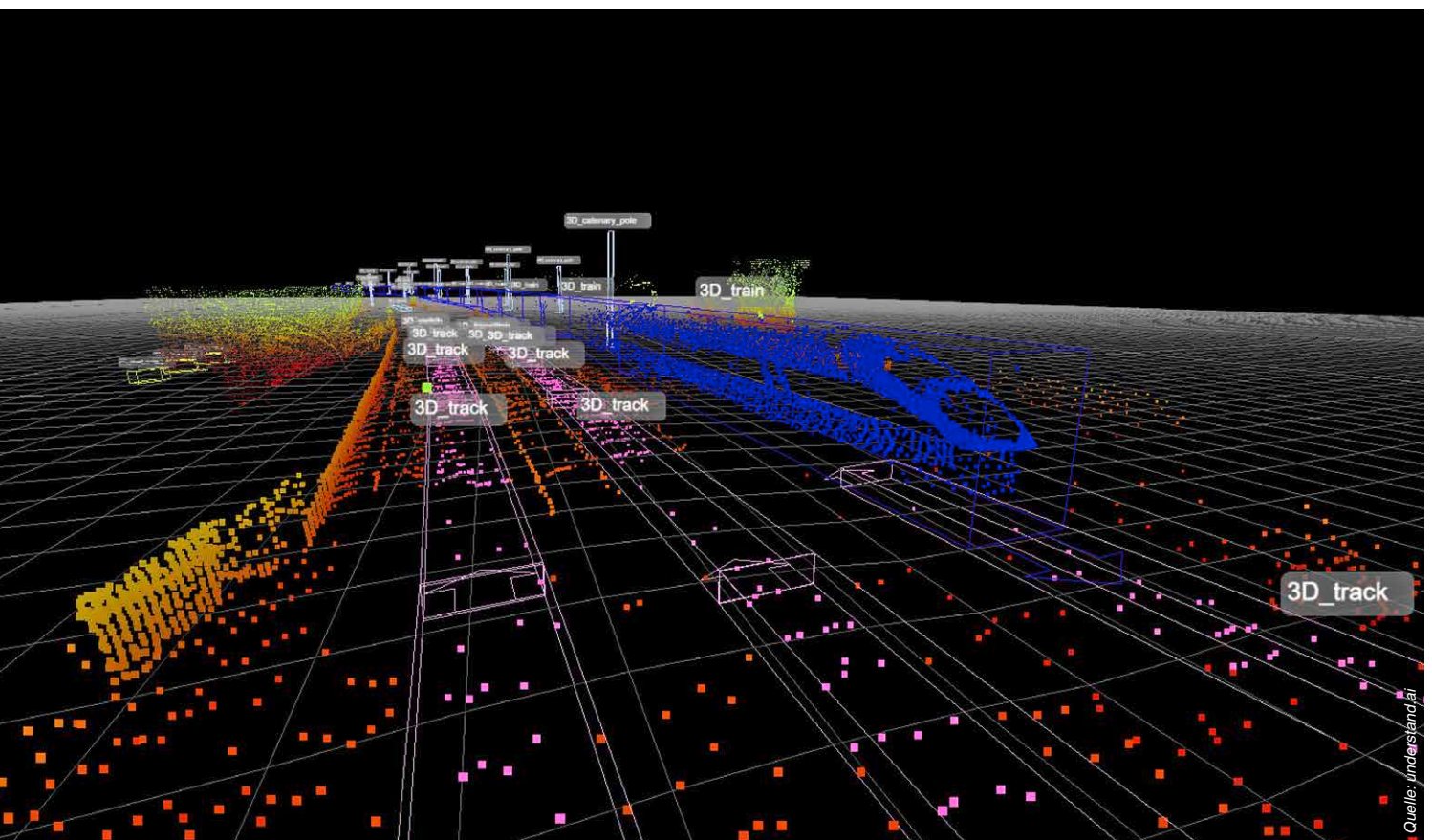


Automatisiertes Fahren

Multisensordatensatz für die Umfeldüberwachung von Schienenfahrzeugen



Claudio Diotallevi, Account Executive, **Rodrigo Gudiño**, Project Manager, **Zaharia Pachalieva**, Solution Architect, understand.ai GmbH, Karlsruhe, **Dr. Philipp Neumaier**, Head of Data Factory, **Patrick Denzler**, Product Owner Data Platform, und **Dr. Martin Köppel**, Project Owner Data Annotation, DB InfraGO AG, Berlin

Die Überwachung von Strecke und Gleisumfeld spielt im Rahmen des automatisierten Fahrens eine wichtige Rolle. So kann sie z. B. im Automatisierungsgrad (Grade of Automation) GoA2, bei dem der Triebfahrzeugführende noch an Bord ist, als Assistenzsystem für die Streckenbeobachtung zum Einsatz kommen. Im vollautomatisierten, fahrerlosen Fahren im GoA4 übernehmen diese Systeme schließlich die Umfeldüberwachung komplett selbstständig. Mit Hilfe von Künstlicher Intelligenz (KI) reagieren sie automatisch auf Risiken und gefährliche Ereignisse auf der Strecke.

Um KI-Verfahren zur Umfeldwahrnehmung entwickeln zu können, sind sehr große Datenmengen erforderlich. Sogenannte Annotationen liefern zusätzliche Informationen, welche die relevanten Objekte in den Daten beschreiben. Im Rahmen der Sektorinitiative Digitale Schiene Deutschland (DSD)^[1] hat die DB InfraGO AG zusammen mit der understand.ai GmbH (UAI),^[2] einem auf diese Thematik spezialisiertes Unternehmen, einen vollumfänglichen Multisensordatensatz erzeugt.

Der Datensatz beinhaltet über 7 Millionen Annotationen von bahnspezifischen und generischen Wahrnehmungsobjekten. Die Arbeiten an dem Datensatz sind abgeschlossen und können nun der Industrie zur Entwicklung von Umfeldüberwachungssystemen bereitgestellt werden.

Schienenfahrzeuge zur Aufzeichnung von Multisensordaten

Um geeignete Daten aufzeichnen zu können, wurden zwei Schienenfahrzeuge mit Sensorik ausgestattet. Das erste Fahrzeug war ein Gleisarbeitsfahrzeug (GAF). Im Rahmen eines internen Projektes der DB InfraGO AG wurden Sensoren an die Front eines GAF montiert. Dann wurden in mehreren Testfahrten in Hamburg und Berlin Sensordaten aufgezeichnet. Die Daten umfassen reguläre und nicht-reguläre Betriebsumgebungen und -situationen. Die Sensor-konfiguration besteht aus insgesamt sechs Farbbildkameras, drei Infrarotkameras (IR), sechs LiDAR-Sensoren mit unterschiedlichen Reichweiten und Erfassungsbereichen, einen Radar-Sensor sowie Positions- und Beschleunigungssensoren. Weitere technischen Spezifikationen sind in [3] aufgeführt.

Das zweite Fahrzeug wurde im Kooperationsprojekt Sensors4Rail^[4] mit Sensorik ausgestattet. Im Rahmen dieses Projektes der Deutschen Bahn, Bosch Engineering, HERE Technologies, Ibeo Automotive Systems (heute: MicroVision) und Siemens Mobility wurden Technologien zur Umfeldwahrnehmung erprobt. Ein

Fahrzeug der S-Bahn Hamburg (BR 472) wurde mit drei Farbbild-, einer Infrarotkamera, sechs Lidaren und vier Radaren ausgerüstet. Die Daten wurden auf der 23 Kilometer langen Strecke der Linie S21 zwischen Hamburg-Berliner Tor und Bergedorf aufgezeichnet.

Datenannotationen und Objektklassen

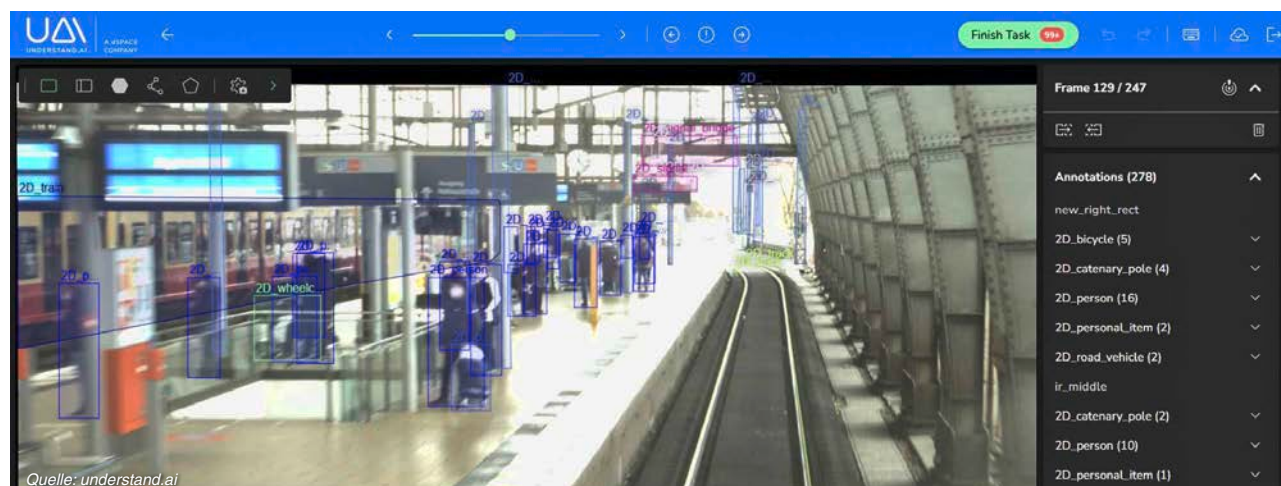
Bei der Datenannotation handelt es sich um Markierungen von Objektklassen in den Sensordaten. Hierbei gibt es verschiedene Arten von Annotationen, z. B. 2D- oder 3D-Bounding Boxen, Segmentierungen, Polygone und Polylinien. Die Annotation enthält auch zusätzliche Informationen zu den Objekten über Attribute. Seit 2021 wurden gemeinsam von UAI und DB InfraGO AG Annotationen für 21 Objektklassen erzeugt.

Die angewandte Annotationsmethodik umfasste sowohl automatische Softwaretools als auch manuelle Annotation. Die Softwaretools wurden in allen Projektphasen (Datenimport, Datenqualitätskontrollen, Annotation und Datenexport) verwendet, um die Effizienz zu steigern und somit die manuellen Annotationszeiten zu reduzieren. Alle Daten wurden zunächst in den – vom Lidar aufgezeichneten – 3D Punktwolken von Experten annotiert. Über Projektionsfunktionen wurden diese Annotationen dann in die weiteren Sensoren projiziert.

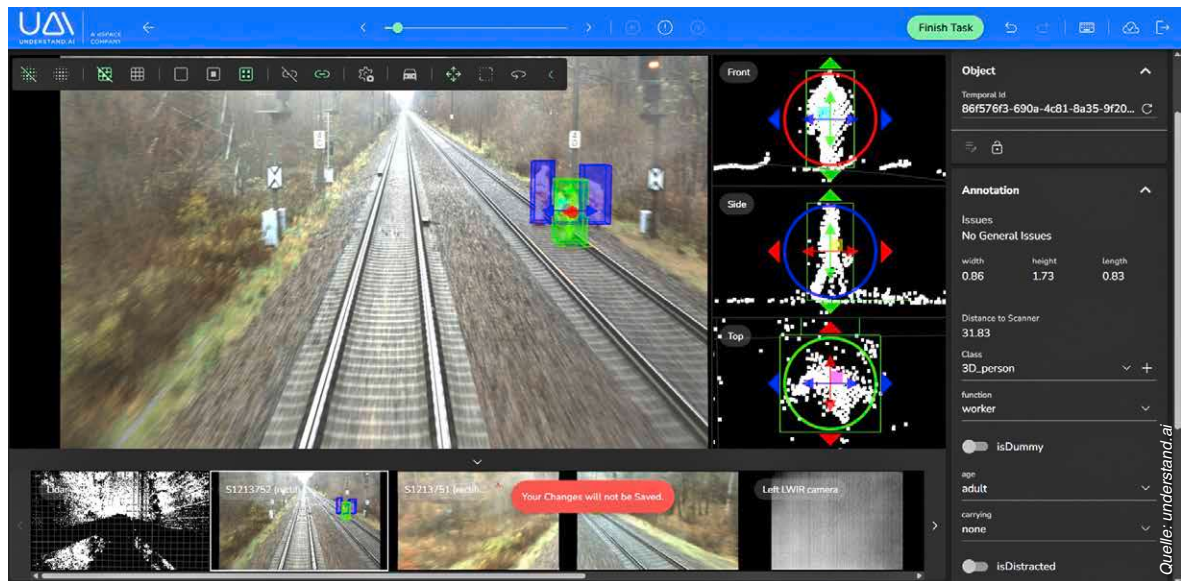
Im Folgenden sind die verschiedenen Projektionsfunktionen beschrieben:

- (1) Projektion von 3D-Bounding Boxen auf 2D-Bounding Boxen
- (2) Projektion von 3D-Bounding Boxen auf 2D-Polygone
- (3) Projektion von 2D-RGB auf 2D-Infrarotkameras (IR)
- (4) Projektion von 3D-Lidar-Punkten auf Radarbilder

2D-Bounding Boxen von Personen, annotiert im Kamerabild



3D-Bounding Boxen
von Arbeitern, projiziert
in das Kamerabild



Im Anschluss wurden alle projizierten Annotationen geprüft und gegebenenfalls nachgearbeitet. Während der Annotationsprojekte konnte UAI einen Spitzenlieferdurchsatz von 140.000 Annotationen/Woche realisieren.

Qualitätskontrolle

Ein strenges Datenqualitätsmanagement war entscheidend, um die Qualitätsziele zu erreichen. Das Projektteam hatte mehrere Qualitätsebenen genutzt, von der anfänglichen Qualitätsbewertung der Rohdaten über die Prototypenentwicklung, die automatisierte Qualitätssicherung in der Produktion bis hin zur abschließenden Qualitätssicherung vor bzw. nach der Auslieferung. Im Folgenden werden die verschiedenen Schritte beschrieben.

- **Qualitätsbewertung Rohdaten:** Es wurden Tools und manuelle Kontrollen angewendet, um Qualitätsprobleme der Rohdaten (Kalibrierung, Odometrie usw.), die möglicherweise zu Verzögerungen und kostspieligen Nacharbeiten führen können, frühzeitig zu erkennen.
- **Prototypenentwicklung:** Für jede neue Version der Annotationsspezifikationen oder Sensorkonfiguration führte das Team mehrere iterative Schleifen aus, um die Einhaltung der Spezifikation anhand von Beispieldatensätzen überprüfen zu können. Erst im Anschluss wurde mit der Datenlieferung im Produktionsmaßstab begonnen.
- **Automatisierte QA-Validatoren für Datenannotationen:** Automatisierte Softwarevalidatoren wurden als Echtzeitunterstützung zur Fehlerreduzierung während der Annotation angewendet, um den Bedarf an umfangreichen Prüfungen zu minimieren und die Zeit bis zur Fertigstellung zu verkürzen.

- **Qualitätssicherung vor der Lieferung:** Vor der Datenlieferung wurden letzte Kontrollen durchgeführt, um eine bewährte Qualität gewährleisten zu können.

- **Qualitätsprüfung durch DB:** Im letzten Schritt wurden die Daten stichprobenartig von der DB InfraGO AG geprüft.

Vorstellung des Datensatzes

Der erstellte Datensatz umfasst insgesamt 88,2 min (5292 s) annotierte Sensordaten. Davon sind 1981 Sekunden und 69 Sequenzen realer Sensordaten vom GAF-Fahrzeug und 3311 Sekunden und 194 Sequenzen realer Sensordaten von der S-Bahn (BR472). Der Datensatz beinhaltet insgesamt 7.052.055 Annotationen. Die detaillierte Aufschlüsselung der Annotationen pro Objektklasse wird in der Tabelle 1 dargestellt.

Eine Nutzung des Datensatzes für die verschiedensten Anwendungsfälle, in Bereichen wie z. B. Infrastruktur- und Umfeldüberwachung, ist möglich. Der Datensatz kann bei der DB InfraGO AG angefragt werden.

Zusammenfassung

Der von den Projektpartnern aufgestellte Multisensordatensatz umfasst 88,2 min (5.292 s) annotierte Sensordaten, die im Rahmen von Kooperations- bzw. Pilotprojekten aufgezeichnet wurden.

Die Annotationen wurden von understand.ai durchgeführt. Basierend auf der Zusammenarbeit zwischen DB InfraGO und UAI wurde ein effizienter Prozess aufgesetzt, um große Mengen hochwertiger Daten liefern zu können. Gleichzeitig wurde die Flexibilität bewahrt, Datenformate und Spezifikationen an die

Entwicklung von zukünftigen Sensorkonfigurationen, Projektbedürfnissen und -anforderungen anpassen zu können.

Es ist möglich den Datensatz für verschiedene Anwendungsfälle zu verwenden. Die DB InfraGO hat damit wichtige Pionierarbeiten geleistet und einen deutlichen Impuls in den Sektor gesendet. Der Datensatz kann Interessierten und Partnern zur Verfügung gestellt werden und die Basis für eine zukünftige Weiterentwicklung durch die Industrie bilden. ■

Referenzen

- [1] Digitale Schiene Deutschland: <https://digitale-schiene-deutschland.de/de>, 2025.
- [2] UnderstandAI: <https://understand.ai/>, 2025.
- [3] Rustam Tagiew, Pavel Klasek, Roman Tilly, Martin Köppel, Patrick Denzler, Philipp Neumaier, Tobias Klockau, Martin Boekhoff, Karsten Schwalbe, OSDaR23: Open sensor data for rail 2023, 2023 8th International Conference on Robotics and Automation Engineering (ICRAE), 2023.
- [4] Hauswald, C.; Skibinski, S.; Reiniger, F.; Euler, T.; Isaac, G.; Irvathraya, A.: Sensors4Rail: Ein Erfolgsprojekt ist zu Ende, El 09/2023 (https://digitale-schiene-deutschland.de/Downloads/55_61_Hauswald_et.al.pdf) und <https://digitale-schiene-deutschland.de/de/projekte/Sensors4Rail#tab-Projektbeschreibung>

Objektklasse	GAF-Fahrzeug	S-Bahn (BR472)
Person	984818	467063
Personal Item	70467	16551
Crowd	13188	4044
Track	738833	522621
Ignore Track	7963	20767
Switch	79811	112093
Train	98057	102878
Train Front	26842	32674
Wagons	39531	24613
Buffer Stop	15166	25252
Signal Pole	326273	312362
Signal Bridge	23862	10885
Signal	148189	447573
Catenary Pole	1494205	516902
Group of Bicycle	5745	1253
Motorcycle	2620	-
Road Vehicle	175788	127544
Bicycle	22622	11576
Pram	4929	4826
Wheelchair	225	-
Animal	7473	3971

Aufschlüsselung der Annotationen pro Objektklasse

Quelle: DB InfraGO AG/
understand.ai

— Anzeige —

Ideal für innovationsinteressierte Führungskräfte und Wissenschaftler:innen

Scientific Railway Signalling Symposium

**Einfach Fahren! –
Digitale Transformation im
Spannungsfeld Automatisierung,
europäische Standardisierung
und schneller Rollout**

24. September 2025

Weitere Informationen unter
www.verkehr.tu-darmstadt.de/symposien



Nur 100 Teilnahmeplätze



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Institut für
Bahnsysteme
und Bahntechnik



TECHNISCHE
UNIVERSITÄT
BERLIN



Bahnbetrieb & Infrastruktur

Bild: TU Darmstadt / Felipe Fernandes

Innovations for Rail
Darmstädter Symposien zum Bahnverkehr



BfV
BAHN
FACHVERLAG



NEXTRAIL

