Business Paper

Automatisierte Kartierung von LST-Elementen zur Bestandsdigitalisierung mittels georeferenzierter Zeilenkamerabilder

Jannes Mennenga[1], Narbota Amanova[1], Paul Oskar Böttger[1], Philipp Britten[1], Christian Foltin[1], Rasmus Kawan[1], Eric Krumbiegel[1], Mario Neumann[1], Julia Richter[1], Tom Rößiger[1], Mark Spitzner[1]

[1] SIGNON Deutschland GmbH, ein Unternehmen der Deutschen Bahn AG

1 Problemstellung

Wie lässt sich im zehntausende Gleiskilometer umfassenden Netz der Deutschen Bahn sicherstellen, dass Veränderungen an technischen Anlagen zeitnah erkannt und korrekt in den relevanten Datenbanken abgebildet werden? Wie können Mitarbeitende unterstützt und ihre begrenzte Zeit möglichst effizient eingesetzt werden? Wie können verlässliche Planungsgrundlagen für Baumaßnahmen schnell geschaffen werden, um Reaktionszeiten in Sanierungsfällen zu verkürzen?

Eine Vorbedingung vieler Antworten auf diese Fragen liegt in einer umfassenden Bestandsdigitalisierung, also dem Schaffen einer aktuellen und vollständigen digitalen Repräsentation der Infrastruktur. Dies ist eine Grundvoraussetzung für die Unterstützung einer Vielzahl von Maßnahmen durch digitale Werkzeuge. Von der Planung von Bauprojekten im Bestand, der Baufortschrittsdokumentation, Abnahmeprüfungen oder der Aktualisierung von Datenbanken für das Instandhaltungsmanagement können viele Prozesse von einer fundierten Datengrundlage profitieren. Viele dieser Vorteile schwinden jedoch, wenn die Erfassung der Daten überwiegend manuell erfolgt. Gleisbegehungen, auch virtuell mittels Sensordaten, sind langsam, teuer und zudem fehleranfällig. Gerade in zeitkritischen oder flächendeckenden Anwendungen sind sie somit zunehmend unpraktikabel, insbesondere vor dem Hintergrund der demographischen Entwicklung, die auch vor dem Bahnsektor nicht Halt macht.

2 Ziel

Im Rahmen des Innovationsprojekts „InfraAI – Virtual Track Inspection“ (VTI) arbeitet die SIGNON Deutschland GmbH an einem System, das in Daten von Kamerabefahrungen relevante Objekte im Gleis lokalisiert. Die relevanten Objekte umfassen sowohl LST-, als auch weitere Oberbau-Elemente. Die Objekte werden mitsamt ihren relevanten Eigenschaften erfasst und kategorisiert. In Abbildung 1 ist die Lokalisierung der Elemente in den Bilddaten beispielhaft dargestellt. Die Bilddaten sind auf wenige Zentimeter genau verortet, so dass Aufnahmen von Elementen aus wiederkehrenden Befahrungen einander zugeordnet werden können.

Ziel ist ein automatischer Abgleich der Erkennungen mit relevanten Datenbanken und die Auflistung erkannter Elemente als Ausgangspunkt für weitere Arbeitsschritte.

[1] Korrespondierender Autor: Jannes Mennenga, jannes.mennenga@deutschebahn.com

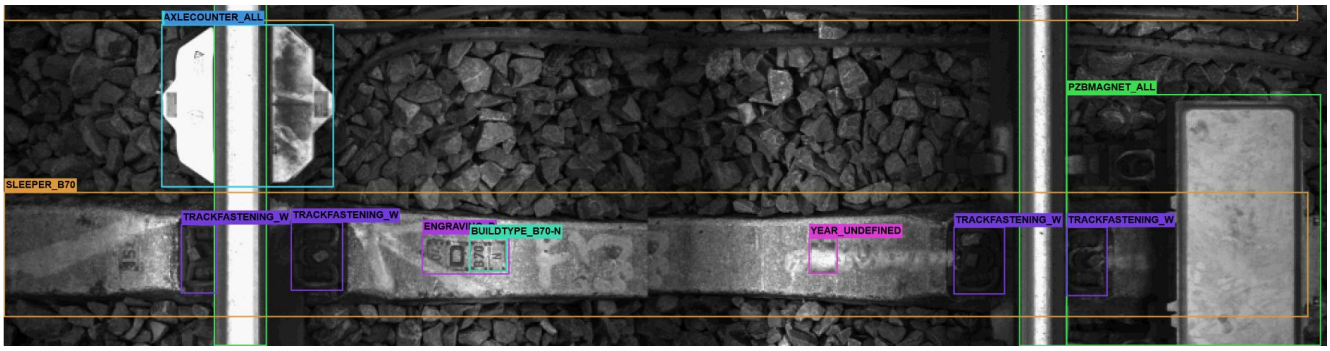


Abb. 1 Aufnahme aus Messfahrt mit hervorgehobenen Objekten

3 Methodisches Vorgehen

3.1 Allgemeiner Ablauf

Die virtuelle Oberbauinspektion beginnt mit der Planung und Durchführung der Messfahrten. Die erhobenen Daten werden nach Abschluss der Messfahrt vom Messzug in ein lokales Datenspeichersystem übertragen und im Folgenden attribuiert, hierbei findet eine Korrektur der GNSS-Daten und eine Verknüpfung mit DB AVANI Gleiskanten statt. Anschließend werden die Bilddaten zur Verarbeitung ins DataNet der DB Systel hochgeladen, bei Daten aus einem Mehrkamerasystem werden die beiden Halbvideos zu einem synchronen Einzelvideo fusioniert. Die so vorprozessierten Daten werden durch eine Datenpipeline aus verschiedenen Machine-Learning-Modellen verarbeitet, um relevante Objekte zu erkennen. Die Ergebnisse werden durch einen Qualitätssicherungsprozess geprüft und über ein Web-Frontend DB-intern zur Verfügung gestellt. Über die dem Frontend zugrundeliegende Datenbank können auch Sonderauswertungen für Spezialanwendungen realisiert werden.

3.2 Messaufbau

Das Messystem besteht aus zwei weggesteuerten Zeilenkameras, die monochrome Bilder bei einer kombinierten Auflösung von 3960 Pixeln in der Breite aufzeichnen. Somit ergibt sich eine räumliche Auflösung von 0,7 mm pro Pixel, was sehr genaue visuelle Inspektionen ermöglicht. Die Kameras sind mit einem Satellitennavigationssystem, einem Odometer und einem Beschleunigungsmesser synchronisiert, über die die Bilddaten georeferenziert werden. Die beiden Kameras sind rechts und links der Gleisachse positioniert, so dass die meisten relevanten Objekte im Fokus befinden. In der Schwellenmitte kann sich ein perspektivischer Versatz ergeben.

3.3 VTI-Pipeline

Die VTI-Pipeline ist eine modulare Softwareanwendung zur automatisierten Erkennung und Klassifikation von Oberbau- und LST-Elementen, deren Merkmalen (z. B. Hersteller, Bauform, Baujahr) sowie möglichen Defekten. Die Pipeline wurde im Kontext des InfraAI-Projekts entwickelt und dient der Unterstützung der Fahrbahn-Instandhaltung der DB InfraGO AG.

Die Pipeline verarbeitet Videos blockweise. Das bedeutet, ein Eingabevideo wird in aufeinanderfolgende Blöcke (Chunks) von konfigurierbarer Größe unterteilt. Jeder Chunk durchläuft sequenziell die folgenden Prozessschritte:

Zunächst werden Einzelbilder aus dem Eingabevideo extrahiert und in Chunks unterteilt. Dabei werden Meta-Daten einbezogen, um irrelevante Abschnitte (z. B. Stillstand oder Rückwärtsfahrt)

auszuschließen. Anschließend wird eine Objektdetektion basierend auf einem neuronalen Netz durchgeführt. Dieses Modell ist auf vollständig gelabelten Orthobildern des Gleisbetts trainiert und in der Lage, neben Schwellen und deren Eigenschaften (z. B. Gravuren und Kleineisen) auch LST-Objekte (aktuell PZB-Magnete, Isolierstöße, Balisen, Achszähler) zu erkennen. Die Detektionen werden mittels konfigurierbarer Filter von unplausiblen Erkennungen bereinigt.

Das positionsbasierte Tracking korreliert Detektionen über mehrere Frames hinweg, um mehrfach gesehene Objekte zu einer Instanz zusammenzuführen. Anschließend werden die erkannten Bereiche wie Gravuren oder LST-Objekte als Bildausschnitte aus den Originalbildern extrahiert und der jeweiligen Feature-Kategorie zugeordnet.

Je nach Merkmalskategorie kommen anschließend verschiedene Verfahren zur genaueren Charakterisierung zur Anwendung. Bildausschnitte, die durch den Objektdetektor einer Schwellen-Gravur zugeordnet wurden, werden in dieser Folgestufe der Pipeline durch einen gravurartspezifischen Klassifikator hinsichtlich der genauen Gravur klassifiziert. Beispielsweise werden Herstellergravuren durch ein individuell für diesen Gravurtyp trainierten Klassifikator-Modell dem konkreten Hersteller zugeordnet. Erkannte Schienenausschnitte wiederum werden durch klassische Verfahren der Computer Vision analysiert, um die Fahrspiegelbreite abzuleiten. Bauteile wie Kleineisen werden mittels Anomalieerkennung auf strukturelle Schäden (z. B. abgebrochene Spannklemmen, fehlende Dübel oder Ähnliches) untersucht. Zusätzlich wird ein Segmentierungsnetz zur Erkennung von Rissen oder Abplatzungen im Beton der Schwellen auf den Bildausschnitt der Schwelle angewandt.

Im Anschluss an diese initiale Erkennungsphase wird eine Nachverarbeitung (Postprocessing) durchgeführt, um die Ergebnisse zu verfeinern und zu plausibilisieren. Zunächst werden konfigurierbare Geschäftsregeln auf die Objekte angewandt. Regeln können Merkmalswerte überschreiben, Labels setzen oder bestimmte Erkennungen löschen, sollten diese einer Plausibilisierungsregel widersprechen. Für Schwellen mit undefinierten Merkmalen werden mit Hilfe eines eigens dafür trainierten neuronalen Netzes die ähnlichste Nachbarschwelle gesucht, auf denen das gesuchte Merkmal erkannt wurde, um dieses auf die Schwelle zu übertragen. Da Schwellen häufig chargenweise verbaut werden, können so fehlende Merkmale ausgefüllt werden, um bspw. Bedeckungen mit Schotter oder Verschmutzungen zu kompensieren.

Nach Verarbeitung aller Chunks werden die Ergebnisse zusammengeführt. Abschließend werden verschiedene Visualisierungen der Ergebnisse generiert. Diese dienen der Qualitätssicherung durch Stichproben, der Plausibilisierung der Anzahl gefundener Elemente pro Streckenabschnitt sowie der Erschließung neuer Klassen und Labels.

3.4 Datensatz

Ein Kernbestandteil des Systems ist der Datensatz, aus dem die Machine-Learning Komponenten trainiert werden und anhand dessen die Funktionsfähigkeit des Gesamtsystems geprüft wird. Der gesamte Datensatz besteht aus rund 44.000 Bildern mit insgesamt 451.724 Labels. Die Labels bestehen aus kategorisierten Umgrenzungsboxen (Boundingboxen), die in einem XML-Format gespeichert sind. Im Datensatz sind die 255 unterschiedlichen Klassen definiert, die das System erkennen soll.

Der Datensatz umfasst für die Schwellenerkennung 33 Schwellenbauformen, 33 Herstellergravuren, 44 verschiedene Bauformgravuren und Baujahrgravuren von 1955 bis 2025. Für die Erkennung von LST-Elementen wurde der Datensatz um Achszähler, Balisen, Isolierstöße und PZB-Magneten erweitert. Für die Zustandsbewertung der Kleineisen wurden 4 Klassen angelernt, die Oberbauten K, KS und W, sowie

eine Sammelklasse für alle weiteren, für die Defekterkennung an Schienen wurden Schienenköpfe binär gelabelt.

Da der Datensatz verschiedene Aufgaben hat, sind die Daten abschnittsweise unterschiedlich strukturiert. Um das Objekterkennungssystem zu trainieren, sind vollständig gelabelte Bilder notwendig, das heißt, alle im Bild sichtbaren Objekte sind mit einem Label versehen. Für die Testdaten kommen zusätzlich Informationen über den Schwierigkeitsgrad der Erkennung und ihre Ableitbarkeit aus der Nachbarschaft hinzu.

Da das vollständige Labeln von Bildern trotz verschiedener technischer Unterstützungsmechanismen sehr aufwändig ist, können Boundingboxen auch einzeln gelabelt werden. Die so nur teilweise gelabelten Bilder werden mit einem Attribut markiert, und nur für das Training der Bildausschnitte verarbeitenden Komponenten verwendet, da diese den Gesamtkontext des Bildes nicht benötigen. So können Bildausschnitte als Labels wiederverwendet werden, die ohnehin von Menschen bewertet werden müssen, beispielsweise in QS-Prozessschritten. Insgesamt sind knapp 10.000 Labels mit einem solchen Attribut versehen.

Einen weiteren Sonderfall stellt der Teildatensatz für die Risserkennung in Betonschwellen dar. Dieser besteht aus über 2.000 Rissinstanzen, bei denen jeder einem Riss zugeordnete Pixel händisch im Bildmaterial maskiert wurde. Diese Masken werden für Training und Auswertung des Segmentierungsmodells verwendet.

3.4.1 Verteilung der Daten im Datensatz

Der Datensatz ist über mehrere Jahre entstanden und spiegelt so auch die Entwicklung der Aufnahmetechnik hin zum heutigen System wider. Ideal für die Erkennung von Infrastruktur mit dem aktuellen System wäre ein rein aus neuen Bilddaten bestehender Datensatz. Bei einigen seltenen Klassen ist dies nicht möglich, da sie nur mit neuen Labels ersetzt werden können, wenn sie in einer Befahrung auch mit dem neuen System erfasst wurden. Dies erfordert einen kontinuierlichen Aufwand zur Datensatzinstandhaltung.

Die Seltenheit bestimmter Klassen führt auch zu einer ungleichmäßigen Verteilung im Datensatz. Beispielsweise ist die Markierung des Herstellers C (Schwellenwerk Coswig) mit fast 6.000 Labels sehr häufig vertreten, während zwei Drittel der Hersteller mit weniger als 1.000 Labels vertreten sind. Die Ausbalancierung des Datensatzes ist, insbesondere bei den Schwellen und ihren Eigenschaften nicht möglich, da diese in Kombination miteinander vorkommen, beispielsweise fügt die Erweiterung um neue Jahreszahlen zwangsläufig neue Hersteller- und Bauform-Labels hinzu. Ein unbalancierter Datensatz erschwert das zu lösende Problem der ML-Komponenten und erfordert eine Robustheit gegen Ausreißer.

3.5 Halbautomatische Erweiterung von Trainingsdaten für die Objekterkennung

Da der Fokus primär auf der Erkennung von Schwellen lag, enthielt der Datensatz keine Annotationen für etwaige im Bild vorhandene LST-Elemente wie Isolierstöße, Achszähler und PZB-Magnete. Eine vollständige manuelle Nachannotation des gesamten Datenbestands wäre jedoch mit erheblichem Aufwand verbunden gewesen. Daher wurden zwei halbautomatische Ansätze zur Identifikation und Annotation der neuen Objekte entwickelt. Die Ansätze werden im Folgenden für Isolierstöße und Achszähler exemplarisch beschrieben.

Zunächst wurde auf Basis domänenspezifischen Wissens eine Vorauswahl geeigneter Bilddaten getroffen. Mithilfe dieses Musters konnten potenzielle Kandidaten in Datenbankeinträgen vorhergehender Befahrungen ermittelt werden. Die korrespondierenden Bilder wurden anschließend

manuell geprüft und halbautomatisch annotiert. Hierzu wurde die bestehende Pipeline auf die Bilder angewendet und die Ergebnisse in ein Annotationstool übertragen, so dass neben den neuen Klassen lediglich fehlerhafte oder fehlende Boundingboxen manuell bearbeitet werden mussten.

Zur Erweiterung des Trainingsdatensatzes wurden in einer weiteren Iteration zusätzliche Bilder automatisch mithilfe eines spezialisierten Modells ausgewertet und so weitere potenzielle Fundstellen identifiziert. Diese wurden erneut manuell überprüft und annotiert. Um den manuellen Aufwand zu reduzieren, wurde im Anschluss ein eigenes Verfahren entwickelt um den bereits existierenden Bestandsdatensatz von rund 40.000 Bildern halbautomatisch um neue Klassen zu erweitern. Dabei werden potenziell neue Labels für eine Klasse mit einer möglichen Boundingbox versehen und zur manuellen Verifizierung vorgeschlagen.

Dieser Ansatz ermöglicht eine deutliche Reduktion des Aufwands bei der Erweiterung und Ausdifferenzierung des Bestandsdatensatzes. Gleichzeitig kann keine vollständige Abdeckung aller im Datensatz enthaltenen Isolierstöße und Achszähler gewährleistet werden. Ungeachtet dieser Einschränkung konnten die Klassen Isolierstoß und Achszähler erfolgreich in das bestehende Erkennungssystem integriert werden, ohne die Erkennbarkeit der bestehenden Klassen zu beeinträchtigen.

3.6 Ende-zu-Ende Auswertung

Damit die Software produktiv genutzt werden kann, muss die entsprechend notwendige Qualität sichergestellt werden. Dafür wird der Testdatensatz verwendet.

Schon während der Entwicklung wird die Ende-zu-Ende Auswertung regelmäßig durchgeführt. Dadurch können Auswirkungen durch neu trainierte Modelle und Änderungen an der Datenverarbeitung direkt ausgewertet werden. Außerdem wird vermieden, dass Änderungen unbeabsichtigten Einfluss auf die Ergebnisse haben.

Für die Auswertung wird jedem Objekt im Testdatensatz die passende Erkennung aus der Software zugeordnet. Dafür werden die Boundingboxen der Objekte mit den Boundingboxen der Testobjekte im selben Bild mittels Jaccard-Koeffizient (IoU) miteinander verglichen. Wenn die IoU einen bestimmten Schwellwert überschreitet, dann gilt das Testobjekt als gefunden. Im Folgeschritt werden die Merkmale der Objekte miteinander verglichen. Bei Schwellen sind das: Hersteller, Baujahr, Bauform und mögliche Risse. Bei den Merkmalen wird normalerweise nur der Wert verglichen, da das zugehörige Objekt einem Testobjekt bereits zugeordnet wurde. Gibt es jedoch mehrere Instanzen eines Merkmals, wie bei den Rissen, dann findet erneut eine Zuordnung über die IoU statt. Das dient der Ermittlung der Anzahl an richtig und falsch erkannten Objekten und Merkmalen. Aus diesen Zahlen werden Recall (Trefferquote), Precision (Präzision) und verschieden gewichtete F-Scores berechnet. Hauptsächlich relevant für die Freigabe zum produktiven Einsatz ist der Recall. Dieser wurde als Entscheidungsgrundlage gewählt, da es deutlich wichtiger ist alle Objekte zu finden, als keine Falscherkennungen zu erzeugen. Das liegt daran, dass es leichter ist, falsch erkannte Objekte zu entfernen, als nicht gefundene Objekte zu suchen, beziehungsweise festzustellen, ob Objekte nicht gefunden wurden.

Neben dem Boundingbox-Vergleich wird auch die Ableitung über die Nachbarschaft ausgewertet. Dafür wird geprüft, ob Merkmale eines Objekts, die über ein Nachbarobjekt abgeleitet wurden, auch im Testdatensatz als „muss über die Nachbarschaft abgeleitet werden“ markiert wurden.

Die Ergebnisse werden in Form einer Konfusionsmatrix dargestellt. Für jeden Typ gibt es eine eigene Matrix.

Tab. 1 Ausschnitt aus Hersteller-Konfusionsmatrix

Erkennung \ Datensatz	LMB	LMH	Präzision
LMB	84	0	100%
LMH	2	96	77%
Trefferquote	98%	100%	-
$F_{0,5}$-Score	100%	81%	-
F_1-Score	99%	87%	-
F_2-Score	98%	94%	-

Die Tabelle 1 ist eine stark reduzierte Version der Konfusionsmatrix. Die Metriken wurden anhand der Zahlen aller Klassen berechnet. Die Zahlen der anderen Klassen wurden für die Abbildung ausgeblendet.

Sollte der Fall eintreten, dass der Recall über alle Klassen gemittelt unter den Schwellwert liegt, wird die Pipeline nach einer Ursachenanalyse iterativ optimiert, bis die Metriken wieder im Zielbereich liegen.

3.7 Qualitätssicherungsprozess

Eine hohe Qualität der Daten sicherzustellen ist wichtig, damit das System für Endanwender:innen verlässliche Ergebnisse liefert. Aus diesem Grund ist der Qualitätssicherungsprozess über alle wichtigen Schritte bis zum Ergebnis allgegenwärtig. Von der Vermessung bis zur Auslieferung der Daten werden alle Prozessschritte geprüft.

Bevor unsere zuvor angesprochene Pipeline gestartet wird, gibt es bereits eine Prüfung der Quelldateien, um etwaige Probleme der Attribuierung aufzudecken.

Es gibt Schwellwerte bei denen eine Prüfung als Warnung oder Fehler gilt. Warnungen werden begutachtet, können aber ggf. toleriert werden, bei Fehlern wird die Annahme des Videos jedoch verweigert.

Geprüft werden Formate innerhalb der Datei aber auch fachliche Themen wie Geolokalisation und gefahrene Kilometer.

Wir unterstützen derzeit für die Messfahrten Ein –und Zweikamera-Setups. Bei letzterem werden linkes und rechtes Video gestitcht (engl. *zusammengenäht*). Wir erzeugen Stichproben, um zu kontrollieren, ob die Kameraaufnahmen synchron bleiben. Bei starken Abweichungen können werden Abschnitte als nicht-relevant markiert. Somit spielen diese bei weiterer Auswertung keine Rolle mehr. Bevor die Inferenz der Messfahrten gestartet wird, gibt es noch kleinere Prüfschritte. Diese behandeln vor allem die Frage: Sind alle notwendigen Daten vollständig vorhanden? Zusätzlich wird der Stand des Quellcodes mit Hilfe eines Git-Tags eingefroren, um eine Reproduzierbarkeit der Ergebnisse herzustellen. Die aktuellen Erkennungsraten für den Stand werden dokumentiert.

Während der Inferenz wird kontinuierlich protokolliert, um den Bearbeitungsstand aller Videos überwachen zu können.

Nach Durchlaufen der Pipeline werden die Ergebnisse gebündelt und eine Gesamtlieferung vorbereitet. Diese enthält umfangreiche Statistiken und weitere Hilfsmittel zur Qualitätskontrolle. Zum einen

erlaubt es notwendige Schnell-Checks. Es gibt eine hochrechenbare erwartete Schwellenzahl der Befahrung. Diese muss mit einer definierten Toleranz erreicht werden. Falls die Abweichung zu hoch ist, muss geprüft werden, wie diese zustande kommt. Eine weitere Prüfung besteht in der Zuordnung der gefundenen Schwellen zu technischen Plätzen in SAP mit ähnlichen Kriterien.

Aus jedem inferierten Video werden manuell zu prüfende Stichproben gezogen. In den Stichproben werden Fehlerkennungen vermerkt und abschließend gesammelt ausgewertet. Die Ergebnisse der Reviews bestimmen unsere finalen Erkennungsraten für die Lieferung anhand ungesehener Daten. Zudem können sie als Quelle für neue Labels verwendet werden. Der letzte große Abschnitt der Qualitätssicherung betrifft den Upload der Inferenzergebnisse auf eine Datenbank und die Bereitstellung im Frontend. Das Frontend ist in Test –und Produktivumgebung unterteilt. Nach dem Datenbankupload werden weiter grobe Plausibilitätsprüfungen durchgeführt, da hier Daten aus wiederkehrenden Befahrungen überschrieben werden. Sobald der Stand auf der Datenbank verifiziert ist, wird das Frontend aktualisiert. Hier gibt es noch einmal Stichprobenprüfungen. Wenn die Abnahme am Testfrontend erfolgreich war, wird das Produktivfrontend aktualisiert und die Lieferung gilt als abgeschlossen.

4 Ergebnisse

Nachdem in den vorherigen Abschnitten die Motivation für die Umsetzung des Projekts sowie die abstrakten Eigenschaften der implementierten Software, der zugehörigen Prozesse und Daten beschrieben wurden, zeigt der folgende Abschnitt Details zur Anwendung der Software an realen Streckenvideos in nennenswertem Umfang. Zunächst werden die Eigenschaften der Strecke beschrieben, anschließend werden die Ergebnisse zur Erkennung von Schwellen und LST-Elementen dargestellt.

4.1 Experimentelles Setup

Zur Evaluation der Software wurde im Dezember 2025 eine Route in der Netzregion Nord befahren: von Minden über Seelze nach Hannover Hbf und südlich um Hannover herum und über Seelze wieder zurück nach Minden. Von der Gesamtlänge von 143,45 km wurden 138,27 km mit der Software ausgewertet und anschließend gemäß dem beschriebenen Qualitätssicherungsprozess geprüft. Ursachen für die Differenz zwischen befahrenen und ausgewerteten Kilometern sind Doppelbefahrungen und Qualitätsmängel des Videomaterials.

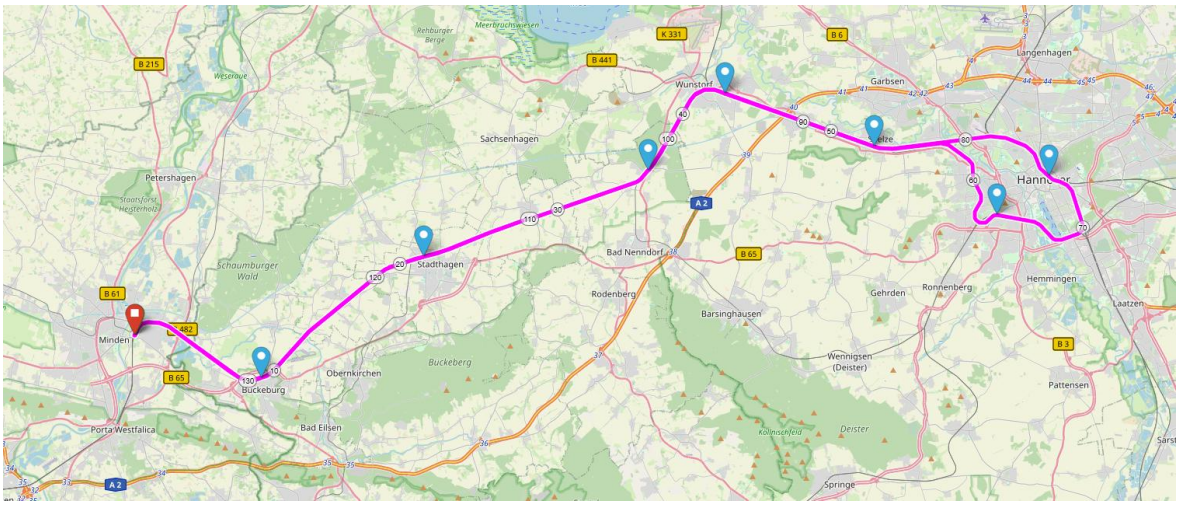


Abb. 2 Visualisierung der befahrenen Strecke von Hannover Hbf. über Seelze und Minden zurück nach Hannover Hbf.

In diesem Abschnitt zeigen wir den Positive Predictive Value (PPV) oder Precision für die durchgeführten Erkennungen auf der befahrenen Strecke. PPV ist der Anteil der tatsächlich vorhandenen Objekte, die korrekt erkannt wurden:$PPV = \frac{TP}{TP+FP}$.

Dabei sind True-Positives (TP) korrekt erkannte Objekte und False-Negatives (FN) tatsächlich vorhandene, aber nicht erkannte Objekte.

Die Daten wurden durch manuelle Stichprobenauswertung erhoben. Die Angabe aller Ergebnisse erfolgt gerundet auf Ganzzahlen.

4.2 Schwellen- und Oberbauerkennung

Die befahrene Strecke wurde mit InfraAI ausgewertet und anschließend dem beschriebenen Qualitätssicherungsprozess unterzogen. Insgesamt wurden 240.589 Schwellen erkannt. Bei lediglich 11 % konnten keine Herstellergravuren, bei 12 % keine Baujahrsangaben und bei 4 % keine Bauformangaben erkannt werden. Hierbei ist jedoch zu beachten, dass z.B. Weichenschwellen nicht immer und Holzschwellen nie entsprechende Gravuren haben.

Für jede Gravurkategorie wurde der PPV auf Basis der Stichproben bestimmt. Für Herstellergravuren ergab sich ein PPV von 92 %, für Jahreszahlen 82 % und für die Bauformgravurerkennung 85 %.

Für den Anwendungsfall der digitalen Erfassung der Bahninfrastruktur ist zudem relevant, welcher Anteil der Schwellen bestehenden technischen Plätzen (SAP) zugeordnet werden kann. Von den erkannten Schwellen konnten 226 025 (94 %) technischen Plätzen zugeordnet werden. Dies ist unter dem Gesichtspunkt von in realem Betrieb befindlichen Stammdaten ein sehr guter Wert.

Die Kleineisen-Erkennungen werden nach Oberbauart unterteilt; zusätzlich werden Defekte als eigene Klasse modelliert. Die manuelle Bewertung der Defekte ergab eine geringe PPV von < 1 %. Eine detaillierte Auswertung der Oberbauklassen fand im Rahmen dieser Befahrung nicht statt, da der Projektfokus auf den Defekten lag.

Für die Risserkennung ergab die manuelle Stichprobenprüfung eine moderate Präzision: Von insgesamt 405 durch die Pipeline gemeldeten Rissfällen wurden 176 als korrekt bestätigt, während 229 Meldungen Falscherkennungen waren. Das führt zu eine PPV von 43%. Der Anteil an False Positives (57 %) führt zu zusätzlichem Aufwand im bei den Anwendern und unterstreicht den Bedarf an Verbesserungen

insbesondere in der Segmentierung und den Vorfilterstufen. Auf der anderen Seite ist das korrekte Detektieren und Verorten von Rissen auf Betonschwellen von 43% der untersuchten Risse eine deutliche Verbesserung gegenüber bisherigen Ansätzen.

4.3 Erkennung von LST-Elementen

Insgesamt wurden 448 PZB-Magnete, 293 Achszähler und 885 Isolierstöße erkannt. Da die Strecke nicht ETCS-ausgerüstet ist, entfällt die Erkennung von Balisen.

Die manuelle Stichprobenprüfung ergab eine PPV von 97 % für PZB-Magnete; für Achszähler und Isolierstöße lagen die Werte bei 74 % bzw. 80 %. Auffällig ist, dass Achszähler häufig mit Schienenstegdämpfern verwechselt werden. Isolierstöße werden bei Falscherkennungen insbesondere mit Schweißstößen, Weichenverschlüssen oder anderen an der Schiene verschraubten Objekten (z. B. PZB-Montageelementen) verwechselt.

Auf der Evaluationsstrecke wurde ebenfalls eine Erkennung des Fahrspiegels durchgeführt. Da diese jedoch aktuell lediglich eine deskriptive Charakterisierung liefert, ist es nicht möglich einen quantitativen Korrektheitsgrad abzuleiten.

5 Diskussion

5.1 Einordnung der Ergebnisse

Die Ergebnisse belegen, dass die Pipeline für die Schwellen- und Oberbauerkennung bereits eine hohe Erkennungsleistung erreicht. Insbesondere die hohen PPV-Werte für Herstellergravuren (92 %), Herstelljahresgravuren (82 %) und Bauformgravuren (85 %) sowie die Zuordnung von 94 % der erkannten Schwellen zu bestehenden technischen Plätzen unterstreichen den praktischen Nutzen für die Bestandsdigitalisierung und nachgelagerte Prozesse. Hervorgehoben werden muss auch die automatische Erkennung von Rissen mit einem PPV von 43%. Dies stellt einen entscheidenden Schritt in der Entwicklung solcher Systeme dar.

Für LST-Elemente wurden in der Stichprobe sehr hohe Werte für PZB-Magnete (97 %) und solide Werte für Achszähler (74 %) und Isolierstöße (80 %) erzielt, wobei typische Verwechslungsmuster identifiziert wurden. Die Detektion von Kleineisen-Defekten weist mit < 1 % PPV dagegen noch deutlichen Entwicklungsbedarf auf.

Die Befahrung erfolgte auf einem spezifischen Streckenabschnitt und in einem begrenzten Zeitraum. Übertragbarkeit auf andere Regionen, Jahreszeiten, Beleuchtungssituationen und Bauzustände ist wahrscheinlich, aber nicht belegt. Bisherige weitere Befahrungen in denen lediglich Schwellen untersucht wurden, zeigen allerdings konsistente Ergebnisse, können allerdings keine qualitätsgesicherten Erkenntnisse für die Erkennbarkeit von LST-Elementen liefern.

Die hohe Abdeckung bei Schwellen und die Zuordnung zu technischen Plätzen zeigen einen direkten Mehrwert für die Bestandsdigitalisierung, Qualitätssicherung und Plausibilisierung von Streckenabschnitten. Der recall-zentrierte Freigabeansatz ist für die Praxis sinnvoll, da nicht entdeckte Objekte in kritischen Prozessen gravierend sein können.

5.2 Fehlerbilder und Ursachenanalyse

Im aktuellen Stand treten vier charakteristische Fehlerbilder hervor. Bei Achszählern kommt es häufig zu Verwechslungen mit Schienenstegdämpfern. Beide Bauteile weisen eine hohe visuelle Ähnlichkeit auf und liegen unmittelbar am Schienensteg, sodass form- und texturbezogene Hinweise allein nicht ausreichen.

Eine weitere Fehlquelle betrifft Isolierstöße, die insbesondere mit Schweißstößen, Weichenverschlüssen oder verschraubten Anbauteilen verwechselt werden. Dieses Verhalten der Software kann durch eine hohe Ähnlichkeit in einzelnen Bildausschnitten zwischen den einzelnen Strukturelementen erklärt werden.

Schließlich zeigt die Defekterkennung an Kleineisen noch deutliche Leistungsgrenzen. Hauptursache ist ein ausgeprägtes Klassenungleichgewicht. Aus etwa 160.000 Kleineisenbeispielen stehen lediglich rund 60 Defektinstanzen zur Verfügung.

Bei der Untersuchung der Risserkennung fällt auf, dass besonders kleine und feine Risse nicht gut segmentiert werden. Dabei handelt es sich um ein bekanntes Problem von Machine-Learning-Ansätzen zur Zustandserkennung, das durch die verwendeten Architekturen und den Einfluss einzelner Pixel in hohen Auflösungen begründet ist.

Diese Fehlerbilder begrenzen derzeit die Detektionsleistung und machen gezielte Datensatzerweiterungen, eine Ansatzverbesserung sowie ein methodisches Feintuning erforderlich.

5.3 Verbesserungspotentiale

Im aktuellen Stand der Software zeigt sich eine deutliche Datenabhängigkeit. Klassenungleichgewichte und wechselnde Aufnahmebedingungen beeinflussen sowohl die Modellstabilität als auch die Generalisierbarkeit der Resultate.

Schließlich sind die Defektklassen noch ausbaufähig. Für Schäden an Kleineisen und bei der Risserkennung besteht erkennbarer Nachbesserungsbedarf in Detektion und Segmentierung. Diese Faktoren begrenzen derzeit die Übertragbarkeit der Befunde und markieren zugleich Prioritäten für die nächste Entwicklungsphase.

6 Ausblick

Zukünftige Entwicklungen des Systems werden neben der Erweiterung des Katalogs an erkennbaren Objekten vor allem die Klassifizierung von Objektzuständen zum Gegenstand haben.

Die Risserkennung soll so weit verbessert werden, dass sie im produktiven Einsatz als zuverlässiges Frühwarnsystem genutzt werden kann.

In Zusammenarbeit mit Fachexpert:innen wurde ein erster Versuch gemacht, um Bilder von Isolierstößen Schadensklassen zuzuweisen und ihnen einen einheitlichen Zustandsbewertungsscore zuzuweisen.