benchmark).

Im Projekt hat **ZF** das Forschungsgebiet maßgeblich vorangetrieben. Die allgemeine Idee für teil- und unüberwachte Lernverfahren besteht darin, eine Folge von Bildern aus einer Sequenz zu verwenden, die Transformationsmatrix zwischen ihnen zu schätzen und diese Information als Überwachungssignal für die Tiefenschätzung zu verwenden. Evaluiert wurden Bildsequenzdaten verschieden komplexer realer Szenarien (KITTI, DDAD). Verbesserungen wurden insbesondere im Bereich texturarmer Regionen erzielt, bei denen klassische Tiefenschätzverfahren (Stereo/Multi-View, Structure from Motion usw.) typischerweise eine geringere Performanz aufweisen.

Als Weiterführung des Themas hat **BMW** die 3D Tiefensegmentierung im Hinblick auf eine damit ermöglichte 3D-Geschwindigkeitsschätzung aus Videodaten untersucht und mit dem KITTI-Datensatz erprobt. Um das Verfahren zur 3D-Geschwindigkeitsbestimmung von Objekten der Fahrzeugumgebung jedoch quantitativ bewerten zu können, sind jedoch Referenzdaten in ausreichender Qualität erforderlich. Diese sind in typischen Datensätzen (z.B. KITTI, aber auch Stereo-Benchmarks) nicht enthalten und sollen in zukünftigen Arbeiten erstellt werden.

3.5.2.6.2.1.4 Objektdetektion

Eine weitere typische Aufgabe im Rahmen der Fahrzeugumfeldwahrnehmung ist die Detektion von Objekten in den Sensordaten. Üblicherweise bedeutet dies zunächst die Erkennung von Objekten und deren Eigenschaften im Bildraum, also bspw. 2D-Objekte in Kamerabildern oder 3D-Objekte in LiDAR-Punktwolken. Zu erfassende Eigenschaften sind bspw. Umriss, geometrische Hülle (z.B. achsenparallele oder gedrehte Bounding Box) und die semantische Objektklassifikation. Oftmals in nachfolgenden Prozessschritten kommen weitere Punkte wie Tracking und Prädiktion oder die Transformation in fahrzeugrelative und geodätische Koordinaten hinzu. Es gibt eine Vielzahl von Spezialisierungen für bestimmte Anwendungsfelder, bspw. Erkennung von Fahrzeugen und deren Trajektorien, Erkennung und Kartierung von Verkehrsschildern und anderen Objekten, Kommunikation und Fusion von Objektdaten in verteilten Systemen. Verfahren zur Objektdetektion werden oftmals generisch untersucht und entwickelt, d.h. häufig auch anhand von Eingabedaten aus anderen Domänen (vgl. Klassifikation). In vielen Fällen wird die Objektdetektion als Erkennung von 2D-Boundingboxen der jeweiligen zu klassifizierenden Objekttypen im Bildraum verstanden.

Zusammen mit der semantischen Segmentierung war das Thema Objektdetektion ein im Projekt sehr umfangreich bearbeitetes Forschungsfeld, bei denen die hier beteiligten Projektpartner unterschiedliche Aspekte betrachtet und weiterentwickelt haben. Folgende Themen wurden untersucht:

- **Mercedes-Benz** betrachtete die Objekterkennung im globalen cross-Sensor Kontext und hat am Beispiel des nuScenes-Datensatzes für Kamerabilder untersucht, inwieweit Sensoreigenschaften und z.B. Kalibrierunsicherheiten die Erkennungsqualität beeinflussen. Hierbei haben Sensormodalitäten wie z.B. die Betrachtungshöhe einen signifikanten Einfluss (z.B. durch Verdeckungen) und es stellt sich die Herausforderung, adaptierbare Verfahren und Modelle für eine große Bandbreite von Kameras und Einbaupositionen zu entwickeln und zu trainieren. Das Potenzial der Generalisierung und Anwendbarkeit auf eine Vielzahl von Sensor- und Fahrzeugtypen ist sowohl für die zur Datenerhebung verwendeten unterschiedlichen Fahrzeuge von Bedeutung, als auch für den späteren ökonomischen Einsatz der Detektionssoftware.
- **BMW** untersuchte am Beispiel Objekterkennung Verfahren zum föderierten Lernen. Am Beispiel der Datensätze BDD und Nulmages erwies sich dies als einfache und effiziente Trainingsmethode wenn Daten auf zwei Datacenter verteilt sind. Allerdings erwies sich auch einfaches Transfer Learning als nicht viel schlechter. Potenzial bietet föderiertes Lernen auch bei Langzeitexperimenten, bei denen ein Versuch mit 100 statt (oft üblichen ca. 30) Epochen Vorteile gegenüber klassischen Trainingsmethoden zeigt. Dieser Effekt und dessen Generalisierbarkeit soll weiter untersucht werden.
- **Porsche Engineering** entwickelte Methoden zur Erhöhung der Robustheit gegenüber unbekanntem Objekttypen. Dabei wird die Hierarchie auf Basis der schon bekannten

Klassen und deren hierarchischen Beziehungen untereinander definiert. Das Klassifikationsmodell lernt hierbei die Objekte aus den unbekanntem Klassen in die ähnlichste Klasse der Hierarchie zu klassifizieren und stellt somit eine semantische Einordnung des unbekanntem Objekts zur Verfügung. Anhand der Datensätze nulmages und den synthetischen Daten aus AP 1.3 wurde gezeigt, dass die vorgestellten Methoden erfolgreich bekannte und unbekanntem Klassenobjekte in verschiedensten Verkehrsszenarien unterteilen und diese in eine hierarchische Klassenhierarchie zuordnen können. Herausfordernd sind reale Eingabedaten mit einer hohen Anzahl von Objekten.

- Die **Hochschule Reutlingen** befasste sich in diesem Themenbereich mit der Erfassung neuartiger Verkehrsobjekte, mit Fokus auf e-Scooter. Solche Fahrzeuge tauchen in älteren Datensätzen nicht auf und sind ein Beispiel für den Anwendungsfall Delta-Zeit. Durch inkrementelles Training konnte, unterstützt durch synthetische Daten, die Detektionsqualität (Präzision, mAP) von 0,52 auf 0,72 gesteigert werden, der Wert für die vortrainierte Klasse (hier: Detektion von Fußgängern) beträgt 0,82. Der Ansatz bietet somit die Chance, neue Verkehrsobjekte bzw. die Domäne Delta-Zeit, bei der vor allem kleine inkrementelle Änderungen der zu detektierenden Umgebung zu erwarten sind, unüberwacht zu lernen. Das Verfahren für die unüberwachte Domänenadaptation wurde zudem auf einen eigenen Datensatz angewendet, um eine Umgebungsanpassung auf degradierte Wetterbedingungen (Nebel, Regen, Schnee) vorzunehmen. Signifikant war vor allem die Verbesserung der Detektion bei Nebel.
- Die **Universität Freiburg** hat ebenfalls Verfahren zur Domänengeneralisierung untersucht, um die Detektion verschiedenen Wetterbedingungen zu verbessern. Die Maßnahme verbessert die Ergebnisse im Fall von ungesesehenen Wetterbedingungen (besonders Sand und Schnee) recht deutlich, ohne dass sich die Ergebnisse unter den vorab trainierten Bedingungen verschlechtern. Dies wurde anhand des DAWN Datensatzes untersucht. Insgesamt können somit unvorhergesehene Änderungen in der Domäne adressiert werden, d.h. dass für die unbekanntem Zieldomäne kein Training erforderlich ist. Hier ist zu erwarten, dass die Änderung nicht allzu groß werden darf, d.h. dieser allgemeine Ansatz kann die Domain Gap nicht vollständig schließen. Bei vorhersehbaren Änderungen der Umgebungsbedingungen ist eine (zusätzliche) explizite Domain Adaptation auf Zieldaten empfehlenswert.
- **ZF** hat Methoden entwickelt, um Erkennungsmethoden auf verschiedene Sichtperspektiven hin zu generalisieren (Cross-Sensor-Anpassung). Hierzu wurden Gradient Reversal Layer (GRL) für die perspektivenübergreifende Anpassung von Objekterkennungsmodellen von EfficientDet verwendet. Hierzu wurde ein eigener Datensatz und der Argoverse Datensatz verwendet, in denen verschiedene Sensorpositionen vorkommen, relevant sind insbesondere die Höhe und Ausrichtung. Nach aktuellem Stand sind Datensätze mit vielen Sichtperspektiven selten, und so bietet der im späteren Projektverlauf aufgenommene Datensatz, welcher synchrone Daten aus unterschiedlichen Kameraperspektiven bietet, einen Ansatzpunkt für Folgeuntersuchungen. Bei den untersuchten Daten wurde eine direkte Korrelation zwischen der Leistung des Modells und der Größe der Boundingboxen von Objekten derselben Klasse festgestellt. Ebenso wurde die Auswirkung von ungleichmäßig über die Domänen verteilten Objektklassen untersucht. Größere Modelle zeigten jedoch unabhängig von der Objektgröße eine bessere Leistung.

- Hierzu verwandt hat **Mercedes-Benz** anhand des nuScenes Datensatzes das Thema Robustheit gegenüber Sensorparametern (hier: verschiedene Kameras und Optiken) bearbeitet. Es zeigt sich als offene Herausforderung, generische bzw. weitestgehend sensorinvariante Verfahren, zu entwickeln.

3.5.2.6.2.1.5 Instanzsegmentierung

Instanzsegmentierung kann als Erweiterung semantischer Segmentierung in Kombination mit Objekterkennung verstanden werden. Das übliche Ziel ist die pixelweise Segmentierung von Kamerabilddaten in Regionen, die einzelnen Objekten zuzuordnen sind. Die Regionen bilden dabei nicht nur Objektklassen, sondern ermöglichen eine Unterscheidung der segmentierten Objekte. Die Aufgabe ist daher als erheblich schwieriger als die reine semantische Segmentierung anzusehen, bietet jedoch das Potenzial der sehr exakten Objekterkennung und und robusten Verfolgung über die Zeit.

Am Beispiel der Datensätze Synscapes und Cityscapes hat **Mercedes-Benz** Verfahren zur Domänenadaptation erforscht. Dabei konnte in einem nur spärlich annotierten Zieldatensatz die Erkennungsleistung erhöht werden, basierend auf trainierten Realdaten und synthetischen Daten als Quelldomäne. Mit 32% Präzision (mAP) sind die Ergebnisse vielversprechend, jedoch noch weit entfernt von der Erkennungsleistung auf einem vollständig trainierten Datensatz. Hier sind umfassende weitere Forschungen erforderlich und sinnvoll.

3.5.2.6.2.1.6 Posen- und Verhaltenserkennung und Prädiktion

Die Erkennung von spezifischeren Objekteigenschaften und die Prädiktion von Bewegungen und Verhaltensmustern kann dazu dienen, nichtlineare und sehr dynamische Bewegungen bestimmter Verkehrsteilnehmer zu erkennen und frühzeitig darauf zu reagieren. Hier kann es bspw. nicht nur darum gehen, Objekte als Ganzes zu erfassen, sondern bestimmte Bewegungen, Gesten, Verhaltensweisen und deren Intention zu erkennen. Das Thema ist an sich mit der Objekterkennung verwandt, teilt Objekte wie Verkehrsteilnehmer in separat zu detektierende Muster auf und ermöglicht daher die Erkennung komplexerer Sachzusammenhänge. Je nach Objektklasse sind spezifischere Einteilungen und Mustererkennungsverfahren erforderlich, die für das jeweilige Objekt typisch sind und objektindividuelle Schlüsse zulassen.

In diesem Bereich beschäftigte sich die **Hochschule Reutlingen** mit der Erkennung von menschlichen Posen, d.h. Bewegungen von z.B. Kopf, Beinen und Armen und deren Bedeutung. Dies kann als Grundlage für die frühzeitige Prädiktion von Verhaltensmustern von Fußgängern im Verkehr dienen, was mit klassischem Objekttracking aufgrund der sehr dynamischen Verhaltensweisen nur sehr unzureichend gelingt. Hierzu wurden zumeist an eigenen Datensätzen Labels für reale und synthetische Daten erzeugt, wobei die zu erkennenden Klassen wesentlich vielfältiger sind (d.h. einzelne Körperteile usw.) als bei den oben genannten Objekterkennungsverfahren. Basierend wurden die bereits genannten Verfahren angewendet und für diesen spezifischen Anwendungsfall weiterentwickelt, insbesondere zur Domänenadaptation (Cross-Sensor-Anpassung und Umgebungs-/Wetteranpassung), teilüberwachtes und unüberwachtes Lernen, Trainingsorganisation für Segmentierungs- und Objekterkennungsverfahren und aktives Lernen. Ein weiterer Schwerpunkt war die Erkennung von Anomalien, wobei die Herausforderung ist, atypische Verhaltensweisen in einen entsprechenden Kontext (u.a. Verkehrssituation) setzen zu können, was bei den üblicherweise verfügbaren Daten und Simulationen nur unzureichend gelingt. Wichtig sind hier längere

Aufzeichnungen von Sequenzen, da viele Verhaltensmuster nur im Zeitverlauf und nicht z.B. im Einzelbild erkennbar werden.

Ebenso beschäftigte sich **Mercedes-Benz** mit der Erkennung von menschlichen Posen und atypischem Verhalten. Besonders hervorzuheben ist der erstellte R-U-MAAD Datensatz, als Benchmark für die unüberwachte Anomaliedetektion in Trajektorien mit mehreren Agenten und mit dem Ziel, die Erkennung normalen Fahrverhaltens zu trainieren und Anomalien erkennen zu können. Herausfordernd ist hierbei, weiteres Wissen, u.a. aus Kartendaten, mit einzubeziehen. Zudem wurden Methoden zur videobasierten Prädiktion menschlicher 3D-Posen entwickelt und anhand des Human3.6M-Datensatzes evaluiert. Die Erkennungsgenauigkeit (Positionsgenauigkeit) unter Testumgebungsbedingungen und unter Verwendung von Pseudo-3D-Labels aus Multikamerasystemen ist bei unüberwachtem Lernen vergleichbar mit der Genauigkeit, wenn überwachte Verfahren, d.h. mit manuell annotierten Trainingsdaten, eingesetzt werden. Eine offene Herausforderung ist hierbei die Anwendung auf realistische Verkehrsszenen und bei bewegter Kamera. Entwicklungen zu LiDAR-basierten Verfahren zur Trajektorienprädiktion von Fahrzeugen wurden mit dem Argoverse-Datensatz evaluiert, hier zeigen sich Fortschritte bei Prädiktionsverfahren, die die Bewegungsmodelle mitlernen und dabei keine hochauflösenden Karteninformationen benötigen.

3.5.2.6.2.2 Bewertungs- und Optimierungsmethoden

Dieser Abschnitt umfasst die Entwicklungen im Bereich von Methoden zur Bewertung, Vorhersage und Optimierung. Die zu lösende Aufgabe (hier idR. eine der o.g. Perzeptionsaufgaben) wird dabei eher generisch betrachtet, Fokus sind hier die Methoden, welche die Daten bzw. Datenauswahl und die Systemanforderungen von KI-Verfahren thematisieren.

3.5.2.6.2.2.1 Bewertung von Eingabedaten

Ziel ist hierbei die Bewertung von Eingabedaten wie Bildern und anderen Sensordaten, sowie nach Bedarf abgeleiteten Daten wie Trajektorien, auf Ihre Eignung für Trainingsverfahren. Hierdurch soll die Auswahl der Datenstichprobe optimiert werden, die für das KI-Training den bestmöglichen Nutzen hat oder notwendiges (idR. aufwändiges) manuelles Labeling minimiert. Eine derartige Bewertung ist eine typische Teilkomponente des aktiven und kontinuierlichen Lernens, kann jedoch auch bei anderen Lernverfahren zum Einsatz kommen.

Hierzu hat **Porsche Engineering** eine Methode für eine verbesserte Datenauswahl für Replay-Methoden im kontinuierlichen Lernen entwickelt. Ziel ist es, mit einer durch entsprechende Bewertung gezielten Bildauswahl bessere und stabilere Ergebnisse zu erzeugen, als dies mit einer zufälligen Stichprobe der Fall ist. Am Beispiel der semantischen Segmentierung mit Cityscapes und BDD konnte die Performanz deutlich gesteigert werden (klassen-inkrementell: von 55% auf 60% mIoU, domänen-inkrementell: von 48% auf 57% mIoU), allerdings treten Effekte des katastrophalen Vergessens auf. Um dies zu reduzieren, wurde analysiert, wie sich katastrophales Vergessen im Klasse-inkrementellen Lernen verhält und anhand von geeigneten Daten (hier: PascalVoc) untersucht. Insbesondere wurde mithilfe von Layer Stitch bestimmt, dass wichtige Merkmale für alte Klassen, die vergessen worden sind, nicht vollständig überschrieben werden, sondern stattdessen im Klassifikationslayer neuen Klassen oder dem Hintergrund zugewiesen werden. Weiterhin wurde mithilfe von einer speziellen Fehlerfunktion, die die Hintergrundklasse speziell berücksichtigt, katastrophales Vergessen weiter reduziert. Anhand der gefunden Ursachen und Effekte für katastrophales Vergessen müssen weiterhin Verfahren entwickelt werden, die diese Effekte weiter vermindern können.

Ein weiterer von Porsche Engineering untersuchter Punkt ist die systematische Analyse der möglichen Beeinträchtigungen einer Bilddatenanonymisierung, wie sie aus Datenschutzgründen oftmals angewendet wird. Der Performanceverlust wurde anhand von Gesichtsdaten qualitativ, quantitativ als auch mit separaten statistischen Experimenten untersucht. Es hat sich gezeigt, dass kleine neuronale Netze und das Fehlen von Multiskalenmerkmalen in dem Netz größere Einflüsse durch die Anonymisierung haben, was durch statistische Tests nachgewiesen werden kann. Die Verwendung generativer Modelle zur Regenerierung biometrischer Merkmale können die Nebenwirkungen der Datenanonymisierung abschwächen, da generative Modelle die Verteilung von den Daten während des Trainingsprozesses gelernt haben. Offen ist die Reduzierung der benötigten Ressourcen für die Wiedererstellung der wichtigen Gesichtsmerkmale, sowie eine generische Untersuchung und der Vergleich mit weiteren, für Verkehrsszenen typischen, Anonymisierungen (bspw. Kfz-Kennzeichen).

Das FZI entwickelte Methoden zur Auswahl von Trajektorien und zur Vorhersage von besonders interessanten Trajektorien anhand von mit CARLA simulierten Daten. Die neuen Methoden erlauben eine Bewertung ganzer Trajektorien statt einzelner Transitionen für den Lernprozess und die Vorhersage von Trajektorien mit überdurchschnittlichem Informationsgehalt bzw. Lernwert. Durch die Trajektorienauswahl bieten sich Vorteile bei der Recheneffizienz durch entsprechende Priorisierung im Lernprozess, jedoch keine Vorteile bei der Dateneffizienz. Zudem sind weiterführende Integrationen offen, bspw. Vorhersagen in Koordinaten, um mit einem Navigationssystem zu integrieren.

3.5.2.6.2.2 Optimierung des Lernprozesses

Aufgabenspezifische Aspekte sind im Rahmen der einzelnen Methoden (vgl. Abschnitt Perzeptionsmethoden) betrachtet worden. Eine generische Lösung im Bereich des modellbasierten Reinforcement Learning wurde vom FZI mit Hilfe der DeepMind Control Suite untersucht. Hierzu wurden zum einen verschiedene Strategien untersucht, um das Training effizienter zu gestalten, indem sich auf solche Daten fokussiert wird, die maximalen Nutzen für das Lernen der Parameter der neuronalen Modelle haben (vgl. Abschnitt zur Bewertung von Eingabedaten). Zum anderen wurde untersucht, inwiefern Strategien zum kontinuierlichen Lernen in einen modellbasierten Reinforcement Learning Ansatz integriert werden können und was die Schwierigkeiten dabei sind. Kern der Optimierung ist die Wiederverwendung bereits gelernter Inhalte auch auf anderen Datensätzen, wobei Strategien zur Reduzierung des Vergessens angewandt werden. Ein vielversprechender Ansatz ist Elastic Weight Consolidation (EWC), bei dem Lerninhalte unterschiedlich gewichtet werden, je nach Erfolg und Relevanz in vorhergehenden Iterationsschritten. Gegenüber bisherigen Ansätzen konnte gezeigt werden, dass die Kombination mit EWC durchaus das Vergessen vermindert hat, also zu einem höheren Average Return auf bereits gesehenen Tasks geführt hat, auch wenn schon auf weiteren Tasks trainiert wurde. Hierbei ließ sich feststellen, dass die einzelnen Komponenten des Agenten eine sehr unterschiedliche Wichtigkeit bzw. Rate des Vergessens haben. Dadurch, dass ein modellbasierter RL Agent relativ komplex ist, da er aus vielen einzelnen neuronalen Komponenten besteht, müssen Interaktionen und Verschiebung der internen Repräsentationen bei der Nutzung von EWC berücksichtigt werden. Zwar zeigte sich eine starke relative Verbesserung durch EWC, allerdings sind die absoluten Ergebnisse immer noch zu weit davon entfernt, dass dieser Ansatz tatsächlich das Vergessen bereits gelerntes Wissen effektiv verhindert.

3.5.2.6.2.2.3 Abschätzung und Vorhersage von Laufzeit und Speicherplatz

Als vorbereitende Maßnahme zur Laufzeit- und Speicherplatzoptimierung (siehe folgender Abschnitt) hat das DLR untersucht, inwieweit zu optimierende Parameter bereits während des Trainings vorhersagbar sind. Ergebnis anhand der Untersuchungen mit dem Tensorflow CNN repository ist, dass die Laufzeit der KI schon während des Trainings schnell und genau vorhergesagt werden kann. Eine Modellausführung dauert selbst für große und komplexe Systeme weniger als eine Sekunde. Die Vorhersagegenauigkeit der Laufzeit ist üblicherweise bei etwa unter 10% Standardabweichung. In Einzelfällen kann die Abweichung höher liegen, dennoch ist die Abschätzung neu und bedeutsam für die Bewertung von Methoden und die Auswahl einer geeigneten Zielhardware für eine vorgegebene KI. Insgesamt wurden drei verschiedene Hardwareplattformen untersucht, zukünftig sollte die Methode für eine breite Auswahl an Hardwarekomponenten durchgeführt werden.

Erschwerend können Programmierfehler in der Synthesoftware der untersuchten KI Systeme sein, sie bewirken teilweise sehr hohe, unvorhersagbare Laufzeiten. In diesen Fällen bewirkt eine Vergrößerung des Systems eine Verringerung der Laufzeit. Der Nutzer sollte von unseren Methoden Hinweise bekommen können, welche Metaparameter der KI sich für die Laufzeitbewertung eignen.

3.5.2.6.2.2.4 Laufzeit- und Speicherplatzoptimierung

Typischerweise haben KI-Verfahren im Entwicklungsstadium sehr hohe Hardwareanforderungen (GPU-Cluster, hoher Speicherplatz, usw.) und sind noch nicht im Hinblick auf eine bestimmte Zielhardware hin optimiert. Für höhere Technologiereifegrade werden Verfahren entwickelt, um KI-basierte Prozesse mit möglichst wenig Einbußen in der Ausgabedatenqualität bzw. Perzeptionsleistung zu beschleunigen und den Speicherbedarf zu reduzieren. Die Implementierung für eine konkrete Zielhardware steht dabei jedoch nicht im Vordergrund, sondern die durch Optimierung erreichte mögliche Anwendbarmachung für kleinere Zielsysteme mit entsprechend begrenzten Ressourcen. Darüber hinaus ist die Reduktion der Rechenzeit für das Erreichen von Echtzeitanforderungen ein Ziel.

Zusammen mit den oben genannten Abschätzungsuntersuchungen hat das DLR agnostische Verfahren zur Kompression (Pruning) von neuronalen Netzen entwickelt. Unabhängig von der Trainingsstrategie, und anwendbar für beliebige Tasks, kann jede einzelne Schicht eines neuronalen Netzes so transformiert werden, dass die Größe der Tensoren und die Anzahl der benötigten Rechenoperationen stark reduziert werden kann. Der Tradeoff zwischen Genauigkeit und Laufzeit der KI ist hierbei frei wählbar. Es konnten in der Evaluation Lösungen gefunden werden, die die Größe der Tensoren nur um knapp über 10% reduzieren, dafür aber die Genauigkeit der KI sogar minimal steigern. Auf der anderen Seite konnte die Größe um den Faktor 30 reduziert werden wodurch die Genauigkeit um etwa 2% reduziert wurde. Im Kompressionsbereich x3 bis x10 ist der Ansatz der Network Architecture Search (NAS) um einige Promill-punkte überlegen. Bei geringerer Kompression als Faktor 3 findet das NAS üblicherweise geringfügig bessere Lösungen, bei höherer Kompression als Faktor 9 überwiegt der statische Overhead der Methode, so dass NAS hier teils deutlich bessere Ergebnisse liefert. Gegenüber heuristischer Tensorkompression ist der Ansatz bei gleicher Kompression um etwa 0.3% Genauigkeit besser. Der eigene Ansatz ist derzeit etwa viermal schneller als das NAS System, die Methode hat aber Potential noch wesentlich schneller zu laufen. Hier muss die Transformation (Tucker decomposition) bessere, dem Originallayer ähnlichere Versionen liefern. Der Overhead

bei sehr hohen Kompressionsraten kann durch eine Kombination mit Filter Pruning und der Einführung einer eindimensionalen Tucker-Dekomposition noch reduziert werden.

ZF hat hierzu die Interpretable Pruning (IP) Methode entwickelt und implementiert. Interpretable Pruning benutzt Heatmaps generiert durch Deep Taylor Decomposition (hier spezifisch adaptiert für die Aufgabe der Objektdetektion) um zu entscheiden, welche der Filter gelöscht werden können. Dabei wird ein Ranking anhand eines Bruchteils der Trainingsdaten erstellt und die jeweiligen Filter Kernel anschließend entfernt. Die Ergebnisse sind in erster Linie vielversprechend, haben allerdings noch Verbesserungspotential. Es kommt ab einer bestimmten Kompressionsrate zu einem ungewöhnlich starken Abfall der Genauigkeit.

Der Fokus der bei **BMW** weiterentwickelten Pruning-Verfahren liegt auf der Berücksichtigung der Hardware. Dieses Hardware/Software Co-Design kann somit Verfahren für eine bestimmte Zielhardware hin optimieren

Porsche Engineering und **Mercedes-Benz** entwickelten Verfahren im Bereich der Wissensdestillation (Knowledge Distillation) weiter, bei denen KI-Modelle in reduzierte Modelle überführt werden und dadurch Laufzeit und Speicherplatz sparen. Hier wurde die Detektionsleistung eines Student Networks im Bereich der untersuchten Datensätze und Aufgaben (semantische Segmentierung mit PascalVoc bzw. Objektdetektion mit MS COCO) gegenüber bisherigen Methoden verbessert. Offen ist hierbei die Generalisierung und Skalierbarkeit auf andere Perzeptionsaufgaben.

Die **Universität Wuppertal** hat für ihr Klassifikationssystem (siehe Abschnitt xy) Verfahren zur Reduktion von Überparameterisierung Neuronaler Netze durch eine strukturierte Verringerung der Gewichte entwickelt. Die Ergebnisse zeigen, dass die Architektur einen verbesserten Parameter-Accuracy Trade-off erzielt. Insbesondere für Probleme, die weniger Kapazität brauchen, lässt sich die Genauigkeit bei drastischer Gewichtsverringerng verbessern. Bisher lassen sich diese Beobachtungen auf größere Probleme generalisieren, dies soll aber noch für weitere, größere, Probleme untersucht werden. Ebenso soll die Architektur auch als Backbone für Segmentierungs- und Detektionsnetze eingesetzt werden.

3.5.2.6.2.3 Synthese- und Simulationsmethoden

Dieser Abschnitt umfasst die Entwicklungen im Bereich von Methoden zur Synthese und Simulation und zur Optimierung von Eingabedaten als Schritte, die in der Anwendung den Lern- und Perzeptionsprozessen vorgelagert sind. Der Abschnitt fokussiert auf Themen die über generische Optimierungen hinaus gehen und spezifisch für weitere in diesem Projekt angewendeten Schritte sind.

3.5.2.6.2.3.1 LiDAR Upsampling

BMW hat ein Verfahren zur Erzeugung dichter LiDAR-Punktwolken entwickelt. Ziel ist die verbesserte Nutzarmachung von Punktwolkendaten für nachfolgende KI-Methoden unter Berücksichtigung von Unregelmäßigkeiten, die für LiDAR-Punktwolkendaten typisch sind. Letztendlich wurde auch hier ein KI-basiertes Verfahren eingesetzt, um bestimmte Features aus der Punktwolke zu extrahieren, zu erweitern, und in eine dichtere Punktwolke zu überführen.

3.5.2.6.2.3.2 3D Bildsynthese

Am **Max-Planck-Campus der Universität Tübingen** lag der Fokus im Projekt auf den Themenbereichen Datenaugmentierung mit generativen 3D Modellen sowie Robuste

Repräsentationen für die sensomotorische Regelung (nachfolgender Abschnitt). Im ersten Themenbereich wurde ein datenbasierter Ansatz zur Synthese und Analyse von 3D-Szenen entwickelt sowie geeignete 3D Repräsentationen zur Projektion in neue Sensormodalitäten (z.B. Lidar) untersucht. 3D-bewusste Bildsynthese bezeichnet die Synthese von 3D-konsistenten Bildern mit expliziter Kontrolle über die Kameraposition. Wichtig ist, dass lediglich Bilder und keine 3D-Modelle als Trainingsdaten benötigt werden. Hierzu wurden neue Methoden für die 3D-bewusste Bildsynthese entwickelt, welche eine hohe Bildauflösung ermöglichen und dabei rechnerisch effizient arbeiten. Diese wurden auf verschiedenen Datensätzen validiert (CARLA, FFHQ, FFHQ) und in neue Sensormodalitäten projiziert, z.B. Tiefenkarten. Die Ergebnisse zeigen das Potential von 3D-bewusster Bildsynthese für die Datenaugmentierung sowie zum Erstellen von 3D Repräsentationen, die in neue Sensormodalitäten projiziert werden können. Ein möglicher zukünftiger Anknüpfungspunkt ist die weitere Entwicklung hin zur Synthese von gesamten Szenen.

3.5.2.6.2.3.3 Imitation Learning

Ziel der diesbezüglichen Forschung am **Max-Planck-Campus der Universität Tübingen** ist unter anderem das automatisierte Lernen von automatisiertem Fahrverhalten in einer Closed-Loop Simulation durch Erkennung von Wechselwirkungen zwischen Perception und Fahrzeugführung/-regelung. Für den CARLA-Simulator wurde eine Methode entwickelt, die es ermöglicht, mehrere Sensoren zu fusionieren. Dadurch wird eine multimodale Eingabe ermöglicht, anstatt nur Kamerabilder zu verwenden. Des Weiteren wurde ein auf Transformer basierender Ansatz zur besseren Pfadplanung entwickelt. Bei beiden Methoden ist eine Erklärbarkeit der Entscheidungen durch die Visualisierung der "Attention" der Transformerarchitektur möglich. Die Ansätze wurden mittels eines Simulators im geschlossenen Regelkreis getestet. Die Closed-Loop Evaluation im Bereich des automatisierten Fahren ermöglicht eine realitätsnahe Evaluation, da offline-Metriken oft nicht mit der Online-Performance korrelieren. Ein Nachteil ist das Antrainieren des Fahrverhaltens durch menschliches Fahren im Simulator. Der Experte, der in unseren Imitation-Learning-basierten Trainingsstrategien verwendet wird, erreicht keine perfekte Punktzahl und weist bestimmte konsistente Fehler auf. Eine menschliche Datenerfassung, um dies zu beheben, wäre sehr zeitaufwändig. Verfahren des Imitation Learnings sind prinzipiell auch für Closed-Loop-Experimente an Realfahrzeugen geeignet, dies ist jedoch im Vergleich zur Simulation wesentlich anspruchsvoller und aufwändiger.

3.5.2.6.2.3.4 Adversarial Training

Ziel der Arbeiten von **Mercedes-Benz** ist die Generierung sicherheitskritischer Szenarien, bei denen aktuelle Systeme (z.B. Perzeptionsverfahren) versagen. Die Resultate sind Verbesserungen im Vergleich zu früheren Arbeiten, welche eher eine generische Black-Box-Optimierung waren. Hier wurden mit der CARLA-Simulation kritische Ereignisse gezielt erzeugt und angewendet, und die Kollisionsrate eines simulierten Fahrzeuges gezielt verringert. Der Ansatz funktioniert in den untersuchten Einzelszenarien, allerdings ist eine umfassende Optimierung sehr schwierig, da die Vielfältigkeit und Komplexität möglicher Szenarien sehr hoch ist. Automatisiert erzeugte Szenarien sind also möglicherweise häufig unrealistisch und wenig für weitere Tests geeignet, was eine weitere Auswahl notwendig macht. Zukünftige Arbeiten könnten hier ansetzen, in dem der Suchraum zur Erzeugung sinnvoller Szenarien eingegrenzt wird.

3.5.2.6.3 Gegenüberstellung der vorgegebenen Ziele

Im Hinblick auf die Vorhabenbeschreibung haben wir uns im Projektkonsortium dafür entschieden, die Evaluation innerhalb der einzelnen Arbeitspakete vorzunehmen. Die

ausführlichen Darstellungen der einzelnen Partner sind daher in den entsprechenden Arbeitspaketen aufgeführt und werden dort zusammengefasst. Das DLR hat sich als AP-Leitung in Abstimmung mit den Projektpartnern daher die Aufgabe gegeben, die Ergebnisse zusammenzutragen und aus einer AP-unabhängigen Sichtperspektive zusammenzuführen.

3.5.3 Zielerreichung

Im Teilprojekt 5 Evaluation wurden die in der VHB beschriebenen Ziele erreicht. Die eng an die zeitliche Projektstruktur angelehnte Zeitschiene wurde an die Änderungen der Gesamtprojektplanung angepasst. Ausgehend vom internen Workflow wurden die Ergebnisse somit fristgerecht produziert und bereitgestellt.

Die von **AP5.1 Review und Refinement** durchgeführten Treffen fanden je nach Reisemöglichkeiten physisch oder virtuell statt. Nachdem die Ergebnisse des jeweils abgelaufenen Projektinkrements vorgestellt und diskutiert wurden, konnte die APs dann auf dieser Grundlage die Arbeiten der kommende PI-Phase konkret planen.

Die State-of-the-Art-Sammlungen und die ebenfalls in **AP5.2 Baseline und Training der Referenzmodelle** verorteten Defizitanalysen wurden durchgeführt und dokumentiert.

Hauptergebnis von **AP5.3 Evaluation und Demonstration** ist ein abschließender Bericht zu den drei Themen Transfer Learning, Didaktik und Automotive Tauglichkeit. Dieser ist in deutscher Sprache als Kapitel 3.5.2.3 Teil des vorliegenden Berichts, zusätzlich ist eine englische Version erstellt worden, die als Ergebnis E4.1.5.3a-c Bestandteil des Deliverable 15 "Eine Validierung mit vollständigem Datensatz des Delta Learning-Projekts ist für die entwickelten Methoden des Delta Learning vorhanden" ist.

3.6 TP6: Projektorganisation

3.6.1 Ausgangssituation und Zielsetzung

In einem gemeinsamen Forschungsprojekt mit einer hohen Zahl an Beteiligten, die verteilt und vernetzt arbeiten, ist die Koordination und Dokumentation der Arbeiten sowie des Arbeitsfortschritts notwendig. Die Kommunikation unter den Beteiligten und die Sicherung der angestrebten Qualität der Projektergebnisse ist essentiell, um die Projektziele zu erreichen. Zu diesem Zweck mussten geeignete Maßnahmen des technischen und operativen Managements etabliert und gepflegt werden.

Als Projektvorschlag der VDA LI hatte das Vorhaben KI Delta Learning den Anspruch einer kohärenten und konvergenten Technologieentwicklung und einer schnellen industriellen Verwertung. Unter dieser Maßgabe sollte das Projekt Wissen und Ergebnisse für weitere Projekte bereitstellen und auch von anderen Projekten des Innovationsfeldes „Künstliche Intelligenz und maschinelles Lernen im automobilen Umfeld“ (KI-Familie) erhalten. Um die intendierten Synergiepotentiale zu heben wurde eine Zusammenarbeit mit dem bereits laufendem Projekt KI Absicherung verfolgt, wie auch mit den zeitgleich bzw. später gestarteten Projekte der KI-Familie, KI Data Tooling und KI Wissen.

Als öffentlich gefördertes Projekt in der vorwettbewerblichen Forschung war es für KI Delta Learning von großer Bedeutung die Ergebnisse in der Fachöffentlichkeit zu verbreiten, damit die erarbeiteten Maßnahmen und Methoden weiter diskutiert werden und so zu anerkannten Bausteinen auf dem Weg zum übergeordneten Ziel, dem autonomen Fahren werden können.

Deswegen war die zielgruppengerechte Aufbereitung der Ergebnisse und das Ausspielen über die verschiedenen Kanäle in einem eigenen Arbeitspaket verankert.

3.6.2 Strukturierung

Teilprojekt 6 bestand aus 3 Arbeitspaketen:

- AP6.1: Management und Koordination
- AP6.2: Austausch mit anderen Projekten
- AP6.3: Ergebnisverbreitung

3.6.2.1 AP6.1 Management und Koordination

3.6.2.1.1 Zielsetzung

Hauptaufgabe von AP6.1 war es, sicherzustellen, dass die übergeordneten Projektziele innerhalb des vereinbarten Budget- und Kostenrahmens und in der erforderlichen Qualität erreicht werden. Hierzu wurden die Arbeiten der einzelnen Arbeitspakete inhaltlich abgestimmt und nachverfolgt, um den Arbeitsfortschritt hinsichtlich der definierten Meilensteine sicherzustellen. Der technische Arbeitsfortschritt wurde kontinuierlich überwacht und verfolgt. Entsprechend der zugrunde gelegten Methode der inkrementellen Projektdurchführung erfolgte eine umfassende Fortschrittsbewertung jeweils zum Abschluss der definierten vier Prozessschritte, der Projektinkremente I bis IV. Bei inhaltlichen Konfliktsituationen und Planabweichungen wurden bereits im laufendem Arbeitsprozess und grundsätzlich zu den vier Gate-Terminen Vorschläge für adäquate Gegenmaßnahmen und/oder Planänderungen erarbeitet und den Steuergremien des Projektes zur Entscheidung vorgelegt.

Die angestrebte Qualität wurde durch Aufsetzen von Prozessen und Durchführung des Qualitätsmanagements gesichert, die auf das inkrementelle Vorgehen abgestellt sind. Hier standen insbesondere die für den jeweiligen Prozessschritt definierten Gütekriterien und Reifegrade im Mittelpunkt. Das Risikomanagement wurde implementiert, Risiken wurden regelmäßig überprüft, bewertet und Mitigationsmechanismen entwickelt, sodass zu den vier Gate-Terminen die erforderliche Transparenz über Projektrisiken und damit die Voraussetzung für strategische Weichenstellungen gegeben war. Die Durchführung des technischen Qualitätsmanagements, die Kommunikation innerhalb des Konsortiums wurde sichergestellt und der Arbeitsfortschritt wurde dokumentiert. AP6.1 hat außerdem die Berichtspflichten gegenüber dem Fördermittelgeber und Steuergremien sichergestellt.

Diese Aufgaben werden durch die Teilprojektleiter-Runde mit Unterstützung des Projektmanagement-Dienstleisters EICT (GUA1) wahrgenommen.

Der aktuelle Stand der Arbeiten und Ergebnisse, sowie im Falle von inhaltlichen oder zeitlichen Abweichungen vorgeschlagene Gegenmaßnahmen wurden periodisch für den Steuerkreis und den Fördermittelgeber aufbereitet und vorgestellt.

Der Austausch in der Teilprojektleiter-Runde hat bis zur Covid-19-Pandemie in Form von physischen und virtuellen Projekttreffen (Web-/Telefonkonferenzen) stattgefunden, danach wurde auf eine rein virtuelle Durchführung umgestellt. Die Organisation und Koordination dieser Projekttreffen umfasste Aufgaben der Terminplanung und -abstimmung, die Unterstützung der

inhaltlichen Vorbereitung der Treffen (Erstellen einer Agenda) sowie deren Nachbereitung (Erstellen von Protokollen inklusive Formulierung von Aufgaben sowie deren Nachverfolgung).

Die gesamte Ergebnisdokumentation, die Erstellung von Berichten sowie die Delegation von Aufgaben - innerhalb der Teilprojektleiter-Runde sowie für das gesamte Konsortium - wurde technisch über eine Projektmanagement-Plattform, konkret über eine Confluence-Wiki-Plattform ermöglicht.

3.6.2.1.2 Ergebnisse im Einzelnen

Unter der Leitung des Konsortialführers Mercedes-Benz wurden in AP6.1 die verschiedenen Aufgaben bearbeitet, die im Rahmen einer effizienten und effektiven Projektleitung angefallen sind. Dazu beigetragen haben neben dem Koordinator vor allem die Teilprojektleiter und der GUA1-Auftragnehmer EICT GmbH.

Um die Zusammenarbeit zwischen den Partnern bzw. deren Mitarbeitern zu erleichtern wurde die Confluence-**Projektplattform** zur Verfügung gestellt. Confluence stellt im Grundprinzip eine Wiki-Plattform dar, bei der die Teilnehmer Seiten erstellen und bearbeiten können. Durch eine Vielzahl von zusätzlichen Apps können auch Funktionalitäten weit über ein reines Wiki-/Blogsystem hinaus realisiert werden. So wurde zum Beispiel auf Wunsch aus den Projekten ein Formeditor hinzugefügt, eine App für Literaturverzeichnisse und ein Graphikmodul. Mittels einer weiteren App können Seiten mit Unterseiten zu umfangreichen Dokumenten zusammengefügt und als Worddokument ausgespielt werden; dies erlaubt eine Nachbearbeitung in Word um das Layout weiter anpassen zu können. Auch der hier vorliegende Bericht wurde auf diese Weise produziert.

Nach einer Anpassung an die projektspezifischen Corporate Identity wurde eine transparente Struktur aufgesetzt, die den verschiedenen Bedarfen innerhalb der Teilprojekte, aber auch den Erfordernissen projektübergreifender Themen oder GUA-Nehmer gerecht wird. Bis zum Projektende wurden so über 3.400 Confluence-Seiten erstellt, teils durch das Projektmanagement aber auch durch Mitarbeiter aus den Teilprojekten und den Arbeitspaketen. Eine durchgehende Aufgabe für AP6.1 bestand im Anlegen und Verwalten der Logins, deren Anzahl am Ende über 150 betragen hat.

Um eine zielgenaue und effiziente Kommunikation zu ermöglichen wurden 32 **Emailverteiler** eingerichtet und gepflegt.

Die Planung, Vorbereitung, Durchführung und Nachbereitung der zahlreichen unterschiedlichen **Treffen und Telefon- bzw. Videokonferenzen** war eine weitere über die gesamte Projektlaufzeit andauernde Aufgabe. Als Regeltermine wurden wöchentliche Abstimmung zwischen Mercedes-Benz und EICT abgehalten, zweiwöchentliche Teilprojektleitertreffen und vierwöchentliche Treffen der Firmenprojektleiter. Dazu kamen sechs Steuerkreistreffen sowie zahlreiche Treffen zu Sonderthemen wie dem Joint Controller Vertrag, der Vorbereitung von Zwischen- und Abschlusspräsentation oder zur Mitigation der Verzögerungen durch die Covid-19-Pandemie oder den verzögerten Start der Datenaufnahme.

In den verschiedenen Regelterminen konnten die jeweiligen Gremien über die aktuellen Themen diskutieren. So wurde sichergestellt, dass die gefundenen Lösungen und die projektierten Maßnahmen im Konsortium verankert sind und von Diesem mitgetragen werden. Der periodische Abgleich wurde bei allen Themen mit projektweiter Relevanz durchgeführt, wie zum Beispiel zur Projektstrategie, zum Ergebnistransfer oder natürlich zum Austausch über operative Lösungen

bei Problemen mit den gemeinsamen Unteraufträgen oder den nötigen vertraglichen Abstimmungen um Konformität mit der DSGVO sicherzustellen.

Die Erstellung der regelmäßig fälligen **Berichte** wurde über die Confluence-Projektplattform effizient organisiert, auf der die Inhalte durch die jeweils Verantwortlichen bereitgestellt und gereviewed wurde. Daraus wurden dann Dokumente erstellt, die in das angestrebte Layout gebracht und versendet wurden. Alle Berichte für den Projektträger wurden im vorgegebenen Zeitplan eingereicht. Während der Projektlaufzeit wurden so 19 Deliverables, 4 Meilensteinberichte, 6 zusammenfassende Zwischenberichte und der hier vorliegende zusammenfassende Schlussbericht zusammengestellt und ihrer jeweiligen Nutzung zugeführt.

Während der Projektlaufzeit auftretende **Abweichungen vom Projektplan** oder von außen auf das Projekt einwirkende Entwicklungen wurden kontinuierlich gemonitort und es wurden entsprechende Maßnahmen zu **Mitigation** beschlossen und durchgeführt. Damit konnte das Projekt trotz verschiedener Widrigkeiten seine Ziele erreichen obwohl verschiedene verzögernde Umstände eingetreten sind, die Abweichungen von der VHB nötig gemacht haben. Die Covid-19-Pandemie ist hier zuallererst zu nennen aber auch die deutlich über das geplante Maß hinausgehenden Schwierigkeiten rund um die Erstellung der Datenschutzvereinbarungen stellen Ereignisse dar, die in den Leitungsrunden besprochen wurden und für die jeweils geeignete Mitigationsmaßnahmen beschlossen und umgesetzt wurden. Als zum Beispiel der bilaterale Abschluss einer Auftragsdatenvereinbarung zwischen GUA4-Nehmer DXC und jedem einzelnen Partner nicht im nötigen Zeitraum abgeschlossen wurde, hat TP6.1 diesen Vorgang an sich gezogen und beschleunigt, um die weitere Verzögerung so gering wie möglich zu halten.

Als eine zentrale Mitigationsmaßnahme sei hier die **Projektverlängerung** genannt. Nachdem der nicht zu vermeidende Verzug feststand wurde von AP6.1 eine Projektverlängerung von vier Monaten im Projekt abgestimmt. Als Vermittler zwischen Projektträger und Konsortium hat AP6.1 dann die nötigen Schritte abgesprochen und effizient mit den Partnern umgesetzt.

Zur effizienten Durchführung von Standardaufgaben wurden **Prozesse** definiert und wo möglich durch geeignete Hilfsmittel wie die Confluence-Projektplattform und entsprechende Templates unterstützt, dies betraf zum Beispiel die Freigabe von Veröffentlichungen oder die Erstellung von Berichten.

3.6.2.2 AP6.2 Austausch mit anderen Projekten

3.6.2.2.1 Zielsetzung

Das AP6.2 war die Schnittstelle zu den anderen Projekten der Projektfamilie und weiteren externen Projekten und Organisationen des internationalen Innovationsfeldes KI. Es hat sichergestellt, dass KI Delta Learning an den Ergebnissen der KI-Projektfamilie partizipiert und im Projekt erarbeitetes Wissen transferiert. Dazu mussten einerseits die für den Transfer nach außen identifizierten Ergebnisse entsprechend aufbereitet werden, andererseits mussten die Fortschritte und relevanten Ergebnisse der anderen Projekte systematisch für das Konsortium verfügbar gemacht werden. In AP6.2 wurde eine Transfer-Roadmap entlang der Deliverable-Liste erstellt und getrackt, die dazu nötigen Freigabeprozesse wurden definiert und implementiert. Außerdem wurden kontinuierlich weitere Forschungsbedarfe erhoben und konsolidierte Wissensbestände für die Publikation und industrielle Verwertung weitergegeben.

Die definierten Aktivitäten sind im Wesentlichen - unter Anleitung jeweils eines Partners aus dem Konsortium aus der Gruppe der OEM sowie der Tier-1 - durch das Projektmanagement erfolgt.

3.6.2.2.2 Ergebnisse im Einzelnen

Es wurden verschiedene Aktivitäten gemeinsam mit anderen Projekten aus der KI-Familie durchgeführt:

- Bei einer gemeinsam vom damaligen BMWi und dem VDA organisierten High-Level-Veranstaltung am 02.03.2021 zum Thema autonomes Fahren hatten die Projekte der KI-Familie die Gelegenheit sich diesem Publikum vorzustellen.
- Im Rahmen der ITS Europe am 11.-15.10.2021 in Hamburg konnte die KI-Familie eine eigene Session gestalten, in der die Projekte und die Familie auf dieser renommierten Plattform der Fachöffentlichkeit nähergebracht wurden.
- Bei einem gemeinsamen Workshop mit KI Wissen und KI Data Tooling am 14.10.2022 wurden vier "Common Topics" fokussiert: ML-Ops picture/Joint DevOps, Active Learning (semi-supervised learning), Ontology und Synthetic Data. Nach Berichten aus den Projekten wurden im Anschluss jeweils die übergreifenden Ziele und entsprechende Potenziale für Kooperationen diskutiert.
- KI Wissen hat unter Teilnahme von Projektmitarbeitern der anderen Projekte der KI-Familie zwei "Open Project Days" durchgeführt (22.02.2022 und 17.01.2023) in denen der Stand der Ergebnisse für Projektexterne aufbereitet und im projektübergreifenden Kontext diskutiert wurde.
- Zu Zwischen- und Abschlusspräsentationen der KI-Familie wurden explizit auch die Mitarbeiter der jeweils anderen Projekte eingeladen. Vor allem bei den inhaltlich tiefergehenden Teilen der Veranstaltungen wurde die Gelegenheit für Austausch und Fachdiskussionen vielfach und gern genutzt.
- Bis zum Ende von KI Absicherung Mitte 2022 wurden auch Regeltelkos mit den Koordinatoren der KI-Projekte durchgeführt um projektübergreifende Themen auch auf dieser höheren Ebene zu diskutieren und abzustimmen.
- Um die projektübergreifend Neuigkeiten zu verbreiten wurde ein gemeinsamer Newsletter für alle Projekt der KI-Familie etabliert der so Highlights und übergeordnete Informationen für alle leichtgängig verfügbar gemacht hat.
- Im Rahmen der Abschlussdokumentation von KI Absicherung wurden 139 finale Ergebnissteckbriefe zur internen Verwendung erstellt. Für einen Teil davon wurden weitergehende Freigaben erteilt: 58 Ergebnisse wurden zur allgemeinen Veröffentlichung freigegeben, für die Mitglieder von Projekten der VDA-Leitinitiative autonomes und vernetztes Fahren waren es mit 101 Ergebnissen deutlich mehr, sodass hier ein verbesserter Ergebnistransfer zu einem fruchtbaren Zusammenspiel der Projekte beitragen konnte.

3.6.2.3 AP6.3 Ergebnisverbreitung

3.6.2.3.1 Zielsetzung

Ziel von AP6.3 war die Kommunikation des Vorhabens und dessen Ergebnissen nach außen. Damit war es für die zielgruppengerechte Aufbereitung und Verbreitung von Projektinformationen durch geeignete Kommunikationsmittel verantwortlich. Die Ziele im Einzelnen waren:

- Planung, Durchführung und Nachbereitung der projekteigenen Veranstaltungen, z.B. Halbzeit- und Abschlusspräsentation inkl. Einladung an unterschiedliche Adressatengruppen aus dem projektexternen Umfeld und Abstimmen der Beteiligungen im Rahmen der KI-Familie oder anderer Akteure
- Schaffung einer Projektidentität in Anlehnung an die Corporate Identity von KI Absicherung um die Familienzugehörigkeit und damit die Gesamtwahrnehmung zu optimieren
- Entwicklung und Pflege einer Kommunikationsstrategie (inklusive der Abstimmung mit den Zeitplänen der Schwesterprojekte),
- Konzeption, Gestaltung, Umsetzung und Pflege von Basisinstrumenten der Kommunikation,
- Implementierung und Durchführung von Pressearbeit sowie
- Entwicklung und Umsetzung geeigneter Formate zum Dialog mit Stakeholdern

Die Formulierung der inhaltlichen Anforderungen an die Ergebnisverbreitung fanden in starker Abstimmung mit dem AP6.1 an statt. Die konkrete Umsetzung von AP6.3 ist im Rahmen des GUA2: Ergebnisverbreitung durch die EICT GmbH erfolgt. Die formale Verantwortung für den GUA, die Priorisierung der Aufgaben und die Abnahme der Leistungen erfolgte durch den Projektkoordinator.

3.6.2.3.2 Ergebnisse im Einzelnen

3.6.2.3.2.1 Projektidentität, CI und Templates

Um die ganze Breite der Aktivitäten der KI-Familie auch für Außenstehende wahrnehmbar zu machen, wurden für die vier Projekte und die Familie selbst jeweils Corporate Designs geschaffen, die eigenständig funktionieren aber die Zusammengehörigkeit klar ersichtlich machen. Die individuellen Signets basieren auf sechseckigen Grundformen und das "KI" im Namen der jeweiligen Projekte werden in derselben Farbe dargestellt. Jedes Einzelprojekt hat zusätzlich eine eigene Projektfarbe, die im Signet und im Namen auftaucht.



Abbildung 73: Logos der KI Familie und der vier Einzelprojekte

Für die Projekte wurden zusätzlich individuelle Hintergrundbilder mit den jeweiligen Projektfarben bereitgestellt, die zum Beispiel für Titelfolien oder Webseiten genutzt wurden und so verstärkt zur Projektidentität beitragen.



Abbildung 74: Projektlogos mit Hintergrundbildern

3.6.2.3.2.2 Abgestimmte Außenkommunikation im Rahmen der KI-Familie

Mit der KI-Familie wurde erstmals in diesem Kontext ein zusammenhängendes Cluster von Förderprojekten von einer gemeinsamen, abgestimmten Außenkommunikation begleitet. Die oben beschriebene, gemeinsame CI der Projekte wurde von den Projekten konsequent an den potenziellen Einsatzstellen der genutzten Medien verwendet und mit Verweisen auf die KI-

Familie versehen, um die Rezipienten immer auch auf den übergeordneten Ansatz mit komplementären, sich ergänzenden Einzelprojekten hinzuweisen. Dadurch konnte eine starke Sichtbarkeit geschaffen werden, weil jede Kommunikation der Einzelprojekte immer auch die Familie und die anderen Projekte mitbeworben hat. Hinzu kamen die Kanäle der KI-Familie selbst wie die Homepage <https://ki-familie.vdali.de/>, der Familien-Newsletter oder die KI-Familien-LinkedIn-Gruppe.

Die dadurch erzielte Aufmerksamkeit wurde auch bei der bereits in AP5.2 berichteten Veranstaltung am 02.03.2021 genutzt. Bei einer gemeinsam von BMW und VDA organisierten High-Level-Veranstaltung hatten die Projekte im Rahmen eines Beitrags der KI-Familie die Chance sich einem hochrangigen Publikum bekannt, beziehungsweise bekannter zu machen. Neben Bundesminister Peter Altmaier waren Vortragende und Teilnehmer aus der Vorstandsebene anwesend, wie z.B. Ola Källenius und Oliver Zipse, die Vorstandsvorsitzenden der damaligen Daimler AG und der BMW AG.

3.6.2.3.2.3 Templates

Die projekttypischen Elemente wurden dann bei der Erstellung der Office-Templates verwendet. Es wurden verschiedene Templates für Word-Dokumente bereitgestellt sowie für Power-Point-Präsentationen, beides jeweils in Deutsch und auf Englisch.

3.6.2.3.2.4 Standardpräsentation, Factsheet, Projekthandbuch

Die Projektidentität wurde auch für das Fact Sheet und die Standardpräsentation verwendet, die jeweils auf Deutsch und Englisch erstellt und dem Konsortium zur Verfügung gestellt wurden. Zu Projektbeginn wurde zudem ein Handbuch erstellt und auf der Projektplattform veröffentlicht, in dem zahlreiche Informationen bereitgestellt wurden, die für Projektmitarbeiter relevant sind. Neben der Projektstruktur und den Kontakten der verschiedenen Ansprechpartner wurden dort Berichtspflichten und Prozesse erklärt. Ebenso wurden die Emailverteiler vorgestellt, die Deliverables und die Meilensteine sowie zum Beispiel auch die Hinweispflicht auf den Fördergeber bei Veröffentlichungen aus dem Projekt. Die Erstellung des Projekthandbuches ist in enger Kooperation mit AP6.1 erfolgt.

3.6.2.3.2.5 Internetauftritt

Ebenfalls zu Projektbeginn wurde eine öffentliche Webseite geschaffen (Deutsch und Englisch) die unter der URL <https://www.ki-deltalearning.de/> zu erreichen ist. Sie wird vom der EICT GmbH auch noch einige Zeit, vermutlich bis ca. zwei Jahre nach Projektende weiterbetrieben werden.

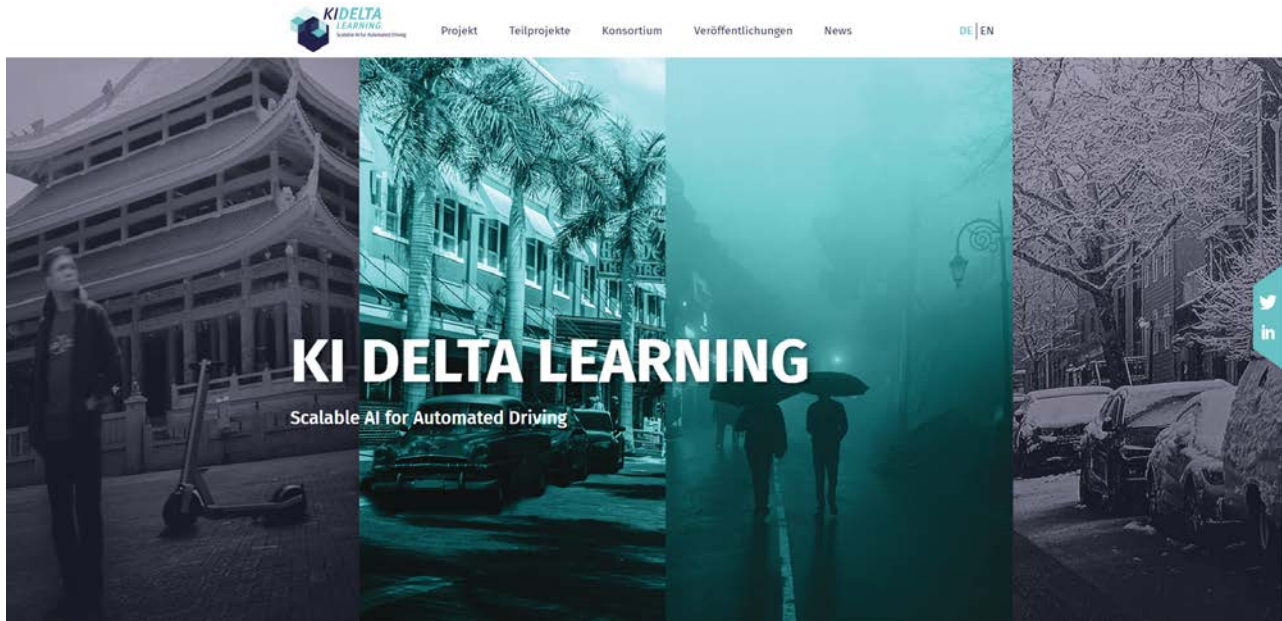


Abbildung 75: Startansicht der KI Delta Learning-Webseite

Dort wird das Projekt vorgestellt, es wurden Neuigkeiten zu öffentlichen Veranstaltungen gepostet und es wurden Materialien und Ergebnisse für die Öffentlichkeit verfügbar gemacht wenn eine entsprechende Freigabestufe vorlag. Mittlerweile sind dort verfügbar:

- Materialien für die Presse, Links zum Pressemitteilungen der Partner und zu externen Berichten über KI Delta Learning
- Projektpräsentationen
- Liste der wissenschaftlichen Veröffentlichungen mit pdfs oder Links zu den Originalpapern
- 13 Präsentationen, Keynote und ca. 50 Poster der Zwischenpräsentation
- Deliverables aus dem dritten (und nach Fertigstellung auch aus dem vierten) Projektinkrement soweit eine entsprechende Freigabe vorliegt
- ca. 50 Poster und nach Klärung der Freigabe auch Keynote und die Präsentationen der Abschlussveranstaltung

3.6.2.3.2.6 Kick-off, Zwischen- und Abschlussveranstaltung

Zu den Aufgaben von AP6.3 gehörte auch die Organisation der großen Projektevents Kick-Off, Zwischenpräsentation und Abschlussveranstaltung.

3.6.2.3.2.6.1 Kick-off

Das Kick-off Treffen wurde als zweitägige Veranstaltung am 22. und 23.01.2020 beim Projektkoordinator Mercedes-Benz in Sindelfingen durchgeführt. Unter den 92 Teilnehmern waren Mitarbeiter aller Partner, Vertreter des Fördermittelgebers BMWi (jetzt BMWK) und des Projektträgers TÜV Rheinland. Am ersten Tag wurden unter anderem die Projektstruktur mit den vier Projektinkrementen vorgestellt, anschließend wurden die TPs präsentiert, um allen Mitwirkenden zu vermitteln, wie der individuelle Beitrag in das Gesamtprojektergebnis

einfließen wird. Am zweiten Tag wurden Ziele des ersten Projektinkrements wurden abgestimmt, auf dieser Basis wurden dann TP-Kick-offs in parallelen Sessions durchgeführt. Nach einer gemeinsamen Zusammenführung wurde abschließend noch ein Workshop zu Use Cases und Data Specification durchgeführt.



Abbildung 76: Gruppenfoto vom Kick Off 22/23.01.2020

3.6.2.3.2.6.2 Zwischenpräsentation

Die Zwischenpräsentation wurde am 07.10.2021 aufgrund der anhaltenden Covid-19-Pandemie als virtuelle Veranstaltung durchgeführt. Dabei wurde Veertly genutzt, ein professionelles Konferenztool, mit dem eine Hauptbühne aber auch verschiedene parallele Sessions bespielt werden können. Durch individuelle Gestaltungsmöglichkeiten war es möglich die Corporate Identity über Farben und Logo auch hier als Wiedererkennungsmerkmal fortzuführen, die Konferenzplattform verfügt außerdem über ein differenziertes System für Teilnehmermanagement und Informationsoptionen, sodass eine attraktive Veranstaltung entstanden ist, die deutlich mehr bieten konnte als einen Stream, den man sich ansehen kann: In einer ersten parallelen Session haben die TPs 1 bis 4 je drei oder vier Highlightvorträge präsentiert und die Teilnehmer konnten die Themen ihrer Wahl in jeweils eigenen virtuellen Räumen besuchen, bei Bedarf wechseln und dort auch direkt mit dem Vortragenden oder den anderen Besuchern diskutieren. Bei einem zweiten parallelen Teil war dies ähnlich, hier wurden 50 Highlightposter vorgestellt, die nach Partner geordnet in 20 Räumen zur Auswahl standen. Zusätzlich konnten die Teilnehmer über große oder kleine Chatrunden kommunizieren oder auch weitere Räume eröffnen und spontan in einen individuellen, intensiveren Austausch treten.

Durch diese Umsetzung konnte die Zwischenpräsentation auch virtuell mit über 300 Teilnehmern sehr erfolgreich durchgeführt werden, die zum Beispiel mit 5.000 Views bei LinkedIn auch dort einen nennenswerten Widerhall hinterlassen hat.

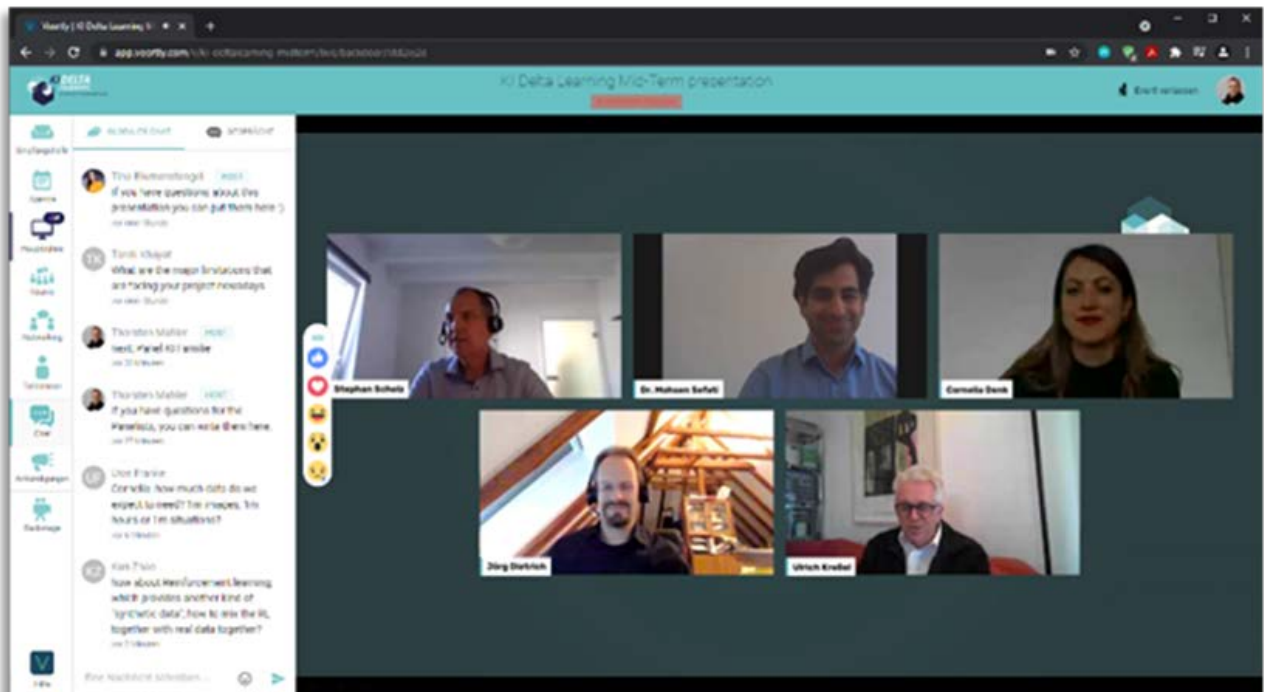
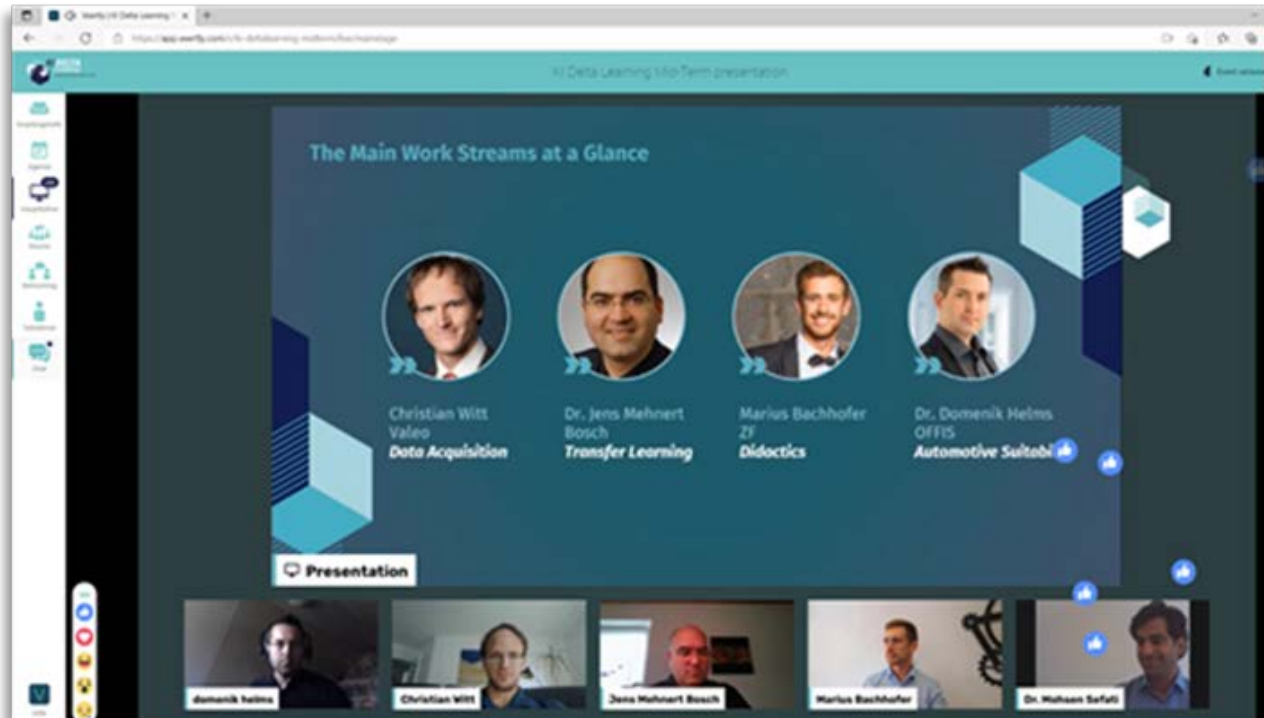


Abbildung 77: Zwei Screenshots aus der virtuellen Zwischenpräsentation

3.6.2.3.2.6.3 Abschlussveranstaltung

Die Abschlussveranstaltung konnte dann wieder als physisches Treffen organisiert werden, am 09. und 10.03.2023 wurden die Ergebnisse bei Mercedes-Benz in Stuttgart-Vaihingen der

Fachöffentlichkeit vorgestellt. Am ersten Tag, der auf internationales Publikum und eher höherrangige Firmenvertreter zugeschnitten war, konnten die Zuschauer nach Vorträgen über das Gesamtprojekt und die Teilprojekte (beziehungsweise deren Ergebnisse) dann aus jedem TP je ein bis zwei Fachthemen per Abstimmung auswählen, die dann in Vertiefungsvorträgen präsentiert wurden. Ein weiteres Highlight stellte am ersten Tag sicherlich die Keynote von Magnus Östberg dar, dem Chief Software Officer (CSO) von Mercedes-Benz.

Der zweite Tag war als "Expert Day" angelegt, der Fokus lag auf "Deep Dives" und Gelegenheit zu Austausch und Diskussion. Davon wurde bei den Fachvorträgen und der Postersession regen Gebrauch gemacht, aber auch in den Pausen entstanden zum Beispiel rund um die Poster regelmäßig Fachdiskussionen. Unter den ca. 150 Teilnehmern vor Ort waren Vertreter aller Partner, des BMWK und des Projektträgers, zusätzlich konnte das Geschehen auf der Hauptbühne auch über einen Live-Stream verfolgt werden.

Neben der Organisation von zum Beispiel Veranstaltungsort, Catering und Technik wurde viel Arbeitsleistung in die Optimierung von Postern und Vorträgen gesteckt. Wie auch schon bei der Vorbereitung der Zwischenpräsentation musste nach Bereitstellung der Templates der Rücklauf geprüft und überarbeitet werden, um die durchgehende Verwendung der CI und ein ansprechendes, professionelles Layout sicherzustellen. Angesichts der Volumina an Material für beide Veranstaltungen stellt das eine Aktivität dar die gesondert erwähnt werden sollte. Dies wurde gemeinsam mit AP6.1 durch die EICT GmbH durchgeführt.





Abbildung 78: Bilder von der Abschlussveranstaltung

3.6.2.3.2.7 Veröffentlichung von Projektergebnissen

Die im Projekt definierten Einzelergebnisse wurden in 19 Deliverables zusammengefasst, jeweils am Ende der Projektinkremente erstellt wurden. Den gereiften Einzelergebnisse, die ab dem dritten Projektinkrement entstanden sind, wurden dann drei möglichen Publikationsklassen zugewiesen: Rein intern, für die VDA Leitinitiative oder ganz öffentlich. Die öffentlichen Versionen wurden beziehungsweise werden auf <https://www.ki-deltalearning.de/downloads> hochgeladen, die etwas umfangreicheren Versionen für die LI können über geteilte Bereiche der gemeinsamen Projektplattform dem vorgesehenen Kreis zur Verfügung gestellt werden. Diese weitreichende Ergebnisverbreitung - zum Teil noch während der Projektlaufzeit - geht über das normalerweise übliche Maß hinaus und stellt eine Besonderheit von KI Delta Learning und den anderen Mitgliedern der KI-Familie dar. Sie ist Teil des Konzeptes der VDA Leitinitiative autonomes und vernetztes Fahren, um mit der sich schnell entwickelnden Forschung zu autonomen Fahren und dem dafür nötigen Einsatz von KI schritt halten und auch Impulse geben zu können.

3.6.2.3.2.8 Wissenschaftliche Veröffentlichungen, Workshop on Autonomy@Scale auf IEEE Intelligent Vehicles Symposium

Die Ergebnisse die im Laufe von KI Delta Learning entstanden sind wurden zum allergrößten Teil in wissenschaftlichen Artikeln veröffentlicht. Dabei mussten verschiedenen Vorgaben bezüglich der Urheberschaft und der Freigabe als offizielle Projektergebnis eingehalten werden. Dazu wurde ein Confluence-basierter Prozess entwickelt und gepflegt, mit dem über 100 Veröffentlichungen systematisch erfasst und entsprechend kanalisiert wurden. Die Liste der Veröffentlichungen ist als Kapitel 5 Teil dieses Berichts, die Beiträge wurden zu einem guten Teil auf sehr

renommierten Konferenzen gehalten, wie dem IEEE Intelligent Vehicles Symposium. Dort konnte aus KI Delta Learning heraus sogar eine eigene Reihe etabliert werden, die Workshops on Autonomy@Scale. Dort wurden beziehungsweise werden seit 2021 und 2023 nun zu dritten Mal explizit Ergebnisse in der Fachöffentlichkeit vorgestellt und diskutiert, die sich mit der Erweiterung bestehender KI-Funktionen um Deltas beschäftigen.

3.6.2.3.2.9 LinkedIn-Gruppe KI Familie

Als weiteres Disseminationsmedium wurde eine LinkedIn-Gruppe "KI Familie" gegründet, die Personen außerhalb der Konsortien offen steht. Auf diese Weise konnte die Wahrnehmung weiter erhöht werden und die Projekte haben einen zusätzlichen Kanal zu Verfügung über den Informationen ausgespielt werden konnten beziehungsweise können.

3.6.3 Zielerreichung

Das Teilprojekt 6 - Projektorganisation hat die gesetzten Ziele erreicht. Die auftretenden Herausforderungen durch die Covid-19-Pandemie oder die unvorhergesehen komplexe Ausarbeitung einer Datenschutzvereinbarung wurden zeitnah nach dem Auftreten besprochen und es wurden Mitigationsmaßnahmen angestoßen, die letztlich dazu geführt haben, dass KI Delta Learning mit nur vier Monaten Projektverlängerung die vorgesehenen Ergebnisse erzielen konnte. Bei der inneren Organisation im Konsortium konnte mit transparenten Prozessen und der vielseitig nutzbaren Projektplattform eine effiziente Struktur geschaffen werden, die es ermöglicht hat, dass die Partner und das Konsortium als Ganzes die administrativen Pflichten mit minimalem Aufwand erfüllen konnte.

Die zahlreichen in KI Delta Learning erzielten Ergebnisse wurden in über 100 wissenschaftlichen Veröffentlichungen in die Fachwelt getragen, die Deliverables aus den beiden letzten Projektinkrementen 3 und 4 sind auf der externen Webseite verfügbar und stellen zum Teil sehr umfangreiche Dokumentationen des erreichten Fortschritts dar, die öffentliche Version der Deliverable 11 hat zum Beispiel 358 Seiten. Die Zwischen- sowie naturgemäß vor allem die Abschlusspräsentationen waren attraktive Veranstaltungen, die dort gezeigten Inhalte sind weiterhin über die Webseite abrufbar, die Außenwirkung ist also nicht auf die Veranstaltungen oder die Projektlaufzeit begrenzt. Die Verwendung der Projektidentität und der Corporate Identity von Projekt und Projektfamilie haben zusätzlich dazu beigetragen, dass KI Delta Learning einen bleibenden Eindruck in der Fachwelt hinterlassen wird und für kommende Mitglieder der KI Familie von vornherein in ein positives Licht setzen kann.

4 Zusammenfassung

4.1 TP1 - Datengewinnung

Im Projekt wurde eine Plattform für integriertes Datenmanagement und -verarbeitung entwickelt, die aus einer Benutzeroberfläche und einer zugrundeliegenden Plattform besteht. Die Benutzeroberfläche bietet eine Übersicht über alle verfügbaren Datensätze, Metadaten und den Status der Datenverarbeitung. Es ermöglicht auch das Filtern und Suchen von Datensätzen. Die Plattform für die Datenverarbeitung führt automatisierte Schritte zur Aufbereitung der Rohdaten durch.

Für die Datengewinnung wurden verschiedene Arbeitsgruppen eingerichtet, um die Definition des Sensor Setups, die Routendefinition, die Annotationen und die Qualitätsvalidierung der Daten zu bearbeiten. Die Zusammenarbeit zwischen den Partnern und Sensorherstellern wurde intensiviert, um die Anforderungen zu erfüllen und die Integration und Kalibrierung der Sensoren sicherzustellen. Es wurden insgesamt 240 Recordings in Deutschland und Italien erstellt, mit einer Gesamtdauer von 193 Stunden. Die Annotationen und Labeling-Formate wurden in der Arbeitsgruppe spezifiziert und validiert, wobei die Qualität der annotierten Daten als gut bewertet wurde. Zur weiteren Anreicherung der Daten wurden Algorithmen entwickelt, um Informationen wie den Sonnenstand und Daten aus der Open-Source-Karte "OpenStreetMap" zu extrahieren. Ein Algorithmus wurde implementiert, um eine automatisierte Auswahl von Frames zu ermöglichen, die im Datensatz unterrepräsentiert sind und dadurch eine iterative Verbesserung des Datensatzes ermöglichen.

Es wurden auch neue Modelle für Lidar- und Radar-Sensoren in der Simulationsumgebung CARLA entwickelt, um realistischere Daten zu erzeugen. Ein Ansatz auf Basis von Neural Radiance Fields (NeRF) wurde verwendet, um virtuelle Welten basierend auf Kamera- und Lidar-Daten einer realen Szene zu konstruieren. Um eine optimale Kombination von synthetischen und realen Daten zu ermöglichen, wurden die Modalitäten der Sensoren und Labeling-Spezifikationen angeglichen. Es wurden verschiedene Strategien entwickelt, um gezielt Key-Frames zu generieren und spannende Verkehrssituationen zu erzeugen. Im Rahmen der Evaluierung wurden Tasks ausgewählt und Metriken zur Evaluation der Modelle definiert. Es wurden Delta-Labeling-Strategien, Anonymisierungsmethoden und Algorithmen zur Auswahl von geeigneten Aufnahmen für das manuelle Annotieren entwickelt.

Das Teilprojekt hat wichtige Fortschritte im Bereich des integrierten Datenmanagements und der Datenverarbeitung erzielt, was zu einer verbesserten Datenauswertung und Validierung der Modelle führt. Die entwickelten Algorithmen und Methoden tragen zur Erhöhung der Vielfalt und Qualität der Daten bei und ermöglichen eine effiziente Kombination von synthetischen und realen Daten.

4.2 TP2 - Transfer Learning

Im Teilprojekt Transfer Learning (TL) wurden die Fähigkeit von neuronalen Netzwerken zur Verarbeitung von Daten verschiedener Sensor- und Ortsdomänen in den vier Arbeitspaketen Kontinuierliches Lernen/Continuous Learning (CL), Synthetic Data (SD), Cross-Sensor-Adaption (CSA) und Orts-/Zeitdomäne, grundlegend verbessert.

Kontinuierliches Lernen beschreibt das Hinzufügen neuer Lerndaten in ein schon trainiertes KI-Modell, was man auch unter Begrifflichkeiten wie "online learning" und "continuous learning"

kennt. Catastrophic forgetting" (CF) beschreibt das Phänomen, dass KI-Modelle während des kontinuierlichen Lernens (CL) das Wissen über zuvor gelernte Aufgaben nach dem Lernprozess einer neuen Aufgabe mit hoher Wahrscheinlichkeit vergessen werden. Durch die Implementierung von CL-Verfahren in KI-Modellen konnten Probleme wie "catastrophic forgetting" (CF) gemildert werden und die Lerneffizienz dieser Netze gesteigert werden. Folglich ist es mit solchen Verfahren möglich Wissen aus verschiedenen Datensätzen miteinander zu kombinieren, was die Kernaufgabe von Transfer Learning ist. Im Projekt wurden dazu zahlreiche Methoden zu CL ausprobiert, gegeneinander verglichen und auf die für die Projektpartner relevanten "use cases" angewendet.

Durch den Einsatz von synthetischen Daten konnten KI-Modelle auf eine größere Varietät von Daten vorbereitet werden, was zu einer Verbesserung bezüglich nicht gesehener Daten führte. Synthetische Daten sind von Künstlicher Intelligenz (KI) erzeugte, strukturierte Daten, die aus realen Informationen abgeleitete Fake-Daten enthalten. Sie gelten als anonym, wenn kein Personenbezug besteht. Konkret haben Partner beispielsweise mit dem CARLA-Simulator komplette synthetische Welten erzeugt, in denen virtuelle Fahrzeuge das autonome Fahren trainieren können. In dieser Umgebung wurden zudem, mit wenigen Parametern, verschiedene Wettersituationen konfiguriert und damit weitere Daten, für eine bessere Generalisierung des KI-Modells, erzeugt.

Cross-Sensor-Adaption ist eine Methode zur Anpassung von KI-Modellen an unterschiedliche Sensoren. Die CSA ermöglichte eine Übertragung von Wissen aus einem Trainingsdatensatz auf einen anderen, was die Effizienz des Trainingsprozesses weiter erhöhte. Konkret wurden CSA-Verfahren entwickelt, welche mit Hilfen von CL-Methoden, durch Anpassung der Verlustfunktion und Manipulation von Bild- und Annotations-Daten das Delta unterschiedlicher Sensoren reduziert haben.

Die Implementierung der Orts- und Zeitdomänen-Features schließlich ermöglichte es den Netzen, Daten aus verschiedenen Orten und Zeiten zu integrieren, was zu einer verbesserten Generalisierung und einer höheren Fähigkeit zur Vorhersage führte. Hier wurden konkret SD verwendet um leicht verschiedene Wettersituationen augmentieren zu können, was insbesondere für Kameradaten gut funktioniert. Auch bezüglich LIDAR bestand grosse Interesse an Verfahren zur Reduktion des Deltas zwischen Seriensensoren. Hier wurden immer wieder "generative adversarial network"-KI-Modelle verwendet, welche auch bei der Reduktion von Kameradeltas im P gute Ergebnisse lieferten.

Im Hinblick auf unsere Projektziele können wir somit feststellen, dass das TL-Teilprojekt ein bedeutender Erfolg war und wesentlich dazu beigetragen hat, die Effektivität von KI-Modellen zu erhöhen und deren Einsatzmöglichkeiten in verschiedenen Bereichen zu erweitern.

4.3 TP3 - Didaktik

Das Teilprojekt Didaktik hat durch seine entwickelten Methoden wesentlich dazu beigetragen den Trainingsprozess neuronales Netze zu optimieren.

Im Gebiet teilüberwachtes und unüberwachtes Lernen wurden effizientere Lernmethoden entwickelt. Spezieller Fokus wurde hier auf die semantische Segmentierung gelegt, da dort der Labelingaufwand besonders hoch ist. Ebenso wird es notwendig sein in Zukunft neue Klassen dazuzulernen. Hierzu wurden zwei Methoden für das unüberwachte Auffinden und Nachlernen neuer Klassen entwickelt. Eine weitere Möglichkeit Labelingaufwand zu reduzieren ist es, Labels

mit geringeren Detailgrad zu verwenden. Beispielsweise Bounding Boxen anstelle von pixelgenauen Segmentierungen. Hierzu wurden mehrere Konzepte zum schwach überwachten Lernen entwickelt, implementiert, experimentell ausgewertet, die Schwächen analysiert und eine abschließende Bewertung vorgenommen.

Die Generierung von Tiefendaten aus einzelnen Kamerabildern würde die Umfeldwahrnehmung deutlich erleichtern und teure 3D-Sensoren wie Lidare obsolet machen. Es wurden Methoden für die unüberwachte monokulare Tiefenschätzung entwickelt, die die Generierung von Tiefendaten aus monokularen Bildern ermöglicht.

Das nächste Forschungsfeld hat Ansätze für die Erweiterung der Performanz von neuronalen Netzen über über die Trainingsorganisation erarbeitet.

Um die Überparameterisierung, also eine ineffiziente Nutzung der Gewichte, von neuronalen Netzen zu reduzieren, wurden die Analogien von hierarchischen Iterationsverfahren und residual Netzen ausgenutzt. Der Fokus bei den untersuchten Iterationsverfahren lag bei Mehrgitter Methoden, welche hierarchische Verfahren sind um große (nicht-) lineare Gleichungssysteme effizient zu lösen.

Für das Training eines 3D Objektdetektors sind, wie bei Deep Learning-Verfahren üblich, eine große Anzahl annotierter Daten notwendig. Da das Aufnehmen und Annotieren von Daten kosten- und zeitintensiv ist, wird Augmentierung verwendet. Um die Auswirkungen von verschiedenen Schwierigkeitsgraden der Augmentierung während des Trainings zu untersuchen, wurde ein Hyperparameter eingeführt, der die Modulation der Schwierigkeit für alle Augmentierungsmethoden gleichermaßen ermöglicht.

Es wurde eine Simulation entwickelt mit der komplexe Animationen im virtuellen Kontext aufgezeichnet werden können. Es können verschiedene Ground Truth Arten (RGB, Depth, Semantic/Instance Segmentation, Bounding Box, Human Joints 2D/3D, Meta Daten) generiert werden. Mit Hilfe der Aufzeichnung von realistischen Motion Capture Animationen wurden eigene Datensätze erzeugt.

Im nächsten Teilbereich wurden Ansätze für die effizientere Nutzung von Trainingsdaten entwickelt. Hierdurch werden sowohl der Labeling-Aufwand reduziert als auch die Trainingszeit verkürzt. Zwei Ansätze sind hier besonders hervorzuheben. Solche, die auf der Grundlage von unterschiedlichen Metriken Situationen herausfiltern, die nicht mit der bisherigen Verteilung der Trainingsdaten übereinstimmen und Solche, die speziell die aktuell falsch klassifizierten Daten berücksichtigen.

Es wurden verschiedene Aktive Learning Verfahren für die semantische Segmentierung und Objekterkennung entwickelt, deren Aquisefunktionen auf Inkonsistenzen und Unsicherheiten in den Prädiktionen beruhen.

Im Rahmen des Projektes wurde ebenfalls eine modulare halb-überwachte Domain-Adaptation-Methode mit CycleGAN entwickelt, welche mit Hilfe einer ausgewählten Menge an annotierten Realdaten zur Zielaufgabe (z.B. semantische Segmentierung) geleitet wird.

Zur Identifikation von Datenpunkten mit Potenzial zur Performancesteigerung eines Algorithmus wurde ein Ensemble-Model entwickelt, aus welchem die aleatorische und epistemische

Unsicherheit einer Trajektorie approximativ extrahiert werden kann und es somit erlaubt vor allem bisher schlecht gelernte Daten in das Training zu integrieren.

Ein anderer Schwerpunkt der diversen Arbeiten des dritten Teilbereiches lag auf dem Thema Transfer Learning zwischen der Kamera- und LiDAR-Domäne. Dazu wurde eine Methodik entwickelt, um Ansätze des Self-Trainings, die im Detail in Arbeitspaket unüberwachtes Lernen untersucht wurden, mittels dieses Transfers zu verbessern. Ergebnis der Arbeiten ist ein Graph Convolutional Network basierter Algorithmus, welcher eine maschinell lernbare Non-Maximum-Suppression durch Message Passing zwischen Detektionskandidaten modelliert.

Eine weitere Arbeit aus dem Bereich Trajektorienprädiktion brachte zukunftsweisende Erkenntnisse im Bereich kartenloser Bewegungsvorhersage und setzt neue Maßstäbe. Ergebnis ist ein CRAT-Pred, ein Expertenalgorithmus auf Basis eines Crystal Graph Convolutional Neural Networks mit Multi-Head Self-Attention, welches den Einfluss zwischen Verkehrsteilnehmern im Hinblick auf ihre zukünftigen Trajektorien schätzt.

Ziel einer anderen Arbeit war es zu untersuchen weshalb *Knowledge Distillation* Verfahren in der kontinuierlichen Semantischen Segmentierung so viel effektiver sind wie in der Klassifikation und wie Vergessen in kontinuierlichen trainierten Netzwerken auftritt. In den Untersuchungen wurde besonderes Augenmerk auf die „Hintergrund-Klasse“ gelegt, welche für irrelevante Objekte im Bild verwendet wird. Die Ergebnisse zeigten, dass der semantische Wechsel der Bedeutung der Hintergrund-Klasse einer der Hauptgründe für Vergessen im Klassen-Inkrementellen Lernen ist. Dies äußert sich, indem die ursprünglichen Klassen nun als Hintergrund erkannt werden, während ausschließlich die zuletzt gelernten Klassen korrekt klassifiziert werden.

Im Bereich der Multitask Netze wurde ebenfalls aktiv Forschung betrieben. Es wurde ein Algorithmus zur gleichzeitigen Prädiktion von Semantischer Segmentierung, Bounding Box Detektion und Posenschätzung entwickelt.

Darüber hinaus lässt sich das Netz auch mit Daten aus verschiedenen Datensätzen trainieren. Es wird hierfür ein Spezialistennetzwerk auf jedem Datensatz trainiert, welcher Inferenzen auf dem Ziel-Datensatz durchführt. Die Zusammenlegung der Inferenzen aller Spezialistennetzwerke lässt sich nun verwenden, um ein Multitask Netzwerk für alle Tasks zu trainieren.

Für automobile Anwendungen bringt das überwachte Training neuronaler Netze mehrere Nachteile mit sich, die sich in zwei Kernprobleme komprimieren lassen: Enormer Zeit- und Ressourcenverbrauch sowohl für das Training als auch das Data-Labeling. Zusammenfassend lässt sich sagen, dass das Teilprojekt Didaktik mit seinen entwickelten Methoden wesentlich dazu beigetragen hat den Trainingsprozess, ökonomischer und auch ökologischer zu betreiben.

4.4 TP4 - Automotive Tauglichkeit

Im Bereich der Robustheit wurden zwei unterschiedliche Ziele erreicht: Quantifizierung der Robustheit und Steigerung der Robustheit von KI-Systemen.

Bei der Quantifizierung wurden zwei Verfahren entwickelt, um die Robustheit mittels Novelty-Detektion zu untersuchen. Es wurden Ansätze basierend auf neuronalen Netzen entwickelt, um die Novelty in Verkehrssituationen zu bewerten. Die Unsicherheit in der Tiefenschätzung mittels Lidar-Punktwolken und Kamerabildern wurde untersucht, ebenso wie die Unsicherheit in der geschätzten Pose von Verkehrsteilnehmern. Diese Modelle helfen dabei, die Robustheit von KI-

Systemen im Fahrzeug in neuen und ungewohnten Situationen zu bewerten und sichere Fahrmanöver in unsicheren Situationen zu planen.

Um die Robustheit zu steigern, wurden Tiefenschätzungsverfahren entwickelt, die auch unter neuen Bedingungen wie unterschiedlichem Wetter robuste Vorhersagen treffen können. Eine neue Trainingsstrategie mit Daten-Augmentierung wurde verwendet, um die Regressionsleistung in unvorhergesehenen Situationen aufrechtzuerhalten. Des Weiteren wurde eine hierarchische Klassifikationsmethode entwickelt, die dem KI-Modell ermöglicht, auf übergeordnete Klassen zurückzugreifen, falls die Klassifikation in einer gegebenen Situation unsicher ist.

Im Bereich der Embedded Systems wurden Ziele in drei Kategorien erreicht: Bewertung von KI-Hardware, Vorhersage von Metriken und Optimierung von Metriken.

Die Bewertung von KI-Hardware umfasste eine Befragung von industriellen und akademischen Anwendern, um die Anforderungen an Hardwareplattformen zu ermitteln. Eine Marktrecherche wurde durchgeführt, um die Daten der Befragung zu ergänzen und Embedded AI-Entwicklern bei der Auswahl der optimalen Hardwareplattform für ihre Anwendungen zu unterstützen.

Die Vorhersage von Metriken ermöglichte es KI-Entwicklern, relevante Aspekte wie die erwartete Latenzzeit unter Berücksichtigung der Hardware frühzeitig im Entwurfsprozess zu berücksichtigen. Eine Hardware-in-the-Loop (HIL) basierte Methode wurde entwickelt, um die Leistungskennzahlen der Hardware bei der Inferenz von neuronalen Netzen zu messen und die Vorhersagemethodik zu überprüfen.

Die Optimierung von Metriken umfasste das Entfernen nicht relevanter Netzwerkteile (Pruning) unter Verwendung von Reinforcement Learning-Agenten. Es wurden Methoden entwickelt, um den Zusammenhang zwischen Komprimierungstechniken und KI-Beschleunigern optimal zu nutzen. Quantisierungstechniken wurden erweitert, um Speicherbedarf und Rechenaufwand zu reduzieren. Schließlich wurde eine Optimierungstechnik basierend auf Tensor Compression entwickelt, um große Convolution-Schichten durch kleinere und weniger speicherintensive Schichten zu ersetzen. Diese Optimierungen tragen zur Verbesserung der Eigenschaften der gewählten neuronalen Netze für die spezifische Hardware bei.

5 Veröffentlichungen

Im Projekt sind zahlreiche Veröffentlichungen entstanden, die folgenden Listen zeigen alle Veröffentlichungen, die im Rahmen des Projektes erarbeitet wurden.

5.1 Veröffentlichungen im Rahmen von Konferenzen

Titel der Publikation	Konferenz/ Veranstaltung	Datum der Veranstaltung
Detection and Retrieval of Out-of-Distribution Objects in Semantic Segmentation	Computer Vision and Pattern Recognition (CVPR) 2020, SAIAD Workshop	14-19 Jun 2020
Counterfactual Generative Networks	International Conference on Learning Representations (ICLR) 2021	12 Jan 2021
ConTraKG: Contrastive-based Transfer Learning for Visual Object Recognition using Knowledge Graphs	International Semantic Web Conference 2021	17 Feb 2021
Semi-Supervised Disparity Estimation with Deep Feature Reconstruction	Computer Vision and Pattern Recognition (CVPR), WiCV Workshop 2021	19 Jun 2021
Self-Supervised 3D Human Pose Estimation with Multiple-View Geometry	Computer Vision and Pattern Recognition (CVPR) 2021	19 - 25 Jun 2021
Improving robustness to distribution shift by combining frequency biased models	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021
Multi-Modal Fusion Transformer for End-to-End Autonomous Driving	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021
Seeing Behind Objects for 3D Multi-Object Tracking in RGB-D Sequences	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021
M3: Monocular Self-Supervised Depth, Pose and Motion	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021
Neural Attention Fields for End-to-End Autonomous Driving	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021

A Study of Learning in a Separate World - Federated Learning for Object Detection in Automated Driving	Computer Vision and Pattern Recognition (CVPR) 2021	19 Jun 2021- 25 Jun 2021
Learning Cascaded Detection Tasks with Weakly-Supervised Domain Adaptation	32nd IEEE Intelligent Vehicles Symposium (IV) 2021	10 Jul 2021
Continual Learning for Class- and Domain-Incremental Semantic Segmentation	32nd IEEE Intelligent Vehicles Symposium (IV) 2021	10 Jul 2021
A Survey on Deep Domain Adaptation for LiDAR Perception	32nd IEEE Intelligent Vehicles Symposium (IV), Workshop on Autonomy@Scale 2021	11 Jul 2021
Combining Semantic Self-Supervision and Self-Training for DomainAdaptation in Semantic Segmentation	32nd IEEE Intelligent Vehicles Symposium (IV), Workshop on Autonomy@Scale 2021	11 Jul 2021
Optimizing Neural Networks for Embedded Hardware	IEEE International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD) 2021	19 Jul 2021
Safety Assessment for Safety-Critical Systems with Rare Events Using Optimistic Optimization	19th International Conference on Formal Modeling and Analysis of Timed Systems (FORMATS) 2021	23 - 27 Aug 2021
Learning Temporal 3D Human Pose Estimation with pseudo-Labels	IEEE Int'l Conf on Advanced Video and Signal-based Surveillance	16 – 19 Nov 2021
Unsupervised Domain Adaptation via Shared Content Representation for Semantic Segmentation	24th IEEE International Conference on Intelligent Transportation (ITSC) 2021	19 - 22 Sep 2021
Continual Unsupervised Domain Adaptation for Semantic Segmentation by Online Frequency Domain Style Transfer	24th IEEE International Conference on Intelligent Transportation (ITSC) 2021	19 - 22 Sep 2021
Anomaly Detection in Multi-Agent Trajectories for Automated Driving	Conference on Robot Learning (CoRL) 2021	08 Nov 2021
Anomaly Detection in Multi-Agent Trajectories for Automated Driving	Conference on Robot Learning (CoRL) 2021	08 Nov 2021

On Exposing the Challenging Long Tail in Future Prediction of Traffic Actors	International Conference on Computer Vision (ICCV) 2021	10 - 17 Oct 2021
Is In-Domain Data Required to Learn Multi-View Depth Estimation?	International Conference on Computer Vision (ICCV) 2021	10 - 17 Oct 2021
SegmentMelfYouCan: A Benchmark for Anomaly Segmentation	International Conference on Computer Vision (ICCV) 2021	10 - 17 Oct 2021
Neural Attention Fields for End-to-End Autonomous Driving	International Conference on Computer Vision (ICCV) 2021	10 - 17 Oct 2021
ProAI: An Efficient Embedded AI Hardware for Automotive Applications - a Benchmark Study	International Conference on Computer Vision (ICCV), ERCVAD 2021	11 Oct 2021
MultiTask-CenterNet (MCN): Efficient and Diverse Multitask Learning using an Anchor Free Approach	International Conference on Computer Vision (ICCV), ERCVAD 2021	11 Oct 2021
Self-Supervised 3D Human Pose Estimation with Multiple-View Geometry	IEEE International Conference on Automatic Face & Gesture Recognition 2021	20 Oct 2021
Deep Neural Network Trustworthiness Estimation from Representation Anomalies	Neural Information Processing Systems 2021	06 Dec 2021
Projected GANs Converge Faster	Neural Information Processing Systems 2021	06 Dec 2021
On the Frequency Bias of Generative Models	Conference on Neural Information Processing Systems 2021	06 - 14 Dec 2021
SegmentMelfYouCan: A Benchmark for Anomaly Segmentation	Conference on Neural Information Processing Systems 2021	06 - 14 Dec 2021
AI as Key Technology for Pattern Recognition and Situational Awareness - How does the Vehicle get the Knowledge it needs?	TÜV Süd Tagung Automatisiertes Fahren 2022	29 Mar 2022

Understanding the Domain Gap in LiDAR Object Detection Networks	14. Uni-DAS e.V. Workshop Fahrerassistenz und automatisiertes Fahren 2022	09 - 11 May 2022
Wavelet-based Unsupervised Label-to-Image Translation	EEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2022	22 May 2022
A Downstream Task Informed Domain Adaptation GAN for Semantic Street Scene Segmentation	DataNinja Spring School 2022	23 - 25 Mar 2022
CRAT-Pred: Vehicle Trajectory Prediction with Crystal Graph Convolutional Neural Networks and Multi-Head Self-Attention	IEEE International Conference on Robotics and Automation (ICRA) 2022	23 - 27 May 2022
Capsule Networks for Hierarchical Novelty Detection in Object Classification	IEEE Intelligent Vehicles Symposium (IV), Workshop on Autonomy@Scale 2022	05 Jun 2022
MEAT: Maneuver Extraction from Agent Trajectories	33rd IEEE Intelligent Vehicles Symposium 2022	05 - 09 Jun 2022
Impacts of Data Anonymization on Semantic Segmentation	33rd IEEE Intelligent Vehicles Symposium 2022	05 - 09 Jun 2022
Towards Unsupervised Open World Semantic Segmentation	Conference on Computer Vision and Pattern Recognition (CVPR) 2022	19 - 24 Jun 2022
Anomaly Heatmaps for Out-of-Distribution Detection	Conference on Computer Vision and Pattern Recognition (CVPR) 2022	19 - 24 Jun 2022
Towards Understanding Adversarial Robustness of Optical Flow Networks	Conference on Computer Vision and Pattern Recognition (CVPR) 2022	19 - 24 Jun 2022
Continual Unsupervised Domain Adaptation for Semantic Segmentation using a Class-Specific Transfer	IEEE World Congress on Computational Intelligence (WCCI) 2022	18 - 23 Jul 2022
Contrastive Learning and Self-Training for Unsupervised Domain Adaptation in Semantic Segmentation	IEEE World Congress on Computational Intelligence (WCCI) 2022	18 - 23 Jul 2022

MotionMixer: MLP-based 3D Human Body Pose Forecasting	31st International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial Intelligence 2022	23 - 29 Jul 2022
Measuring the Effects of Catastrophic Forgetting in Neural Networks	Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory 2022	31 Jul - 05 Aug 2022
A Benchmark and a Baseline for Robust Multi-view Depth Estimation	International Conference on 3D Vision 2022,	12 - 15 Sep 2022
Verifying Safety of Safety-Critical Systems with Rare Events via Optimistic Optimization	19 th International Conference on Quantitative Evaluation of SysTems (QEST) 2022	12 - 16 Sep 2022
Using Network Architecture Search for Optimizing Tensor Compression	Workshop on IoT, Edge, and Mobile for Embedded Machine Learning (ITEM) 2022	19 - 23 Sep 2022
Hardware Execution Time Prediction for Neural Network Layers	Workshop on IoT, Edge, and Mobile for Embedded Machine Learning (ITEM) 2022	19 - 23 Sep 2022
Robustness under Domain Shift via Anisotropic Diffusion	Robust Understanding of Street Scenes Using Computer Vision 2022	22 - 23 Sep 2022
MGiaD: Multigrid in all Dimension	Robust Understanding of Street Scenes Using Computer Vision 2022	22 - 23 Sep 2022
Segment-wise Uncertainty Quantification for Self-Training for Domain Adaptation	Robust Understanding of Street Scenes Using Computer Vision 2022	22 - 23 Sep 2022
End-to-End Single Shot Detector using Graph-based Learnable Duplicate Removal	German Conference on Pattern Recognition (GCPR) 2022	27 - 30 Sep 2022
SF2SE3: Clustering Scene Flow into SE(3)-Motions via Proposal and Selection	German Conference on Pattern Recognition (GCPR) 2022	27 - 30 Sep 2022
Uncertainty Quantification and Resource-Demanding Computer Vision Applications of Deep Learning	11th The John von Neumann Institute for Computing Symposium in Jülich 2022	29 - 30 Sep 2022

A Benchmark for Unsupervised Anomaly Detection in Multi-Agent Trajectories	25th IEEE International Conference on Intelligent Transportation Systems 2022	08 Oct 2022
Improving Replay-Based Continual Semantic Segmentation with Smart Data Selection	25th IEEE International Conference on Intelligent Transportation Systems 2022	08 - 12 Oct 2022
Infusing Contextual Knowledge Graphs for Visual Object Recognition	International Semantic Web Conference (ISWC) 2022	23 Oct 2022
SpatialDETR: Robust Scalable Transformer-Based 3D Object Detection from Multi-View Camera Images with Global Cross-Sensor Attention	European Conference on Computer Vision (ECCV) 2022	23 - 27 Oct 2022
Gradient-based Uncertainty for Monocular Depth Estimation	European Conference on Computer Vision (ECCV) 2022	23 - 27 Oct 2022
KING: Generating Safety-Critical Driving Scenarios for Robust Imitation via Kinematics Gradients	European Conference on Computer Vision (ECCV) 2022	23 - 27 Oct 2022
An Unsupervised Domain Adaptive Approach for Multimodal 2D Object Detection in Adverse Weather Conditions	International Conference on Intelligent Robots and Systems (IROS) 2022	23 - 27 Oct 2022
On the calibration of underrepresented classes in LiDAR-based semantic segmentation	International Conference on Intelligent Robots and Systems (IROS), Workshop: Probabilistic Robotics In the age of Deep Learning 2022	27 Oct 2022
Semi-supervised domain adaptation with CycleGAN guided by a downstream task loss	34 th IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2022	31 Oct 2022
Structural Knowledge Distillation for Object Detection	Conference on Neural Information Processing Systems 2022	28 Nov - 09 Dec 2022
Two Video Data Sets for Tracking and Retrieval of Out of Distribution Objects	Asian Conference on Computer Vision (ACCV2022) 2022	04 - 08 Dec 2022
Causes of Catastrophic Forgetting in Class-Incremental Semantic Segmentation	Asian Conference on Computer Vision (ACCV2022) 2022	04 - 08 Dec 2022

Domain Adaptation and Generalization: A Low-Complexity Approach	Conference on Robot Learning (CoRL) 2022	14 - 18 Dec 2022
Heatmap-based Out-of-Distribution Detection	IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2023	03 - 07 Jan 2023
Auxiliary Task-Guided CycleGAN for Black-Box Model Domain Adaptation	IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2023	03 - 07 Jan 2023
Automated Detection of Label Errors in Semantic Segmentation Datasets via Deep Learning and Uncertainty Quantification	IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2023	03 - 07 Jan 2023
An Approach for Fusing Two Training-Datasets with Partially Overlapping Classes	18th IEEE International Conference on Semantic Computing 2023	01 - 03 Feb.2023
Severity of Catastrophic Forgetting in Object Detection for Autonomous Driving	International Conference on Pattern Recognition Applications and Methods (ICPRAM) 2023	22 - 24 Feb 2023
Exploring Navigation Maps for Learning-Based Motion Prediction	IEEE International Conference on Robotics and Automation 2023	29 May - 02 Jun 2023
RESET: Revisiting Trajectory Sets for Conditional Behavior Prediction	IEEE Intelligent Vehicles Symposium, Anchorage 2023	04 - 07 Jun 2023
LMR: Lane Distance-Based Metric for Trajectory Prediction	IEEE Intelligent Vehicles Symposium, Anchorage 2023	04 - 07 Jun 2023
Effects of Architectures on Continual Semantic Segmentation	IEEE Intelligent Vehicles Symposium, Anchorage 2023	04 - 07 Jun 2023
Survey on LiDAR Perception in Adverse Weather Conditions	IEEE Intelligent Vehicles Symposium, Anchorage 2023	04 - 07 Jun 2023
Augmentation-based Domain Generalization for Semantic Segmentation	IEEE Intelligent Vehicles Symposium, Anchorage 2023	04 - 07 Jun 2023

Category Differences Matter: A Broad Analysis of Inter-Category Error in Semantic Segmentation	Conference on Computer Vision and Pattern Recognition (CVPR) 2023	18 Jun 2023
Principles of Forgetting in Domain-Incremental Semantic Segmentation in Adverse Weather Conditions	Conference on Computer Vision and Pattern Recognition (CVPR) 2023	18 Jun 2023
Who breaks early, loses: goal oriented training of deep neural networks based on port Hamiltonian dynamics	International Conference on Artificial Neural Networks (ICANN), Kreta, Griechenland 2023	26 – 29 Sep 2023

5.2 Veröffentlichungen in Journals, Buchprojekt

Titel der Publikation	Journal	Datum der Veröffentlichung
Night-to-day: Online image-to-image translation for object detection within autonomous driving by night	IEEE Transactions on Intelligent Vehicles	07 Dec 2020
Cuepervision: Self-supervised learning for continuous domain adaptation without catastrophic forgetting	Image and Vision Computing	17 Dec 2020
A Survey on Visual Transfer Learning using Knowledge Graphs	Semantic Web Journal	15 Feb 2021
USIS: Unsupervised Semantic Image Synthesis	ArXiv	29 Sep 2021
Quantifying point cloud realism through adversarially learned latent representations	ArXiv / GCPR	29 Sep 2021
Lidar Upsampling with Sliced Wasserstein Distance	IEEE Robotics and Automation Letters	14 Okt 2022
MGiAD: Multigrid in all dimensions. Efficiency and robustness by coarsening in resolution and channel dimensions	ArXiv	10 Nov 2022
Towards Rapid Prototyping and Comparability in Active Learning for Deep Object Detection	ArXiv	21 Dez 2022
SCENE: Reasoning about Traffic Scenes using Heterogeneous Graph Neural Networks	IEEE Robotics and Automation Letters	08 Jan 2023
LU-Net: Invertible Neural Networks Based on Matrix Factorization	ArXiv	21 Feb 2023

Survey on Unsupervised Domain Adaptation for Semantic Segmentation for Visual Perception in Automated Driving	IEEE Access	14 Apr 2023
Identifying Label Errors in Object Detection Datasets by Loss Inspection	ArXiv	13 Mar 2023
Pixel-wise Gradient Uncertainty for Convolutional Neural Networks applied to Out-of-Distribution Segmentation	ArXiv	13 Mar 2023
Robust Mean Teacher for Continual and Gradual Test-Time Adaptation	ArXiv	22 Mar 2023
Detecting Novelties with Empty Classes	ArXiv	30 Apr 2023
Perception Datasets for Anomaly Detection in Autonomous Driving: A Survey	ArXiv	31 Mar 2023

Außerdem wurde aus KI Delta Learning ein Kapitel zu einem Buch beigesteuert, das hauptsächlich im Rahmen des Schwesterprojekt KI Absicherung heraus entstanden ist. Für **“Deep Neural Networks and Data for Automated Driving - Robustness, Uncertainty Quantification, and Insights Towards Safety”** (Springer, <https://link.springer.com/content/pdf/10.1007/978-3-031-01233-4.pdf>, Editoren: Fingscheidt, Gottschalk, Houben) wurde das Kapitel **“Detecting and Learning the Unknown in Semantic Segmentation”** von Mitarbeiterinnen und Mitarbeitern des KI Delta Learning Partners BUW erstellt.