

Schriftenreihe, Band 8

Hochschule  
Kempten

University of Applied Sciences



ADAC Stiftung & Hochschule Kempten (Hrsg.)

# Unfallforschung 2021

Symposium für Unfallforschung  
und Sicherheit im Straßenverkehr



Cuvillier Verlag Göttingen  
Internationaler wissenschaftlicher Fachverlag

ADAC Stiftung



## LKW Abbiegeassistent mithilfe KI Szenenverständnis

Johannes Traxler Geschäftsführung, EYYES GmbH	Michael Breiter Entwicklungsleitung, EYYES GmbH	Philipp Zanitzer Projektleitung, EYYES GmbH
---	---	---

### 1. Projektbeschreibung

#### 1.1. Ausgangslage / Problemstellung

Mit einer nach wie vor hohen Anzahl an Unfalltoten hat der Einzug der LKW Abbiegeassistenten eine Verbesserung für Personen und Radfahrer gebracht. Die positive Grundidee eines Assistenten, der vor Personen und Radfahrern im toten Winkel warnen soll, wird allerdings überschattet von der hohen Anzahl an Fehlalarmen der bestehenden Systeme. Durch fehlendes Szenenverständnis bzw. Auswertelogik ist eine Unterscheidung zwischen gefährdeten Personen und Personen, die es nicht sind (da allgemeines Objekt, Geschwindigkeit, Richtung von Person) oft nicht ausreichend möglich und treibt eine Desensibilisierung der Fahrer voran, sodass der Assistent seinen Nutzen verlieren könnte. Personen, die durch ihr Verhalten (Geschwindigkeit, Richtung) nicht gefährdet sind oder auch allgemeine Objekte (bewegte Sträucher,...) werden als häufigster Alarmgrund gesehen. Eine Alarmierung auf tatsächlich gefährdete Personen/Radfahrer bietet einen tatsächlichen Mehrwert für die Fahrer.

Die Anforderungen an LKW Abbiegeassistenten sind sowohl durch die Prüfvorschrift „Allgemeinen Betriebserlaubnis (ABE) nach § 22 StVZO und zu Prüfungen Nr. 149, Verkehrsblatt (VkB), Heft 19-2018 des BMVI, Deutschland“ als auch die seit 2020 gültige UN ECE R151 für die Integration in Neufahrzeugen definiert. Die ABE für die Nachrüstung von bestehenden LKW mit einem Abbiegeassistent stand bei diesem Projekt im Vordergrund, da dies auch kommerziell der Eintrittsmarkt ist. Die Weiterentwicklung in Richtung der UN ECE R151 mit den wesentlich größeren Detektionsbereichen und damit verbundenen Komplexität wurde in der Entwicklung bereits mit berücksichtigt, jedoch nicht gleich mit umgesetzt. Grund dafür ist, dass es für die UN ECE R151 keine allgemeine Betriebserlaubnis gibt und daher eine Zulassung pro Fahrzeugtype erfolgen muss. Dies kann EYYES als Kleinunternehmen nicht voreilend durchführen, sondern es wird erst mit einem Umsetzungsprojekt durch einen OEM sinnvoll machbar sein.

Generell muss bei der UN ECE R151 auch noch mit betrachtet werden, dass die geforderten Detektionsbereiche von bis zu 30m neben dem Fahrzeug und 7m vor dem Fahrzeug in 4,25m Entfernung quer zum Fahrzeug für einen sinnvollen Einsatz eine entsprechendes Szenenverständnis benötigt. D.h. ohne dass die Algorithmik des Systems erkennt ob eine Kollisionsgefahr besteht und z.B. nur eine Präsenz meldet ist das System für den Fahrer unbrauchbar. Während einer Fahrt im innerstädtischen Bereich würde jedes System de facto ununterbrochen Radfahrer und Personen die sich in dieser großen Entfernung befinden melden. Dies war auch eine der wesentlichen Anforderungen des Marktes die bei dem Entwicklungsprojekt zum „CAREYE SAFETY ANGLE“ (CSA) von EYYES zu den wesentlichsten Innovationen geführt haben.



## 1.2. Ziel

Es wird/wurde untersucht, wie eine bildbasierte KI in ein echtzeitfähiges System integriert werden kann, um somit ein hochqualitatives Szenenverständnis erreichen zu können. Das dadurch erklärte Ziel ist die Reduktion der Fehlalarmhäufigkeit in Hinsicht auf tatsächliche Gefährdung von Personen und Radfahrern und die Anwendung eines hochperformanten KI-Gesamtsystems.

Mit dem System soll es möglich sein einerseits die geltenden Vorschriften als auch die zukünftigen zu erfüllen. Die erste Stufe beinhaltet klar die für eine Allgemeine Betriebserlaubnis notwendigen Anforderungen. Die Plattform soll bereits von Beginn an so konzipiert sein, dass auch die UN ECE R151 in weiterer Folge damit erfüllbar sein wird.

## 1.3. Vorgehensweise / Methodik

Es wurde ein bildbasiertes Verfahren gewählt, da dieses eine flächige Analyse der Umgebung des LKW ermöglicht und einen hohen Informationsgehalt aufweist. Ein neuartiges Verfahren auf Basis von künstlichen neuronalen Netze (KNN) mit vielen Schichten "Deep Learning" genannt, wurde eingesetzt. Die hohe Robustheit von KNN resultiert aus der Invarianz gegenüber Objektformen gleicher Kategorie und Umwelteinflüssen. Damit sind die KNN konventionellen Bildverarbeitungsmethoden bei weitem überlegen und können mit vergleichbaren Sensoren wie RADAR oder LIDAR konkurrieren.

Es wurde ein Verfahren der „Object Detection“ ausgewählt, da dies einen guten Kompromiss aus Informationsgehalt, Verarbeitungsgeschwindigkeit und Machine Learning Aufwand bildet. Durch die Information, welche Objekte sich wo im Bild befinden, kann die KI vor schwächeren Verkehrsteilnehmern im Abdeckungsbereich warnen. Die größte Herausforderung eines Abbiegeassistenten besteht darin, die benötigte hohe Sensitivität zu erreichen, ohne dass die Spezifität zu gering wird. Es wurden eigene Trainings- und Testdaten durch reale Aufnahmefahrten akquiriert. Der „Detector“ wurde für die beiden Objektklassen Person und Fahrrad trainiert. Das System ist dadurch in der Lage, sowohl vor Radfahrern als auch vor Fußgängern zu warnen, wenn es zu einer gefährlichen Situation kommt. Es wurde streng darauf geachtet, dass Trainingsdaten und Testdaten sich nicht überschneiden. Da kein „Detector“ frei von falsch Positiv und falsch Negativ Meldungen ist, wurde das KNN auf Sensitivität optimiert. Um eine hohe Spezifität zu erzielen, wurde ein Trackingverfahren nachgeschaltet. Das Tracking-by-Detection Verfahren ermittelt den zeitlichen Verlauf, also die Objektbewegung im Bild. Führt die Relativbewegung zu einer Entfernung des Objekts vom Fahrzeug, kann keine gefährliche Situation entstehen. Bei sich nähernden oder sich parallel bewegenden Fußgängern/Radfahrern ist das System in der Lage zu warnen.

## 2. Technologie

### 2.1. Sensorik und Abdeckungsbereiche

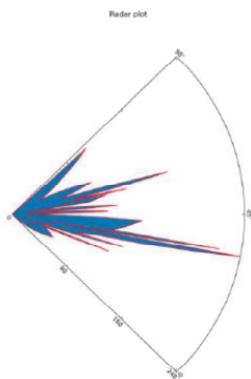
Die Technologien die für die Erfüllung der Anforderungen in Frage kommen können in

- Aktive (RADAR, LiDAR, Ultraschall)



- Passive (Videokamera)

Unterschieden werden. Bei den aktiven Sensoren sendet der Aktor eine elektromagnetische oder akustische Welle aus und misst über Laufzeit und Reflexionsmuster die Präsenz von Objekten. Über nachgelagerte Algorithmen wird aus den Einzelmessergebnissen über die Laufzeit der Abstand und auch die Bewegung relativ zum Fahrzeug gemessen. Diese Sensoren sind davon abhängig, welches Reflexionsverhalten bzw. Querschnitt das angestrahlte Objekt liefert. D.h. die Oberfläche bzw. auch die optischen/elektromagnetischen Eigenschaften spielen hier eine wesentliche Rolle. Darüber hinaus verfügen die Sensoren auf Grund der möglichen Auflösung über kein Szenenverständnis. D.h. sie können nicht unterscheiden, ob sich z.B. ein Radfahrer hinter einer geparkten Autoreihe bewegt oder ob es sich um eine Hauswand mit Fenstern handelt, durch welche immer wieder Objekte erkannt werden. Ebenso ist keine Klassifizierung nach Lichtmast, Personen, Hydranten, Poller zuverlässig möglich, weil die Auflösung dafür nicht ausreicht.



**Abbildung 1: Beispielbild einer FMCW Radarsignalantwort**

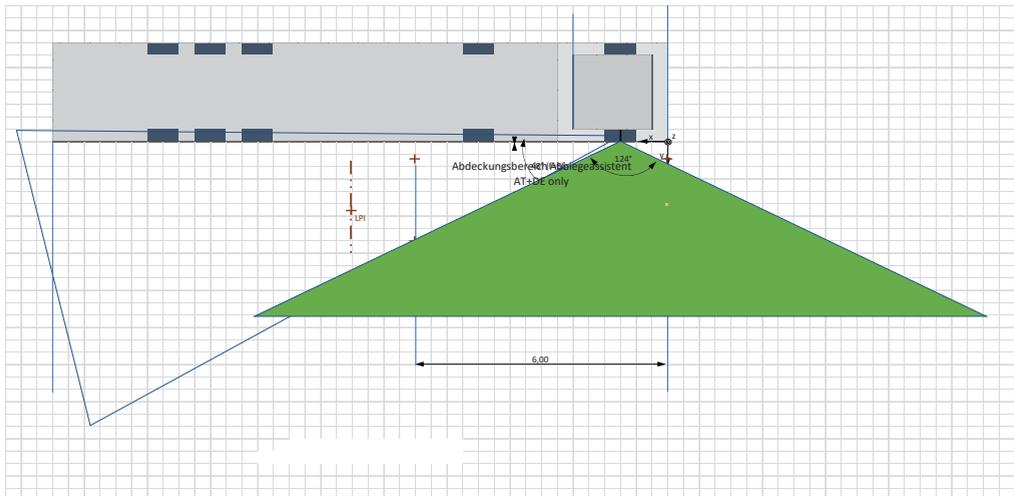
Bei Kamerasensoren wird das von den Objekten reflektierte Licht von einem photoempfindlichen Sensor eingefangen und in ein elektrisches Signal gewandelt. Der wesentliche Unterschied ist, dass die Kamera mit einem Bild die gesamte Umgebung abbildet und somit auch sie Szenen erkennbar macht. Die Kamera sendet jedoch nicht aktiv Licht aus, weshalb sie von anderen Lichtquellen (Sonne, Scheinwerfer, Straßenbeleuchtung etc.) abhängig ist. Alternativ gibt es auch Infrarotkameras, die im fernen Infrarotbereich (8 – 15  $\mu\text{m}$  Wellenlänge) die abgestrahlte Wärme von Objekten aufnehmen können. Diese sind jedoch aus Kostengründen für den Anwendungsfall LKW Abbiegeassistenten derzeit nicht weiter unter Betrachtung gezogen worden.

Die im konkreten Fall eingesetzten Kamerasensoren sind nach dem Stand-der-Technik neueste Modelle von Full-HD Kamerasensoren mit 1920x1200 Bildpunkten und einem High-Dynamic-Range Verhalten. D.h. sie haben eine Bittiefe pro Bildpunkt von  $\geq 12$  Bit und daher sind sie besonders lichtempfindlich und liefern auch bei Dämmerungssituationen bzw. hohen Lichtschwankungen exzellente Bilder. Trotzdem gibt es Situationen, wo die Kamera nicht mehr ausreichend Informationsgehalt im Bild liefert. In diesen Fällen



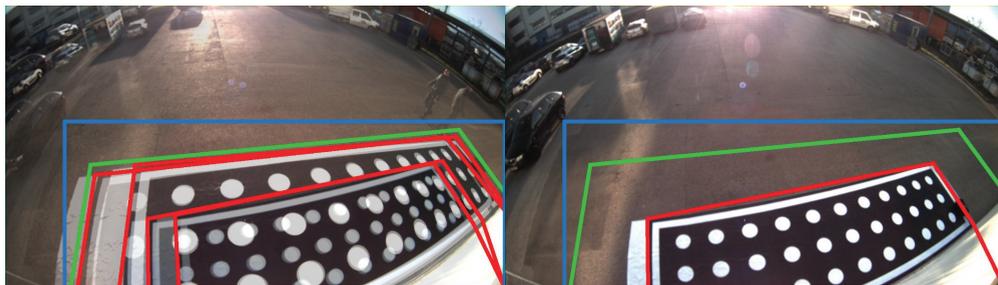
muss dann eine integrierte Algorithmik das System gezielt deaktivieren bzw. den Fahrerinnen informieren, dass keine zuverlässige Funktion mehr möglich ist.

Im Rahmen des Entwicklungsprojektes CSA wurde auf Kamerasensoren gesetzt, weil sie die Möglichkeit bieten komplexe Situationen, wie der Erkennung von Radfahrern und Radfahrerinnen hinter geparkten Autoreihen oder Kreuzungssituationen, gut zu erfassen. Entscheidend ist für eine optimale Erfassung des Überwachungs- bzw. Gefahrenbereichs, dass die richtige Brennweite und Montageposition mit Spielraum für unterschiedliche Fahrzeugtypen gefunden wird. Dazu wurde im Rahmen des Projektes eine Lösung für den LKW Abbiegeassistenten laut „Allgemeinen Betriebserlaubnis (ABE) nach § 22 StVZO und zu Prüfungen Nr. 149, Verkehrsblatt (VkBt.), Heft 19-2018 des BMVI, Deutschland“ (kurz KBA) als auch für die neue UN ECE R151 betrachtet. In der Abbildung 2 ist dieser Unterschied erkennbar. Während die KBA Version einen 1,6x6m breiten Bereich neben der rechten Fahrzeugfront fordert, ist der Bereich für die UN ECE R151 von 30m hinter dem Fahrzeug bis zu 7m vor dem Fahrzeug in einem Abstand von 4,25 m abzudecken.



**Abbildung 2: Übersicht Abdeckungsbereiche KBA Nr. 149 (orange) und UN ECE R151**

Im ersten Schritt wurde mit einer Detektionskamera der KBA Bereich als Primärziel definiert. Dafür wurden bereits zwei Systemvarianten TASET001 und mit April 2021 auch die Weiterentwicklung TASET002 entsprechend den Prüfvorschriften vom technischen Dienst erfolgreich überprüft.



**Abbildung 3: Geforderter Abdeckungsbereich (rot) und verwendeter überwachter Bereich (blau/grün) für verschiedene Kameramontagepositionen überlagert (links) und für eine konkrete Montageposition in 3,8m Höhe und 1,8m hinter der Fahrzeugfront (rechts).**

Da sich der sichtbare Bereich mit der Montageposition der Kameras am Fahrzeug verändert, wurden Kalibrierungen für verschiedene Montagepositionen am Fahrzeug durchgeführt. Das System wurde schließlich für die großzügigste Auslegung der möglichen Montagepositionen parametrieren, um die Installation durch den Endkunden zu vereinfachen und Parametrierfehler zu umgehen. Dies bedeutet jedoch, dass der durch das System überwachte Bereich für gewöhnlich größer ist als der geforderte Abdeckungsbereich, was zu Warnung vor Objekten führen kann, die sich außerhalb des geforderten Abdeckungsbereiches befinden. In Abbildung 3 wird dieser Einfluss gezeigt: Um alle Montagepositionen abzudecken wird der überwachte Bereich im schlimmsten Fall stark vergrößert.

Für die Abdeckung des UN ECE R151 Bereichs (siehe Abbildung 2) sind mehrere Sensoren erforderlich. Mit einem einzigen Sensor (unabhängig von der Art des Sensors) können die benötigten Auflösungen (= Mindestfläche des kleinsten Objektes das noch automatisch erkannt werden muss [11]) bzw. Öffnungswinkel nicht mehr erreicht werden. Im Fall des CSA sind zwei Kameras dafür geplant. Der benötigte Überwachungsbereich führt zu Herausforderungen, speziell bei der Relevanz der Meldungen.

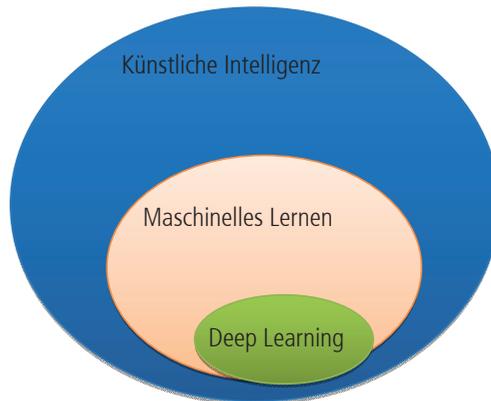
Es muss laut UN ECE R151 ein Radfahrer bereits in einer Entfernung von 7m vor dem Fahrzeug gemeldet werden. Man denke an Kreuzungssituationen, wo die gegenüberliegenden Gehsteige oft nur < 5m entfernt und keine Kollisionsgefahr besteht. Auch die Meldung eines Radfahrers in der Entfernung von 30m hinter der Fahrzeugfront wird im täglichen Straßenverkehr zu sehr häufigen, korrekten, aber nutzlosen Meldungen führen. Diese Anforderungen führen zu der Notwendigkeit, dass das System den Radfahrer zumindest in einem Abstand von > 10m vor und 40m hinter dem Fahrzeug detektiert um zuverlässig das Bewegungsprofil bestimmen zu können. Ansonsten kommt es speziell im innerstädtischen Bereich zu enorm häufigen Meldungen, weil sich oft Radfahrer in einem solchen Abstand bewegen werden.

Diese Rahmenbedingungen waren die Ursache für die Entwicklung des KI-Szenenverständnisses für den CSA. Nur damit sehen die Experten von EYES eine Chance eine robuste und für den täglichen Einsatz brauchbare Lösung für die Erhöhung der Sicherheit von Radfahrerinnen und Radfahrern als auch eine Unterstützung für Lenkerinnen und Lenker anzubieten.

## 2.2. Künstliche Intelligenz



Das Forschungsgebiet der Künstliche Intelligenz besteht aus mehreren Teilbereichen (siehe Abbildung 4), wobei maschinelles Lernen sicherlich einer der wesentlichsten davon ist. Dabei geht es darum Muster oder Strukturen durch adaptive Parameter bzw. Matrizen unterstützt von entsprechenden Optimierungsalgorithmen so gut als möglich zu „erlernen“, um diese in weiterer Folge wiederzuerkennen und einen Ähnlichkeitswert auszugeben.



**Abbildung 16: Übersicht Künstliche Intelligenz**

Allein der Gedanke an die Möglichkeit eine in vielen Bereichen des täglichen Lebens, in der Wirtschaft, oder in der Wissenschaft jederzeit einsetzbare mit der menschlichen Denkleistung vergleichbare „künstliche Intelligenz“ schaffen zu können fasziniert schon lange die Menschheit und bringt sie dazu, sich mit diesem hochinteressanten Wissensgebiet zu beschäftigen.

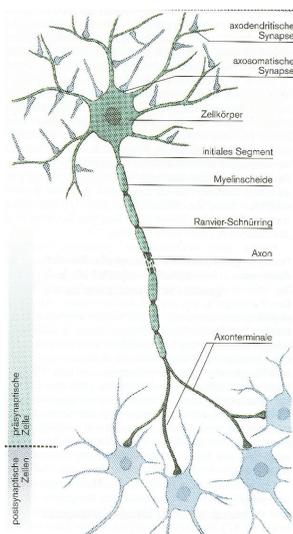
Seit den 1950er Jahren haben sich unterschiedlichste mathematische Lösungsansätze für diese Aufgabenstellung entwickelt. Ein Teilgebiet dieser Entwicklung bilden die künstlichen neuronalen Netze. Dieses Untergebiet des „Machine Learning“, versucht durch mathematische Funktionen die Eigenschaften von biologischen neuronalen Netzen (siehe Abbildung 5) zu imitieren. Der Denkansatz, dass man die Funktionen des menschlichen Gehirnes modellhaft in einzelne Verarbeitungsschritte (Algorithmen) zerlegt und diese mit Hilfe der Methoden der Mathematik und Physik nachempfendet, bildet die Grundidee der in vielen Medien bereits zitierten „künstlichen Intelligenz“.

Seit 2009/2010 sind diese durch eine neue Methode, dem „Deep-Learning“ reicher geworden. Dabei waren die eigentlichen Algorithmen gar nicht neu. Die erste Umsetzung eines künstlichen Neurons wurde bereits in den 60er Jahren des 20ten Jahrhunderts mit analoger Technologie durchgeführt (siehe Abbildung 6). Das erste Perceptron wurde 1957 von Frank Rosenblatt entwickelt. Damals war schon die Meinung, dass innerhalb weniger Jahre der erste Computer gehen, sprechen, sehen, schreiben und sich selbst reproduzieren wird können. Das dies zu optimistisch war denn erst heute, also 63 Jahre später werden die ers-



ten vier Schritte erreicht. Von einer Reproduktion im Sinne von evolutionärer Weiterentwicklung sind wir noch weit entfernt.

Neu an Deep-Learning ist jedoch, dass Trainingsdaten in einem Umfang zur Verfügung gestanden sind, welcher wenige Jahre zuvor noch undenkbar gewesen wäre. Mit Millionen Bilddaten mit einer entsprechenden „Grundwahrheit“, d.h. Trainingsinformationen für das Neuronale Netz, und Rechenleistungen die entsprechende Trainingsläufe in realistischen Zeiträumen ermöglichten zeigten, dass die zuvor geglaubten Grenzen der Neuronalen Netze nur durch zu wenig Trainingszyklen und Trainingsdaten bedingt waren.



**Abbildung 5: Abbildung eines biologischen Neurons**

Als wesentliche Innovation des Deep-Learning kann die Verwendung von immer „Tiefer=Deep“ werden den Netzen angesehen werden. Während bis dahin wenige Schichten an Neuronen die gelebte Praxis waren, wurden die Architekturen der Neuronalen Netze um immer mehr Schichten erweitert und es wurde festgestellt, dass die Qualität der Detektionsergebnisse damit immer mehr steigt. Dies war ein wesentlicher Durchbruch, denn bei den vorherigen Modellen ging man immer davon aus, dass die Netze beim Training übertrainiert werden und dann nicht mehr besser werden. Damit war auch die Grundlage für die dann folgende 10-jährige Aufholjagd der „Deep-Neural-Networks“ in unterschiedlichsten Anwendungsgebieten.

Allgemein sind die Neuronalen Netze also Optimierungsverfahren, welche durch Erlernen von Informationen in der Lage sind ähnliche Muster wiederzuerkennen. Das bedeutet jedoch, dass für eine zuverlässige und sichere Funktion des Neuronalen Netzes auch ausreichend Information oder in Nachrichtentechnischen Worten gesprochen, der Signal-Rausch-Abstand, ausreichend sein muss. Für die Videotechnik ist dies eine einfach zu verstehende Grundlage, denn wenn man als Mensch in dem Bild das gesuchte Objekt oder die Information nicht mehr erkennen kann, dann wird dies auch kein Algorithmus schaffen.

Als Ziel der „Machine Learning“ Experten Weltweit steht das „generalisierte“ Künstliche Intelligenz (Artificial general intelligence – AGI). Es handelt sich dabei um eine hypothetische Intelligente Maschine mit der



Fähigkeit jede intellektuelle Aufgabe, die ein Mensch lösen kann, zu lösen. Eine solche Maschine wird den „Turing Test“, ein Test bei dem ein Mensch mit einer Maschine und einem Menschen kommuniziert und feststellen muss wer von den Beiden eine Maschine ist, bestehen. Bis dato wird eine solcher Grad an Intelligenz nicht erreicht. Nimmt man die Anzahl der Neuronen des leistungsfähigsten Neuronalen Netzes besitzen 16 Millionen Neuronen, was in etwa der Leistungsfähigkeit eines Froschs entspricht. Dieses Netz läuft den leistungsfähigsten Supercomputern. Um einen Vergleich mit dem menschlichen Gehirn zu ziehen, entspricht dies in Etwa 0,016%. Man sieht also, der zu einer „generalisierten“ Künstliche Intelligenz ist noch weit.

Anwendungsmöglichkeiten für künstliche Intelligenz gibt es in den verschiedensten wissenschaftlichen und industriellen Disziplinen. Erwähnt seien an dieser Stelle die medizinische Technik, die Messtechnik und die Sicherheitstechnik. Je nach Einsatzgebiet für die Mustererkennung werden speziell geeignete Detektoren, welche die Informationen für den jeweiligen Algorithmus aufnehmen, benötigt. Diese unterschiedlichen Aufnahmeverfahren werden auch „bildgebende Verfahren“ genannt. Sie können sowohl optische, akustische als auch atomare Eigenschaften des zu beobachtenden Objektes oder Raumes zur Aufnahme heranziehen.



**Abbildung 6: Bild des ersten künstlichen Neurons**

Prinzipiell gibt es keine Beschränkung auf welche Art von Daten Tiefe Neuronale Netze angewandt werden. Am bekanntesten sind jedoch die Anwendungen im Bereich „Computer Vision“ und Spracherkennung wo Fahrerassistenzsysteme in heutigen Kraftfahrzeugen bereits auf Deep-Learning aufsetzen oder Dienste wie SIRI<sup>1</sup> oder ALEXA<sup>2</sup> bereits große Verbreitung erreicht haben. Die Entwicklung hört hier jedoch nicht auf und jedes Jahr kommen neue Techniken und mathematische Funktionen hinzu die weitere Verbesserungen ermöglichen.

### 2.3. Embedded Deep Learning

Die herausragenden Qualitäten der Erkennung und auch Stabilität von KNN stellt die Technologie jedoch auch vor Herausforderungen. Ein großes neuronales Netz benötigt im Durchschnitt für die Verarbeitung

<sup>1</sup> Copyright von APPLE Inc.

<sup>2</sup> Copyright von AMAZON Inc.



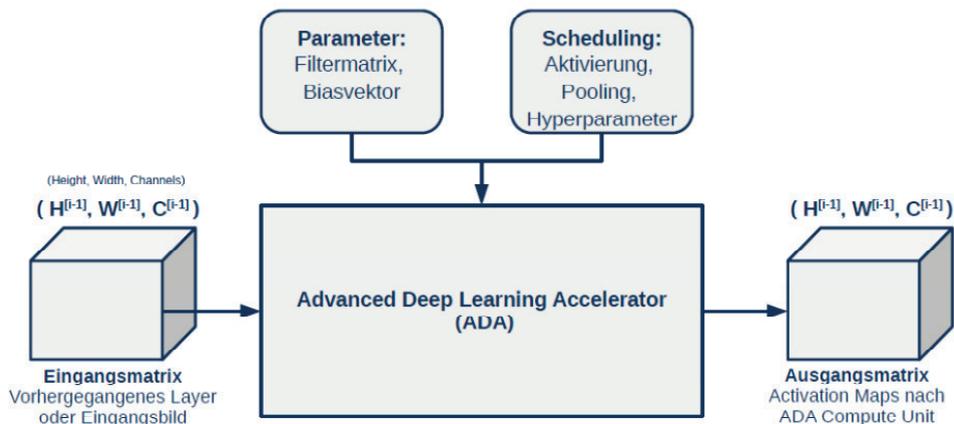
eines einzigen Bildes mit einer Auflösung von 416x416 Bildpunkten 100x10<sup>9</sup> Rechenoperationen. Die meisten davon sind Multiplikationen und Additionen. Eine solche Menge von Berechnungsschritten ist jedoch selbst für Hochleistungsprozessoren noch immer eine Aufgabe. Im Zusammenhang mit mobilen Anwendungen ist es natürlich noch viel schwieriger, entsprechende leistungsfähige Systeme von der Energieversorgung bis hin zur Kühlung in Fahrzeuge zu integrieren.

Deshalb hat sich speziell seit 2018 der Markt der so genannten „Deep-Learning-Accelerator“ (DLA) entwickelt. Dabei wird versucht die Rechenoperationen einerseits von energieaufwändigen Gleitkommaoperationen in Fixkommaoperationen zu konvertieren ohne große Qualitätseinbußen (Loss) in Kauf nehmen zu müssen. Andererseits wird durch unterschiedlichste Ansätze der Verarbeitungsprozess so weit als möglich parallelisiert, um mit weniger Frequenzen bzw. elektrischer Verlustleistung ähnliche Verarbeitungsgeschwindigkeiten zu erzielen.

DLA lassen sich zunächst in zwei Kategorien teilen. Die erste Kategorie umfasst alle Hardwarebeschleuniger, die für den Trainingsprozesse geeignet sind. Hierfür verfügen diese über zusätzliche Funktionalitäten, die für die Backpropagation benötigt werden. Dabei handelt es sich um ein Verfahren zum Trainieren von KNN, bei dem basierend auf der Abweichung zwischen einem berechneten Wert und dem erwarteten Wert Anpassungen an den Parametern des KNN vorgenommen werden. Derartige Beschleuniger, wie etwa Googles Tensor Processing Unit (TPU) [7] und die Deep Learning Unit von Wave Computing [5], sind meist für die Anwendung in großen Rechenzentren konzipiert. Dabei steht vor allem ein hoher Datendurchsatz von mehreren tausend Bildern pro Sekunde im Fokus des Designs. Dieser Datendurchsatz wird meistens auf Kosten einer hohen Leistungsaufnahme erzielt.

Die zweite Kategorie umfasst die Hardwarebeschleuniger, die für die Verwendung mit bereits trainierten Netzen gedacht sind. Anders als die Beschleuniger der ersten Kategorie verfügen diese nicht über die Funktionalitäten, die für die Backpropagation benötigt werden. Dadurch ergibt sich für Hardwarebeschleuniger dieser Kategorie schlankeres, kompakteres Design. Hardwarebeschleuniger dieser Kategorie finden überwiegend in mobilen Geräten Anwendung. Die Leistungsaufnahme sollte möglichst gering sein, wobei gleichzeitig die Datenrate hoch genug sein muss, um zeitkritische Aufgaben erfüllen zu können.

EYYES hat eine eigene Lösung für einen DLA entwickelt. Dieser so genannte „Advanced Deep-Learning Accelerator“ (ADA) ist ein Beschleuniger der zweiten Kategorie. Er verfügt über separate Module für die Berechnung von Convolutional, Activation und Pooling Layer. Dabei können das Convolutional Modul und das Pooling Modul unabhängig von den anderen Modulen angesteuert werden. Lediglich das Activation Modul kann nur in Kombination mit dem Convolutional Modul aktiviert werden. Das Convolutional Modul setzt sich aus acht einzelnen Convolutional Einheiten zusammen, von denen jede einen Kanal eines Filters berechnet. Die Anzahl der Convolutional Einheiten ist parametrierbar und prinzipiell unbeschränkt. Auf das Convolutional Modul folgt ein Caching Modul, in dem die Ergebnisse des Convolutional Moduls für Filter mit mehr als acht Kanälen zunächst zwischengespeichert und später zusammengefasst werden. Die zusammengefassten Daten werden an das Activation Modul übergeben.



**Abbildung 7: Ein- und Ausgänge der ADA Compute Unit**

Durch die maximal parallelisierte Verarbeitung ist es möglich mit einem leistungssparenden FPGA mit nur 100 MHz Taktfrequenz bis zu  $10^{12}$  INT8 Rechenoperationen pro Sekunde zu verarbeiten. Die Innovation besteht dabei in der Art und Weise wie die Daten durch den ADA Kern miteinander für Faltungs-, Additions- und Vergleichsoperationen verschränkt werden. Deshalb ist es auch möglich mit einem Gerät mit einem Leistungsverbrauch von nur 2W pro Videokamera eine Applikation wie den LKW Abbiegeassistent mit Deep-Learning zu realisieren.

## 2.4. Objektdetektion

Objektdetektion beschreibt die Funktion, Informationen über Ort und Art eines Objektes im Bild zu erkennen. Als Objekt kann dabei jede Entität angesehen werden, also z.B. Personen, Tiere, Fahrzeuge, Gegenstände. Objektdetektion bildet einen wichtigen Grundbaustein für das Szenenverständnis aus Kamerabildern, stellt jedoch gleichzeitig ein schwieriges Problem der Computer Vision dar. Durch Deep Learning konnten jedoch seit 2013 große Erfolge erzielt werden. Wichtige Errungenschaften sind Neuronale Netze wie „Region based CNN“ (R-CNN) [13] oder „Single Shot Detector“ (SSD) [14], jedoch stellt der erhöhte Berechnungsaufwand dieser Methoden eine nicht zu unterschätzende Hürde dar. Das Neuronale Netz „Yolo v3“ [15] stellt einen sehr guten Kompromiss aus Qualität und Laufzeitverhalten vor, während „Tiny Yolo v3“ [15] mit deutlich weniger Rechenoperationen auskommt und in der Literatur ordentliche Ergebnisse liefert, die jedoch deutlich unter der State-of-the-Art liegen. Wir haben eine ähnliche Architektur entworfen, trainiert und für den Anwendungsfall Abbiegeassistent optimiert, die die Laufzeitperformanz eines Tiny Yolo v3 mit der Qualität eines Yolo v3 für die Detektion von Personen und Fahrrädern vereint.

### 2.4.1 Training der Objektdetektion

Mit der Entwicklung der tiefen KNN für Training und Qualitätsoptimierung wurde auch die Miniaturisierung erforscht. Die Mittel zur Reduktion der benötigten Speicherbandbreiten, der Anzahl an Rechenoperationen und der Nutzung von Fixkommaoperationen spielt dabei eine wesentliche Rolle. Methoden wie Pruning [9] sind zusätzlich wichtig, um unwesentliche Teile des KNN zu entfernen. Für eine Realisierung einer mobilen



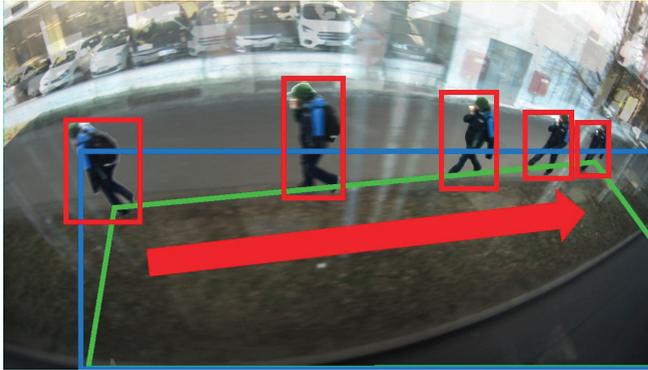
Applikation für Fahrzeuge auf DLA, wie sie in Abschnitt 2.3 beschrieben sind, müssen sämtliche Methoden angewandt werden um den Rechenaufwand für die Berechnung der Aufgabe des KNN um einen Faktor 10 oder mehr zu reduzieren. Dies ist EYYES mit der Entwicklung des EYYESNET gelungen.

Obwohl für den Fall Abbiegeassistenten optimiert, wurde im Training möglichst diverses Bildmaterial verwendet, um den Detektor robust gegenüber Veränderungen zu machen. Der Fokus bei der Akquise neuen Bildmaterials lag dennoch in der Perspektive in der die Kamera von zwischen 2m und 4m Höhe auf die Bodenebene herabschaut, da diese Perspektive in üblichen Bilddatenbanken deutlich unterrepräsentiert ist, aber im Abbiegeassistenten zum Einsatz kommt. Durch Data Augmentation wurde im Training das verfügbare Bildmaterial zufällig verfremdet, um Overfitting zu umgehen und eine höhere Robustheit gegenüber Störeinflüssen zu erzielen. Neben Helligkeits- und Auflösungsvariationen wurden auch Bildrotationen um bis zu +/- 45° augmentiert, die der Verzerrung eines Fischaugenobjektivs im Randbereich ähnlich sind. Das Training hat in zwei Schritten stattgefunden: Zuerst wurde das neuronale Netz mit mit 64764 Bildern vortrainiert, wobei allgemeines Bildmaterial zum Einsatz kam. In diesem Schritt werden vor allem die ersten Layer des Feature Extractor Netzes definiert. Anschließend wurde ein Fine-Tuning Training mit 15459 Bildern durchgeführt, bei dem vermehrt Bildmaterial für den Anwendungsfall „Abbiegeassistent“ zum Einsatz kam. In diesem Schritt werden vorwiegend die hinteren Layer des neuronalen Netzes erlernt. Für die Validierung und Identifikation der finalen Trainingsiteration, bei der noch kein Overfitting stattfand, wurden 6953 Bilder verwendet, die nicht im Trainingsprozess eingesetzt wurden. Die Generalisierung und Verwendbarkeit für den Abbiegeassistenten wurde mit einem vom Training vollständig unabhängigen Testdatensatz aus Abbiegeassistentenaufnahmen geprüft. Die erzielte Trainingsqualität wird in Kapitel 3 dargestellt. Gezielt akquirierte Aufnahmen zeigen die geforderte Funktionalität nach KBA. Zusätzliche Aufnahmen unter erschwerten Bedingungen zeigen Funktionalitäten nach UN-ECE R151.

Im Abbiegeassistenten „TASET“ von EYYES der Produktgruppe CSA wird eine eigens entwickelte KI-Beschleuniger Plattform eingesetzt, um Transparenz zu ermöglichen und das Thema funktionale Sicherheit in der eigenen Hand zu haben.

## 2.5. Tracking

Objektdetektion allein reicht nur für die Erkennung, dass sich eine Person oder ein Radfahrer an einer Stelle im Bild befindet. Dies ist zwar ausreichend für einen Abbiegeassistenten um die KBA Anforderungen zu erfüllen, indem die Präsenz von Personen/Radfahrern im Abdeckungsbereich den Fahrer warnt. Jedoch führt dies selbst bei perfekter Detektorqualität noch zu unnötig vielen Warnungen, da sich Personen häufig im Abdeckungsbereich aufhalten können, auch wenn im Abbiegefall keine kritische Situation vorliegt. In Abbildung 8 ist ein solcher Fall beispielhaft dargestellt. Ähnliche Situation treten oft auch näher am Fahrzeug auf, so z.B. bei Personen/Radfahrer, die auf dem Gehweg an einer Ampel stehen und unbewegt warten.



**Abbildung 8: Zeitlich überlagerte Bilder einer unkritischen Situation: Das Fahrzeug überholt eine Person, die sich im Abdeckungsbereich befindet. Die Relativbewegung der Person führt von der Fahrzeugfront weg.**

Eine hohe Falsch-Positivrate im Alltag kann jedoch zu einer Belastung des Fahrers und zu einem Vertrauensverlust in den Abbiegeassistenten führen, der wiederum zu einem erhöhten Risiko in echten Gefahrensituationen führt. Stattdessen sollte ein optimaler Abbiegeassistent nur dann warnen, wenn es zu einer kritischen Situation kommen kann, d.h. wenn die relative Objektbewegung im zeitlichen Verlauf einer Abbiegesituation zur Kollision führen kann. Die Erkennung der relativen Objektbewegung im zeitlichen Verlauf wird in der Computer Vision als „Tracking“ bezeichnet. Eine kritische Situation kann im Fall des EYYES CarEye Safety Angle genau genommen dann auftreten, wenn sich ein Objekt zum Bildrand links oder unten links bewegt. Beispiele für solche Objekt-Trajektorien sind in Abbildung 9 dargestellt.

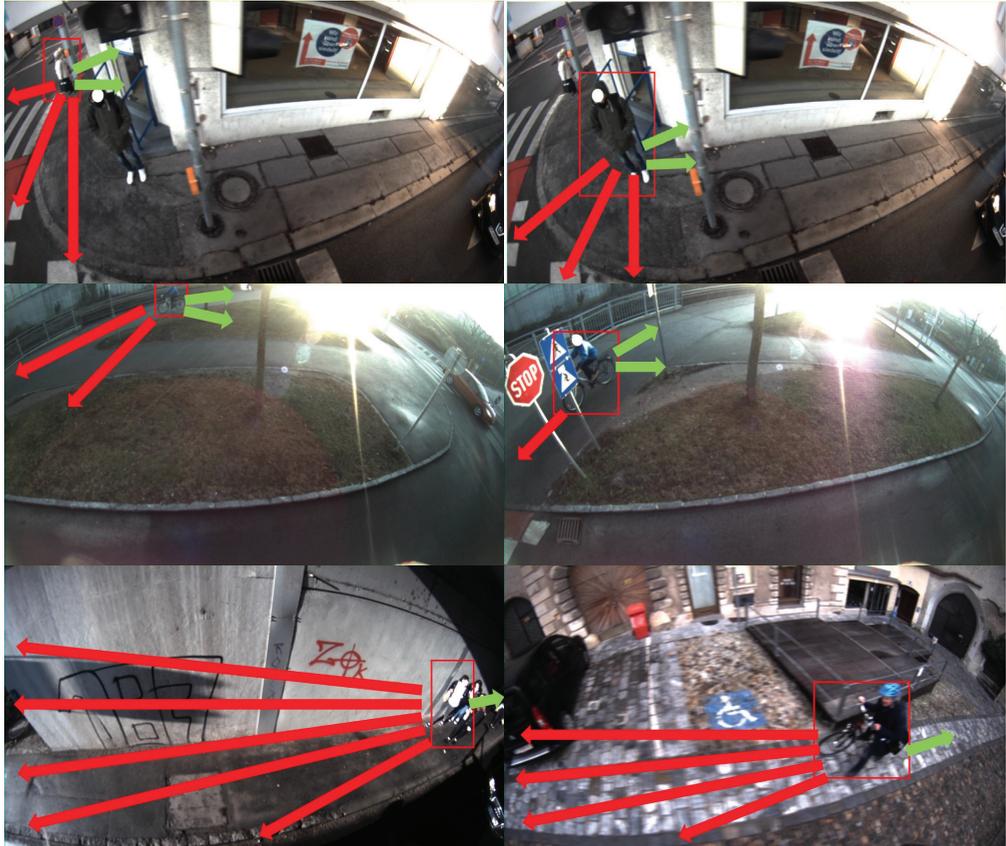


Abbildung 9: Beispiele für Objekt-Trajektorien die zu kritischen (rot) und unkritischen (grün) Situationen führen.



Abbildung 10: Kritischer Bereich (rot), der während des Abbiegevorgangs eine Rolle spielt.

Aus der Objekttrajektorie und dem Bildbereich bei durch die Eigenbewegung im Abbiegevorgang durchquert wird, wurden Gefahrenzonen ermittelt. Der Bereich, der für einen Abbiegevorgang von Bedeutung ist, wird in Abbildung 10 rot hervorgehoben. Hierbei wird ein geringer Kurvenradius angenommen. Bei Bekanntheit des tatsächlichen Kurvenradius könnte theoretisch der Bereich auch dynamisch adaptiert wer-



den. Außerdem könnte der Bereich im zeitlichen Verlauf eines Abbiegevorgangs unterschiedlich variiert werden. Wir verwenden jedoch jederzeit den großzügigen Gesamtbereich, um die höchstmögliche Sicherheit zu erhalten. Führt die Vorhersage der Objektbewegung aus der Objekt-Trajektorie in einer Zeitspanne von 3 Sekunden in den kritischen Bereich, so wird im Falle eines Abbiegevorgangs sofort vor dem Objekt gewarnt. Beim Tracking gibt es mögliche Risiken, die behandelt werden müssen.

1. Fehler im Tracking: Objekte werden im zeitlichen Verlauf falsch zueinander zugeordnet oder einzelne Detektionen/Zuordnungen fehlen
2. Trajektorie liegt zu spät vor
3. Trajektorien sind zu ungenau und führen zu falschen Vorhersagen der Objektbewegung
4. Objekte verändern ihre Bewegungsrichtung
5. Zeitliche Betrachtung ist zu optimistisch oder ungenau

Um diese Risiken zu vermindern, wurden alle Verarbeitungsschritte möglichst großzügig dimensioniert. Die Trajektorien werden von Bild zu Bild über die Detektionen aufgebaut und ausgebaut. Liegen dabei Mehrdeutigkeiten vor, sodass nicht sicher entschieden werden kann, welche Detektionen in zwei Bildern zueinander gehören, werden mehrere mögliche Trajektorien erstellt. Sämtliche Grenzwerte wurden großzügig ausgelegt, um keine Objektzuordnungen bei schnellen Relativgeschwindigkeiten zu übersehen. Auch die Vorhersage der zukünftigen Objektpositionen wurde möglichst großzügig ausgelegt. Zum einen wird die bekannte Objektbewegung verstärkt, sodass das Objekt als schneller angenommen wird als bis zu dem Zeitpunkt bekannt. Zum anderen werden mehrere Vorhersagen für eine Trajektorie erstellt, um die tatsächliche Bewegungsrichtung sicherer abzudecken. Auch der Vorhersagezeitraum wurde großzügiger ausgelegt, als es in der Minimalbetrachtung nötig wäre. Der Algorithmus ist so ausgelegt, dass er bereits nach zwei Detektionen eines Objektes eine Trajektorie schätzt. Alle diese Maßnahmen tragen dazu bei, eine möglichst hohe Sensitivität zu erhalten und keine Sicherheitsrisiken einzugehen, die durch Ungenauigkeit und Messfehler in der Detektion oder dem Tracking auftreten können.

## 2.6. Qualitätskriterien

Die Erkennung von Ereignissen mit einem Detektor kann als binärer Klassifikator angesehen werden, der insgesamt vier Fälle unterscheidet [10]:

- **richtig positiv ( $r_p$ ):**  
Ein Ereignis wird korrekt erkannt.
- **richtig negativ ( $r_n$ ):**  
Eine „Einheit“, welche kein Ereignis darstellt, wird nicht als Ereignis klassifiziert.
- **falsch positiv ( $f_p$ ):**  
Eine Meldung wird generiert, obwohl kein offensichtlich zutreffendes Ereignis stattgefunden hat
- **falsch negativ ( $f_n$ ):**  
Ein stattgefundenes Ereignis wird nicht erkannt, es wird keine Meldung generiert

Für die Beurteilung der Detektionsgüte sind die folgenden Kriterien relevant.



- Die **Richtigpositivrate** (auch: „Sensitivität“, „Detektionsrate“, „Trefferquote“, „Recall“) gibt den Anteil der korrekt als positiv klassifizierten Ereignisse an der Gesamtheit der tatsächlich positiven Ereignisse an (z. B. eine vorhandene Störung wird richtig erkannt).

$$\frac{r_p}{r_p + f_n}$$

- Die **Spezifität** ist definiert durch das Verhältnis der richtig-negativ erkannten Störungen und Ereignisse zur Summe der richtig-negativ und falsch-positiv klassifizierten Störungen und Ereignisse.

$$\frac{r_n}{r_n + f_p}$$

- Die **Falschpositivrate** (auch: „Fehlalarmrate“) ist der Anteil an „Einheiten“, die fälschlicherweise eine Meldung auslösen. (z. B. eine Meldung wird generiert, obwohl keine Störung vorliegt)

$$= \frac{f_p}{r_n + f_p}$$

- Die **Relevanz** (auch: „Genauigkeit“) gibt den Anteil der korrekt als positiv erkannten Ergebnisse an der Gesamtheit der als positiv erkannten Ergebnisse an, und stellt somit die Richtigpositivrate und Falschpositivrate ins Verhältnis.

$$\frac{r_p}{r_p + f_p}$$

Die Akquise von geeigneten Testdaten kann wie im Falle eines Abbiegeassistenten schwierig sein: Im realen Fahrbetrieb kommt es glücklicherweise selten zu kritischen Situationen. Kritische Situationen nachzustellen kann gefährlich sein, daher werden Systemtests, wie z.B. von Prüfstellen, mit stationärem Fahrzeug oder einem bewegten Radfahrerdummy durchgeführt. In welchem Umfang solche limitierten Prüfungen geeignet sind, um die Funktionsfähigkeit in realen Fahrsituationen zu prüfen ist unklar. Im Straßenverkehr kann es zu komplizierten und unübersichtlichen Situationen kommen, bei der auch Sensoren wie Radar oder Kamera in der Erfassung der Szene eingeschränkt werden, z.B. durch Verdeckungen oder andere Störgrößen.

Wir haben kritischen Situationen wie Überholmanöver durch Radfahrer kontrolliert mit stationärem Fahrzeug hergestellt, z.B. Radfahrer mit variierendem Abstand, variierender Geschwindigkeit, beispielhaft in Abbildung 11 dargestellt.



**Abbildung 11: Kontrolliert hergestellte kritische Situationen ähnlich KBA Anforderungen: Ein Radfahrer mit unterschiedlichen Geschwindigkeiten und Abständen am Abbiegeassistenten vorbei.**

Weiter wurden umfangreiche Aufnahmefahrten im öffentlichen Raum durchgeführt. In diesen Aufnahmen kam es jedoch zu sehr wenigen kritischen Situationen, in denen das Fahrzeug im Abbiegevorgang von einem Fußgänger oder einem Radfahrer überholt wurde. Die Falsch-Positiv-Rate kann jedoch mit solchen Aufnahmen sehr gut untersucht werden. Zusätzlich können – eigentlich unkritische – Fahrsituationen umgedeutet werden, um kritische Situationen nachzustellen. So finden Überholmanöver durch Fußgänger und Radfahrer außerhalb von Abbiegevorgängen deutlich häufiger statt. Wird nun in einer solchen Videosequenz ein direkt bevorstehender Abbiegevorgang implizit angenommen und wird auch der Straßenverlauf außer Acht gelassen, kann diese Sequenz als kritische Situation interpretiert werden. In Abbildung 12 sind solche Fälle dargestellt, das Fahrzeug befindet sich nicht vor Abbiegevorgängen (es existieren nicht einmal Wege auf denen abgebogen werden könnte), aber durch den Überholvorgang der Fußgänger/Radfahrer kann die Situation als Testfall für eine kritische Situation herangezogen werden.



**Abbildung 12: Beispiele von uminterpretierbaren Fahrsituationen. Die roten Kreise zeigen Bewegungsvorhersagen aus den Objekt-Trajektorien.**

Eine quantitative Offline-Untersuchung in Simulation wurde durchgeführt, mit aufgezeichneten Videodaten aus zufälligen Fahrten durch Stadt, Ort, Landstraße und Autobahnen. Dabei wurde wie folgt vorgegangen:

- Videodaten wurden akquiriert und in kurze Videosequenzen aus bis zu 10 Sekunden Länge aufgeteilt.
- Für jede Sequenz wurde bewertet, ob in der Videosequenz eine kritische Situation aufgetreten ist, oder die Sequenz zu einer kritischen Situation wie oben beschrieben uminterpretiert werden kann.



- Fällt die Bewertung mehrdeutig aus, z.B. wenn die Sequenz zu Beginn einer möglicherweise kritischen Situation abbricht, wird die Sequenz vollständig aus der Auswertung entfernt.
- Eine Erstellung der Ground Truth wurde anhand dieser der Bewertung erstellt. Jede Sequenz ist
  - o POSITIVE, wenn eine kritische Situation vorliegt oder die Sequenz zu einer kritischen Situation uminterpretiert werden kann.
  - o NEGATIVE, sonst.
- Jede Sequenz wird in einer Offline-Simulation der Abbiegeassistentenfunktionalität ausgewertet: Eine Objekt- Detektion mit 4-5 fps; Tracking + Vorhersage der Objektbewegung; Einordnung der Sequenz in „POSITIVE“ oder „NEGATIVE“ je nachdem, ob die Vorhersage in einen kritischen Bereich fällt oder nicht (Klassifizierung).
- Zum Vergleich wird das Verfahren auch ohne Tracking simuliert: Wenn sich in der Sequenz eine Detektion im überwachten Bereich (Abdeckungsbereich) befindet, wird die Sequenz als „POSITIVE“, sonst als „NEGATIVE“ klassifiziert.
- Der Vergleich zwischen Klassifizierung und Ground Truth wird durchgeführt. Sequenz ist:
  - o True Positive/Negative (TP/TN), wenn für die Sequenz die Klassifizierung und die Ground Truth übereinstimmen.
  - o False Positive, wenn Klassifizierung POSITIVE, aber Ground Truth NEGATIVE (ungerechtfertigter Alarm).
  - o False Negative, wenn Klassifizierung NEGATIVE, aber Ground Truth POSITIVE (fehlender Alarm)

Die Ergebnisse dieser Untersuchung werden in 3.1 beschrieben. Es ist zu beachten, dass in jeder Fahrsituation das Abbiegekriterium (Blinker gesetzt oder geringe Fahrgeschwindigkeit bei eingeschlagener Lenkwinkel) angenommen wurde. Die Auswertung analysiert also nicht das Gesamtsystem, sondern die reine Algorithmen, das Szenenverständnis aus Detektion, Tracking und Vorhersage.

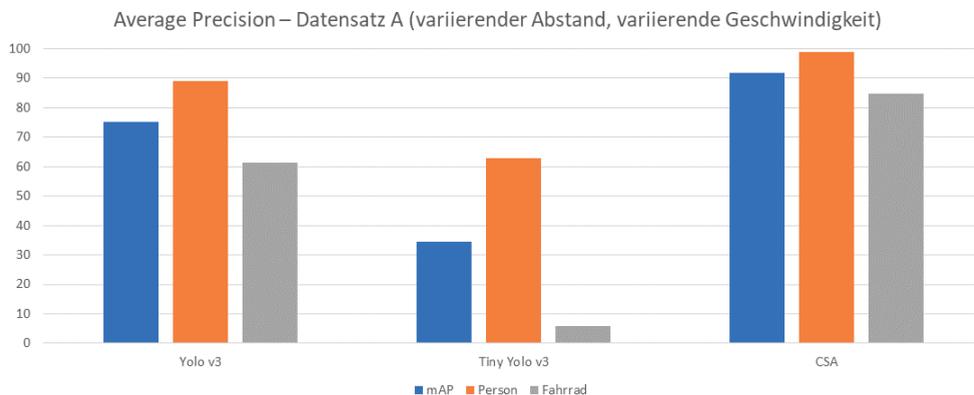
Weiter wurde eine quantitative Online-Untersuchung durchgeführt: Ein Abbiegeassistent mit „embedded Artificial Intelligence“: TASET001 wurde an einem Fahrzeug montiert und Testfahrten wurden durchgeführt. Das Verhalten des Abbiegeassistenten (Monitorausgabe und Signalgeber) wurden mittels zusätzlicher Kamera festgehalten und im Nachgang ausgewertet. Die Ergebnisse dieser Felduntersuchung werden in 3.2 beschrieben und stellen die Qualität und das Verhalten des Gesamtsystems dar, also inklusive der Verwendung der zusätzlichen Informationen wie Lenkwinkel, Blinker und Fahrzeugeigengeschwindigkeit, um den Abbiegevorgang zu erkennen.

### 3. Ergebnis

Vier Iterationen aus Datenakquise, Annotierung, Training und Evaluierung haben zu einer schrittweisen Verbesserung der Detektionsqualität hin zu 98,83% für Personen und 84,79% für Fahrräder im Einzelbild geführt. Vortrainierte Detektoren aus der Literatur mit ähnlicher Komplexität und ähnlichem Ressourcenbe-

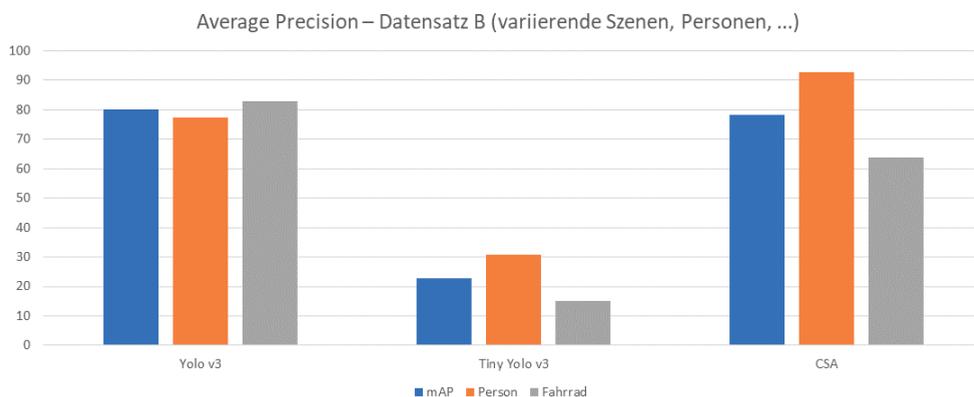


darf des neuronalen Netzes erzielen auf diesen Testdaten eine Detektionsqualität von 62,98% für Personen und 5,75% für Fahrräder aus der Kameraperspektive des LKW-Abbiegeassistenten. Insgesamt wurden mehr als 100K Bilder im Raum Krems, St. Pölten und Wien akquiriert.



**Abbildung 13: Ergebnisse Training Objektdetektion: Datensatz A (in %)**

In Abbildung 13 werden die Trainingsergebnisse des CSA den mit dem COCO Datensatz [12] vortrainierten Objektdetektoren Yolo v3 und Tiny Yolo v3 gegenübergestellt. Testdatensatz A besteht aus gezielt akquirierten Testdaten für einen Abbiegeassistenten: Ein Radfahrer fährt in insgesamt 9 Sequenzen am Fahrzeug vorbei, dabei werden drei verschiedene Abstände zwischen Radfahrer und Fahrzeug und drei verschiedene Geschwindigkeiten verwendet. Die Detektionsqualität wird in Average Precision (AP) bemessen, die sowohl die Sensitivität als auch die Relevanz in einem Wert je Objektklasse zusammenfasst. Die Mean Average Precision (mAP) ist der Mittelwert der APs aller Objektklassen. Jedes einzelne Bild jeder Sequenz wird in der Bewertung herangezogen. Beispiele für Bildmaterial aus Datensatz A sind in Abbildung 11 enthalten. In Abbildung 14 werden die Qualitätswerte für einen weiteren Datensatz dargestellt. Die hier aufgenommenen Sequenzen sind komplexer aufgebaut und Objekte können weiter entfernt liegen. Anonymisierte Beispiele sind in Abbildung 15 dargestellt.



**Abbildung 14: Ergebnisse Training Objektdetektion: Datensatz B (in %)**



**Abbildung 15: Beispieldaten Datensatz B. Die Gesichter der Personen wurden aus Datenschutzgründen anonymisiert.**

In beiden Fällen wurden die Ergebnisse „Yolo v3“ und „Tiny Yolo v3“ mit dem jeweiligen vortrainierten Objektdetektor aus der Literatur erzielt. Die Ergebnisse „CSA“ wurden mit dem von EYES trainierten Objektdetektor erzielt. Die Ergebnisse zeigen, dass mit einer kompakten Deep Neural Network Architektur ein Objektdetektor für einen konkreten Anwendungsfall trainiert werden kann, der die Qualität eines in der Literatur deutlich mächtigeren General Purpose Deep Neural Network Architektur erreicht.

Durch das Tracking-by-Detection Verfahren wird im Alltag eine Reduktion der Warnungen bei unkritischen Fahrsituationen um ca. 66,2565% erreicht. Einerseits bildet das Tracking eine einfache Form der Plausibilisierung, andererseits das Ignorieren von Fußgängern oder Radfahrern, die sich vom Fahrzeug entfernen. Ein Tracking über zwei Einzelbilder genügt bei einer Auswertung von 5 Bildern pro Sekunde.

### 3.1. Quantitative Offline-Untersuchung:

Die Ergebnisse der quantitativen Offline-Untersuchung werden in Tabellen zusammengefasst. In der zweiten Zeile stehen die Anzahl von Positive und Negative Sequenzen, sowie die Zahl an Sequenzen, die für die Auswertung entfernt wurden, da bei Erstellung der Ground Truth keine klare Einordnung als Positive oder Negative möglich war. Die Spalte „Detection“ bezeichnet das Verfahren ohne Tracking, die Spalte „Tracking“ stellt das Verfahren mit Tracking dar. Es ist wichtig zu betonen, dass diese Auswertung nur kritische Situationen betrachtet unter der Annahme, dass ein Abbiegevorgang stattfinden würde, selbst wenn das nicht der Fall ist. Wird z.B. das an einer Ampel stehende Egofahrzeug von einem Fußgänger überholt, ohne dass ein Abbiegevorgang stattfindet (weder Blinker gesetzt, noch Lenkwinkel eingeschlagen), wird es in dieser Auswertung dennoch als kritische Situation eingeordnet. Umgekehrt wird eine Situation auch dann als Falsch-Positiv gewertet, wenn in der Realität weder Abbiegesituation noch kritische Situation vorliegt, das System aber warnen würde, wenn eine Abbiegesituation vorläge. In einem realen System würden diese Situationen jedoch zu keiner Warnung führen, da nur im Abbiegefall gewarnt wird. Die hier genannten Werte für False Positive, Spezifität und Relevanz lassen sich also nicht direkt auf den realen Fahrbetrieb übertragen, sondern zeigen die Qualität der Algorithmik. Die Sequenzen wurden in der Auswertung mit einer Abtastung von 5 Bildern pro Sekunde (Aufnahmezeit) verarbeitet, was in etwa der Verarbeitungsleis-



tung der im Livesystem verfügbaren Hardware entspricht. In Abbildung 16, Abbildung 17, Abbildung 18, Abbildung 19 und Abbildung 20 werden die Ergebnisse der Offline-Untersuchung dargestellt.

Datensatz 1: 20200108			
Sequenzen: 82x Negative, 17x Positive, 2x entfernt			
	Detection	Tracking	Change
TP:	17	17	
FP	61	23	-62.29%
TN	21	59	+180.95%
FN	0	0	
Sensitivity:	100%	100%	
Specificity:	25.6%	71.95%	
Relevance:	21.79%	42.5%	

**Abbildung 16: Ergebnisse Offline-Untersuchung: Datensatz 1**

Datensatz 2: 20200311_15			
Sequenzen: 145x Negative, 4x Positive, 4x entfernt			
	Detection	Tracking	Change
TP:	4	4	
FP	56	26	-53.57%
TN	89	119	+33.7%
FN	0	0	
Sensitivity:	100%	100%	
Specificity:	61.37%	82.06%	
Relevance:	6.67%	13.3%	

**Abbildung 17: Ergebnisse Offline-Untersuchung: Datensatz 2**



Datensatz 3: 20200311_16			
Sequenzen: 55 Negative, 2x Positive			
	Detection	Tracking	Change
TP:	2	2	
FP	29	5	-82.27%
TN	26	50	+92.3%
FN	0	0	
Sensitivity:	100%	100%	
Specificity:	42.27%	90.9%	
Relevance:	6.45%	28.57%	

**Abbildung 18: Ergebnisse Offline-Untersuchung: Datensatz 3**

Datensatz 4: 20200715_2903			
Sequenzen: 29 Negative, 0x Positive			
	Detection	Tracking	Change
TP:	0	0	
FP	7	1	-85.7%
TN	22	28	+27.27%
FN	0	0	
Sensitivity:	100%	100%	
Specificity:	75.86%	96.55%	
Relevance:	0%	0%	

**Abbildung 19: Ergebnisse Offline-Untersuchung: Datensatz 4**



Datensatz 5: 20200715_2904			
Sequenzen: 17 Negative, 0x Positive			
	Detection	Tracking	Change
TP:	0	0	
FP	6	1	-83.3%
TN	11	16	+45.45%
FN	0	0	
Sensitivity:	100%	100%	
Specificity:	64.7%	94.11%	
Relevance:	0%	0%	

**Abbildung 20: Ergebnisse Offline-Untersuchung: Datensatz 5**

Die eingesetzte Tracking-Strategie ist in der Lage, die Zahl der Falsch-Positiven deutlich zu reduzieren, ohne die Sensitivität einzuschränken. So werden insbesondere Einzelobjekte, die sich nicht in eine Richtung bewegen, die zu einer kritischen Situation führt, erfolgreich herausgefiltert. Auch einzelne Falsch-Positiv-Detektionen werden erkannt und entfernt. Die großzügige Zuordnung und Vorhersage hat jedoch Schwierigkeiten mit Gruppen von Personen, da die Zuordnung von Detektionen dann leicht uneindeutig wird und zumeist eine der möglichen Trajektorien zu einer kritischen Situation führen könnte. So sind beispielsweise bei über 50% der nach Tracking verbleibenden 23 FP Sequenzen im Datensatz 20200108 Personengruppen oder größere Menschenmengen in der Sequenz enthalten. Ein weniger großzügiges Tracking ist vermutlich in der Lage in solchen Fällen deutlich mehr richtige Entscheidungen zu treffen, jedoch steigt möglicherweise auch das Risiko für Falsch-Negative. Diese Vermutung soll in Zukunft untersucht werden. Mögliche Ansätze für weniger großzügiges Tracking beinhalten

- Zuordnungsstrategien wie den Hungarian Algorithm
- Matching Techniken, die Pixelbildinhalte der Detektionen vergleichen
- Probabilistisches Tracking

Weitere Verbesserungen für Tracking und Vorhersage können voraussichtlich durch Bildkorrektur erzielt werden, jedoch steigen dadurch auch Aufwand und Fehleranfälligkeit der Kalibrierung.

### 3.2. Quantitative Online-Untersuchung:

Auswertungen Stand Ende 2020:

Krems: 45 min, Schneefall, Winter (< 100lx Lichtintensität)

<b>Information</b>	<b>Sensitivität</b>	97%
<b>Information</b>	<b>Specificität</b>	94%
<b>Warning</b>	<b>Sensitivität</b>	100%
<b>Warning</b>	<b>Specificität</b>	100%



Wien Innenstadt: 180min (Sommer, Abend, > 10klx Lichtintensität)

<b>Information</b>	<b>Sensitivität</b>	100%
<b>Information</b>	<b>Specificität</b>	69%
<b>Warning</b>	<b>Sensitivität</b>	100%
<b>Warning</b>	<b>Specificität</b>	100%

Tabelle 1: Beispielbilder von Feldversuchen als Basis für die Qualitätsauswertung

Bild	Beschreibung
	Bild einer Person inkl. Testfigur „ROTAKIN“ im Abstand von 10m hinter der Fahrzeugfront. Der „orange“ Zonenrand markiert, dass das KNN die Person aktiv erkannt und innerhalb der Gefahrenzone klassifiziert hat.
	Radfahrer im Abstand von 8m hinter der Fahrzeugfront mit aktivem „orange“ Informationssignal.
	Fahrrad in der Gefahrenzone. Das KNN erkennt, dass sich ein Fahrrad in der Gefahrenzone befindet. Dies führt zu einer „orange“ Information, keiner Warnung.
	Radfahrer im Bereich der Totwinkelkamera mit Detektion von Fahrrad und Radfahrerperson. Führt zu einem „orange“ Informationssignal und noch keiner Warnung. Erst wenn der Lenker eingeschlagen wird oder der Blinker aktiviert wird, dann wird eine Kollisionsgefahr errechnet und kann eine Warnung mit „rot + Akustik“ entstehen.

### 3.3. Ausblick/ weiterer Forschungsbedarf

Die Ergebnisse der Forschungsarbeit sind in das Produkt TASET001 eingeflossen, welche bereits im Dezember 2019 die ABE erhalten hat. Mit der Weiterentwicklung TASET002 hat EYYES im Frühjahr 2021 eine noch performantere Plattform zur Anwendung als Abbiege- und Rückfahrassistent auf den Markt gebracht. Die nächste Weiterentwicklung wird die Anforderung der UN-ECE R151 vollständig erfüllen. Daran arbeiten die Expertinnen und Experten von EYYES aktuell mit Hochdruck. Die nächste Generation des CSA soll



derart in das Fahrzeug integriert werden, dass die Sicherheit der Radfahrerinnen und Radfahrer noch weiter erhöht und den Fahrerinnen und Fahrern das Lenken im täglichen Straßenverkehr vereinfacht wird.

Weiterführender Forschungsbedarf besteht vor allem in der Bewertung und dem Benchmarking von bildbasierten Abbiegeassistenten. Die Bewertung bildet immer eine Momentaufnahme und die realen Testmöglichkeiten sind begrenzt. Zwar sind die gewählten Verfahren der künstlichen Intelligenz sehr robust gegenüber Kultur- und Umwelteinflüssen, jedoch können neue Herausforderungen in ungetesteter Umgebung nicht ausgeschlossen werden. Neben Verfahren zur explainable AI könnten Simulationsumgebungen und virtuelle Prüfstände solche Systeme auf Herz und Nieren testen. Ein groß angelegtes Forschungsprojekt zur Herstellung eines Visual Blind Spot System Benchmarks könnte sowohl für die Wissenschaft als auch die Übertragung in die Praxis einen enormen Mehrwert darstellen.

In Zukunft soll untersucht werden, wie das KI basierte Szenenverständnis durch weitere Sensormodalitäten wie Radar, Lidar oder Ultraschall unterstützt werden kann, um die Robustheit durch Redundanz und Störsicherheit zu erhöhen.

Wir sind davon überzeugt, dass ein enormes Potenzial im Szenenverständnis durch Kamerabildauswertung liegt. Die Verwendung von Deep Learning Artificial Intelligence bringt einen großen Qualitätssprung, gegenüber herkömmlichen Techniken der Computer Vision und Image Processing, ohne sein Potenzial bereits völlig auszuschöpfen. Die Zulassung und Prüfung solcher Systeme bleibt jedoch eine Hürde. Die weitere Verbesserung der Detektions- und Trackingalgorithmen unter Berücksichtigung der Sicherheitsanforderungen soll in Zukunft erfolgen.

## **4. Bewertung**

### **4.1. Nutzen für die Gesellschaft**

Der gesellschaftliche Nutzen besteht in dem Zugewinn des Einsatzes von bildbasierten KIs im Alltag und somit im Zugewinn des Vertrauens in die Abbiegeassistenzsysteme. Wenn die Fahrer bei Ertönung der Alarmmeldung sich darauf verlassen kann, dass es zu einer kritischen Situation kommt, sollte nicht eine Handlung eintreten, bildet sich der Mehrwert in Form von weniger verursachten Unfällen im Bereich von Personen/Radfahrern und Lastkraftwagen ab.

### **4.2. Innovationsgehalt**

Deep-Learning in einem Fahrerassistenzsystem im speziellen einem LKW Abbiegeassistent stellt eine Weltneuheit dar. Die Innovation besteht dabei darin mittels tiefer neuronaler Netze, welche sehr hohe Robustheit bei der Erkennung von Personen und Radfahrern erreichen, die erste Verarbeitungsstufe zu realisieren. Damit hat man gegenüber klassischen Feature-Tracking oder hintergrundbasierten Objektverfolgungsalgorithmen enorme Vorteile. Die Innovation von EYES besteht dabei darin die enorm hohen Rechenaufwände in einem kompakten Gerät zu integrieren.



### 4.3. Übertrag-/Skalierbarkeit

Ein solches System lässt sich auf alle beweglichen Sondermaschinen (Landwirtschaft, Kräne), die potenziellen Unfälle durch tote Winkel mit Personen generieren können, übertragen. Die erhöhten Kamerapositionen und somit die Anwendung von einer solch spezialisierten KI mit Tracking sind in vielen Bereichen gegeben. Die Basistechnologie der bildbasierten KI und embedded „Deep-Learning“ Beschleunigerhardwareplattformen, wie von EYYES im Rahmen von CAREYE SAFETY ANGLE entwickelt - werden in naher Zukunft für sich selbst in immer mehr Bereichen Anwendung finden.



## Quellenangaben, Zitierweise, Verzeichnisse:

- [1] Suryavansh, Manu: How to make your own deep learning accelerator chip! Towards Data Science, September 2018, URL: <https://towardsdatascience.com/how-to-make-your-own-deep-learning-accelerator-chip-1ff69b78ece4>
- [2] Wong, William G.: Habana Enters Machine-Learning Derby with Goya Platform ElectronicDesign, 19.September 2018, URL: <https://www.electronicdesign.com/industrial-automation/habana-enters-machine-learning-derby-goya-platform>
- [3] Chen; Yang; Emer; Sze: Eyeriss v2: A Flexible Accelerator for Emerging Deep Neural Networks on Mobile Devices, URL: [http://www.rle.mit.edu/eems/wp-content/uploads/2019/04/2019\\_jetcas\\_eyerissv2.pdf](http://www.rle.mit.edu/eems/wp-content/uploads/2019/04/2019_jetcas_eyerissv2.pdf)
- [4] NVIDIA: NVDLA Primer 2018, URL: <http://nvdla.org/primer.html>
- [5] Nicol, Chris: A Dataflow Processing Chip for Training Deep Neural Networks Wave Computing, 2017, URL: [https://www.hotchips.org/wp-content/uploads/hc\\_archives/hc29/HC29.22-Tuesday-Pub/HC29.22.60-NeuralNet1-Pub/HC29.22.610-Dataflow-Deep-Nicol-Wave-07012017.pdf](https://www.hotchips.org/wp-content/uploads/hc_archives/hc29/HC29.22-Tuesday-Pub/HC29.22.60-NeuralNet1-Pub/HC29.22.610-Dataflow-Deep-Nicol-Wave-07012017.pdf)
- [6] Danon, Orr: Introducing Hailo-8: The Most Efficient Deep Learning Processor for Edge Devices Embedded Vision Summit, Mai 2019, URL: <https://www.slideshare.net/embeddedvision/emerging-processor-architectures-for-deep-learning-options-and-tradeoffs-a-presentation-from-hailo>
- [7] Sato; Young; Patterson: An in-depth look at Google's first Tensor Processing Unit (TPU), URL: <https://cloud.google.com/blog/products/gcp/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>
- [8] Habana Labs Ltd.: Goya Inference Platform White Paper August 2019, URL: [https://habana.ai/wp-content/uploads/pdf/habana\\_labs\\_goya\\_whitepaper.pdf](https://habana.ai/wp-content/uploads/pdf/habana_labs_goya_whitepaper.pdf)
- [9] Breslow, Aha: Simplifying Decision Trees, A Survey, The Knowledge Engineering Review, Vol 12(1), 1997, pp. 1-47
- [10] FGSV, AK-3.2.9, Hinweise zur Videodetektion in Verkehrsbeeinflussungsanlagen HVVBA, FGSV 304/2
- [11] J. Traxler, Johnson Criteria applied for Traffic Incident Detection Systems, MTITS-2013, Dresden
- [12] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.
- [13] He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [14] Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.



- 
- [15] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

Am 16. und 17. Juni 2021 fand das „Symposium für Unfallforschung und Sicherheit im Straßenverkehr“ (UFO 2021) der ADAC Stiftung als Online-Veranstaltung statt. Schwerpunktthema des alle zwei Jahre stattfindenden Symposiums war 2021 die Verkehrssicherheit von ungeschützten Verkehrsteilnehmern.

Dieser Tagungsband enthält die Langfassungen der Beiträge, welche von einem Expertenbeirat ausgewählt wurden. Inhaltlich befassen sie sich mit den für die Erhöhung der Verkehrssicherheit relevanten Bereichen Mensch, Fahrzeug und Infrastruktur.

Mit dem Symposium und diesem Tagungsband wird Experten und Nachwuchswissenschaftlern die Möglichkeit gegeben, ihre Forschungsarbeiten der Fachwelt zu präsentieren und sich zu informieren.

