

Einsatz von Deep Learning zur automatischen Detektion und Klassifikation von Fahrbahnschäden aus mobilen LiDAR-Daten

Deep Learning for Automatic Detection and Classification of Road Damage from Mobile LiDAR Data

Maximilian Sesselmann¹, Ronny Stricker², Markus Eisenbach²

¹GINGER Lehmann+Partner GmbH, Erfurt · sesselmann@lehmann-partner.de

²Neuroinformatics and Cognitive Robotics Lab, TU Ilmenau

Zusammenfassung: Im Kontext automatisierter Datenauswertung sind künstliche neuronale Faltungsnetzwerke und der Einsatz von Deep-Learning-Ansätzen mittlerweile Stand der Technik. Im Bereich der Zustandserfassung und -bewertung von Straßen wurde die Leistungsfähigkeit tiefer neuronaler Netze zur Analyse von Kamerabildern bereits demonstriert. Im vorliegenden Beitrag soll diese Methodik nun erstmals auf hochgenaue mobile LiDAR-Daten des Fraunhofer Pavement Profile Scanners in Form von 2.5D-Oberflächenmodellen übertragen werden, um eine automatische Schadensdetektion und -klassifikation auf Basis von radiometrischen und geometrischen Merkmalen zu realisieren. Damit ist eine automatisierte Erfassung von Fahrbahnschäden in Form von präzise verorteten Geoobjekten möglich.

Schlüsselwörter: Mobile LiDAR, Deep Learning, Straßenzustandserfassung

Abstract: *In the context of automated data analysis, convolutional neural networks and the use of deep learning approaches have become state of the art. In the field of road condition assessment and evaluation, the performance of deep neural networks for the analysis of camera image data has already been demonstrated. For the first time, this methodology is to be applied to high-precision mobile LiDAR data of the Fraunhofer Pavement Profile Scanner in the form of 2.5D surface models in order to realize automatic road damage detection and classification on the basis of radiometric and geometric features. Thus, an automated detection of road damage in the form of precisely located geo objects is possible.*

Keywords: *Mobile LiDAR, Deep Learning, Road Condition Assessment*

1 Motivation und Stand der Technik

Vor dem Hintergrund des fortschreitenden Substanzverzehr² bei der Straßeninfrastruktur sind Ausbau, Erhalt und Unterhalt des bestehenden Netzes wichtige Aufgaben für die jeweiligen Baulastträger (Bundesministerium für Verkehr und digitale Infrastruktur, 2016). Die Grundvoraussetzung für ein solides Erhaltungsmanagement ist eine möglichst detaillierte Kenntnis über den Straßenzustand. Bislang dominiert im Bereich der Erfassung von Substanzmerkmalen auf Fahrbahnoberflächen die manuell-visuelle Inspektion von Kamerabildern. Dieser Prozess ist nicht nur zeit-, personal- und kostenintensiv, sondern nutzt darüber hinaus ausschließlich radiometrische Informationen für den Inspektionsvorgang. So werden Oberflächenschäden im Rahmen der „ZTV ZEB-StB“ (Forschungsgesellschaft für Straßen- und Verkehrswesen, 2006) ausschließlich auf Basis von Grauwertunterschieden in Kamerabildern erkannt. Dabei sind die Erfassungssysteme, die im Rahmen der Zustandserfassung

und -bewertung von Straßen zum Einsatz kommen, in der Lage, mithilfe von hochgenauen Laserscannern, digitale Topographien von Fahrbahnoberflächen schnell, flächendeckend und präzise zu erfassen. Wie bei Sesselmann (2018) gezeigt, erlauben die mit dem Fraunhofer Pavement Profile Scanner (Reiterer et al., 2013) erfassten Oberflächendaten eine automatisierte Analyse von Substanzmerkmalen sowohl entlang radiometrischer als auch geometrischer Merkmale.

Im Kontext automatisierter Datenauswertung sind künstliche neuronale Faltungsnetzwerke und der Einsatz von Deep-Learning-Ansätzen mittlerweile Stand der Technik. In vielen Anwendungsdomänen der Objekterkennung und Bildsegmentierung zeigten sie sich konventionellen maschinellen Lernverfahren überlegen (Russakovsky et al., 2015; Liu et al., 2016; Karaca et al., 2017; Rachmadi et al., 2017). Nachdem Eisenbach et al. (2017) und Stricker et al. (2019) das Leistungsvermögen von Deep-Learning-Ansätzen auch im Bereich der Zustandserfassung und -bewertung von Fahrbahnoberflächen auf Basis von hochauflösenden Kamerabilddaten demonstriert haben, soll diese Methodik im vorliegenden Beitrag auf die Daten des Pavement Profile Scanners adaptiert werden. Ziel ist es, eine automatisierte Erkennung, Lokalisierung und Unterscheidung von Straßenschäden auf Basis von mobilen Laserscanningdaten zu realisieren und somit das Potenzial der Deep-Learning-Technologie im Feld der Geoprozessierung aufzuzeigen. Hierzu wird zunächst dargestellt, wie Laserscans in Form von unorganisierten 3D-Punktwolken in ein adäquates Format übertragen werden können, welches von den kontemporären Deep-Learning-Architekturen performant verarbeitet werden kann. Anschließend wird auf einem umfangreichen Datensatz von 2.5D-Oberflächenmodellen systematisch untersucht, mit welchem lokalen Einzugsbereich und welchen Datenkanälen der leistungsfähigste Detektor trainiert werden kann, um Schäden als georeferenzierte Einzelobjekte zu detektieren und verschiedene Schadenstypen zu klassifizieren.

2 Datengrundlage

2.1 Datenaufnahme: mobile Laserscanningsysteme

Die im Beitrag untersuchten Messdaten wurden mit den Mobile-Mapping-Systemen S.T.I.E.R und I.R.I.S der Firma LEHMANN+PARTNER GmbH aufgenommen (Abb. 1). S.T.I.E.R ist ein System zur Erfassung der Längs- und Querebeneheit, der Textur, der dreidimensionalen Oberfläche sowie des Oberflächenbilds von Straßen und Verkehrsflächen. Die Kernkomponenten von S.T.I.E.R sind ein Positionierungssystem, Laserdistanzsonden zur Messung von Ebenheit und Texturparametern, ein Laserscanner sowie verschiedene Kamerasysteme zur Erfassung der Fahrzeugumgebung und der Fahrbahnoberfläche. Das I.R.I.S-System dagegen ist für die bildhafte wie auch dreidimensionale Erfassung des Straßenraumes im Allgemeinen konfiguriert. So ist hier neben dem Laserscanner zur Erfassung der Fahrbahnoberfläche auch der Fraunhofer Clearance Profile Scanner zur Erfassung des befahrenen Korridors verbaut.

Für diesen Beitrag sind der auf die Fahrbahn ausgerichtete Laserscanner und das Positionierungssystem die relevanten Datenquellen. Bei ersterem handelt es sich um einen Pavement Profile Scanner (PPS) des Fraunhofer-Instituts für Physikalische Messtechnik (Reiterer et al., 2013). Die Messgenauigkeit des PPS liegt, gemittelt über ein $10\text{ cm} \times 10\text{ cm}$ großes Oberflächenelement, im Submillimeterbereich. Um diese extrem hohe Präzision bei großen Messge-

schwindigkeiten zu gewährleisten, nutzt der PPS als Messprinzip das Phasenvergleichsverfahren. Wichtige technische Spezifikationen sind bei Fraunhofer (2017) zu finden. Die absolute Lagegenauigkeit der Trajektorie nach Post-Processing liegt im Mittel bei ca. 0.02 m, die Höhengenaugigkeit bei etwa 0.05 m, die relative Genauigkeit aufeinanderfolgender Trajektorienpunkte ist dabei deutlich höher (Applanix, 2015). Über die zeitliche Synchronisierung der Messungen des Positionierungssystems und des Laserscanners entsteht das grundlegende Datenprodukt für die weiterführenden Analysen: eine 3D-Punktwolke, die neben der geometrischen auch eine radiometrische Information in Form der Reflexionsintensität pro Lasermesspunkt enthält.



Abb. 1:
Das Mobile-Mapping-System I.R.I.S mit dem Laserscanner PPS des Fraunhofer-Instituts für Physikalische Messtechnik, der zentral im hinteren Bereich des Trägerrahmens montiert ist

Da S.T.I.E.R für Messungen im Kontext der Zustandserfassung und -bewertung von Straßen konzipiert ist, sind die Anforderungen an die Genauigkeit der Sensoren und die Qualität der mit ihnen aufgenommenen Daten sehr hoch. Seit dem Jahr 2012 hat das Messsystem S.T.I.E.R mit dem PPS jedes Jahr die zeitbefristete Zulassung zur Erfassung von Ebenheitsparametern durch die Bundesanstalt für Straßenwesen erhalten. Eine vergleichende Studie zur Ebenheitserfassung, in der das I.R.I.S-System inklusive PPS erfolgreich getestet wurde, ist bei Lundberg et al. (2018) zu finden.

2.2 Oberflächenmodellierung aus 3D-Punktwolken

Im Bereich der Bildverarbeitung haben sich Deep-Learning-Verfahren und insbesondere der Einsatz von Faltungsnetzwerken, welche lokale Zusammenhänge in Bilddaten ausnutzen können, bereits stark etabliert. Einen Überblick hierzu bieten Zhao et al. (2017) und Pathak, Pandey & Rautaray (2018). 3D-Punktwolken weisen oft keine triviale lokale Ordnung auf, wodurch der direkte Einsatz von Faltungsnetzen schwierig oder nicht möglich ist. Um die im Kontext visueller Erkennungsaufgaben ausgereiften Verfahren nutzen zu können, müssen die 3D-Daten des PPS zunächst in ein Bilddatenformat überführt werden. Ein probates Format stellen hier gerasterte 2.5D-Oberflächenmodelle dar. Da der PPS sowohl geometrische als auch radiometrische Informationen der Fahrbahnoberfläche aufnimmt, ist jeder Messpunkt durch die 3D-Koordinaten X [m], Y [m] und Z [m] sowie einen Intensitätswert [12 bit] beschrieben. Für eine 2D-Position können demnach sowohl die entsprechende Höheninformation als auch die Information über die Reflexionsintensität als Attribut gespeichert werden.

Da jeder Position nur exakt eine Höhe bzw. eine Intensität zugeordnet wird, spricht man von 2.5D-Modellen.

Im Zuge der Oberflächenmodellierung aus dem Laserscan stellt sich zunächst die Frage, welches Interpolationsverfahren dem Untersuchungsgegenstand, den Wirkungszusammenhängen und den Messdaten angemessen ist. Um die für die gegebene Fragestellung geeignete räumliche Interpolationsmethode aus der Vielzahl existierender Ansätze (Li & Heap, 2008) zu selektieren, wurde eine Nutzwertanalyse durchgeführt. Folgende Kriterien wurden hierfür definiert: die Art des Oberflächenmodells sollte ein Gittermodell sein, die Oberflächeneigenschaften sollten erhalten bleiben, die Interpolationsfunktion also möglichst durch die Messpunkte verlaufen, die Parametersensitivität sowie Rechenkomplexität sollten gering und das visuelle Erscheinungsbild ansprechend sein. Folgende deterministische Interpolationsverfahren wurden aufgrund der Nutzwertanalyse ausgewählt und entlang der definierten Kriterien untersucht: Inverse Distance Weighted, Modified Quadratic Shepard, Natural Neighbour und Thin Plate Spline. Um die Genauigkeit zu bewerten, mit der ein Oberflächenmodell die originalen Messwerte wiedergibt, wurden 3D-Punktwolken repräsentativer Fahrbahnausschnitte in zwei Gruppen unterteilt: zum einen Punkte, die tatsächlich für die Oberflächeninterpolation entlang des Attributs Höhe genutzt werden und zum anderen Punkte, die lediglich der Kontrolle interpolierter Bereiche dienen. An den Kontrollpunkten wurden dann die Genauigkeitsmaße Root Mean Square Error (RMSE) und Mean Average Error (MAE) berechnet. Das visuelle Erscheinungsbild wurde qualitativ auf Basis der Darstellung von markanten Objekten, wie z. B. Fahrbahnmarkierungen, im Intensitätsbild bewertet. Kriterien waren hier beispielsweise eine möglichst scharfe Abgrenzung zwischen Flächen unterschiedlicher Reflexionsintensität. Tabelle 1 bietet einen Überblick über die verglichenen Interpolationsmethoden. Hinsichtlich der Interpolationsgenauigkeit und dem visuellen Erscheinungsbild sind die Natural-Neighbor-Methode sowie die Inverse-Distance-Weighted-Methode am geeignetsten. Unter Berücksichtigung des Kriteriums Rechenzeit wurde die Inverse-Distance-Weighted-Methode mit dem Exponenten 4 als für die Fragestellung optimale Interpolationsmethode zur Erzeugung von 2.5D-Oberflächenmodellen aus 3D-Punktwolken gewählt. Im Ergebnis der Oberflächenmodellierung, welche mit dem freien GIS „SAGA“ (Conrad et al., 2015) durchgeführt wurde, liegen pro 3D-Punktwolke Oberflächenmodelle mit einer Zellauflösung von 0.01 m für die Höheninformation und die Intensitätsinformation vor.

Um die Oberflächenmodelle hinsichtlich einer für die Weiterverarbeitung potenziell problematischen Abhängigkeit von der absoluten Höhenlage zu entkoppeln, wurden aus den absoluten Höhenmodellen relative Oberflächen gerechnet. Hierzu wurde der sogenannte Topographic Position Index (TPI) genutzt (Weiss, 2001). Dieser Algorithmus berechnet innerhalb einer lokalen Nachbarschaft pro Zelle eines Eingabe-Rasters die Differenz aus der Höhe einer Zentral-Zelle zu dem mittleren Höhengiveau der Nachbarzellen. TPI-Werte um den Wert null repräsentieren ein flaches Gelände, negative Werte repräsentieren Hohlformen und positive Werte relative Erhebungen. Schließlich wurden die jeweils zusammengehörigen relativen Höhenmodelle und Intensitätsbilder zu einem 2-Kanal-Komposit verrechnet.

Tabelle 1: Kriterien für die Auswahl einer Interpolationsmethode zur Berechnung der 2.5D-Oberflächenmodelle auf Basis von 3D-Punktwolken des PPS. Die Werte in den geschweiften Klammern stehen für das verwendete Parametersetting: Exponent (*), Quadric Neighbors und Weighting Neighbors (**), sowie Regularisation (***) Die Rechenzeit ist als Faktor zur schnellsten Methode angegeben, die hier als Referenz dient (****).

	RMSE [mm]	MAE [mm]	Rechenzeit [Faktor]****	Visuelles Erscheinungsbild [--- bis +++]
Natural Neighbor {Sibson}	0.79	0.59	2.0	++
Natural Neighbor {Non-Sibsonian}	0.78	0.59	2.1	+++
Natural Neighbor {Linear}	0.82	0.60	1.4	+
Inverse Distance Weighted {2} *	0.79	0.59	1.4	++
Inverse Distance Weighted {4} *	0.81	0.57	1.4	++
Inverse Distance Weighted {6} *	0.84	0.57	1.4	++
Modified Quadric Shepard {5,15} **	4.20	2.01	1.0	---
Modified Quadric Shepard {13,19} **	1.00	0.65	1.0	--
Modified Quadric Shepard {23,29} **	0.86	0.60	1.6	-
Thin Plate Spline {0.001} ***	0.83	0.61	1.8	-
Thin Plate Spline {0.01} ***	0.81	0.60	1.9	-
Thin Plate Spline {0.1} ***	0.78	0.59	1.8	-

2.3 Datensatz

Der Datensatz basiert auf insgesamt zehn verschiedenen Messkampagnen, die im nachgeordneten Netz auf Asphaltflächen mit dem S.T.I.E.R- und I.R.I.S-System aufgenommen wurden. Insgesamt wurden Messdaten auf einer Länge von 30 nicht zusammenhängenden Netzkilometern entsprechend der in Kapitel 2.2 beschriebenen Methodik aufbereitet, sodass die resultierenden Oberflächenmodelle in einem 10-Meter-Intervall vorlagen. Aus dieser Gesamtheit wurden zunächst automatisch zufällige Beispiele und daraus wiederum 220 Beispiele mit repräsentativen Schadensbildern manuell ausgewählt. Diese Auswahl wurde schließlich anhand eines umfangreichen Klassenkatalogs händisch annotiert. Die hierbei verwendeten Polygon-Masken wurden mithilfe der Software QGIS erzeugt (QGIS Development Team, 2019). Insgesamt wurden beim Labelling 6 Schad- und 9 Objektklassen unterschieden. Abb. 2 zeigt unter C ein Beispiel für ein gelabeltes Oberflächenmodell. Für das Training der Detektoren, welches in Kapitel 4.1 beschrieben ist, wurden die insgesamt 15 verwendeten Klassen zunächst auf zwei Klassen reduziert: Dabei wurden alle Schadklassen, wie z. B. „Riss“, „Ausbruch“ oder „Flickstelle“, zu einer Klasse „Schaden“ zusammengefasst. Alle Objektklassen, wie z. B. „Ablauf“, „Markierungslinie“ oder „Bordstein“, wurden zur Klasse „Schadfrei“ hinzugefügt. Dieses Labelling adressiert in erster Linie die reine Detektion von Schadstellen. In einem zweiten Schritt wurde der originale Klassenkatalog so zusammengefasst, dass auch die Klassifikation verschiedener Schadenstypen untersucht werden kann. Hierzu wurden alle rissähnlichen Schadklassen, wie z. B. „Riss“ oder „Offene Naht“ zur Klasse „Riss“ und alle flächenhaften Schadklassen, wie z. B. „Ausbruch“ oder „Ausmagerung“, zur Klasse „Ausbruch“ zusammengefasst. Die Schadklasse „Flickstelle“ blieb als sol-

che erhalten. Hinsichtlich der Objektklassen wurde die Klasse „Inventar“ aus den Einzelklassen „Schieber“, „Ablauf“ und „Schacht“ zusammengeführt, die Klasse „Markierungslinie“ blieb erhalten und alle restlichen Objektklassen wurden in der Klasse „Schadfrei“ konzentriert.

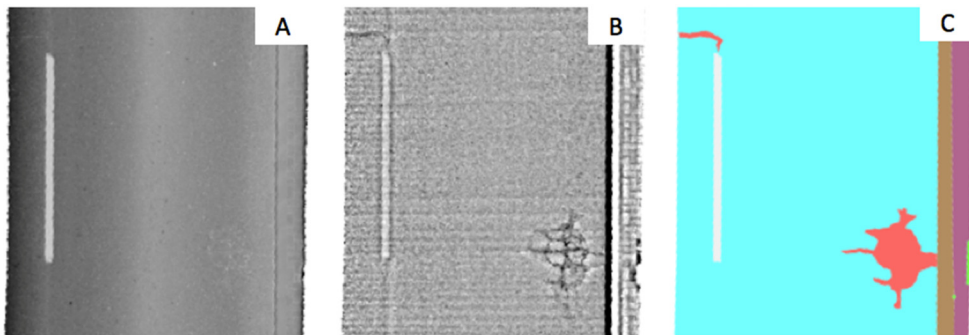


Abb. 2: Von links nach rechts sind die einzelnen im Beitrag verwendeten Daten beispielhaft dargestellt: Intensitätsbild (A), relatives Oberflächenmodell (B) sowie Objekt- und Schadklassen als Polygon-Masken (C)

3 Methodik

3.1 Schadstellendetektion mittels Faltungsnetzwerken

Die automatisierte Detektion von Schadstellen der Straßenoberfläche stellt ein Forschungsfeld dar, welches, beginnend mit ersten Arbeiten von Lee (1991), seit fast drei Dekaden eine Vielzahl unterschiedlicher Verfahren hervorgebracht hat. Bis vor wenigen Jahren kamen dabei vor allem Verfahren der klassischen Bild- und Datenverarbeitung zum Einsatz, welche sich grob in zwei Gruppen aufteilen lassen.

Die erste Gruppe an Verfahren sind auf die Detektion von Rissen spezialisiert und versuchen diese durch die Suche nach lokalen Intensitätsminima zu extrahieren (Oliveira et al., 2014; Peng et al., 2015). Die zweite Gruppe an Verfahren basiert auf der Extraktion von Merkmalsvektoren für lokale Bereiche des Bildes, welche anschließend an einen Klassifikator übergeben werden. Für die Extraktion der Merkmalsvektoren kommen beispielsweise Histogramme von orientierten Gradienten (Kapela et al., 2015), Local-Binary-Pattern (Quintana et al., 2016) oder andere handdesignte Merkmalsextraktionsverfahren zum Einsatz. Als Klassifikatoren werden vorzugsweise Support Vector Machines oder flache Neuronale Netzwerke eingesetzt, welche auf einem Trainingsdatensatz überwacht trainiert werden. Diese klassischen Methoden des Maschinellen Lernens haben den Vorteil, dass sie, durch das händische Merkmalsdesign, mit einer relativ geringen Menge an Trainingsdaten trainiert werden können. Nachteilig wirkt sich dabei jedoch aus, dass durch die händisch designte Merkmalsextraktion für den Klassifikator wichtige Merkmale unbeachtet bleiben können, wodurch die Leistungsfähigkeit beeinträchtigt wird.

Durch den Einsatz von tiefen Faltungsnetzwerken (Convolutional Neural Networks – CNNs) kann das explizite Design einer Merkmalsextraktion entfallen. Das Neuronale Netzwerk wird direkt auf Bildern oder Bildausschnitten trainiert und lernt durch eine Vielzahl an Trainingsbeispielen eine implizite Merkmalsextraktion innerhalb der Schichten des Neuronalen Netzwerks.

Um Trainingsdaten zu generieren, werden verschiedene Teilbilder gleicher Größe, welche auch als Patches bezeichnet werden, aus Oberflächenbildern ausgeschnitten (Abb. 3). Diesen Patches wird jeweils eine Zielklasse, zum Beispiel „Schaden“ oder „Schadfrei“, zugewiesen und das Netzwerk damit trainiert. Anschließend kann das Netzwerk durch Ausschneiden von Patches auf einem unbekanntem Oberflächenbild angewendet werden und erzeugt dafür eine Schadstellenkarte (Abb. 4). Weitere Details zu diesem Trainingsvorgang und wie das Netzwerk transformiert werden kann, sodass die Detektion auf einem ganzen Bild erfolgen kann, ohne dass dazu Teilbilder ausgeschnitten werden müssen, können in Eisenbach et al. (2017) nachgeschlagen werden. Wichtig ist für das Training, dass der Datensatz eine möglichst große Varianz aufweist und alle relevanten Ausprägungen von Schäden und Objekten abdeckt.

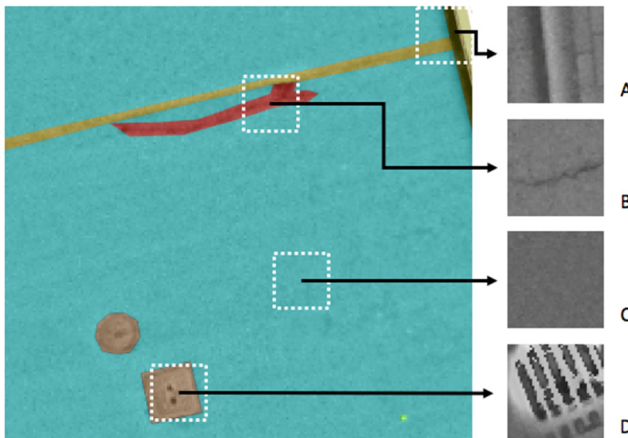


Abb. 3: Schematische Darstellung der Extraktion von Patches. Die linke Teilabbildung zeigt ein mit Polygon-Masken gelabeltes Oberflächenmodell. Rechts sind extrahierte Patches dargestellt: ein Bereich mit Bordstein (A), ein Bereich mit einem Riss (B), ein schadffreier Bereich (C) und ein Bereich mit einem Ablauf (D).

3.2 Residual Neural Networks

Aufgrund der großen Anzahl an Schichten, haben sich bei Faltungsnetzwerken unterschiedliche Netzwerkarchitekturen durchgesetzt. VGG-basierte Netzwerke stellen z. B. eine klassische Netzwerkarchitektur für Faltungsnetzwerke dar, welche auch in Eisenbach et al. (2017) für die Schadstellenklassifikation in Kamerabildern verwendet wurde. Aufgrund der großen Anzahl an Gewichten müssen bei dieser Netzwerkarchitektur jedoch viele Regularisierungstechniken eingesetzt werden, damit sich das Netzwerk nicht zu stark auf die Trainingsdaten spezialisiert und auch auf Testdaten eine gute Klassifikation ermöglicht.

Für die Experimente kommt daher mit Residual Neural Networks (He et al., 2016) eine modernere Netzwerkarchitektur zum Einsatz, welche für die Problemstellung deutlich schneller als VGG-basierte Netzwerke trainiert werden kann und darüber hinaus bereits durch Stricker et al. (2019) erfolgreich für die Schadstellendetektion aus Kamerabildern eingesetzt wurde. Die grundlegende Idee von Residual Networks ist die Einführung von Shortcut-Connections. Das Netzwerk wird dabei in Blöcken von Faltungsschichten organisiert, wobei jeweils zwischen zwei aufeinanderfolgenden Blöcken eine zusätzliche Shortcut-Connection existiert (Abb. 4). Dadurch wird der Gradientenfluss während des Trainings verbessert und ein Block kann sich jeweils auf die Lösung eines Teilproblems konzentrieren. Da Residual Networks außerdem auf vollverschaltete Schichten weitestgehend verzichten, kommen sie, trotz größerer Tiefe, mit deutlich weniger Gewichten als VGG-basierte Netzwerke aus. In ersten Vor-Tests hat sich ein Residual Netzwerk mit 20 Faltungsschichten als guter Kompromiss zwischen Laufzeit und Genauigkeit herausgestellt und wurde daher auch für die Experimente dieses Beitrags verwendet (Abb. 4).

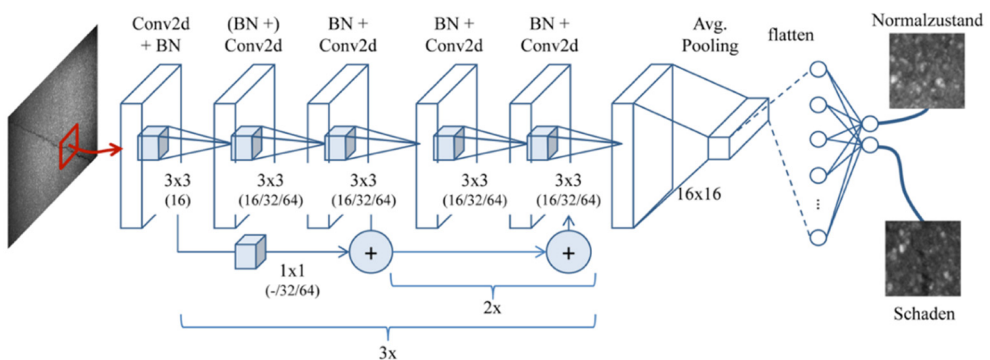


Abb. 4: Schematische Darstellung eines Residual Networks mit 20 Faltungsschichten (Conv2d). Ein Block besteht jeweils aus zwei Faltungsschichten und die Ausgaben des Vorgängerblocks sind immer mit der Ausgabe des aktuellen Blocks durch eine Additionsschicht verbunden. Nach jeder Faltungsschicht folgt jeweils eine Batch-Normalization-Schicht (BN). Jeder Patch des Oberflächenbildes, wird jeweils einer Klasse zugeordnet (z. B. Normalzustand/Schaden), sodass sich als Ergebnis eine Schadstellen- oder Objektkarte für das Oberflächenbild ergibt.

4 Ergebnisse

In diesem Abschnitt werden Untersuchungen bzw. Experimente zur Analyse der Problemstellung und des Datensatzes vorgestellt. Dabei wird untersucht, wieviel Kontext der Klassifikator benötigt und welche Inputarten am besten geeignet sind, um eine sichere Schadstellendetektion zu erzielen. Anschließend wird die Leistungsfähigkeit des Systems in Hinblick auf die Klassifikation unterschiedlicher Schadens- und Objekttypen analysiert.

4.1 Training und Validierung von Detektoren

Für die durchgeführten Untersuchungen werden die Neuronale Netzwerke jeweils ausschließlich auf dem Trainingsdatensatz trainiert und haben somit niemals explizit die Daten des Validierungs- und Testdatensatzes erlernt. Die Auswahl des besten Klassifikators erfolgte anschließend auf dem Validierungsdatensatz, um zu gewährleisten, dass der Klassifikator nicht zu stark an die Trainingsdaten adaptiert ist. Der Testdatensatz wird anschließend nur verwendet, um die Ergebnisse auf unbekanntem Daten zu simulieren und dient nicht der Auswahl einzelner Netzwerke. In Tabelle 2 ist die Verteilung der Datensätze zu sehen, die für die Experimente verwendet wurden. Während des Trainings wurde Data-Augmentation genutzt, um die Anzahl der Trainingsdaten synthetisch zu erhöhen. Dabei werden die Trainingspatches zufällig translatiert und rotiert.

Aufgrund der unbalancierten Daten, erfolgte die Bewertung der Güte des Klassifikators anhand des F1-Scores. Dieser berücksichtigt sowohl Falsch-Positiv- als Falsch-Negativ Klassifikationen und ist somit deutlich besser für die Beurteilung der Güte des Klassifikators geeignet als beispielsweise die Accuracy. Jedes Training wurde außerdem dreimal durchgeführt, wobei, wenn nicht anders genannt, im vorliegenden Beitrag jeweils der Mittelwert des erzielten F1-Scores angegeben ist.

Tabelle 2: Übersicht über die Klassenbesetzung der Datensätze für die reine Detektion von Schadstellen (oben) und für die Klassifikation von Schadstellen und Objekten (unten). Die Zahlenwerte entsprechen der Anzahl verwendeter Patches

Datensatz	Objekt-/Schadklasse	
	Schadfrei	Schaden
Training	44956	19448
Validierung	5253	1802
Test	5991	1235

Datensatz	Objekt-/Schadklasse					
	Schadfrei	Ausbruch	Inventar	Markierung	Riss	Flickstelle
Training	36653	3100	1214	7089	14111	2237
Validierung	4674	333	281	298	1234	235
Test	4890	163	98	1003	1007	65

4.2 Schadstellendetektion

Zunächst wurde mit dem Datensatz für das Zweiklassenproblem der Schadstellendetektion untersucht, wie viel Kontext für den Detektor erforderlich sind, um sicher zwischen schadhafte und schadfreien Bereichen unterscheiden zu können. Dazu wurden die Patches des Trainingsdatensatzes an jeweils den gleichen Positionen, aber mit unterschiedlichen Größen aus den Oberflächenbildern gezogen. Die Größen variierten dabei in 8-Pixel-Schritten von 16 bis 96 Pixeln, was, auf die zugrunde liegende Straßenoberfläche bezogen, einer Ausdehnung von 0.16 m bis 0.96 m entspricht.

In Abbildung 5 ist zu erkennen, dass die Detektionsleistung bis zu einer Patchgröße von 32 Pixeln auf den Validierungsdaten sehr stark ansteigt. Darüber hinaus fällt der Anstieg deutlich geringer aus und nimmt ab 80 Pixeln sogar leicht ab. Ein ähnliches Bild bietet sich bei den Testdaten. Hier ist allerdings ersichtlich, dass bereits ab einer Patchgröße von 64 Pixeln der Detektor keinen Nutzen aus zusätzlichen Kontextinformationen ziehen kann. Viel mehr zeigt sich, dass zu große Patchgrößen die Detektion verschlechtern, da sich ab dieser Größe sehr oft mehrere Schadstellen in einem Patch befinden. Für die weiteren Untersuchungen wird daher eine Patchgröße von 64 Pixeln als ideal angenommen.

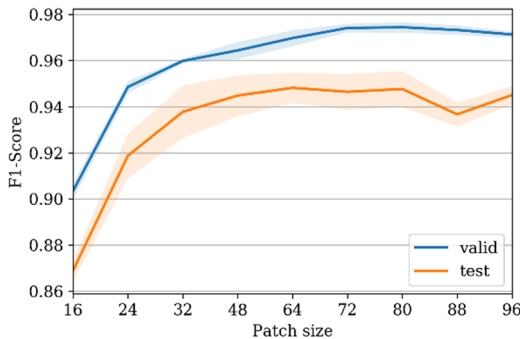


Abb. 5: Detektionsleistung auf dem Validierungs- und Testdatensatz in Abhängigkeit der Patchgröße. Angegeben sind jeweils der F1-Score und die Schwankungsbreite der Ergebnisse über die drei Durchläufe.

In einem weiteren Experiment wurde untersucht, welche Art der Inputmodalität für die Schadstellendetektion am besten geeignet ist. Dazu wurden Netzwerke ausschließlich auf Intensitätspatches, ausschließlich auf Höhenpatches und auf der Kombination beider Inputmodalitäten trainiert. Wie in Tabelle 3 zu sehen ist, sind die auf den Höhendaten trainierten Detektoren bereits auf den Validierungsdaten deutlich besser für die Detektion geeignet, als Detektoren, die nur den Intensitätskanal verwenden. Noch deutlicher wird dieser Unterschied auf Testdaten, da hier die Detektion auf reinen Intensitätsbildern noch einmal deutlich abfällt. Es ist zu vermuten, dass die Varianz in den Intensitätsbildern durch unterschiedliche Straßenoberflächen deutlich größer ist, was die Detektion erschwert. Dass die Intensitätsdaten dennoch zu einer Verbesserung beitragen können, zeigt sich bei den auf beiden Inputmodalitäten trainierten Netzwerken, welche gegenüber den rein auf Höhendaten trainierten Detektoren noch einmal eine leichte Steigerung erzielen können.

Inputmodalität	Validierung (F1-Score)	Test (F1-Score)
Intensität	0.88	0.78
Höhe	0.94	0.93
Intensität + Höhe	0.97	0.95

Tabelle 3: Vergleich der Detektionsleistung bei Training auf unterschiedlichen Inputmodalitäten

Um einen Eindruck von der Detektionsleistung für die Schadstellendetektion zu bekommen, ist die Konfusionsmatrix des besten Detektors, der auf beiden Inputkanälen trainiert ist, in Abbildung 6 dargestellt.

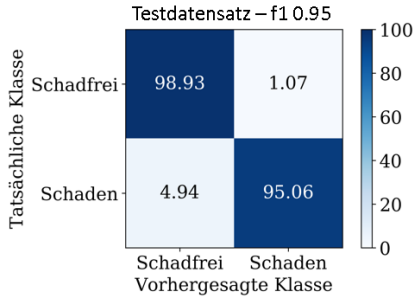


Abb. 6: Konfusionsmatrix für die Schadstellendetektion auf Testdaten. Das Training erfolgte auf Intensitäts- und Höhenwerten.

Hierbei ist zu berücksichtigen, dass ein Schaden typischerweise aus mehreren Patches besteht. Wenn ein Patch nicht als Schaden detektiert wurde, heißt dies daher nicht zwangsläufig, dass der komplette Schaden übersehen wurde. Ein beispielhaftes Detektionsergebnis – im Folgenden als Konfidenzraster bezeichnet – ist in Abbildung 7 gegeben.

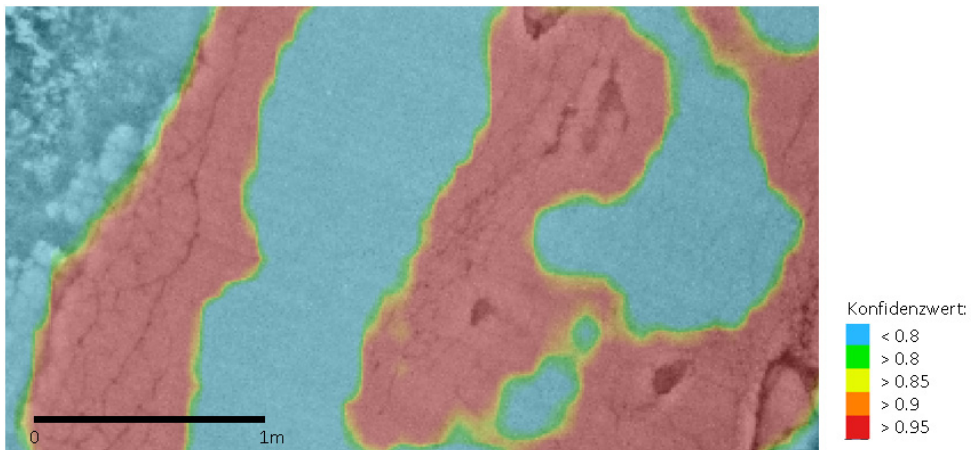


Abb. 7: Beispiel für die Visualisierung der Ergebnisse der Schadensdetektion. Das Bild zeigt ein Konfidenzraster, wobei der Konfidenzwert für die Klasse „Schaden“ mit zunehmender räumlicher Nähe zum schadhafte Bereich zunimmt. Der maximale Konfidenzwert ist 1.0 und ist in der Abbildung rot dargestellt. Das Konfidenzraster ist als Überlagerung über dem Eingabedatensatz dargestellt. Die dunklen flächigen Areale sind dabei Ausbrüche und die dunklen linearen Strukturen sind Risse. Beiden texturierten Bereichen in der oberen linken Ecke handelt es sich um Bankett-Bereiche, die korrekt als schadfrei klassifiziert wurden.

4.3 Schadstellenklassifikation

Da für eine Zustandsbewertung der Straßenoberfläche eine reine Schadstellendetektion oft nicht ausreichend ist, wurde weiterhin untersucht, ob mit dem Ansatz auch eine Untergliederung in einzelne Schad- und Objektarten möglich ist. Es wurden daher Netzwerke mit 64er Patchgröße auf den kombinierten Intensitäts- und Höhendaten trainiert, welche eine Unter-

scheidung in 6 Klassen ermöglichen (Tabelle 2 unten). Mit einem erreichten F1-Score von 0.91 auf den Testdaten schneidet der Klassifikator nur wenig schlechter ab als bei der reinen Schadensdetektion. Die in Abbildung 8 gegebene Konfusionsmatrix zeigt, dass die Klassifikation für alle Klassen auf einem guten Niveau liegt. Wesentliche Verwechslungen treten in manchen Fällen nur zwischen „Ausbruch“ und „Riss“ und zwischen „Inventar“ und „Schadfrei“ auf. Hier ist allerdings zu bedenken, dass Fehldetektionen auf Patchebene oftmals durch eine geeignete Nachverarbeitung beseitigt werden können. Eine beispielhafte Klassifikation auf den Testdaten ist in Abbildung 9 gegeben.

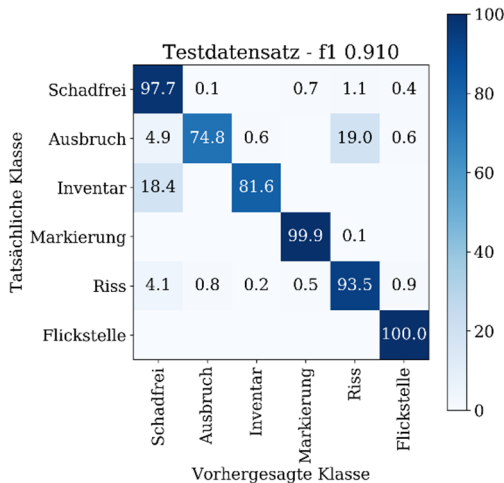


Abb. 8: Konfusionsmatrix für die Schadstellendeckung auf Testdaten. Als Inputmodalitäten wurden die Intensitäts- und Höhenwerten verwendet.

4.4 Fahrbahnschäden als georeferenzierte Einzelobjekte

Als Ausgabe produzieren die vorstehend beschriebenen Detektoren sogenannte Konfidenzraster. Ein solches bildet pixelweise den jeweiligen Konfidenzwert für die entsprechende Ausgabe Klasse ab. Dieser kann einen Wert zwischen 0 und 1 annehmen und – sehr vereinfacht ausgedrückt – als eine Art Sicherheitsindikator für die Klassen-Entscheidung des Netzes interpretiert werden. Für eine weiterführende Analyse im GIS oder in Pavement Management Systemen bietet sich eine Umwandlung der Konfidenzraster in ein Vektordatenformat an. So lassen sich die Ergebnisse einerseits performant verwalten, es können darüber hinaus aber auch weitere Attribute, wie z. B. Fläche, Orientierung, oder Form-Deskriptoren, wie der Shape Index, mit Standard-GIS-Tools berechnet werden.

Das Beispiel in Abbildung 9 zeigt die Ergebnisse des 6-Klassen-Detektors für die Klasse Riss auf einem 10 Meter langen und 4 Meter breiten Fahrbahnabschnitt. Ganz links sind unter A die manuell gesetzten Polygone für die Klasse Riss zu sehen, die als Referenz dienen sollen. Teilabbildung B zeigt das Klassifikationsergebnis in Form eines Konfidenz-Rasters. Im rechten Teil der Abbildung ist unter C das Ergebnis nach dem Beschränken auf Konfidenzwerte größer als 0.85, sowie dem Vektorisieren und der Attributberechnung zu sehen.

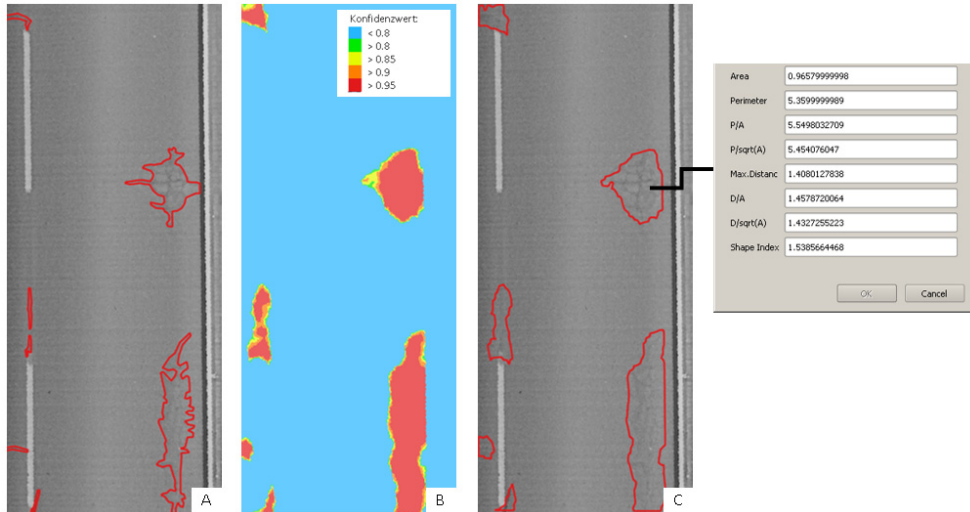


Abb. 9: Exemplarische Visualisierung der Klassifikationsergebnisse in Form eines Konfidenzrasters (B) und von attribuierten Vektorgeometrien (C) im Vergleich zu manuell definierten Schadenspolygonen für die Klasse Riss

Auch wenn in diesem Beispiel alle tatsächlich betroffenen Riss-Areale von den automatisch detektierten Polygonen abgedeckt werden, sieht man deutlich, dass der Detektor einen gewissen Bereich um die eigentliche Tiefenlinie des Risses ebenfalls als betroffenen Fläche klassifiziert. Im gegebenen Beispiel sind in der Summe 1.92 m² tatsächlich vom Merkmal Riss betroffen. Die automatische Analyse liefert nach der Vektorisierung hier eine Schadensfläche von 3.67 m².

5 Schlussfolgerung

Im vorliegenden Beitrag wurde gezeigt, dass die Inverse Distance Weighted Interpolation in Bezug auf die Kriterien Genauigkeit, Rechenzeit und visuelles Erscheinungsbild zur Erzeugung von Intensitätsbildern und Höhenmodellen aus 3D-Punktwolken des Fraunhofer Pavement Profile Scanners im Kontext der Problemstellung am besten geeignet ist. In Form dieser 2.5D-Oberflächenmodelle können LiDAR-Daten von Fahrbahnoberflächen als Eingabe für Faltungsnetzwerke dienen und mithilfe tiefer neuronaler Netze automatisiert hinsichtlich Schäden und Objekten analysiert werden.

Mit dem trainierten Detektor für ein Zwei-Klassen-Problem (Schaden vs. schadfrei) wurde auch auf unabhängigen Testdaten eine sehr hohe Detektionsleistung demonstriert. Im Zuge der Experimente zeigte sich, dass jene Detektoren, die eine Kombination von radiometrischer und geometrischer Information als Datengrundlage nutzten, die höchsten F1-Scores bei der Schadensdetektion erreichten. Darüber hinaus war das verwendete tiefe neuronale Netz in der Lage, aus den Oberflächenmodellen robuste Merkmale zu extrahieren, die auch eine automatisierte Unterscheidung von Schäden und Objekten auf einem sehr hohen Niveau ermöglichen. Da die mit dem I.R.I.S- und S.T.I.E.R-System erfassten LiDAR-Daten präzise

georeferenziert sind, können die Detektions- bzw. Klassifikationsergebnisse als einzelne Geobjekte z. B. in Form von Polygonen extrahiert und so für weiterführende Analysen in GIS und Pavement Management Systeme integriert werden. Künftige Arbeiten werden sich darauf konzentrieren, die Datenbasis für Straßen in Asphaltbauweise weiter zu verbreitern und somit noch leistungsfähigere Detektoren zu trainieren. Vor allem wird eine sichere Klassifikation der „ZTV ZEB-StB“-relevanten Schadenstypen angestrebt.

Literatur

- Applanix (2015). *POS LV specifications – Mobile geospatial data acquisition*. Retrieved Jan 25, 2019, from https://www.applanix.com/pdf/specs/ POSLV_Specifications_dec_2015.pdf.
- Bundesministerium für Verkehr und digitale Infrastruktur (2016). *Bundesverkehrswegeplan 2030*. Retrieved Jan 25, 2019, from https://www.bmvi.de/SharedDocs/DE/Anlage/VerkehrUndMobilitaet/BVWP/bvwp-2030-gesamtplan.pdf?__blob=publicationFile.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, *Geoscientific Model Development*, 8, 1991–2007.
- Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stöckert, U., & Gross, H.-M. (2017). How to Get Pavement Distress Detection Ready for Deep Learning? A Systematic Approach. In: *Int. Joint Conf. on Neural Networks* (pp. 2039–2047). New York, NY: IEEE.
- Forschungsgesellschaft für Straßen- und Verkehrswesen (2006). *Zusätzliche Technische Vertragsbedingungen und Richtlinien zur Zustandserfassung und -bewertung von Straßen ZTV ZEB- StB*, Köln: FGSV-Verlag.
- Fraunhofer (2017). *Pavement Profile Scanner*. Retrieved Jan 25, 2019, from <https://www.ipm.fraunhofer.de/content/dam/ipm/de/PDFs/produktblaetter/OF/MTS/Pavement-Profile-Scanner-de.pdf>.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity Mappings in Deep Residual Networks. *European Conference on Computer Vision* (pp. 630–645). Cham: Springer International Publishing.
- Kapela, R., Sniatała, P., Turkot, A., Rybarczyk, A., Pozarycki, A., Rydzewski, P., Wyczałek, M., & Błoch, A. (2015). Asphalt surfaced pavement cracks detection based on histograms of oriented gradients. In: *Int. Conference on Mixed Design of Integrated Circuits & Systems*, June, 2015, Toruń, Poland (pp. 579–584).
- Karaca, Y., Cattani, C., & Moonis, M. (2017). Comparison of Deep Learning and Support Vector Machine Learning for Subgroups of Multiple Sclerosis. In: *Computational Science and Its Applications – ICCSA 2017*, July 3–6, 2017, Trieste, Italy (pp. 142–153).
- Lee, H. (1991). Application of Machine Vision Techniques for the Evaluation of Highway Pavements in Unstructured Environments. In: *Proceedings of the Fifth International Conference on Advanced Robotics* (pp. 1425–1428).
- Li, J., & Heap, A. D. (2008). A Review of Spatial Interpolation Methods for Environmental Scientists. *Geoscience Australia Record*, 23.
- Liu, P., Choo, K.-K. R., Wang, L., & Huan, F. (2016). SVM or deep learning? A comparative study on remote sensing image classification. *Soft Computing*, 21(23), 7053–7065.

- Lundberg, T., Andrén, P., Wahlman, T., Eriksson, O., Sjögren, L., & Ekdahl, P. (2018). New technology for road surface measurement. *VTI rapport 961A*.
- Oliveira, H., & Correia, P. L. (2014). CrackIt – An Image Processing Toolbox for Crack Detection and Characterization. In: *International Conference on Image Processing* (pp. 798–802). New York, NY: IEEE.
- Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Application of Deep Learning for Object Detection. *Procedia Computer Science*, 132, 1706–1717.
- Peng, L., Chao, W., Shuangmiao, L., & Baocai, F. (2015). Research on Crack Detection Method of Airport Runway Based on Twice-Threshold Segmentation. In: *Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control* (pp. 1716–1720). New York, NY: IEEE.
- QGIS Development Team (2019). *QGIS Geographisches Informationssystem. Open Source Geospatial Foundation Projekt*. Retrieved Jan 25, 2019, from <http://qgis.osgeo.org>.
- Quintana, M., Torres, J., & Menendez, J. M. (2016). A Simplified Computer Vision System for Road Surface Inspection and Maintenance. *Transactions on Intelligent Transportation Systems*, 17(3), pp. 608–619.
- Rachmadi, M. F., Valdés-Hernández, M. C., Agan, M. L. F., & Komura, T. (2017). Deep Learning vs. Conventional Machine Learning: Pilot Study of WMH Segmentation in Brain MRI with Absence or Mild Vascular Pathology. *Journal of Imaging*, 3(66), 1–19.
- Reiterer, A., Dambacher, M., Maindorfer, I., Höfler, H., Ebersbach, D., Frey, C., Scheller, S., & Klose, D. (2013). Straßenzustandsüberwachung in Sub-Millimeter. In: T. Luhmann, & C. Müller (Eds.), *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik: Beiträge der Oldenburger 3D-Tage 2013* (pp. 78–85). Berlin/Offenbach: Wichmann.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252.
- Sesselmann, M. (2018). Automatische Detektion von Substanzmerkmalen auf 3D-Fahrbahnoberflächen. *AGIT – Journal für angewandte Geoinformatik*, (4), 65–74. doi: 10.14627/537647009.
- Stricker, R., Eisenbach, M., Sesselmann, M., Debes, K., & Gross, H.-M. (2019). Improving Visual Road Condition Assessment by Extensive Experiments on the Extended GAPS Dataset. In: *Int. Joint Conf. on Neural Networks (IJCNN)*, IEEE, Budapest, Hungary, 2019, to appear.
- Weiss, A. (2001). Topographic position and landforms analysis (Poster presentation). In: *ESRI User Conference*, July 9–13, 2001, San Diego, CA (pp. 227–245).
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2017). *Object Detection with Deep Learning: A Review*. Retrieved Jan 25, 2019, from <https://arxiv.org/pdf/1807.05511v1.pdf>.