



Abschlussbericht

28.07.2025

Hybrid AI towards Large Scale Data and Behavior Modeling for Automated Driving (HAIBrid)

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

| | |
|---|---|
| Zuwendungsempfänger: Karlsruher Institut für Technologie (KIT) | Förderkennzeichen: 01IS21096A |
| Vorhabenbezeichnung: Hybrid AI towards Large Scale Data and Behavior Modeling for Automated Driving (HAIBrid) | |
| Laufzeit des Vorhabens: 01.10.2021 – 30.09.2024 | |

Teil I: Kurzbericht

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS21096A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/beim Autor.

1.1 Die ursprüngliche Aufgabenstellung

Dieses Projekt wurde im Rahmen einer bilateralen Forschungs Kooperation zwischen Frankreich und Deutschland durchgeführt und vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Beteiligt waren Universitäten und Unternehmen aus beiden Ländern: Auf universitärer Seite das Karlsruher Institut für Technologie (KIT) und Mines Paris; auf industrieller Seite FZI, SafeAD und Valeo.

Ziel dieser Forschungs Kooperation ist die Verbesserung der Verhaltensprognose für automatisiertes Fahren durch den Einsatz unüberwachter Lernverfahren. Im Fokus stehen dabei Trajektorien und Interaktionen zwischen Verkehrsteilnehmenden, die aus großen Datenmengen extrahiert werden. Grundlage dafür ist die langjährige Erfahrung des KIT sowie dessen Infrastruktur zur Erfassung umfangreicher Verkehrsdaten mittels Straßenkameras und mit Sensorik ausgestatteten Versuchsfahrzeugen.

Die manuelle Annotation von Trajektorien und die Klassifizierung von Interaktionen ist aufgrund ihrer Komplexität, Vielfalt und zeitlichen Variabilität äußerst aufwendig und wenig praktikabel. Im Bereich der Bildverarbeitung haben unüberwachte Methoden inzwischen jedoch eine Leistungsfähigkeit erreicht, die mit der von überwachten Verfahren vergleichbar ist.

Auf Basis der neuesten Entwicklungen in diesem Forschungsfeld wurden effiziente Methoden zur automatischen Trajektorienerkennung, Interaktionsgruppierung und zur Generierung von 3D-Karten entwickelt. Die statische Umgebungsinformation, die typischerweise in semantischen 3D-Karten gespeichert wird, ist nicht nur essenziell für die Bewegungsplanung automatisierter Fahrzeuge, sondern auch für das Verständnis des Verhaltens anderer Verkehrsteilnehmender.

1.2 Der Ablauf des Vorhabens

Die Arbeitspakete der Projektpartner waren vielfältig und komplementär. Das KIT war für die Konzeption und Umsetzung einer intelligenten Infrastruktur mit Kameras zur Datenerfassung verantwortlich und entwickelte ein Modul zur Trajektorienvorhersage. Mines Paris übernahm die Verarbeitung der vom KIT aufgenommenen Bilddaten zur Extraktion von Trajektorien für Fußgänger, Fahrzeuge, Radfahrende und weitere Verkehrsteilnehmende. FZI erstellte eine hochauflösende Karte (HD-Map) der mit Kameras ausgestatteten Kreuzung. SafeAD entwickelte Werkzeuge zur automatisierten Kartenerstellung, mit dem Ziel, das Mapping auf größere urbane Gebiete skalierbar zu machen. Valeo führte Datenerhebungen mit einem experimentellen Fahrzeug durch und setzte zusätzlich einen Simulator ein, um die Interaktion zwischen verschiedenen Agenten zu modellieren und Trajektorien mithilfe von Reinforcement Learning vorherzusagen.

Das Projekt erwies sich als äußerst fruchtbar, da die Zusammenarbeit zwischen den Partnern sehr effektiv verlief und bis heute anhält – unter anderem in Form gemeinsamer neuer europäischer Projektanträge. Darüber hinaus markierte dieses Projekt den Beginn einer neuen Forschungslinie in unserer Arbeitsgruppe, die bereits zahlreiche neue Ideen hervorgebracht hat. Zu den Herausforderungen, denen wir uns in Zukunft widmen möchten, zählen etwa die Einzelpersonenerkennung innerhalb von Gruppen, da aktuell ganze Gruppen als eine Person erkannt werden, sowie die spezifische Detektion von Einsatzfahrzeugen, da diese das Verkehrsverhalten erheblich beeinflussen. Solche besonderen Szenarien sollen künftig gezielt annotiert und analysiert werden.

1.3 Die wesentlichen Ergebnisse

Die aus über 500 Stunden Videomaterial extrahierten Trajektorien wurden öffentlich zugänglich gemacht und sind auf der Projektwebseite <https://haibrid.gitlab.io> für die wissenschaftliche Gemeinschaft verfügbar.

Im Rahmen des Projekts wurden vom KIT mehrere wissenschaftliche Veröffentlichungen erarbeitet und publiziert. Dazu gehören:

- “Road Barlow Twins: Redundancy Reduction for Motion Prediction”. In: ICRA2023 Workshop on Pretraining for Robotics (PT4R). 2023.
- “JointMotion: Joint Self-supervision for Joint Motion Prediction”. In: 8th Annual Conference on Robot Learning (CoRL) 2024.
- “RedMotion: Motion Prediction via Redundancy Reduction”. In: Transactions on Machine Learning Research (2024).
- “SceneMotion: From Agent-centric Embeddings to Scene-wide Forecasts”. In: 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC). 2024.

Die aufgezeichnete Kreuzung erwies sich als sehr dynamisch und interaktionsreich, wodurch das erzeugte Datenvolumen unsere ursprünglichen Erwartungen deutlich übertraf. Insgesamt wurden 1,25 Millionen Trajektorien generiert, darunter ca. 513.000 für Fußgänger, 576.000 für Autos, 50.000 für Radfahrende, 45.000 für LKW, 20.000 für Busse, 25.000 für Züge sowie 25.000 nicht klassifizierte Objekte.

Die im Projekt erzielten Ergebnisse waren für alle Beteiligten äußerst positiv und wir sind überzeugt, dass sie einen nachhaltigen Beitrag zur internationalen Forschung im Bereich des automatisierten Fahrens und der modellbasierten Verkehrsanalyse leisten werden.

Teil II: Eingehende Darstellung

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS21096A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/beim Autor.

2.1 Wichtigsten Positionen des zahlenmäßigen Nachweises

Der größte Teil des Budgets, 92%, wurde für Personalkosten aufgewendet. Weniger als 1% wurde für den Kauf von Kabeln und Materialien ausgegeben, die für die Installation der Sensoren auf dem Dach des Gebäudes benötigt wurden. Etwa 4% des Budgets wurden für die Teilnahme an Konferenzen (ICRA in London und ITSC in Bilbao) und internen Projekttreffen in Paris verwendet. Die verbleibenden 3% wurden schließlich in die Anschaffung eines Grafikprozessors für die Entwicklung von Methoden zur Vorhersage von Flugbahnen sowie in Speichergeräte für die Datenaufzeichnung investiert.

2.2 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die geleistete Arbeit orientierte sich an den Projektzielen und den einzelnen Arbeitspaketen und war zur Erreichung dieser Ziele notwendig. Ein Teil der Projektentwicklung, wie z.B. die Bahnvorhersage und die Verhaltensanalyse, wurde an öffentlichen Datensätzen trainiert und evaluiert und nicht am projektinternen Datensatz, um unsere Methoden mit denen des Stands der Technik vergleichen zu können. Im Folgenden finden Sie eine detaillierte Beschreibung der Aufgaben, die für jedes Arbeitspaket entwickelt wurden.

2.2.1 WP0 Verwaltung

T0.1 Konsortialvertrag: Der Konsortialvertrag wurde von allen Projektteilnehmern unterzeichnet.

T0.2 Projektverwaltung: Das KIT hat ein gemeinsames Gitlab-Projekt eingerichtet, in dem alle Partner gemeinsam diskutieren können, wie anstehende Probleme zu lösen sind, und in dem sie auch Code untereinander austauschen können. Wir haben auch regelmäßige Treffen organisiert, um die Aufgaben während der Entwicklung des Projekts zu synchronisieren.

T0.3 Dissemination: Die vierte Auflage der Summer School on Cooperative Interacting Vehicles war für den Sommer 2023 geplant. Aufgrund von Problemen in der Agenda der Organisatoren fand sie jedoch nicht. Ziel der Summer School war es, Doktoranden und Young Professionals von Universitäten und Forschungseinrichtungen zu Tutorien, Keynotes, Posterpräsentationen und

interaktiven Workshops über kooperative Wahrnehmung, kooperative Bewegungsplanung, implizite und explizite Interaktion zusammenzubringen. In Zusammenarbeit mit unseren französischen Partnern von Mines Paris haben wir einen Workshop auf dem IEEE Intelligent Vehicles Symposium mit dem Titel “Interaction driven Behavior Prediction and Planning for Autonomous Vehicles” organisiert.

2.2.2 WP1 Datenerhebung

T1.1 Erfassung der genauen Flugbahndaten: Ursprünglich war geplant, das Testfeld Autonomes Fahren in Karlsruhe zu nutzen. Letztlich stellte sich jedoch heraus, dass die Bildfrequenz der dort installierten Kameras sehr niedrig war (nur 1–2 Bilder pro Sekunde) und die Verfügbarkeit der Ressourcen begrenzt war. Daher entschieden wir uns für einen neuen Standort zur präzisen Datenerfassung (Hirtenberg/Haid-und-Neu-Straße), an dem wir die Aufnahmen mit unserer eigenen Drohne anstelle der Testfeld-Kameras durchführten.

Um die Aufzeichnung zu verbessern und den Mehrwert unseres Datensatzes zu erhöhen, haben wir mit unserer Drohne + experimentelles Fahrzeug mit Sensoren + statische Lidars an der Kreuzung während 90 Minuten (Juli-2022) aufgezeichnet. Aufgrund der Batterien der Drohne konnten wir nicht länger aufzeichnen. Nach der Verarbeitung der Daten stellten wir fest, dass die Bewegung der Drohne nur schwer zu kompensieren ist und auch die Sicht von Drohne anders ist als die Sicht einer Überwachungskamera. Da wir diese Daten als Bodenwahrheit verwenden könnten, dachten wir, dass die Bewegung der Drohne es nicht erlaubt, sie wirklich als Bodenwahrheit zu verwenden. Wir waren besorgt, dass, wenn wir die erwarteten 6 Stunden mit der Drohne aufzeichnen und den großen Aufwand für die Verarbeitung der Trajektorien betreiben, das trainierte Modell vielleicht nicht auf den 500-Stunden-Datensatz anwendbar ist, der aus anderen Blickwinkel (Dach eines Gebäudes) aufgenommen wurde. Daher war es unsere Absicht, eine Kreuzung zu finden, an der wir den 500-Stunden-Datensatz aufzeichnen und einen 6-Stunden-Datensatz mit Kameras + Versuchsfahrzeugen + statischen Lidars an der gleichen Stelle erstellen können. Wir haben ein Gebäude auf unserem Campus gefunden, in dem wir Kameras zur Überwachung des Verkehrs installieren können. Wir haben die Kameras im Januar 2023 installiert, sie kalibriert und die Aufzeichnungstools und das Synchronisationssystem vorbereitet. Die folgenden Bilder zeigen, wie die Kameras nach ihrer Kalibrierung in eine Draufsicht projiziert werden können.

Das KIT hat in Zusammenarbeit mit unseren französischen Partnern Valeo und Mines Paris sowie mit Unterstützung von SafeAD und FZI die Aufzeichnung von sechs Stunden Daten koordiniert. Dabei kamen drei Versuchsfahrzeuge sowie stationäre LiDARs und Kameras zum Einsatz, um präzise Verkehrsdaten zu erfassen. Diese Daten können für zukünftige Forschungsarbeiten genutzt werden, beispielsweise im Bereich der kollaborativen Wahrnehmung und Verfolgung.

T1.2 Aufzeichnung von Verkehrsvideos: Wir verfügen über ein synchronisiertes Kamerasystem, das bis zu 120 Stunden ohne Komprimierung aufzeichnen kann. Wir haben mehr als 500 Stunden komprimiertes Video unter verschiedenen Wetterbedingungen aufgezeichnet, darunter Tag, Nacht und starker Regen.

T1.4 Generierung von Karten-Trainingsdaten: Das KIT hat mit SafeAD zusammengearbei-

tet, um die erforderlichen Daten für die kartografische Ausbildung zu erstellen.

2.2.3 WP2 Groß angelegte Datenkommentierung

T2.1 Datenstrukturen: Auf der Grundlage der Analyse der im Juli-2022 aufgezeichneten Daten haben wir eine erste Version der Datenstruktur für den Datensatz entworfen. Wir sollten 2D (und 3D) Annotationen berücksichtigen, einschließlich Objektklasse, Geschwindigkeit, eindeutige ID und zusätzliche Informationen wie Zeitstempel, Wetter, Interaktionsgrad usw. , wenn wir in der Lage sind, Tausende von Trajektorien zu generieren, müssen die Nutzer des Datensatzes möglicherweise zwischen Trajektorien oder Merkmalen unterscheiden, die spezielle Verhaltensweisen kennzeichnen.

T2.3 Training eines ML-Modells für 3D-Karten: Das KIT unterstützte SafeAD bei der Erkennung von sich bewegenden Verkehrsteilnehmern aus Luftbildern mit Hilfe von KI-Modellen durch die Erstellung einer Masterarbeit zum Thema.

T2.4 Trajektoriendatensatz mit semantischen Strukturen: Wir haben eine semantische Karte erstellt, mit der wir die extrahierten Trajektorien aufwerten können. Wir können Verhaltensweisen wie einen Fußgänger finden, der die Straße außerhalb des Zebrastreifens überquert, oder spezielle Fahrzeuge, die auf dem Bürgersteig fahren.

2.2.4 WP3 Verhaltensmodellierung

T3.1 Unabhängige Vorhersage: Das KIT hat eine Basismethode entwickelt, die Karteninformationen und die letzten Positionen eines Verkehrsteilnehmers verwendet, um eine Vorhersage für die nächsten Sekunden zu erstellen. Die Methode wird derzeit mit bestehenden öffentlichen Datensätzen im Bereich der Trajektorienvorhersage getestet. Sie wird jedoch auch auf unserem eigenen Datensatz getestet und mit Methoden verglichen, die Wechselwirkungen berücksichtigen.

T3.2 Identifizierung von Wechselwirkungen: Das KIT hat im Rahmen dieses Projekts eine Methode zur Identifizierung von Interaktionen entwickelt. Die Methode wurde erfolgreich auf der IEEE 27th International Conference on Intelligent Transportation Systems (ITSC) 2024 veröffentlicht. Das Paper [4] trägt den Titel "SceneMotion: From Agent-centric Embeddings to Scene-wide Forecasts". Selbstfahrende Fahrzeuge sind auf multimodale Bewegungsvorhersagen angewiesen, um effektiv mit ihrer Umgebung zu interagieren und sichere Manöver zu planen. Wir stellen SceneMotion vor, ein aufmerksamkeitsbasiertes Modell zur Vorhersage von szeneweiten Bewegungsmodi von mehreren Verkehrsagenten. Unser Modell transformiert lokale agenten-zentrierte Einbettungen in szeneweite Prognosen unter Verwendung eines neuartigen latenten Kontextmoduls. Dieses Modul lernt einen szeneweiten latenten Raum aus mehreren agenten-zentrierten Einbettungen, was eine gemeinsame Vorhersage und Interaktionsmodellierung ermöglicht. Die konkurrenzfähige Leistung in der Waymo Open Interaction Prediction Challenge beweist die Effektivität unseres Ansatzes. Darüber hinaus clustern wir zukünftige Wegpunkte in Zeit und Raum, um die



(a) Semantische Karte



(b) Luftbildaufnahme

Abbildung 2.1: Wir haben eine semantische Karte erstellt, auf der wir den extrahierten Trajektorien zusätzlichen Kontext hinzufügen können.

Interaktion zwischen Agenten zu quantifizieren. Wir führen alle Modi zusammen und analysieren jeden Modus unabhängig, um festzustellen, welche Cluster durch Interaktion aufgelöst werden oder zu Konflikten führen. Unsere Implementierung ist verfügbar unter: <https://github.com/kit-mrt/future-motion>

T3.3 Vorhersage kollektiver Interaktionen: Wir haben eine Vorhersagemethode entwickelt und sie in einem Artikel auf dem IEEE ICRA2023 Workshop on Pretraining for Robotics (PT4R), London, UK, veröffentlicht. In dem Artikel [1] mit dem Titel "Road Barlow Twins: Redundancy Reduction for Motion Prediction" (Redundanzreduktion für Bewegungsvorhersage) stellen wir eine neuartige selbstüberwachte Pre-Training-Methode für die Bewegungsvorhersage vor. Unsere Methode basiert auf Barlow Twins und wendet das Prinzip der Redundanzreduktion auf Einbettungen an, die aus HD-Maps generiert werden. Durch unsere Methode lernen Deep-Learning-Modelle augmentationsinvariante Merkmale von HD-Maps. Wir stellen die Hypothese auf, dass ein Verständnis der Umgebung mit diesen Merkmalen schneller erlernt werden kann. Wir trainieren mehrere große Transformer-Modelle vor und stimmen sie anschließend auf die Bewegungsvorhersage ab. Unsere Experimente zeigen, dass die vorgeschlagene Pre-Training-Methode mADE und mFDE um 12% und 15% verbessern kann und das kontrastive Lernen mit PreTraM und SimCLR in einer semi-supervised Umgebung übertrifft.

T3.4 Bewertung der Vorhersage anhand des Verhaltens: Wir haben dazu zwei Methoden

entwickelt, die erste Methode [3] wurde in Transactions on Machine Learning Research (2024). issn: 2835-8856 unter dem Titel: "RedMotion: Motion Prediction via Redundancy Reduction". Die Vorhersage der zukünftigen Bewegung von Verkehrsteilnehmern ist für selbstfahrende Fahrzeuge unerlässlich, um deren sicheren Betrieb zu gewährleisten. Wir stellen RedMotion vor, ein Transformatormodell zur Bewegungsvorhersage, das zwei Arten der Redundanzreduktion beinhaltet. Die erste Art der Redundanzreduzierung wird durch einen internen Transformator-Decoder induziert und reduziert eine variable Menge von Straßenumgebungs-Token, wie z.B. Straßengraphen mit Agenten-Daten, auf eine Einbettung fester Größe. Die zweite Art der Redundanzreduktion ist ein selbstüberwachtes Lernziel und wendet das Prinzip der Redundanzreduktion auf Einbettungen an, die aus erweiterten Ansichten von Straßenumgebungen erzeugt werden. Unsere Experimente zeigen, dass unser Ansatz des Repräsentationslernens PreTraM, Traj-MAE und GraphDINO in einer halbüberwachten Umgebung übertreffen kann. Unser RedMotion-Modell erzielt Ergebnisse, die mit denen von Scene Transformer oder MTR++ konkurrenzfähig sind.

Die zweite Methode [2] wurde auf der 8th Annual Conference on Robot Learning (2024) veröffentlicht. Die Methode mit dem Titel "JointMotion: Joint Self-supervision for Joint Motion Prediction" ist ein selbstüberwachter Pre-Training-Ansatz zur Vorhersage gemeinsamer Bewegungen in autonomen Fahrzeugen. Unsere Methode optimiert gemeinsam ein Ziel auf Szenenebene, das Bewegung und Umgebung verknüpft, sowie ein Ziel auf Instanzebene zur Verfeinerung der gelernten Repräsentationen. Die Repräsentationen auf Szenenebene werden durch nicht-kontrastives Ähnlichkeitslernen vergangener Bewegungssequenzen und des Umgebungskontexts erlernt. Auf Instanzebene verwenden wir maskierte automatische Kodierung, um multimodale Polyliniendarstellungen zu verfeinern. Ergänzt wird dies durch einen adaptiven Pre-Training-Decoder, der JointMotion in die Lage versetzt, über verschiedene Umgebungsrepräsentationen, Fusionsmechanismen und Datensatzmerkmale hinweg zu generalisieren. Unsere Methode reduziert den gemeinsamen finalen Positionsfehler (displacement error) der Modelle Wayformer, HPTR und Scene Transformer um 3%, 8% bzw. 12% und ermöglicht Transferlernen zwischen den Datensätzen Waymo Open Motion und Argoverse 2 Motion Forecasting.

Die vorgenommenen Änderungen haben dazu beigetragen, die Projektziele fristgerecht zu erreichen und damit unsere Erwartungen in Bezug auf mehrere Veröffentlichungen in Konferenzen und hochrangigen achzeitschriften zu erfüllen. Die Aussichten, die Projektziele zu erreichen, haben sich seit des Projekts nicht grundlegend geändert. Im Folgenden finden Sie eine Zusammenfassung aller Projektziele und Stand.

2.2.5 Ziele in Bezug auf Rohdaten zum Verkehr

Quantitative Zielsetzung

- Rohverkehrsdatensätze D1.1 und D1.2: Erfassung des Umfangs potenziell extrahierbarer Trajektorien.
- Quantifizierungskriterium: Aufzeichnungsdauer in Stunden. Während des Projekts wäre es sinnvoll, die tatsächliche Dauer mit der geschätzten zu gewichten.
- Quantitatives Ziel: D1.1 - 6 Stunden; D1.2 - 500 Stunden.

Erreichtes Ergebnis

Der Datensatz D1.1 wurde im Juli 2022 (intern) und D1.2 (öffentlich) im Juli 2023 unter verschiedenen Wetterbedingungen wie Tag, Nacht und Starkregen aufgezeichnet. D1.2 wird unter <https://haibrid.gitlab.io> verarbeitet und veröffentlicht.

2.2.6 Ziele für den Flugbahndatensatz

Quantitative Zielsetzung

- Flugbahndatensatz M1.2: Schätzung des Interaktionspotenzials
- Quantifizierungskriterien: Anzahl der Trajektorien. Das Interaktionspotenzial würde wahrscheinlich besser durch die Anzahl der interagierenden Trajektorien beschrieben werden, wenn wir bereits wüssten, wann eine Interaktion stattfindet. Wir glauben, dass dies proportional zur Anzahl der Trajektorien und der Art des Ortes sein sollte. Dieses Kriterium kann im Laufe des Projekts erweitert werden.
- Quantifiziertes Ziel: M1.2 enthält 100 000 Flugbahnen

Erreichtes Ergebnis

An der Kreuzung gab es mehrere interaktive Situationen, und wir erhielten viel mehr Trajektorien als erwartet. Die verarbeiteten Daten summierten sich auf insgesamt 1,25 Mio. Trajektorien mit Klassenverteilung: Person (513K), unbekannt (25K), Auto (576K), LKW (45K), Bus (20K), Radfahrer (50K), Zug (25K).

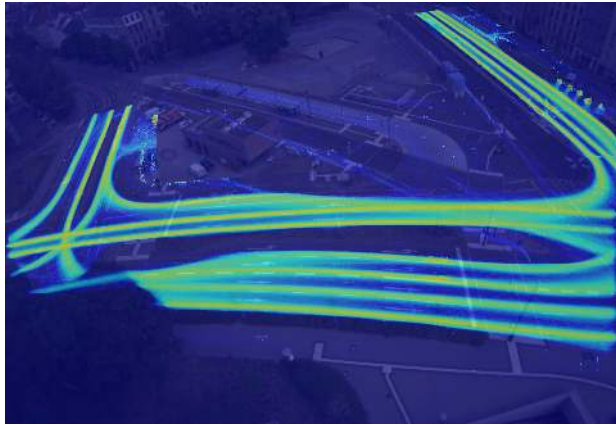
2.2.7 Zielsetzungen für Methoden zur Extraktion von Flugbahnen aus Videos

Quantitative Zielsetzung

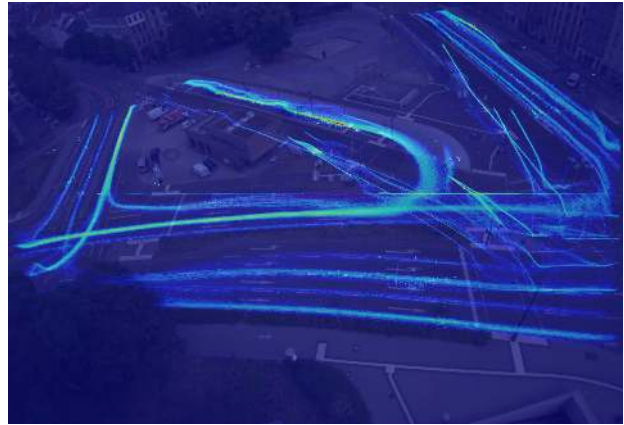
- Methoden zur Extraktion von Trajektorien aus Videos: Minimierung des menschlichen Eingriffs.
- Quantifizierungskriterium: Prozentsatz der automatisch extrahierten Trajektorien. Obwohl dieses Quantifizierungskriterium keine exakte Metrik des Ziels ist (eine kleine Menge an menschlicher Intervention könnte es wert sein), werden wir es durch Messungen an (langen) Videos mit Bodenkontakt bewerten. Wahrheit (wie von WP2 bereitgestellt).
- Quantifiziertes Ziel: 95% der Flugbahnen in M1.2 werden automatisch extrahiert.

Erreichtes Ergebnis

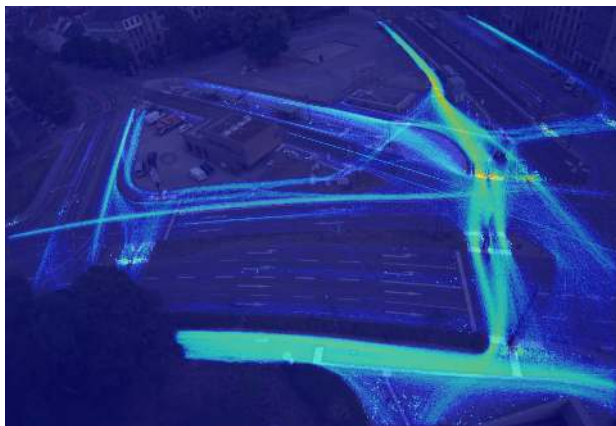
Der Datensatz wurde automatisch verarbeitet, und die resultierenden Trajektorien wurden mithilfe von Visualisierungstools wie Heatmaps über Hintergrundbildern analysiert. Zusätzlich wurden bestimmte Daten wie Beschleunigungen, Trajektorienlängen und andere Merkmale untersucht, um heuristische Regeln zum Herausfiltern falscher Trajektorien aufzustellen. Wir können die Anzahl der menschlichen Eingriffe nicht quantifizieren, ohne 1,2 Millionen Trajektorien manuell zu



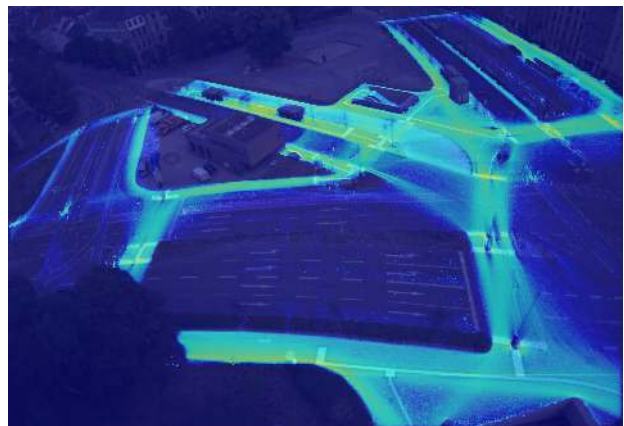
(a) Auto



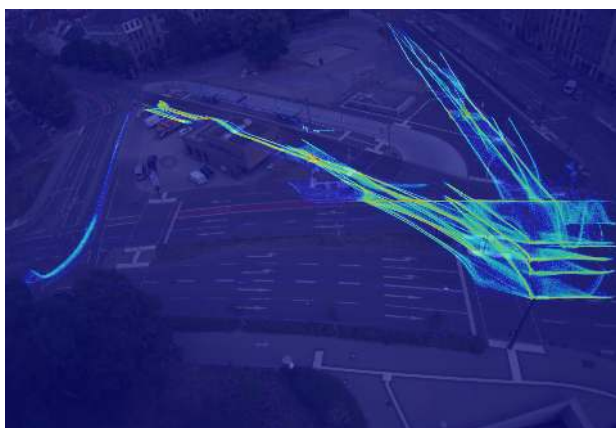
(b) Bus



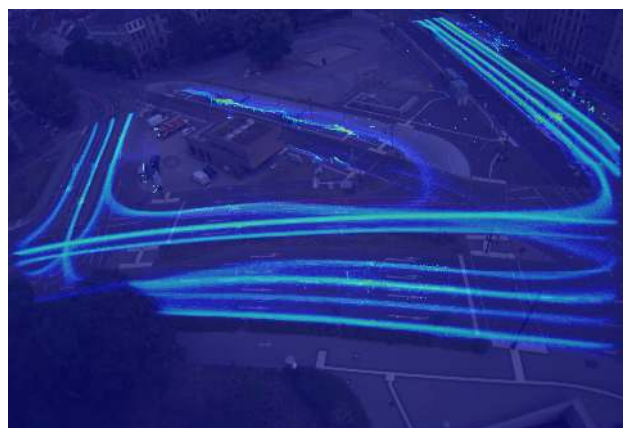
(c) Radfahrer



(d) Person



(e) Zug



(f) LKW

Abbildung 2.2: Verteilung der berechneten Trajektorien nach Verkehrsteilnehmerkategorie.

überprüfen. Stattdessen können wir durch die Analyse der Heatmaps beurteilen, ob die erhaltenen Trajektorien angemessen sind oder ob eine manuelle Überwachung erforderlich ist.

Wir haben eine mögliche Überschneidung zwischen den Klassen Bus und Bahn festgestellt, da einige Trajektorien in den jeweiligen Heatmaps auftauchen, aber nicht den erwarteten Mustern

entsprechen. Wir werden diese Fälle manuell überprüfen, was immer noch eine automatische Verarbeitung von mehr als 94% der Trajektorien ermöglicht. Unserer Meinung nach ist die Datenmenge, die wir erhalten haben, recht gut, wir haben weit mehr Flugbahnen erhalten, als wir ursprünglich erwartet hatten. Unsere Pläne für die Zukunft sehen vor, die Erkennungs- und Verfolgungspipeline weiter zu verbessern, um bessere Ergebnisse zu erzielen und 3D-Projektionen einzubeziehen, indem wir die neuesten Fortschritte bei der monokularen kamerabasierten Entfernungsschätzung nutzen.

2.2.8 Zielvorgaben für die Qualität der Extraktion von Flugbahnen

Quantitative Zielsetzung

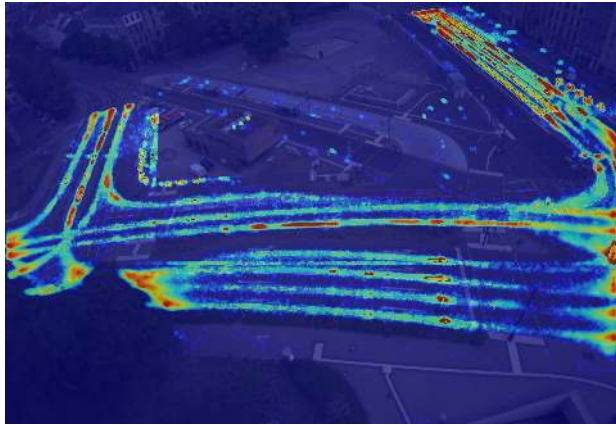
- Qualität der Trajektorienextraktion: Vollständigkeit und Korrektheit der extrahierten Trajektorien.
- Quantifizierungskriterien: Die Vollständigkeit wird anhand des prozentualen Anteils der einzelnen Flugbahnen gemessen, die nicht in Teile zerlegt sind. Die Korrektheit wird anhand des Durchschnitts (über die Zeit) der Genauigkeit der Position des Objekts gemessen. Diese beiden Kriterien werden durch den Vergleich mit präzisen, aus der Datenfusion (T2.1) gewonnenen Trajektorien mit einer kleinen Menge manuell beschrifteter Bodenwahrheits-Trajektorien und Lidar-basierten Positionsschätzungen, sofern verfügbar, gemessen.
- Quantifiziertes Ziel: 95% Vollständigkeit; Korrektheit: 60 cm (als INTERACTION).

Erreichtes Ergebnis

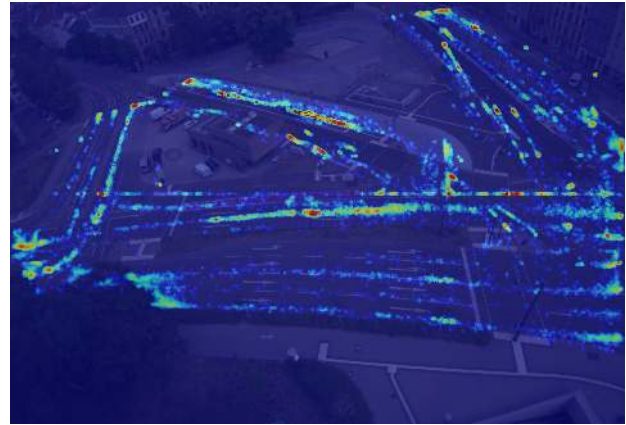
Was die Qualität der Trajektorienextraktion betrifft, so basiert das Qualitätskriterium auf der Anzahl der Trajektorien, die nicht in mehrere Teile fragmentiert wurden. Um dies zu bewerten, haben wir eine Heatmap erstellt, die die Bereiche zeigt, in denen die Trajektorien beginnen und enden.

Bei Autos beispielsweise beginnen oder enden die meisten Trajektorien an den Rändern des Bildes. Bei anderen Klassen, wie z. B. Fußgängern, gibt es jedoch viele Erkennungen, die in Bereichen außerhalb der Bildgrenzen beginnen oder enden. Im oberen Bereich (Straßenbahnhaltestelle) haben viele Trajektorien ihren Ursprung, da dort erwartungsgemäß Menschen ein- und aussteigen. Außerdem beobachten wir Erfassungen in der Nähe von Fußgängerüberwegen, was darauf hindeuten könnte, dass einige Fußgänger verloren gegangen sind und später in diesen Bereichen erfasst wurden. Dies könnte darauf hindeuten, dass eine bestimmte Anzahl von Trajektorien in mehrere Teile zerlegt wurde. Zukünftige Verbesserungen im Tracking sind notwendig, um die Anzahl der fragmentierten Trajektorien zu reduzieren. Daher werden wir das gesamte System weiter erforschen und verfeinern (Erkennung+ Verfolgung + Klassifizierung).

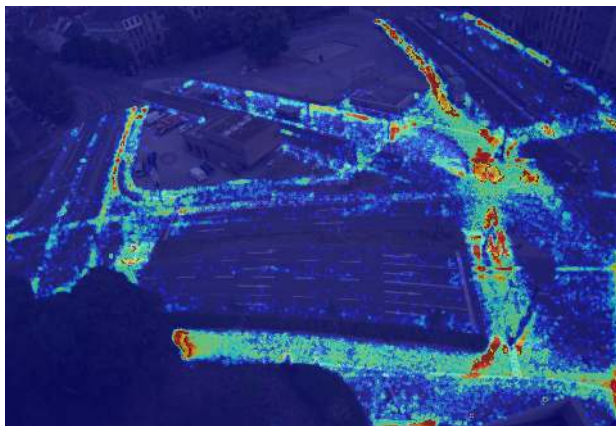
Der andere Maßstab, der zur Bewertung des Ziels herangezogen wird, ist die Erkennungsgenauigkeit. Mit der aktuellen Kamera Einrichtung erreichen wir eine Genauigkeit von 2 cm/px in kameranahen Bereichen und 20 cm/px in weiter entfernten Regionen. Als Kriterium haben wir eine Korrekturschwelle von 60 cm festgelegt, was einem Fehler von 30 px in nahe gelegenen und 3 px in weiter entfernten entspricht. Aufgrund des großen Datenvolumens in diesem Projekt konnten wir die Genauigkeit der Erkennungen nicht manuell überprüfen, da dies nicht durchführbar wäre.



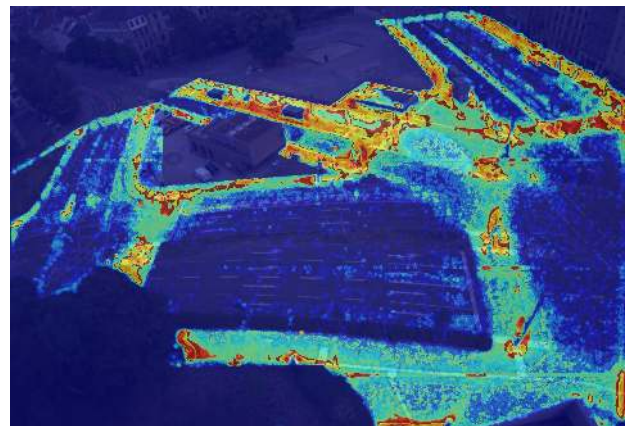
(a) Auto



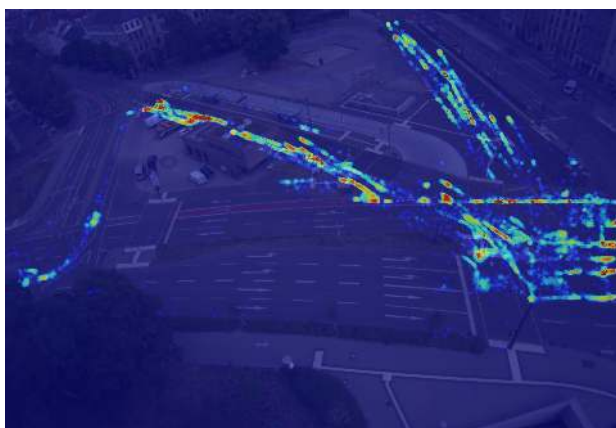
(b) Bus



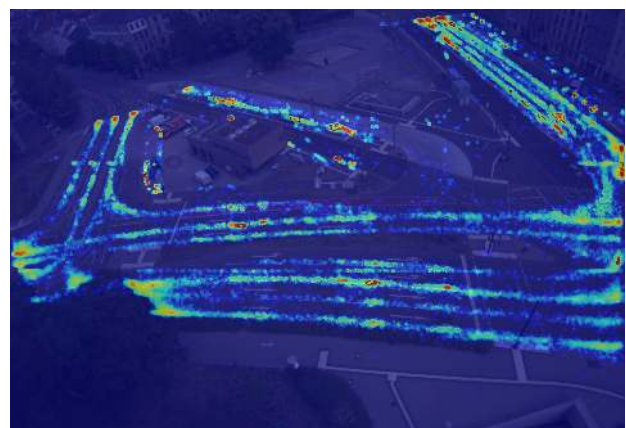
(c) Radfahrer



(d) Person



(e) Zug



(f) LKW

Abbildung 2.3: Anfangs- und Endpunkte der nach Verkehrsteilnehmern geordneten Flugrouten.

Angesichts der Qualität der verwendeten Modelle sind diese Werte jedoch sehr realistisch, und wir glauben, dass sie in den meisten Fällen erreicht wurden.

2.3 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die Identifizierung von Out-of-Distribution-Ereignissen (OOD) birgt ein erhebliches Potenzial sowohl für die akademische Forschung als auch für praktische Anwendungen im Automobilssektor – insbesondere für die Weiterentwicklung von Technologien für autonomes Fahren und fortschrittliche Fahrerassistenzsysteme (ADAS), die darauf abzielen, Unfälle zu vermeiden und das Verhalten der Verkehrsteilnehmer besser zu verstehen. Darüber hinaus könnte unser Ansatz kommunale Dienste unterstützen, indem er die Erkennung von Verkehrsstörungen oder ungewöhnlichen Ereignissen erleichtert und so ein effizienteres Verkehrsmanagement sowie eine schnellere Reaktion auf Notfälle ermöglicht.

Um unseren Datensatz weiter anzureichern und die Fähigkeiten zur Verhaltensprognose zu verbessern, planen wir eine Zusammenarbeit mit der Stadt Karlsruhe und dem Unternehmen, das die Ampelinfrastruktur verwaltet, um Zugang zu Echtzeit-Signaldaten zu erhalten. Parallel dazu wollen wir unsere Partnerschaft mit Mines Paris ausbauen, da ihr Fachwissen eng mit einer unserer neuen Forschungsrichtungen übereinstimmt und eine ideale Grundlage für künftige gemeinsame Projekte bildet.

Die Zusammenarbeit mit Mines Paris war äußerst gewinnbringend. Gemeinsam haben wir Workshops auf internationalen Konferenzen organisiert und arbeiten derzeit an der Ausarbeitung von Anträgen für europäische Projekte, die auf Ideen aus diesem Projekt basieren. Wir beabsichtigen, diese Kooperation auch in Zukunft fortzusetzen. Solche bilateralen Projekte zwischen Frankreich und Deutschland sind zweifellos sehr wertvoll, und falls in Zukunft neue Ausschreibungen veröffentlicht werden, würden wir gerne erneut daran teilnehmen.

2.4 Relevante Ergebnisse von dritter Seite

Während der Entwicklung des Projekts haben wir neue Datensätze gefunden, die unseren Zielen in Haibrid nahe kommen. Einige Beispiele sind:

- **LUMPI** wurde 2022 auf dem IEEE Intelligent Vehicles Symposium (IV) veröffentlicht. Sie bieten 3D Bounding Boxes von Autos, Fußgängern, Radfahrern und Lastwagen. Sie verfügen über 2 statische Kameras auf verschiedenen Gebäuden und 5 statische Lidars, um die Trajektorien der Verkehrsteilnehmer zu erstellen. Die Dauer des Datensatzes beträgt 2h25 Minuten und ist damit kürzer als unsere 6h-Aufnahme. Unser Datensatz D1.1 enthält auch Lidar- und Kamerainformationen von 3 verschiedenen Versuchsfahrzeugen.
- **MONA** wurde 2022 auf der IEEE 25th International Conference on Intelligent Transportation Systems (ITSC) veröffentlicht. Sie liefern 2D-Erkennungen auf statischen Kameras mit einem durchschnittlichen Fehler von 0,51 m. Wir erwarten präzisere Trajektorien durch die Verwendung von statischen Lidars und Kameras, die näher an den zu erkennenden Objekten sind. Die Aufnahmezeit beträgt 130 Stunden. Der Standort befindet sich jedoch auf einer Autobahn, auf der es weniger Interaktionen gibt als an unserem Aufnahmeort, und die hauptsächlich aus Autos und Lastwagen, aber nicht aus Fußgängern und Radfahrern besteht.

- **TUMTraf** ist ein Multisensor-Datensatz für die Mobilitätsforschung (veröffentlicht in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)). Mit einer Konstellation von 7 Sensorstationen, die mit mehr als 60 hochmodernen und multimodalen Sensoren ausgestattet sind, und mit einer Straßennetzabdeckung von etwa 3,5 Kilometern, ist der TUMTraf-Datensatz bietet Mobilitätsforschern, Industriepartnern, Behörden und politischen Entscheidungsträgern hochauflösende kuratierte Datensätze, die durch die Erfassung des realen Verkehrs auf Autobahnen, Landstraßen und städtischen Verkehrsknotenpunkten erstellt wurden. Die Datensätze enthalten einerseits beschriftete, zeitsynchronisierte und anonymisierte multimodale Sensordaten von Flächenkameras, Dopplerradars, Lidars und ereignisgesteuerten Kameras für eine Vielzahl von verkehrs- und wetterbedingten Szenarien und andererseits abstrakte digitale Zwillinge der Verkehrsobjekte mit positions- und trajektorienbezogenen Informationen.

2.5 Veröffentlichungen

Das Ergebnis wurde in den folgenden Publikationen veröffentlicht:

- [1] Royden Wagner, Marvin Klemp, Carlos Fernandez Lopez und Omer Sahin Tas. "Road Barlow Twins: Redundancy Reduction for Motion Prediction". In: *ICRA2023 Workshop on Pretraining for Robotics (PT4R)*. 2023. URL: <https://openreview.net/forum?id=HI12AzV3ZA>.
- [2] Royden Wagner, Omer Sahin Tas, Marvin Klemp und Carlos Fernandez. "JointMotion: Joint Self-supervision for Joint Motion Prediction". In: *8th Annual Conference on Robot Learning (CORL)*. 2024. URL: <https://openreview.net/forum?id=0znxxPLiH>.
- [3] Royden Wagner, Omer Sahin Tas, Marvin Klemp, Carlos Fernandez und Christoph Stiller. "RedMotion: Motion Prediction via Redundancy Reduction". In: *Transactions on Machine Learning Research* (2024). ISSN: 2835-8856.
- [4] Royden Wagner, Ömer Şahin Taş, Marlon Steiner, Fabian Konstantinidis, Hendrik Königshof, Marvin Klemp, Carlos Fernandez und Christoph Stiller. "SceneMotion: From Agent-Centric Embeddings to Scene-Wide Forecasts". In: *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. 2024, S. 812–818. DOI: 10.1109/ITSC58415.2024.10919533.