Abteilung Signalverarbeitung und Machine Learning (Fingscheidt)

1. Forschungsfelder der Abteilung

Die Abteilung Signalverarbeitung und Machine Learning arbeitet mit Methoden des Deep Learning in den Forschungsfeldern Sprachverarbeitung und Computer Vision.

Im Bereich der Sprachverarbeitung erforschen wir Verfahren zur Störgeräuschreduktion, akustischen Echokompensation, künstlichen Sprach-Bandbreitenerweiterung, In-Car-Kommunikationssysteme sowie Strategien zum Training neuronaler Netze zu allen vorgenannten Algorithmen. Weitere Themen sind Beamforming sowie (neuronale) Nachfilter für höherqualitative, aber standardkonforme Sprach- und Audio*de*coder. Darüber hinaus befassen wir uns mit Emotionserkennung, multikanaliger Sprachaktivitäts-Erkennung und akustischer Event-Erkennung sowie mit der automatischen Spracherkennung und Informationsfusion.

Im Bereich der Computer Vision forschen wir im Wesentlichen an Perzeptionsmethoden für das autonome Fahren. Dazu gehören semantische Segmentierung, Tiefenschätzung, gelernte Bildkompression, Corner-Case-Detektion sowie Fragestellungen des Domain Transfers und adversarialer Angriffe.

Wir arbeiten im Kontext von Fahrzeug-, Office- und Consumer-Anwendungen im Bereich von Smartphones, Hörgeräten/Cochlea-Implantaten, Freisprechsystemen, Überwachungs- bzw. Produktionstechnologien bis hin zum autonomen Fahren.

2. Projekte

Unsere erfolgreichen Forschungsarbeiten im Bereich der iterativen Turbo-Informationsfusion in der automatischen Spracherkennung werden seit diesem Berichtsjahr von der DFG durch eine Sachbeihilfe gefördert (DFG-Kennzeichen FI 1494/6-1). Im Rahmen dieser zweijährigen DFG-Einzelförderung sollen weitere Möglichkeiten zur Informationsfusion untersucht werden und die Fusion von statistisch abhängigen Emissionen durch theoretische Analysen untersucht werden.

Das gut zweijährige Kooperationsprojekt MindMarker wurde zusammen mit der Firma Mind Intelligence UG, Berlin, beim Bundesministerium für Bildung und Forschung (BMBF) eingeworben und gestartet. Ziel des Projekts ist es, ein System zur Detektion von Depressionen anhand der Stimme eines Anrufers zu ent-

wickeln. Der Beitrag des IfN liegt in einer dafür notwendigen robusten Emotionserkennung.

In diesem Jahr startete das dreijährige Projekt KI-Absicherung, welches vom Bundesministerium für Wirtschaft und Energie (BMWi) gefördert wird. In diesem dreijährigen Förderprojekt werden Ansätze zur Erhöhung der Robustheit von Methoden der künstlichen Intelligenz gegenüber Veränderungen der Eingangswerte im Kontext des autonomen Fahrens erforscht. Das IfN ist über einen Unterauftrag der Volkswagen Group Innovation eingebunden.

Ebenfalls neu gestartet wurde ein mehrjähriges Kooperationsprojekt mit IAV GmbH zur KI-basierten, automatisierten Fehlergeräuscherkennung in der Fahrzeugdiagnose. Mit dem Schwerpunkt auf der Erkennung akustischer Events sollen dabei fehlerhafte Zustände eines Fahrzeugs durch Machine-Learning-Methoden basierend auf akustischen Sensoren (z.B. Freisprechmikrofone) detektiert und klassifiziert werden.

In einer einjährigen Vertikalstudie, beauftragt von der Volkswagen Group Innovation, wurden Teacher-Student-Netzwerke wie auch neuartige Netztopologien und Netzkonfigurationen auf ihre Möglichkeiten zur Erhöhung der Robustheit gegenüber adversarialen Angriffen erforscht.

Das mit der Volkswagen Group Innovation weitergeführte Kooperationsprojekt zu sog. generativen adversarialen Netzwerken (GANs) im automatischen Fahren thematisierte wie schon im Vorjahr gelernte Bildkompressionsverfahren. Es wurden verschiedene Methoden zur Perzeption mit verteilten Sensornetzwerken, Datenübertragung im Vehicle-to-anywhere (V2X)-Kontext und zur generellen Effizienzsteigerung z.B. durch Vektorquantisierung entwickelt.

Weitergeführt wurden – ebenfalls in Kooperation mit der Volkswagen Group Innovation – unsere methodischen Ansätze zur Corner-Case-Detektion. Dabei handelt es sich um Online- und Offline-Ansätze zum Auffinden von kurzen Bildsequenzen, bei denen kritische bzw. seltene Ereignisse in einem Verkehrsszenario zu sehen sind. Neben der Entwicklung eines Frontends und eines Tools sind grundlegende Arbeiten zum Domain Transfer gelungen, damit das Tool auf unterschiedlichen Datenbanken lauffähig ist.

In einem Förderprojekt des Landes Niedersachsen (NBank) arbeiteten wir nun im zweiten Jahr gemeinsam mit der Firma Pan Acoustics GmbH aus Wolfenbüttel an einem drahtlosen Mikrofonarray. Nachdem im ersten Projektjahr maßgeblich adaptives Beamforming für das Mikrofonmodul im Fokus stand, lag der Fokus in diesem Projektjahr auf niedriger Verzögerungszeit für das Gesamtsystem und einer Soft-Audio-Decodierung für die Funkstrecke.

Mit der R&D-Gruppe der Firma NXP Semiconductors, Product Line Voice and Audio Solutions, Belgien, wurden im Berichtszeitraum Forschungsarbeiten zur Störgeräuschreduktion für Sprachsignale weitergeführt. Der Fokus des Vorhabens im Berichtsjahr lag auf der Erforschung mehrstufig aufgebauter Systeme basierend auf tiefen neuronalen Netzen sowie auf der Betrachtung verschiedener neuartiger Modelltopologien und Fehlerfunktionen für das Training der neuronalen Netze. Ein zweites Projekt in diesem Themenbereich wurde im Juli dieses Jahres aufgenommen. Ziel der Forschungsarbeiten ist es, die im ersten Projekt erarbeiteten Modelle im Hinblick auf Parameter und Rechenkomplexität zu reduzieren.

Zu einem erfolgreichen Abschluss geführt wurde im Berichtszeitraum das BMBF-geförderte Projekt „KI-Plattform-Konzept“. Gemeinsam mit zahlreichen renommierten Partnern aus Industrie und Forschungsinstitutionen wurde die Konzeptionierung einer nationalen institutionsübergreifenden Datenplattform für die Entwicklung von Methoden der künstlichen Intelligenz (KI) beim Einsatz für das autonome Fahren erstellt, die in einem geplanten Nachfolgeprojekt implementiert und betrieben werden soll. Die Beiträge des IfN umfassten die automatische (Video-)Datenselektion, Bild- und Video-Kompression sowie Aspekte einer Audio-Datenakquisition.

Im Rahmen des Förderprogrammes Zentrales Innovationsprogramm Mittelstand (ZIM) des Bundesministeriums für Wirtschaft und Energie (BMWi) wurde im Berichtszeitraum das Projekt „Erkennung von Emotionen in akustischen Signalen“ mit der Firma eye square GmbH in Berlin erfolgreich abgeschlossen. Innerhalb des Projekts wurde ein Expertensystem zur echtzeitigen und dynamischen Beratung von Nutzern von Online-Handelsplattformen auf Basis der Analyse ihrer Präferenzen durch Verhaltensmuster und Erkennung von Emotionen in lautsprachlichen Äußerungen entwickelt.

Das auch vom BMWi geförderte ZIM-Projekt “HIFI-AEC“ mit der Firma Linguwerk GmbH in Dresden, in dessen Rahmen die Sprachbedienung bei HiFi- und Infotainment-Systemen durch eine am Institut für Nachrichtentechnik entwickelte akustische Echokompensation für Stereo-Signale aufgewertet wurde, ist ebenfalls erfolgreich abgeschlossen worden.

Weiterhin abgeschlossen wurde das vom BMBF unterstützte zweijährige Projekt zur Förderung von Qualifizierungsmaßnahmen und Forschungsvorhaben im Bereich Maschinelles Lernen im Rahmen des Förderprogramms „IKT 2020 – Forschung für Innovationen“. Im Berichtszeitraum hat das zweite *Deep Learning Lab* mit insgesamt 30 Studierenden stattgefunden. Insgesamt konnte die Qualität der Lehr- sowie der Abschlussveranstaltung gegenüber dem ersten Durchgang deutlich gesteigert werden. Ein Sonderbericht zu der Abschlussveranstal-

tung, die durch Sponsoren aus der Industrie unterstützt wurde, befindet sich auf Seite 128.

Eine vom China Scholarship Counsil (CSC) unterstützte Forschungsarbeit zur Sprachdecodierung mit verbesserter Qualität wurde zu Ende geführt. Die entwickelten maschinengelernten Verfahren können standardkompatibel deutliche Qualitätsverbesserung bei einer Vielzahl von Sprach(de)codierern erzielen.

Das vom Niedersächsischen Ministerium für Wissenschaft und Kultur geförderte dreijährige Promotionsprogramm „Konfigurationen von Mensch, Maschine und Geschlecht – Interdisziplinäre Analysen zur Technikentwicklung“, kurz „KoMMa.G“, kam im Berichtszeitraum zu einem Ende. Vom IfN sind hier Arbeiten zum Thema Sprachverbesserung und Emotionserkennung für die automatische Sprachanalyse von Team-Meetings erfolgt.

3. Mitarbeiterinnen und Mitarbeiter der Abteilung

Das Forschungsfeld der Computer Vision ist aufgrund der Vielzahl KI-bezogener Projekte weiter auf Expansionskurs. Zu den bereits länger bei uns aktiven Herren Bär, Bolte und Löhdefink sind neu dazu gestoßen Marvin Klingner (seit 01.03.2019) und Jasmin Breitenstein (seit 01.10.2019). In der Sprachverarbeitung forschen weiterhin die Herren Elshamy, Franzen, Lohrenz, Meyer, Strake, Xu und Zhao. Ebenfalls neu an Bord ist Jan Baumann (seit 01.04.2019) im Bereich der Erkennung von akustischen Events. Damit arbeiteten zum Ende des Berichtszeitraums in der Abteilung Signalverarbeitung und Machine Learning neben Prof. Fingscheidt und Frau Erichsen-Rua 13 Wissenschaftler/innen mit. Bei der Volkswagen Group Innovation betreut Prof. Fingscheidt vier weitere Doktorand/innen im Forschungsfeld Vision: Antonia Breuer, Nikhil Kapoor, John Serin Varghese, Christopher Plachetka. Im Berichtszeitraum haben bei uns sieben Studierende eine Masterarbeit und drei Studierende eine Bachelorarbeit bzw. Projektarbeit abgeschlossen. Weiterhin hat uns noch eine Vielzahl studentischer Hilfskräfte unterstützt.

4. Sprachverbesserung

4.1 Störgeräuschreduktion

Algorithmen zur Störgeräuschreduktion sind notwendig, um gute und verständliche Telekommunikation zu ermöglichen, unabhängig vom Umfeld der Teilnehmer. Am Institut wird daher aktiv auf unterschiedlichen Ebenen an der Verbesserung solcher Verfahren geforscht. Ein wichtiger Bestandteil solcher Verfahren sind Gewichtungsfunktionen, die letztlich im Frequenzbereich das Störgeräusch

unterdrücken sollen. Diese können durch traditionelle Verfahren statistisch oder auch direkt durch moderne neuronale Netzwerke geschätzt werden.

Im Berichtszeitraum hat sich Herr Elshamy weiterhin mit der Schätzung der A-Priori-Signal-to-Noise Ratio (A-Priori-SNR) beschäftigt, welche ein wichtiger Bestandteil der traditionellen, statistischen Störgeräuschreduktion ist. Für die Schätzung des A-Priori-SNRs wurde hierbei zunächst das bereits bestehende Verfahren zur Manipulation des Sprachanregungssignals von 2017 durch den Einsatz von neuronalen Netzwerken deutlich verbessert. Die Ergebnisse wurden als Zeitschriftenaufsatz in der November-Ausgabe der IEEE/ACM Transactions on Audio, Speech, and Language Processing publiziert [ELS/FIN1]. Es wurden Untersuchungen zu unterschiedlichen Features sowie zur Normalisierung von Zielvektoren durchgeführt. Das Verfahren wurde anschließend in Kombination mit dem 2018 publizierten Ansatz zur Verbesserung der Sprach-Einhüllenden ausgewertet und konnte auch dort eine stark verbesserte Dämpfung des Störgeräusches erreichen, bei überwiegend gleichbleibender Qualität der Sprachkomponente.

Als Variante wurde eine weitere Version entwickelt, die die Qualität der Sprachkomponente besonders bei schlechtem SNR objektiv verbessert. Dies wurde erzielt bei gleichzeitiger Verbesserung der Dämpfung. Zusätzlich wurde ein subjektiver Hörtest durchgeführt, der zeigte, dass das neue Verfahren von den Hörern gegenüber einem traditionellen deutlich bevorzugt wird. Die Ergebnisse wurden auf dem IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York mit einem Posterbeitrag publiziert [ELS/FIN2].

Herr Xu entwickelte ein maskenbasiertes Sprachverbesserungs-Framework, bei dem identische tiefe neuronale Netze (Deep Neural Networks, DNNs) in Reihe miteinander verkettet werden (sog. Concatenated Identical DNNs, CI-DNNs). Die Idee ist, dass die Sprache durch identische DNNs stufenweise verbessert wird, um trainierbare Parameter zu sparen und auch ein Gleichgewicht zwischen Sprachkomponentenqualität und Rauschunterdrückung zu erreichen. Ein besonderes Ergebnis ist, dass nicht-stationäre Störgeräusche mit dieser Methode ein subjektiv angenehmes Restgeräusch ergeben. Die Ergebnisse wurden in einer Publikation auf der 2019 European Signal Processing Conference (EUSIPCO) vorgestellt [XU/STRA/FIN1].

Um während des Trainings von neuronalen Netzen zur maskenbasierten Sprachverbesserung die Erhaltung der Sprachkomponenten-Qualität, die Unterdrückung der Restrausch-Komponente und die Erhaltung einer natürlich klingenden Restrausch-Komponente voneinander getrennt zu kontrollieren, wurde ein neuartiger Komponentenfehler (engl. Component Loss, CL) vorgeschlagen. Die Ergebnisse dieser neuartigen Methode wurden in Form eines Zeitschriften-

artikels „Components Loss for Neural Networks in Mask-Based Speech Enhancement“ bei den IEEE/ACM Transactions on Audio, Speech, and Language Processing eingereicht und als arXiv-Preprint hochgeladen [XU/ELS/ZHA/FIN1].

Herr Zhao arbeitete an modernen subjektiv motivierten Fehlerfunktionen für neuronale Netze in der Sprachverbesserung. Um die menschliche Wahrnehmung in die Sprachverbesserung einzubeziehen, wird anstelle der herkömmlichen Fehlerfunktion, die den mittleren quadratischen Fehler minimiert, ein Wahrnehmungs-basiert gefiltertes Fehlersignal der Fehlerfunktion während des Trainings des neuronalen Netzwerks zugrunde gelegt. Die sich ergebende vorgeschlagene Fehlerfunktion wird durch ein aus der Sprachcodierung mit code-excited linear prediction (CELP) bekanntes Gewichtungsfilter motiviert. Die experimentellen Ergebnisse zeigten, dass die neue Fehlerfunktion in Bezug auf die Wahrnehmungsqualität der verbesserten Sprache eine bessere Leistung im Vergleich zur herkömmlichen Fehlerfunktion hat. Diese Arbeit wurde auf dem IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York [ZHA/ELS/FIN1] veröffentlicht.

Herr Strake führte das Projekt zum Thema Störgeräuschreduktion für Mobiltelefone in Kooperation mit der R&D-Gruppe der Firma NXP Semiconductors, Product Line Voice and Audio Solutions, Belgien im nunmehr insgesamt fünften Projektjahr fort. Im Berichtszeitraum lag der Fokus der Arbeiten auf der Erforschung zweistufiger Systeme basierend auf neuronalen Netzen und der Untersuchung von Fehlerfunktionen für das Netztraining im Frequenz- sowie Zeitbereich. Es wurde ein Verfahren entwickelt, bei welchem zunächst die Störgeräuschunterdrückung mit Hilfe eines long short-term memory (LSTM)-Netzwerks durchgeführt wird, wobei die rekurrente Netzstruktur im Vergleich zu traditionellen Verfahren eine deutlich verbesserte Unterdrückung nichtstationärer Störgeräusche ermöglicht. Da die Sprachqualität durch eine solche erste Stufe beeinträchtigt werden kann, findet in der zweiten Stufe ein weiteres, auf direkter Schätzung des störungsfreien Sprachsignal-Spektrums mittels Faltungsnetzen (engl. convolutional neural networks, CNNs) basierendes Modell Anwendung, welches in der Lage ist, unterdrückte Sprachanteile wiederherzustellen. Es konnte gezeigt werden, dass durch die Anwendung einer solchen zweiten Stufe eine deutlich verbesserte Sprachqualität sowie eine stärkere Unterdrückung der Störgeräusche erreicht werden kann. Diese Ergebnisse wurden auf dem IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York publiziert [STRA/FIN1] und fanden großes Interesse.

4.2 Echokompensation / In-Car-Kommunikation

Sprachverbesserung in Fahrzeugen erforschen wir aktuell im Bereich der akustischen Echokompensation (engl. acoustic echo cancellation, AEC) für Freisprech-Telefonie und In-Car-Kommunikationssysteme (ICC-Systeme) zur Verbesserung von Konversationen zwischen Passagieren innerhalb einer Fahrzeugkabine. In beiden Anwendungsfällen wird die akustische Rückkopplung von Fahrzeug-Lautsprechern in das Fahrzeug-Mikrofon geschätzt und unterdrückt. Der robuste AEC-Algorithmus, den wir hierfür nutzen, ist das frequenzbasierte Kalman-Filter (engl. frequency domain adaptive Kalman filter, FDAKF). Im Kontext von ICC-Systemen wird der FDAKF als akustische „Feedback"-Unterdrückung (engl. acoustic feedback cancellation, AFC) genutzt. In diesem Kontext wurde die Publikation [FRA/FIN1] auf der DAGA in Rostock vorgestellt. Hierbei wurde untersucht, welche Auswirkung die geschätzte Übertragungsfunktion des FDAKF auf ein ICC-System hat. Die Übertragungsfunktion kann als Ganzes oder aber in unterschiedlich vielen Unterteilungen (Partitionierung) betrachtet werden. Eine gut gewählte Partitionierung kann die Komplexität des Algorithmus deutlich reduzieren. Weiterhin wurde eine verbesserte Methode zur Schätzung des sogenannten Messrauschens für den FDAKF entwickelt. Die Methode ist explizit für den Einsatz in AFC-Systemen ausgelegt. Sie ermöglicht eine stärkere Unterdrückung des Störsignals und verbessert gleichzeitig die Qualität des gewünschten Zielsignals erheblich. Die Ergebnisse wurden auf der International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in Brighton, Vereinigtes Königreich, vorgestellt [FRA/FIN2].

Nachdem Herr Abel das Institut im letzten Jahr verlassen hat, haben die Herren Meyer und Lohrenz das zweite Projektjahr des ZIM-Projekts HIFI-AEC zusammen mit der Firma Linguwerk GmbH aus Dresden bestritten und im Berichtszeitraum erfolgreich abgeschlossen. Im Rahmen dieses Projekts wurde eine Stereo-Echokompensation (engl. stereo acoustic echo cancellation (SAEC)) für den automotiven Bereich entwickelt, um auch bei laufendem Radio eine robuste Spracherkennung zur Bedienung eines Infotainmentsystems im Kfz zu ermöglichen. Das entwickelte Gesamtsystem, bestehend aus SAEC und einem weiteren Modul zur Unterdrückung des Restechos eliminiert dabei zuverlässig das Radiosignal aus dem Eingangssignal der Spracherkennungs-Software. Im zweiten Projektjahr lag der Fokus auf der Implementierung des Gesamtsystems in C++, nachdem dieses im letzten Jahr bereits in Form von MATLAB-Code an Linguwerk übergeben wurde. Außerdem wurden realistischere Anwendungsfälle wie die Hinzunahme von Fahrgeräuschen simuliert und erfolgreich getestet. Die Ergebnisse der Arbeiten wurden in einem Abschlussbericht [MEY/LOH/FIN1] festgehalten.

4.3 Beamforming

Im Rahmen des NBank-Förderprojektes mit der Firma Pan Acoustics wurde die Entwicklung eines adaptiven Beamformers erfolgreich abgeschlossen. Der Beamformer basiert auf einem zirkulären Mikrofonarray mit 13 Mikrofonen, folgt einem differentiellen Entwurf mit weitreichend frequenzunabhängiger Richtwirkung und kann die akustische Ausrichtung (Richtkeule) des verwendeten Mikrofonmoduls ganze 360° um das Modul herum ausrichten. Da sich mehrere Sprecher um das Mikrofonmodul befinden können, in der Regel jedoch nur einer davon aktiv spricht, ist die Algorithmik des Beamformers so konzipiert, dass sich die Richtkeule selbstständig auf den gerade aktiven Sprecher ausrichtet. Dies geschieht jedoch nur dann, wenn durch den Sprecher auch ein bestimmter Energie-Schwellenwert überschritten wird. Dadurch werden kurze, ungewünschte Störgeräusche (wie zum Beispiel das Herunterfallen eines Stiftes auf die Tischfläche) nicht zusätzlich verstärkt und es kann eine gute Sprachqualität gewährleistet werden. Weitere Details finden sich im Projekt-Zwischenbericht [FRA/ZHA/FIN1].

4.4 Künstliche Sprachbandbreiten-Erweiterung

Das Themenfeld der künstlichen Sprachbandbreitenerweiterung (artificial bandwidth extension, ABE) geht am IfN auf die Zielgerade. Zwei wichtige Publikationen sind nach Weggang von Johannes Abel im Berichtszeitraum noch zu verzeichnen.

Zum einen konnte in Kooperation mit der Arbeitsgruppe Auditorische Prothetik des Exzellenzclusters „Hearing4all“ an der Medizinischen Hochschule Hannover (MHH), Prof. Dr.-Ing. Waldo Noguiera, der Einsatz von ABE für hörgeschädigte Menschen mit Cochlea-Implantat (CI) im Kontext der Telefonie untersucht werden. Aufgrund von physikalischen Einschränkungen kann die Cochlea durch den implantierten Elektrodenstrang nur an wenigen (etwa 10-20) Orten zeitgleich angeregt werden, so dass nach der Orts-Frequenz-Transformation in der Cochlea nur eine sehr grobe Frequenzauflösung über den Hörnerv an das Gehirn weitergegeben werden kann. Erschwerend kommt hinzu, dass viele Telefonsignale eine beschränkte akustische Bandbreite aufweisen, so dass nur ein Teil der Elektroden während eines Telefonats angesteuert wird. In den gemeinsamen Untersuchungen wurde das Telefonsignal mittels ABE in der akustischen Bandbreite erweitert und so ermöglicht, dass mehr Elektroden angesteuert werden. In subjektiven Tests mit neun CI-Nutzern konnte eine statistisch signifikante Erhöhung der Sprachqualität sowie eine Verbesserung der Sprachverständlichkeit um über 17 % durch das ABE-Verfahren gezeigt werden. Die Arbeit wurde im Journal of the Acoustical Society of America veröffentlicht [FIN1].

In Telefonsignalen sind tiefe Sprecher-Frequenzen unterhalb von 300 Hz oft nicht mehr enthalten oder stark gedämpft. Abhilfe schafft eine künstliche Erweiterung in eben diesen Frequenzbereich. Bei einer solchen Bandbreitenerweiterung kommt es vornehmlich auf eine exakte und instantane Schätzung der Sprachgrundfrequenz des Sprechers an. Mittels eines ausgeklügelten Zustandsmodells werden die fehlenden Frequenzkomponenten dann rahmenweise synthetisiert. In einem subjektiven Hörtest konnte so in einem direkten Vergleich von ABE mit und ohne Erweiterung im unteren Band eine Verbesserung der Sprachqualität um 0,26 CMOS-Punkte erreicht werden (CMOS: Comparison mean opinion score). Der dazugehörige Journal-Artikel wurde bei den IEEE/ACM Transactions on Audio, Speech, and Language Processing veröffentlicht [FIN2].

4.5 Sprach(de)codierung

Im Bereich der Sprach(de)codierung veröffentlichte Herr Zhao einen Zeitschriftenartikel in den IEEE/ACM Transactions on Audio, Speech, and Language Processing [ZHA/FIN1]. In diesem Artikel wird ein Postfilter mit faltenden neuronalen Netzwerken (CNNs) vorgeschlagen, um codiert übertragene Sprachsignale für verschiedene Schmalband- wie auch Breitband-Codecs zu verbessern. Das Verfahren liefert derart gute Ergebnisse, dass es die TU Braunschweig als Weltpatent angemeldet hat. Darüber hinaus wurde eine Gemeinschaftsarbeit mit dem Institut für Analysis und Algebra auf der International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in Brighton, UK, veröffentlicht [ZHA/FIN2]. Ziel dieser Arbeit war es, niedrigratig quantisierte Sprache zu rekonstruieren. Dabei wurde eine besondere Form der Netzwerke, sogenannte primal-dual-Netzwerke, angewendet. Mit der Nutzung des Gewichtungsfilters (siehe 4.1) zur Berechnung der Fehlerfunktion wird das Netzwerk zusätzlich Wahrnehmungs-effizienter trainiert.

In der Sprachcodierung mit generativen adversarialen Netzwerken (GANs) wurde eine Masterarbeit erstellt [MA 19/003]. In dieser Arbeit werden Sprachcodierer und -decodierer unter Verwendung von GAN-basierten Frameworks entwickelt und verschiedene Netzwerktopologien sowie Verlustfunktionen untersucht und auf ihre Tauglichkeit für Sprachcodierung hin verglichen.

Im Rahmen des NBank-Förderprojektes mit der Pan Acoustics GmbH arbeitete Herr Zhao an einer intelligenten Soft-Audio-Decodierung für eine Funkstrecke. In Matlab wurde hierfür eine echtzeitfähige Sprachübertragungsplattform im 2,4 GHz ISM-Band mit Software-Defined-Radio-Geräten aufgebaut. Weitere Experimente zur Soft-Audio-Decodierung am Empfänger werden auf dieser Plattform durchgeführt. Darüber hinaus wird ein Signal-zu-Rauschleistungs-Schätzer (SNR-Schätzer) auf Basis der Empfangspegelwerte der verfügbaren

ISM-Chipsätze entwickelt. Der SNR-Schätzwert wird dann im Zuge der Soft-Audio-Decodierung verwendet.

4.6 Automatische Spracherkennung

Ein Kernforschungsgebiet im Bereich der automatischen Spracherkennung ist die Turbo-Informationsfusion, die es ermöglicht, mittels eines iterativen Austauschs von probabilistischer Information die Erkennungsgenauigkeit von zwei oder mehreren Erkennern zu verbessern.

Zu Beginn des Berichtszeitraums im Dezember haben wir mit der Publikation eines Artikels auf dem IEEE Workshop on Spoken Language Technology (SLT) in Athen eine wichtige neue internationale Bestmarke auf der TIMIT-Datenbank zur Phonemerkennung für kontextunabhängige und sprecherunabhängige akustische Modellierung erzielt. Dies wurde durch die Nutzung moderner hierarchischer Faltungsnetze zur akustischen Modellierung der einzelnen Amplituden- und Phaseninformation und der darauf folgenden Turbo-Informationsfusion erreicht [LOH/FIN1].

Durch die erfolgreiche Beantragung einer DFG-Sachbeihilfe innerhalb des Berichtszeitraums können die Arbeiten im Bereich der Turbo-Informationsfusion in der zweijährigen Projektlaufzeit deutlich intensiviert werden. Erste Ergebnisse im Rahmen des DFG-Forschungsvorhabens konnten in einer Masterarbeit [MA 19/005] erbracht werden, in der die Turbo-Fusion erstmalig erfolgreich auf die Nutzung von beliebig vielen Erkennern erweitert wurde. Zuvor wurden ausschließlich zwei Komponenten-Erkenner genutzt. Dies ermöglicht insbesondere die Anwendung der Turbo-Fusion für mehrere im Raum verteilte Erkenner. In ersten Experimenten konnte bereits gezeigt werden, dass durch die Nutzung unterschiedlicher Impulsantworten ein Fusionsgewinn erzielt werden kann.

Zur Erweiterung der für verteilte Spracherkennung notwendigen Modularität wurde außerdem erforscht, inwieweit bidirektionale neuronale Netze in der Lage sind (ähnlich wie der bisher genutzte Turbo-FBA), zeitlich unbegrenzte zeitliche Information in die Erkennung mit einzubeziehen. Dabei arbeiten diese BLSTMs mit Phonemwahrscheinlichkeiten sowohl am Eingang als auch am Ausgang und erlauben damit sehr flexible Verschaltungen, zukünftig hoffentlich auch im Sinne einer iterativen Turbo-Decodierung. Überraschende Erkenntnisse, die sich durch die Nutzung solcher BLSTMs ergeben haben, wurden in Form einer Publikation für den IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU) in Singapur angenommen.

Im Bereich der mehrkanaligen robusten automatischen Spracherkennung auf Grundlage der vierten CHiME-Challenge (CHiME: Computational Hearing in Multisource Environments) betreute Herr Strake bereits während des vorhe-

rigen Berichtszeitraums eine Masterarbeit zur Verbesserung des verwendeten akustischen Modells. Durch die Nutzung eines neuartigen dicht verbundenen Faltungsnetzes (engl. densely connected neural network, DenseNet) und die Kombination mit einem rekurrenten Modellanteil, sowie eines speziellen, auf ganzen Sätzen basierenden Trainings, konnte ein hochperformantes Modell entwickelt werden. Die zugehörigen Ergebnisse wurden bereits auf dem 2018 IEEE Workshop on Spoken Language Technology (SLT) in Athen mit einer Posterpräsentation publiziert [STRA/LOH/FIN1].

4.7 Emotionserkennung

Das ZIM-Projekt „Erkennung von Emotionen aus akustischen Signalen“ wurde von Herrn Meyer in Kooperation mit der Firma eye square GmbH aus Berlin Anfang diesen Jahres erfolgreich abgeschlossen. Entwickelt wurde ein System zur Emotionserkennung, welches auf modernen Methoden des maschinellen Lernens basiert. Das Besondere an dem System ist, dass die Erkennung der Emotionen auf der Analyse eines 2-dimensionalen Log-Mel-Spektrogramms basiert, anstelle von tausenden speziell aus den Sprachsignalen extrahierten Merkmalen, wie es sonst üblich ist. Das entwickelte System stellt dadurch im Grunde einen Bildklassifikator für Bilder unterschiedlicher Breite dar und kann bis zu sechs Emotionen aus der menschlichen Stimme erkennen. Dabei schließt der Erkenner auf einer renommierten Datenbank mit dem derzeitigen internationalen Spitzenreiter auf. Die Resultate und Erkenntnisse aus diesem Projekt wurden in einem Abschlussbericht [MEY/FIN1] dokumentiert. Außerdem ist auf Basis der Resultate noch eine Publikation für nächstes Jahr vorgesehen.

Nach Abschluss der Graduiertenschule KoMMa.G im Herbst 2019 arbeitet Herr Xu nun mit Herrn Meyer im neu gestarteten BMBF-Förderprojekt „MindMarker“ in Kooperation mit der Firma Mind Intelligence UG aus Berlin. In diesem Projekt soll ein automatisches System zur Depressionserkennung auf Basis der Stimme eines Telefonnutzers entwickelt werden, welches zusätzlich einen Sprachroboter umfasst, um Nutzern per Telefon oder App automatisch relevante psychologische Fragen zu stellen. Mit Hilfe des künstlichen Sprachdialogs können persönliche Sprachäußerungen erhoben werden, um die Depressions-Detektion anhand der Stimme durchführen zu können. Damit dieser Dialog erfolgreich verläuft, soll die Auswahl der Fragen intelligent auf Basis der emotionalen Stimmung der Nutzer beruhen. Das IfN wird in diesem Zusammenhang einen sprachbasierten Emotionserkenner sowie ein Modul zur Verbesserung der Sprachqualität entwickeln, da diese je nach Aufenthaltsort der Nutzer stark gestört sein kann.

4.8 Akustische Diagnose in Fahrzeugen

In Zusammenarbeit mit IAV GmbH wurde ein Kooperationsprojekt „AcuDia“ gestartet, in dessen Rahmen ein System entwickelt werden soll, welches als KI-basierte, automatisierte Fehlergeräuscherkennung in der Fahrzeugdiagnose eingesetzt werden soll. Dadurch soll mit Hilfe von akustischen Signalen eine Diagnose von Fahrzeugen während des Betriebs erfolgen. Im Betrieb befindliche Fahrzeuge und ihre Komponenten geben Geräusche ab, welche sowohl vom Betriebszustand als auch vom Zustand der Komponenten abhängig sind. Ein aufmerksamer Fahrer, der das normale Betriebsgeräusch seines Fahrzeugs kennt, wird durch solche Geräusche oftmals aufmerksam, bevor ein Totalausfall eines Bauteils oder gar des Fahrzeugs eintritt. In Ergänzung dazu kann ein Servicemitarbeiter mit Kenntnis über diverse Fehlerzustände an unterschiedlichen Fahrzeugen und Fahrzeugtypen anhand einer Testfahrt eines fehlerhaften Fahrzeuges aus dem Betriebsgeräusch verschiedene Fehlertypen erkennen und oft auch deren Quelle lokalisieren.

Im Themenfeld der Detektion und Klassifizierung akustischer Events (acoustic event detection, AED) sind neuronale Netze imstande, Erkennungsraten vergleichbar mit der menschlichen Performanz zu erzielen. Im Projekt AcuDia soll nun die Expertise eines erfahrenen Servicemitarbeiters mit neuronalen Netzen nachgebildet werden, indem ein Klassifikator mit Audioaufnahmen aus Betriebsgeräuschen mit und auch ohne Fehlerereignis von Fahrzeugen trainiert wird. Damit soll ein Diagnosesystem realisiert werden, welches während des Fahrzeugbetriebs Informationen zum Zustand der Fahrzeugkomponenten liefert und vor drohenden Ausfällen warnt.

Da die Akquise von ausreichenden Datenmengen für einen Fahrzeuggeräusch-Datensatz zeitgleich abläuft, wurde zu Beginn des Projektes ein akustischer Event-Erkenner auf Audiodaten einer allgemeinen Geräusch-Datenbank umgesetzt. Der Erkenner wurde mit gemischten Audiodaten aus nicht-stationärem Hintergrundgeräusch und störungsfreien Events trainiert und getestet. Ein Visualisierungstool ist entstanden, und es sind geeignete Qualitätsmetriken entwickelt worden. Im weiteren Projektverlauf sollen erst das Hintergrundgeräusch und später, sobald geeignete Daten in ausreichender Menge vorhanden sind, auch die Events gegen Fahrzeugaufnahmen ausgetauscht werden, um so graduell ein System für den realen Einsatz im Fahrzeug zu erhalten.

5. Computer Vision

5.1 Robuste Semantische Segmentierung und Tiefenschätzung

Das Themenfeld „semantische Segmentierung“ hat sich mittlerweile zu einer der Kernkompetenzen in unserem Forschungsgebiet Computer Vision entwi-

(a) Eingangsbild (b) Semantische Segmentierung

Abbildung 11: Qualitative Ergebnisse des im IfN entwickelten Netzwerks zur semantischen Segmentierung; im Bild rechts entsprechen bestimmte Farben (Grauwerte) bestimmten Objektklassen.

ckelt. Dies zeigt sich an diversen Publikationen, die wir im Berichtszeitraum auf internationalen Tagungen vorgestellt haben [BAE/FIN1], [BOL/FIN1], [BOL/BAE/FIN1], [LOE/BAE/FIN1]. Ein Beispiel zur Performanz des IfN-Segmentierungsnetzwerks ist in **Abbildung 11** zu sehen.

Des Weiteren haben wir das Themenfeld um Teacher-Student-Ansätze erweitert. Dabei wird ein kleineres Netzwerk (Student) von einem größeren Netzwerk (Teacher) trainiert. Das IfN konzentriert sich hier auf Ansätze zur Erhöhung der Robustheit gegenüber sogenannten adversarialen Angriffen – Szenarien, in denen das Eingangsbild von einem potentiellen Angreifer bewusst verändert wird, um das jeweilige neuronale Netzwerk zu täuschen. Eine erste Veröffentlichung zu dieser Thematik hat Herr Bär auf dem Safe Artificial Intelligence for Autonomous Driving (SAIAD) Workshop der Conference on Computer Vision und Pattern Recognition (CVPR) 2019 in Long Beach, USA, vorgestellt. Es trägt den Titel „On the Robustness of Redundant Teacher-Student Frameworks for Semantic Segmentation“ [BAE/FIN1]. Dabei wird ein Teacher-Student-Paar, bestehend aus einem statischen (d.h. mit einmalig trainierten Parametern) Teacher-Netzwerk und einem parametrisch adaptiven Student-Netzwerk, um ein parametrisch statisches Student-Netzwerk erweitert. Mithilfe einer neu entworfenen Kostenfunktion wird dabei das adaptive Student-Netzwerk (gedacht als z.B. im Fahrzeug online lernendes Netzwerk) dazu gezwungen, Merkmale zu extrahieren, die sich von den extrahierten Merkmalen des statischen Student-Netzwerks unterscheiden. Das führt dazu, dass mit diesem Netz-Triplet Mehrheitsentscheidungen getroffen werden können, die robust gegenüber adversarialen Angriffen sind.

Neben der semantischen Segmentierung entwickelte das IfN nun auch eine Instanz-Segmentierung, die in einer studentischen Arbeit [MA 19/011] imple-

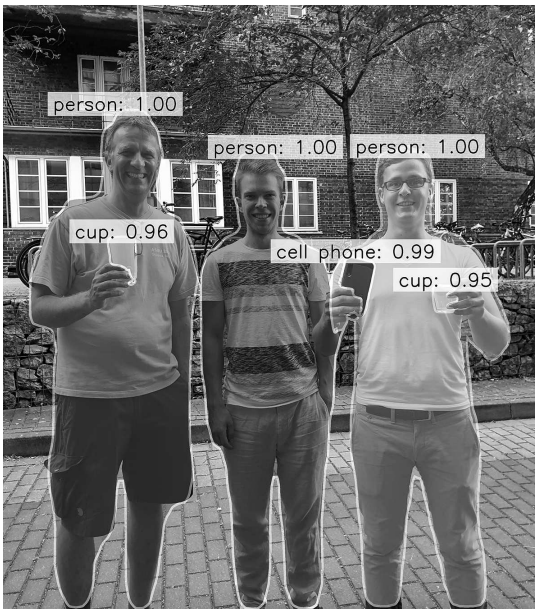

Abbildung 12: Ein Beispiel der auf der TU Night gezeigten Demo zur instanzsbasierten Segmentierung. Zu sehen sind von links nach rechts Prof. Tim Fingscheidt zusammen mit einem Studenten und Marvin Klingner. Alle Personen im Bild werden vom neuronalen Netzwerk pixelweise als Klasse „person“ erkannt. Zusätzlich werden einige Objekte, wie Becher (“cup“) oder Mobiltelefone (“cell phone“), richtig erkannt und segmentiert.

mentiert wurde. In einer weiteren studentischen Arbeit [MA 19/014] wurde das hochkomplexe Netzwerk zur Instanz-Segmentierung auf Ansätze zur Erhöhung der Effizienz untersucht. Eine Instanz-Segmentierung unterscheidet sich von der semantischen Segmentierung in zwei wesentlichen Aspekten. Während die semantische Segmentierung eine pixelweise Klassifikation auf dem Gesamtbild durchführt, fokussiert sich die Instanz-Segmentierung nur auf spezielle Objektklassen, wie Personen, Fahrzeuge oder auch Verkehrsschilder. Des Weiteren findet im Gegensatz zur semantischen Segmentierung auch eine Unterscheidung zwischen Objekten gleicher Klasse statt. Zwei nebeneinander befindliche Fahrzeuge werden als Fahrzeug 1 und Fahrzeug 2 klassifiziert. Neben Verkehrsszenen können mit den richtigen Datensätzen auch andere Szenarien instanzweise segmentiert werden. Ein anschauliches Beispiel dieses Systems von der diesjährigen TU Night ist in **Abbildung 12** illustriert.

Ein neueres Forschungsfeld am IfN ist die Tiefenschätzung aus einzelnen Kamerabildern. Hierbei wird jedem Pixel innerhalb eines Bildes ein Abstand relativ zu der Kameraebene zugeordnet. Damit wird es ermöglicht, ein zweidimensionales Kamerabild in eine Punktwolke im dreidimensionalen Raum umzurechnen und so detektierte Objekte (z.B. aus einer Instanz-Segmentierung) im drei-

dimensionalen Raum zu positionieren. Frühere klassische Verfahren beruhen auf Stereo-Bildern oder Video-Sequenzen. Aktuell ist es mit Hilfe von neuronalen Netzwerken möglich, diese nur während des Trainings zu verwenden und während der Inferenz auf Einzelbildern zu arbeiten. Bei geeignet gewähltem Training umgeht man so Probleme, die durch bewegte Objekte innerhalb von Video-Sequenzen oder durch unterschiedliche Verdeckungen bei Stereo-Bildern entstehen. Herr Klingner versucht des Weiteren, durch sogenannte Multi-Task-Learning-Ansätze Synergien aus der Kombination von semantischer Segmentierung und Tiefenschätzung zu erzielen. Eine Publikation ist in Vorbereitung.

Im Rahmen des Erasmus+-Programms wurde im Berichtszeitraum unter Betreuung von Herrn Elshamy eine Masterarbeit angefertigt, die unterschiedliche Datencontainer für Deep Learning auf ihre unterschiedlichen Eigenschaften bzgl. Zugriffsgeschwindigkeit, Kompatibilität und Handhabbarkeit untersucht [MA 19/012].

5.2 Unüberwachte Domain Adaptation (UDA)

Ein zweites strategisch wichtiges Themenfeld des IfN im Bereich Computer Vision ist die sog. Domain Adaptation. Wenn die Daten, die einem neuronalen Netz im Betrieb zugeführt werden, stark von den Daten des Trainingsmaterials abweichen, dann kommt es aufgrund dieses sog. Domänen-Mismatches zu einem Einbruch der Performanz. Um dem entgegenzuwirken, gibt es typischerweise zwei verschiedene Methoden der Domänenanpassung. In der ersten werden die Bilder der (annotierten) Trainingsdomäne durch einen Stiltransfer so verändert, dass diese aussehen wie die Bilder, die im Online-Betrieb durch das Netz verarbeitet werden. Ein weiterer Ansatz ist es, domänen-invariante Merkmale zu extrahieren. Letzteren Ansatz verfolgt Herr Bolte.

Im Rahmen einer Masterarbeit [MA 18/020] wurde ein Verfahren implementiert, das durch ein diskriminatives Training auf beiden Domänen domäneninvariante Merkmale extrahiert. Es zeigte sich, dass sich dadurch nicht nur die Performanz auf der anvisierten Zieldomäne, sondern auch auf der ursprünglichen Trainingsdomäne verbessert. Wie in **Tabelle 2** zu sehen, ist es durch die unüberwachte Domänenanpassung zum einen möglich, die Performanz auf dem (ungelabelten) Zieldatensatz KITTI von 44,1 % auf 59,5 % zu erhöhen. Außerdem konnte durch die domänen-invarianten Merkmale sogar die Performanz auf den Daten der Quelldomäne Cityscapes von 56,0 % auf 59,8 % erhöht werden, was den großen Vorteil dieser Methode gegenüber anderen Verfahren ausmacht.

Aufbauend auf den Ergebnissen wurde zudem ein erstes Abstandsmaß zwischen den Domänen entwickelt. Auf Grundlage dieser Arbeiten wurde ein Paper mit dem Titel „Unsupervised Domain Adaptation to Improve Image Segmentation Quality Both in the Source and Target Domain“ [BOL/FIN1] verfasst und auf

Trainiert auf	**Test Set Ergebnisse [mIoU]**	
	$\mathcal{D}_{\text{CS}}^{\text{test}}$	$\mathcal{D}_{\text{KITTI}}^{\text{test}}$
$\mathcal{D}_{\text{CS}}^{\text{train}}$ (ohne UDA)	56.0 %	44.1 %
$\mathcal{D}_{\text{all}}^{\text{train}}$ (mit UDA)	**59.8** %	**59.5** %

Tabelle 2: Ergebnisse auf den test sets für Cityscapes ($\mathcal{D}_{\text{CS}}^{\text{test}}$) und KITTI ($\mathcal{D}_{\text{KITTI}}^{\text{test}}$) mit und ohne unüberwachter Domänenanpassung (UDA)

dem Safe Artificial Intelligence for Automated Driving (SAIAD) Workshop, der im Rahmen der Conference on Computer Vision and Pattern Recognition (CVPR) 2019 in Long Beach, USA, stattfand, veröffentlicht. Das Paper wurde mit dem Best Paper Award ausgezeichnet, siehe auch den Sonderbericht auf Seite 119.

5.3 Gelernte Bildkompression

Das von Herrn Löhdefink bearbeitete Grundlagenforschungsprojekt zu generativen adversarialen Netzwerken (GANs) im automatischen Fahren haben wir verstärkt in Richtung gelernter Bildkompression mittels neuronaler Netzwerke entwickelt. Es unterteilt sich in drei Themenkomplexe.

Zunächst ging es um die Bildübertragung zwischen Steuergeräten auf dem Fahrzeugbussystem zur verteilten Perzeption. Hierzu wurde eine Veröffentlichung auf dem Intelligent Vehicles (IV) Symposium 2019 mit dem Titel „On Low-Bitrate Image Compression for Distributed Automotive Perception: Higher Peak SNR Does Not Mean Better Semantic Segmentation“ [LOE/BAE/FIN1] eingereicht und vorgestellt. In der Veröffentlichung wurde gezeigt, dass gelernte Kompression mit GANs gut für nachfolgende Funktionen geeignet ist, die ebenfalls mit neuronalen Netzen umgesetzt werden. Im Bereich niedriger Bitraten konnten mit diesem Ansatz bereits etablierte Baselines zur Bildcodierung wie z.B. JPEG2000 in der Performanz übertroffen werden.

Das zweite Thema ist die Fokussierung der Kompression auf individuelle Bereiche des Bildes. Hierbei wird eine Region of Interest (ROI) in Form einer Maske in die komprimierte Repräsentation eingebettet. Hierzu sind verschiedene Methoden entwickelt worden, basierend entweder auf einer frühen oder späten Fusion. Der Ansatz führte im Bereich der ROI zu einer Verbesserung der Bildqualität, wobei diese in der ROI selbst, aber auch im Gesamtbild ausgewertet wird. Eine weitere Erkenntnis ist, dass die Methode unabhängig von der verwendeten mathematischen Operation in der Fusion gut funktioniert.

Das dritte Thema befasst sich mit der Verwendung der Vektorquantisierung anstelle der Skalarquantisierung im Training der GANs. Vektorquantisierung ermöglicht eine geringere Bitrate bei gleicher Bildqualität bzw. eine höhere Bild-

qualität bei vergleichbarer Bitrate. Für neuronale Netze ist während des Trainings eine differenzierbare Näherung der Quantisierungsfunktion erforderlich, da der Backpropagation-Algorithmus Gebrauch von Fehlergradienten macht, die bei der echten Quantisierung nicht verfügbar sind. Die Bitrate beim Gebrauch von Vektorquantisierung kann im Vergleich zur unquantisierten Variante (8 Bit pro Pixel im sog. latent space) bei fast gleicher Bildqualität ca. um den Faktor 10 reduziert werden. Es hat sich gezeigt, dass die implementierte Vektorquantisierung in allen Metriken besser als die Skalarquantisierung abschneidet [MA 19/010].

5.4 Corner Cases

Zum Thema Corner Cases hat Herr Bolte eine Veröffentlichung auf dem Intelligent Vehicles (IV) Symposium 2019 in Paris vorgestellt [BOL/BAE/FIN1]. In dem Paper wurde eine erste Definition erbracht, die einen Corner Case auf Metriken mehrerer Teilsysteme zurückführen lässt und jedem Bild eines Videos einen Corner-Case-Score zuordnet. Dabei wurde ein Corner Case definiert, wenn sich ein nicht-prädizierbares, relevantes Objekt in relevanter Position zum eigenen Fahrzeug befindet. Die Nicht-Prädizierbarkeit wird dabei durch einen auf der Bildsequenz operierenden Bildprädiktor messbar gemacht. Die Relevanz eines Objekts ergibt sich aus einer semantischen Segmentierung: Fußgänger, Fahrradfahrer und andere Verkehrsteilnehmer sind relevant. Die relevante Position ergibt sich aus der Trajektorienprädiktion des sog. Ego-Fahrzeugs. Zu letzterem Aspekt der Trajektorienprädiktion wurden im Berichtzeitraum in Kooperation mit Volkswagen zwei weitere Beiträge auf internationalen Fachtagungen publiziert [FIN3], [BOL/FIN2].

Aufbauend auf den Arbeiten aus dem Paper wurde in einer Bachelorarbeit [BA 19/708] ein Tool zur Detektion von Corner Cases auf großen Datenmengen entwickelt. Das Tool ist über eine grafische Benutzeroberfläche bedienbar und so aufgebaut, dass es durch Plug-Ins erweitert werden kann. In weiteren Arbeiten im Bereich der Corner-Case-Detektion können somit schnell und einfach neue Metriken in das vorhandene Tool hinzugefügt werden.

Für die weiteren Arbeiten an der Corner-Case-Detektion wäre ein annotierter Datensatz mit tatsächlichen Corner Cases sehr wünschenswert. Hierfür wurden ein Car-PC und Fahrzeugkameras beschafft, mit denen über einen längeren Zeitraum ein eigener Datensatz aufgenommen werden soll. Um reale Corner Cases aufzuzeichnen, werden definierte Szenarien von Freiwilligen des Instituts gespielt. Es ist geplant, den so entstandenen Datensatz nach Möglichkeit zu veröffentlichen. Erste Aufnahmen unter alleiniger Verwendung der Kameras wurden bereits erstellt.

5.5. KI-Projektfamilie

Das Bundesministerium für Bildung und Forschung (BMBF) hat auf Anregung der VDA-Leitinitiative „Autonomes und vernetztes Fahren“ im Kooperationsvorhaben „KI-Plattform-Konzept“ mit vielen Partnern aus Industrie und Wissenschaft die Entwicklung einer öffentlichen Daten- und Lernplattform für das maschinelle Lernen mit besonderem Fokus auf das autonome Fahren vorbereitet. Neben wirtschaftlichen und juristischen Fragestellungen, die für eine solche Plattform behandelt werden müssen, spielen auch hardware-, sicherheits- und datentechnische Aspekte eine Rolle. Das IfN hat in dem Projekt besonders an den Themen zu Datentypen, Formaten und Metainformationen mitgearbeitet. Der Fokus lag in dem Projekt auf Konzepten und Methoden zur Selektion und Kompression anfallender Daten.

In dem dreijährigen BMWi-Förderprojekt „KI-Absicherung“ ist das IfN von der Volkswagen Group Innovation unterbeauftragt. In dem Projekt geht es vor dem Hintergrund des autonomen Fahrens zunächst allgemein um die Erhöhung der Robustheit von künstlicher Intelligenz gegenüber Veränderungen der Eingangswerte. Hier steht u.a. auch die Robustheit gegenüber adversarialen Angriffen speziell im Fokus. Das IfN beteiligt sich im Projekt mit Beiträgen zur Absicherung von KI-Blackbox-Funktionen mithilfe von Teacher-Student-Netzwerken und GAN-Autoencodern.