

Frederic Hake

**Schadenserkennung an Bauwerken
mittels maschinellen Lernens**

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

Schadenserkennung an Bauwerken mittels
maschinellen Lernens

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte
Dissertation

von

Frederic Hake, M. Sc.

München 2025

Verlag der Bayerischen Akademie der Wissenschaften, München

Adresse des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften:



Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften

Alfons-Goppel-Straße 11 • D – 80 539 München
Telefon +49 – 89 – 23 031 1113 • Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail post@dgk.badw.de • <http://www.dgk.badw.de>

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. habil. Jürgen Müller
Referent: PD Dr.-Ing. Hamza Alkhatib
Korreferenten: Prof. Dr.-Ing. Ingo Neumann
Prof. Dr.-Ing. Vincent Oettel
Prof. Dr.-Ing. Harald Sternberg

Tag der mündlichen Prüfung: 20.09.2024

Diese Dissertation ist auf dem Server des Ausschusses Geodäsie (DGK)
der Bayerischen Akademie der Wissenschaften, München unter <http://dgk.badw.de/>
sowie unter Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik
der Leibniz Universität Hannover (ISSN 0174-1454), Nr. 403,
unter <https://repo.uni-hannover.de/handle/123456789/18854>, Hannover 2025 elektronisch publiziert

© 2025 Ausschuss Geodäsie (DGK) der Bayerischen Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus zu vervielfältigen.

Erklärung

Hiermit erkläre ich, dass ich die Regeln der geltenden Promotionsordnung kenne und eingehalten habe. Ich bin mit einer Prüfung nach den Bestimmungen der Promotionsordnung einverstanden.

Ich habe die Dissertation selbst verfasst und keine Textabschnitte von Dritten oder eigenen Prüfungsarbeiten ohne Kennzeichnung übernommen. Alle von mir benutzten Hilfsmittel und Quellen habe ich in meiner Arbeit angegeben.

Dritten habe ich weder unmittelbar noch mittelbar geldwerte Leistungen für Vermittlungstätigkeiten oder für die inhaltliche Ausarbeitung der Dissertation zukommen lassen.

Ich habe die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht.

Frederic Hake

4. Juni 2024

Kurzfassung

Der Güterumschlag in den 60 deutschen Seehäfen hat in den letzten Jahren zugenommen und soll bis 2030 auf etwa 468 Millionen Tonnen steigen. Der steigende Bedarf des Güterumschlags stellt eine erhebliche Belastung für die bestehende Infrastruktur dar. Die vorhandene Infrastruktur unterliegt bereits durch Überbelastung, Salzwassereinwirkung und dem Alter der Bauwerke einem hohen Verschleiß. Die Überprüfung der Verkehrswasserbauwerke ist für den Erhalt einer wettbewerbsfähigen Volkswirtschaft unerlässlich. Frühzeitige Schadenserkenkung ist die Grundlage für kostengünstige Erhaltungsmaßnahmen, da frühzeitig erkannte Schäden in der Regel geringere Instandsetzungskosten verursachen. Infrastrukturbauwerke werden derzeit regelmäßig durch manuelle Kontrollen überprüft. Dies ist besonders unter Wasser mit erheblichem Aufwand verbunden, da Taucher eingesetzt und komplette Bereiche gesperrt werden müssen. Dies verdeutlicht den hohen Bedarf an automatisierten, digitalen Prozessen in der Bauwerksprüfung, um diese zukünftig vollständig, transparent und kosteneffizient gestalten zu können.

Die vorliegende Dissertation konzentriert sich auf die automatisierte Erkennung von Schäden aus 3D-Punktwolken und Farbbildern. Die Daten werden mit einem kinematischen Multisensorsystem auf einer schwimmenden Plattform erfasst. Aufgrund der Plattformbewegungen durch Wellengang und Strömungen sind die entstehenden 3D-Punktwolken und Farbbilder sehr herausfordernd. Die automatisierte Segmentierung der Daten in beschädigte und unbeschädigte Bereiche stellt aufgrund der unterschiedlichen Schadenstrukturen und der Datenkomplexität eine große Herausforderung dar. Die automatisierte Schadenserkenkung ist ein wichtiger Aspekt im Gesamtprozess der digitalen Bauwerksprüfung, da Schadensverdachtspläne erzeugt werden können, die von Bauwerksprüfern für eine schnellere und kosteneffizientere Prüfung genutzt werden. Durch die vollständige Erfassung der Infrastruktur und der automatisierten Schadenserkenkung wird die Bauwerksprüfung auf ein völlig neues Niveau gebracht. Aufgrund der Komplexität des Gesamtprozesses ist die Lösung der Problemstellung als langfristige Aufgabe zu betrachten. Diese Dissertation zielt darauf ab, die Vorgehensweise durch zielgerichtete Beiträge konstruktiv zu unterstützen und bietet gleichzeitig eine Basis für zukünftige Forschungsarbeiten, die die Integration weiterer Sensordaten und fortschrittlicherer Algorithmen zur Schadenserkenkung untersuchen könnten:

- Es wird ein Verfahren entwickelt, um geometrische Bauwerksschäden in 3D-Punktwolken mithilfe von Algorithmen des Machine Learning automatisiert zu erkennen. Dabei werden verschiedene Methoden zur Erzeugung von Höhenfeldern aus 3D-Punktwolken als Vorverarbeitungsschritt für die automatisierte Schadenserkenkung analysiert. Hierfür werden Computer-Aided Design (CAD)-Modelle oder aus den Daten bestmöglich geschätzte geometrische Modelle verwendet. Die darauf aufbauende automatisierte Schadenserkenkung nutzt generische Merkmalsgeneratoren und Ausreißererkenkungsmethoden. Die Methode erreichte einen F1-Score von 96,3 % für einen simulierten Datensatz und 72,4 % für einen realen Datensatz.
- Es wird ein Verfahren entwickelt, um Schäden, welche eine Farbänderung hervorrufen, automatisiert aus Bildern von Verkehrswasserbauwerken durch einen Deep Learning Ansatz zu erkennen. In diesem Kontext wird auch untersucht, ob sich automatisiert generierte, schwach überwachte Labelbilder für das Training der Algorithmen eignen. Zusätzlich zu den segmentierten Segmenten werden Unsicherheitswerte prädiziert. Dies wird am Beispiel von Fotos des JadeWeserPorts in Wilhelmshaven evaluiert. Für die Segmentierung der Bilder in Korrosion und unbeschädigte Bereiche wird ein F1-Score von 98,0 % für unbeschädigte Bereiche und von 71,7 % für die Korrosionsschäden erreicht.

Die Ergebnisse dieser Arbeit zeigen, dass die gezielten Beiträge zur automatisierten Schadenserkenkung den Prozess der digitalen Bauwerksprüfung vollständiger, transparenter und effizienter machen können. Dies trägt dazu bei, die Sicherheit und Lebensdauer der Infrastruktur zu erhöhen und die Betriebskosten zu senken. Die verbesserte Überwachung kann in die statische Beurteilung einfließen, was dazu führen kann, dass Bestandsbauwerke im Mittel länger genutzt werden und somit die Produktivität der Häfen dauerhaft erhöht werden kann. Zudem werden Inspektionsprozesse effizienter und kostengünstiger, was langfristig zur Nachhaltigkeit der Infrastruktur beiträgt.

Abstract

The volume of goods handled in German seaports has significantly increased in recent years and is projected to reach approximately 468 million tonnes by 2030. This growing demand for cargo handling is placing considerable strain on the existing port infrastructure, which is already subject to severe wear and tear due to overloading, saltwater exposure, and aging. Regular inspection of maritime transport structures is essential for maintaining a competitive economy. Early damage detection is the basis for cost-effective maintenance measures, as early damage detection generally results in lower repair costs. Currently, infrastructure is inspected through regular manual checks, which are particularly labor-intensive underwater, requiring the deployment of divers and closure of entire areas. This illustrates the high demand for automated, digital processes in structural inspection to ensure a thorough, transparent and cost-effective inspections in the future.

The focus of this dissertation is the automatic detection of damage from 3D point clouds and color images. Data acquisition is performed using a kinematic multi-sensor system mounted on a floating platform. Due to the platform motion caused by waves and currents, the resulting 3D point clouds and colour images are complex and unstructured. Automated segmentation of the data into damaged and undamaged areas is a major challenge due to the different damage structures and data complexity. Comprehensive digital recording of infrastructure combined with automated damage detection elevates structural inspection to a new level of efficiency and accuracy. The system can generate detailed damage probability maps, enabling structural inspectors to conduct faster and more cost-effective inspections. Due to the complexity of the overall process, the solution to the problem must be seen as a long-term task. This dissertation aims to constructively support this approach through targeted contributions while also providing a basis for future research that could investigate the integration of additional sensor data and more advanced damage detection algorithms:

- This research develops a method to automatically recognize geometric structural damage in 3D point clouds using machine learning algorithms. The research analyzes various methods for generating height fields from 3D point clouds as a pre-processing step for automated damage detection. For this purpose, CAD or geometric models best estimated from the data are used. The automated damage detection based on this uses generic feature generators and outlier detection methods. The method achieved an F1 score of 96.3 % for a simulated data set and 72.4 % for a real data set.
- Using a deep learning approach, a method is being proposed to automatically detect damage that causes a colour change from images of port infrastructure. In this context, the suitability of automatically generated, weakly supervised label images for training the algorithms is investigated. In addition to the segmented segments, uncertainty values are predicted. This is evaluated using the example of photos of the JadeWeserPort in Wilhelmshaven. For the segmentation of the images into corrosion and undamaged areas, an F1 score of 98.0 % is achieved for undamaged areas and 71.7 % for corrosion damage.

The results of this work show that targeted contributions to automated damage detection can make the process of digital structural inspection more thorough, transparent and efficient. This helps to increase the safety and service life of infrastructure and reduce operating costs. The enhanced monitoring capabilities can be integrated into structural assessments, potentially extending the serviceable life of existing structures and consequently increasing the long-term productivity of ports. Furthermore, the optimized inspection processes enhance efficiency and cost-effectiveness, thereby contributing to the long-term sustainability and resilience of the infrastructure.

Abkürzungsverzeichnis

ASPP	Atrous Spatial Pyramid Pooling	10
BMDV	Bundesministerium für Digitales und Verkehr	1
BIM	Building Information Modeling	8
CAD	Computer-Aided Design	8
CNN	Convolutional Neural Network	3
CRF	Conditional Random Field	27
DHM	Digitales Höhenmodell	19
FN	False Negative	23
FP	False Positive	23
HED	Holistically-nested edges	38
ICP	Iterative Closest Point	15
IHLS	improved hue, saturation, lightness	28
IoU	Intersection over Union	4
k-MSS	kinematisches Multisensorsystem	11
LOF	Local Outlier Factor	9
MBES	Multibeam echosounder	5
MCC	Mathews correlation coefficient	23
MCG	Multiscale Combinatorial Grouping	37
MSS	Multisensorsystem	2
PCA	Principal Component Analysis	4
RGB	Rot, Grün, Blau	3
STRUM	Spatially Tuned Robust Multifeature	4
SVM	Support Vector Machine	4
TN	True Negative	23
TP	True Positive	23
UAS	Unmanned Aircraft Systems	4

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Stand der Technik	2
1.2.1	Passive Messverfahren zur Schadenserkennung	3
1.2.2	Aktive Messverfahren zur Schadenserkennung	5
1.3	Forschungsfragen	7
1.4	Workflow	8
1.5	Daten	11
1.6	Originalveröffentlichungen	16
2	Zusammenfassung der relevanten Publikationen	19
2.1	Automatisierte Erkennung von geometrischen Schäden	19
2.2	Automatisierte Erkennung von Farbänderungen	26
3	Weiterführende Betrachtungen	33
3.1	Höhenfelderstellung	33
3.1.1	3D-Gerade	34
3.1.2	Ebene	35
3.1.3	Gekrümmte Fläche	36
3.1.4	Vergleich der Verfahren der Höhenfelderstellung	36
3.2	Segmentierung mit schwachen Labeln	37
3.3	Unsicherheit der Segmentierung	39
4	Fazit und Ausblick	45
4.1	Fazit	45
4.2	Ausblick	47
5	Auflistung weiterer Publikationen	49
6	Eigene Veröffentlichungen	51
6.1	Publikation A	52
6.2	Publikation B	72
6.3	Publikation C	82
6.4	Publikation D	97
	Literaturverzeichnis	105
	Abbildungsverzeichnis	117
	Tabellenverzeichnis	119

1 Einleitung

1.1 Motivation

Eine moderne und leistungsfähige Verkehrsinfrastruktur ist maßgebend für eine wettbewerbsfähige Volkswirtschaft. Für den Erhalt des Bestandsnetzes und die Beseitigung von Engpässen sind im Bundesverkehrswegeplan 2033 (BMDV, 2016) 141,6 Mrd. Euro vorgesehen. Diese Mittel werden auf die drei Hauptverkehrswege – Bundesfernstraßen, das Schienennetz und die Bundeswasserstraßen – aufgeteilt. Erheblicher Erhaltungsbedarf besteht sowohl bei Brückenbauwerken als auch bei Verkehrswasserbauwerken wie Hafenanlagen und Schleusen. So sind von den 39.500 Brücken an Bundesfernstraßen ca. 4.000 Stück in marodem Zustand (BMDV, 2022b). Die Prüfungen von Verkehrsinfrastrukturbauwerken werden regelmäßig durch humansensorische Prüfungen durchgeführt. Dabei führen Prüfengeure visuelle und mechanische Überprüfungen der Bauwerksteile durch, was ein sehr zeit- und kostenintensives Verfahren darstellt. Dies verdeutlicht den enormen Bedarf an automatisierten Prüfungen von Verkehrsinfrastrukturen, um eine frühzeitige Erkennung von Schäden und deren Entwicklung kosteneffizient und reproduzierbar gewährleisten zu können. Im Folgenden wird dies am Beispiel von Verkehrswasserbauwerken dargestellt.

Im Jahr 2020 wurden in deutschen Seehäfen 275,7 Mio. Tonnen Güter umgeschlagen (Destatis, 2021). In der Seeverkehrsprognose 2030, die vom Bundesministerium für Digitales und Verkehr (BMDV) herausgegeben wurde, wird eine anhaltende Wachstumsrate von jährlich 4,3 % erwartet. Dadurch soll der Güterumschlag bis 2030 auf 468 Mio. Tonnen steigen (BMDV, 2014). Um diesen Wachstumsraten gerecht zu werden, sollen 16,2 Mrd. Euro für Erhaltungs- und Ersatzinvestitionen sowie 2,7 Mrd. Euro für Aus- und Neubau investiert werden. Aus diesem Grund ist es auch besonders wichtig, die Verfügbarkeit bestehender Infrastrukturen zu erhöhen. Nach BAW (2018, 2019) sind 80 % der Schleusen und Wehre älter als 50 Jahre und 30 % sogar älter als 100 Jahre, bei einer theoretischen Lebenserwartung von 100 Jahren. Von diesen bestehenden Infrastrukturen befinden sich nur etwa 70 % in ausreichendem Zustand. Dies bedeutet, dass in den kommenden Jahren immer mehr personelle und technische Ressourcen erforderlich sein werden, um den Betrieb solcher Infrastrukturen zu gewährleisten. Zusätzlich steigen die Anforderungen an ein sicheres und zuverlässiges Management der Hafeninfrastrukturen aufgrund des globalisierten Handels und des sich verändernden Schiffsverkehrs stetig. Diese Situation führt zu der Notwendigkeit, die Lebensdauer und somit die sichere Nutzbarkeit der bestehenden Hafeninfrastrukturen zu erhöhen. Durch eine längere Nutzbarkeit der Bestandsanlagen kann die Gesamtverfügbarkeit an Hafeninfrastuktur erhöht und somit die Wirtschaftlichkeit gesteigert werden. Dafür sind regelmäßige Überprüfungen des Bauwerkszustandes und frühzeitige Instandsetzungsarbeiten unabdingbar.

Bei Verkehrswasserbauwerken ist die Prüfung unter Wasser mit erheblich höherem Aufwand verbunden. Durch den Einsatz von Tauchern und die schlechte Sicht im Hafenbereich sind Unterwasserprüfungen und Schadensaufnahmen in Qualität und Quantität starken Variationen unterworfen. Aufgrund der subjektiven Wahrnehmung der Taucher sind die Schadensaufnahme und Klassifizierung bei wiederholten Prüfungen nicht vollständig reproduzierbar. Aufgrund des kosten- und zeitintensiven Prozesses wird zudem keine flächenhafte Prüfung des Unterwasserbereiches durchgeführt, sondern es wird stichprobenartig alle 50 - 100 m geprüft.

Grundlage für die Erhaltungsmaßnahmen ist dabei die frühzeitige Erkennung von strukturellen Bauwerksschäden. Je früher ein Schaden erkannt wird, desto geringer fallen in der Regel die Kosten für die Instandsetzung aus. Auch das Monitoring von vorhandenen Schäden spielt für die Erhaltung eine wichtige Rolle. Insbesondere Bauwerke, die seit vielen Jahren in Betrieb sind, müssen regelmäßig überprüft werden, da durch Dauerbelastungen und Materialermüdung Schäden entstehen können.

Dies ist nicht nur für die Verkehrssicherung relevant, sondern auch für die Industrie von großem Nutzen. Auch für ältere Verkehrswasserbauwerke ist es im Hinblick auf Industrie 4.0, den digitalen Zwilling und der fortschreitenden Digitalisierung notwendig, neue Methoden und Techniken des Lifecycle-Managements zu entwickeln und zu implementieren.

Automatisierte Verfahren sind hierbei notwendig, um Infrastrukturen kosteneffizient überprüfen zu können. Gerade bei Hafenanlagen kommt der zeitliche Aspekt erschwerend hinzu, da Sperrzeiten für Inspektionen durch Taucher zu erheblichen Verzögerungen im Logistikablauf und dadurch zu hohen Kosten führen. Daher sind auf Messdaten basierende automatisierte Verfahren notwendig, um sowohl kosten- als auch zeiteffizient Bauwerke zu überprüfen und Sperrzeiten zu vermeiden.

Diese automatisierten Verfahren bilden die Grundlage für die digitale Bauwerksprüfung, in der die automatisiert erkannten Schäden dem Prüflingenieur als Schadensverdachtsplan zur Verfügung gestellt werden. So muss nicht mehr die komplette Struktur humansensorisch überprüft werden, sondern nur die Bereiche, in denen durch die automatisierten Verfahren Schäden erkannt wurden. Dadurch kann der Zeiteinsatz für eine Bauwerksprüfung und die damit verbundenen Kosten erheblich gesenkt werden. Zusätzlich wird die Vollständigkeit der Prüfung gerade unter Wasser deutlich erhöht sowie die Reproduzierbarkeit der Schadenserkenkung gewährleistet. Durch die erhöhte Vollständigkeit der Bauwerksprüfung steigt im Mittel die zu erwartende Restnutzungsdauer. In Einzelfällen kann sie jedoch aufgrund starker Beschädigungen auch sinken. Im Rahmen dieser Dissertation werden Beiträge für eine digitale Bauwerksprüfung in Form von automatisierten Methoden zur Erkennung von Schädigungen aus Messdaten vorgestellt, die im Gesamtprozess der digitalen Bauwerksprüfung die Sicherheit und Effizienz der Verkehrsinfrastruktur erhöhen können. Die Forschungsfragen in Abschnitt 1.3 sind explizit auf die Herausforderungen der Schadenserkenkung ausgerichtet.

1.2 Stand der Technik

Technologische Entwicklungen und innovative Ansätze spielen eine Schlüsselrolle in der Bauwerksüberwachung. Dieser Abschnitt bietet einen Überblick über den aktuellen Stand der Technik. Aufgrund des hohen Sedimentgehaltes in Seehäfen ist eine visuelle Prüfung nicht möglich, weshalb die Strukturen per Abtasten überprüft werden. Dies führt zu subjektiven, nicht reproduzierbaren Ergebnissen und einem sehr zeit- und kostenintensiven Prozess (Hesse et al., 2019).

Zur Beschleunigung des Prozesses und zur Steigerung der Wirtschaftlichkeit, Häufigkeit und Vollständigkeit der Inspektionen werden in verschiedenen Forschungsprojekten und Veröffentlichungen neue Methoden und Verfahren untersucht. Einerseits erfolgt die Zustandserfassung durch bildgebende Verfahren oder andere Messtechnik, andererseits werden automatisierte Verfahren zur Auswertung und Schadenserkenkung eingesetzt. In der digitalen Bauwerksprüfung bezeichnet die Schadenserkenkung das automatisierte Auffinden von Schadstellen, jedoch ohne eine Einteilung in die Schadensart, oder -klasse vorzunehmen. Schadensklassifizierung bedeutet die Einteilung der erkannten Schäden in verschiedenen Schadensarten und -klassen, welche von den individuellen Charakteristiken der Schäden wie Größe, Tiefe, Form und den Materialeigenschaften abhängen. Für die Datenaufnahme werden häufig Multisensorsysteme (MSS) eingesetzt, die mit verschiedenen Sensoren zur Objekterfassung ausgestattet sein können. Im Folgenden wird zwischen passiven Verfahren, die auf Farbbildern basieren (siehe Abschnitt 1.2.1), und aktiven Messverfahren, die meist 3D-Punktwolken verwenden (siehe Abschnitt 1.2.2), unterschieden. Zu den aktiven Messverfahren gehören alle Messungen, welche ein Messsignal aktiv aussenden, dies können z. B. Laser, Schallwellen oder andere elektromagnetische Wellen sein. Passive Verfahren hingegen empfangen natürliche Signale wie z. B. reflektiertes Sonnenlicht. Diese Einteilung erfolgt, weil verschiedene Schadensarten von jeweils einem der Verfahren besser erkannt werden können. So eignen sich auf 3D-Punktwolken basierende Verfahren gut zur Erkennung von Schäden, welche eine geometrische Änderung hervorrufen, wie Abplatzungen, Anprallschäden oder große Risse. Zu den Nachteilen bei der Erfassung mit aktiven Verfahren zählt, dass in der Regel keine Farbinformationen vorhanden sind und somit

einige Schäden schlechter erkannt werden können. Schäden, die sich durch Farbänderungen zeigen, lassen sich gut mit Farbbildern aus passiven Messverfahren erkennen. Dazu gehören Aussinterungen, Korrosion sowie Bewuchs durch Pflanzen. Ein Nachteil passiver Verfahren ist, dass sie eine Beleuchtung erfordern. Unter Wasser im Hafengebiet können aufgrund des Sedimentgehaltes in der Regel keine Bilder aufgenommen werden.

1.2.1 Passive Messverfahren zur Schadenserkennung

Rot, Grün, Blau (RGB)-Kameras werden häufig zur Erkennung von Rissen in Beton oder anderen Materialien eingesetzt (Kim et al., 2017; Jiang & Zhang, 2020). Ein weiterer Ansatz ist die zusätzliche Verwendung von Thermalkameras zur Erfassung des Bauwerkszustandes (Schischmanow et al., 2022), insbesondere zur Detektion von Wärmebrücken und dem allgemeinen Isolationszustand von Bauwerken. RGB-Bilder eignen sich gut zur Erkennung von Schäden, die keine geometrische Änderung, sondern eine Farbänderung hervorrufen. Im Verkehrswasserbaubereich werden primär Aussinterungen und Roststellen erkannt, wobei verschiedene Publikationen sich mit dieser Thematik befassen. Liao & Lee (2016) stellen einen Algorithmus vor, der drei verschiedene automatische Ansätze beinhaltet: die K-Means-Methode in der Hue-Komponente, der DCDR-Algorithmus (Double-Center-Double-Radius) im RGB-Farbraum und DCDR im HSI-Farbraum (Hue-Saturation-Intensity). Ein auf Schwellwerten und Kantenerkennungsalgorithmen basierender Ansatz wurde von Diaz et al. (2017) in MATLAB implementiert und erreichte eine Erfolgsquote von 90 % bei der Rosterkennung auf Bildern. In den letzten Jahren gewinnen auch Verfahren des Maschinellen-Lernens (ML) zunehmend an Bedeutung. So stellen Jin Lim et al. (2021) ein Verfahren vor, bei dem sie zunächst die RGB- und Infrarotbilder (IR) in ein zweidimensionales Chroma-Blau- und Rot-Differenzbild (CbCr) mit einem zusätzlichen Infrarotkanal (CbCr-IR) umwandeln. Für die automatische Korrosionserkennung und -klassifizierung wird ein regionsbasiertes neuronales Netzwerk aufgebaut und anhand der kombinierten dreidimensionalen CbCr-IR-Bilder von Korrosionen trainiert. Um bei schlechter Beleuchtung bessere Ergebnisse erzielen zu können, schlagen Chen et al. (2012) einen Algorithmus vor, welcher Fourier-Transformationen und Support Vector Machines kombiniert, um eine effektive Methode zur Erkennung von ungleichmäßig beleuchteten Rostbildern zu entwickeln. Im Vergleich zum klassischen K-Means Ansatz erzielt die vorgeschlagene Methode bessere Ergebnisse bei ungleichmäßiger Beleuchtung. Pakrashi et al. (2010) schlagen eine Methode zur Erkennung und Quantifizierung von Korrosion auf Aluminiumoberflächen vor, die auf dem optischen Kontrast des korrodierten Bereichs im Vergleich zu den ihn umgebenden Bereichen beruht. Der korrodierte Bereich wurde mithilfe der Kantenerkennung identifiziert. Ghanta et al. (2011) war in der Lage, Schäden auf Bildern von stahlbeschichteten Brückenoberflächen zu erkennen, indem er eine einstufige Haar-Wavelet-Transformation auf ein RGB-Teilbild angewendet hat und Entropie und Energie im Wavelet-Bereich als Klassifizierungsmerkmale einsetzte. Jahanshahi & Masri (2013) verbesserten die Zuverlässigkeit der Farb- und Texturanalyse, indem sie die Auswirkungen verschiedener Parameter in auf Farb-Wavelets basierenden Texturanalyse-Algorithmen bewerteten. Es zeigte sich, dass die aus CbCr-Farbkanälen gewonnenen Wavelet-basierten Merkmale im Vergleich zu anderen Farbraumkanälen (z. B. Luminanz, Cyan-Orange Balance, Magenta-Grün Balance (YIQ), RGB, Buntwert, Sättigung, Intensität (HSI)) die beste Leistung erbringen. Choi & Kim (2005) unternahmen einen Versuch, die Art der Korrosion anhand der Morphologie der korrodierten Oberfläche zu erkennen. Eine Stichprobe von 197 Bildern wurde zum Trainieren und Testen des Klassifikators verwendet. Achtzehn Attribute, die Farbe, Textur und Form repräsentieren, wurden als Merkmale verwendet. Der Merkmalsraum wurde mittels multidimensionaler Skalierung auf zwei Dimensionen reduziert, und die Klassifizierung der Korrosionsarten wurde mit linearen Klassifikatoren durchgeführt. Zhao et al. (2018) untersuchten Convolutional Neural Network (CNN) für die Inspektion von Brücken. Für die Klassifizierung von Bauwerksschäden an Brücken wurde zunächst ein AlexNet-basiertes CNN mit mehr als 3800 Bildern verschiedener Brücken trainiert. Zur Erkennung von Brückenkomponenten wurde ein ZF-Net-basiertes schnelleres Region Based Convolutional Neural Network mit 600 Brückenbildern trainiert. Zur Erkennung von Rissen wurde ein

GoogleNet-basiertes CNN mit 60.000 Bildern von geschädigten und unbeschädigten Brücken trainiert. Xu et al. (2019) schlugen ein Fusions-CNN für die mehrstufige und mehrskalige Schadenserkenkung in Stahlkastenträgern vor, ohne dass zuvor Annahmen über die Rissgeometrie getroffen wurden. Die vorgeschlagene CNN-Architektur bestand aus mehreren Schichten von Faltung, batch normalization, ReLU, max pooling und Softmax, und wurde mit MatConvNet implementiert. Jedes Bild, das einen oder mehrere Risse, Graffiti und Hintergrundrauschen enthält, wurde mit einer Verbraucherkamera aufgenommen, die für Training und Validierung verwendet wurde. Zhang et al. (2017) schlugen ein CNN auf Pixelebene vor, um Risse auf 3D-Fahrbahnoberflächen zu erkennen. Das vorgeschlagene CNN, „CrackNet“, bestand aus zwei vollständig verknüpften Schichten, einer Faltungsschicht, einer $1 * 1$ -Faltungsschicht und einer Ausgabeschicht. Dieses Netzwerk war effizienter als herkömmliche CNNs, da es keine Pooling-Schichten gab, die die Ausgabe der vorherigen Schichten verkleinert hätten. Abdel-Qader et al. (2006) führten drei auf der Hauptkomponentenanalyse (Principal Component Analysis (PCA)) basierende Riss-Cluster-Methoden ein, darunter die ursprüngliche PCA-Methode, PCA mit linearen Merkmalen aus Rohdaten und PCA mit Merkmalen aus lokalen Regionen, um den Rissstatus von Bildabschnitten zu klassifizieren. Sie stellten fest, dass die dritte Methode die beste Gesamterkennungsgenauigkeit erzielte. Liu et al. (2002) extrahierten die Kanten von Bildern als Merkmale und importierten sie in die Support Vector Machine (SVM)-Trainingspipeline, um einen Rissklassifikator zu trainieren. Sie kategorisierten kleine Bildbereiche in drei Klassen: „Riss, Nicht-Riss und Zwischenbereich“. Der Klassifikator wurde auf das gesamte Bild angewendet, um die möglichen Rissregionen vorherzusagen. Prasanna et al. (2016) entwickelten einen Spatially Tuned Robust Multifeature (STRUM)-Klassifizierungsalgorithmus, um den Rissstatus in Bildfeldern zu klassifizieren. Das Eingangsmerkmal des vorgeschlagenen Klassifikators war der Riss-Merkmalvektor, der eine Kombination aus drei verschiedenen Merkmalen war, wie z. B. dem intensitätsbasierten Merkmal, dem gradientenbasierten Merkmal und dem Skalenraummerkmal. Für die grundlegende Architektur des STRUM-Klassifikators können SVM, AdaBoost und Random Forest ausgewählt werden. Sie verglichen die drei verschiedenen grundlegenden Klassifikatorarchitekturen und stellten fest, dass Random Forest bei dem ausgewählten Datensatz am besten abschnitt. Yamaguchi & Hashimoto (2010) extrahierten Risse aus großformatigem Beton mithilfe von Perkulationsverfahren. Die auf Perkulation basierenden Methoden umfassten eine Reihe lokaler Bildfilter- und Schwellenwert-Iterationsverfahren. Die auf Perkulation basierenden Methoden erkannten schließlich die Rissform, wenn die Kontrollparameter den vorgegebenen Schwellenwert für die Beendigung erreichten. In Alamouri et al. (2024) werden RGB-Bilder von Hafeninfrastrukturen zur Risserkennung genutzt. Dafür wird eine auf dem Deepcrack-Datensatz vortrainierte UNet-Architektur mit VGG19 Backbone für die Risserkennung verwendet, wobei ein Intersection over Union (IoU)-Score von 81,7 % erreicht wird. Morgenthal et al. (2019) stellen ein Framework vor, das den Zustand von Brücken anhand von Bildern bestimmt, die von Unmanned Aircraft Systems (UAS) aufgenommen wurden. Das Verfahren umfasst die automatisierte Flugplanung, Modellierung der Brücke und eine automatisierte Risserkennung.

Werden Stereokameras verwendet oder ein Bildverband aus überlappenden Bildern ausgewertet, können aus den Bildern auch 3D-Informationen wie Tiefenbilder oder Punktwolken erstellt und für die Schadenserkenkung verwendet werden. Helfrick et al. (2009) untersuchten ein digitales 3D Bild-Korrelation-Verfahren zur Erkennung von Schäden durch Krümmungsänderungen unter Verwendung eines Stereokamerapaars. Mit diesem Ansatz konnte die Lage unterschiedlich starker Schäden in Form eines Risses mit einer bestimmten Tiefe vorhergesagt werden. In Kim et al. (2018) wurde ein Ansatz zur Rissidentifizierung und -quantifizierung für Betonstrukturen mithilfe von UAS untersucht. Zur Erstellung einer Inspektionskarte wurde zunächst durch einen vorbereitenden Flug ein punktwolkenbasiertes Hintergrundmodell der Struktur erstellt. Aus hochauflösenden Bildern wurden Regionen mit CNN für tiefes Lernen zur Risserkennung verwendet. Eine Kombination aus 384 Rissbildern und einem durch Transferlernen trainiertes Modell wurde für die Klassifizierung und Lokalisierung von Schäden an einer Brücke verwendet. Die Erkennung und Volumenschätzung von Betonabplatzungen aus der 3D-Rekonstruktion von Bildern wurde in Beckmann et al. (2019)

vorgestellt. Dabei wird ein aus 1091 Bildern bestehender Datensatz mittels eines faster region-based convolutional neural networks segmentiert, wobei eine mittlere Präzision von 90,79 % erreicht wurde.

1.2.2 Aktive Messverfahren zur Schadenserkenkung

Eine weitere Möglichkeit, mit ungleichmäßiger oder fehlender Beleuchtung und geringer Oberflächenstruktur umzugehen, sind aktive Messverfahren. So können über Wasser z. B. Laserscanner und unter Wasser Multibeam echosounder (MBES) eingesetzt werden, welche in der Regel 3D-Punktwolken erzeugen. Beshr et al. (2022) scannen eine große Betonfläche mittels terrestrischem Laserscanning (TLS), um Defekte in den Fahrbahnplatten zu erkennen. Die erfasste Punktwolke wurde in neun Teilwolken zerlegt, um jede Platte darzustellen, und die Defekte wurden anhand von Ebenenparametern charakterisiert. Ein Verfahren zur Messung der Neigung und der seitlichen Verschiebung von Stützmauern mithilfe von mobilem Laserscanning wurde von Kalenjuk et al. (2021) vorgestellt. Aus der erfassten Punktwolke wurde die Stützmauer durch eine binäre Klassifikation aus den Bodenpunkten extrahiert. Im Gegensatz zur Berechnung von Punkt-zu-Punkt-Abweichungen wurden die verankerten Betonplatten einzeln segmentiert und mit ebenen Flächen modelliert. Darüber hinaus zeigten Änderungen der Schlüsselparameter der Ebene Verformungen zwischen zwei Epochen an. In Suchocki et al. (2021) wurde ein Verfahren zur Verringerung der Größe von Punktwolkendatensätzen bei gleichzeitiger Erkennung von Oberflächenschäden (z. B. Risse und Hohlräume) und physikalisch-chemischen Schäden (z. B. Feuchtigkeit, Verwitterung, Salzausblühungen) in Gebäudewänden untersucht. Ein automatisiertes Verfahren zur Erkennung des Massenverlustes von Betonbrücken mithilfe von Laserscans wurde von Teza et al. (2009) beschrieben. Die Schadenserkenkung wurde in drei Schritten durchgeführt: (1) Die Punktwolken wurden in Teilbereiche mit einer bestimmten Größe unterteilt, (2) für jeden Punkt in jedem Teilbereich wurde eine vorläufige Gauß-Filterung und eine parabolische Anpassung durchgeführt und (3) jeder Teilbereich wurde anhand der entsprechenden Krümmungsverteilung als beschädigt oder unbeschädigt klassifiziert. Chen et al. (2013) verwendeten 3D-Punktwolken eines terrestrischen Laserscanners sowie einfache Differenzalgorithmen zur Identifizierung verschiedener Schadensmerkmale, wie z. B. Betonverlust, Bewehrungskorrosion und Oberflächenerosion, an Brücken. Ein vollautomatisches Punktwolken-Segmentierungsverfahren wurde von Riveiro et al. (2016) für die Bauwerksprüfung von gemauerten Bogenbrücken entwickelt. Durch einen Voxelisierungsprozess wurden zunächst die redundanten Daten aus der Punktwolke herausgefiltert, woraufhin die Segmentierung durch die Kombination einer heuristischen Methode mit Bildverarbeitung erfolgte. Zur Identifizierung der einzelnen Strukturelemente wurden topologische Beschränkungen verwendet, um die räumliche Beziehung und die Reihenfolge der Elemente zu bestimmen. Fünf Datensätze wurden verwendet, um den vorgeschlagenen Algorithmus zu testen, und zeigten übereinstimmende Ergebnisse. Ein automatisiertes Verfahren zur Verarbeitung von 3D-Punktwolken für die Überprüfung von Pfeilern in gemauerten Bogenbrücken wurde von Sánchez-Rodríguez et al. (2018) vorgeschlagen. Eine Brücke wurde zunächst anhand ihrer Strukturelemente (unter Wasser, Pfeiler, Brüstungsmauern, Fahrbahn und Gewölbe) segmentiert, sodass die Datenmenge reduziert wurde und jede Pfeilerfläche einzeln analysiert werden konnte. Strukturelle Fehler wurden anhand der geometrischen Parameter der Pfeilerflächen und ihrer jeweiligen topologischen Beziehung zu den anderen Brückenelementen ermittelt. In Yang & Xu (2020) wurde eine Methode zur Risserkennung in Tunneln vorgestellt, die Dilatation und den Canny-Algorithmus auf der Grundlage von TLS-Punktwolken kombiniert. Der Graustufen-Dilatationsprozess wurde eingesetzt, um unterschiedliche Texturen in jedem Bild mithilfe eines scheibenförmigen Strukturierungselements zu eliminieren. Der Canny-Detektor wurde für die Risskantenerkennung eingesetzt. Bei der vorgeschlagenen Methode mussten die Parameter des Canny-Detektors nicht bestimmt werden, sodass kein größerer Benutzereingriff erforderlich war. Für die Erkennung und Kartierung von Rissen in Beton aus 3D-Punktwolken wurde eine automatisierte Methode von Rabah et al. (2013) implementiert. Um das Rauschen zu entfernen, wurden zunächst verschiedene Filter- und Bildverarbeitungsprozesse durchgeführt. Die probabilistische Re-

laxationstechnik wurde angewendet, um die Rissspur aus dem gefilterten Bild zu extrahieren und die Pixelkoordinaten wurden in globale 3D-Koordinaten transformiert. Eine Technik zur gleichzeitigen Lokalisierung und Quantifizierung von Abplatzungen auf Betonoberflächen aus 3D-Punktwolken wurde von Kim et al. (2015) entwickelt. Zwei empfindliche Merkmale für Schäden (Winkel- und Abstandsabweichung) wurden kombiniert, um Schäden zu identifizieren. Ein Schadensklassifikator wurde entwickelt, um das Ausmaß, die Lage und die Größe der Schäden zu beurteilen. Eine Reihe von Scan-Parametern (z. B. Scan-Abstand, Winkelauflösung, Auftreffwinkel) wurden an simulierten und realen Daten eines Testobjektes untersucht. Die Ergebnisse zeigten, dass die vorgeschlagene Technik die Automatisierung und Genauigkeit der Erkennung von Betonschäden verbessert. Eine auf Oberflächennormalen basierende Schadenserennungsmethode wurde von Guldur Erkal & Hajjar (2020) zur Erkennung und Quantifizierung von Schadensarten wie Rissen, Abplatzungen, Korrosion, Delaminierung und Bruch mit Hilfe 3D-Punktwolken implementiert. Die schadhaften Bereiche wurden anhand der Modelleigenschaften lokalisiert und mit einer silhouettenbasierten Methode in einzelne Schadenscluster gruppiert. Die quantitativen Informationen der Schadenscluster wurden mit den vorgestellten Strategien zur Berechnung von Schadensflächen und -volumen erfasst. Ein neuartiger 3D-Oberflächendeskriptor wurde von Madrigal et al. (2017) für die automatische Identifizierung und Klassifizierung von Oberflächendefekten unter Verwendung von Punktwolken entwickelt. Die lokalen Modelle auf jeder 3D-Oberfläche wurden zunächst geschätzt und dann zur Berechnung der Unterschiede zwischen den Modellnormalen in einer lokalen Region verwendet. Ein auf Machine-Learning basierender Vorprozessierungsansatz wurde von Zhao et al. (2022) vorgestellt. Dabei wird der CANUPO-Algorithmus (Brodu & Lague, 2012) zur automatischen Bereinigung der Punktwolken verwendet, um anschließend Veränderungen mittels Punktwolkenvergleichen zu detektieren. Xia et al. (2022) stellen einen Ansatz zur automatischen, semantischen Segmentierung von 3D-Punktwolken am Beispiel von Brücken vor. Dabei verwenden sie zum Training eines Deep Convolutional Networks einen lokalen Descriptor. Dabei erreichen sie eine durchschnittliche Präzision von 97,26 %, einen Recall von 98,00 % und eine IoU von 95,38 %, was PointNet deutlich übertrifft. Ein Nachteil von Machine Learning Ansätzen ist, dass eine große Menge an Trainingsdaten benötigt wird. Eine effiziente Möglichkeit die Menge an Trainingsdaten zu reduzieren ist die Verwendung von vortrainierten Netzwerken. Mithilfe von Transfer-Learning-Techniken (TL) kann die Trainingszeit minimiert werden, indem die Koeffizienten aus dem Basismodell übertragen werden, anstatt mit zufällig zugewiesenen Gewichten zu beginnen. Vortrainierte Modelle können sogar für neue Probleme mit völlig anderen Ausgangsklassen verwendet werden. Bei der Erkennung von Fahrbahnrisse hat sich TL als effizienter Ansatz zur Verbesserung der Genauigkeit des Klassifizierungsproblems erwiesen (Zhang et al., 2018). Gopalakrishnan et al. (2017) waren in der Lage, mithilfe der TL-Technik Risse in Asphalt und Beton zu erkennen. Sie verwendeten das VGG16-Netzwerk, das auf ImageNet-Daten trainiert wurde.

Im Kontext dieser vielfältigen Ansätze und Techniken fokussieren sich viele Arbeiten auf spezifische Bauwerkstypen oder Schadensarten und beschränken sich häufig auf Daten, die über Wasser aufgenommen wurden. Diese Verfahren bieten jedoch nur begrenzte Lösungen für die umfassende Untersuchung von Wasserbauwerken, insbesondere wenn es darum geht, sowohl über als auch unter Wasser liegende Strukturen zu analysieren.

Hier setzt die vorliegende Dissertation an: Sie zielt darauf ab, diese Lücken zu schließen, indem sie innovative Ansätze zur hybriden digitalen Bauwerksprüfung entwickelt und integriert. Ein zentrales Anliegen dieser Dissertation ist es, die Verknüpfung verschiedener Datenquellen sowohl über als auch unter Wasser zu erforschen und automatisierte Schadenserennungsmethoden für eine digitale Bauwerksprüfung zu entwickeln. Dabei werden die häufigsten Schäden an Verkehrswasserbauwerken erkannt und analysiert. Diese neuen Methoden versprechen im Kontext des Gesamttablaufs, eine effizientere, vollständigere, objektivere und besser wiederholbare Bauwerksprüfung zu ermöglichen und stellen somit eine bedeutende Neuerung in diesem Feld dar. Im folgenden Abschnitt 1.3 werden die zentralen Forschungsfragen formuliert, die sich aus den identifizierten Problemstellungen ergeben und die Grundlage für die Entwicklungen in dieser Dissertation bilden.

1.3 Forschungsfragen

Diese Arbeit konzentriert sich auf die Erkennung struktureller Schäden an Verkehrswasserbauwerken anhand von Punktwolken und RGB-Bildern. Die Herausforderung bei der Verwendung von Punktwolkendaten besteht darin, die Punktwolken in zwei Klassen zu unterteilen: beschädigt und unbeschädigt. Aufgrund der überwiegenden Anzahl unbeschädigter Bereiche sind die Klassen ungleich verteilt. Außerdem ist die Menge der verfügbaren Trainingsdaten mit echten Labels begrenzt, sodass vortrainierte Netzwerke als generische Merkmalsgeneratoren verwendet werden müssen. Vor der Anwendung der Algorithmen zur Erkennung von Anomalien müssen die Punktwolken angemessen vorverarbeitet werden, damit sie für die Analyse geeignet sind. Daraus ergeben sich die folgenden Forschungsfragen:

- Forschungsfrage 1: Wie können Punktwolken von unregelmäßigen Hafengebäuden bestmöglich für die digitale Bauwerksprüfung vorprozessiert und durch automatisierte Verfahren auf Schädigungen überprüft werden?
- Forschungsfrage 2: Können Schäden automatisiert aus Farbbildern strukturierter Oberflächen von Hafengebäuden erkannt werden und kann die Anzahl an manuell erzeugten Labels für das Training signifikant reduziert werden?

Abbildung 1.1 gibt einen Überblick, wie die in dieser Arbeit verwendeten Publikationen A-D (siehe Abschnitte 2.1 und 2.2) sowie die weiterführenden Betrachtungen (siehe Abschnitt 3) in die dargestellten Forschungsfragen einfließen. Publikation A und B (Abschnitt 2.1) sowie die weiterführenden Betrachtungen (Abschnitt 3.1) bilden die Grundlage für die erste Forschungsfrage. Die zweite Forschungsfrage wird in Publikation C und D (Abschnitt 2.2) sowie in den weiterführenden Betrachtungen (Abschnitte 3.2 und 3.3) behandelt.

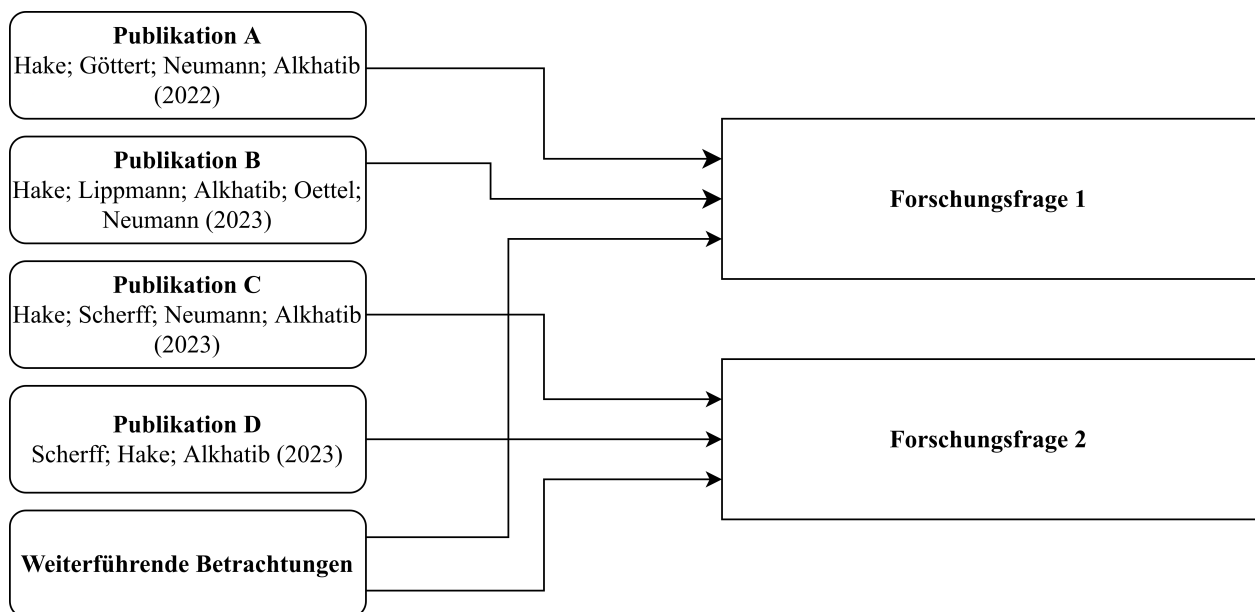


Abbildung 1.1: Übersicht der Originalveröffentlichungen zur kumulativen Dissertation

Im folgenden Abschnitt 1.4 werden ausgehend von den in den beiden Forschungsfragen verwendeten Datentypen zwei allgemeine Workflows zur Erkennung von Schäden vorgestellt. Dabei wird der Gesamtprozess von der Vorprozessierung der Rohdaten, über das Training und die Evaluierung allgemein beschrieben.

1.4 Workflow

In diesem Abschnitt wird ein allgemeiner Prozess zur Schadenserkenkung vorgestellt, der im weiteren Verlauf dieser Arbeit auf spezifische Schäden angewendet wird.

Basierend auf den beiden Forschungsfragen werden zwei getrennte Abläufe für die unterschiedlichen Datentypen 3D-Punktwolke und Farbbild verwendet. Dies liegt zum einen an der stark unterschiedlichen Struktur und Eigenschaften der Daten, aber auch an den Schadenstypen, die primär durch die Charakteristik der Daten erkannt werden können. So lassen sich geometrische Schäden deutlich besser durch 3D-Punktwolken erkennen, da hier die Geometrie direkt gemessen wird. Schäden, die eine Änderung der Farbe hervorrufen, können sehr gut mit Farbbildern erkannt werden.

Was beide Workflows gemein haben, ist das unausgewogene Verhältnis von Schäden und nicht beschädigten Bereichen in den Daten. Dies ist herausfordernd für die Segmentierung der Infrastrukturen und der daraus folgenden Erstellung von Schadensverdachtsplänen. Um diesem Problem zu begegnen, werden Standardverfahren des maschinellen Lernens für unbalancierte Klassen eingesetzt. Dazu zählen Anomaliedetektionsverfahren (Chandola et al., 2009) (siehe auch Abschnitt 2.1), die Verwendung von Klassengewichten (höhere Gewichtung der unterrepräsentierten Klasse (Megahed et al., 2021; Bishop, 2006), siehe auch Abschnitt 2.2), sowie angepasste Verlustfunktionen und Data Augmentation Techniken (Shorten & Khoshgoftaar, 2019) (siehe auch Abschnitte 2.1 und 2.2).

Die Punktwolken müssen zunächst vorverarbeitet werden, da sie im Originalzustand unstrukturiert sind und aufgrund ihrer Größe schwierig automatisiert zu verarbeiten sind. Daher wird ein Ansatz gewählt, bei dem Distanzen zwischen der Punktwolke und einem Modell des Objektes berechnet und verwendet werden. Dazu wird die orthogonale Distanz zwischen jedem Punkt und der nächstgelegenen Fläche im Objekt bestimmt. Diese Distanzen lassen sich in einem Distanzfeld darstellen (siehe auch Abschnitte 2.1 und 3.1). Im weiteren Schritt werden die Distanzfelder gerastert, das heißt die Distanzen in feste Rasterzellen unterteilt. Dabei werden mehrere Distanzen in einer Rasterzelle gemittelt und fehlende Zellen aus den Nachbarn interpoliert. Das gerasterte Distanzfeld lässt sich somit als Grauwertbild speichern, wobei die Grauwerte den Distanzen entsprechen. Je heller der Grauwert, desto größer ist die Distanz. Solche Distanzfelder werden auch Höhenfeld oder digitales Oberflächenmodell (Skidmore, 1989) genannt. Für das Modell des Objektes gibt es verschiedene Möglichkeiten: Wenn ein bestehendes Computer-Aided Design (CAD)- oder Building Information Modeling (BIM)-Modell von dem Bauwerk vorhanden ist, können die Abstände zwischen Modell und Punktwolke direkt für jeden Punkt berechnet und somit das Höhenfeld erstellt werden. Liegt ein solches Modell nicht vor, muss das Modell aus den Daten erstellt werden. Hierzu können zum Beispiel Geometrien wie quadratische Formen oder im einfachsten Fall Ebenen verwendet werden. Die Punktwolke wird dafür in Segmente unterteilt und für jedes Segment eine Fläche nach Drixler (1993) geschätzt. Die ausgeschnittenen Segmente überlappen sich jeweils um 50 %. Damit soll dem Problem entgegengewirkt werden, dass Schäden schwieriger zu erkennen sind, wenn sie direkt am Rand der Segmente liegen oder zerschnitten sind. Die Distanzen werden dann zwischen der Punktwolke und den geschätzten Flächen bestimmt. Im Ergebnis erzeugen beide Methoden ein geometriebereinigtes Höhenfeld.

Aus den so erstellten Höhenfeldern werden mittels eines generischen Merkmalsgenerators automatisiert Merkmale erzeugt, die für die Erkennung von Schäden verwendet werden. Als Merkmalsgenerator wird ein Deep-Learning-Netzwerk verwendet. Dies benötigt eine große Menge an qualitativ hochwertigen, annotierten Daten. Da Schäden an Hafenanlagen bisher nicht ausreichend durch Messtechnik erfasst wurden, dauert es sehr lange und erfordert viele Messungen, bis eine ausreichende Menge an annotierten Daten verfügbar ist. Um dieses Problem zu lösen, wird eine verkürzte Version des VGG19-Netzwerks als Grundgerüst für die Merkmalsextraktion verwendet (siehe auch Abschnitt 2.1). Das neuronale Netz VGG19 ist ein Standard-CNN, das auf natürlichen Bildern vortrainiert wurde (Wan et al., 2020; Simonyan & Zisserman, 2014). In der Literatur wird das VGG19 häufig in verschiedenen Modifikationen zur Bildsegmentierung, Schadenserkenkung oder dem industriellen Monitoring verwendet (Khan et al., 2021; Apostolopoulos & Tzani, 2022; Meena

et al., 2022; Akula & Sardana, 2019; Wan et al., 2021). Aufgrund der sehr guten Ergebnisse in der Bildsegmentierung und der Übertragung der vortrainierten Gewichte auf andere Bereiche wurde in dieser Arbeit das VGG19-Netzwerk für den weiteren Ablauf ausgewählt. Das Netz besteht aus 19 Schichten und wurde speziell für Klassifizierungsaufgaben trainiert. Es ist dafür bekannt, dass es sehr gute Leistungen in der umfangreichen Bilddatenbank ImageNet (Deng et al., 2009) erzielt. Da sich die Bilder des ImageNet Datensatzes (natürliche Gegenstände wie Tiere, Häuser, Autos, etc.) stark von den Schäden im Hafengebiete unterscheiden, wird nicht die originäre Netzstruktur verwendet, sondern eine modifizierte Version. Konkret werden nur die ersten Faltungsschichten des Netzes verwendet. Dies soll eine Überanpassung des Netzes verhindern, da die ersten Schichten eher grundlegendere Merkmale wie zum Beispiel Kanten lernen und die tieferen Schichten höherwertige Merkmale wie Gesichter, Autos, etc. Dabei werden die am ImageNet Datensatz vortrainierten Parameter des VGG19-Netzwerkes auf den Datensatz von Hafengebiete übertragen. Für jede Ausgabe wird die Grammatrix berechnet (Gatys et al., 2015), dabei wird nur die Diagonale der Matrix betrachtet, die sich auf die Stärke eines Merkmals bezieht. Aus diesem Merkmalsvektor wird mittels Local Outlier Factor (LOF) (Breunig et al., 2000) die Klassifizierung in geschädigte und unbeschädigte Bereiche erreicht. LOF ist in der Lage, Ausreißer in Daten zu erkennen. Unter Ausreißern wird in dieser Dissertation nicht wie in der Statistik üblich ein Datenpunkt verstanden, der signifikant vom Normalwert abweicht, sondern Datenpunkte, welche nicht den erwarteten Mustern und Verhaltensweisen entsprechen. In diesem Fall handelt es sich um die Erkennung von Anomalien, bei der es darum geht, seltene Ereignisse oder Datenpunkte zu finden, also Datenpunkte, die erheblich von der Mehrheit der Daten abweichen und nicht einer genau definierten Vorstellung von normalem Verhalten entsprechen (Chandola et al., 2009; Barnett & Lewis, 1994; Hawkins, 1980). Um Ausreißer zu erkennen, konstruiert die LOF-Methode einen Graphen im Merkmalsraum, um die Dichte der Nachbarschaft zu schätzen. Anschließend wird aus dieser Dichte ein Ausreißer-Score für jeden Datenpunkt berechnet. Der gesamte Prozess ist in MATLAB und Python implementiert und in Abbildung 1.2 in Form eines Ablaufdiagramms zusammengefasst.

Die relevanten Parameter, welche für dieses Verfahren gewählt werden müssen, reduzieren sich aufgrund der adaptierten Version des VGG19-Netzes im Wesentlichen auf die Größe der Eingabebilder. Diese hängt von der Auflösung der Punktwolke und der damit verbundenen räumlichen Größe einer Rasterzelle ab. Bei einer sehr kleinen Rasterweite können größere Patches als Eingabebilder verwendet werden. Die minimale Größe eines Eingabebildes beträgt 32 x 32 Pixel.

Im Folgenden werden die wichtigsten Schritte im Bearbeitungsablauf für Farbbilder dargestellt. Dieser unterscheidet sich aufgrund des unterschiedlichen Datentyps und der zu erkennenden Schäden von dem Ablauf der Punktwolkenauswertung.

Im ersten Schritt nach der Datenerfassung werden zunächst in einem Teildatensatz manuell Schäden markiert. Die so entstehenden Labelbilder sind für die spätere Evaluation des Verfahrens notwendig. Der Teildatensatz ist dabei über den gesamten Bereich verteilt, um eine möglichst gleichmäßige Abdeckung zu erhalten. Im Ergebnis entstehen schwarz-weiße Bilder, wobei weiß für die Klasse Schaden und schwarz für unbeschädigte Bereiche steht (siehe auch Abschnitt 2.2). Ein Beispiel für den Bilddatensatz aus dem Jade-Weser-Port in Wilhelmshaven mit markierten Korrosionsschäden ist in Abbildung 1.3 dargestellt. Korrosion ist hierbei nur beispielhaft für die Schadensart, das Verfahren funktioniert für alle Schadensarten, welche eine Änderung der Oberflächenfarbe hervorrufen.

Für die automatische Segmentierung in beschädigte und unbeschädigte Bereiche wird eine modifizierte Variante des DeepLabV3+ (Chen et al., 2018) verwendet. Das DeepLabV3+ ist ein in der Literatur sehr bekanntes und häufig eingesetztes Netzwerk zur semantischen Segmentierung von Bildern insbesondere zur Erkennung von Korrosion (Wang et al., 2023; Lin et al., 2023; Pan & Zhang, 2022; Zeng et al., 2023; Hartz et al., 2022; Garcia et al., 2021). Dabei wurden sehr gute Ergebnisse in der Segmentierung von Korrosion erzielt, weshalb es in dieser Arbeit ausgewählt wurde. Beim DeepLabV3+-Netz handelt es sich um eine Encoder-Decoder-Netzwerkstruktur. Der Encoder extrahiert Merkmale aus dem Bild und der Decoder rekonstruiert das Bild, indem er jedem Pixel Klassenwahrscheinlichkeiten zuordnet. Um es an die spezifische Aufgabe der Schadensdetektion in

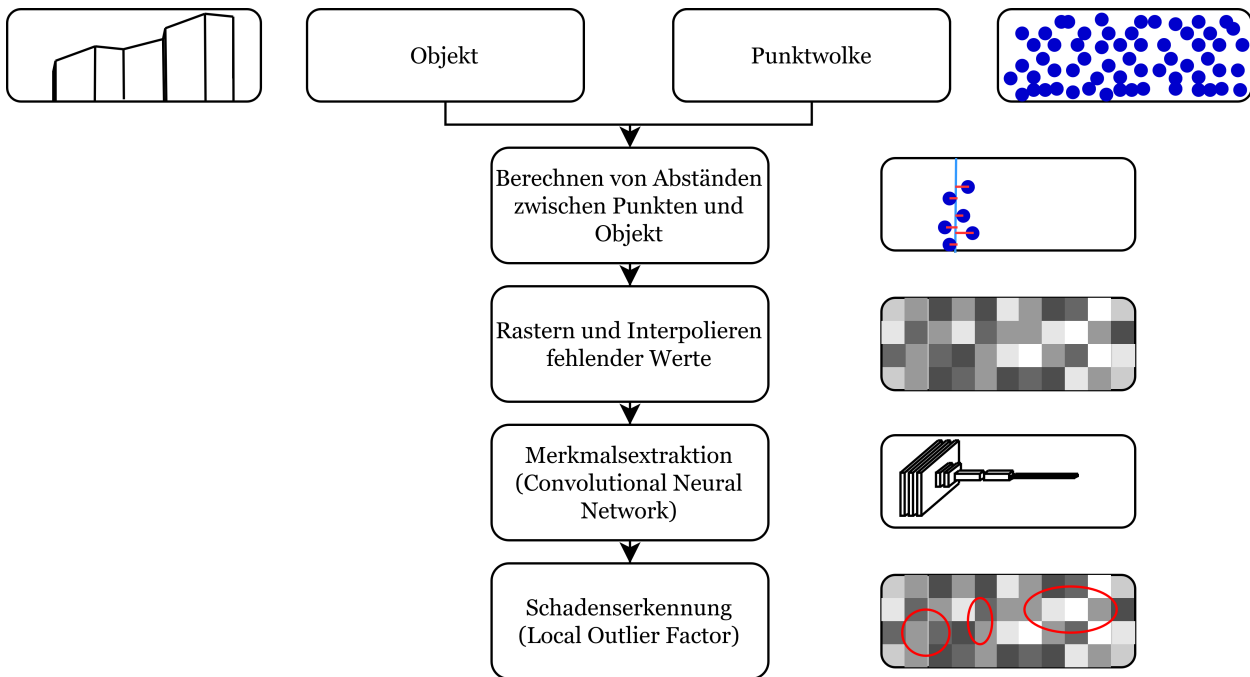


Abbildung 1.2: Ablaufdiagramm des automatischen Schadenserkennungsprozesses aus 3D-Punktwolken. Ausgehend von der originären 3D-Punktwolke und dem mathematischen Modell der Struktur werden zunächst Punkt zu Objekt Distanzen für jeden Punkt berechnet. Diese Distanzen werden in ein Höhenfeld konvertiert, indem sie gerastert und als Grauwertbild gespeichert werden. Mit einem CNN als generischen Merkmalsgenerator und folgendem Local Outlier Factor Algorithmus können Schäden in dem Höhenfeld erkannt werden.



Abbildung 1.3: Ausgeschnittenes RGB-Bild (links) und erstelltes Labelbild (rechts) des Jade-Weser-Ports in Wilhelmshaven. Die Klassen unbeschädigt und Korrosion sind in Schwarz bzw. Weiß dargestellt (Hake et al., 2023b).

den Bildern anzupassen, wurden die Struktur und einige Parameter des Netzes geändert (siehe auch Abschnitt 2.2). Eine dieser Änderungen ist die Verdoppelung der Dilatationsparameter innerhalb des Atrous Spatial Pyramid Pooling (ASPP) Blocks. Um die Genauigkeit des DeepLabV3+ Netzwerks zu verbessern, werden die Eingabebilder in kleinere Bereiche von 512×512 Pixel aufgeteilt. Dies wird gemacht, da das Netzwerk nicht in der Lage ist, größere Bilder effektiv zu nutzen. Die einzelnen Bereiche überlappen sich jeweils um 25 %. Die Überlappung beugt abgeschnittenen Schäden an den Rändern der Bereiche vor und hat in verschiedenen Tests zu den besten Metriken geführt. Die erhaltenen segmentierten Bilder werden durch nahtloses Zusammenfügen der kleineren Bereiche erstellt (Chollet, 2017). Die beiden Klassen Schaden und kein Schaden treten in stark unterschiedlicher Anzahl im Datensatz auf, das heißt, es gibt deutlich mehr unbeschädigte Bereiche

als Schäden. Um dieses Ungleichgewicht zu berücksichtigen, werden Klassengewichte in den Fehlerfunktionen verwendet. Dabei wird die Klasse Schaden mit dem Gewicht von 5 versehen, da etwa fünfmal mehr Bereiche unbeschädigt sind als Schäden vorliegen. Eine weitere Anpassung der DeepLabV3+ Architektur besteht darin, ein anderes CNN als Backbone-Netzwerk zu verwenden und eine andere Fehlerfunktion. Hierbei wurde ResNet-50 (He et al., 2016) mit der IoU-Fehlerfunktion verwendet, da es im Vergleich zu anderen Architekturen die besten Ergebnisse erzielt hat (siehe Abschnitt 2.2).

Der gesamte Prozess ist in MATLAB und Python implementiert und in Abbildung 1.4 in Form eines Ablaufdiagramms zusammengefasst. Ausgehend von der Datenerfassung wird zunächst ein Teildatensatz mit manuellen Labeln für die Optimierung versehen und der Gesamtdatensatz in Train- und Testset aufgeteilt. Nach der Auswahl des Modells wird das Netz trainiert und mit den manuell erzeugten Labeln des Trainingssets optimiert. Im letzten Schritt folgt die Evaluierung anhand des Testsets.

Die wichtigsten Parameter für diese Methode sind in folgender Tabelle 1.1 zusammengefasst:

Tabelle 1.1: *Parameter der DeepLabV3+ Architektur.*

Eingabegröße	512x512 Pixel
Gewichte	Initialisiert auf ImageNet
Batchgröße	12
Epochen	30
Klassengewichte	(1:5)
Verlustfunktion	IoU-Loss

In diesem Abschnitt wurde ein allgemeiner, aus den Forschungsfragen abgeleiteter Workflow für die beiden verwendeten Datentypen 3D-Punktwolke sowie Farbbilder vorgestellt. Dabei wurde bisher nicht auf konkrete Beispiele und Daten eingegangen. Im folgenden Abschnitt 1.5 werden die dieser Dissertation zugrunde liegenden Datensätze vorgestellt und beschrieben. Dabei wird auch auf die spezifischen Schäden eingegangen, welche durch die Daten erkannt werden können.

1.5 Daten

Die in dieser Arbeit verwendeten Analysen und Methoden wurden so ausgelegt, dass die Daten mit dem im Verbundprojekt „3DHydroMapper – Bestandsdatenerfassung und modellgestützte Prüfung von Verkehrswasserbauwerken“ entwickelten kinematisches Multisensorsystem (k-MSS) erfasst werden können. (Hesse et al., 2019). Das Projekt mit dem Förderkennzeichen 19H18011C bestand aus fünf Partnern und einem assoziierten Partner: Dr. Hesse und Partner Ingenieure (Multisensorsystem und kinematisches Laserscanning), WK Consult (Bauwerksinspektion, BIM und Instandhaltungsplanung), Niedersachsen Ports (See- und Binnenhafenbetrieb), Fraunhofer IGP (automatische Modellierung und BIM), Leibniz Universität Hannover (Routenplanung und Schadenserkennung) und die Wasserstraßen- und Schifffahrtsverwaltung des Bundes (Verwaltung der Bundeswasserstraßen) als assoziierter Partner. Die Laufzeit des Projektes war vom 1. Dezember 2018 bis zum 31. Mai 2022.

Bei der Messplattform handelt es sich um ein mit verschiedener Sensorik ausgestattetes Boot, das sowohl während als auch nach der Projektlaufzeit kontinuierlich weiterentwickelt wurde (Abbildung 1.5 & 1.6). Das Boot wurde so konstruiert, dass es möglichst langsam aber voll manövrierfähig ein Hafengebäude entlang fahren kann. Dafür wurden neben dem Hauptantrieb Querstrahlsteueranlagen verbaut, um auch bei langsamen Geschwindigkeiten die Ausrichtung und den Abstand zum Messobjekt präzise einhalten zu können.

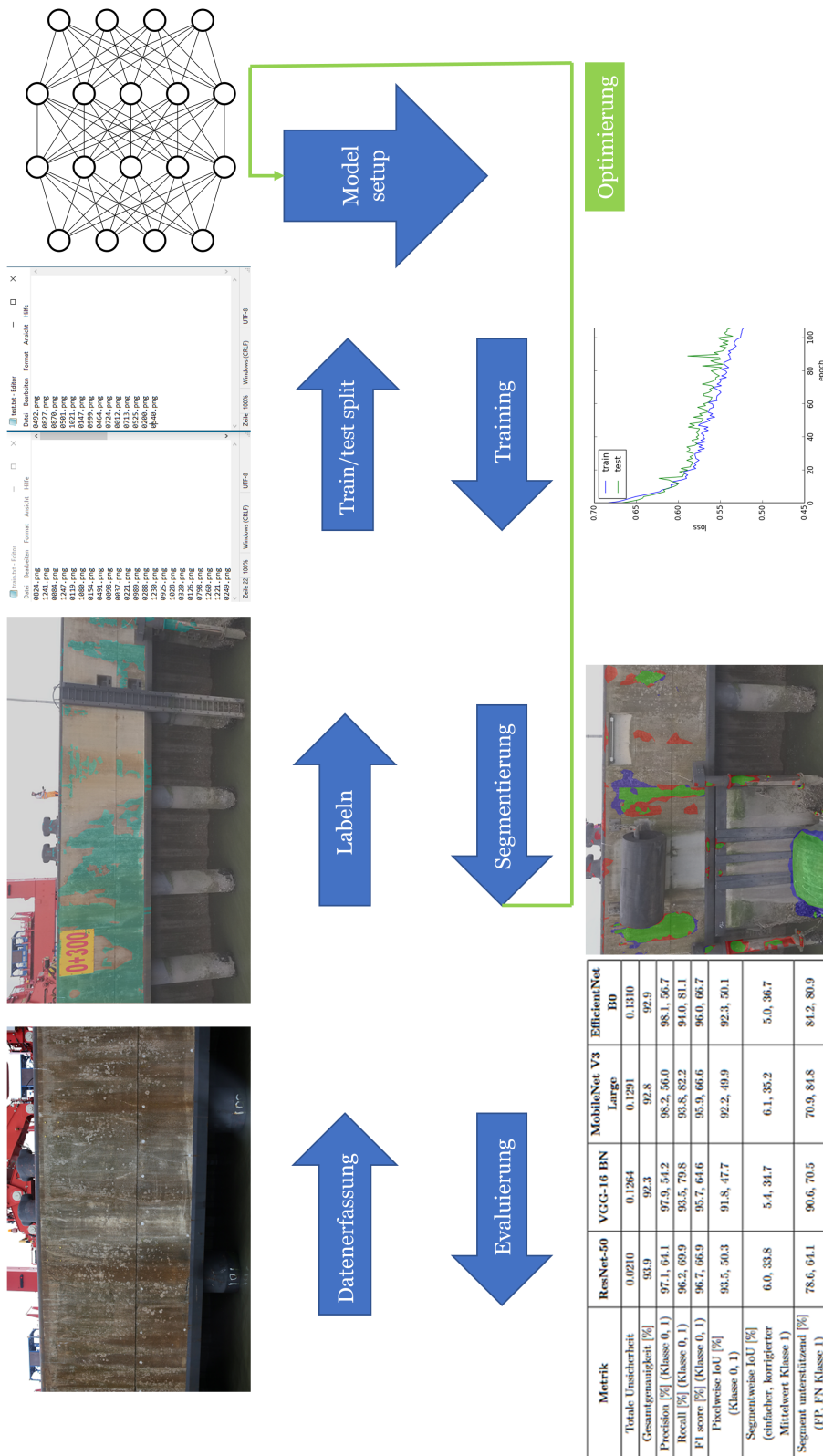


Abbildung 1.4: Ablaufdiagramm des automatischen Schadenserkennungsprozesses aus Farbbildern. Ausgehend von der Datenerfassung wird zunächst ein Teildatensatz mit manuellen Labels für die Optimierung versehen und der Gesamtdatensatz in Trainings- und Testdatensatz aufgeteilt. Nach der Auswahl des Modells wird das Netz trainiert und mit den manuell erzeugten Labels des Trainingsdatensatzes optimiert. Im letzten Schritt folgt die Evaluierung anhand des Testdatensatzes (Hake et al., 2023b).

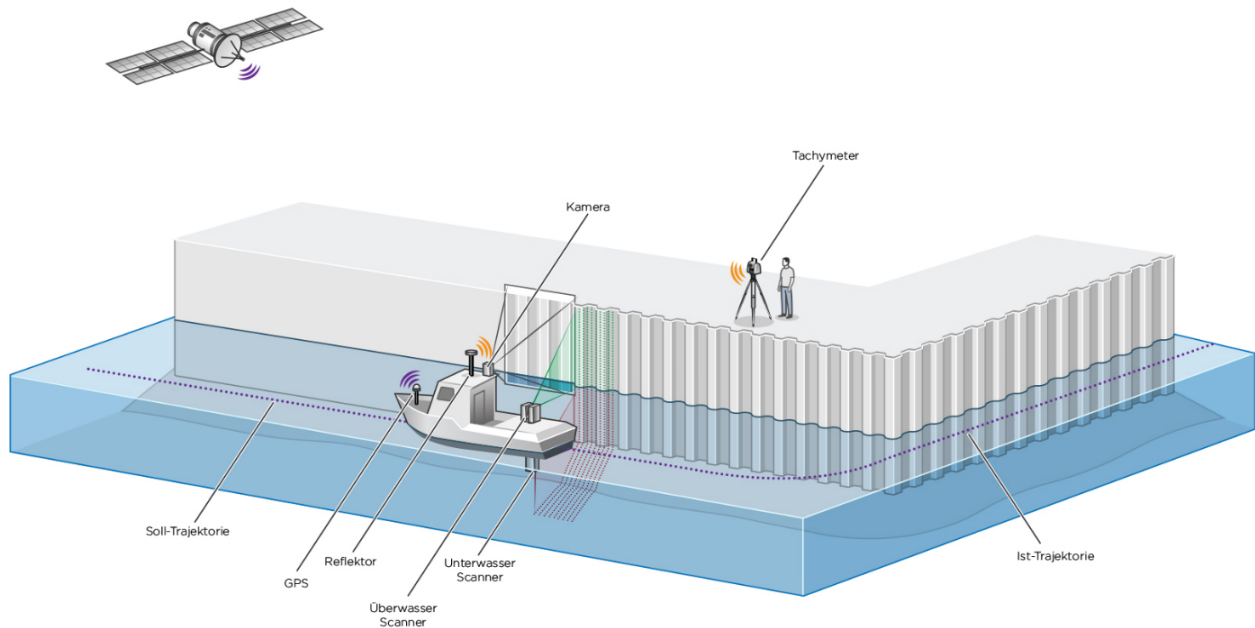


Abbildung 1.5: 3D-Kartierung einer Hafenstruktur über und unter Wasser (Hesse et al., 2019).



Abbildung 1.6: Entwickelte Messplattform (Foto: Karsten Holste).

Für die Positionierung des Bootes wird eine Kombination aus einem zwei Antennen GNSS System zusammen mit einer inertialen Messeinheit verwendet. Optional besteht die Möglichkeit, das Boot mit einem Tachymeter von Land aus zu verfolgen. Dies ist insbesondere im Hafenbereich relevant,

da dort durch die hohen Kranaufbauten und anderen Strukturen die GNSS Abschattung signifikant sein kann.

Für die Objekterfassung ist für den Unterwasserbereich ein MBES verbaut. Dabei handelt es sich um einen im akustischen Bereich arbeitenden Sensor, dessen Ergebnis eine 3D-Punktwolke ist. Dieser ist an einem Teleskopstab montiert, um die Messtiefe optimal an die Gegebenheiten anpassen zu können. Für die über Wasser liegenden Bereiche sind ein Profillaserscanner sowie ein Multi-Kamera-System verbaut. Der Profillaserscanner liefert eine sehr dichte und hochgenaue Punktwolke, die Kameras liefern hoch aufgelöste Farbbilder.

Da während der Projektlaufzeit des Forschungsprojektes die Messplattform entwickelt wurde, konnten noch keine Daten damit aufgenommen werden. Zur Entwicklung der Methoden und Algorithmen wurden daher die folgenden Datensätze sowohl durch Simulation als auch mit klassischen terrestrischen Verfahren erstellt.

Zur Evaluierung der grundsätzlichen Eignung des Verfahrens wird ein simulierter Datensatz einer Spundwand verwendet. Dies bietet den Vorteil, dass sowohl die wahre Struktur durch ein mathematisches Modell als auch die Position und Größe der Schäden bekannt sind. Weiterhin kann so das Problem umgangen werden, dass nicht ausreichend reale Messdaten mit manueller Segmentierung für das Training der Algorithmen vorhanden sind, indem eine große Anzahl an Datensätzen mit zufällig verteilten Schäden erzeugt werden kann. Für die Simulation wird ein mathematisches Modell einer Spundwand verwendet. Zunächst werden kartesische Koordinaten jedes Punktes auf einer Ebene der Spundwand durch Projizieren der Messstrahlen des k-MSS auf die nächste Fläche des mathematischen Modells in vertikalen Inkrementen erzeugt. Die dritte Dimension ergibt sich aus der Bewegung des k-MSS entlang einer Trajektorie, in diesem Fall zur Vereinfachung eine gerade Linie mit äquidistanten Abständen. Zusätzlich wird eine zufällige Anzahl an Schäden hinzugefügt. Jeder Schaden hat eine elliptische Form mit zufälligen Werten für die Hauptachsen. Das Ergebnis ist eine wahre Punktwolke, auf deren einzelne Messwerte zufälliges Messrauschen addiert wird. Die Messunsicherheit der Entfernungsmessung wird als normalverteilt angenommen und gemäß den Angaben des Herstellers auf 20 mm ($k=2$) festgelegt. Die Unsicherheiten der Winkel werden vernachlässigt, da sie zu einer kleinen Verschiebung in der Ebenenrichtung und einem daraus resultierenden kleinen Abstandsfehler führen, der bereits in der 20 mm-Messunsicherheit nicht signifikant ist. Die resultierende Auflösung der Punktwolke beträgt etwa 2 cm. Zusätzlich wird ein zweiter Datensatz generiert: Er gibt für jede Position an, ob an dieser Stelle ein Schaden vorliegt oder nicht, und ist somit das wahre Labelbild. Dies wird in Form eines binären 2D-Label-Bildes erstellt, bei dem ein Wert von eins für beschädigte und null für unbeschädigte Bereiche steht. Abbildung 1.7 veranschaulicht das Simulationsverfahren sowie ein Beispiel der Punktwolke und des dazugehörigen Labelbildes.

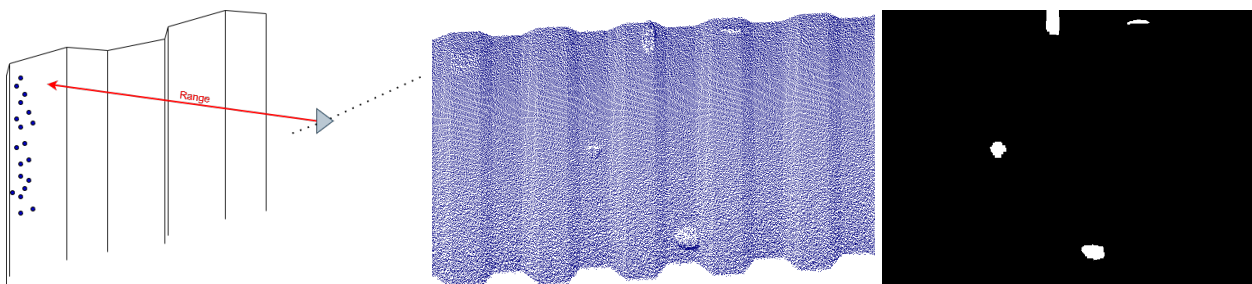


Abbildung 1.7: Links ist das in dieser Arbeit verwendete Simulationsprinzip zu sehen: An jeder Sensorposition werden alle Strahlen in vertikalen Inkrementen gemäß den Angaben des Herstellers mit den Ebenen der Spundwand verschnitten. In der Mitte ist die simulierte Punktwolke zu sehen und rechts das entsprechende Labelbild, wobei weiß für beschädigte und schwarz für unbeschädigte Bereiche steht (Hake et al., 2020).

Ein Beispiel eines Punktwolken-Datensatzes aus dem Stadthafen von Lübeck ist in Abbildung 1.8 dargestellt. Dabei zeigt das obere Bild ein Foto der Kaimauer und das untere Bild die kor-

respondierende Punktwolke. Der Datensatz wurde mit einem Z+F Imager 5016 Laserscanner auf 13 Standpunkten erfasst. Die Erfassung fand an einem Tag unter guten Bedingungen und sonnigem Wetter statt. Die Registrierung der Punktwolken wurde über schwarz-weiße Zielzeichen und Punkt zu Punkt Identitäten berechnet. Dabei werden zunächst die sechs Transformationsparameter (Translation und Rotation) über identische Zielmarken in beiden Scans mittels einer Helmert-Transformation (Stegner, 1984) berechnet. Im Anschluss daran werden in beiden Scans die jeweils nächsten Punktnachbarn bestimmt und mittels des Iterative Closest Point (ICP) Algorithmus (Besl & McKay, 1992) die Transformationsparameter verfeinert. Der Punktabstand innerhalb der Punktwolke variiert aufgrund unterschiedlicher Aufnahmepositionen und liegt im Mittel bei etwa 1 cm. Das Messrauschen der Punktwolke liegt im Mittel bei 1-2 mm und die zu erwartende Genauigkeit der Registrierung in den einzelnen Koordinatenkomponenten liegt bei 1,2 mm. Im oberen Teil der Kaimauer zwischen den Fendern sind offensichtlich Abplatzungen vorhanden. In Abbildung 1.9 sind beispielhaft zwei Abplatzungsschäden an einer Betonkaimauer dargestellt und in rot umrandet. Sie sind bis zu 1,5 m breit und erreichen eine Höhe von 50 cm. Der Punktwolken-Datensatz wird in

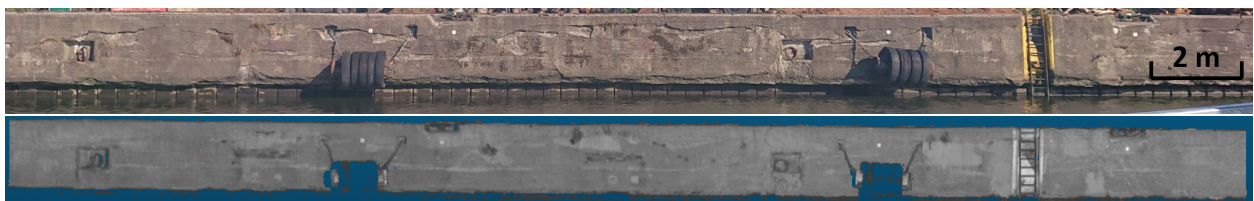


Abbildung 1.8: *Beispiel eines realen Datensatzes aus dem Stadthafen von Lübeck. Das obere Bild zeigt ein Foto der Kaimauer und das untere Bild die entsprechende Punktwolke (Hake et al., 2022).*

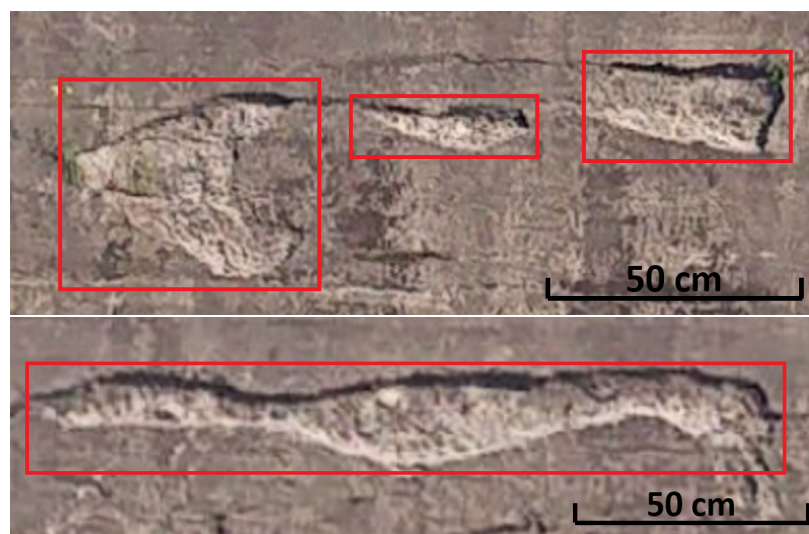


Abbildung 1.9: *Zwei Beispiele für beschädigte Bereiche im realen Datensatz. Beide Bilder zeigen Betonabplatzungen (rot markiert). Das obere Bild zeigt drei kleinere Schäden mit einer Größe von bis zu 50 cm × 50 cm. Das untere Bild zeigt einen größeren Abplatzungsschaden, der 1,5 m breit und 30 cm hoch ist (Hake et al., 2022).*

den Publikationen A und B (Abschnitt 2.1), sowie den weiterführenden Betrachtungen (Abschnitt 3.1 verwendet. Der Gesamtworkflow für die Schadenserkenkung aus 3D-Punktwolken ist allgemein in Abschnitt 1.4 beschrieben.

Der verwendete Bilddatensatz wurde im Jade-Weser-Port in Wilhelmshaven mit einem UAS erfasst. Die Bilder wurden orthogonal zur Kaimauer des Jade-Weser-Ports in einem Abstand von einigen Metern aufgenommen. So kann ein hochauflösendes digitales Bild des Hafens erstellt werden. Für die Bildaufnahme wurde eine Canon EOS 5D Mark III Kamera mit 35 mm Festbrennweite verwen-

det. Alle Einstellungen wurden nicht verändert und die Bilder wurden an einem Tag unter guten Bedingungen aufgezeichnet. Für die Georeferenzierung des Systems wurde ein zwei Antennen GNSS System und eine inertielle Messeinheit verwendet. Ein Pixel deckt den Objektbereich zwischen 2 und 5 mm in den beiden Raumrichtungen ab. Der Hauptzweck dieses Datensatzes ist die Erkennung von Korrosion an einer Betonwand. Der komplette Datensatz des Hafens in Wilhelmshaven besteht aus 1300 Bildern. Ein Beispiel für ein RGB Bild ist in Abbildung 1.10 dargestellt.



Abbildung 1.10: *Ausgeschnittenes RGB-Bild des Jade-Weser-Ports in Wilhelmshaven (Hake et al., 2023b).*

Der Bilddatensatz aus dem Jade-Weser-Port in Wilhelmshaven wird in den Publikationen C und D (Abschnitt 2.2), sowie den weiterführenden Betrachtungen (Abschnitte 3.2 und 3.3) verwendet. Der Gesamtworkflow für die Schadenserkenkung aus Farbbildern ist allgemein in Abschnitt 1.4 beschrieben.

1.6 Originalveröffentlichungen

Die vorliegende kumulative Dissertation umfasst methodische Beiträge und Untersuchungen zur Schadenserkenkung am Beispiel von Verkehrswasserbauwerken basierend auf 3D-Punktwolken und Farbbildern. Die relevanten eigenen Publikationen des Autors sind in Kapitel 6 aufgelistet. Sie wurden in peer-reviewed Medien veröffentlicht und auf zahlreichen Konferenzen und Workshops präsentiert. Die Abhängigkeiten zwischen den Veröffentlichungen und den Forschungsfragen ist in Abb. 1.1 dargestellt. Dieser Dissertation liegen vier wissenschaftlich begutachtete Publikationen zugrunde:

- **Publikation A:**
Hake, F.; Götttert, L.; Neumann, I. & Alkhatib, H. (2022): Using Machine-Learning for the Damage Detection of Harbour Structures, In: Remote Sensing, Jg. 14, Nr. 11, 2518. <https://doi.org/10.3390/rs14112518>, lizenziert unter CC BY 4.0, <https://creativecommons.org/licenses/by/4.0>
- **Publikation B:**
Hake, F.; Lippmann, P.; Alkhatib, H.; Oettel, V. & Neumann, I. (2023): Automated damage detection for port structures using machine learning algorithms in heightfields, In: Applied Geomatics, Jg. 2023. <https://doi.org/10.1007/s12518-023-00493-z>, lizenziert unter CC BY 4.0, <https://creativecommons.org/licenses/by/4.0>
- **Publikation C:**
Hake, F.; Scherff, M.; Neumann, I. & Alkhatib, H. (2023): Using Semantic Segmentation

for the Damage Detection of Port and Marine Infrastructures, In: A. Wieser (Hrsg.), Ingenieurvermessung 23: Beiträge zum 20. Internationalen Ingenieurvermessungskurs Zürich, 2023. Wichmann, VDE, S. 93 - 106, Zürich, Schweiz. <https://doi.org/10.15488/14198>

- **Publikation D:**

Scherff, M.; **Hake, F.**; & Alkhatib, H. (2023): Adaption of deeplab V3+ for damage detection on port infrastructure imagery, In: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-1-2023, 301–308. <https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-301-2023>, lizenziert unter CC BY 4.0, <https://creativecommons.org/licenses/by/4.0>

Im Folgenden sind die Beiträge der einzelnen Autoren an den Publikation A-D im Detail dargestellt:

Publikation A

- **Hauptautor (F. Hake):** Konzeption und Design der Studie, Durchführung der Experimente, Datenanalyse, Verfassung des Manuskripts, Korrespondenz während des Begutachtungsprozesses.
- **L. Götttert:** Unterstützung bei der Durchführung der Experimente und der Datenaufbereitung und Visualisierung.
- **I. Neumann:** Überprüfung und Korrektur des Manuskripts.
- **H. Alkhatib:** Methodische Beratung und fachliche Expertise, Überprüfung und Korrektur des Manuskripts.

Publikation B

- **Hauptautor (F. Hake):** Konzeption und Design der Studie, Durchführung der Experimente, Datenanalyse, Verfassung des Manuskripts, Korrespondenz während des Begutachtungsprozesses.
- **P. Lippmann:** Unterstützung bei der Durchführung der Experimente und der Datenaufbereitung und Visualisierung.
- **V. Oettel:** Überprüfung und Korrektur des Manuskripts.
- **H. Alkhatib:** Methodische Beratung und fachliche Expertise, Überprüfung und Korrektur des Manuskripts.
- **I. Neumann:** Überprüfung und Korrektur des Manuskripts.

Publikation C

- **Hauptautor (F. Hake):** Konzeption und Design der Studie, Durchführung der Experimente, Datenanalyse, Verfassung des Manuskripts, Korrespondenz während des Begutachtungsprozesses.
- **M. Scherff:** Unterstützung bei der Durchführung der Experimente und der Datenaufbereitung und Visualisierung.
- **I. Neumann:** Überprüfung und Korrektur des Manuskripts.
- **H. Alkhatib:** Methodische Beratung und fachliche Expertise, Überprüfung und Korrektur des Manuskripts.

Publikation D

- **Hauptautor (M. Scherff):** Konzeption und Design der Studie (gemeinsam mit F. Hake), Durchführung der Experimente, Datenanalyse (gemeinsam mit F. Hake), Verfassung des Manuskripts (gemeinsam mit F. Hake) Hauptverantwortlich für Kapitel 2, 3 und 4, Korrespondenz während des Begutachtungsprozesses.
- **F. Hake:** Konzeption und Design der Studie (gemeinsam mit M. Scherff), Datenanalyse (gemeinsam mit M. Scherff), Verfassung des Manuskripts (gemeinsam mit M. Scherff) Hauptverantwortlich für Kapitel 1, 5 und 6.
- **H. Alkhatib:** Methodische Beratung und fachliche Expertise, Überprüfung und Korrektur des Manuskripts.

Die in dieser Dissertation verwendeten Publikationen bilden die Grundlage zur Beantwortung der zwei Forschungsfragen (siehe Abschnitt 1.3). Die erste Forschungsfrage wird in den Publikationen A und B (siehe Abschnitt 2.1) und die zweite Forschungsfrage in den Publikationen C und D (siehe Abschnitt 2.2) behandelt. Somit liefern die Publikationen zusammen mit den weiterführenden Betrachtungen (siehe Abschnitt 3) jeweils einzelne Beiträge für den Gesamtprozess der digitalen Bauwerksprüfung. Dabei werden die Bereiche der Datenvorprozessierung und der automatisierten Erkennung von Schäden aus 3D-Punktwolken und Farbbildern behandelt. Im folgenden Kapitel 2 werden die vier Publikationen zusammengefasst und die Hauptaspekte herausgearbeitet.

2 Zusammenfassung der relevanten Publikationen

In diesem Abschnitt werden die vier wissenschaftlichen begutachteten Publikationen (A - D) zusammengefasst, die die Grundlagen für diese kumulative Dissertation bilden. Die Publikationen lassen sich, basierend auf den Forschungsfragen in Abschnitt 1.3, in zwei Hauptthemenbereiche gliedern. Zum einen handelt es sich um die Segmentierung von Punktwolken und die damit verbundene Erkennung von geometrischen Bauwerksschäden wie Abplatzungen oder größeren Rissen (A & B). Zum anderen um die Segmentierung von Farbbildern zur Erkennung von Bauwerksschäden mit einer Änderung der Farbe wie Rostfahnen oder Aussinterungen (C & D). Beide Themenbereiche wurden in den Publikationen A bis D sukzessive weiterentwickelt, wobei die Publikationen A und C den Ausgangspunkt bildeten. Die vier Publikationen bilden einen Grundbaustein im Prozess der digitalen Bauwerksprüfung. Ziel des Gesamtprozesses ist es, Zeit und Kosten bei Bauwerksprüfungen einzusparen, indem regelmäßig stattfindende, manuelle Prüfungen durch gezielte, digitale Methoden wie die automatisierte Schadenserkenkung unterstützt werden. Die in den Publikationen abgedeckten Themenbereiche bilden die Grundlage zur Beantwortung der beiden in Abschnitt 1.3 aufgestellten Forschungsfragen. Die in den Publikationen nicht ausreichend thematisierten Themen werden in den weiterführenden Betrachtungen in Abschnitt 3 detaillierter beschrieben. Im Folgenden werden die relevanten Publikationen A - D in den beiden Themenschwerpunkten automatisierte Erkennung von geometrischen Schäden (Abschnitt 2.1) und automatisierte Erkennung von Farbänderungen (Abschnitt 2.2) zusammengefasst und die Hauptaspekte herausgearbeitet.

2.1 Automatisierte Erkennung von geometrischen Schäden

In diesem Abschnitt werden die Ergebnisse der Publikationen A und B zusammengefasst, die sich mit der ersten Forschungsfrage beschäftigen: Wie können Punktwolken von unregelmäßigen Hafengebäuden bestmöglich für die digitale Bauwerksprüfung vorverarbeitet und durch automatisierte Verfahren auf Schädigungen überprüft werden? In Publikation A wird die grundsätzliche Eignung von Punktwolkendaten und maschinellem Lernen zur Erkennung geometrischer Schäden an Hafengebäuden anhand eines simulierten Datensatzes untersucht. Publikation B baut auf Publikation A auf und untersucht die Vorverarbeitung von Punktwolkendaten, insbesondere die Erstellung von Höhenfeldern, für die Schadenserkenkung an realen Hafengebäuden.

Für die automatisierte Auswertung müssen die Rohdaten zunächst vorverarbeitet werden und somit bestmöglich an die Algorithmen angepasst werden. Da unstrukturierte Punktwolken für die meisten Anomaliedetektionsalgorithmen ungeeignet sind, wird die Punktwolke in ein Höhenfeld konvertiert. Ein Höhenfeld (auch als Digitales Höhenmodell (DHM) (Skidmore, 1989) bezeichnet) ist in der Computergrafik ein Rasterbild, das hauptsächlich als diskretes globales Gitter in der Höhenmodellierung verwendet wird. Jedes Pixel enthält Daten über die Oberflächenhöhe. In dieser Arbeit werden nicht die Oberflächenhöhen, sondern Distanzen zwischen den einzelnen gemessenen Punkten der Punktwolke und der wahren, schadensfreien Struktur verwendet. Diese Distanzen können auf verschiedene Arten bestimmt werden: wenn ein mathematisches Modell, wie z. B. ein CAD, vorliegt, können die Distanzen direkt zwischen den Punkten und dem Modell berechnet werden. Liegt kein solches Modell vor, können lokal bestangepasste einfache Geometrien verwendet werden. Im Falle des simulierten Datensatzes (siehe Abschnitt 1.5) wird ein mathematisches Modell einer Spundwand zur Erstellung des simulierten Datensatzes und des entsprechenden CAD verwendet.

Für den realen Datensatz aus dem Stadthafen von Lübeck (siehe Abschnitt 1.5) liegt kein mathematisches Modell vor, sodass dieses aus den Daten bestmöglich geschätzt werden muss. Der Datensatz wird zunächst manuell mit Klassen für die Punkte versehen. Dabei werden die Klassen unbeschädigt, Betonabplatzung und Störobjekte verwendet. Störobjekte sind alle Objekte, die nicht in die Berechnung des Höhenfeldes einfließen sollen, da sie zu einer höheren Distanz führen würden, aber unbeschädigt sind. Beispiele hierfür sind Leitern, Fender oder starker Bewuchs. Das Ergebnis ist eine Punktwolke mit unterschiedlichen Indizes für die Kaimauer, Schäden und andere Objekte. Ein Beispiel hierfür ist in Abbildung 2.1 zu sehen. Dabei sind die drei Klassen unbeschädigt in Blau, Schaden in Grün und andere Störobjekte in Rot dargestellt. Dieser Vorprozessierungsprozess ist für eine genaue und zuverlässige Analyse erforderlich.

Für die Schätzung des Modells aus den Daten werden die Punktwolken zunächst durch eine Hauptkomponentenanalyse (Dunteman, 1989) in eine einheitliche Richtung gedreht. Dies ist notwendig, damit die Punktwolke automatisiert in kleinere, quadratische Bereiche zugeschnitten werden kann. Diese kleineren Quadrate überlappen sich jeweils zu 50 % in X- und Y-Richtung. Durch die kleinen Quadrate kann die Geometrie des Objektes besser angenähert werden. Hierzu werden Ebenen in jedes ausgeschnittene Quadrat nach Drixler (1993) (siehe Gleichung (2.1)) geschätzt. Dabei werden nur die Punkte verwendet, welche zu den Klassen unbeschädigt und Schaden gehören, damit andere Störobjekte die Schätzung nicht verfälschen.

Die Abstände von jedem Punkt zur entsprechenden Ebene im Modell werden nach Drixler (1993) mit Gleichung (2.1) bestimmt und dann in ein Höhenfeld mit einer Rastergröße von 2 cm überführt,

$$dist = \frac{n_x \cdot p_x + n_y \cdot p_y + n_z \cdot p_z - d}{|\mathbf{n}|} \quad (2.1)$$

wobei \mathbf{n} der Normalenvektor der Ebene mit den Einträgen n_x , n_y und n_z ist. d ist der Abstand zum Ursprung und p_x , p_y und p_z sind die Koordinaten des Punktes. Im Falle des realen Datensatzes werden für die Punkte, welche zu anderen Störobjekten gehören, die Distanzen manuell auf kleine Werte gesetzt, um die Erkennung der Schäden nicht zu beeinflussen. Die Rastergröße ist abhängig von der Auflösung der Punktwolke und muss an den jeweiligen Datensatz so angepasst werden, dass in möglichst jeder Rasterzelle mindestens ein Messwert liegt. Leere Zellen, die aufgrund von Datenlücken oder ungünstiger Punktverteilung entstehen, werden mittels „natural neighbour interpolation“ (Sibson, 1981) interpoliert, um Beeinträchtigungen bei der Merkmalsextraktion zu vermeiden (vgl. Gleichung (2.2)):

$$G(x) = \sum_{i=1}^n \frac{A(x_i)}{A(x)} f(x_i). \quad (2.2)$$

$G(x)$ ist der Schätzwert in Rasterzelle x , $f(x_i)$ die bekannten Daten in x_i , $A(x)$ das Volumen der neuen Zelle in der Mitte von x und $A(x_i)$ das Volumen der Schnittmenge zwischen der neuen Zelle in der Mitte von x und der alten Zelle in der Mitte von x_i . In überbesetzten Zellen wird der Median der Distanzen verwendet. Das Höhenfeld des erhaltenen Infrastrukturobjektes wird als skalare Funktion interpretiert, die auf einem 2D-Gitter definiert ist und mit $H(x, y)$ bezeichnet wird. Anschließend werden Patches aus dem Gitter extrahiert und in Datenvektoren überführt, welche als Matrix \mathbf{X} mit der Form $N \times p$ gespeichert werden, wobei N die Anzahl an Patches und p die Anzahl der Pixel ist. Abbildung 2.1 zeigt ein Beispielausschnitt der realen Punktwolke aus Lübeck. In der oberen Reihe ist die Punktwolke nach den manuell eingeteilten Klassen unbeschädigt in Blau, Schaden in Grün und andere Störobjekte in Rot eingefärbt dargestellt. Die mittlere Reihe zeigt das erzeugte Höhenfeld in Graustufen, wobei ein hellerer Grauwert eine größere Abweichung vom mathematischen Modell repräsentiert. Außerdem wird ein entsprechendes Labelbild (untere Reihe) erstellt, bei dem es sich um ein binäres Bild handelt, in dem Punkte mit Schadensindizes weiß und alle anderen Punkte schwarz dargestellt sind.

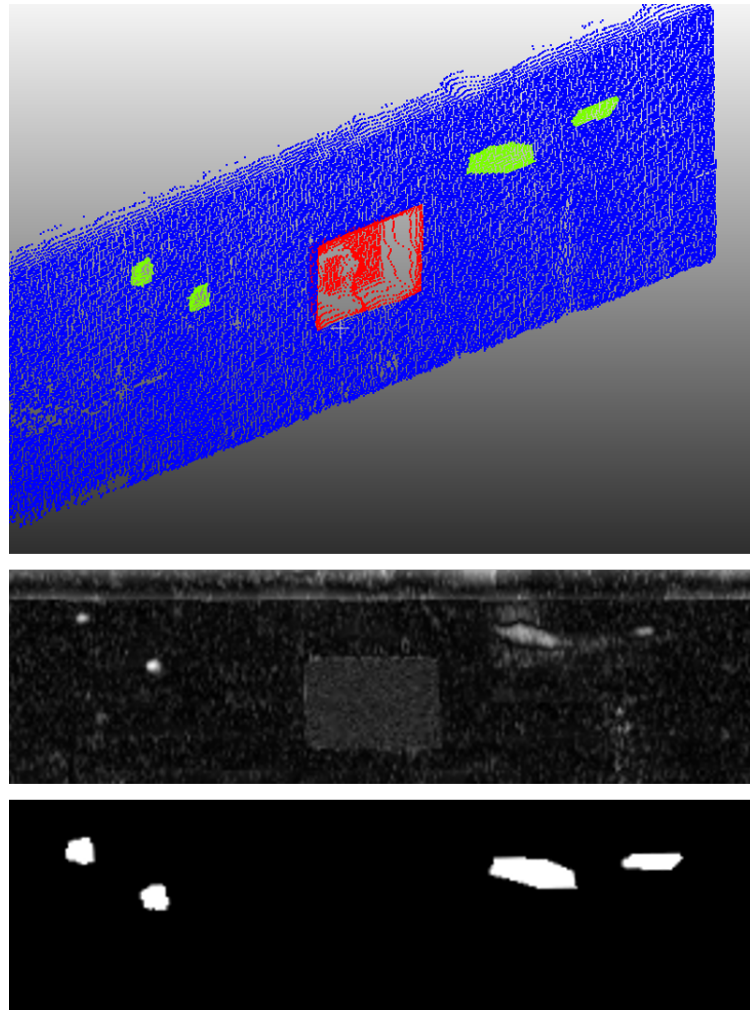


Abbildung 2.1: *Reale Daten aus dem Stadthafen Lübeck mit Höhenfeld und Labelbild. Die obere Reihe zeigt die Punktwolke in den drei Klassen unbeschädigt (blau), Schaden (grün) und andere Störobjekte (rot). Die mittlere Reihe zeigt das daraus entstandene Höhenfeld in Grauwerten, wobei hellere Grauwerte für größere Distanzen stehen. In der unteren Reihe ist das binäre Labelbild, welches für die Evaluierung benötigt wird. Dabei entspricht schwarz unbeschädigt und weiß Schaden (Hake et al., 2023a).*

Zur Extraktion relevanter Merkmale aus dem Höhenfeld wird eine modifizierte Version des VGG19-Netzes (Simonyan & Zisserman, 2014) eingesetzt, das zuvor auf dem umfangreichen ImageNet-Datensatz (Deng et al., 2009) vortrainiert wurde. Der Vorteil des an den mehr als 14 Mio. Bildern des ImageNet Datensatzes vortrainierten Gewichten ist, dass sie schon sehr gut initialisiert sind. Dadurch kann die Anzahl an benötigten Trainingsdaten für die Schadenserkenkung deutlich reduziert werden. Das VGG19 Netz wird nach dem vierten Pooling Layer abgeschnitten, da die tieferen Schichten dazu neigen, hochwertige Merkmale wie Gesichter, Autos etc. zu lernen während die ersten Schichten eher grundlegendere Merkmale wie Kanten und Sprünge lernen. Die erzeugten Höhenfelder werden zur effizienteren Verarbeitung in kleinere Abschnitte unterteilt, wobei sich die Abschnitte jeweils um 50 % überlappen. Dies schafft auch die Möglichkeit, mehr als ein Label für den gesamten Bereich zu erhalten. Infolgedessen können Defekte aufgrund der geringeren Größe der Kacheln effizienter lokalisiert werden. Jeder Abschnitt wird durch das Netz propagiert, und die Aktivierung gespeichert. Für jede Aktivierung wird die Diagonale der Gramian Matrix (Gatys et al., 2015) berechnet, die sich auf die Energie pro Merkmal bezieht. Dies führt zu dem Merkmalsvektor $\mathbf{z} \in \mathbb{R}^k$, wobei k die Anzahl der Merkmalskarten in der j -ten Schicht des Netzes ist. Die Merkmalsvektoren werden als Zeilen in einer Matrix organisiert, was zu einer Merkmalsmatrix \mathbf{Z}

mit der Form $N \times k$ führt.

Für die eigentliche Erkennung der Schäden durch eine Segmentierung hat sich die LOF-Methode (Breunig et al., 2000) als am besten geeignet erwiesen. Im Gegensatz zu klassischen Cluster- oder Klassifizierungsmethoden, die mehrere Cluster suchen, erkennt LOF Datenpunkte, die von der Mehrheit abweichen und nicht dem erwarteten Verhalten entsprechen. Somit erfüllt die LOF-Methode genau die Anforderungen, da sie wenige, unterschiedlich geformte Schäden in einem großen unbeschädigten Bereich erkennen kann. Dabei wird im Merkmalsraum ein Erreichbarkeitsgraph gebildet, um somit die Dichte der Nachbarschaften schätzen zu können. Aus dieser Dichte wird dann für jeden Datenpunkt ein Ausreißer-Score berechnet. Als Eingabe wird die Merkmalsmatrix \mathbf{Z} verwendet, welche einem Patch des Höhenfeldes entspricht. Diese Merkmale werden zu einem einzelnen Wert zusammengefasst, der die Wahrscheinlichkeit angibt, ob das jeweilige Patch einen Schaden aufweist. Die Patches überlappen zu 50 %, was dazu führt, dass jedes Pixel viermal ausgewertet wird. Aus diesen vier Werten wird der Mittelwert berechnet und in einem regelmäßigen, nicht überlappenden Raster zusammengefasst, um eine bessere Interpretation, Analyse und Bewertung der Klassifizierungsergebnisse zu ermöglichen. Um aus den skalaren Ausreißerwerten ein Label für Schaden oder unbeschädigt zu erzeugen, ist ein Schwellwert erforderlich. Dieser kann beliebig gewählt werden, um die gewünschte Falsch-negativ-Rate zu erzeugen. Mit einem größeren Schwellwert werden mehr Bereiche als Schaden segmentiert. Dabei sinkt jedoch auch die Präzision, da mehr erkannte Schäden in Wahrheit keine Schäden sind. Da es deutlich mehr unbeschädigte Bereiche als Schäden gibt, kann der Wert für die Genauigkeit nah an 100 % liegen, obwohl der Recall schlechter ist. Die Genauigkeit alleine ist daher nicht geeignet, um die Leistung der Anomalieerkennung zu bewerten. Abbildung 2.2 zeigt, wie die üblicherweise verwendeten Klassifizierungsbewertungsmetriken vom Schwellwert für den LOF-Ansatz abhängen.

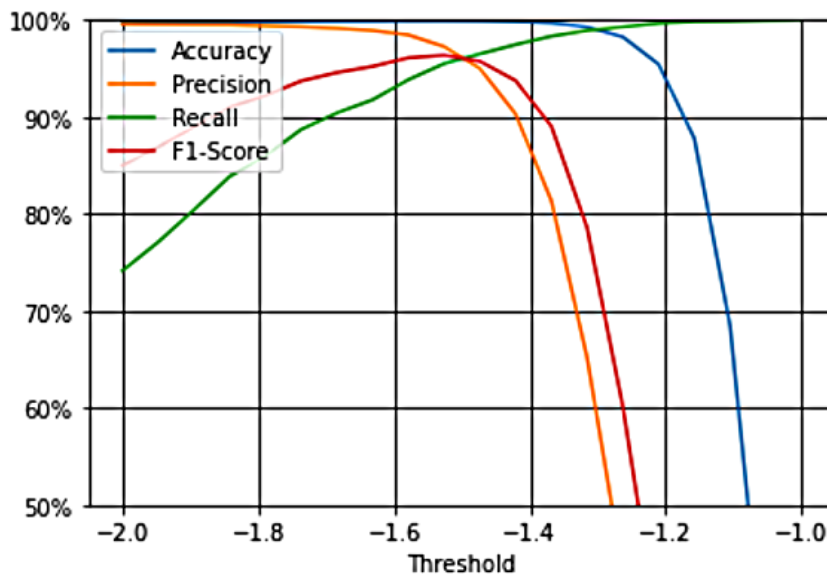


Abbildung 2.2: Abhängigkeit der Bewertungsmetriken vom Schwellwert (Hake et al., 2023a).

Beschädigte Bereiche haben niedrigere (negativere) Werte als unbeschädigte. Daher führt ein höherer Schwellwert zu einer Verringerung des Recall (mehr gefundene fehlerhafte Beispiele). Gleichzeitig werden aber auch mehr unbeschädigte Bereiche als Schaden erkannt, was die Präzision verringert. Das primäre Ziel unserer Anwendung ist es, die meisten Schäden zu finden, das heißt ein hoher Recall. Gleichzeitig muss aber die Genauigkeit niedrig genug sein, um den nachträglichen Arbeitsaufwand zu reduzieren. Alles als schadhaft zu kennzeichnen, ist nicht sinnvoll. Der Schwellwert muss so gewählt werden, dass beide Bedingungen erfüllt sind.

Die besten Ergebnisse für den simulierten Datensatz (500 Höhenfelder) wurden mit einer Kachel-

größe von 32 x 32 Pixeln als Eingabe für das CNN erzielt. Bei einem Schwellwert von -1,5 ergibt sich die in Tabelle 2.1 dargestellte Konfusionsmatrix, die eine Genauigkeit von 99,9 %, eine Präzision von 96,3 % und einen Recall von 96,2 % aufweist. Dies führt zu einem F1-Score von 96,3 %.

Tabelle 2.1: Konfusionsmatrix für Recall = 96,2 % (Schwellwert = -1,5).

		wahre Klasse	
		Schaden	unbeschädigt
vorhergesagte Klasse	Schaden	52217	2054
	unbeschädigt	2008	4050050

Accuracy beschreibt den Anteil der korrekt klassifizierten Vorhersagen des Modells und wird nach Gleichung (2.3) berechnet.

$$\text{Accuracy} = \frac{\#\text{correct predictions}}{\#\text{all predictions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2.3)$$

Mit True Positive (TP), False Positive (FP), True Negative (TN) und False Negative (FN). Das Verhältnis der wahren positiven Ergebnisse, die vom Modell vorhergesagt wurden, wird durch die Precision nach Gleichung (2.4) angegeben.

$$\text{Precision} = \frac{\#\text{relevant retrieved instances}}{\#\text{all retrieved instances}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.4)$$

Recall (auch Sensitivität genannt) gibt nach Gleichung (2.5) die Wahrscheinlichkeit an, mit der ein positives Objekt korrekt als positiv klassifiziert wird.

$$\text{Recall} = \frac{\#\text{relevant retrieved instances}}{\#\text{all relevant instances}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.5)$$

Das harmonische Mittel aus Recall und Precision wird durch den F1-Score nach Gleichung (2.6) beschrieben.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \quad (2.6)$$

Der Mathews correlation coefficient (MCC)-Wert wurde nach Gleichung (2.7) ebenfalls mit 0,96 ermittelt, was auf eine gute Klassifizierung hinweist.

$$\text{MCC} = \frac{\text{TN} \cdot \text{TP} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{TN} + \text{FN})(\text{FP} + \text{TP})(\text{TN} + \text{FP})(\text{FN} + \text{TP})}} \quad (2.7)$$

Aus der Konfusionsmatrix in Tabelle 2.1 geht hervor, dass wenige Schäden nicht erkannt wurden, jedoch die Anzahl an false positive Ergebnissen höher ist. Dies könnte durch eine Veränderung des Schwellwertes reduziert werden, wodurch jedoch auch die Anzahl an nicht erkannten Schäden steigen würde.

Nach der erfolgreichen Evaluierung an simulierten Daten wird der Algorithmus im Folgenden auf reale Daten angewendet, um seine Praxistauglichkeit zu überprüfen. Hierzu wird der Datensatz aus dem Stadthafen von Lübeck verwendet (siehe Abschnitt 1.5), wobei sich die in Abbildung 2.3 dargestellten Kurven für Genauigkeit, Präzision, Recall und F1-Score ergeben, welche ein ähnliches Verhalten bei geringerer Gesamtgenauigkeit zeigen. Auch hier wird ein Schwellwert gewählt, bei dem Präzision und Recall im Wesentlichen gleich sind, was einen guten Kompromiss zwischen true und false positives im wirtschaftlichen Sinne darstellt.

Die entsprechende Konfusionsmatrix für den gewählten Schwellwert von -1,55 ist in Tabelle 2.2 zu

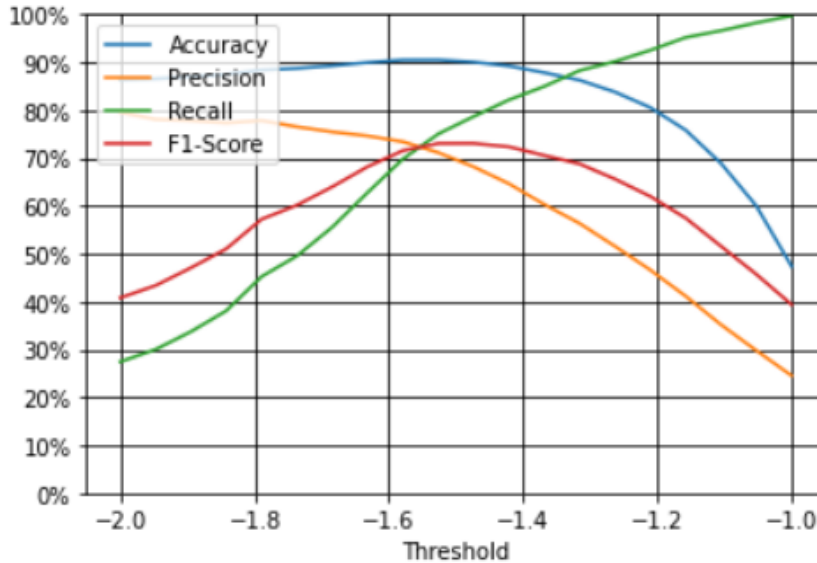


Abbildung 2.3: Abhängigkeit der Bewertungsmetriken vom Schwellenwert für reale Daten (Hake et al., 2023a).

sehen. Wie aus der Tabelle ersichtlich ist, besteht erneut ein starkes Ungleichgewicht zwischen den beiden Klassen. Die Anzahl der Falsch-Positiven und Falsch-Negativen ist im Wesentlichen gleich groß.

Tabelle 2.2: Konfusionsmatrix für Recall = 72,6 % (Schwellenwert = -1,55).

		wahre Klasse	
		Schaden	unbeschädigt
vorhergesagte Klasse	Schaden	9128	3437
	unbeschädigt	3506	57225

Für den realen Datensatz aus dem Stadthafen Lübeck liegt bei einem Schwellenwert von -1,55 die Genauigkeit bei 90,5 %, die Präzision bei 72,2 %, der Recall bei 72,6 % und der F1-Score bei 72,4 %. Der MCC-Wert liegt bei 0,66, was auf eine gute Klassifizierung hinweist, die jedoch etwas niedriger ist als die mit simulierten Daten erzielten Ergebnisse. Obwohl die Ergebnisse nicht so gut waren wie die, die mit simulierten Daten erzielt wurden, scheint die Methode bei der Anwendung auf reale Daten dennoch effektiv zu sein. Das vorgestellte Verfahren soll dazu dienen, aus Punktwolken automatisch einen Verdachtsplan mit vermuteten Schadensregionen zu erstellen. Dieser wird von den Bauwerksprüfern verwendet, um gezielt die Schäden kontrollieren zu können. Somit findet nach der automatisierten Schadenserkenner eine manuelle Qualitätskontrolle der Ergebnisse statt. Um das Verfahren in der Realität anwenden zu können, müssen möglichst alle Schadensregionen gefunden werden. Darüber hinaus sollen nur die Schadensregionen als solche erkannt werden. Das bedeutet, dass Präzision und Recall zusammen möglichst hoch sein sollten.

Das Klassifizierungsergebnis für zwei Beispielbilder ist in Abbildung 2.4 dargestellt. Grün, Rot und Gelb stehen für richtig-positiv, falsch-negativ bzw. falsch-positiv. Das Höhenfeld ist mit einem höheren Kontrast in der Mitte dargestellt, um Grauwertunterschiede im Höhenfeld für das menschliche Auge besser sichtbar zu machen. In der obersten Zeile von Abbildung 2.4 ist zu erkennen, dass alle Schäden korrekt erkannt und klassifiziert werden. Es gibt keine falsch-positiven oder falsch-negativen Ergebnisse. Lediglich die beiden kleinen Schäden am unteren Rand werden nicht erkannt, aber das liegt daran, dass die Randbereiche bei der Klassifizierung abgeschnitten werden,

da es dort zu Artefakten aus der Interpolation leerer Zellen kommen kann. Das Beispiel in der unteren Reihe zeigt ein schwächeres Klassifikationsergebnis. Zwei beschädigte Bereiche werden nicht erkannt und darüber hinaus werden zwei Bereiche fälschlicherweise als Schaden erkannt. Gründe hierfür können in der unterschiedlichen Art und Tiefe der Schäden, in einer ungünstigen Schätzung des Modells oder anderweitigen Störobjekten wie zum Beispiel Pflanzen liegen. Ein weiterer Grund liegt in den für das Training verwendeten Labelbildern. Diese wurden nicht von Experten für die Bauwerksprüfung erstellt und weisen dadurch ein recht hohes Label Noise auf, was in der Folge zu Fehlklassifizierungen führen kann.

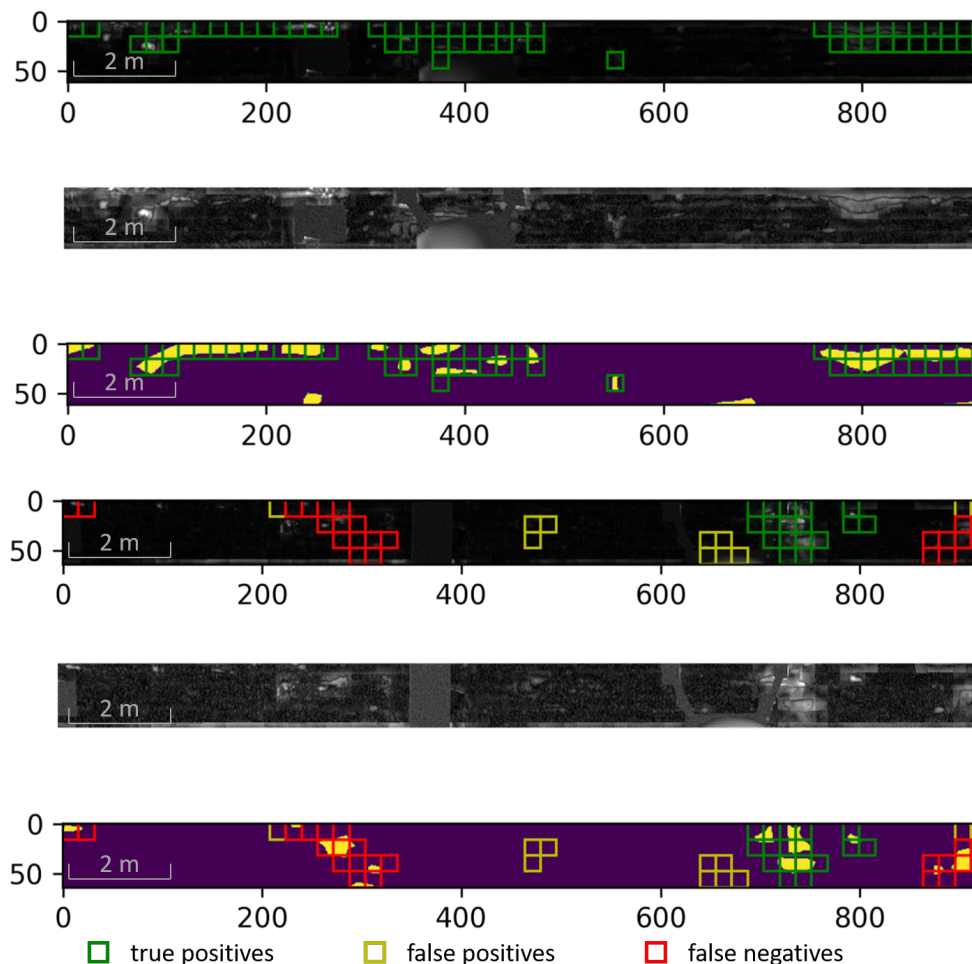


Abbildung 2.4: Zwei Beispiele für die Schadenserkenkung anhand des realen Datensatzes aus dem Lübecker Stadthafen. Die oberste Zeile ist jeweils die Punktwolke mit überlagerten Schadenserkennung. Die Farben Grün, Rot und Gelb kennzeichnen jeweils richtig-positive, falsch-negative und falsch-positive Ergebnisse. Die mittlere Reihe ist jeweils das Höhenfeld und die untere Reihe ist das Labelbild mit überlagerten erkannten Bereichen (Hake et al., 2022).

Die unterschiedlichen Ergebnisse zwischen dem simulierten und realen Datensatz lassen sich auf verschiedene Gründe zurückführen. Zum einen sind im realen Datensatz Störobjekte wie Leitern, Fender oder Pflanzen vorhanden, welche auch zu höheren Distanzen zwischen Objekt und Punktwolke führen. Es wird aber auch keine Bereinigung und Filterung der Daten vorgenommen, sodass auch Ausreißer und grobe Messfehler in den Daten verbleiben. Ein weiterer Grund für die unterschiedlichen Ergebnisse liegt in der Erstellung der Höhenfelder, welche bei den simulierten über CAD-Modelle und bei den realen über ausgleichende Ebenen erstellt werden. An dieser Stelle scheint noch Potenzial zur Verbesserung vorhanden zu sein, in dem komplexere Geometrien oder

andere Größen verwendet werden. Eine Auswahl an weiteren Geometrien und verschiedenen Größen wurden in Abschnitt 3.1 untersucht.

Insgesamt wurden vergleichbare Ergebnisse zu anderen Studien erzielt. Point CNN ergibt ein mean interval over unit value von 74,68 % für Brückeninspektionen mit Punktwolkenklassifizierungen (Kim & Kim, 2020). Eine Kombination aus Bildern und Punktwolken, die auf dem Algorithmus von Otsu zur automatischen Erkennung von Betonrissen basiert, erreicht einen durchschnittlichen F1-Score von 86,7 % (Chen et al., 2021).

2.2 Automatisierte Erkennung von Farbänderungen

In diesem Abschnitt werden die Ergebnisse der Publikationen C und D zusammengefasst, welche sich mit der zweiten Forschungsfrage beschäftigen. Diese lautet: Können Schäden automatisiert aus Farbbildern strukturierter Oberflächen von Hafeninfrastrukturen erkannt werden und kann die Anzahl an manuell erzeugten Labeln für das Training reduziert werden? Publikation C untersucht die automatisierte Schadenserkenkung in Farbbildern von Hafeninfrastrukturen, wobei der Fokus auf Korrosionsschäden liegt. Dabei werden verschiedene Bildklassifizierungs- und Segmentierungsmethoden evaluiert. Publikation D baut auf Publikation C auf und untersucht die Verwendung von schwach überwachten Labelbildern, um den Aufwand der manuellen Labelerstellung zu reduzieren. Dazu wurden verschiedene Bildklassifizierungs- und Segmentierungsmethoden aus dem Bereich des maschinellen Lernens und des Deep Learning untersucht und optimiert, um diese Aufgabe zu automatisieren. Zur Validierung des Verfahrens wird der Datensatz an Farbbildern des JadeWeserPort aus Wilhelmshaven (siehe Abschnitt 1.5) verwendet. Dieser besteht insgesamt aus 1300 Farbbildern, von denen 80 mit manuellen Labeln für Schaden und unbeschädigt versehen wurden. Der Hauptzweck dieses Datensatzes ist die Erkennung von Korrosion an einer Betonwand. Andere Arten von Schäden, wie z. B. Abplatzungen, Risse und Aussinterungen, treten so selten auf (maximal eine niedrige zweistellige Zahl in der gesamten Anlage), dass die Modelle nicht in der Lage waren, sie zu lernen oder effektiv zu erkennen. Die 80 manuell segmentierten Bilder wurden so gewählt, dass sie gleichmäßig über die gesamte Kaimauer verteilt sind und alle relevanten Merkmale und Ausprägungen von Rostschäden repräsentieren. Die manuelle Segmentierung in Schaden und unbeschädigt wurde in Form eines gleich großen Schwarz-Weiß-Bildes gespeichert. Dies wurde zum einen durch ein selbst entwickeltes Matlab Programm und durch die Nutzung des „Supervisely Service“ (Supervisely contributors, 2023) umgesetzt. Abbildung 1.3 (rechts) zeigt ein Beispiel für ein RGB-Bild und das dazugehörige manuell erstellte Labelbild. Die Trainings- und Testdatensätze wurden im Verhältnis 80/20 aufgeteilt und nach dem Zufallsprinzip zugewiesen. Die Datensätze wurden überprüft, um sicherzustellen, dass sie die spezifischen Merkmale des Hafens, wie etwa die Lichtverhältnisse, abdecken.

Zur Bewertung der Leistung der verschiedenen Segmentierungsverfahren mittels Kreuzvalidierung wird der Datensatz zufällig in sechs Teilmengen mit jeweils 14 Bildern aufgeteilt. Vier Teilmengen bilden das Trainingsset, die restlichen zwei werden jeweils als Validierungs- und Testset verwendet. Das Validierungs-Set bestimmt den optimalen Parameterzustand eines jeden Modells auf der Grundlage des entsprechenden niedrigsten epochenweisen Verlustwerts, und das Test-Set dient als unabhängige Bewertungsmaßstab.

Da die manuelle Erstellung der Labelbilder sehr zeitaufwendig ist (bis zu 30 Minuten für ein Bild), wird zusätzlich die Eignung schwach überwachter Labelbilder untersucht. Die schwach überwachten Labelbilder werden mithilfe eines Region-Growing-Algorithmus erstellt. Da nur ein kleiner Teil des Bildes mit einem Pinselmarkierungstool markiert wird, anstelle des pixelgenauen Polygons, kann der Prozess deutlich beschleunigt werden. Diese schnell markierten Bereiche werden anschließend mit dem Region-Growing-Algorithmus erweitert.

Die hier verwendete Version des Region Growing basiert auf der unabhängigen Standardisierung jedes Farbkanals c , um zu bestimmen, ob ein benachbartes Pixel zur Startregion hinzugefügt werden soll. Ein Pixel x_i wird zur Startregion hinzugefügt, wenn die maximale standardisierte absolute

Abweichung von der erwarteten Farbe innerhalb eines bestimmten Schwellwertes t liegt. Gleichung (2.8) definiert die Abweichung:

$$\Delta x_{max} = \max \left(\left| \frac{x_i - \mu_c}{\sigma_c} \right| \right) \leq t \quad (2.8)$$

Dabei stellt Δx_{max} die Abweichung dar, μ_c und σ_c enthalten die kanalweisen statistischen Größen und die Funktion \max wählt das größte Element des standardisierten RGB-Vektors. Der Schwellwert t wird auf einen Wert von 1,25 gesetzt, um ein Überlaufen der Startregion in andere Farbbereiche zu verhindern. Zur Verhinderung von falsch erstellten Korrosionssegmenten wird ein Conditional Random Field (CRF) (Krähenbühl & Koltun, 2011) in jeder Iteration angewandt. Der Schwellwert wird nicht speziell auf die Optimierung bestimmter Metriken abgestimmt, sondern wird durch den visuellen Vergleich verschiedener Ergebnisse für eine kleine Gruppe von Bildern ausgewählt.

Damit Region Growing den vollständig überwachten Labeln qualitativ nahe kommt, müssen die Schadenssegmente homogene Farben haben und sich deutlich vom Hintergrund und den umgebenden Klassen unterscheiden. Dies trifft im Durchschnitt für Korrosion und zum Beispiel Ausblühungen zu, gilt jedoch beispielsweise nicht für Betonabplatzungen oder Risse. In Abbildung 2.5 ist ein originales Farbbild mit den manuell erstellen Labeln und den automatisiert durch Region Growing erstellten Labeln dargestellt.



Abbildung 2.5: *Beispielbild des JadeWeserPorts, oben das originale Farbbild, in der Mitte die manuell erzeugten Label und unten die Label vom schwach überwachten Region Growing Verfahren (Scherff et al., 2023).*

Für die automatisierte Segmentierung der Bilder in Korrosion und unbeschädigte Bereiche wurden

zwei Methoden evaluiert. Zunächst wird mit den manuell erzeugten Labeln das beste Modell ermittelt. Im Anschluss wird an diesem Modell die Eignung der schwach überwachten Labelbilder evaluiert.

Als klassische Methode des maschinellen Lernens wurde in Liu et al. (2019) Random Forest als bester Klassifizierungsalgorithmus zur Erkennung von Rostschäden vorgeschlagen. Der Random Forest Klassifikator besteht aus mehreren Entscheidungsbäumen, die jeweils den Merkmalsraum in klassifizierte Regionen unterteilen. Jeder Baum trifft seine Entscheidung basierend auf einer zufälligen Auswahl von Merkmalen. Diese Regionen werden durch Blattknoten in jedem Baum dargestellt. Um ein Klassifikationsergebnis für eine neue Stichprobe zu erhalten, wird jeder Baum von oben nach unten durchlaufen. Auf jeder Tiefenebene folgt die Probe einem von zwei Pfaden (Kanten im Graphenkontext), bis ein Blattknoten erreicht wird. Der Random Forest Klassifikator verwendet RGB-Pixelwerte und den improved hue, saturation, lightness (IHLS)-Farbraum als Merkmale, um zwischen zwei Klassen zu unterscheiden. Der IHLS-Farbraum wird mithilfe nichtlinearer Funktionen aus den ursprünglichen Pixelwerten erstellt. Um das Random Forest Klassifizierungsmodell zu optimieren, wird ein grid search für die Anzahl der Entscheidungsbäume und die Klassengewichte durchgeführt. Die Ausgangswerte basieren auf Liu et al. (2019) und der Einstellung „balanced“ der Random Forest Klasse im Scikit-learn Python-Paket (Pedregosa et al., 2011). Die Suchintervalle für die Anzahl der Bäume lagen zwischen 8 und 14 mit einer Schrittweite von 2, und die Gewichte für die Korrosionsklasse reichten von 2 bis 5. Das Gewicht für die Klasse 0 wurde immer auf 1 gesetzt. Diese Optimierung erfolgte im Hinblick auf den gewichteten F1-Score zwischen den beiden stark unausgewogenen Klassen. Weitere Parameter, die dem Bootstrapping, den minimalen Stichproben pro Blattknoten, der maximalen Baumtiefe, den berücksichtigten Merkmalen und der Menge des Informationsgewinns entsprechen, wurden für den gesamten Prozess als „False“, 0,01 %, 14, None (alle Merkmale) bzw. „Entropie“ festgelegt. Es wurden mehrere Modelle mit verschiedenen Parameterpaaren trainiert, wobei 10 Entscheidungsbäume mit einem Klassengewicht von 1:3 den höchsten F1-Score erzielten. Von diesem Modell wurden zwei weitere Instanzen trainiert, eine nur mit dem IHLS-Farbraum und die andere mit allen sechs Farbmerkmalen. Zur Validierung des Modells wurden fünf verschiedene Modelle auf verschiedenen Teilen des Trainingsdatensatzes trainiert und auf dem Testsatz bewertet.

Die Gesamtunsicherheit, die sich aus Modell- und Datenunsicherheit zusammensetzt und mithilfe der Entropie berechnet wird, ist bei allen drei Modellen vergleichbar. Das auf den IHLS-Farbraum trainierte Modell erreicht 0,2733 im Vergleich zu 0,2538 beim RGB-bezogenen Klassifikator. Dies spiegelt sich auch in den anderen Metriken wie der Gesamtgenauigkeit, dem F1-Score und den pixelweisen IoU-Scores wider. Das auf dem RGB-Farbraum trainierte Modell übertrifft das IHLS-Modell um bis zu 2,0 % und sogar 8,0 % für die Hintergrund- bzw. Korrosionsklasse. Die einzelnen Metriken der Modelle, die mit dem RGB- und den beiden Farbräumen verbunden sind, sind in der Tabelle 2.3 aufgeführt.

Tabelle 2.3: *Ergebnisse der ausgewählten Random Forest Klassifikatoren.*

Metric	RGB	RGB & IHLS
Total uncertainty	0,2538	0,2618
Overall accuracy [%]	94,1	93,8
F1-Score [%] (class 0; 1)	96,8; 59,9	96,6; 59,3
Pixel-wise IoU [%] (class 0; 1)	93,8; 42,8	93,5; 42,2

Um die Ergebnisse des klassischen Random Forest Ansatzes zu vergleichen, wurde zusätzlich ein Deep Learning Ansatz evaluiert. Hier wird auf die bekannte DeepLabV3+-Architektur (Chen et al., 2018) aufgebaut, da sie in anderen Arbeiten als sehr gut für Segmentierungsarbeiten geeignet beschrieben wird. Dabei handelt es sich um eine Encoder-Decoder-Netzstruktur. Um es an die spezifische Aufgabe der Erkennung von Rostschäden an Hafengebäuden anzupassen, wurden die Struktur

und einige Parameter des Netzes geändert. Eine dieser Änderungen war die Verdoppelung der Dilationsparameter innerhalb des ASPP Blocks. Das modifizierte Netz wurde dann mit verschiedenen Merkmalsextraktionsnetzen („Backbones“) und Verlustfunktionen trainiert und bewertet, um den IoU-Score für die Korrosionsklasse zu maximieren. Als Backbones wurden neben dem in Chen et al. (2018) verwendeten Modified Aligned Xception Net unter anderen das in He et al. (2016) vorgestellte ResNet-50 (version 1.5) sowie VGG-16 BN (Simonyan & Zisserman, 2014) getestet. Um den unausgewogenen Daten Rechnung zu tragen, wurde die Fehlerfunktionen mit Klassengewichten versehen, mit einem Gewicht von 5 für die Korrosionsklasse im ResNet-Modell, das sich nach dem IoU-Score als das beste Modell erwies. Die Leistung des Modells wurde mit demselben Verfahren wie im Fall des Random Forest validiert, wobei die Modellparameter aus der Epoche mit dem niedrigsten Verlustwert in der verbleibenden Trainingsmenge übernommen wurden. Um die Genauigkeit des DeepLabV3+ Netzwerks zu verbessern, werden die Eingabebilder in kleinere Bereiche von 512 x 512 Pixel aufgeteilt. Dies geschieht, zum einen aus Speicheroptimierungsgründen, aber auch, weil das Netzwerk aufgrund des begrenzten rezeptiven Feldes in der Ausgabeschicht oder der maximalen Größe der korrodierten Bereiche nicht in der Lage ist, größere Bilder effizient zu verarbeiten. Die einzelnen Patches überlappen sich zu 25 % in jeder Richtung, da dies zu 1 bis 2 % Verbesserung der Segmentierungsergebnisse führt.

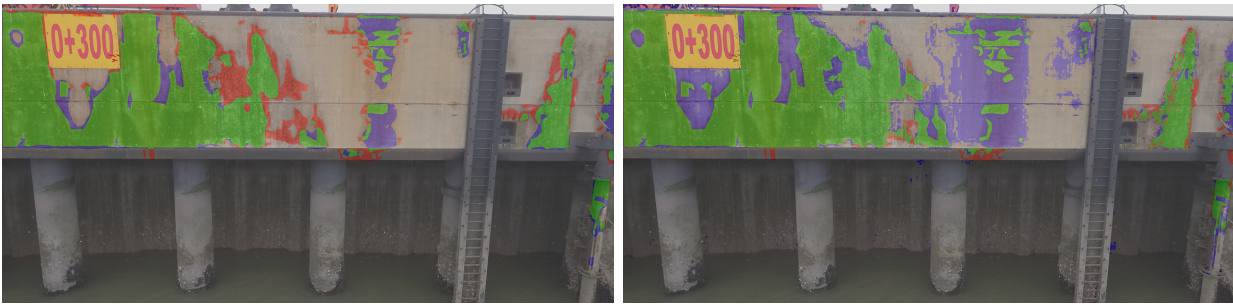
Die Ergebnisse der verschiedenen DeepLabV3+-Versionen werden im folgenden Abschnitt verglichen, und die Tabelle 2.4 wird durch das beste und das schlechteste Segmentierungsnetzwerk ergänzt, um die große Bandbreite der Ergebnisse zu demonstrieren. Die Deep Learning Modelle wurden für maximal 35 Epochen trainiert. Die Zustände der trainierbaren Parameter wurden gespeichert und dort verwendet, wo der mittlere Fehler in Bezug auf die Testmenge am kleinsten war. In Bezug auf die Gesamtunsicherheit schneidet das ResNet-Modell deutlich besser ab als die anderen Modelle, insbesondere aufgrund der Verwendung der IoU-Verlustfunktion. Bei der Betrachtung der Gesamtgenauigkeit schneidet das ResNet-Modell am besten ab, gefolgt von den Modellen MobileNet und EfficientNet mit 95,5 % bzw. 95,4 % Genauigkeit. Das VGG-16-Modell hat mit 95,2 % die geringste Genauigkeit. Die Mehrheitsklasse (Klasse 0) hat bei allen Modellen einen F1-Score von 97,4 % oder höher. Der Wert für Klasse 1 ist jedoch geringer, mit F1-Scores zwischen 68,4 % und 71,7 %. Die IoU-Werte folgen einem ähnlichen Trend, weisen aber im Allgemeinen niedrigere Werte auf.

Tabelle 2.4: *Ergebnisse der ausgewählten DeepLabV3+ Klassifikatoren.*

Metric	ResNet-50	VGG-16 BN
Total uncertainty	0,0161	0,1121
Overall accuracy [%]	96,2	95,2
F1-Score [%] (class 0; 1)	98,0; 71,7	97,4; 68,3
Pixel-wise IoU [%] (class 0; 1)	96,0; 55,9	95,0; 51,9

Abbildung 2.6 zeigt ein Beispielbild aus dem Datensatz mit überlagerten Klassifizierungsergebnissen beider Segmentierungsansätze. Die korrekte Klassifizierung von Korrosion ist grün eingefärbt, blau ist false-positive (Korrosion wurde klassifiziert, obwohl keine vorhanden ist) und rot zeigt false-negative Ergebnisse (Korrosion wurde nicht erkannt). Im Vergleich der Abbildungen 2.6a und 2.6b wird deutlich, dass die mit Random Forest erzeugten Segmente deutlich größer sind, was zu mehr false positive Ergebnissen führt. Die in Abbildungen 2.6a dargestellten Segmentierungsergebnisse der DeepLabV3+-Methode erzeugen eher kleinere Segmente, was an den Rändern zu mehr false negative Bereichen führt. Im Kontext der digitalen Bauwerksprüfung sind nicht erkannte Bereiche an den Rändern von erkannten Schäden allerdings unkritisch, da der Bereich durch einen Bauwerksprüfer aufgrund der erkannten Schäden kontrolliert wird.

Um die trainierten Modelle zu validieren, wurde die Trainingsmenge in fünf Teilmengen unterteilt und die Methode der Kreuzvalidierung angewandt. Die Ergebnisse für das beste Modell sind in



(a) Segmentierte Bilder von ResNet-50 DeepLabV3+ (b) Segmentierte Bilder von Random Forest Klassifikator basierend auf RGB-Merkmalen.

Abbildung 2.6: Die korrekte Klassifizierung von Korrosion ist grün eingefärbt, blau sind falsch-positive (klassifizierte Korrosion, wo keine vorhanden ist) und rot falsch-negative (die Korrosion wurde nicht erkannt) (Hake et al., 2023b).

Tabelle 2.5 dargestellt.

Tabelle 2.5: Ergebnisse der Kreuzvalidierung des RGB Random Forest und des ResNet-50 DeepLabV3+ Modells.

Metric	RGB Random Forest		ResNet-50 DeepLabV3+	
	Mean	Standard dev.	Mean	Standard dev.
Total uncertainty	0,2519	0,0084	0,0181	0,0015
Overall accuracy [%]	94,1	0,3	95,8	0,3
F1-Score [%] (class 0; 1)	96,8; 59,7	0,2; 1,0	97,8; 69,4	0,2; 0,5
Pixel-wise IoU [%] (class 0; 1)	93,8; 42,5	0,3; 1,0	95,4; 53,2	0,2; 0,6

Die Ergebnisse der Kreuzvalidierung zeigen, dass sowohl das Random-Forest- als auch das DeepLabV3+-Modell insgesamt ähnlich gut abschneiden. Das DeepLabV3+-Modell zeigt jedoch eine signifikante Steigerung der Erkennungsrate von Korrosion (2,7 % Steigerung) im Vergleich zum Random Forest-Modell. Darüber hinaus sind die Ergebnisse des DeepLabV3+-Modells über die fünf verschiedenen Modelle hinweg konsistenter, wie die niedrigeren Standardabweichungen zeigen. Insbesondere variieren die Werte für Klasse 1 für das Random-Forest-Modell um den Faktor zwei, und seine Vorhersageunsicherheit ist im Vergleich zu den trainierten DeepLabV3-Modellen fast sechsmal höher. Dies beweist die Überlegenheit der selbst erlernten Merkmale des Deep Learning-Ansatzes.

Für das Training der schwach überwachten Labelbilder wird die Kombination aus DeepLabV3+-Modell und Verlustfunktion verwendet, welche die besten Ergebnisse mit den manuell erzeugten Labeln erzielt hat. Dies geschieht in einem iterativen Prozess, wobei die Trainings- und Validierungs-Sets durch die mit einem Potts-Modell nachbearbeiteten Modellvorhersagen erstellt werden. Das iterative Training wird abgebrochen, sobald sich der IoU-Score zur vorherigen Epoche nicht mehr signifikant verbessert. Die ersten Modelle erreichen nur wenig überlappende Segmente und erreichen zum Beispiel einen IoU-Score von etwa 30 % für die Korrosionsklasse. Die Modelle in den folgenden Iterationen profitieren von einheitlicheren Segmenten, welche durch einen CRF-Ansatz nachbearbeitet werden, wodurch die Schadenssegmente schrittweise verfeinert werden. Durch dieses gesamte Verfahren wird verhindert, dass das endgültige Modell selbst erzeugtes Labelrauschen lernt. Zur Überprüfung des Einflusses der schwach überwachten Labelbilder werden verschiedene Anteile an überwachten und schwach überwachten Labeln getestet. In Abbildung 2.7 sind die Ergebnisse der verschiedenen Anteile schwach überwachter Label und die Modellverbesserung über die Iterationen dargestellt.

Tabelle 2.6 zeigt die Ergebnisse des DeepLabV3+-Modells, das mit 70 % schwach überwachten

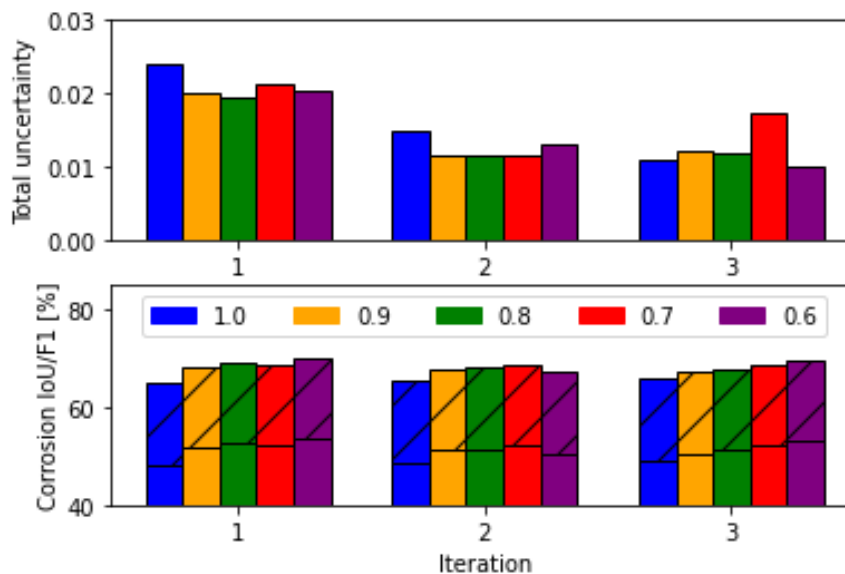


Abbildung 2.7: Gesamtunsicherheiten und korrosionsklassenbezogene Metriken von optimalen DeepLabV3+-Modell in Bezug auf den Testsatz, trainiert auf schwach überwachten Daten. Die verschiedenen Farben stehen für den Anteil der verwendeten schwachen Labels. Die schraffierten Balkenabschnitte zeigen den Unterschied zwischen IoU- und F1-Score (Scherff et al., 2023).

Labelbildern trainiert wurde.

Tabelle 2.6: Ergebnisse des DeepLabV3+-Segmentierungsnetzwerks, trainiert auf 70 % schwach überwachter Bilder in Bezug auf die Testmenge.

Total uncertainty	0,02248
Total Accuracy [%]	95,2
F1-Score [%] (class 0; 1)	97,4; 68,5
Pixel-wise IoU [%] (class 0; 1)	94,9; 52,1

Der Einsatz von schwach überwachter Labelbilder, welche mit dem Region-Growing-Ansatz erstellt wurden, kann den manuellen Aufwand der Erstellung der Labelbilder erheblich verringern. Ein Drittel manuell erzeugter Labelbilder ist ausreichend, um etwa 90 % der vollständig überwachten Erkennungsfähigkeiten zu erreichen.

Ein Aspekt, der in diesen Arbeiten nicht berücksichtigt wurde, ist die Einbeziehung der Unsicherheiten der Labelbilderstellung sowie die Vorhersage von Unsicherheiten für die erzeugten Segmentierungsergebnisse. So könnten zum Beispiel für die manuell erzeugten Label Klassen hinzugefügt werden, wie sicher sich die Person über den Schaden an dieser Stelle ist, da manche Schäden selbst für Experten nicht eindeutig auf Bildern zu identifizieren sind. Automatisiert erzeugte, schwach überwachte Label könnten so auch eine höhere Unsicherheit als manuell erzeugte Label bekommen. Auch für die erzeugten Segmentierungsergebnisse könnte eine Unsicherheit bestimmt werden. So lassen sich die gefundenen Schäden im Verdachtsplan in weitere Klassen unterteilen, wie sicher sich der Algorithmus mit der Erkennung ist. Die Prädiktion der Unsicherheiten als weitere Metrik für die digitale Bauwerksprüfung wurde in Abschnitt 3.3 untersucht.

3 Weiterführende Betrachtungen

In diesem Kapitel werden weiterführende Betrachtungen und Analysen vorgestellt, die auf den Erkenntnissen aus den Publikationen A bis D (siehe Abschnitt 2.1 & 2.2) aufbauen. Ziel ist es, die Methoden zur Schadenserkenkung weiter zu verfeinern und neue Ansätze zu evaluieren.

Dazu zählen weitere Untersuchungen zur Erstellung von Höhenfeldern aus 3D-Punktwolken von Hafeninfrastrukturen (Abschnitt 3.1), eine genauere Betrachtung des Einflusses von schwach überwachten Labelbildern (Abschnitt 3.2) sowie die Analyse der Unsicherheiten in den Segmentierungsergebnissen (Abschnitt 3.3). In den Publikationen A und B (siehe Abschnitt 2.1) wurden zur Vorprozessierung der Punktwolken und Erstellung der Höhenfelder Ebenen für die Distanzberechnung genutzt. Diese haben den Nachteil, dass sie bei gekrümmten Bauwerken keine gute Approximation darstellen und somit ungewollt größere Distanzen hervorrufen. Daher werden in Abschnitt 3.1 zwei weitere Geometrien zur Erstellung der Höhenfelder untersucht.

Das in Abschnitt 2.2 vorgestellte Verfahren zur Reduzierung der manuellen Labelerstellung hat sich als effizient erwiesen. In Abschnitt 3.2 wird ein weiterer Ansatz zur Generierung schwach überwachter Label evaluiert, welcher in der Literatur zu besseren Ergebnissen geführt hat.

Ein bisher weniger beachteter Aspekt sind die Unsicherheiten bei der Segmentierung und Schadenserkenkung. Jeder Schritt in der Verarbeitungskette, von der Erfassung der Rohdaten bis hin zur Erstellung des segmentierten Schadensverdachtsplans, ist mit inhärenten Unsicherheiten behaftet. Dies kann die Modellauswahl bei der Höhenfelderstellung, die Wahl der Netzarchitektur oder die Erstellung der Labels sein. Die Prädiktion von Unsicherheiten für das segmentierte Ergebnis stellt einen Mehrwert für die Bauwerksprüfung dar, indem die Bauwerksprüfer für jeden erkannten Schaden auch einen Wert erhalten, der angibt, mit welcher Zuverlässigkeit die Erkennung erfolgte. Die Prädiktion der Unsicherheiten wird in Abschnitt 3.3 beispielhaft für den Bilddatensatz aus dem JadeWeserPort in Wilhelmshaven (siehe Abschnitt 1.5) beschrieben.

3.1 Höhenfelderstellung

Die Rohdaten in Form unstrukturierter 3D-Punktwolken werden so aufbereitet, dass Schäden automatisiert erkannt werden können. Die Erstellung von Höhenfeldern ist ein elementarer Schritt in der Vorprozessierung für die automatisierte Schadenserkenkung. Daher sind die Methoden zur Erstellung der Höhenfelder ein wichtiger Punkt zur Beantwortung der ersten Forschungsfrage in Abschnitt 1.3.

Im Rahmen dieser Dissertation wurden verschiedene Ansätze zur Erstellung von Höhenfeldern verfolgt, um die vorhandene Infrastruktur bestmöglich abzubilden. Geometrische Schäden in den Punktwolken, wie Abplatzungen, Risse oder Anprallschäden, können dann durch den Vergleich mit einer Referenzgeometrie erkannt werden. Die Idee dahinter ist, dass die Referenzgeometrien die Struktur und Form des Bauwerks so gut wie möglich approximieren. Sie sollen jedoch keine Schäden enthalten, so dass sie den fehlerfreien Zustand des Bauwerks widerspiegeln. Dies wird erreicht, indem Geometrien verwendet werden, die keinen hohen Komplexitätsgrad aufweisen und somit die Struktur gut approximieren, aber die Schäden generalisieren. Geometrische Schäden werden dann durch ihre Abweichung von der angenäherten Geometrie sichtbar.

In der Literatur wird bei der Oberflächenapproximation zumeist eine Unterscheidung zwischen deterministischen und stochastischen Methoden getroffen. Beispiele für stochastische Verfahren sind z. B. kleinste Quadrate Kollokation (Straub, 1996), Gaußsche Prozesse (Rasmussen, 2004) oder Kriging (Montero et al., 2015). Der Vorteil dieser stochastischen Methoden gegenüber den deter-

ministischen besteht darin, dass sie durch die Modellierung von Unsicherheiten oftmals robustere Schätzungen liefern, insbesondere in Gegenwart von verrauschten Daten oder Ausreißern. Um dem Ausreißerproblem zu begegnen, können heuristische Verfahren wie Random Sample Consensus (Fischler & Bolles, 1981) oder robuste Schätzer wie Huber oder Biber (Kanani, 2000; Huber & Ronchetti, 2009) eingesetzt werden. Allerdings sind mit diesen Verfahren auch einige Nachteile verbunden, darunter der hohe Berechnungsaufwand und die Komplexität der Modelle. Deterministische Ansätze fokussieren sich in erster Linie auf den Trend in den Daten, Beispiele dafür sind herkömmliche Polynom- und Freiformflächen wie Bézier, B-Splines und ungleichmäßige rationale B-Splines (Piegl & Tiller, 2012; Bureick et al., 2016). Sie sind in der Regel effizient zu berechnen und liefern präzise, reproduzierbare Approximationen der Oberfläche mit einem festen Parametersatz. Hierarchische Verfahren wie Multilevel B-Splines (Lee et al., 1997; Mohammadivojdan et al., 2020) liefern auch bei inhomogenen Daten und Datenlücken gute Approximationen. Ein Nachteil von B-Splines liegt in der komplexen Modellauswahl und der Bestimmung eines optimalen Gitters an Kontrollpunkten (Mohammadivojdan et al., 2020). Auf Grund der Nachteile der beschriebenen Verfahren wurden in den Publikationen A und B (siehe Abschnitt 2.1) Höhenfelder durch Schätzung einfacher Ebenen in kleinen Ausschnitten der Punktwolke erstellt. Dieser Ansatz funktioniert gut bei langgestreckten, flachen Infrastrukturen wie Betonwänden oder Fahrbahnen. Durch Imperfektionen beim Bau oder aufgrund von Alterung kann die Form jedoch stärker von einer Ebene abweichen, was zu ungewollt größeren Distanzen zwischen Modell und Punktwolke führt. Dies beeinflusst die Erkennung der Schäden, da höhere Distanzen zwischen Modell und Punktwolke die eigentlichen Schäden überdecken könnten. Noch gravierender werden die Abweichungen bei gekrümmten Bauteilen wie Säulen oder Spundwänden sein. Aus diesem Grund wurden zwei weitere Verfahren zur Erstellung der Distanzen zwischen Punktwolke und Modell in dieser Arbeit untersucht. Dabei handelt es sich um 3D-Geraden und gekrümmte Flächen. Für die Schätzung dieser Geometrien werden die Punktwolken in kleinere Abschnitte unterteilt. Diese Segmentierung der Punktwolken soll es ermöglichen, bestangepasste Referenzgeometrien zu schätzen. Dabei hat sich herausgestellt, dass die Approximation mit einer Gerade zu unsteten Ergebnissen führt, welche sich nicht so gut für die Schadenserkenkung eignen. Die Verwendung gekrümmter Flächen bietet eine gute Approximation, ist aber sehr zeitaufwendig, weshalb die Nutzung von Ebenen einen guten Kompromiss zwischen Rechenleistung und Ergebnis darstellt. Wie die Höhenfelder aus den einzelnen Geometrien erstellt werden, ist im folgenden kurz dargestellt. Für den Vergleich der Methoden wurde der Punktwolkendatensatz aus Lübeck verwendet (siehe Abschnitt 1.5).

3.1.1 3D-Gerade

Die einfachste Geometrie zur Approximation der Infrastruktur sind 3D-Geraden. Hierzu wird die Punktwolke entlang einer Achse in schmale Streifen geschnitten und jeder Streifen durch eine 3D-Gerade approximiert. Dieses Vorgehen ist beispielhaft in Abbildung 3.1 dargestellt. Die Punktwolke wird in schmale Streifen (blauer Kasten) mit der Breite Δs eingeteilt. Δs hängt von der Punktdichte des jeweiligen Datensatzes ab und sollte möglichst klein gewählt werden, wobei jedoch noch genügend Punkte für eine Geradenschätzung enthalten sind. Für den Datensatz aus Lübeck wurde Δs mit 4 cm festgelegt. Mit den Punkten innerhalb dieser Streifen wird die 3D-Gerade geschätzt (rote Linie). Abschließend werden die Punktabstände zur 3D-Geraden berechnet. Die Ausschnitte der Punktwolke überlappen sich nicht.

Die Verwendung einer 3D-Geraden zur Höhenfelderstellung bietet den Vorteil eines einfachen mathematischen Modells und ermöglicht die Approximation von Rundungen, Knicken orthogonal zur Geradenrichtung sowie leichten Verdrehungen der Struktur. Nachteilig ist jedoch, dass die Verwendung einer 3D-Geraden zu einem unsteten Höhenfeld führt, das sich weniger gut für die Schadenserkenkung eignet. Der F1-Score der Segmentierung ist bei Höhenfeldern aus dem Datensatz aus Lübeck, die mithilfe von 3D-Geraden erstellt wurden, 5 % niedriger als bei Höhenfeldern, die mit Ebenen erzeugt wurden. Dieses schlechtere Ergebnis und das unstete Höhenfeld sind darauf zurückzuführen, dass für die Geradenapproximation weniger Punkte als für eine Ebene verwendet werden.

Dadurch können Ausreißer oder Punkte mit größerer Distanz die gesamte Linie verschieben.

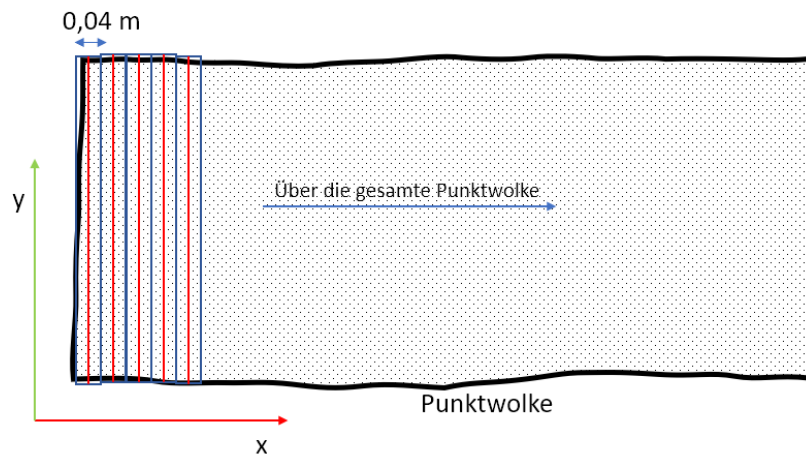


Abbildung 3.1: Vorgehen zur Generierung der Höhenfelder aus mehreren 3D-Geraden als Referenzgeometrien (Lippmann, 2021).

Die unsteten Ergebnisse im Höhenfeld durch die Erstellung mittels einer 3D-Gerade können durch größere Geometrien wie Ebenen oder gekrümmte Flächen reduziert werden. Diese werden in den nächsten beiden Abschnitten beschrieben.

3.1.2 Ebene

Sollen größere Bereiche der Punktwolke approximiert werden, bieten sich Ebenen an, da sie im Gegensatz zur 3D-Geraden weniger anfällig für Ausreißer sind und daher einheitlichere Ergebnisse erzielen. Die Punktwolke wird in Quadrate mit einer Breite Δs . Die Größe von Δs hängt von der Punktdichte und der Größe der zu erwartenden Schäden ab. Liegen große Schäden vor und wird Δs zu klein gewählt, kann es passieren, dass die gesamte Ebene innerhalb des Schadens liegt und die berechneten Distanzen zu klein sind. Wird Δs zu groß gewählt, können die Abweichungen der Geometrie zur Ebene auf Grund von Krümmungen und Rundungen im Bauwerk größer werden, was zu einer Verzerrung der Schadenserkenkung führen kann. In dieser Arbeit wurden verschiedene Größen für Δs zwischen 0,1 und 1 m untersucht, wobei die besten Ergebnisse beim Datensatz aus Lübeck für die Schadenserkenkung bei einer Größe von 0,5 m erzielt wurden. Zusätzlich wird eine Überlappung der Quadrate von 50 % in X- und Y-Richtung eingeführt. Dies soll bewirken, dass Schäden möglichst komplett in einem Ausschnitt liegen und nicht abgeschnitten werden. Für die Punkte eines Quadrats wird eine Ebenenschätzung nach Drixler (1993) durchgeführt und die Punktabstände zur geschätzten Ebene berechnet. Die Ebenenschätzung nach Drixler (1993) wurde gewählt, da sie keine Näherungswerte für die Parameter benötigt und die Parameter der Ebene so geschätzt werden, dass die Summe der quadratischen Abstände der Punkte zur Ebene minimiert wird. Das Vorgehen ist beispielhaft in Abbildung 3.2 dargestellt.

Der Vorteil bei der Verwendung von Ebenen für die Höhenfelderstellung ist neben dem einfachen mathematischen Modell und der durch Drixler (1993) effizienten Schätzung der Ebenenparameter die gute Approximation größerer, ebener Bereiche und damit der Erzeugung von gleichmäßigen Höhenfeldern. Nachteilig ist jedoch, dass Krümmungen, Sprünge oder Verdrehungen des Bauwerkes nicht approximiert werden können und bei solchen Strukturen größere Abweichungen auftreten, welche das Ergebnis der Schadenserkenkung negativ beeinflussen können. Für diese Bauwerke eignen sich gekrümmte Flächen besser zur Erzeugung der Höhenfelder, welche im folgenden Abschnitt beschrieben werden.

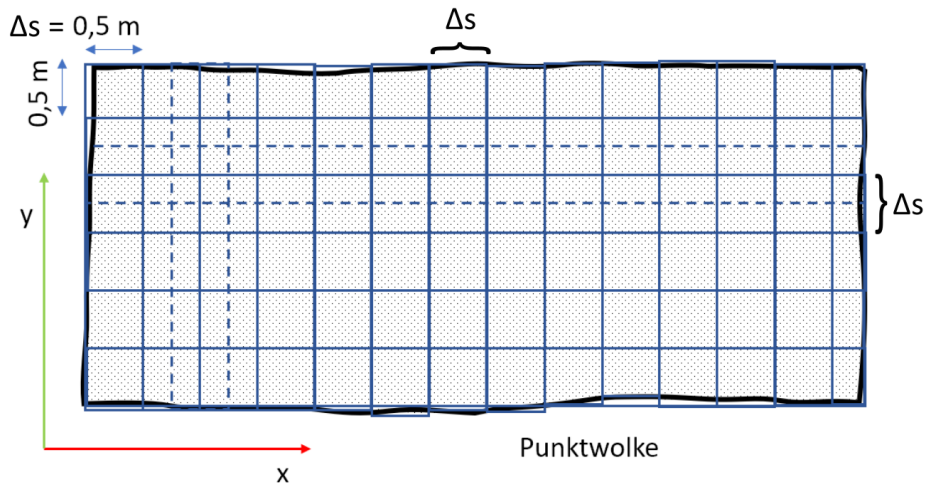


Abbildung 3.2: Vorgehen zur Generierung der Höhenfelder mit quadratischen Ausschnitten der Punktwolken (Lippmann, 2021).

3.1.3 Gekrümmte Fläche

Gerade ältere Verkehrswasserbauwerke lassen sich aufgrund ihrer Form nicht ausreichend durch Geraden oder Ebenen approximieren. Auch für Spundwände oder gekrümmte Bauteile wie Pfeiler ist die Approximation mit einer Ebene unpassend. Daher können für die quadratischen Ausschnitte der Punktwolke auch gekrümmte Flächen geschätzt und für die Abstandsberechnung verwendet werden. Dies kann zum Beispiel nach Dreßler (2006) mittels Quadriken gelöst werden. Durch den Normalenvektor eines Punktes werden alle umgebenden Punkte in ein lokales Koordinatensystem $(u, v, h(u, v))$ transformiert. Über eine Gauß-Markov-Ausgleichung wird eine zweifach gekrümmte Fläche durch die Punkte geschätzt. Die gekrümmte Fläche wird durch Gleichung (3.1) beschrieben.

$$h(u, v) = au^2 + buv + cv^2 + du + ev + f \quad (3.1)$$

wobei a, b, c, d, e, f die Koeffizienten der gekrümmten Fläche sind, u, v einen parametrisierten Punkt beschreiben und $h(u, v)$ einen Punkt auf der Oberfläche der gekrümmten Fläche darstellt.

Vorteil der Verwendung von gekrümmten Flächen zur Erstellung der Höhenfelder ist, dass Rundungen, Verdrehungen und Krümmungen des Bauwerks gut durch das mathematische Modell approximiert werden können. Nachteilig ist die komplexere mathematische Beschreibung und der höhere Rechenaufwand in der Schätzung der Flächen.

Für den verwendeten Punktwolkendatensatz aus Lübeck (siehe Abschnitt 1.5) zeigten die mit gekrümmten Flächen erzeugten Höhenfelder keinen signifikanten Unterschied zu den mit Ebenen erzeugten Höhenfeldern. Die Ausschnitte der Punktwolke haben hier eine Breite von 0,5 m. Da der verwendete Datensatz eine Betonwand darstellt, welche wenig gekrümmt ist, kommt die Stärke der gekrümmten Flächen vermutlich nicht zum Tragen und eine Approximation mit Ebenen ist ausreichend. Gerade bei Spundwänden oder bei älteren Bauwerken, welche häufig gekrümmte Bereiche aufweisen, ist aber mit einer deutlichen Verbesserung zu rechnen, da die Struktur genauer approximiert werden kann.

3.1.4 Vergleich der Verfahren der Höhenfelderstellung

In der Gegenüberstellung der drei Geometrien zur Generierung von Höhenfeldern hat sich gezeigt, dass die Verwendung von 3D-Geraden (siehe Abschnitt 3.1.1) eine einfache und effiziente Berechnung ermöglicht, jedoch im Resultat zu einem unsteten Höhenfeld führt. Die Schätzung der 3D-Gerade basiert auf einer geringen Anzahl von Punkten, wodurch einzelne Ausreißer eine signi-

fikante Beeinflussung der Schätzung verursachen können. Es besteht jedoch die Möglichkeit, dass die einzelnen 3D-Geraden dennoch Sprünge zueinander aufweisen, was zu Unstetigkeiten im Höhenfeld führt.

Ein Vergleich der Ergebnisse für die Ebene (siehe Abschnitt 3.1.2) mit denen für die gekrümmte Fläche (siehe Abschnitt 3.1.3) zeigt, dass die gekrümmten Flächen im vorliegenden Datensatz aus dem Stadthafen von Lübeck (siehe Abschnitt (1.5)) keine Verbesserung der Ergebnisse bewirken, wie man sie eigentlich erwarten würde. Der Grund hierfür ist, dass der Datensatz eine Betonkaimauer darstellt, die eine sehr ebene Oberfläche aufweist. Folglich erzeugt auch die Schätzung gekrümmter Flächen an diesem Beispiel eine sehr ebene Geometrie. Die Stärken des Verfahrens, durch gekrümmte Geometrien Rundungen und Verwindungen der Struktur zu approximieren, kommen hier nicht zum Tragen.

Die Vor- und Nachteile der vorgestellten Geometrien sind in folgender Tabelle 3.1 noch einmal zusammenfassend dargestellt.

Tabelle 3.1: Vergleich der verschiedenen Geometrien zur Höhenfelderstellung.

Methode	Vorteil	Nachteil
3D-Gerade	einfache Berechnung	unstetes Höhenfeld
Ebene	gute Approximation ebener Bereiche	nicht für gekrümmte Strukturen geeignet
Gekrümmte Fläche	gute Approximation gekrümmter und verdrehter Bereiche	hohe Komplexität

3.2 Segmentierung mit schwachen Labeln

Die Erstellung von manuellen Labelbildern, welche für das Training und die Evaluierung von automatisierten Segmentierungsverfahren genutzt werden, ist ein elementarer, aber sehr zeitaufwendiger Prozess. Im Zuge der Effizienzsteigerung durch eine digitale Bauwerksprüfung wurde in den Publikationen C und D (siehe Abschnitt 2.2) ein Verfahren untersucht, mit dem sich die für das Training wichtigen Labelbilder teilweise automatisiert erzeugen lassen. Dadurch kann der Aufwand der manuellen Erstellung der Labelbilder erheblich reduziert werden, was zu einer Reduzierung des Zeit- und Kosteneinsatzes führt und somit die Wirtschaftlichkeit und Effizienz direkt steigert. In diesem Abschnitt wird eine weitere Methode zur automatisierten Erzeugung von Labelbildern untersucht. Bei der Segmentierung mit schwachen Labeln werden die Label für das Training teilweise automatisiert erstellt. Dies senkt den manuellen Aufwand der Labelerzeugung enorm, wodurch meist mehr Daten für das Training verwendet werden können und somit die Leistung der Segmentierung gesteigert werden kann. Neben dem in Publikation D (siehe Abschnitt 2.2) beschriebenen Region-Growing-Ansatz wurde auch die „Simple Does It Methode“ (Khoreva et al., 2017) evaluiert, da sie in der Literatur mit guten Ergebnissen beschrieben wurde. Diese Methode kombiniert zwei Verfahren zur Erstellung von Labelbildern: GrabCut+ und Multiscale Combinatorial Grouping (MCG). MCG ist ein leistungsfähiger Algorithmus, der darauf abzielt, Bilder auf der Grundlage mehrstufiger Analyse zu segmentieren. Der Prozess beginnt mit der Erstellung einer Bildpyramide, bei der ein Bild auf verschiedene Auflösungen transformiert wird. Diese mehrskalige Darstellung ermöglicht es, bedeutende Strukturen des Bildes über verschiedene Ebenen hinweg zu erfassen. In jedem Maßstab werden relevante Konturen erkannt und als potenzielle Objektkanten identifiziert. Hierbei findet ein spezieller Graphcut-Algorithmus Anwendung, der das Bild als Graph modelliert. In diesem Graphen repräsentieren die Knoten die Pixel, während die Kanten die Verbindungen zwischen diesen darstellen. Der Graphcut-Algorithmus partitioniert den Graphen so, dass zusammenhängende Objekte zuverlässig isoliert werden können. Die ermittelten Segmente werden im Anschluss auf die

ursprüngliche Bildgröße zurückgeführt und zueinander ausgerichtet. Ein binärer Begrenzungsklassifikator führt die Umrisse dieser Segmente zusammen, um konsistente und präzise Segmentierungen zu erzielen (Pont-Tuset et al., 2017). Diese Segmente werden letztlich mit vorhandenen Begrenzungsrahmen verglichen, wodurch relevante Objekt- oder Schadenlabels abgeleitet werden. Diese Begrenzungsrahmen werden hier aufgrund der vorhandenen Label aus den segmentierten Bildern abgeleitet. Für neue Datensätze müsste nur ein Rechteck um die Schäden gezogen werden, was im Vergleich zur präzisen Labelerstellung deutlich weniger Zeit in Anspruch nimmt.

GrabCut+ ist eine erweiterte Version der ursprünglichen GrabCut-Methode, die ursprünglich als interaktive Technik zur Vorder- und Hintergrundsegmentierung entwickelt wurde. Zu Beginn erfordert GrabCut eine grobe manuelle Markierung des Zielobjekts innerhalb eines umgebenden Begrenzungsrahmens. Dieser Rahmen dient als anfängliche Grundlage für die Klassifizierung des Vordergrundes. Der iterative GraphCut-Prozess analysiert dann das Bild, indem er Pixel basierend auf ihrer Farbe und ihrem Kontext entweder als Vordergrund oder als Hintergrund klassifiziert. Die Erweiterung GrabCut+ integriert zusätzlich sogenannte Holistically-nested edges (HED), die von einem CNN bereitgestellt werden und als zusätzliche Kanteninformationen dienen. Diese Kanteninformationen erhöhen die Genauigkeit der Trennung erheblich, indem sie strukturelle Details des Bildes stärker betonen. Folglich werden die Bildbereiche verfeinert klassifiziert, wobei die Schadensbereiche dem Vordergrund zugeordnet werden.

In der „Simple Does It Methode“ wird zur Erstellung von präzisen schwachen Labels eine gezielte Kombination der komplementären Fähigkeiten der MCG und GrabCut+ Methodik genutzt. Der Prozess beginnt mit der initialen Segmentierung eines Bildes unter Anwendung des MCG-Algorithmus. Nach dieser anfänglichen Segmentierung durch MCG wird das Verfahren durch den Einsatz von GrabCut+ verfeinert. Die endgültige Erstellung der schwachen Label erfolgt durch die Konsolidierung der Ergebnisse beider Verfahren. Hierbei wird ein Bereich nur dann als spezifisches Objekt oder Schaden klassifiziert, wenn beide Methoden, MCG und GrabCut+, übereinstimmende Ergebnisse hinsichtlich dieser Klassifizierung liefern. Diese Schnittmengenbildung zwischen den Ergebnissen der beiden Algorithmen gewährleistet, dass die generierten Labels sowohl robust als auch akkurat sind, da sie auf der doppelten Validierung basieren.

Die „Simple Does It Methode“ erzielt jedoch im Ergebnisse einen etwa 0,5 % geringeren F1-Score als der Region-Growing-Ansatz aus Abschnitt 2.2. Der IoU-Score liegt knapp 1 % unter dem Wert vom Region-Growing-Ansatz. Aufgrund dieser Ergebnisse und der Tatsache der höheren Komplexität der „Simple Does It Methode“ im Vergleich zum einfacheren Region-Growing wurde letztere für die Erstellung der schwach überwachten Labelbilder verwendet.

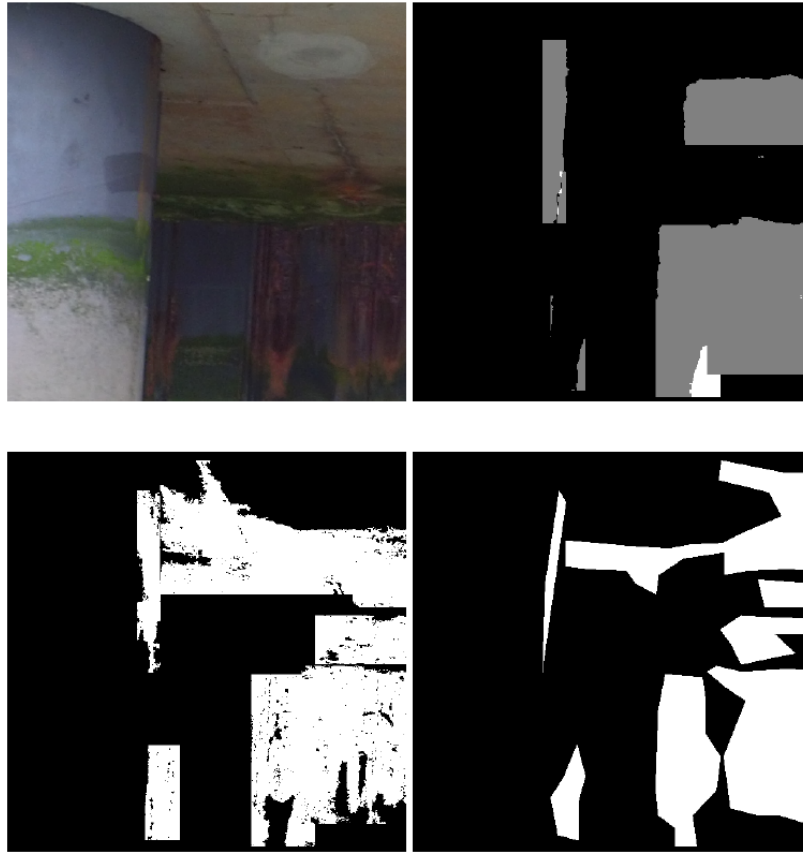


Abbildung 3.3: *Beispielsergebnisse der schwachen Label. Oben Links das Originalbild, oben rechts der Simple Does It Ansatz, unten links der Region-Growing-Ansatz und unten rechtes die händisch erzeugten Label. (Scherff, 2022)*

3.3 Unsicherheit der Segmentierung

Die vorgestellten Schadenserkennungsverfahren für geometrische Schäden aus Punktwolken (siehe Abschnitt 1.4, 2.1 und 3.1) sowie für Schäden mit einer Farbänderung aus Bildern (siehe Abschnitt 1.4, 2.2 und 3.2) nutzten angepasste Varianten des VGG19- sowie DeepLabV3+-Netzes. Das Ziel lag bisher rein in der Erkennung von Schäden, das heißt der Segmentierung der Daten in Schadensflächen und unbeschädigte Bereiche. Nicht beachtet wurden dabei bisher die Unsicherheiten. Diese können jedoch gerade im Kontext der digitalen Bauwerksprüfung einen besonderen Mehrwert bieten. Werden neben den potenziellen Schadensflächen auch die Unsicherheiten der Segmentierung prädiziert, können diese von den Bauwerksprüfern verwendet werden. So könnten Schäden, bei denen sich der Algorithmus sehr sicher ist, das heißt, wo die Unsicherheit sehr klein ist, mit weniger manuellem Aufwand kontrolliert werden als bei Schäden mit einer hohen Unsicherheit, also wo der Algorithmus sich unsicher über das Ergebnis ist. Somit fließt die Prädiktion der Unsicherheiten in die beiden in Abschnitt 1.3 beschriebenen Forschungsfragen ein, in dem die Effizienz und Zuverlässigkeit der digitalen Bauwerksprüfung gesteigert werden kann. In diesem Abschnitt wird am Beispiel des Bilddatensatzes aus dem JadeWeserPort in Wilhelmshaven exemplarisch die Prädiktion der Unsicherheiten für das Segmentierungsergebnis dargestellt.

Die Unsicherheit bezieht sich im Falle der Segmentierung also auf das Vertrauen in die einzelnen Ergebnisse des maschinellen Lernens. Da kein Algorithmus absolute Sicherheit über seine Ergebnisse garantieren kann, ist die präzise Abschätzung und Interpretation von Unsicherheiten von entscheidender Bedeutung. Unsicherheiten können in jedem Schritt der Prozesskette auftreten: Die Wahl der Geometrie für die Höhenfelderstellung (siehe Abschnitt 3.1) kann suboptimal sein, die Netzarchitektur (siehe Abschnitt 1.4) kann den funktionalen Zusammenhang nur teilweise abbilden,

oder die erstellten Label (siehe Abschnitt 3.2) können inkonsistent sein. Für eine sichere, fundierte und zuverlässige Beurteilung des Bauwerkszustandes müssen Modelle nicht nur Ergebnisse liefern, sondern auch den Grad ihres Vertrauens in diese Ergebnisse möglichst genau quantifizieren. Die Unsicherheit von Verfahren des maschinellen Lernens werden nach Matthies (2007) typischerweise in die beiden Arten epistemische und aleatorische Unsicherheit unterschieden.

Die aleatorische Unsicherheit, auch als stochastische Unsicherheit oder zufällige Abweichung benannt, ist die Unsicherheit, die durch die Zufälligkeit eines Ereignisses entsteht. Beispielhaft für ein aleatorisches Ereignis sei hier das Werfen einer Münze genannt. Im Kontext des maschinellen Lernens bezieht sich die aleatorische Unsicherheit normalerweise auf die Zufälligkeit der Eingabedaten, die durch verschiedenen Faktoren beeinflusst werden kann. Im Fall von Farbbildern können dies beispielsweise wechselnde Lichtverhältnisse oder allgemeines Sensorrauschen sein. Folglich können die Trainingsdaten die Beziehung zwischen Eingabe und Ziel möglicherweise nicht genau genug abbilden, was zu einer geringeren Modellgenauigkeit führt. Daher führt eine größere Menge an Trainingsdaten auch nicht zu einer Reduzierung der aleatorischen Unsicherheit.

Die epistemische Unsicherheit hingegen beschreibt systematische Effekte, die theoretisch bekannt sein könnten, in der Praxis jedoch nicht bekannt sind. Dies kann beispielsweise auftreten, wenn das Modell bestimmte Effekte vernachlässigt oder ein vorhandener Offset nicht berücksichtigt wird. Die epistemische Unsicherheit lässt sich somit auf die Modellunsicherheit beziehen. Sie entsteht also zum Beispiel aus unzureichenden Trainingsdaten oder zu starken Modellvereinfachungen. Anders als die aleatorische Unsicherheit lässt sich die epistemische Unsicherheit durch eine größere Menge an Trainingsdaten reduzieren. Daher ist sie auch stark von der Qualität der Trainingsdaten abhängig. Durch die Analyse und Quantifizierung dieser Unsicherheiten kann die Effizienz und Zuverlässigkeit der digitalen Bauwerksprüfung weiter erhöht werden. Die quantifizierten Unsicherheiten könnten zum Beispiel in Form der Ampelfarben Grün für Segmente, die mit hoher Sicherheit erkannt wurden, Gelb für Segmente, die mit mittlerer Unsicherheit erkannt wurden und Rot für Segmente mit einer hohen Unsicherheit angegeben werden. Somit ist für die Bauwerksprüfer auf einen Blick ersichtlich, welche Schäden genauer händisch untersucht werden müssen.

Für die Prädiktion von Unsicherheiten in Machine Learning Verfahren gelten Bayesian Neural Networks (Blundell et al., 2015) als State of the Art. Bei diesen Ansätzen werden gemäß der Bayes'schen Theorie (Bernardo & Smith, 2009) a-priori Verteilungen für die Parameter eines neuronalen Netzes festgelegt und mithilfe der Trainingsdaten die a-posteriori Verteilung berechnet. Diese wird am Ende zur Quantifizierung der Unsicherheiten verwendet. Die Berechnung der exakten posterioren Wahrscheinlichkeitsverteilung erfordert jedoch umfangreiche Modifikationen im Trainingsverfahren und ist für Deep-Learning-Modelle aufgrund der hohen Parameteranzahl und des damit verbundenen hohen Rechenaufwands nicht praktikabel (Lakshminarayanan et al., 2017; Blundell et al., 2015; Loquercio et al., 2020).

Da die exakte Berechnung nicht möglich ist, wurden verschiedene Approximationsmethoden entwickelt, um die Unsicherheiten bestmöglich abzuschätzen. Eine Studie von Nash et al. (2022) befasste sich mit der Erkennung von Korrosion unter Berücksichtigung von Unsicherheiten unter Anwendung der Bayes'schen Variational Inference Methoden. Hierbei werden einige Gewichte des Netzwerkes durch Verteilungen, oft Gauss'sche Verteilungen, ersetzt, die durch ihren Mittelwert und ihre Standardabweichung charakterisiert sind. Anschließend werden die Daten mehrfach durch das Bayes'sche neuronale Netzwerk geleitet, wobei die Netzwerkgewichte jedes Mal aus ihrer Gauss'schen Verteilung gezogen werden.

Eine andere Approximationsmethode zur Unsicherheitsquantifizierung sind Deep Ensembles, die aus einem Ensemble von trainierten Modellen bestehen, welche unterschiedliche Vorhersagen erzeugen (Lakshminarayanan et al., 2017). Die Zufälligkeit wird dabei durch verschiedene Gewichtsinitialisierung oder unterschiedliche Datenaugmentationstechniken erzeugt (Fort et al., 2020). Deep Ensemble-Methoden können andere Methoden bei verschiedenen Aufgaben zur Quantifizierung von Unsicherheiten übertreffen (Gustafsson et al., 2020; Ovadia et al., 2019; Wursthorn et al., 2022). Allerdings ist dieser Leistungsgewinn mit hohen Rechenkosten und umfangreichen Modellanpassungen

verbunden.

Zusätzlich zu den erwähnten approximierenden Methoden gibt es ein wachsendes Interesse an deterministischen Ansätzen mit einem einzigen Vorwärtsdurchlauf, die Vorteile hinsichtlich Speicherverbrauch und Inferenzzeit bieten. So wird z. B. das Konzept der entfernungsabhängigen Ausgabeschichten in (Van Amersfoort et al., 2020; Liu et al., 2020) untersucht. Diese Methoden zeigen zwar eine gute Leistung, sind aber nicht mit dem aktuellen Stand der Technik konkurrenzfähig und erfordern erhebliche Änderungen am Trainingsprozess (Mukhoti et al., 2023). Ein weiterer von Mukhoti et al. (2023) vorgeschlagener Ansatz vereinfacht die beiden vorherigen Methoden, indem er die Gaußsche Diskriminanzanalyse für die Schätzung der Merkmalsraumdichte nach dem Training einsetzt. Obwohl sie in einigen Situationen mit Deep Ensembles gleichziehen, erfordert ihr Ansatz immer noch einen ausgefeilteren Trainingsansatz. Im Gegensatz dazu schlägt Valdenegro-Toro (2023) eine einfache, aber effektive Annäherung an Deep Ensembles vor, bei der das Ensemble nur eine Teilmenge von Schichten anstelle des gesamten Modells abdeckt. Diese sogenannten Deep Sub-Ensembles (DSE) ermöglichen einen Kompromiss zwischen Unsicherheitsqualität und Rechenkosten. Andere Ansätze nutzen Markov Chain Monte Carlo Methoden (Neal, 1995), variational Bayesian Methoden (Graves, 2011) oder Laplace Approximationsmethoden (MacKay, 1992).

Eine weitere Möglichkeit, Unsicherheiten in einem Machine-Learning-Algorithmus vorherzusagen, ohne größere Änderungen an der Architektur vornehmen zu müssen, ist Monte Carlo Dropout (Gal & Ghahramani, 2016). Diese Methodik hat sich für den in dieser Arbeit beschriebenen speziellen Fall als die bevorzugte Option herausgestellt, da sie einfach zu handhaben ist, minimale Auswirkungen auf den Trainingsprozess hat und im Vergleich zu Deep Ensembles rechnerisch effizient ist. Dabei werden die eigentlich zur Regularisierung während des Trainings eingeführten Dropout Layer dafür eingesetzt, um im Prädiktionsschritt Variationen in der Modellausgabe zu erzeugen. Im Trainingsschritt sorgen die Dropout Layer für eine Regularisierung und verhindern somit eine Überanpassung an die Daten. Dazu werden zufällig verschiedene Knoten des neuronalen Netzwerkes deaktiviert. Im Prädiktionsschritt werden die Dropout Layer normalerweise deaktiviert, um wiederholbare Ergebnisse der Prädiktion zu erzielen. Werden Sie jedoch aktiviert, erzeugt jeder Durchlauf leicht unterschiedliche Ergebnisse. Aus den Unterschieden können die Unsicherheiten abgeleitet werden. Ein wichtiger Parameter ist hier die Dropoutrate. Höhere Dropoutraten führen dazu, dass mehr Neuronen zufällig deaktiviert werden was zu einer größeren Streuung des Ergebnisses führen kann. Kleinere Dropoutraten hingegen verringern die Anzahl an zufällig deaktivierten Neuronen, was die Streuung der Ergebnisse zu reduzieren scheint. Die Dropoutrate sollte an die Daten und die Aufgabestellung angepasst werden (Pauls & Yoder, 2018). Die Unsicherheiten sollen für das segmentierte Ergebnis zur Erkennung von Korrosionsschäden erzeugt werden. Da hierfür das Netz schon vollständig trainiert und optimiert wurde, werden keine Anpassungen der Netzstruktur und Hyperparameter durchgeführt, sondern alle Parameter wie in den Abschnitten 2.2, 6.3 und 6.4 beschrieben verwendet.

Die von Gal & Ghahramani (2016) eingeführte Idee des Monte Carlo Dropout ist dabei eine Approximation der Verteilung eines Bayesian Neural Networks. In einem Bayesian Neural Network lässt sich die posteriori Dichte $p(y|\mathbf{x}, D)$ nach Gal & Ghahramani (2016) wie folgt präzisieren:

$$p(y|\mathbf{x}, D) = \int p(y|\mathbf{x}, \omega)p(\omega|D)d\omega \quad (3.2)$$

mit y beobachtete Ausgabe, \mathbf{x} Eingabe, D Trainingsdaten, W_i Gewichtsmatrix des Netzes und $\omega = \{W_i\}_{i=1}^{Layers}$.

Gal (2016) schlägt vor, die Entropie der Vorhersageverteilung (auch predictive entropy genannt) und mutual Information als Metriken für die Unsicherheiten zu verwenden, da sie verschiedene Aspekte der Unsicherheit beinhalten. Predictive entropy enthält die prädizierte Unsicherheit, welche eine Kombination aus aleatorischen und epistemischen Unsicherheiten darstellt. Durch Monte Carlo Dropout kann die Entropie der Vorhersageverteilung $\hat{\mathbb{H}}[y|\mathbf{x}, D]$ nach Gal & Ghahramani (2016) wie

folgt approximiert werden:

$$\hat{\mathbb{H}}[y|\mathbf{x}, D] = - \sum_c \left(\frac{1}{T} \sum_t p(y = c|\mathbf{x}, \hat{\omega}_t) \right) \log \left(\frac{1}{T} \sum_t p(y = c|\mathbf{x}, \hat{\omega}_t) \right) \quad (3.3)$$

mit y als Ausgabevariable, c umfasst alle Klassen, T ist die Anzahl der Monte Carlo Iterationen, $p(y = c|\mathbf{x}, \hat{\omega}_t)$ ist die Wahrscheinlichkeit, dass die Eingabe \mathbf{x} zur Klasse c gehören und $\hat{\omega}_t$ sind die Modellparameter der t -ten Monte Carlo Iteration. In ähnlicher Weise kann die mutual Information $\hat{\mathbb{I}}[y, \omega|\mathbf{x}, D]$ zwischen der Vorhersageverteilung und dem Posteriorwert der Modellparameter wie folgt approximiert werden:

$$\hat{\mathbb{I}}[y, \omega|\mathbf{x}, D] = \hat{\mathbb{H}}[y|\mathbf{x}, D] + \frac{1}{T} \sum_{c,t} p(y = c|\mathbf{x}, \hat{\omega}_t) \log p(y = c|\mathbf{x}, \hat{\omega}_t) \quad (3.4)$$

Die mutual Information beinhalten die epistemische Unsicherheit oder auch Modellunsicherheit. Abbildung 3.4 zeigt beispielhaft an drei Bildern aus dem Bilddatensatz aus Wilhelmshaven (siehe Abschnitte 1.5 und 2.2) die Ergebnisse der semantischen Segmentierung von Roststellen sowie die prädizierte und epistemische Unsicherheiten. Diese wurden mithilfe von 1000 Monte-Carlo-Iterationen prädiziert. Zu beachten sind hierbei die Unterschiede zwischen den mutual Information und der predictive Entropy. Die predictive Entropy zeigt auch innerhalb der erkannten Objekte höhere Unsicherheiten, während die mutual Information hauptsächlich an den Segmentgrenzen hohe Unsicherheiten aufweist. Dies lässt darauf schließen, dass die Grenzen der Segmente im Datensatz mit einem hohen Rauschen vorliegen und somit das Netzwerk diese nur mit geringerer Genauigkeit prädizieren kann.

Durch die Analyse und Quantifizierung dieser Unsicherheiten kann der Algorithmus gezielt verbessert und robuster gegenüber Variabilitäten und Unsicherheiten in den Daten gemacht werden. Dies ermöglicht eine zuverlässigere und effizientere digitale Bauwerksprüfung.

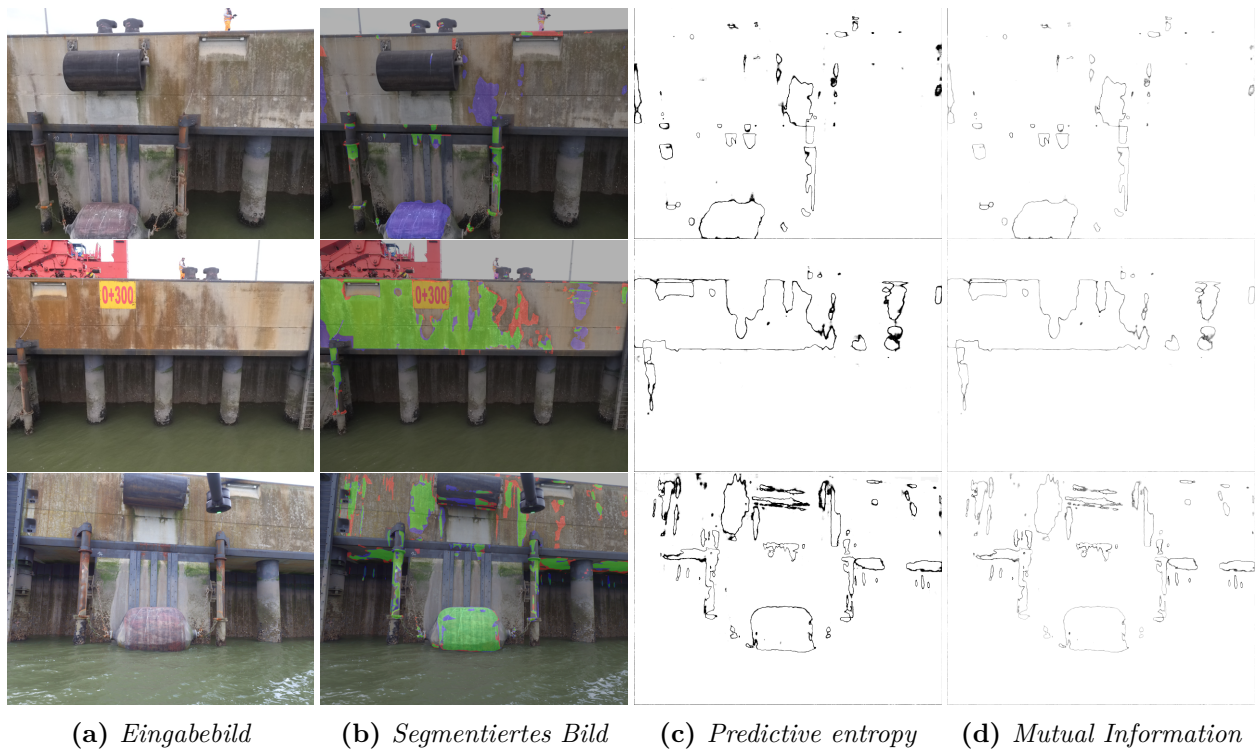


Abbildung 3.4: *Beispielergebnisse der semantischen Segmentierung zur Erkennung von Roststellen mit Prädiktion der Unsicherheit. Die Ergebnisse enthalten das Originalbild, die dazugehörige semantische Segmentierung und die prädizierte und epistemische Unsicherheiten, die durch die Entropie der Vorhersageverteilung bzw. die mutual Information geschätzt werden. Die Unsicherheiten werden als normierte Grauwerte im Bereich von 0 bis 255 dargestellt.*

4 Fazit und Ausblick

4.1 Fazit

Multisensorsysteme haben sich in der Ingenieurgeodäsie als etabliertes Verfahren zur (kinematischen)-Erfassung von Objekten, wie zum Beispiel Bauwerken, etabliert. Sie werden in vielen Bereichen wie dem Monitoring, der Bestandsdatenaufnahme sowie Überwachungsmessungen eingesetzt. Obwohl sie in vielen Bereichen der Geodäsie bereits Stand der Technik sind, stellt die automatisierte Auswertung der erhobenen (Massen)-Daten weiterhin eine offene Herausforderung dar. Dies liegt daran, dass zwischen den erfassten 3D-Punktwolken oder Farbbildern und den abzuleitenden Zielgrößen ein komplexer funktionaler Zusammenhang besteht, welcher teilweise noch nicht vollständig verstanden ist. Infolgedessen werden häufig nur kleine Teile der erfassten Daten manuell ausgewertet, wodurch das große Potenzial der verfügbaren Daten und Informationen ungenutzt bleibt. Ein Anwendungsgebiet, in dem diese Herausforderungen besonders deutlich werden, ist die digitale Bauwerksprüfung und Schadenserkennung. Da große Bereiche auf Schädigungen überprüft werden müssen, ist die manuelle Auswertung sehr zeitaufwendig und ineffizient. Zusätzlich ist die manuelle Auswertung der Daten und die Erkennung von Schäden sehr subjektiv und führt zu Ergebnissen, die weder reproduzierbar noch leicht vergleichbar sind.

Aufgrund der Komplexität dieser Problemstellung ist eine vollständige, allgemeingültige Lösung ein langfristiger Prozess. Ziel ist es, das Verständnis der Verarbeitungskette zu verbessern und geeignete Verfahren zur Erkennung und Bewertung von Schäden an Infrastrukturbauwerken zu entwickeln. Ziel der vorliegenden Dissertation ist es, durch gezielte Beiträge im Bereich der Datenvorverarbeitung und der Erkennung von einzelnen Schadensarten diesen Prozess konstruktiv voranzutreiben. Die wesentlichen Beiträge dieser Dissertation lassen sich basierend auf den Zielsetzungen aus dem Abschnitt 1.3 wie folgt zusammenfassen:

1. Automatisierte Erkennung von geometrischen Schäden

Im Rahmen dieser Dissertation wurde ein Verfahren entwickelt, um geometrische Bauwerkschäden wie Abplatzungen oder größere Risse aus unstrukturierten Punktwolken eines mobilen Scansystems zu erkennen. Dies war notwendig, da die bisherige manuelle und stichprobenartige Bauwerksüberprüfung keine einheitliche Schadenserkennung ermöglichte. Dazu wird die Punktwolke in ein Höhenfeld konvertiert und mittels einer Kombination aus CNN und LOF verarbeitet. Die vorgestellte Methode ermöglicht die Digitalisierung der Gebäudeinspektion und hebt die automatisierte Erstellung von Schadensverdachtsplänen auf ein neues Niveau.

Das vorgestellte Verfahren dient der automatischen Erstellung eines Verdachtsplans mit vermuteten Schadensregionen aus Punktwolken. Damit das Verfahren in der Praxis für die Bauwerksüberprüfung einen Mehrwert bietet, müssen möglichst alle Schadensregionen erkannt werden. Darüber hinaus sollen ausschließlich beschädigte Regionen als solche erkannt werden. Das bedeutet, dass sowohl die Genauigkeit als auch der Wiedererkennungswert möglichst hoch sein sollten. Das Verfahren wurde zunächst an simulierten Daten getestet und anschließend auf reale Daten angewandt.

Die Schadenserkennung basiert auf der Erzeugung von Höhenfeldern, die durch den Vergleich der 3D-Punktwolke mit einem geometrischen Modell des Bauwerks erstellt werden. Insbesondere bei älteren Infrastrukturbauwerken, aber auch durch Umbauten an neueren Bauwerken, fehlen häufig CAD-Modelle, die für die Erzeugung der Höhenfelder erforderlich wären. Daher

müssen für die automatisierte Schadenserkennung an diesen Bauwerken die Modelle bestmöglich aus den vorhandenen Daten geschätzt werden. Daher wurden in dieser Arbeit verschiedene Ansätze zur Vorprozessierung und der Erstellung von Geometrien untersucht. Dabei wurde lediglich eine grobe Datenbereinigung durchgeführt, ohne statistische Ausreißerererkennung oder weitgehende Datenbereinigung.

Um Höhenfelder aus den Punktwolkendaten abzuleiten, müssen diese durch mathematisch beschreibbare Geometrien bestmöglich approximiert werden. Aufgrund der Komplexität von Verkehrswasserbauwerken werden die Daten zunächst in kleinere Bereiche unterteilt. Um Probleme mit abgeschnittenen Schäden an den Rändern zu vermeiden, überlappen sich die Bereiche vertikal und horizontal zu 50 %. Für jeden Bereich wird dann eine Geometrie angenähert.

In dieser Arbeit wurden als Geometriotyp für die Schätzung eine einfache Linie, eine Ebene sowie eine gekrümmte Fläche untersucht. Linien und Ebenen sind sehr einfach zu berechnen und liefern gute Ergebnisse für wenig komplexe und neuere Strukturen, wie zum Beispiel Betonholme. Bei komplexeren oder älteren Strukturen weichen die Formen jedoch stärker von einfachen Geometrien wie Ebenen ab. Daher liefern gekrümmte Flächen in diesen Fällen eine bessere Annäherung an die tatsächliche Form des Bauwerkes.

Insgesamt konnten durch die Vorprozessierung der Daten und die Schätzung des Modells aus den Daten gute Segmentierungsergebnisse erzielt werden. Die Analyse der simulierten Daten ergab einen sehr guten F1-Score von 96,3 %. Für die realen Daten aus dem Stadthafen von Lübeck (siehe Abschnitt 1.5) sind die Ergebnisse etwas schlechter. Der F1-Score beträgt 72,4 %. Bei der Betrachtung von Schadensbeispielen in den Daten, die der Algorithmus nicht erkannt hat, zeigt sich, dass die meisten davon am Rand der Höhenfelder liegen. Die Ergebnisse ließen sich möglicherweise weiter verbessern, indem die Ränder separat behandelt werden, da diese in der Regel eine deutlich andere Verteilung als der Rest des Scans aufweisen. In Bezug auf die oben genannten Anforderungen eignet sich die Methode zur Erstellung von Verdachtsplänen von Schadensregionen an Kaimauern. Die vorgeschlagene Strategie lässt sich auch auf andere Infrastrukturobjekte, wie z. B. Brücken, Hochhäuser und Tunnel anwenden.

Die Verwendung von Netzwerken, die mit natürlichen Bildern vortrainiert wurden, um Schäden aus 3D-Punktwolken zu erkennen, ist ein neuer Ansatz im Bereich der Systeme zum Monitoring von Bauwerken. Dies ermöglicht die automatische Erkennung von Schäden und somit eine vollständige, objektive und effiziente digitale Bauwerksprüfung.

Im Vergleich zu der manuellen Methode mit Tauchern ermöglicht der Einsatz des k-MSS in Kombination mit dem hier vorgestellten Verfahren eine deutlich höhere Vollständigkeit der Bauwerksüberprüfung. Der Einsatz von Laserscannern und hydrographischen Messungen ermöglicht eine qualitätsgesicherte und reproduzierbare Kartierung der Infrastruktur. Die flächenhafte Vermessung der Bauteiloberflächen über und unter Wasser in Kombination mit der hier vorgestellten automatisierten Auswertung ermöglicht die zuverlässige Erkennung und Verifizierung von vermuteten Schäden. Dadurch wird der Vergleich verschiedener Messepochen, wie sie im Rahmen der Bauwerksprüfung alle sechs Jahre durchgeführt werden müssen, auch für Bauwerke unter Wasser ermöglicht. Dies erlaubt eine vollständigere, transparentere und effizientere Beobachtung und Bewertung der Schadensentwicklung und bildet die Grundlage für eine verbesserte Lebensdaueranalyse dieser volkswirtschaftlich wichtigen Bauwerke

2. Automatisierte Erkennung von Farbänderungen

Im Rahmen dieser Dissertation wurde ein Verfahren zur automatisierten Erkennung von Schäden, die durch Farbänderungen charakterisiert sind, aus RGB-Bildern entwickelt. Zusätzlich wurde untersucht, inwieweit sich der Aufwand für die manuelle Erstellung von Labelbildern reduzieren lässt, indem automatisch generierte, schwach überwachte Labelbilder verwendet werden. Ziel ist es, den Aufwand der manuellen Labelbilderstellung zu reduzieren, ohne die Qualität der Segmentierungsergebnisse signifikant zu beeinträchtigen.

Zur Reduktion der benötigten Labelbilder wurden zwei verschiedene Ansätze zur Erzeugung von schwach überwachten Labelbildern untersucht. Dabei wurde auch der Einfluss des Verhältnisses zwischen automatisch generierten und manuell erstellten Labelbildern auf die Segmentierungsqualität evaluiert. Dabei erwies sich die Erstellung der schwach überwachten Labelbilder mit einem Region-Growing-Algorithmus als das Verfahren mit dem besten Verhältnis zwischen Aufwand und Ergebnis.

Zur Segmentierung in beschädigte und unbeschädigte Bereiche wurden in der DeepLabV3+-Architektur verschiedene Kombinationen aus CNN als Backbone und Verlustfunktion untersucht. Nachdem die beste Kombination aus CNN-Modell und Verlustfunktion anhand der manuell erzeugten Labelbilder ermittelt wurde, wurden die schwach überwachten Labelbilder hinzugefügt. Dabei zeigte sich, dass bereits 30 % manuell erstellte Labelbilder an der Gesamtmasse ausreichend sind, um etwa 90 % der Erkennungsfähigkeit eines vollständig überwachten Modells zu erreichen. Dadurch lässt sich der Aufwand für die Erstellung der Labelbilder erheblich reduzieren, was zu einer deutlichen Zeit- und Kostenersparnis führt (siehe auch Abschnitt 2.2).

Am Beispiel von Korrosionsschäden wurde ein Deep-Learning-Ansatz vorgestellt, der vortrainierte Netzwerkstrukturen nutzt, um eine pixelgenaue Segmentierung der Bilder in beschädigte und unbeschädigte Bereiche durchzuführen. Die Analyse der Daten aus dem JadeWeserPort in Wilhelmshaven (siehe Abschnitt 1.5) zeigt eine sehr gute Erkennungsleistung für Schäden. Dabei wurde ein F1-Score von 98,0 % für die unbeschädigten Bereiche und 71,7 % für die beschädigten Bereiche erzielt (siehe auch Abschnitt 2.2).

Neben der automatisierten Schadenserkenkung, die bereits eine erhebliche Effizienzsteigerung gegenüber der manuellen Methode darstellt, kann die Effizienz der digitalen Bauwerksprüfung durch die Vorhersage von Unsicherheitsmetriken für das Segmentierungsergebnis weiter erhöht werden. So können die Bauwerksprüfer gezielt die Schadensverdachtsfälle genauer untersuchen, welche von den automatisierten Verfahren mit einer höheren Unsicherheit erkannt wurden (siehe auch Abschnitt 3.3).

4.2 Ausblick

Hinsichtlich des Gesamtprozesses der digitalen Bauwerksprüfung besteht bei der automatisierten Schadenserkenkung noch großes Entwicklungspotential. So sind die beiden in den Abschnitten 1.4, 2.1 und 2.2 vorgestellten automatisierten Schadenserkenkungsverfahren aktuell darauf beschränkt, geometrische Schäden und Schäden mit einer Farbänderung isoliert in individuellen Prozessketten zu betrachten. Ein wichtiger nächster Schritt wäre die gemeinsame Nutzung von 3D- und Farbinformationen. Dadurch könnten nicht nur geometrische Schäden besser erkannt, sondern auch weitere Schadensarten im selben Verfahren identifiziert werden. Durch die Kombination der Daten, beispielsweise in Form von eingefärbten 3D-Punktwolken, erhöht sich die Anzahl der verfügbaren Merkmale, was die Segmentierungs- und Klassifizierungsfähigkeiten verbessert. Dies ermöglicht eine präzisere Charakterisierung der einzelnen Schäden aufgrund der höheren Anzahl verfügbarer Merkmale. Dies ließe sich durch eine Einfärbung der 3D-Punktwolken und die Nutzung der Farbinformationen als zusätzliche Merkmale erreichen. Alternativ könnte die Fusion von 3D-Punktwolken und Farbbildern in einer gemeinsamen Netzwerkarchitektur die kombinierte Auswertung von Farb- und Geometriemerkmale ermöglichen.

Die Erkennbarkeit geometrischer Bauwerksschäden wird maßgeblich von der eingesetzten Messsensorik beeinflusst. Über Wasser ermöglicht der Einsatz von Laserscannern ein geringes Messrauschen im unteren einstelligen Millimeterbereich und einen Punktabstand von etwa 1 cm. Dadurch können auch kleine Schäden gut erkannt und Fehlinterpretationen reduziert werden. Unter Wasser liegt das Messrauschen im Bereich von mehreren Zentimetern, wodurch kleinere Schäden im Messrauschen untergehen und somit nur größere Schäden zuverlässig erkannt werden können. Eine Verringerung

des Messrauschens unter Wasser würde die zuverlässige Erkennung kleinerer Schäden ermöglichen. Störobjekte in den Daten, wie Anbauteile (Leitern, Fender, Poller, etc.) oder Pflanzen, können die automatisierte Schadenserkenkung ebenfalls beeinflussen, da sie zu Abweichungen im Höhenfeld führen können, obwohl sie keine Schäden darstellen. Die Berücksichtigung von Anbauteilen in zukünftigen CAD- oder BIM-Daten könnte die Probleme bei der Distanzberechnung reduzieren. Ein Bewuchs durch Pflanzen könnte mithilfe spezifischer automatischer Verfahren in einem Vorverarbeitungsschritt erkannt und in der weiteren Prozesskette berücksichtigt werden. Dadurch ließe sich der manuelle Aufwand der Vorsegmentierung reduzieren und der Gesamtprozess effizienter gestalten. Eine weitere Einschränkung der in dieser Arbeit vorgestellten Beiträge besteht darin, dass sie sich auf die reine Erkennung der Schäden beschränken. Eine anschließende Klassifizierung der einzelnen Schäden in die verschiedenen Schadensklassen würde die Effizienz und die Zuverlässigkeit der digitalen Bauwerksprüfung weiter erhöhen und stellt somit einen wichtigen Ansatzpunkt für zukünftige Forschung dar. Die erkannten Schäden könnten anhand abgeleiteter Merkmale beispielsweise mithilfe von Entscheidungsbäumen in die vier in der Bauwerksprüfung verwendeten Schadensklassen kategorisiert werden. Dadurch stünde den Bauwerksprüfern eine weitere automatisch generierte Information zur Verfügung, die die Bewertung der Infrastruktur erleichtert und den Gesamtprozess zuverlässiger und effizienter macht.

Hinsichtlich der prädizierten Unsicherheiten liegt der hier vorgestellte Fokus darauf, neben der Segmentierung eine Metrik zu erzeugen, die widerspiegelt, wie sicher sich der Algorithmus mit den erzeugten Segmenten ist. Dies soll den Bauwerksprüfern erleichtern, die Schäden zu identifizieren, die einer genaueren manuellen Überprüfung bedürfen. Die prädizierten Unsicherheiten könnten jedoch auch dazu genutzt werden, die Verfahren durch weitere Untersuchungen und Analysen robuster gegenüber Variabilitäten in den Daten und unzureichenden Modellannahmen zu machen. Beispielsweise könnten die initial prädizierten Unsicherheiten in einem zweiten Schritt als zusätzliche Eingabe für das Training verwendet werden, um die Gesamtgenauigkeit und Präzision des Netzes zu verbessern.

Die automatisierte Schadenerkennung bildet zusammen mit der vollständigen Erfassung der Infrastrukturen durch Messdaten die Grundlagen für neue, digitale Prozesse im Lifecycle-Management der Verkehrswasserinfrastrukturen. Aus den erfassten 3D-Punktwolkendaten können BIM-Modelle erstellt werden, die den aktuellen Zustand des Bauwerks abbilden. Die automatisiert erkannten Schäden können im BIM-Modell direkt den Bauteilen zugeordnet werden und sind somit für alle Beteiligten transparent und effizient für die Bewertung, Planung und Instandsetzung zugänglich. An dieser Stelle könnten weitere Arbeiten ansetzen, um die BIM-Strategien für Bundeswasserstraßen und Bundesbauten weiter voranzutreiben (BMI, 2021; BMDV, 2022a).

Durch die Verknüpfung von automatisierter Schadenerkennung und vollständiger Erfassung der Infrastrukturen mittels Messdaten werden neue, digitale Prozesse im Lifecycle-Management der Verkehrswasserinfrastrukturen ermöglicht. Die vorgestellten Verfahren und Untersuchungen der vorliegenden Dissertation tragen zur Steigerung der Effizienz, Vollständigkeit, Transparenz und Zuverlässigkeit der Bauwerksprüfung bei. Vor diesem Hintergrund leistet die Dissertation einen wichtigen Beitrag zur fortschreitenden Digitalisierung der Bauwerksprüfungen.

5 Auflistung weiterer Publikationen

In diesem Kapitel werden weitere Publikationen des Autors dieser Dissertation aufgelistet. Diese Publikationen unterscheiden sich von den relevanten Publikationen in Abschnitt 1.6 dahingehend, dass die Inhalte entweder bereits durch die relevanten Publikationen abgedeckt sind oder die Publikationen nicht in direktem Zusammenhang zu den dissertationswürdigen Inhalten stehen.

Hake, F.; Alkhatib, H & Neumann, I (2022): Automatische Schadenserkennung an Wasserbauwerken mittels hybrider Messplattform und Machine-Learning. In: Hydrographie – Messen mit allen Sinnen: Beiträge zum 204. DVW-Seminar und 35. Hydrographentag. Bd. 102/2022, DVW-Schriftenreihe, Wissner Verlag, S. 147-151.

Lippmann, P.; **Hake, F.**; Alkhatib, H. & Neumann, I. (2022): Schadenserkennung mittels Transfer-learning und Local-Outlier-Factors am Beispiel von Hafeninfrastrukturen. VDVMagazin, 73(2), 128.

Hake, F.; Alkhatib, H.; Hesse, C.; Holste, K. & Neumann, I. (2022): KI-basierte Schadensdetektion für das innovative Über- und Unterwasserscansystem 3D HydroMapper. In: D. Schermer, & E. Brehm (Hrsg.), Mauerwerk-Kalender 2022: Fassadengestaltung, Bauphysik, Innovationen.

Hake, F.; Alkhatib, H & Neumann, I (2021): Automatische Schadenserkennung bei Infrastrukturprojekten mittels Deep-Learning. In: Terrestrisches Laserscanning 2021 (TLS 2021): Beiträge zum 203. DVW-Seminar. Bd. 99/2021, DVW-Schriftenreihe, Wissner Verlag, S. 125 - 135.

Hake, F.; Hermann, M.; Alkhatib, H.; Hesse, C.; Holste, K.; Umlauf, G.; Kermarrec, G. & Neumann, I. (2020): Damage Detection for Port Infrastructure by Means of Machine-Learning-Algorithms. Beitrag in FIG Working Week 2020, Amsterdam, Niederlande.




Hesse, C.; Krause, N. L.; Frenz, M.; Neumann, I.; **Hake, F.**; & Paffenholz, J-A. (2020): Kinematisches Laserscanning zur statischen Bewertung großer Brückenbauwerke. in Tagungsband GeoMonitoring 2020, S. 65 - 78.

Hesse, C.; Holste, K.; Neumann, I.; **Hake, F.**; Alkhatib, H.; Geist, M.; Knaack, L. & Scharr, C. (2019): 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadenserkennung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. Hydrographische Nachrichten, 113, 26 - 29.

6 Eigene Veröffentlichungen

Article

Using Machine-Learning for the Damage Detection of Harbour Structures

 Frederic Hake , Leonard Göttert, Ingo Neumann  and Hamza Alkhatib 

Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany; leonard.goettert@stud.uni-hannover.de (L.G.); neumann@gih.uni-hannover.de (I.N.); alkhatib@gih.uni-hannover.de (H.A.)

* Correspondence: hake@gih.uni-hannover.de; Tel.: +49-511-762-4469

Abstract: The ageing infrastructure in ports requires regular inspection. This inspection is currently carried out manually by divers who sense the entire below-water infrastructure by hand. This process is cost-intensive as it involves a lot of time and human resources. To overcome these difficulties, we propose scanning the above and below-water port structure with a multi-sensor system, and by a fully automated process to classify the point cloud obtained into damaged and undamaged zones. We make use of simulated training data to test our approach because not enough training data with corresponding class labels are available yet. Accordingly, we build a rasterised height field of a point cloud of a sheet pile wall by subtracting a computer-aided design model. The latter is propagated through a convolutional neural network, which detects anomalies. We make use of two methods: the VGG19 deep neural network and local outlier factors. We showed that our approach can achieve a fully automated, reproducible, quality-controlled damage detection, which can analyse the whole structure instead of the sample-wise manual method with divers. We were able to achieve valuable results for our application. The accuracy of the proposed method is 98.8% following a desired recall of 95%. The proposed strategy is also applicable to other infrastructure objects, such as bridges and high-rise buildings.



check for updates

Citation: Hake, F.; Göttert, L.; Neumann, I.; Alkhatib, H. Using Machine-Learning for the Damage Detection of Harbour Structures.

Remote Sens. **2022**, *14*, 2518. <https://doi.org/10.3390/rs14112518>

Academic Editor: Fernando Vicente-Guijalba

Received: 31 March 2022

Accepted: 21 May 2022

Published: 24 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: damage detection; machine-learning; infrastructure; laserscanning; multibeam echo-sounder

1. Introduction and Motivation

The ageing infrastructure of sea and inland ports requires new technologies and methods in the preparation and implementation of life cycle management processes. The traditional processes are usually time- and labour-intensive, and should be replaced by new automated, smart and innovative measurement and analysis processes to ensure transparency, resource efficiency and reliability for a more dependable lifetime prediction.

Port infrastructure, such as quay walls for loading and unloading ships, bridges, locks and flood gates, are mostly made of concrete, bricks, steel and, in the case of very old structures, wood. They are subject to severe degradation due to especial environmental conditions and human activities throughout their lifetime. The material of seaports is especially profoundly affected by saltwater, which damages the concrete structures, sheet pile walls or wooden structures. It is crucial to detect any damage and categorize its importance to ensure the safety and stability of the infrastructure. Identifying structural damage in time allows early maintenance and avoids expensive repairs and the collapse of the infrastructure.

Nowadays, the monitoring of port infrastructural buildings is divided into the parts above and below water. The structural testing of port infrastructure above water is carried out by manual and visual inspections. The recording and documentation of the condition of damage below water involve considerably more effort; the infrastructure is tested sample-wise every 50 to 100 m; the divers slide down the structure and try to sense the wall with

their hands. The results depend directly on human sensory tests. Therefore, damage inspections below water with divers are highly variable in quality and quantity. Damage classification and development are not reproducible due to the subjective perception. In addition, there is usually no comprehensive inspection below water, thus, only a small percent of the structure can be inspected by divers. One way to deal with this problem is by utilising sensors that detect the shape of the object. Such sensors provide point clouds and include laser scanners for surfaces above water and echo-sonars for those below water. The focus in this paper is on the general process of damage detection in point clouds. We use two datasets to validate the overall procedure. The first is a simulated dataset of a sheet pile wall below water. The second is a real dataset of a concrete quay wall from the northern German city harbour of Lübeck, measured with a laser scanner above water. It is not possible to detect small damages, such as cracks, especially in the area below water, due to the point spacing, which is 20 mm. Therefore, the main focus is on the detection of spalling damages larger than 20 mm.

It is essential when monitoring harbour structures to assure a transparent, efficient and quality-controlled process. This can be achieved by a comprehensive visual inspection at short time intervals during the whole life cycle of the structure. However, a quality-controlled visual inspection is nearly impossible in regions such as the Ems, Weser and Elbe due to the high level of sedimentation. In this research, a fully automated, quality-controlled and reproducible three-dimensional (3D) sensing and damage detection of port infrastructures, above and below water, is proposed. Based on the results obtained, the port operator has more reliable information to efficiently plan maintenance and construction work. This approach will reduce the expenses significantly by lowering the downtimes of the port facilities and well-planned construction. Damage detection is usually performed in modern data processing based on pattern recognition methods (see [1] for more information). This is a reliable approach to detect any damages and make a well-founded assessment of the current state of the structure. Not only exact but also high-resolution 3D data for the above and below water parts of the building are required for the acquisition of the building geometry and condition.

Various publications deal with comprehensive sensing methods for the structural health monitoring of concrete or other materials above water or in clear offshore regions. A static underwater multibeam scanner for 3D reconstruction of subaquatic environments was introduced by [2]. Robert et al. [3] used a multibeam echo-sounder and underwater stereo cameras to create a 3D point cloud of vertical marine structures. Hadavandsiri et al. [4] introduced a new approach for the automatic, preliminary detection of damage in concrete structures with terrestrial ground scanners and a systematic threshold. An automatic classification for underwater geomorphological bed forms was presented by Ref. [5], which achieved an overall accuracy of 94%. A long-term monitoring approach for zigzag-shaped retaining structures is proposed by Ref. [6]. Aldosari et al. [7] used an ultra-high accuracy wheel-based mobile LiDAR mapping system for monitoring mechanically stabilized earth walls. O'Byrne et al. [8] detected disturbances by the texture segmentation of colour images. Gatys et al. [9] showed that neural networks trained on natural images learn to represent textures in such a way that they can synthesize realistic textures and even entire scenes. Neural networks, as feature extraction, are thus preferred over hand-crafted features [10–12]. A novel sensor data-driven fault diagnosis method is proposed based on convolutional neural networks (CNNs) by [13]. However, the limitation of such a transfer of features remains an open research question, especially when the input domain has the same topological structure but different statistical behaviour. The detection of non-normal instances within datasets is often called anomaly detection.

The definition of outliers first mentioned by [14] for outlier detection varies widely nowadays. Anomalies are no longer just understood as incorrect readings, but are often associated with particularly interesting events or suspicious datasets. The original definition was, therefore, extended by [15]:

1. Anomalies are different from the norm regarding their features;
2. Anomalies are rare compared to normal instances in a dataset.

Two widely used methods in anomaly detection are transfer learning and local outlier factors (LOF). Transfer learning adopts pretrained neural networks based on a different domain [16]. This results in advantages such as faster creation, better model quality, and less use of resources (training data). Breunig et al. [17] describe a method called LOF, which judges a sub-element on how isolated it is regarding the local neighbourhood.

Nowadays, anomaly detection algorithms are often used in many application domains. García-Teodoro et al. [18] describe a method using anomaly detection algorithms to identify network-based intrusion. In this context, anomaly detection is also often called behavioural analysis, and these systems typically use simple but fast algorithms. Other possible scenarios are fraud detection [19], medical applications (such as patient monitoring during electrocardiography [20]), data leakage prevention [21] and other more specialised applications, such as movement detection in surveillance cameras [22] or document authentication in forensic applications [23].

In this work, we aim to detect structural damages in infrastructures based on point clouds. We use anomaly detection algorithms due to the large imbalance between damaged and undamaged areas and the small amount of training data for the damaged areas. The novel detection approach we use can classify defective from non-defective features in a simulated data environment. The procedure of transferring features from natural images to point clouds and then performing a novel detection is totally new in the context of structural health monitoring systems. It is now for the first time possible to detect damages in an automated manner. This opens the door for further research into the use of pretrained neural networks for range sensor data. Therefore, the approach developed is applicable in all areas of damage detection for infrastructure objects.

2. Methodology

We first need to preprocess the data, because unstructured, large 3D point clouds are unsuitable for most anomaly detection algorithms. Therefore, we transfer features learnt from natural images to height maps from a range sensor. A height map or height field (also called a digital elevation model (DEM) [24]) in computer graphics is a raster image that is mainly used as a discrete global grid in secondary height modelling. Each pixel records values, such as surface elevation data. In contrast to natural images, the characteristics of height maps depend on the scan resolution and the object scanned itself, which makes transferability difficult. A way to overcome this drawback is to train height map neural networks from scratch [25].

In our system, three different sensor types are merged into one kinematic multi-sensor system (k-MSS) for the mapping task: a high-resolution hydro-acoustic underwater multi-beam echo-sounder, an above-water profile laser scanner and five high dynamic range cameras. In addition to the IMU-GNSS-based georeferencing method known from various applications, hybrid referencing with automatically tracking total stations is used for positioning (Figure 1). Although the individual sensors record in a grid pattern, the resulting point cloud is not grid-shaped due to the movements of the carrier platform.

The choice of damage types depends on the application and the relevant task within the life cycle management. In this study, we focus on geometrical damages and, for the time being, we only use point cloud data and no images from the cameras. The point clouds should have the smallest possible distance between the points, but still large enough so that there is no correlation between the points due to overlapping laser footprints. The head of the laser scanner rotates with 100 hz, the platform moves as slow as possible and the position and orientation is obtained from the GNSS/IMU system. Furthermore, the following state-of-the-art does not address mapping and data collection due to the fact that the research contribution lies in the damage detection area. It focuses on damage detection (see [1] for more details on mapping and data collection).

The method starts with a point cloud of typical structures (see Sections 3 and 5 for details). Firstly, we transform the point cloud into a height field, which is described in Section 2.1. Secondly, in Section 2.2, we extract features with a CNN. The third step is the defect detection using two different approaches: transfer learning and LOF (Section 2.3).

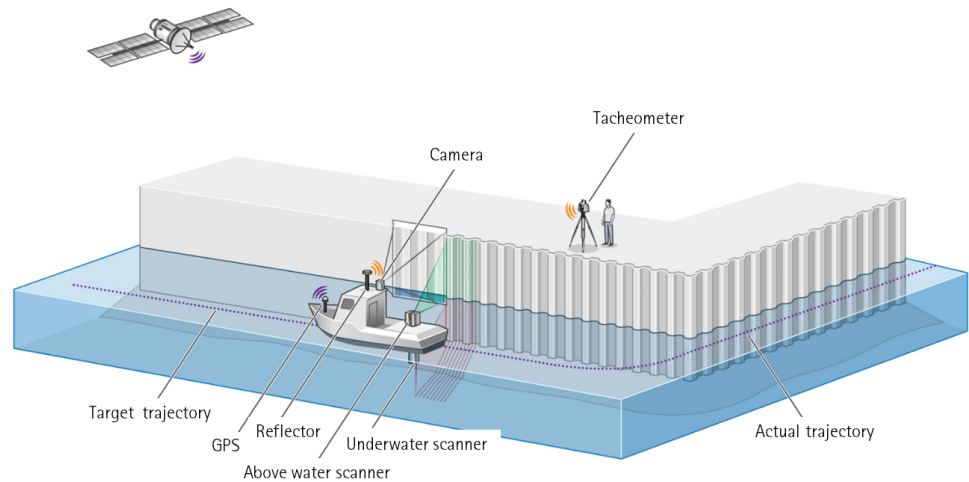


Figure 1. 3D mapping of a port structure above and below water [1].

Both methods yield outlier scores, which can be thresholded to achieve a binary classification. In contrast to other common outlier detection methods, these do not make any assumptions about the distribution of the outliers. They are, thus, well-suited for port infrastructural monitoring where each damage is expected to be unique.

2.1. Height Field Generation

Input variables for the machine learning approach are equally sized and rasterised distances between the point cloud and the original damage-free structure. In an optimal scenario, one can use a computer-aided design (CAD) or building information model (BIM) and determine deviations between the model and the point cloud. Unfortunately, no models are available for most existing infrastructural objects. There are two possibilities to overcome this challenge: the manual or (semi-)automatic generation of a CAD model or the use of an approximated local surface, for example, using a moving-window approach (e.g., [26]).

In the case of the simulated dataset, we use a mathematical model of a sheet-pile wall to create the simulated dataset and the corresponding CAD model. The distances from each point to the corresponding plane in the model are determined according to [27] with Equation (1) and are then rasterised into a two-dimensional (2D) height field with an equal 2 cm raster size,

$$dist = \frac{n_x \cdot p_x + n_y \cdot p_y + n_z \cdot p_z - d}{|\mathbf{n}|}, \quad (1)$$

where \mathbf{n} is the normal vector of the plane with the entries n_x , n_y and n_z . d is the distance to the origin and p_x , p_y and p_z are the co-ordinates of the point.

There is no existing CAD model of the quay wall for the real dataset, therefore, we had to create the model ourselves. For this purpose, regular shapes are fitted into the point cloud and the distance from the points to the geometry is determined. A simple plane according to [27] is used as reference geometry in this work. Firstly, the point clouds are rotated in a consistent direction using principal component analysis. Regular square sections are then cut from the point cloud. These sections overlap by 50% each in the X and Y directions. The cutting into smaller sections is useful to be able to estimate (well) fitting geometries into the point cloud. After cutting, a plane is estimated in each of the

sections. We only used the points of the quay wall and the damaged areas for the plane estimation. The distance to the plane is set manually to small values for points that are located on additional objects, such as ladders or fenders. This allows deviations due to damage in the grey value differences to be more clearly visible.

The raster size depends on the resolution of the point cloud and must be adapted to the respective dataset. Empty cells, which occur due to data gaps or inappropriate point distribution, are interpolated according to [28] with natural neighbour interpolation to avoid interference in the feature extraction step (cf. Equation (2))

$$G(x) = \sum_{i=1}^n \frac{A(x_i)}{A(x)} f(x_i) \quad (2)$$

where $G(x)$ is the estimate at x , $f(x_i)$ the known data at x_i , $A(x)$ is the volume of the new cell centred in x , and $A(x_i)$ is the volume of the intersection between the new cell centred in x and the old cell centred in x_i .

The median value of the distances is used in overpopulated cells, which occur due to inappropriate point distribution. The whole process is implemented in MATLAB and Python and summarised in a flowchart form in Figure 2.

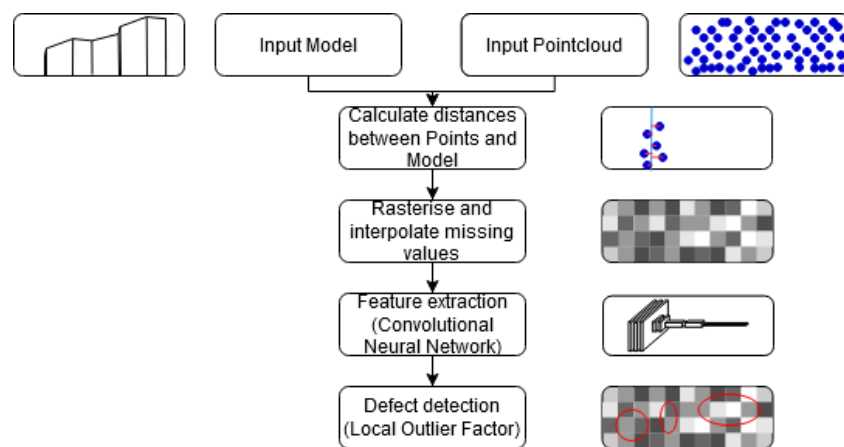


Figure 2. Flow chart of the automatic damage detection process.

The height field of the infrastructure object obtained is interpreted as a scalar function defined on a 2D grid, denoted by $H(x, y)$. Afterwards, patches are extracted from the grid and rearranged into data vectors. The latter x are organised as matrix X with shape $N \times p$, where N is the number of patches and p the number of pixels. Figure 3 shows an example of such a height field in grey scale, where a lighter grey value represents a greater deviation from the nominal CAD model.

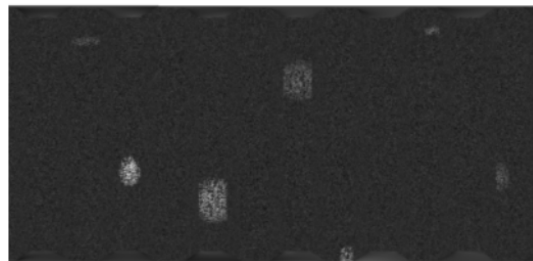


Figure 3. Visualising an exemplary height field on a 2D grid in grey scale.

2.2. Feature Extraction

A deep learning network requires a large amount of high-quality annotated data. However, as damage to port structures is relatively rare, it takes a very long time and a lot of measurements until a sufficient amount of annotated data is available. To overcome this problem, we chose a truncated version of the VGG19 network as the basic backbone for feature extraction and transferred its pretrained parameters on ImageNet to our dataset of port structures. The VGG19 neural network is a standard CNN, pretrained on natural images [29,30]. The network consists of 19 layers and is trained in a classification scenario. It is well-known for achieving superhuman performance on the extensive scale image database ImageNet [31]. The latter consists of more than a million labelled natural images of life scenes (such as cats, people, bicycles), which is very different from the dataset used in this paper. Therefore, we do not use the original network, but a variant that we modified. We only keep the first convolutional layers of the network, including layer pool_4, to prevent overfitting. The reason for this is that deeper layers in the network tend to learn higher order features, such as objects and faces, than lower layers that learn lower order features, such as edges and structures. A comprehensive visualisation can be found in [32]. We focused in this work on the detection of geometric damage such as spalling, which can be described well with lower order features, therefore, we obtained the best results with the network truncated after layer pool_4.

In contrast to the scalar function of the height field, the VGG19 network requires three-channel input (RGB-colour). Therefore, the signal is broadcast over three channels. We may encounter large height fields depending on the length of the wall scanned. We split the height fields into smaller tiles to compensate for hardware limitations. Dividing a large scan into smaller tiles not only increases the computational efficiency but also creates the possibility of achieving more than one label for the whole area. As a result, defects can be more efficiently located based on the smaller size of the tiles. If a defect is located at the border of a tile, affecting more than one tile, a criterion of 50% overlap in both directions is defined for a more reliable defect detection. Every vector x_i is propagated through the network, and the intermediate activation of the j th layer is stored.

Afterwards, the Gramian matrix of each activation is computed (see [9] for details). We only keep the diagonal of the Gramian matrix, which relates to the energy per feature, for computational efficiency and because we are not interested in synthesising new data. This leads to the feature vector $z \in \mathbb{R}^k$, where k is the number of feature maps in the j th layer of the network. Note that this procedure always leads to a dimensionality k independent of the input size p . Again, we organise all feature vectors as rows in a matrix, resulting in a feature matrix Z with shape $N \times k$.

2.3. Defect Detection

The last step of the damage detection process transforms the features that were extracted from the height fields into a single prediction label. Two different but interchangeable methods were used: transfer learning and LOF. Their performance is evaluated and compared in Section 4.

2.3.1. Transfer Learning

We use a three-layer feed-forward neural network to transform the extracted features into a single output label.

Firstly, it consists of two fully connected layers of the same size as the extracted features. These layers use the widely used ReLu (rectified linear unit) activation function [33] to allow for non-linear modelling. Furthermore, a dropout rate of 20% was chosen to help prevent overfitting during the training of the neural network. Secondly, there is a layer with a single neuron that is fully connected to the previous layer. This network is appended to the feature extraction network from Section 2.2. The value of the single output neuron is then used for threshold-based classification.

2.3.2. Local Outlier Factor

This second type of discriminator uses a standard approach called a LOF [17]. It is capable of detecting outliers in data (outliers are data points that do not fit in with the rest of the data). In order to achieve outlier detection, the LOF method constructs a reachability graph in feature space to estimate the density of the neighbourhood. It then computes an outlier score for each data point from this density. There are two different ways of using this discriminator:

1. The first way is to feed an untrained LOF discriminator with new data and let it compute outlier scores for each data point. Using these scores, outliers can be found in new data without any prior training;
2. The second way uses training on clean data (only showing the normal state without any defects) to create the reachability graph. Afterwards, outlier scores can be computed for new data by comparing it with the reachability graph of the trained normal case.

Either way, the outlier score is then used for threshold-based classification.

3. Simulation and Application

3.1. Creating the Dataset

Using a machine learning approach requires a large set of labelled training data. Four steps are necessary for the generation of a simulated point cloud of a sheet pile wall with damages:

1. Generation of a large number of datasets with randomly located and sized damages with a mathematical model for a sheet pile wall;
2. Computation of Cartesian co-ordinates of each point on the planes of the sheet pile wall by projecting rays from the k-MSS in vertical increments. The third dimension results from the movement of the sensor along a given trajectory. We used a straight line with equidistant sensor positions in this study for the sake of simplicity;
3. Addition of a random number of damages onto the planes of the sheet pile wall. Each damage has an ellipsoidal shape with random values for the principal axis;
4. The result is a noise-free point cloud. The ranges of each beam are then contaminated with random instrumental measurement uncertainty. The distance measurement uncertainty is assumed to be a normally distributed standard deviation and set according to the manufacturer's specifications to 20 mm. The uncertainties of the angles are neglected, because they lead to a small shift in the plane direction and a resulting small distant error, which is already included in the 2 cm noise. The resulting resolution of the point cloud is around 2 cm. Figure 4 illustrates the simulation procedure.

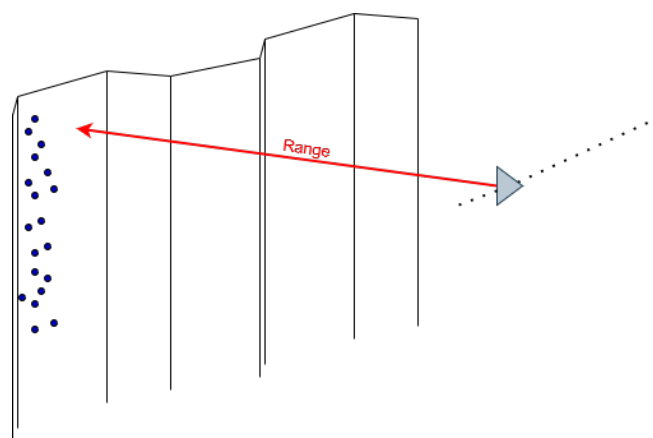


Figure 4. Simulation principal used in this study: at each sensor position, all rays in vertical increments according to the manufacturer's specifications are intersected with the planes of the sheet pile wall.

A second dataset was generated: it indicates for every position whether it is damaged or not, and is this the ground truth. Since the height field is a 2D raster image for the anomaly detection, we also use a 2D binary label image where a value of one stands for damage and zero for undamaged zones (right-hand side of Figure 5).

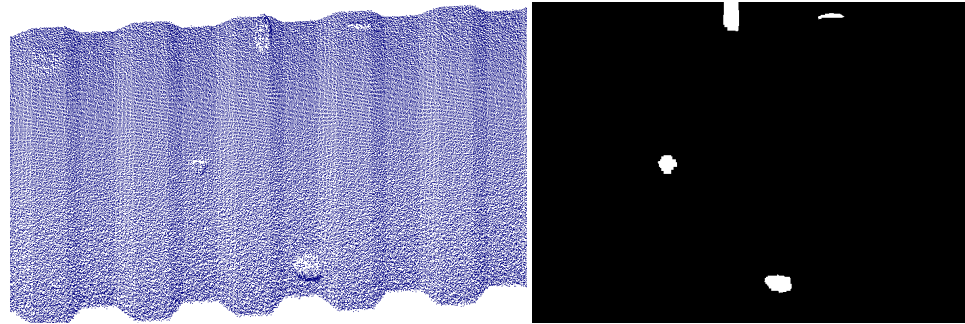


Figure 5. Left is the simulated point cloud and right is the corresponding label image where a value of one stands for damage and zero for undamaged zones.

The raster is generated in the XZ-plane of the point cloud. The image is filtered with the morphological operators erosion and dilation to avoid coarse shapes or misclassified raster cells [34]. Erosion is used to separate two near clusters, and dilation fills small holes. Using erosion after dilation is also called ‘opening’ in mathematical morphology. Figure 5 shows the simulated point cloud with the corresponding label image.

3.2. Training

The defect detection algorithms were implemented in Python 3.7. The TensorFlow library [35] was used for the pretrained VGG19 network and transfer learning approach. The implementation provided by the library sklearn [36] was used for the LOF approach. The datasets are split into 70% training and 30% test data for both approaches.

3.2.1. Transfer Learning

The training of this network requires height field samples without damages. The network needs to be trained on examples of normal and abnormal data to learn how to discriminate between the two. A different approach had to be used because there are no large datasets with an even amount of normal and abnormal data available. A new dataset was created where half of the examples are undamaged and the other half had the pixels of the height map inside a rectangular region of random size rearranged randomly. Additionally, some additive white Gaussian noise was added in that region. The network is expected to learn the statistical properties of a normal height field.

The network was trained on this new dataset. The unaltered, normal examples get a label of zero, and a damaged one gets a label of one. The network was trained to learn these labels using transfer learning. This means that the pretrained neurons of the truncated VGG19 network are kept fixed, while only the new neurons are trained. A dropout rate of 20% was used to decrease the risk of overfitting.

3.2.2. Local Outlier Factor

The LOF approach can be used in two different ways (see Section 2.3.2):

1. With training: A dataset without defects is required for training the LOF discriminator. The large height field strips are tiled and propagated through the feature extraction stage, as described in Section 2.2. The LOF learns the distribution of the features of the normal case. Even if the training data was not completely free of defects, the training of the LOF discriminator can compensate for a small amount of contamination in the training data. This feature was not used since we have full control of our training data

because it is created artificially through simulation. The trained LOF discriminator can then be saved for inference on new data;

2. Without training: The LOF discriminator can also be used without training. In this case, the LOF discriminator is trained on-the-fly at inference time. In this manner, the LOF discriminator can detect anomalies in the new data without prior training of the normal case.

3.3. Output Merging

Both methods take a large height map image as input, which is split into smaller overlapping tiles. A pretrained truncated VGG19 network extracts features from those tiles. Subsequently, either a neural network or a LOF discriminator is used to convert these features into a scalar value that indicates whether the respective tile shows a defect or not (see Section 2).

The tiles overlap by 50%, thus, each source pixel from the original height map is evaluated four times. The scores for the tiles are merged onto a regular non-overlapping grid of values to allow for a better interpretation, analysis and evaluation of the classification results. The grid cells have the same size as the individual overlapping regions.

Two functions were tested for merging the tiles' values into the grid: min and mean. These functions operate on the values of all tiles, which share the overlapping region at the position of the grid cell and directly determine its value.

The advantage of this method is that the results can be interpreted as a grey-scale image and, thus, can be visualised easily. Figure 6 shows an example of a simulated height map with defects applied (top) and the corresponding outlier activation scores that result from merging the scores of the tiles using the mean function (bottom). The threshold-based classification is also performed on this raster as opposed to directly on the tile's score. This may improve the classification results as it reduces noise. Oversampling is eliminated, therefore, each defect is only considered once, making the evaluation more accurate.

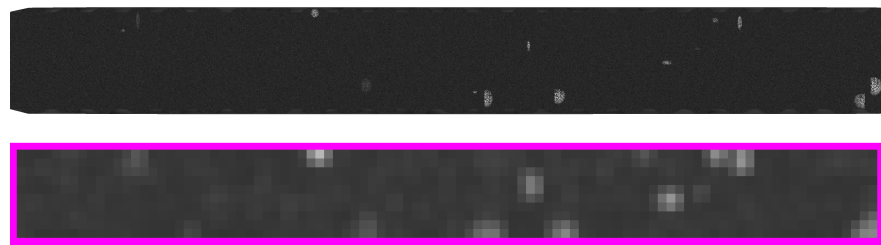


Figure 6. (Top) example of a simulated height map with defects applied. (Bottom) corresponding outlier activation scores that result from merging the scores of the tiles using the mean function.

4. Evaluation

We have evaluated our algorithms on our own synthetic datasets, which we created as explained in Section 3.1. The damages in this dataset are ellipsoidal in shape and the three axes of the ellipsoid are randomly sized between 5 and 50 cm.

4.1. Tile Score Merging Method

As discussed in Section 3.3, the outlier score computed for each individual tile needs to be merged into a non-overlapping grid. Two different functions were used. Their performance was compared on a dataset with known labels.

Figure 7 shows how the false positive and false negative rate relate to each other for varying chosen threshold values using the LOF approach. It can be seen that for any false negative rate, the false positive rate of the mean function is lower than for the min function. Lower false positive rates are of interest for this application, as a higher value may mean more manual work to check a larger number of candidates for damage that are, in fact, in a good state. Thus, the mean function is found to be superior.

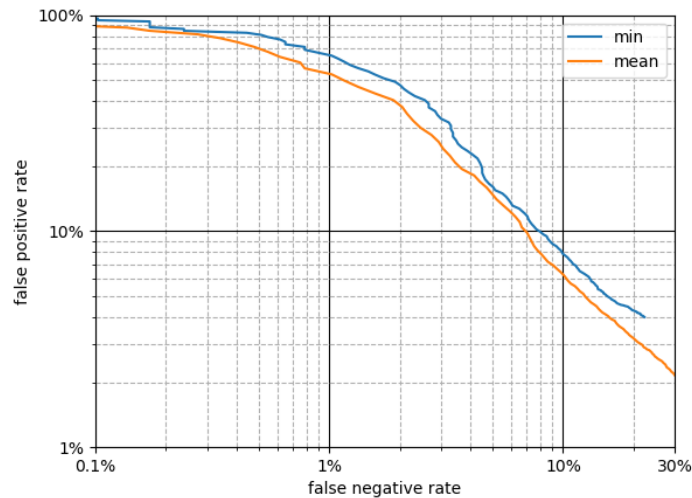


Figure 7. Comparison of tile score merge mode performance.

4.2. Evaluation Method

The size of a defect in our training data is typically larger than the size of one grid tile. In other words, a defect usually spreads over several grid tiles. It is important for our application that we can find all the defects, but we do not need to know the exact extent of each defect. Because of this, it is sufficient if only one (or more) of the tiles that a defect covers is classified as defective. Thus, for evaluating the usefulness of our algorithms regarding their intended application, we consider a defect as detected even if not all the tiles that it covers were classified as defective. Additionally, we ignore the border region of a labelled defect.

4.3. Threshold Selection

After merging the tile values, a grid of outlier values remains. The threshold required to detect these values into the Boolean labels *normal* and *abnormal* is not known a priori. It can be chosen arbitrarily to achieve a required maximum false negative rate. Figure 8 shows how the commonly used classification evaluation metrics are dependent on the threshold for the LOF approach. Outliers are assigned smaller numbers. Thus, with a larger threshold, more tiles will be classified abnormal. The curve for the recall shows that more actual defects can be detected. On the other hand, raising the threshold will decrease the precision, which means that many of the tiles that are classified defective are, in fact, in a good state. The graph for the accuracy is also worth noting. The dataset contains far more normal than abnormal examples, therefore, the accuracy can be close to 100% even when the recall is not that good. This metric is, thus, not suited to evaluating the performance of anomaly detection. Figure 9 shows the same metrics for the transfer learning approach. The graphs are generally lower, which shows that our transfer learning discriminator is inferior to our LOF discriminator.

Abnormal tiles have lower (more negative) scores than normal tiles. Thus, a larger threshold leads to a decrease in the recall (more defective examples found). However, at the same time, more false positive results are generated, which decreases the precision.

The primary goal with our application is to find most defects, i.e., a large recall. However, at the same time, the precision must be low enough to reduce the work that is required afterwards. Labelling everything as defective is useless. The threshold has to be selected to fulfil both conditions.

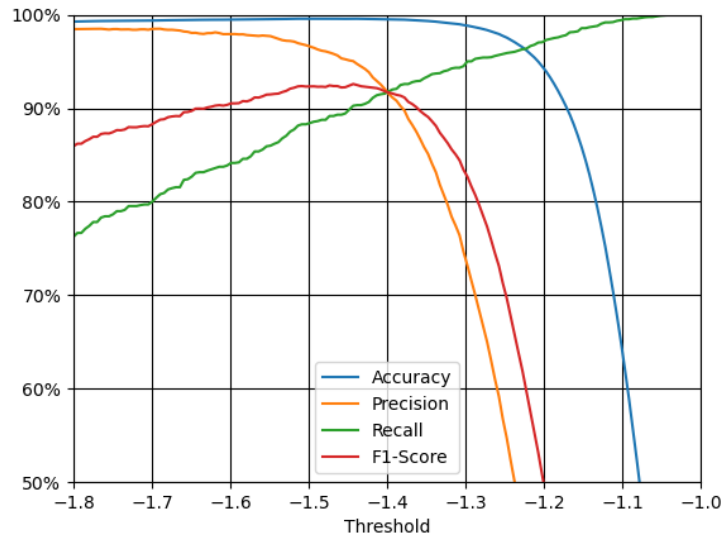


Figure 8. Dependence of evaluation metrics on threshold, LOF.

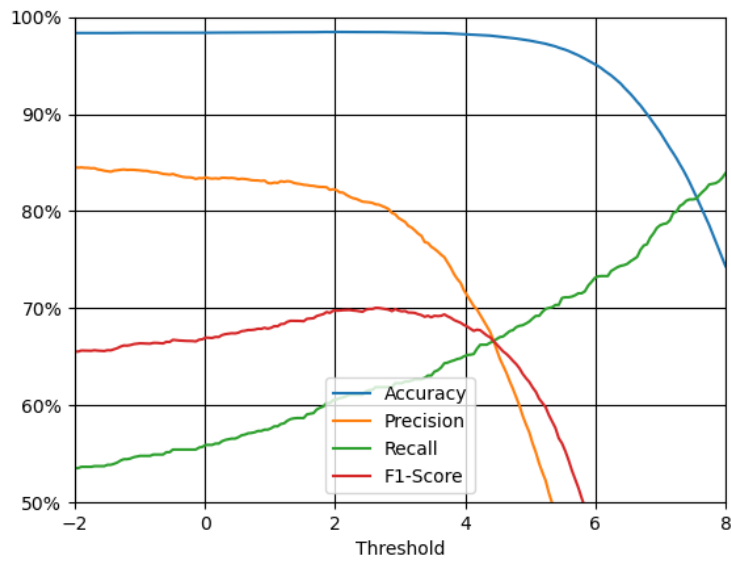


Figure 9. Dependence of evaluation metrics on threshold, Transfer Learning.

We were able to achieve much better results using the LOF approach, as can be seen when comparing the graphs of Figures 8 and 9. Thus, we will focus only on that from now on. If we chose a desired recall of 95%, we would achieve the following average results from our data as shown in the confusion matrix in Table 1 and evaluation metrics in Table 2:

Table 1. Confusion matrix for recall = 95% (threshold = -1.298).

		Actual Label	
		abnormal	normal
Predicted label	abnormal	2781	1017
	normal	146	95,820

Table 2. Evaluation metrics for recall = 95% (threshold = -1.298).

Accuracy	98.8%
Precision	73.2%
Recall	95.0%
F1-Score	82.7%

Figure 10 shows a few examples from our dataset with overlaid classification results. Correct classification of damage is coloured in green, yellow are false positives (classified damage where there is none) and red are false negatives (did not detect the damage). The magenta colouring shows the border areas that were excluded from our analysis.

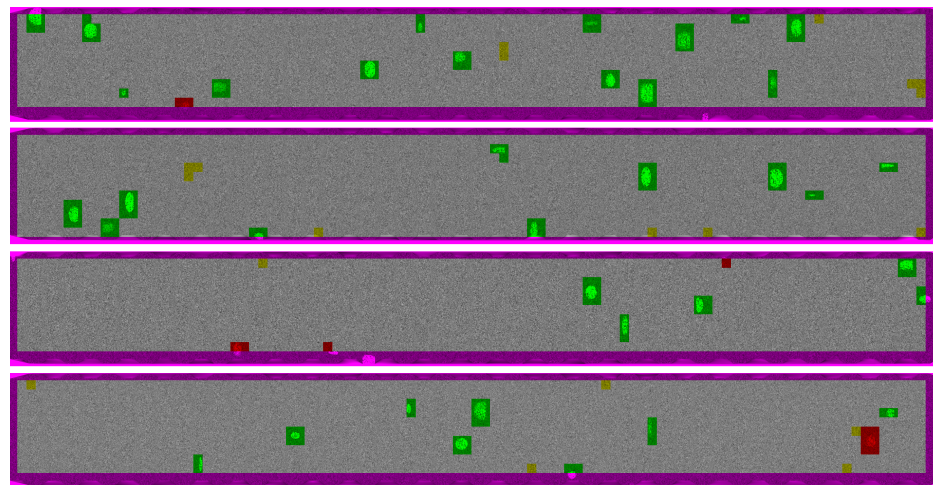


Figure 10. Examples of classification results on test data. Correct classification of damage is coloured in green, yellow are false positives (classified damage where there is none) and red are false negatives (did not detect the damage). The magenta colouring shows the border areas that were excluded from our analysis.

As can be seen from the examples, most damage that was not detected is close to the borders of the scan strips. They are covered by fewer tiles and, thus, have a lesser chance of being detected.

The reason for the rather worse results in transfer learning is possibly that the trained three-layer network is too small to detect all randomly sized damages.

5. Application to Real Data

Since the method appears fundamentally suitable for identifying potential damage areas when using simulated data, the next step is to analyse its application to the real dataset. The real data represents a quay wall above water. It is surveyed with a terrestrial laser scanner of type Z + F Imager 5016 in Lübeck city port, Germany on 13 sensor positions. We fused the 13 sensor positions on point clouds to achieve a small spacing between the points. The point spacing within the point clouds varies due to different scanning positions but is around 1 cm. The noise reaches 1–2 mm. Figure 11 shows a photo of the quay wall and the corresponding point cloud. There is obviously spalling in the upper part of the quay wall between the fenders. Two examples of damaged areas are shown in Figure 12. Both are spalling in the concrete. They are up to 1.5 m wide and reach a height of 50 cm.

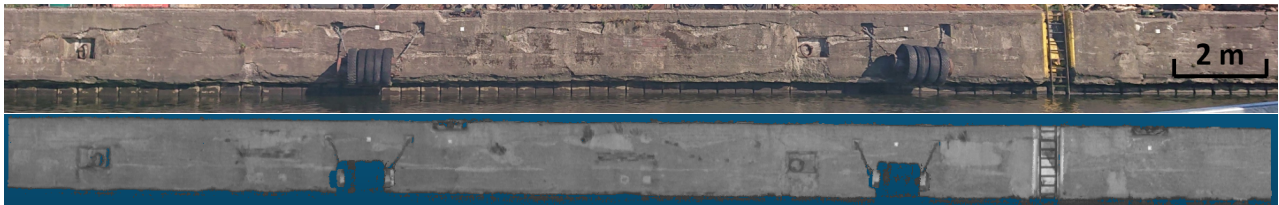


Figure 11. Real dataset in Lübeck city port. The upper row shows a photo of the quay wall and the lower row the corresponding point cloud. Spalling is in the upper part of the quay wall between the fenders.

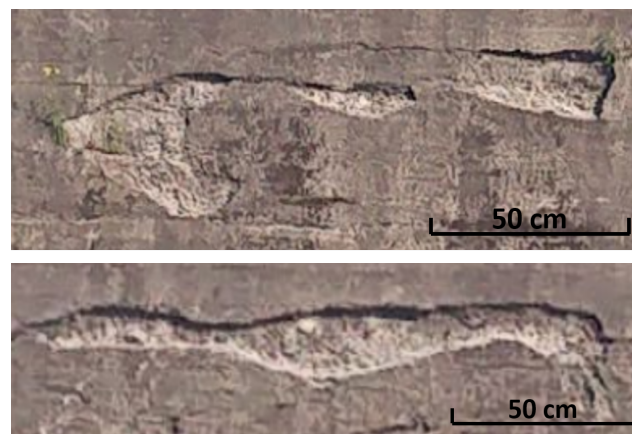


Figure 12. Two examples of damaged areas in the real dataset. Both pictures show the concrete spalling damage. The **upper** image shows three smaller damages with a size of up to 50 cm × 50 cm. The **lower** image shows larger spalling damage, which is 1.5 m in width and 30 cm in height.

The point clouds were manually divided into three categories to generate ground truth: quay wall (blue), concrete spalling (green), and additional objects (red). This classification and corresponding depth and label image are shown in the top row of Figure 13.

Since the methodology performs much better with LOF, as can be seen in the comparison of Figures 8 and 9, we use only LOF for the real data. The average result can be seen in Figure 14. When comparing the average results from the simulated dataset (Figure 8) with the real dataset (Figure 14), it can be seen that the curves for accuracy, precision, recall, and F1 score show similar behaviour with a lower overall accuracy.

Again, a threshold is chosen, where precision and recall are essentially equal, which is a good compromise between true and false positives in an economic sense. The corresponding confusion matrix for the threshold of -1.55 selected can be seen in Table 3. As can be seen from the table, there is again a strong imbalance between the two classes. The number of false positives and false negatives is essentially the same.

Table 3. Confusion matrix for recall = 72.6% (threshold = -1.55).

		Actual Label	
		abnormal	normal
Predicted label	abnormal	9128	3437
	normal	3506	57,225

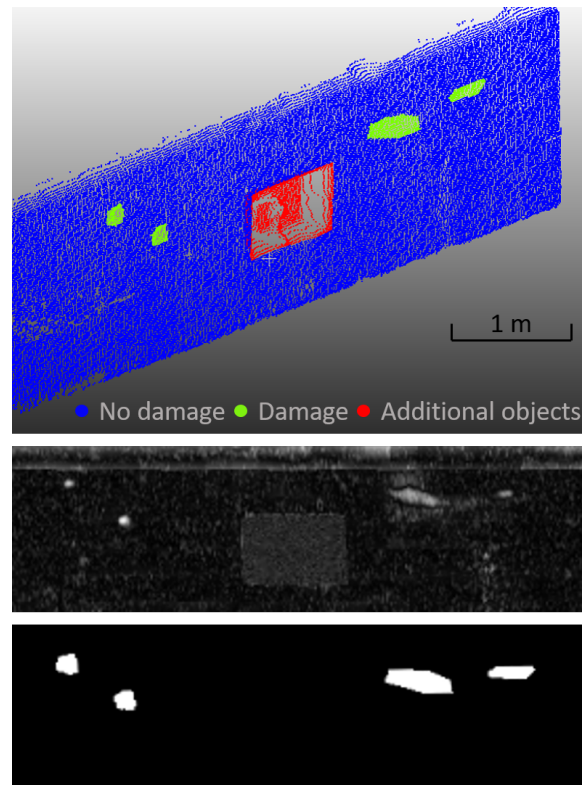


Figure 13. Real dataset with manually classified ground truth in the top row. No damage is in blue, concrete spalling is in green and additional objects, such as ladders or fenders, are in red. The middle row shows the corresponding height field in greyscale, with lighter areas indicating greater distances between the point cloud and the model. The bottom row shows the binary label image, with white indicating damage and black indicating no damage.

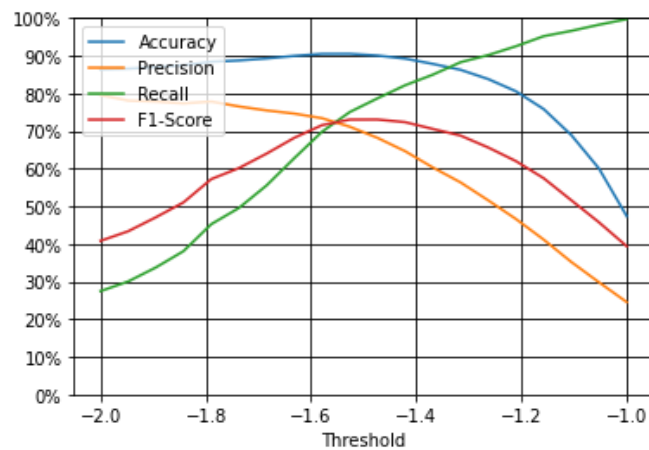


Figure 14. Dependence of evaluation metrics on the threshold, LOF for real data.

The evaluation metrics for the threshold of -1.55 selected are shown in Table 4. Accuracy reaches 90.5%. Precision and recall are with 72.2% and 72.6% mainly at the same level. This results in an average F1 score of 72.4%. This is still an indication of a good classification, but somewhat worse than in the simulated data.

Table 4. Evaluation metrics for recall = 72.6% (threshold = -1.55).

Accuracy	90.5%
Precision	72.2%
Recall	72.6%
F1-Score	72.4%

The classification result for two exemplary images is shown in Figure 15. Green, red and yellow indicate true positives, false negatives and false positives, respectively. Here, the original height field is shown with a higher contrast in the middle to make grey value differences in the height field more visible to the human eye. It can be seen in the top row of Figure 15 that all damages are detected and classified correctly. There are no false positives or false negatives. Only the two small damages at the bottom edge are not recognised, but this is because the edge areas are cut off during classification. The example in the bottom row shows a weaker classification result. Two damaged areas are not detected and, furthermore, two areas are falsely detected as damage.

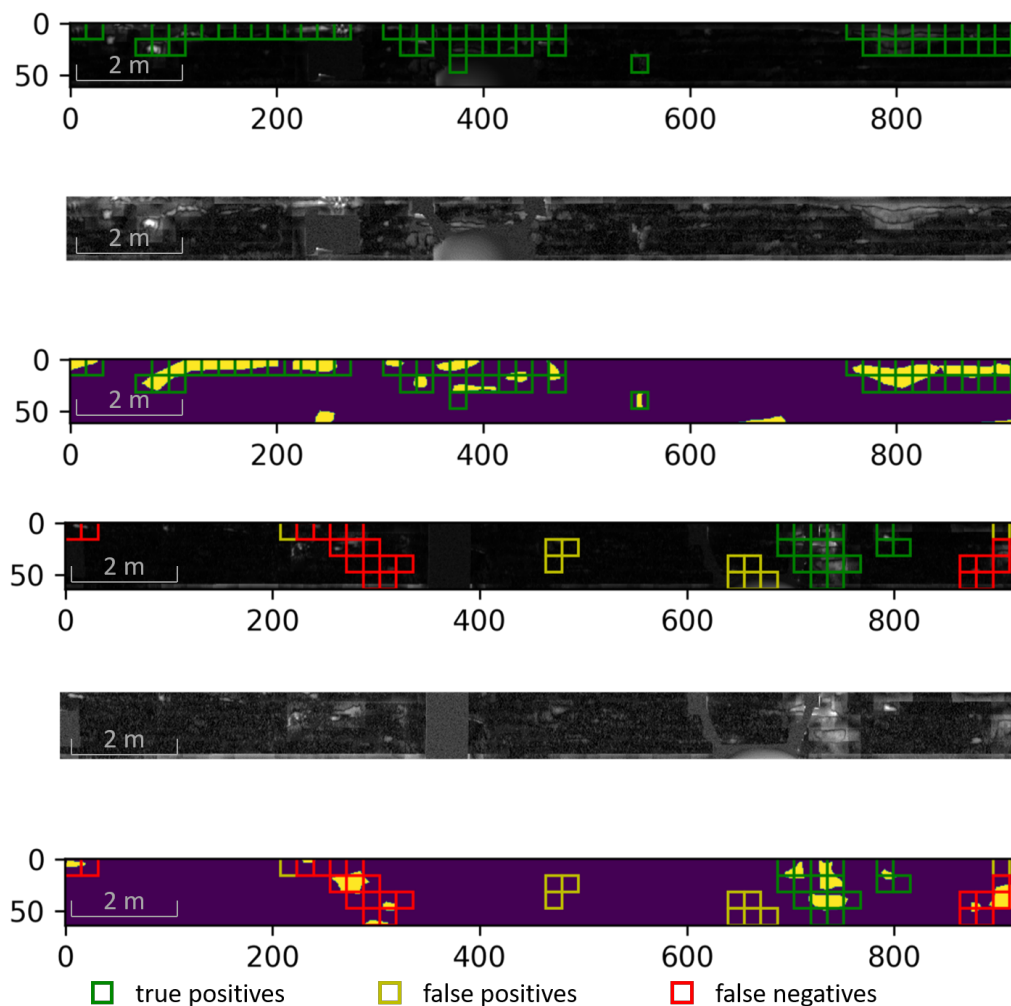


Figure 15. Two examples of damage detection using the real dataset from Lübeck city port. The **top** row in each case is the point cloud with overlaid damage detection. Green, red and yellow indicate true positives, false negatives and false positives, respectively. The **middle** row in each case is the height field and the **bottom** row is the label image with overlaid detected areas.

Therefore, the result is worse than for the simulated data (cf. Table 2). Nevertheless, the method seems to give good results when applied to real data.

6. Discussion

We assume several reasons for the different results. The first and probably the most important point is that there are other disturbing objects in the data, such as ladders, fenders, plants and ropes. The objects also lead to a higher distance in the height fields, which are currently not separable from higher distances based on real damage. The impact, particularly of plants, can be reduced by measurements in seasons with little vegetation, such as winter. The second point is that we do not clean or filter the data at the beginning. Only a rough manual cutting into the area of interest is carried out. The dataset still contains outliers and sensor artefacts that lead to false measurements. These artefacts are particularly strong where the structure comes into contact with the water. Therefore, the optimal time for the measurement is when the water level is as low as possible. In addition, the noise in the real data may not be normally distributed and still contain systematic components, unlike the simulated data. The threshold value for the separation into damaged and undamaged zones is chosen in the present study in such a way that a good trade-off between detected damage and actually correct classifications is achieved. The threshold may differ from another data set and has to be chosen again. In a very sensible and non-economic approach, one would choose a different threshold value that would give a higher recall.

The method presented is currently limited to geometrical damages, such as spalling and large cracking. The reason is that only 3D point clouds are used, and no colour information, which would be necessary to also detect small cracks and sintering. The method cannot detect damages smaller than the decimetre range because the point spacing and noise level underwater from a multibeam echo-sounder is much higher than a laser-scanned point cloud above water. Damages above water can be detected from the centimetre range due to the higher accuracy of the laser scanner and the smaller point spacing. Nevertheless, the results still contribute to automated damage detection and a digitally guided building inspections process.

The presented method gives similar results to other studies. Point CNN gives a mean interval above the unity score of 74.68% for bridge inspections with point cloud classification [37]. A combination of images and point clouds based on Otsu's algorithm for automatic concrete crack-detection achieves an average F1 score of 86.7% [38].

7. Conclusions and Outlook

The point clouds are converted into depth images and processed in a pretrained CNN with two extensions. Regarding the classification, firstly, an NN is attached to the CNN and, secondly, the LOF is calculated. Building inspection can be digitalised and taken to a completely new level with the method presented. We achieve a significantly higher completeness of the infrastructure inspections with the k-MSS used compared to the manual method with divers. We obtain a quality-controlled and reproducible mapping of the infrastructure by using laser scanners and hydrographic measurements. Suspected damage can be reliably detected and verified through the area-based measurement of the component surfaces above and below water. A comparison of different measurement epochs—as they have to be carried out every six years within the framework of the building inspection—is, thus, also possible for structures below water, so that the damage development and the service life of these economically important structures for our national economy can be better observed and evaluated in the future.

The procedure of transferring the features from natural images to point clouds and then performing a novel detection is totally new in the context of structural health monitoring systems. It is now possible for the first time to detect damage automatically.

The methodology presented is intended to automatically create a suspicion plan with suspected damage regions from point clouds. To be able to apply the methodology in reality, all damaged regions must be found as far as possible. Furthermore, only the damaged

regions should be recognised as such. This means that the accuracy and the recall value together should be as high as possible. The methodology was first tested on simulated data and then applied to real data.

The analysis of the simulated data resulted in a very good classification with an F1 score of 96.3%. Concerning the requirements mentioned above, the method is suitable for creating suspicion plans of damage regions on quay walls. The result is slightly less effective for the real data. The F1 score is 72.4%. When looking at examples of damage in the data that has not been detected by our algorithm, it can be seen that most of them are at the border of the scanned depth map.

The results could be further improved by handling the edges separately, as those typically show a significantly different distribution compared to the rest of the scan. So far, we have only been able to test our algorithm on simulated data and one real dataset of a concrete quay wall. However, we are working towards acquiring more real world data with different materials and building types. The proposed strategy is also applicable to other infrastructure objects, such as bridges, high-rise buildings, and tunnels.

Author Contributions: Conceptualization, F.H., L.G. and H.A.; methodology, F.H., L.G. and H.A.; software, F.H. and L.G.; formal analysis, F.H., L.G. and H.A.; investigation, F.H., L.G. and H.A.; writing—original draft preparation, F.H., L.G. and H.A.; writing—review and editing, F.H., L.G., I.N. and H.A.; visualisation, F.H. and L.G.; supervision, H.A. and I.N. all authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Federal Ministry of Transport and Digital Infrastructure grant number 19H18011C. The publication of this article was funded by the Open Access Fund of the Leibniz University Hannover.



IHATEC
Innovative
Hafentechnologien

Supported by:



Federal Ministry
of Transport and
Digital Infrastructure

on the basis of a decision
by the German Bundestag

Acknowledgments: This work was carried out as part of the joint research project “3DHydroMapper—Bestandsdatenerfassung und modellgestützte Prüfung von Verkehrswasserbauwerken”. It consists of five partners and one associated partner: Hesse und Partner Ingenieure (multisensor system and kinematic laser scanning), WK Consult (structural inspection, BIM and maintenance planning), Niedersachsen Ports (sea and inland port operation), Fraunhofer IGP (automatic modeling and BIM), Leibniz University Hannover (route planning and damage detection) and Wasserstraßen-und Schifffahrtsverwaltung des Bundes (management of federal waterways).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hesse, C.; Holste, K.; Neumann, I.; Hake, F.; Alkhatib, H.; Geist, M.; Knaack, L.; Scharr, C. 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadens-erkennung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. *Hydrogr. Nachr.-J. Appl. Hydrogr.* **2019**, *113*, 26–29.
2. Moisan, E.; Charbonnier, P.; Foucher, P.; Grussenmeyer, P.; Guillemain, S.; Samat, O.; Pages, C. Assessment of a static multibeam sonar scanner for 3D surveying in confined suaquatic environments. In Proceedings of the XXIII ISPRS Congress, ISPRS, Prague, Czech Republic, 12–19 July 2016; Volume 41, pp. 541–548.
3. Robert, K.; Huvenne, V.A.; Georgiopoulou, A.; Jones, D.O.; Marsh, L.; DO Carter, G.; Chaumillon, L. New approaches to high-resolution mapping of marine vertical structures. *Sci. Rep.* **2017**, *7*, 1–14. [[CrossRef](#)] [[PubMed](#)]
4. Hadavandsiri, Z.; Lichti, D.D.; Jahraus, A.; Jarron, D. Concrete Preliminary Damage Inspection by Classification of Terrestrial Laser Scanner Point Clouds through Systematic Threshold Definition. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 585. [[CrossRef](#)]

5. Janowski, L.; Wroblewski, R.; Rucinska, M.; Kubowicz-Grajewska, A.; Tysiac, P. Automatic classification and mapping of the seabed using airborne LiDAR bathymetry. *Eng. Geol.* **2022**, *301*, 106615. [[CrossRef](#)]
6. Seo, H. Long-term Monitoring of zigzag-shaped concrete panel in retaining structure using laser scanning and analysis of influencing factors. *Opt. Lasers Eng.* **2021**, *139*, 106498. [[CrossRef](#)]
7. Aldosari, M.; Al-Rawabdeh, A.; Bullock, D.; Habib, A. A Mobile LiDAR for Monitoring Mechanically Stabilized Earth Walls with Textured Precast Concrete Panels. *Remote Sens.* **2020**, *12*, 306. [[CrossRef](#)]
8. O'Byrne, M.; Schoefs, F.; Ghosh, B.; Pakrashi, V. Texture analysis based damage detection of ageing infrastructural elements. *Comput.-Aided Civ. Infrastruct. Eng.* **2013**, *28*, 162–177. [[CrossRef](#)]
9. Gatys, L.; Ecker, A.S.; Bethge, M. Texture synthesis using convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 262–270.
10. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: San Francisco, CA, USA, 2014; pp. 3320–3328.
11. Carvalho, T.; de Rezende, E.R.S.; Alves, M.T.P.; Balieiro, F.K.C.; Sovat, R.B. Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 866–870.
12. Abati, D.; Porrello, A.; Calderara, S.; Cucchiara, R. Latent Space Autoregression for Novelty Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 215–232.
13. Li, G.; Deng, C.; Wu, J.; Xu, X.; Shao, X.; Wang, Y. Sensor Data-Driven Bearing Fault Diagnosis Based on Deep Convolutional Neural Networks and S-Transform. *Sensors* **2019**, *19*, 2750. [[CrossRef](#)] [[PubMed](#)]
14. Grubbs, F.E. Procedures for Detecting Outlying Observations in Samples. *Technometrics* **1969**, *11*, 9005. [[CrossRef](#)]
15. Goldstein, M.; Uchida, S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLoS ONE* **2016**, *11*, e0152173. [[CrossRef](#)] [[PubMed](#)]
16. Andrews, J.; Tanay, T.; Morton, E.J.; Griffin, L.D. Transfer representation-learning for anomaly detection. In Proceedings of the 33rd International Conference on Machine Learning New York, NY, USA, 19–24 June 2016; JMLR: New York, NY, USA, 2016.
17. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 93–104.
18. García-Teodoro, P.; Díaz-Verdejo, J.; Maciá-Fernández, G.; Vázquez, E. Anomaly-based network intrusion detection: Techniques, systems and challenges. *Comput. Secur.* **2009**, *28*, 18–28. [[CrossRef](#)]
19. Li, S.H.; Yen, D.C.; Lu, W.H.; Wang, C. Identifying the signs of fraudulent accounts using data mining techniques. *Comput. Hum. Behav.* **2012**, *28*, 1002–1013. [[CrossRef](#)]
20. Lin, J.; Keogh, E.; Ada Fu.; Van Herle, H. Approximations to magic: Finding unusual medical time series. In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, Ireland, 23–24 June 2005; pp. 329–334. [[CrossRef](#)]
21. Sigholm, J.; Raciti, M. Best-effort Data Leakage Prevention in inter-organizational tactical MANETs. In Proceedings of the MILCOM 2012–2012 IEEE Military Communications Conference, Orlando, FL, USA, 29 October–1 November 2012; pp. 1–7. [[CrossRef](#)]
22. Basharat, A.; Gritai, A.; Shah, M. Learning object motion patterns for anomaly detection and improved object detection. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8. [[CrossRef](#)]
23. Gebhardt, J.; Goldstein, M.; Shafait, F.; Dengel, A. Document Authentication Using Printing Technique Features and Unsupervised Anomaly Detection. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 479–483. [[CrossRef](#)]
24. Skidmore, A.K. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *Int. J. Geogr. Inf. Syst.* **1989**, *3*, 323–334. [[CrossRef](#)]
25. Simony, M.; Milzy, S.; Amendey, K.; Gross, H.M. Complex-YOLO: An Euler-Region-Proposal for Real-time 3D Object Detection on Point Clouds. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
26. Haas, T.C. Kriging and automated variogram modeling within a moving window. *Atmos. Environ. Part A Gen. Top.* **1990**, *24*, 1759–1769. [[CrossRef](#)]
27. Drixler, E. Analyse der Form und Lage von Objekten im Raum. München, DGK Reihe C, Heft Nr. 409. Ph.D. Thesis, Karlsruhe Institute of Technology, Karlsruhe, Germany 1993.
28. Sibson, R. A brief description of natural neighbour interpolation. In *Interpreting Multivariate Data*; Barnett, V., Ed.; John Wiley & Sons: Chichester, UK, 1981; pp. 21–36.
29. Wan, X.; Liu, L.; Wang, S.; Wang, Y. A Transfer Learning Strip Steel Surface Defect Recognition Network Based on VGG19. In *Advanced Manufacturing and Automation IX*; Lecture Notes in Electrical Engineering; Wang, Y., Martinsen, K., Yu, T., Wang, K., Eds.; Springer: Singapore, 2020; Volume 634, pp. 333–341.
30. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
31. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

-
32. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833.
 33. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the International Conference on Machine Learning (ICML)*, Haifa, Israel, 21–24 June 2010 .
 34. Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press, Inc.: Cambridge, MA, USA, 1983.
 35. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
 36. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
 37. Kim, H.; Kim, C. Deep-Learning-Based Classification of Point Clouds for Bridge Inspection. *Remote Sens.* **2020**, *12*, 3757. [[CrossRef](#)]
 38. Chen, X.; Li, J.; Huang, S.; Cui, H.; Liu, P.; Sun, Q. An Automatic Concrete Crack-Detection Method Fusing Point Clouds and Images Based on Improved Otsu’s Algorithm. *Sensors* **2021**, *21*, 1581. [[CrossRef](#)] [[PubMed](#)]



Automated damage detection for port structures using machine learning algorithms in heightfields

Frederic Hake¹ · Paula Lippmann¹ · Hamza Alkhatib¹ · Vincent Oettel² · Ingo Neumann¹ Received: 7 November 2022 / Accepted: 30 January 2023
© The Author(s) 2023

Abstract

Marine infrastructures such as harbours, bridges, and locks are particularly exposed to salt water and are therefore subject to increasing deterioration. This makes regular inspection of the structures necessary. The inspection is carried out manually, using divers under water. To improve this costly and time-consuming process, we propose to scan the surface and underwater structure of the port with a multi-sensor system (MSS) and classify the obtained point cloud into damaged and undamaged areas fully automatically. The MSS consists of a high-resolution hydro-acoustic underwater multi-beam echo-sounder, an above-water profile laser scanner, and five HDR cameras. In addition to the IMU-GPS/GNSS method known from various applications, hybrid referencing with automatically tracking total stations is used for positioning. The key research idea relies on 3D data from TLS, multi-beam or dense image matching. For this purpose, we build a rasterised heightfield of the point cloud of a harbour structure by reducing the CAD-based geometry from the measured 3D point cloud. To do this, we fit regular shapes into the point cloud and determine the distance of the points to the geometry. To detect anomalies in the data, we use two methods in our approach. First, we use the VGG19 Deep Neural Network (DNN), and second, we use the Local-Outlier-Factors (LOF) method. To test and validate the developed methods, training data was simulated. Afterwards, the developed methods were evaluated on real data set in Lübeck, Germany, which were acquired with the developed Multi-Sensor-System (MSS). In contrast to the traditional, manual method by divers, we have presented an approach that allows for automated, consistent, and complete damage detection. We have achieved an accuracy of 90.5% for the method. The approach can also be applied to other infrastructures such as tunnels and bridges.

Keywords Damage detection · Machine learning · Laser scanning · Multibeam echo sounder · Infrastructure

Introduction

The infrastructure at ports in Germany, both on the sea and inland, is aging and in need of new technologies and methods for managing its lifespan. Traditional processes, which are labour-intensive and time-consuming, need to be replaced with automated, smart, and innovative measurement and analysis processes to improve transparency, efficiency, and reliability for more accurate lifetime predictions.

The infrastructure at ports is prone to degradation over time due to human activities and environmental factors, particularly the impact of saltwater on the material at seaports. This can lead to damage to concrete structures, sheet pile walls, and wooden construction. It is important to identify and assess the severity of this damage in order to take timely maintenance measures and prevent costly repairs or even infrastructure collapses.

✉ Frederic Hake
hake@gih.uni-hannover.de

Paula Lippmann
paula.lippmann@stud.uni-hannover.de

Hamza Alkhatib
alkhatib@gih.uni-hannover.de

Vincent Oettel
oettel@ifma.uni-hannover.de

Ingo Neumann
neumann@gih.uni-hannover.de

¹ Geodetic Institute Hannover, Leibniz University Hannover, Nienburger Straße 1, Hannover, Germany

² Institute of Concrete Construction, Leibniz University Hannover, Appelstraße 9a, Hannover, Germany

Testing and monitoring of port infrastructural buildings typically involves both above and underwater parts. The above water portion is usually tested through manual and visual inspections, while the underwater portion is more difficult to assess. Divers must manually inspect the structure by touching it with their hands, but this method is unreliable and subjective, making it difficult to accurately classify and track damages. Additionally, the underwater inspection process is typically limited in scope, meaning that only a small percentage of the structure is actually examined. As a result, damage is often not detected until it has become severe or is not discovered at all.

Structural health monitoring with kinematic multi-sensor system

Sea and inland ports need to be inspected comprehensively on a regular basis. However, it is nearly impossible to visually inspect underwater areas, especially in river regions due to the high sediment content.

For this reason, we use a kinematic multi-sensor system (k-MSS) to record the object's surface both above and below water.

Only in this way is it possible to scan the entire structure, reliably detect damage, and subsequently assess the current condition of the building in detail. In order to accurately record the building's geometry and condition, high-resolution 3D data is required for both the underwater and above-water parts of the building.

The carrier platform, which functions as a boat-like vehicle, plays a central role in recording measurement data. However, it is also exposed to variables such as wind, waves, and currents. The carrier platform also serves to hold the sensor platform, and its design aims to prevent flow-related influences from causing quality-reducing deformation and constraining forces on the sensor platform. For this purpose, a drift and torsion compensation is carried out.

In addition to recording data, structural analyses are conducted to assess the service life and lifecycle of the infrastructures.

Port operators can use the results of these analyses to implement maintenance concepts and construction measures in accordance with the structure inspection as transparently as possible. Cost-intensive maintenance measures and long downtimes are significantly reduced by this procedure.

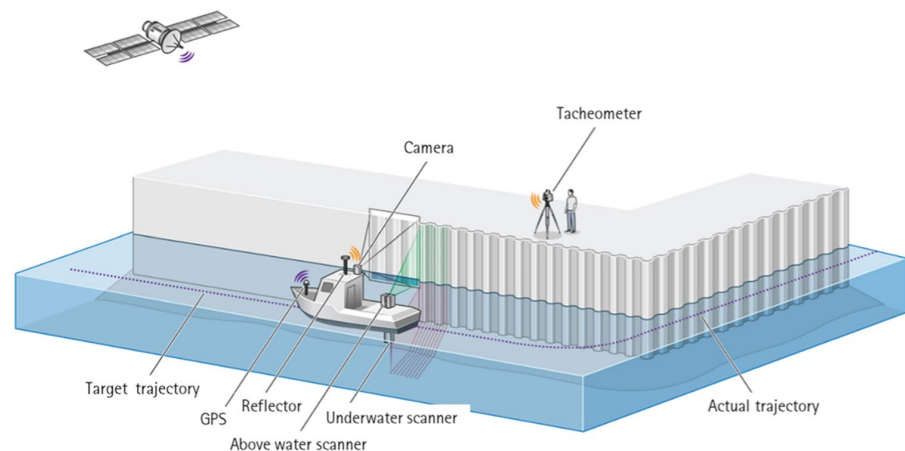
The k-MSS integrates various sensors for object recording, including a high-resolution hydroacoustic underwater multibeam echo sounder, a surface profile laser scanner, and five HDR cameras. Positioning is achieved through a combination of an IMU and GNSS combination, and there is the option for hybrid positioning with an automatically measuring total station from the shore (see Fig. 1).

The underwater multibeam echo sounder and the profile laser scanner measure 2D profiles, and the movement of the carrier platform creates an unstructured point cloud. The expected noise of the point clouds is in the millimetre range above water and in the centimetre range underwater.

In modern data processing, damage detection is often done using pattern recognition methods (see Hesse et al 2019). This work focuses on ensuring the quality, completeness, reproducibility, and automation of damage detection from 3D point clouds of port infrastructures above and underwater.

The main focus of this study is the detection of geometric damage in point clouds. In order to fully investigate this topic, we decided to only use point cloud data in our analysis. This decision was made in order to provide a more focused and in-depth examination of the subject and to avoid any potential complications or distractions that might arise from using additional data sources. As a result, the captured images from the installed cameras were not included in our analysis. The selection of damage types is highly dependent on the intended use and the specific application. In this study, we tested a methodology using two different

Fig. 1 A 3D recording of a port facility above and below water (Hesse et al. 2019)



data sets: a synthetic set and a real set from the harbour of Lübeck, Germany. The results of this testing will help us better understand the effectiveness of the methodology. The following state of the art is limited to damage detection and does not include data collection and mapping (for more details on mapping and data collection, see Hesse et al. 2019).

Structural monitoring by detection methods is addressed in several publications over water or in clear offshore areas.

Damage detection is mostly divided into two approaches: classification based on images or based on point clouds. In the case of point clouds, as in the present work, a distinction is made between point-based, area-based, and geometry-based methods (Neuner et al. 2016). A deformation monitoring and damage detection of large retaining structures using motor-vehicle-based mobile mapping systems was presented by Kalenjuk et al. (2021). Hadavandsiri et al. (2019) introduced a new approach for automatic, preliminary detection of damage in concrete structures with terrestrial ground scanners and a systematic threshold using a point-based approach. In Zhang et al. (2013), a geometry-based (using quadrics) approach is implemented for the segmentation of point clouds.

Image-based methods are divided into various approaches and methods. In recent years, neural networks have been increasingly used for classification and segmentation tasks. Tung et al. (2013) take repeated images of a retaining wall with a standard camera to perform digital image correlation. O'Byrne et al. (2013) are detecting disturbances by texture segmentation of colour images, while Gatys et al. (2015) showed that neural networks trained on natural images learn to represent textures in a way that enables synthesising realistic textures and even whole scenes. Neural networks, as feature extraction, are thus preferred over hand-crafted features (Yosinski et al. 2014; Carvalho et al. 2017; Abati et al. 2019).

When detecting damage in point clouds or images, one very often has to deal with the problem of having very little damage and very many undamaged areas. This leads to a great imbalance of classes. One possible way around this problem is to look for instances that are most different from the majority of the data. Such methods that distinguish non-normal instances from the majority are often called anomaly detection (AD).

Anomaly detection with transfer learning

Anomaly detection is a technique used to identify unusual or suspicious events or data sets. It was first mentioned by Grubbs in 1969 for identifying outlying data points, but the definition has since evolved to include a wider range of applications. Anomalies are now understood as data that is different from the norm and is rare compared to other

instances in a data set. Two commonly used methods for anomaly detection are transfer learning and local outlier factors (LOF). Transfer Learning involves using pre-trained neural networks to create models more quickly and efficiently (see, e.g. Andrews et al. 2016), while LOF evaluates an object based on its isolation from other data in its local environment (Breunig et al. 2000).

Nowadays, anomaly detection algorithms are often used in many application domains. García-Teodoro et al. (2009) describe a method to use anomaly detection algorithms to identify network-based-intrusion. In this context, anomaly detection is also often called behavioural analysis, and these systems typically use simple but fast algorithms. Other possible scenarios are Fraud detection (Li et al. 2012), medical applications like patient monitoring during electrocardiography (Lin et al. 2005), data leakage prevention (Sigholm and Raciti 2012), and other more specialised applications like movement detection in surveillance cameras (Basharat et al. 2008) or for document authentication in forensic applications (Gebhardt et al. 2013).

Contribution

This work focuses on using anomaly detection methods to segment point clouds into two classes: damaged and not damaged. Because there are significantly more areas without damage, the classes are unequal in size. Additionally, the amount of available training data with true labels is limited, so we use pre-trained networks as generic feature generators. Before applying the anomaly detection algorithms, the point clouds must be pre-processed to make them suitable for analysis. We create a height field (also known as a digital elevation model (DEM) (Skidmore 1989) from the point clouds. Unlike natural images, the statistics of height fields depend on the scan resolution of the sensor, which makes it difficult to transfer pre-trained networks. The anomaly detection method we present can successfully differentiate between damaged and undamaged areas in point cloud-derived height fields. This is a novel approach to automated damage detection in structural monitoring, and it provides a foundation for further research in this field.

Methodology

In this paper, we convert point clouds into heightfields and use them as input for a deep neural network (DNN). The features extracted by the DNN are then used to calculate the local outlier factor (LOF) for individual sections (or "tiles") of the heightfield. By setting a threshold, we can classify each tile as either a damaged or undamaged region. This allows us to identify damage in the heightfield using the LOF and DNN.

Pre-processing of the point clouds

To accurately evaluate the results, we need to manually label the damage (concrete spalling) in the existing data. Additionally, objects such as fenders or ladders appear in the data and should not be included in the calculation of the heightfield. These objects are also manually labelled and assigned an index. This results in a point cloud with different indices for the quay wall, damage, and other objects. An example of this can be seen in Section III. This labelling process is necessary for accurate and reliable analysis.

Generation of the heightfield

To create the heightfield, regular shapes are fitted into the point cloud, and the distance of the points to the geometry is calculated. In this study, we use a simple plane according to Drixler (1993) as the reference geometry. We first rotate the point clouds in a consistent direction using principal component analysis and then cut regular square sections from the point cloud. These sections overlap by 50% in the X and

Y directions. This allows us to more easily fit geometries into the point cloud. After cutting, a plane is estimated in each section.

When estimating the plane, we only use the points corresponding to the quay wall and damaged areas, taking into account the different indices. We then calculate the distance of each point to the estimated plane using the formula:

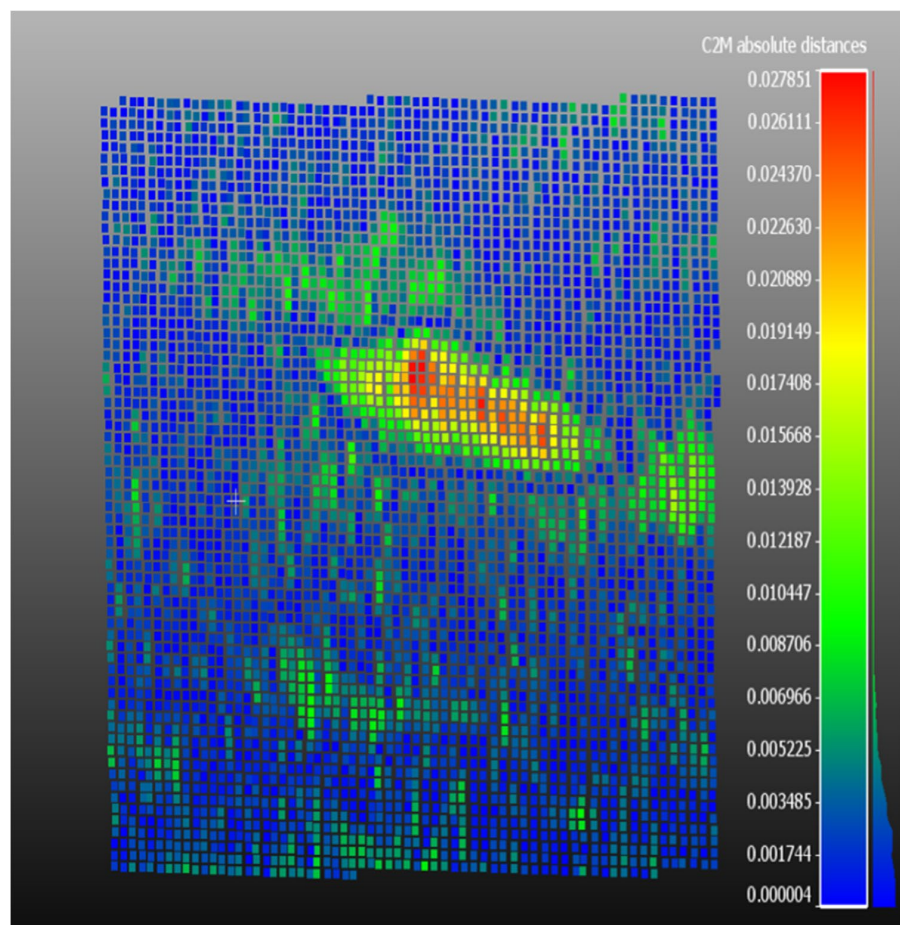
$$dist = \frac{n_x \cdot p_x + n_y \cdot p_y + n_z \cdot p_z - d}{|\vec{n}|}$$

This process enables us to generate a heightfield from the point cloud data.

The vector \vec{n} is the normal vector of the plane with the entries n_x , n_y , and n_z . d is the distance to the origin, and p_x , p_y , and p_z are the coordinates of the point. An example of the determined distances is shown in Fig. 2. Blue represents a small distance to the plane and green via yellow and red an increasing distance.

To make deviations due to damage more noticeable, the distance to the plane is manually set to small values for

Fig. 2 Distances to the plane



points that indicate the presence of additional objects. The point clouds are then converted into a heightfield by rasterizing them. The raster size should be selected based on the resolution of the point cloud, and distances in overlapping areas are averaged during the rasterization process. The resulting heightfield is represented as a grayscale image, with black indicating a distance of zero and lighter shades indicating increasing distance. A corresponding label image, which is a binary image showing points with damage indices in white and all other points in black, is also created. Examples of the heightfield and label images are shown in Figs. 3 and 4.

Convolutional neural network

The feature extraction from the heightfield is carried out using the VGG19 DNN (Simonyan & Zisserman 2014). Afterwards, the local outlier factors (LOF) is calculated out of the resulting feature maps.

The used VGG19 DNN is a variant of the VGG network which achieved very good results at the *ImageNet Large Scale Visual Recognition Competition* in 2014. The

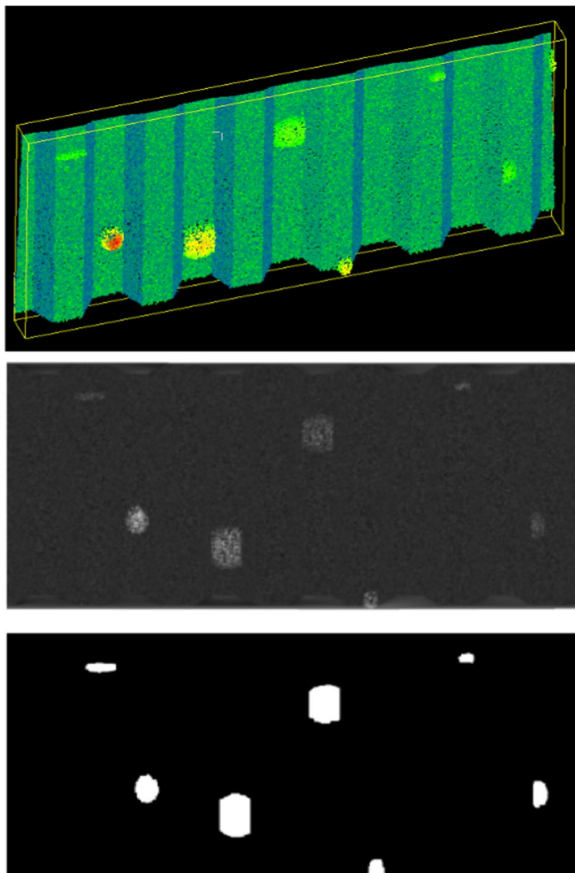


Fig. 3 Simulated data with heightfield and label

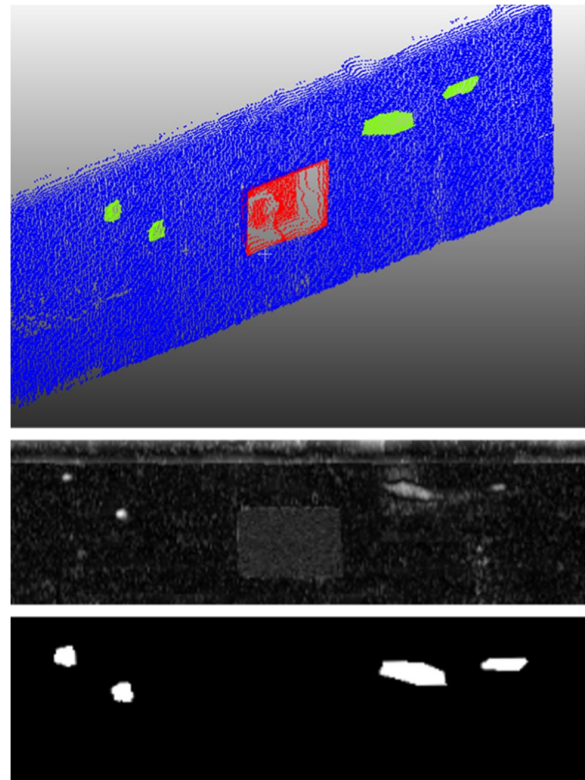


Fig. 4 Real data with heightfield and label image

VGG19 DNN consists of 19 layers, thereof 16 convolutions and 3 fully connected layers. Because of the good results, the VGG19 DNN is used with weights pre-trained on the ImageNet data set (Deng et al. 2009). There is no adjustment of the weights carried out. For the extension with LOF, the DNN is truncated after Layer *pool_4*. At this point, the feature maps are available. Small sections or “tiles” of the original depth image are used as input to the DNN. Accordingly, one output per tile is obtained. The size of the tiles is a variable parameter in the DNN. In addition, the tiles overlap by 50%.

The LOF is then calculated from these different feature maps for each tile. Each of the individual tiles is assigned its value. In the overlapping areas, the average value of the calculated LOF is determined. Using a threshold for the resulting LOF of the sections of the point cloud, the actual binary classification is carried out.

Data

In this work, we will use two data sets for analysis: one consisting of simulated sheet pile walls with randomly distributed damage and another consisting of terrestrially surveyed

quay walls from the port in Lübeck, Germany. The simulated data set will be used to demonstrate the suitability of our chosen methodology, while the real data set will be used to confirm that our method can be applied in practice.

In Fig. 3, there is a point cloud that has been simulated to depict quay walls above the water's surface. This point cloud has been divided into three categories: quay walls (coloured blue), concrete spalling (coloured green), and additional objects (coloured red). These categories can be seen in the label image that accompanies the point cloud. The distance between points in the point cloud can vary, but is generally about 1 cm. There is also some noise present, with a range of 1–2 mm. Figure 4 displays the classified point cloud along with its corresponding depth and label image.

Experiments

In the following section, the results of various experiments will be presented. The experiments were conducted on both simulated and real data. There were a total of 500 simulated images and 395 created heightfield and label images used as the basis for the experiments. The purpose of the experiments is to determine whether the proposed method, which utilizes DNN and LOF, is effective at detecting damage on quay walls. There are two primary criteria that must be considered when assessing the success of damage detection. The first is precision, which refers to the proportion of damage regions identified by the network that are actually damaged in reality. The second is recall, which refers to the proportion of actual damage that is identified by the network. In addition to precision and recall, other measures such as accuracy (the percentage of correctly classified regions in the entire data set), the F1 score (a harmonic mean of precision and recall), and the Mathews correlation coefficient (MCC) will also be reported. The MCC takes all four categories of the confusion matrix into account and ranges from -1 (strong misclassification) to $+1$ (excellent classification). A value of 0 indicates random classification (Chicco and Jurman 2020).

Simulated data

The best results were obtained when using tile sizes of 32×32 pixels and normalized features as output from the DNN. Figure 5 shows the results for different threshold values. Using a threshold of -1.5 , the confusion matrix in Table 1 was obtained, resulting in an accuracy of 99.9%, a precision of 96.3%, and a recall of 96.2%. This leads to an F1 score of 96.3%. The MCC was also found to be 0.96, indicating a good classification. Figure 6 shows the results for two of the simulated images, with green rectangles indicating true positive (TP) classification, red indicating false negative (FN), and yellow indicating false positive (FP).

Table 1 Confusion matrix for threshold -1.5 for simulated Data

	Predicted positive	Predicted negative
Reference positive	52,217	2054
Reference negative	2008	4,050,050

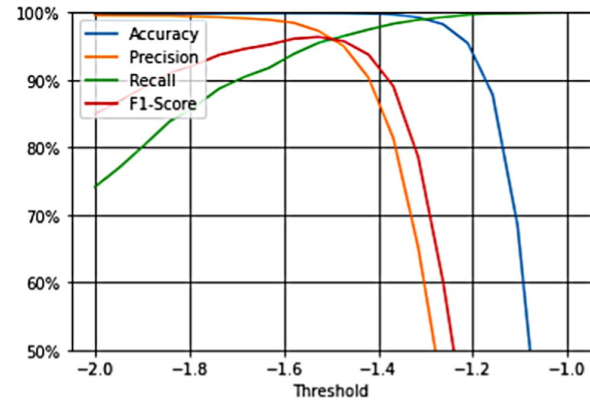


Fig. 5 Simulated data using tiles size 32×32 px and normalized features

Overall, the results using simulated data were very good, and the LOF appears to be effective for the automated classification of quay walls.

Real data

The method has shown promise in identifying potential damage areas, so the next step is to test it on a real data set using various tile sizes. The normalized features from the DNN are used as the output. The best result was achieved using a tile size of 32×32 pixels, as shown in Fig. 7.

A threshold of -1.55 was selected, and the resulting confusion matrix is shown in Table 2. At this threshold, the accuracy is 90.5%, precision is 72.2%, recall is 72.6%, and the F1 score is 72.4%. The MCC is 0.66, indicating good classification, but slightly lower than the results obtained using simulated data.

The classification result for the two images is shown in Fig. 8. Green, red, and yellow again indicate TP, FN, and FP. Here, the original heightfield is shown with higher contrast in the middle again to make grey value differences in the heightfield more visible for human eye.

Although the results were not as good as those obtained using simulated data, the method still appears to be effective when applied to real data.

The presented methodology is supposed to be used to automatically create a suspicion plan with suspected damage regions from point clouds. To be able to use the methodology in reality, all damage regions must be found as far as possible. In addition, only the damaged regions as such should be

Fig. 6 Examples of classification results on test data. Correct classification of damage is coloured in green, yellow are false positives (classified damage where there is none), and red are false negatives (did not detect the damage)

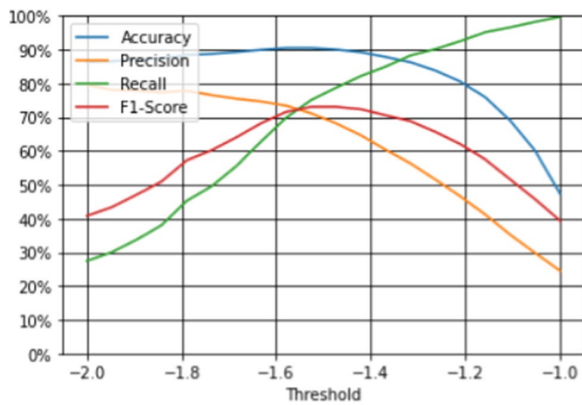
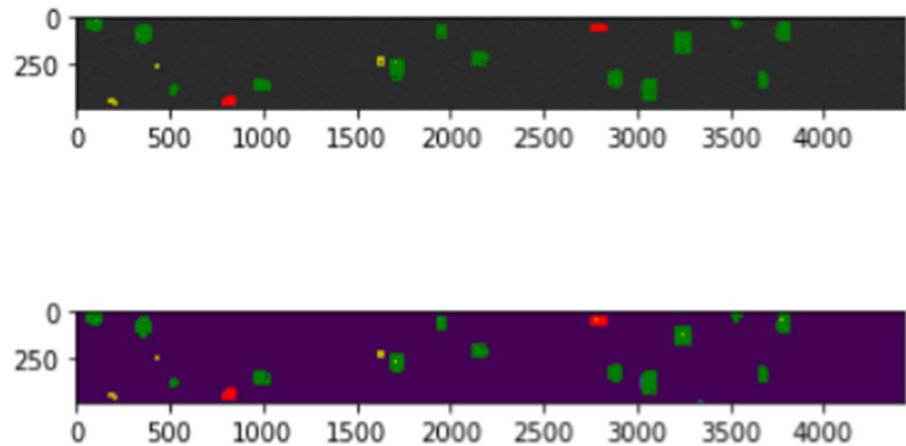


Fig. 7 Real data generated with a segmentation of 50×50 cm and tile size 32×32 px, normalized features

Table 2 Confusion matrix for threshold – 1.55 for real data

	Predicted positive	Predicted negative
Reference positive	9128	3437
Reference negative	3506	57,225

detected. This means that precision and recall together should be as high as possible. The methodology was first tested on simulated data and then applied to real data.

The method demonstrated strong performance on simulated data, with a F1 score of 96.3% and an MCC of 0.96. These results indicate that the method is suitable for generating suspicion maps of damage on quay walls, meeting the desired criteria for accuracy and completeness.

The results from the real data were slightly worse than those from the simulated data, with a F1 score of 72.4% and an MCC of 0.66. There are several potential reasons for this discrepancy. One factor could be the presence of additional objects such as ladders, fenders, plants, and ropes in the real data, which can affect the measurements. Additionally, the data has not been cleaned or filtered, potentially leading to outliers and sensor artefacts that impact the results. The noise in the real data may also not be normally distributed and could contain systematic components, in contrast to the simulated data. Finally, we used a pre-existing CAD model for the simulated data to determine distances, while we created our own model for the real data. Despite these differences, the results still have value in terms of automated damage detection and digitally assisted building inspections.

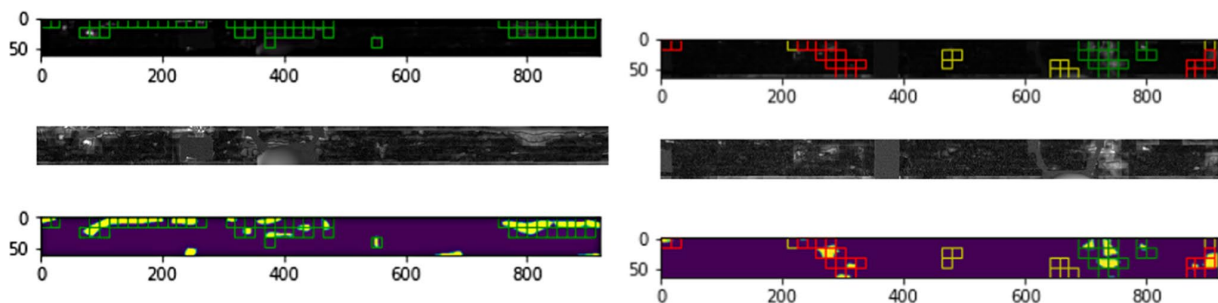


Fig. 8 Example classification real data, in the middle depth image with more contrast

Conclusion and outlook

Identifying damaged areas is important for maintaining aging port infrastructure. In this study, point clouds collected using multi-beam echo sounder and profile laser scanners were converted into depth images and processed using a pre-trained CNN with LOF extension. The automated method of creating heightfields from point clouds and then classifying them using a DNN in combination with LOF was found to be effective. The experiments yielded very good results for the simulated data, and while the classification of the real data was slightly worse, it was still sufficient. Overall, the method is useful for detecting damage and can be applied to improve inspection procedures. In the future, it may be possible to improve classification by more effectively removing additional objects from the data. Another area of focus for future research will be combining 3D point clouds and colour images to detect small cracks and non-geometric damage such as rust or efflorescence, as well as developing more advanced and knowledge-based data cleansing techniques.

Acknowledgements This work was carried out as part of the joint research project “3DHydroMapper–Bestandsdatenerfassung und modellgestützte Prüfung von Verkehrswasserbauwerken.” It consists of five partners and one associated partner: Hesse und Partner Ingenieure (multisensor system and kinematic laser scanning), WK Consult (structural inspection, BIM, and maintenance planning), Niedersachsen Ports (sea and inland port operation), Fraunhofer IGP (automatic modelling and BIM), Leibniz University Hannover (route planning and damage detection), and Wasserstraßen- und Schifffahrtsverwaltung des Bundes (management of federal waterways).

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by German Federal Ministry of Transport and Digital Infrastructure grant number 19H18011C.

Data availability The data sets in this paper were collected and used in previous work by the authors (see Hake et al. (2020) & Hake et al. (2022)).

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abati D, Porrello A, Calderara S, Cucchiara R (2019) Latent space autoregression for novelty detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, pp 481–490. <https://doi.org/10.1109/CVPR.2019.00057>
- Andrews J, Tanay T, Morton EJ, Griffin LD (2016) Transfer representation-learning for anomaly detection. In: Proceedings of the 33rd International Conference on Machine Learning. JMLR, New York
- Basharat A, Gritai A, Shah M (2008) Learning object motion patterns for anomaly detection and improved object detection. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, pp 1–8. <https://doi.org/10.1109/CVPR.2008.4587510>
- Breunig MM, Kriegel HP, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. SIGMOD Rec 29(2):93–104. <https://doi.org/10.1145/335191.335388>
- Carvalho T, de Rezende ERS, Alves MTP, Balieiro FKC, Sovat RB (2017) Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, pp 866–870. <https://doi.org/10.1109/ICMLA.2017.00-47>
- Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom 21, Nr. 1, 1–13.
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Drixler E (1993) Analyse der Form und Lage von Objekten im Raum (German). München, DGK Series C, Volume 409, Dissertation
- Feng C, Zhang H, Wang S, Li Y, Wang H, Yan F (2019) Structural damage detection using deep convolutional neural network and transfer learning. KSCE J Civ Eng 23 10 4493–4502. –ISSN 1226–7988
- García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, Vázquez E (2009) Anomaly-based network intrusion detection: Techniques systems and challenges. Comput Secur 28:18–28
- Gatys L, Ecker AS, Bethge M (2015) Texture synthesis using convolutional neural networks. In: Cortes C, Lawrence N, Lee D, Sugiyama M, and Garnett R, editors, Advances in Neural Information Processing Systems, vol 28. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2015/file/a5e00132373a7031000fd987a3c9f87b-Paper.pdf>
- Gebhardt J, Goldstein M, Shafait F, Dengel A (2013) Document authentication using printing technique features and unsupervised anomaly detection. In: 2013 12th International Conference on Document Analysis and Recognition, Washington, pp 479–483. <https://doi.org/10.1109/ICDAR.2013.102>
- Goldstein M, Uchida S (2016) A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS One 11:1–31. <https://doi.org/10.1371/journal.pone.0152173>
- Gopalakrishnan K, Khaitan SK, Choudhary A, Agrawal A (2017) Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr Build Mater 157:322–330 (–ISSN 09500618)
- Grubbs FE (1969) Procedures for detecting outlying observations. Samples Technometrics 11:1–2
- Hadavandsiri Z, Lichti DD, Jahraus A, Jarron D (2019) Concrete preliminary damage inspection by classification of terrestrial laser scanner point clouds through systematic threshold definition. ISPRS Int J Geo-Inf 8:585. <https://doi.org/10.3390/ijgi8120585>

- Hake F, Göttert L, Neumann I, Alkhatib H (2022) Using machine-learning for the damage detection of harbour structures. *Remote Sens* 14:2518. <https://doi.org/10.3390/rs14112518>
- Hake F, Hermann M, Alkhatib H, Hesse C, Holste K, Umlauf G, Kermarrec G, Neumann I (2020) Damage detection for port infrastructure by means of machine-learning-algorithms. In: FIG Working Week 2020, Amsterdam. https://www.fig.net/resources/proceedings/fig_proceedings/fig2020/papers/ts08b/TS08B_hake_alkhatib_et_al_10441.pdf
- Hesse C, Holste K, Neumann I, Hake F, Alkhatib H, Geist M, Knaack L, Scharr C (2019) 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadenserkennung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. *Hydrographische Nachr-J Appl Hydrography* 113:26–29
- Kalenjuk S, Lienhart W, Rebhan MJ (2021) Processing of mobile laser scanning data for large-scale deformation monitoring of anchored retaining structures along highways. *Comput Aided Civ Inf* 36:678–694. <https://doi.org/10.1111/mice.12656>
- Li SH, Yen DC, Lu WH, Wang C (2012) Identifying the signs of fraudulent accounts using data mining techniques. *Comput Hum Behav* 28:1002–1013
- Lin J, Keogh E, Fu A, Van Herle H (2005) Approximations to magic: finding unusual medical time series. In: 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, pp 329–334. <https://doi.org/10.1109/CBMS.2005.34>
- Neuner H, Holst C, Kuhlmann H (2016) Overview on current modelling strategies of point clouds for deformation analysis. *Allg Vermessungs-Nachr (AVN)* 123:328–339
- O'Byrne M, Schoefs F, Ghosh B, Pakrashi V (2013) Texture analysis based damage detection of ageing infrastructural elements. *Computer-Aided Civil and Infrastructure Engineering* 28:162–177. <https://doi.org/10.1111/j.1467-8667.2012.00790.x>
- Sigholm J, Raciti M (2012) Best-effort Data Leakage Prevention in inter-organizational tactical MANETs. In: MILCOM 2012 - 2012 IEEE Military Communications Conference, Orlando, pp 1–7. <https://doi.org/10.1109/MILCOM.2012.6415755>
- Simony M, Milzy S, Amendey K, Gross HM (2018) Complex-YOLO: an Euler-region-proposal for real-time 3D object detection on point clouds. *Eur Conf Comput Vision (ECCV) Workshops*.
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Preprint at <https://arxiv.org/abs/1409.1556>
- Skidmore AK (1989) A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *Int J Geogr Inf Syst* 3:323–333
- Tung SH, Weng MC, Shih MH (2013) Measuring the in situ deformation of retaining walls by the digital image correlation method. *Eng Geol* 166:116–126
- Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: *Advances in Neural Information Processing Systems 27*; Z. Ghahramani.; M. Welling.; C. Cortes.; N. D. Lawrence.; K. Q. Weinberger., Eds.; Curran Associates, Inc, 3320–3328
- Zhang G, Vela PA, Brilakis I (2013) Detecting, fitting, and classifying surface primitives for infrastructure point cloud data. In: *Computing in Civil Engineering*, pp. 589–596. <https://doi.org/10.1061/9780784413029.074>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Using Semantic Segmentation for the Damage Detection of Port and Marine Infrastructures

Frederic HAKE, Marvin SCHERFF, Ingo NEUMANN and Hamza ALKHATIB

Abstract

The ageing infrastructure in ports requires regular inspection. This inspection is currently carried out manually by sensing the entire infrastructure by hand. Such a process is costly as it requires a lot of time and manpower. To overcome these difficulties, we propose to map the harbour structure above and below water with a multi-sensor system and try to automate the classification process in terms of common damage types using deep learning approaches. In the images taken above water, damaged and undamaged zones are localised using a semantic segmentation approach. We make use of a real data set captured at JadeWeserPort Wilhelmshaven to test our approach. The images are divided into smaller sections of 512x512 pixels and these are propagated through the DeepLabv3+ architecture, a modern convolutional neural network for semantic segmentation tasks, which is trained in particular to detect corrosion or rust anomalies. We achieve with a pre-trained ResNet-50 backbone and fully supervised data set IoU scores of 96.0 % and 55.9 % for undamaged and damaged zones as well as F1-scores of 98.0 % and 71.7 %. We show that our approach can achieve a fully automated and reproducible image segmentation and damage detection which can analyse the whole structure instead of the sample-wise manual method.

1 Introduction and Motivation

The ageing infrastructure of sea and inland ports requires new technologies and methods in the preparation and implementation of life cycle management processes. The traditional processes are usually time- and labour-intensive, and should be replaced by new automated, smart and innovative measurement and analysis processes to ensure transparency, resource efficiency and reliability for a more dependable lifetime prediction.

Port infrastructure, such as quay walls for loading and unloading ships, bridges, locks and flood gates, are mostly made of concrete, bricks, steel and, in the case of very old structures, wood. They are subject to severe degradation due to special environmental conditions and human activities throughout their lifetime. The material of seaports is especially profoundly affected by saltwater, which damages the concrete structures, sheet pile walls or wooden constructions. It is crucial to detect any damage and categorise its importance to ensure the safety and stability of the infrastructure. Identifying structural damage in time allows the possibility of early maintenance to avoid expensive repairs or even more crucial damage or collapse of the infrastructure.

Nowadays, the monitoring of port infrastructural buildings is divided into the parts above and below water. The structural testing of port infrastructure above water is carried out by manual and visual inspections. The recording and documentation of the condition of damage below water involve considerably more effort; the infrastructure is tested sample-wise every 50 to 100 m; the divers slide down the structure and try to sense the wall with their hands. The results depend directly on human sensory tests. Therefore, damage inspections below water with divers are highly variable in quality and quantity. Damage classification and development are not reproducible due to subjective perception. In addition, there is usually no comprehensive inspection below water, thus, only a small percentage of the structure can be inspected by divers. One way to deal with this problem is by utilising cameras that capture the object surfaces.

The focus of this paper is on the general process of damage detection in images. We examined and optimised different image classification and segmentation methods in the field of machine and even deep learning in order to automate this task. For that purpose, one data set is created to validate the overall procedure. It is a real data set showing a concrete quay wall, which is regularly supported by steel pipes, and captured in the harbour of Wilhelms-haven, a city located in north Germany with the main focus on the detection of rust damages.

It is essential when monitoring harbour structures to assure a transparent, efficient and quality-controlled process. This can be achieved by a comprehensive visual inspection in short time intervals during the whole life cycle of the structure. However, a quality-controlled visual inspection is nearly impossible in regions such as the Ems, Weser and Elbe due to the high level of sedimentation. In this research, a fully automated, quality-controlled and reproducible mapping and damage detection of port infrastructures, above and under water, is proposed. Based on the results obtained, the port operator has more reliable information to efficiently plan maintenance and construction work. This approach will reduce the expenses significantly by lowering the downtimes of the port facilities and well-planned construction. Damage detection is usually performed in modern data processing based on pattern recognition methods (see HESSE et al. (2019) for more information). This is a reliable approach to detect any damages and make a well-founded assessment of the current state of the structure. Not only precise but also high-resolution data for the above and below water parts of the building are required for the acquisition of the building geometry and condition.

Various publications deal with comprehensive sensing methods for the structural health monitoring of infrastructural buildings. DIAZ et al. (2017) used edge detection algorithms in MATLAB in order to segment colour images. A segmentation pipeline based on wavelet transforms, principal component analysis and pattern classification was proposed by GHANTA et al. (2011). DUY et al. (2020) introduced a new architecture-based fully convolutional network for semantic segmentation to recognise rust on power electric towers. A Bayesian classifier in combination with Fisher indexes on simulated images with Perlin Noise was presented by ACOSTA et al. (2014). KHAYATAZAD et al. (2020) developed an artificial intelligence-based algorithm that can recognise corrosion damage in a series of photographic images. O'BYRNE et al. (2013) detected disturbances by the texture segmentation of colour images. GATYS et al. (2015) showed that neural networks trained on natural images learn to represent textures in such a way that they can synthesise realistic textures and even entire

scenes. Neural networks, as feature extraction, are, thus, preferred over hand-crafted ones (YOSINSKI et al. 2017, CARVALHO et al. 2017, ABATI et al. 2019). A novel data-driven fault diagnosis method is proposed based on convolutional neural networks (CNNs) by LI et al. (2019). HAKE et al. (2022) proposed a novel anomaly detection approach in which a point cloud is transferred into an image and then processed through a machine learning pipeline to detect concrete spalling.

In this work, we aim to detect structural damages in infrastructures based on colour images. We use anomaly detection algorithms due to the large imbalance between damaged and undamaged areas and the small amount of training data for the damaged areas. The automatic detection of rust damage in images is totally new in the context of structural health monitoring systems. It is now for the first time possible to detect damages in an automated manner. This opens the door for further research into digitally assisted building inspection. Therefore, the approach developed is applicable in all areas of damage detection for infrastructure objects.

The remainder of this paper is organised as follows. Section 2 presents the proposed segmentation pipeline. In Section 3, a real experiment is conducted to evaluate the effectiveness of the proposed method. The conclusions are given in Section 4.

2 Methodology

In this section, all steps from data acquisition to model training and optimisation are listed. All steps in the entire processing pipeline are illustrated in Figure 1. The remaining steps are presented later.

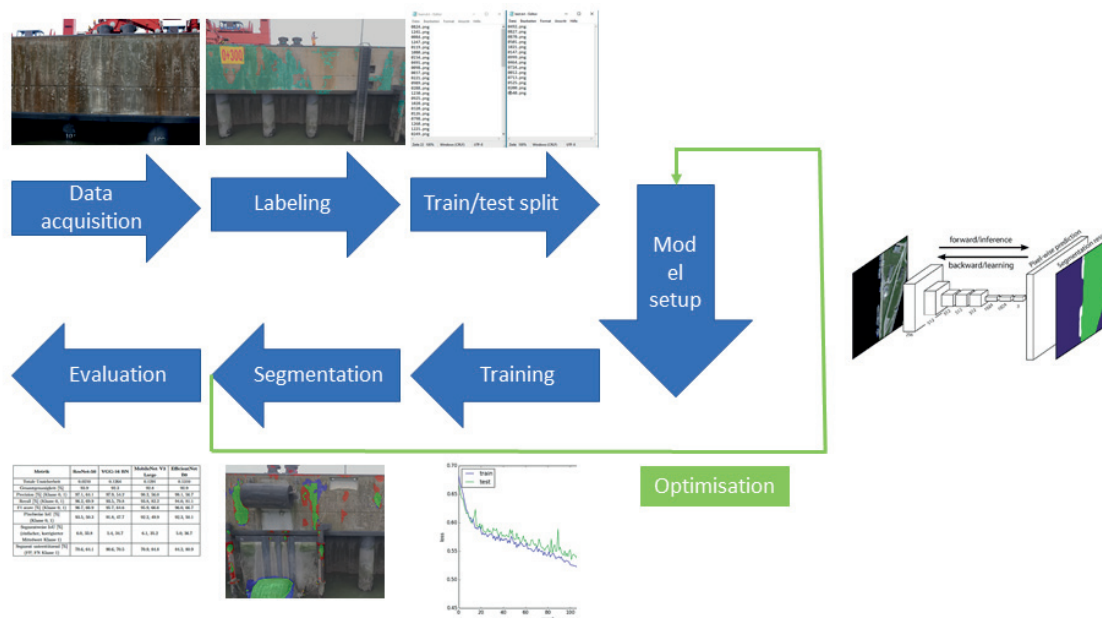


Fig. 1: Processing pipeline for machine learning related projects

For every supervised machine learning related project data availability and preparation are the key elements for a successful product. These steps help to enable the full potential of any considered model architecture. The more diligently the data is collected and pre-processed, the higher the chance to profit from the learning algorithms behind a selected method. After pre-processing, the next challenging step is to adjust the model structure and learning strategy to obtain high-quality results for general applications in the targeted field of work.

In the following subsections, we cover all the steps in Figure 1 except the model evaluation to highlight the results of different trained models afterwards. These steps include data collection of the actual measurement system, the labelling process and optimisations of the considered AI models. We have chosen the Random Forest (RF) (BREIMANN 1984) classifier as a classical machine learning model extensively examined by (TINGTING et al. 2019) and the Deeplab V3+ (CHEN et al. 2018), a state-of-the-art segmentation network developed by Google, to demonstrate the performance of self-learned image features.

2.1 Data acquisition

In order to achieve comparable detection results from the machine learning models with respect to the developed platform, a real-world data set is acquired above water using a camera-equipped drone. The images are taken orthogonally to the quay wall of the Jade-Weser-Port facility at a distance of a couple of meters while the drone hovers freely a couple of meters over the water surface. Therefore, a high-resolution digital image of the port can be analysed. We used a Canon EOS 5D Mark III camera with 35 mm fixed focal length. All settings were unchanged and the pictures were taken on one day. We estimate, that one pixel covers between two to five millimetres in each spatial dimension. The main purpose of this data set is the detection of corrosion on a concrete wall. Other types of damage occur so infrequently (maximum low double-digit number across the entire facility) that the models were unable to learn or to recognize them effectively. The detected damaged areas identified by our trained models can be directly integrated into the digital model of the port facility to track its life cycle status.

2.2 Data Preprocessing

The first task is to label the acquired image data set pixel-wise into two classes. These classes represent undamaged areas or background (sections outside the quay wall) and corrosion, which are treated as class 0 and 1 in the learning algorithms of the models. This time-consuming work was done manually by means of our own developed MATLAB program (MATLAB 2020). Each marked region is therein converted into a white segment and positioned successively onto a dark black/white 8-bit image. An example of this work is shown next to the related RGB image in figure 2.

The complete data set of the port in Wilhelmshaven consists of 1300 images. From all images, 80 are labelled manually. These images are almost uniformly spread over the facility to represent the characteristics of different corrosion variants and their appearing frequencies in a certain neighbourhood. This was done to ensure that the image and ground truth pairs ap-

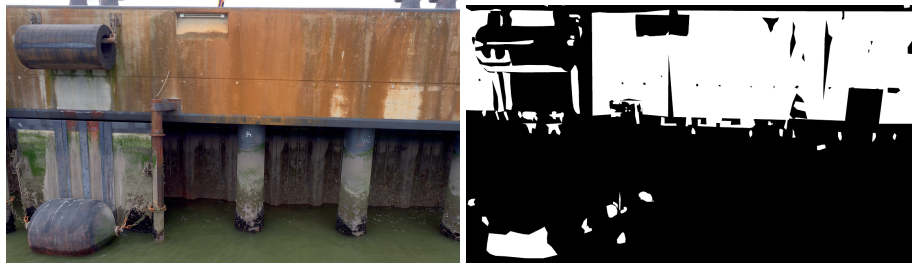


Fig. 2: Cropped RGB (left) and created ground truth image (right) of the Jade-Weser-Port in Wilhelmshaven. Undamaged and corrosion class are shown in black and white, respectively.

proximate the real distribution of RGB pixel values and related classes of this quay wall. The machine learning models were able to recognize the damage type corrosion. We expect that they will be able to differentiate between the two classes on other image data sets as well. However, this claim has not yet been tested. The training and test sets were split 80/20 and randomly assigned. The data sets were checked to ensure that they covered the specific characteristics of the port, such as lighting conditions. To validate the trained models, the training set was divided into five subsets and the cross-validation method was applied.

When the models are evaluated on the test set, the uncertainty of the ground truth images caused by non-expert knowledge has to be taken into account. Thus, small differences in metrics, like accuracy or Intersection over Union (IoU) score, do not reflect objectively, that one model predicts the true corrosion segments with higher quality in terms of shapes or positions in the scene. The same level of label noise is also included in the training data set, but the deep learning approach is able to generalise the appearance of corrosion better due to its ability to extract or learn its own features from the training images. The same phenomenon is even shown with a small corrosion data set in (NASH et al. 2019).

2.3 Segmentation with Random Forest

TINGTING et al. (2019) suggest that Random Forest outperforms other machine learning classifiers, such as Support Vector Machine, Bayesian Network, or even Multilayer Perceptrons, regarding corrosion detection by a significant amount. Thus, we adopted a Random Forest algorithm and applied it to our data set to see the segmentation performance of a classical machine learning method.

The RF classifier consists of several binary decision trees (ordered, directed graphs), which individually recursively separate the feature space into classified regions. These regions are represented by leaf nodes in each tree. To obtain a classification result for a new sample, it goes through each tree from top to bottom (root to leaf). On every depth level, the sample follows one out of two paths (edges in graph context) until a leaf node is reached. The covered way on every decision tree depends on the corresponding feature value compared to the learned threshold. The final classification is determined by the winner class of a majority

vote, in which each tree is treated equally. Every decision tree is trained on a data subset (bootstrapping) and the considered feature on each level is selected randomly. The threshold values are determined such that they lead to the highest information gain ($G(Q_m, \Theta)$) after splitting a subset of the feature space along the individual considered axis. This degree of separation is measured either using entropy or Gini impurity. Thus, they can be substituted for one of the listed H quantities in equation 1 (in accordance to the optimisation purpose).

$$G(Q_m, \Theta) = \frac{n_m^{left}}{n_m} H(Q_m^{left}(\Theta)) + \frac{n_m^{right}}{n_m} H(Q_m^{right}(\Theta)) \quad (1)$$

$$H_{entropy}(Q_m) = -\sum_k p_{mk} \log(p_{mk}) \quad (2)$$

$$H_{gini}(Q_m) = \sum_k p_{mk} (1 - p_{mk}) \quad (3)$$

where Q is a subset of the considered n samples, m corresponds to a particular node in a decision tree, $\Theta = (j, t_m)$ indicates a possible candidate split consisting of the feature j and threshold t_m and, finally, *left* and *right* refer to the split subsets (BREIMANN 1984).

The Random Forest classifier uses RGB pixel values and the improved HLS colour space as features to differentiate between two classes. The IHLS colour space is created using non-linear functions from the original pixel values and has been shown to improve the F1 score by 3.5 % in the experiments of TINGTING et al. (2019). We transformed the data set using Principal Component Analysis (PCA) to improve the RF's performance.

To optimize the Random Forest classification model, we conducted a grid search for the number of decision trees and the class weights. The initial values were based on previous research (TINGTING et al. 2019) and the "balanced" setting of the RF class in the Scikit-learn Python package (PEDREGOSA et al. 2011). The search intervals for the number of trees were from 8 to 14 with a step size of 2, and the weights for the corrosion class ranged from 2 to 5. The weight for class 0 was always set to 1. This optimization was done with respect to the weighted F1 score between the two highly imbalanced classes, using the RGB features. Further parameters corresponding to bootstrapping, minimal samples per leaf node, maximal tree depth, considered features and information gain quantity were fixed for the entire process as 'False', 0.01 %, 14, None (all features) and 'entropy', respectively. To reduce the training time, we used a subset of the data with a few hundred million pixels, and we prevented overfitting by restricting the depth of the decision trees and optimizing the model in terms of the entropy metric (see chapter 4). We trained several models with different parameter pairs, and found that 10 decision trees with a class weight ratio of 1:3 achieved the highest F1 score. We then trained two additional instances of the same model, one using only the IHLS colour space and the other using all six colour features. We did not perform further hyperparameter optimization as we did not expect the model to require a more complex understanding of the data. To validate the model, we trained five different models on different parts of the training data set and evaluated them on the test set.

2.4 Segmentation with DeeplabV3+

The DeepLabV3+ architecture (CHEN 2018) was chosen for this study due to its success in various computer vision tasks. In order to adapt it to the specific task of identifying corroded areas in the images, the structure and some parameters of the network were modified (s. Figure 3). One such modification was the doubling of the dilation parameters within the Atrous Spatial Pyramid Pooling (ASPP) block. The modified network was then trained and evaluated using various feature extraction networks ("backbones") and loss functions with the goal of maximizing the IoU score for the corrosion class. The IoU score is a measure of the overlap between the predicted segment and the corresponding ground truth label, expressed as a percentage of their union. This is a common metric in the field of semantic segmentation.

2.4.1 Structure and Modifications of the DeepLabV3 Network

The Deeplab V3+ network is an encoder-decoder architecture that processes images. The encoder extracts features from the image and the decoder reconstructs the image by assigning class probabilities to each pixel. The encoder reduces the spatial resolution of the original image and increases the number of channels while producing a compressed representation of the image.

Initially, an image is passed through the encoder of a deep convolutional neural network (DCNN) and the extracted feature maps are then analysed in an ASPP block. The CNN in CHEN et al. (2018) is named Modified Aligned Xception Net and represents an optimised version of the original Xception network (CHOLLET 2016). For our testing, we also considered further CNNs, which consist of a significantly lower number of trainable parameters and their architecture are based on other basic blocks. Among them, especially the inverted residual block occurs frequently (MARK et al. 2019), which consists of lower number of parameters than the regular residual block and achieves the same or even superior performance. All of the networks are pretrained on the ImageNet-1K data set (DENG 2009) (s. torchvision documentation (MARCEL 2010)) to be more robust against overfitting in the training procedure and provide suitable low-level features, which are transferred directly to the decoder part in order to maintain the object boundaries in the final segmentation image. The feature maps at the end of the CNN extract more detailed information in regards to the image scale by using different types of convolution operations and a Global Average Pooling (GAP) layer.

2.4.2 Processing and Training the DeepLabV3 Network

To improve the accuracy of the Deeplab V3+ network, the input images are split into smaller patches of 512x512 pixels. This is done to avoid running out of memory during training, but also because the network is not able to effectively utilize larger images due to the restricted receptive field in the output layer or the maximal size of corroded areas. Currently, we construct the segmented image by blending the predicted images smoothly together (CHEVALLIER 2017). These image patches already provide a significant amount of contextual information for the segmentation task, and our experiments have shown that adding an overlap of 25 % to the training images can improve the model's performance by

up to 1-2 % for certain metrics listed in table 1. This is likely due to the different perspectives the model is able to gain on individual corrosion segments, especially when they are located near the edges of the original patch. However, its important to note that the performance may degrade due to overfitting if the overlapped area exceeds around 40 %.

To prepare the images for training with the Deeplab V3+ network, they are normalized based on the mean and standard deviation of the ImageNet data set. This helps to take full advantage of the pretraining process. In addition, we apply various image augmentation techniques to artificially expand the data set and reduce overfitting. These include random resized cropping, rotations, horizontal and vertical flipping, colour jittering, and Gaussian blur. These operations are applied with certain probabilities to each image, and are designed to create realistic phot photos that can be used for training.

2.4.3 Optimization and Evaluation of the DeepLabV3 Network

In order to improve the performance of the Deeplab V3+ network, we made some changes to its structure and tested different feature extraction networks and loss functions. These changes included adding dropout layers and increasing the filter size within the ASPP block. A schematic overview is illustrated in Figure 3.

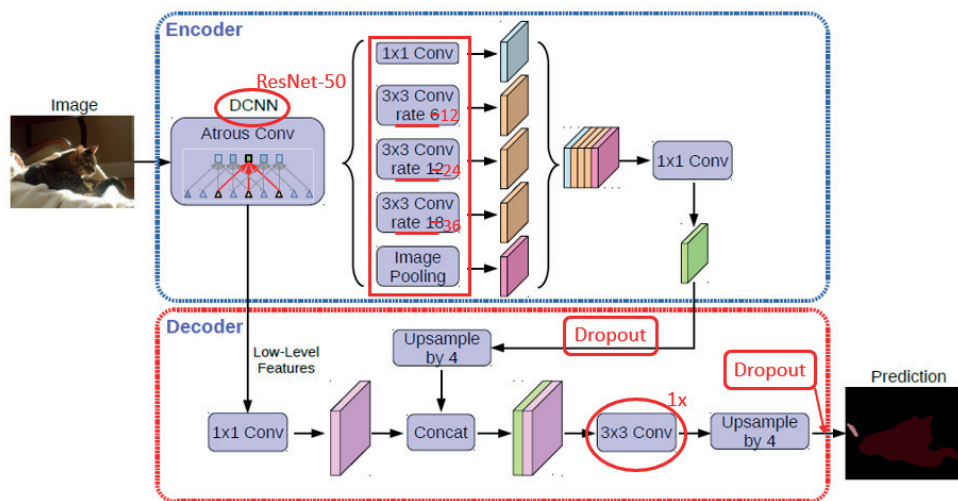


Fig. 3: Modified structure of the Deeplab V3+ architecture (edited figure from CHEN et al. 2018). Changes are highlighted in red.

We tested VGG-16 BN and ResNet-50 (version 1.5) as feature extraction networks and made our decisions based on the IoU score for the minority class. We tested different combinations of loss functions and feature extraction networks to optimize the performance of our deep learning model, specifically focusing on improving the IoU score for the minority class. The loss functions included both distribution-based and region-based methods, such as categorical cross-entropy and IoU loss. The backbones consisted of popular architectures from

2014 to 2019 that have been successful on large classification benchmark data sets, and have fewer parameters than the Xception Net. We tested ResNet-50, VGG-16 BN, MobileNet V3 Large, and EfficientNet B0, which were chosen for their availability in the Pytorch framework (PASZKE 2019) and to limit the number of experiments. To address the imbalanced data, we introduced class weights to the error functions, with a weight of 5 for the corrosion class in the ResNet model, which was found to be the best model according to the IoU score. The best performing combinations of backbones and loss functions were ResNet-50 with IoU loss, VGG-16 BN and MobileNet V3 Large with categorical cross-entropy loss, and EfficientNet B0 with Dual Focal loss (HOSSAIN et al. 2021). The model's performance was validated using the same procedure as in the Random Forest case, with model parameter states taken from the epoch with the lowest loss score on the remaining training subset.

3 Application to real Data

3.1 Segmentation with Random Forest

In this section, we evaluate the three trained Random Forest classifiers to assess how useful the Improved HLS colour space is for the corrosion task. The individual metrics of the models associated with the RGB and both colour spaces are listed in table 1.

The total uncertainty (model and data uncertainty) computed by means of the entropy based on the probabilistic or Softmax output (in case of the segmentation network) is quite similar between the three models. Since these were determined with the same images, this also corresponds directly to the distance in terms of model uncertainty. The model trained on the IHLS colour space achieves 0.2733 compared to 0.2538 by the RGB related classifier. This is also reflected in the other variables such as the overall accuracy, the F1 score and the pixel-wise agreement. In these rather general metrics, the RF trained on the RGB colour space outperforms the IHLS model by up to 2.0 % and even 8.0 % for the background and corrosion class, respectively. We have therefore not observed the same behaviour as TINGTING et al. (2019). While the PCA pre-processing has a positive effect on the RGB model by improving for instance the IoU of corrosion class by approximately 3 %, the IHLS results remain the same.

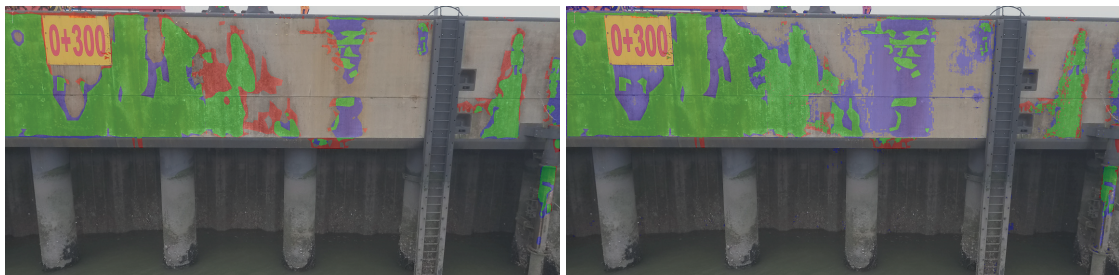
With the classifier trained on both colour spaces, we see results in between these RF classifiers, but closer to the RGB model performance. The three general metrics fall short by up to 0.6 % (only 0.3 % for background class). Further experiments showed that an increased maximum depth of each decision tree in this model had only a minimal effect on improving the results, but the RGB model's performance could not be matched. This shows that more information doesn't have to improve the problem understanding in general as stated in the earlier assumption. Due to the small difference some prediction images may profit from the additional knowledge nevertheless, but we can not say this with for sure. Figure 4b shows one example image from our data set with overlaid classification results. Correct classification of corrosion is coloured in green, blue are false positive (classified corrosion where there is none) and red is false negatives (did not detect the corrosion).

3.2 Segmentation with DeeplabV3+

The results of the different Deeplab V3+ versions are compared in the section below and table 1 is completed by the best and worst segmentation network to demonstrate the wide range of outcomes. The Deep Learning models were trained for a maximum of 35 epochs and their states of trainable parameters were saved and used afterwards where the mean error with respect to the test set was the smallest.

In terms of overall uncertainty, the ResNet model performs significantly better than the other models, particularly due to its use of the intersection over union (IoU) loss function. When considering overall accuracy, the ResNet model performs the best, followed by the MobileNet and EfficientNet models with 95.5 % and 95.4 % accuracy, respectively. The VGG-16 Deeplab model has the lowest accuracy at 95.2 %. The majority class (class 0) has a F1 score of 97.4 % or higher for all models. However, the performance for class 1 is lower, with F1 scores ranging from 68.4 % to 71.7 % (in order from the VGG-16 to the ResNet model). The IoU values follow a similar trend, but generally have lower values. For the damage-free areas, the IoU values are about 2 % lower, and the predicted corrosion class overlaps with the ground truth by $53.9 \% \pm 2.0 \%$.

Figure 4a show a example image from our data set with overlaid classification results. Correct classification of corrosion is coloured in green, blue is false positive (classified corrosion where there is none) and red illustrates false negatives (did not detect the corrosion).



(a) Segmented images from ResNet-50 Deeplab V3+ segmentation model (b) Segmented images from Random Forest classifier based on RGB features

Fig. 4: Correct classification of corrosion is coloured in green, blue are false positive (classified corrosion where there is none) and red false negatives (did not detect the corrosion)

Table 1 give an overview of different trained Random Forest and Deeplab V3+ models regarding common metrics calculated on the same corrosion test set.

3.3 Validation

To evaluate the generalization ability of the trained machine learning models, we applied the cross-validation method described in Chapter 2. The results for the best model are shown in Table 2, along with the same evaluation metrics as before.

Table 1: Results of chosen Random Forest and Deeplab V3+ classifiers

Metric	Random Forest		Deeplab V3+	
	RGB	RGB & IHLS	ResNet-50	VGG-16 BN
Total uncertainty	0.2538	0.2618	0.0161	0.1121
Overall accuracy [%]	94.1	93.8	96.2	95.2
F1 Score [%] (class 0, 1)	96.8, 59.9	96.6, 59.3	98.0, 71.7	97.4, 68.3
Pixel-wise IoU [%] (class 0, 1)	93.8, 42.8	93.5, 42.2	96.0, 55.9	95.0, 51.9

Table 2: Cross validation results of the RGB Random Forest and ResNet-50 Deeplab V3+ model

Metric	RGB Random Forest		ResNet-50 Deeplab V3+	
	Mean	Standard dev.	Mean	Standard dev.
Total uncertainty	0.2519	0.0084	0.0181	0.0015
Overall accuracy [%]	94.1	0.3	95.8	0.3
F1 Score [%] (class 0, 1)	96.8, 59.7	0.2, 1.0	97.8, 69.4	0.2, 0.5
Pixel-wise IoU [%] (class 0, 1)	93.8, 42.5	0.3, 1.0	95.4, 53.2	0.2, 0.6

The results of the cross-validation evaluation show that both the Random Forest and Deeplab models perform similarly overall (as indicated by the mean values in the table). However, the Deeplab model demonstrates a significant increase in the detection rate of corrosion (2.7 %increase) compared to the Random Forest model, which is almost unchanged by the extra data fold (20 % of the training set). Additionally, the results of the Deeplab model are more consistent across the five separate models, as indicated by the lower standard deviation values. In particular, the class 1 scores for the RF model vary around a factor of two and it's prediction uncertainty is almost six times higher compared to the trained Deeplab models. This demonstrates the superiority of self-learned features of the Deep Learning approach.

4 Discussion and Outlook

In this work, different machine and deep learning models have been investigated for the segmentation of damaged areas on a port facility image data set. In an attempt to promptly detect the emergence of damages and in order to be able to follow their emergence and temporal development in a Building Information Modeling (BIM) approach with high accuracy. These models are to be used to automatically evaluate the images captured by an Multi-Sensor-System (MSS) platform and drone, primarily to support the structural inspectors in assessing the facility.

Using the example of a real data set from Wilhelmshaven, which is characterised by various forms of corrosion damage, the machine learning architectures and their training procedures were optimised and evaluated.

Evaluation of the results shows that the approach with Deeplab V3+ and ResNet-50 delivers only 2.1 % better overall accuracy than the adapted approach with the Random Forest classifier. On the other hand, the precision score of the Deeplab V3+ approach with ResNet-50 for class 1 is 11.8 % higher than in the RF classifier and the precision for class 0 is still slightly better with 1.2 %. Similarly, when comparing the pixel-wise IoU score, the Deeplab V3+ approach is 13 % higher for class 1 than the RF classifier and the IoU for class 0 is also roughly 2 % higher. By comparing the results from the two models, it is clear that the Deeplab model performs better in terms of reducing false positives and increasing false negatives. False negatives are areas that were not identified as corrosion but were actually corroded. These areas are typically (in more than 75 % of the cases) located near true positives, which means that they are still within the range of the damage detected by the model. This is not a significant problem in the context of building inspection, as any detected areas of damage will still be examined by a building inspector. Overall, the performance of the Deeplab model demonstrates the superiority of self-learned features in comparison to the traditional machine learning approach.

Therefore, the approach with Deeplab V3+ delivers better results in the segmentation of rust spots in comparison to the classical approach represented by the Random Forest classifier. It is possible to improve the results by including additional scales. For example, in the ASPP block of the Deeplab encoder, it is not only possible to use the kernel spacing once set up to three times. This has so far been increased from six to twelve, as it has led to considerable improvements, but this makes it difficult to detect smaller damage with high geometric accuracy. This can be investigated in more detail in the future with a multi-dimensional grid search. In addition, the number of Atrous convolutions could also be increased to cover multiple scales regardless of the data set.

The comparison of the different feature spaces in table 1 has shown that the use of the IHLS colour space has not achieved the improvements in accuracy shown in (TINGTING et al. 2019). This is probably due to the different data sets and objects. But the RF model trained on the original image data could be improved strongly by means of the PCA pre-processing.

So far, we have only been able to test our algorithm on one real data set of a sheet pile wall with a concrete spar. However, we are working towards acquiring more real-world data with different materials and building types. The proposed strategy is also applicable to other infrastructure objects, such as bridges, high-rise buildings, and tunnels.

Funding: This research was funded by Federal Ministry of Transport and Digital Infrastructure grant number 19H18011C.

Acknowledgments: This work was carried out as part of the joint research project "3DHydroMapper". It consists of five partners and one associated partner: Dr. Hesse und Partner Ingenieure, WK Consult, Niedersachsen Ports, Fraunhofer IGP, Leibniz University Hannover and Wasserstraßen- und Schifffahrtsverwaltung des Bundes.

Bibliography

- ABATI, D., PORRELLO, A., CALDERARA, S. & CUCCHIARA, R. (2019): Latent Space Autoregression for Novelty Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Long Beach, CA, USA, 15-20 June 2019, pp. 215-232.
- ACOSTA, M. R. G., DÍAZ, J. C. V., & CASTRO, N.S. (2014): An innovative image-processing model for rust detection using Perlin Noise to simulate oxide textures. In: Corrosion Science, pp. 141-151, DOI: 10.1016/j.corsci.2014.07.027.
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. & STONE, C. J. (1984): Classification and Regression Trees. Wadsworth 10.2307/2530946.
- BREIMAN, L. (2001): Random Forests. In: Machine Learning 45, pp. 5–32, DOI: 10.1023/A:101093-3404324.
- CARVALHO, T., DE REZENDE, E. R. S., ALVES, M. T. P., BALIEIRO, F. K. C. & SOVAT, R.B. (2017): Exposing Computer Generated Images by Eye's Region Classification via Transfer Learning of VGG19 CNN. In: 2017 16th IEEE International Conference on Machine Learning and Applications, pp. 866-870, DOI: 10.1109/ICMLA.2017.00-47.
- CHEN, L., ZHU, Y., PAPANDREOU, G., SCHROFF, F. & ADAM, H. (2018): Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. arXiv 10.48550/ARXIV.1802.02611.
- CHEVALLIER, G. (2017): Using a U-Net for image segmentation, blending predicted patches smoothly is a must to please the human eye. Github Last visit: 12/22/2022. github.com/Vooban/Smoothly-Blend-Image-Patches.
- CHOLLET, F. (2017): Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 1251-1258, 10.48550/arXiv.1610.02357.
- DENG, J., DONG W., SOCHER, R., LI, L.-J., KAI, L. & LI, F.-F. (2009): ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition 248-255, 10.1109/CVPR.2009.5206848.
- DIAZ, J. A. I., LIGERALDE, M. I., JOSE, A. C. & BANDALA, A. A. (2017): Rust detection using image processing via Matlab. In: 2017 IEEE Region 10 Conference, pp.1327-1331, DOI: 10.1109/TENCON.2017.8228063.
- DUY, L. D., ANH, N. T., SON, N. T., TUNG, N. V., DUONG, N. B. & KHAN, M. H.R. (2020): Deep Learning in Semantic Segmentation of Rust in Images. In: Proceedings of the 2020 9th International Conference on Software and Computer Applications, pp. 129–132, DOI: 10.1145/3384544.3384606.
- GATYS, L., ECKER, A., ALEXANDER S & BETHGE, M. (2015): Texture synthesis using convolutional neural networks. In: Advances in neural information processing systems. pp. 262-270.
- GHANTA, S., KARP, T. & LEE, S. (2011): Wavelet domain detection of rust in steel bridge images. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1033-1036, DOI: 10.1109/ICASSP.2011.5946583.
- HAKE, F., GÖTTERT, L., NEUMANN, I. & ALKHATIB, H. (2022): Using Machine-Learning for the Damage Detection of Harbour Structures. In: Remote Sensing 14, DOI: 10.3390/rs14112518.

- HESSE, C., HOLSTE, K., NEUMANN, I., HAKE, F., ALKHATIB, H., GEIST, M., KNAACK, L. & SCHARR, C. (2019): 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadenserkenkung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. In: *Hydrographische Nachrichten* 113, pp. 26-29.
- HOSSAIN, S., BETTS, J. M. & PAPLINSKI, A. P. (2021): Dual Focal Loss to address class imbalance in semantic segmentation. In: *Neurocomput*, pp. 69-87, DOI: 10.1016/j.neucom.2021.07.055.
- KHAYATAZAD, M., DE PUE, L. & DE WAELE, W. (2020): Detection of corrosion on steel structures using automated image processing. In: *Developments in the Built Environment*, pp. 100022, DOI: 10.1016/j.dibe.2020.100022.
- LI, G., DENG, C., WU, J., XU, X., SHAO, X. & WANG, Y. (2019): Sensor Data-Driven Bearing Fault Diagnosis Based on Deep Convolutional Neural Networks and S-Transform. In: *Sensors* 19, 2750. DOI: 10.3390/s19122750.
- MARCEL, S. & RODRIGUEZ, Y. (2010): Torchvision the Machine-Vision Package of Torch. In: *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 1485-1488. Association for Computing Machinery, Firenze, Italy. DOI: 10.1145/1873951.1874254.
- MATLAB (2020): Version 9.8.0 (R2020a). The MathWorks Inc. Natick, Massachusetts.
- NASH, W. T., POWELL, C. J., DRUMMOND, T. & BIRBILIS, N. (2019): Automated Corrosion Detection Using Crowd Sourced Training for Deep Learning. In: *Corrosion* 76 (2), pp. 135-141.
- O'BYRNE, M., SCHOEFS, F., GHOSH, B. & PAKRASHI, V. (2013): Texture analysis based damage detection of ageing infrastructural elements. In: *Computer-Aided Civil and Infrastructure Engineering* 28 (3), pp. 162-177.
- PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N., ANTIGA, L., DESMAISON, A., KOPF, A., YANG, E., DEVITO, Z., RAISON, M., TEJANI, A., CHILAMKURTHY, S., STEINER, B., FANG, L., BAI, J. & CHINTALA, S. (2019): PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: WALLACH, H., LAROCHELLE, H., BEYGELZIMER, A., D'ALCHÉ-BUC, F., FOX, E. & GARNETT, R. (Eds.): *Advances in Neural Information Processing Systems*, 32. Curran Associates, Inc., Vancouver, Canada.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. (2011): Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12, pp. 2825-2830.
- TINGTING, L., KAI, K., FEN, Z., JIALIANG N. & TIANYUN W. (2019): A corrosion detection algorithm via the random forest model. In: ZHAOHUI L. (Ed.): *17th International Conference on Optical Communications and Networks (ICOON2018)*, International Society for Optics and Photonics, 16-19 November 2018, Zhuhai, China.
- YOSINSKI, J., CLUNE, J., BENGIO, Y. & LIPSON, H. (2014): How transferable are features in deep neural networks? In: GHARAMANI, Z., WELLING, M., CORTES, C., LAWRENCE, N. & WEINBERGER, K. Q. (Eds.): *Advances in Neural Information Processing Systems*, Curran Associates, Inc., San Francisco, USA.

ADAPTATION OF DEEPLAB V3+ FOR DAMAGE DETECTION ON PORT INFRASTRUCTURE IMAGERY

M. Scherff¹, F. Hake¹, H. Alkhatib¹

¹ Geodetic Institute, Leibniz Universität Hannover, Germany - scherff,hake,alkhatib@gih.uni-hannover.de

KEY WORDS: Image segmentation, Damage Detection, Deep Learning, Optimization, Supervised, Weakly Supervised

ABSTRACT:

Regular inspection and maintenance of infrastructure facilities are crucial to ensure their functionality and safety for users. However, current inspection methods are labor-intensive and can vary depending on the inspector. To improve this process, modern sensor systems and machine learning algorithms can be deployed to detect defects based on rapidly acquired data, resulting in lower downtime. A quality-controlled processing chain allows to provide hence informed uncertainty assessments to inspection operators. In this study, we present several Deeplab V3+ models optimized to predict corroded segments of the quay wall at JadeWeserPort, Germany, which is a dataset from the 3D HydroMapper research project. Our models achieve generally high accuracy in detecting this damage type. Therefore, we examine the use of a Region Growing-based weakly supervised approach to efficiently extend our model to other common types in the future. This approach achieves about 90 % of the results compared to corresponding fully supervised networks, of which a ResNet-50 variant peaks at 55.6 % Intersection-over-Union regarding the test set's corrosion class.

1. INTRODUCTION

Infrastructure buildings must meet high construction standards and require regular maintenance throughout their complete life cycle to ensure safe and reliable usage for its users like pedestrians, workers or even machines. Scheduled inspections are performed to detect emerging defects that could affect the facility's functionality and stability, supplemented by additional inspections as needed. National legislation, inspired by norm standards, typically regulates this process. Commissioned inspection companies compile reports detailing damaged parts or sections and providing a complete infrastructure assessment, allowing for informed maintenance decisions.

Besides commonly known infrastructure objects as bridges or tunnels, this concept has also to be applied for port facilities or similar building objects such as watergates or canals. However, the traditional approach of performing inspections by hand, with only basic tools for documentation, causes significant downtime for the facility, especially in underwater sections where divers face complicated visual conditions. This downtime leads to high costs for the facility operator. To shorten this process step, an automated inspection process should be introduced, which rely on the capabilities of today's high accurate sensors. An example of such capabilities can be found within the 3D HydroMapper research project, in which (Hesse et al., 2019) developed a swimming multi-sensor system (MSS) sensing the environment below and above the water surface. By that approach, the complete facility can be captured within several minutes or a few hours depending on its size, allowing to detect for instance regular appearing corrosion spots in the so-called splash zone more objectively afterwards. Therefore, various software systems can assist building inspectors digitally by providing data-driven predictions and optionally uncertainty information for damaged areas in order to assess the infrastructure knowledgeably.

In this study, we aim to enhance the damage detection process by adapting the Deeplab V3+ segmentation network (Chen et al., 2018) to analyze an image corrosion dataset obtained from

the 3D HydroMapper project. Initially, we establish a baseline model, and explore different combinations of backbone networks and loss functions to create a robust feature-rich network. Subsequently, we conducted an extensive performance analysis and examined the impact of introducing weakly supervised ground truth images obtained through the Region Growing algorithm during the training process.

The structure of the paper is as follows: Chapter 2 gives a brief overview on related studies to damage detection of infrastructure buildings, with a focus on corrosion investigations. In Chapter 3, we describe the methodology used for the dataset acquisition and preparation. Chapter 4 presents the process of obtaining the best performing segmentation model in detail. Subsequently, we provide and discuss the evaluation results of the validation and test sets. Finally, Chapter 6 presents the conclusion and an outlook of further investigations.

2. RELATED WORK

The field of infrastructure inspection can nowadays be approached using a wide variety of sensors or tools. Among these, cameras are an excellent acquisition instrument that can effectively examine building surfaces due to their universal and simple deployment. Because of the progress achieved in image processing and computer vision, especially in the last decade, early damage detection became a feasible goal. This particular task has been investigated recently i. a. by (Duy et al., 2020), (Katsamenis et al., 2020) and (Tanveer et al., 2022). They were able to detect different damage types to a certain extent by means of Deep Learning (DL) segmentation models on concrete and metal surfaces, respectively. However, some damage types are more recognizable than others due to their colour, shape or size. While corroded segments can be unambiguously distinguished from the concrete quay wall and background in general at a certain severity level, their shape and size can vary significantly, which may lead to erroneous detections. On the other hand, for instance cracks can have more similar properties overall, but individual examples can be overlooked due to their small width

or low resolution. Nevertheless, DL models like Convolutional Neural Networks (CNN) are currently the dominant objective in this research field and have a high potential to classify, detect, or even segment these damaged areas correctly due to their self-learning feature capabilities. Hence, this image processing strategy is a promising approach to digitally assist the inspection operator effectively.

In the study by (Duy et al., 2020), corroded areas on electric poles are detected using several architectural modified segmentation networks ranging from hundred thousands to ten millions trainable parameters. Three models were evaluated with and without prior background removal, achieving high Intersection-over-Union (IoU) scores for the pole and background class, but struggled to detect corrosion reliably. To overcome this limitation, the dataset was extended with a focus on the corrosion class, and the Mask R-CNN architecture (He et al., 2017) was investigated to partition these segments based on their severity level. But this network type was already examined with common CNNs by (Katsamenis et al., 2020) in the same year achieving 71 % F1 score on another corrosion dataset. In their experiments, Mask R-CNN outperforms the other candidates significantly. Their purposely developed boundary refinement approach, a data projection method dividing the model prediction in confident and fuzzy regions, effects the quantified results of this model barely. In addition, (Tanveer et al., 2022) evaluated the performance of various low parameter segmentation CNNs compared to the Deeplab V3+ architecture with ResNet as backbone network. The group trained these on a concrete dataset containing the damage types cracks, efflorescence, rebar exposure and spalling and found on average a relative maximum discrepancy of around 17 % in terms of F1 score to the reference. All these groups use Cross-entropy loss for the classification purpose in model training and no extensive optimization or evaluation was performed. Moreover, only a single model instance was learned, which limits the uncertainty investigation. Here comes (Nash et al., 2022) into play. This group trained different models of modified HRNetV2 (Sun et al., 2019) networks by Variational Inference, Monte Carlo Dropout and an Ensemble method on a further corrosion dataset. These Bayesian transformed models were evaluated especially in terms of aleatoric and epistemic uncertainty. In our study, we combine these concepts to provide an informed overview of different modified Deeplab V3+ models trained on an optimal loss function that is selected from a pool of distribution- and region-based error terms using our self-labeled port dataset. In contrast, (Hake et al., 2023b) focused only on single trained model instances and restricted architecture changes with the smaller and differently structured 3D HydroMapper dataset.

The most precise network structure is iteratively trained on simulated weakly supervised label images generated once by a Region Growing algorithm (Adams and Bischof, 1994) before any model training begins. However, (Huang et al., 2018) proposed a more sophisticated approach by using this algorithm on seed cues generated by a classification network in their segmentation pipeline. Their Deep Seeded Region Growing method works hence on image-level label information. We aim to estimate still the expected performance decline for this straight-forward approach when the manual labeling effort is significantly reduced in our investigation. These findings will guide our approach to further datasets containing other damage types that will be fed to the same model architecture to support the infrastructure inspection process. It is currently intended to mark a small region in the damage segment instead of bordering it completely by

a polygon if that procedure is proven successfully. Due to the nature of the current dataset, we cannot efficiently benefit from bounding box or class activation map approaches.

3. DATASET

3.1 Data acquisition

The dataset used in this work was specifically acquired within the 3D HydroMapper project to aid the development of AI-assisted software for future port inspections based on MSS data. This image sequence shows the quay wall of the JadeWeserPort in Wilhelmshaven (Germany) in an overlapping manner. We observe similar coverage of up to 80 % between subsequent captures. Occasionally, the perspective is adjusted during the acquisition process by flying closer to the wall. This provides a wide variety of real-world information for the detection models to be trained on. The complete dataset was recorded on a single day using a controlled drone equipped with a Canon EOS 5D Mark III reflex camera, hovering or flying over the water a few meters in front of the wall. For this project the focal length was fixed at 35 mm and the resolution of pre-processed images are either 5760 x 3840 or cropped to 4608 x 3456 pixels. The dataset consists of 1300 high-resolution digital images of the infrastructure, where the object pixel size varies approximately between 2 and 5 mm depending on the individual capture. Due to this, we can find fine or arising defects and achieve accurate results in terms of position and shape of the damaged segments. This allows us to retrace the growing process over time and modify the inspection period or maintenance measures accordingly to related predictions. The labeling process neglected damages besides corrosion, such as cracks, efflorescence, or spalling due to their seldom occurrence. Therefore, our trained models cannot detect them yet. However, (Hake et al., 2023a) have already demonstrated an AI-assisted approach to detect geometrical related damage types in heightfields extracted from point cloud data. Currently, we had images of an efflorescence-damaged watergate available, which were taken by means of a tripod at several locations within this facility. Hence, these are not suitable for the current inspection concept based on a swimming or consistent moving sensor platform.

3.2 Generate ground truth images

To train the supervised segmentation models, we required labeled images indicating which pixels or segments correspond to the corrosion class. Initially, our own Matlab program was used to obtain these ground truth images, which converts manually drawn polygons successively into a 8-bit black/white (b/w) image. Later on, we used the service of Supervise.ly (Supervisely contributors, 2023) to complete our currently considered dataset consisting of 84 image-label pairs. Thereon, we can incorporate model predictions directly into the labeling process. These image pairs are uniformly distributed over the quay wall, showing a big portion of the total infrastructure, because overlapping areas were avoided for the most part. The ground truth images are based on our knowledge in this field and the gained experience over the time of this procedure, but have not been validated yet by experts. However, due to investigations in handling label noise in the DL context especially by (Nash et al., 2019), particular for corrosion detection, we expect no significantly improved corrosion detection capabilities. Nonetheless, we can not ensure that the achieved metric scores listed in chapter 5 correspond to the true values, especially with the provided number of digits.

To adapt the Deeplab V3+ structure for the corrosion segmentation in the beginning by means of cross validation, and evaluate the performance of the various models methodically correct afterwards, we divide the created dataset randomly into six subsets à 14 images each. Four of them form the train set, and each of the remaining ones forms the validation and test sets, respectively. The validation set determines the optimal parameter state of any model based on the corresponding lowest epoch-wise loss value, and the test images act as independent evaluation measures. An example of the aforementioned image-label pairs is illustrated in Figure 1. As visualized by the ground truth image (black: damage-free/background, white: corrosion), corroded segments represent the minority class of this segmentation task. Because damage detection has to deal regularly with highly imbalanced datasets due to its underlying nature, models will by design favor the majority class to minimize the prediction error or maximize the total accuracy. To reorder the model's behaviour in terms of equal class detection abilities, data processing itself can be modified or the optimization process of the considered learning method has to be adjusted. In this study, we focus mainly on the second point by selecting more robust loss functions with (additionally) integrated class weights.

3.3 Simulating weakly supervised images

For the weakly supervised approach, we label the same images as before by simulating a starting region for each corroded segment. These regions are subsequently extended using the Region Growing algorithm. The quality of this process is evaluated using the original label images serving as reference. This procedure was chosen primarily because of the labor-intensive and time-consuming nature of manual labeling, which is especially true for the here investigated segmentation task. By using a brush annotation tool to mark only a small fraction or region, rather than labeling the entire segment pixel-accurately with a polygon, the process can be significantly sped up. Based on our own experience with the port dataset, labeling a single ≥ 16 megapixel image can take up to 30 minutes, even one with only a few damaged areas requires between five to ten minutes. Assuming that the suggested brush method reduces the manual effort by only 50 %, it can still result in time savings of 1-2 working days for the small image dataset we considered. This scales proportionally to the true labeling duration, which may vary between individual samples due to the number of damaged segments, their variable sizes and shapes. The subsequent Region Growing processing can be conducted after working hours to minimize utilization of the available computing capacity, or directly after labeling a single image to reduce the total processing time. On the other hand, this approach can also lead to a higher number of image and ground truth pairs in the same time period, resulting in a more robust trained segmentation network. However, a major disadvantage of the chosen algorithm in particular is that it is not suitable for all kinds of damage types or every segmentation class in general. For Region Growing to approximate fully supervised labels qualitatively, the true damage segments must have homogeneous colours and must clearly distinguish themselves from the background and surrounding classes. While this may be true on average for corrosion and efflorescence in this domain, it does not to concrete spalling, for example.

The version of Region Growing used here is based on standardizing each colour channel c independently to determine whether an adjacent pixel should be added to the growing start region.

A pixel x_i is added if the maximum standardized absolute deviation from the expected colour is within a given threshold t . Equation 1 defines the deviation:

$$\Delta x_{max} = \max \left(\left| \frac{x_i - \mu_c}{\sigma_c} \right| \right) \leq t \quad (1)$$

Here, Δx_{max} represents the deviation, μ_c and σ_c contain the channel-wise statistical quantities, and the \max function selects the largest element of the standardized RGB vector $\in \mathbb{R}^{\mathcal{F}}$. The threshold is set to a strict value of 1.25 to prevent overgrowing damage segments, especially for this pile quay wall and to provide the segmentation network with mostly correctly classified pixels, even if they represent only a small fraction of the true damaged areas. We believe that it is better to add initial missing corrosion pixels along the way than to include a large number of incorrect background pixels in the weakly supervised damage segments. To counteract wrongly generated corrosion segments during training, we use Conditional Random Field (CRF) implementation¹ by (Krähenbühl and Koltun, 2012) in each iteration. The threshold parameter is not specifically tuned to optimize certain metrics based on the original labeled data but is selected by visually comparing different outcomes for a small set of images. At the end of the algorithm, we apply equation 1 once again to every pixel within the grown region to possibly remove outliers. The starting segments for Region Growing are created by a randomly orientated ellipse that satisfies a certain overlapping condition with the reference label. The initial set semi-axes are dependent on the segment's size and are shrunken every time to the condition is not fulfilled. Using this method, the weak labels were evaluated in order to gain insight into the model's performance in the iterative training process. The initial weak labels achieved IoU scores of 92.4 and 36.2 % for the background/undamaged and corrosion class, respectively, with respect to the manual images. In Figure 1, we present the original RGB and both labeled ground truth images for a visually appealing sample of the test dataset.

4. SEGMENTATION NETWORKS

4.1 Baseline model

The Deeplab V3+ architecture is widely recognized for its impressive segmentation capabilities, and its basic modules, such as spatial pyramid pooling (SPP) and encoder-decoder structures, which are broadly applied even outside the core AI research community. The network captures object boundaries to some extent by gradually upsampling concatenated low- and high-level feature maps, and detects variable-sized objects with ease thanks to the Atrous SPP (ASPP) module, which encodes multi-scale contextual information (Chen et al., 2018). Generally, object boundaries can be refined using post-processing steps, but in this work, we orient ourselves on feature attention mechanisms ((Azad et al., 2020), (Hsu et al., 2022), (Zeng et al., 2020)) within the encoder part to achieve this. Such mechanisms recalibrate the functional relationship represented by the model to rely more on meaningful feature extractions. We aim to mimic this behaviour by modifying the dilation rates within ASPP's convolutions. By backpropagating the prediction error through the network during training, the feature maps or responsible filter blocks should be modified accordingly to boost detection capabilities of certain spatial resolutions slightly. This

¹ <https://github.com/lucasb-eyer/pydensecrf>



Figure 1. Example of a cropped RGB image showing the east-facing sheet pile wall of the JadeWeserPort and corresponding weakly and manually supervised ground truth images (top to bottom).

approach is expected to be beneficial due to the limited variability of occurring damaged areas in the particular port images we use, and for further related datasets that will be captured by similar systems to (Hesse et al., 2019). In other words, the damaged areas on quay walls or buildings can be closely sensed with high-resolution, allowing the ASPP module, regardless of the damage size, to segment the border area of different classes with high quality by taking optimized local context into account.

To improve segmentation results, we investigate different dilation rates of the corresponding Atrous convolution operation. The combined pyramid pooled features regain spatial context when merged with low-level features from the feature extraction network (alias backbone), primarily consisting of object boundaries. In the decoder, two 4x upsampling operations reconstruct the original image size, and contiguous segments are produced. To prevent information loss in the network structure, we apply some convolutions to combat the simplicity of linear interpolation. We examine whether adding additional blocks (0, 1 and 2) in between the Upsampling layers can significantly improve the segmentation result. For the dilation rates we consider the single to triple amount of 3, 6 (default), 12 and 24 for the Atrous convolutions. These hyperparameters are tested with three different models: VGG-16 (Simonyan and Zisserman, 2014) (Batch normalizations included), ResNet-50 (He et al., 2015) (version 1.5), and EfficientNet V2 (Tan and Le, 2021) (Small). VGG-16 has the most basic structure due to its early development in 2014 and consists of around 16 million of the lowest number of trainable parameters. The other two models are built with more advanced concepts in mind, such as Residual blocks, Neural Architecture Search optimized model structure, and Stochastic Depth, but can still be considered as rather basic or small DL models, with approximately 23 and 20 million parameters, respectively. We individually modify the models such that the output stride is set to 16 by removing the fifth

MaxPooling layer or setting the step size of the last strided convolution(s) to one. Furthermore, we freeze the initial layers up to Deeplab's low-level feature output to preserve the extensively learned parameters from the ImageNet database (used weight version 1 according to the torchvision (Marcel and Rodriguez, 2010) documentation). While the purpose of this work is to assist inspection operators, fast inference times are not necessarily required. However, to prevent overfitting or long training duration for new acquired datasets in the future, we have not examined significantly larger or more complex backbones.

To train our models, we initially divide the image-label pairs in (non-)overlapping patches of 512x512 pixels. This approach allowed the models to be fed with rich semantic context that exceeded their output layer's receptive field, enabling them to learn from different perspectives of cut-off corrosion segments located near the edges. We use a 25 % overlap for the train dataset but nothing is utilized otherwise. The available 24 GB VRAM of the GeForce RTX 3090 and later 4090 is fully utilized by setting the batch size to 8 (for the EfficientNet) or 12. For an efficient training advancement, the initial learning rate is determined to be $5e-4$ and $1e-3$, respectively, by the most negative slope of a learning rate finder curve.

To counteract the imbalance between undamaged and corroded areas, we use image augmentation methods such as affine transformation, colour jitter and Gaussian blur, with individual associated probabilities and class weights (1:1 for baseline determination to emphasize architecture changes effectively) within the different considered loss functions. To optimize the network architecture robustly, we employ k-fold Cross-Validation with the four aforementioned training subsets extensively. Each combination of dilation rates and number of additional convolutions is trained on three of these subsets and subsequently evaluated on the validation set to determine the optimal model parameter state based on the loss value. Through this procedure, the maximum number of epochs has never exceeded 30 for all types of model trainings.

In our experiments, we employ the combination of distribution- and region-based loss terms in the form of Categorical Cross-Entropy and IoU loss. To identify the optimal model structure, we calculate the total uncertainty (entropy of softmax output), accuracy, F1 and IoU scores based on the validation patches. The average uncertainty of this subset ranges from 0.06 to 0.12, while the total accuracy varies only between 95.2 and 96.1 %. The class-specific F1 and IoU scores reveal the details in the detection capabilities on the other hand. Overall, undamaged areas are recognized with more than 97 and 94 %, respectively, with deviations of around 1 % across all experiments. The most significant performance differences occur in the corrosion class. The F1 and IoU scores for this class vary by 8 % between the best and worst performing Deeplab network, with upper bound of 59 and 42 %, respectively. By counting the best performing models for both parameter sets individually, we determined that no additional convolution leads to a performance boost, and the standard dilation rates of 3 and 12 are almost equally useful for the segmentation task. Therefore, we changed the kernel spaces within the Atrous convolutions to 5, 10 and 20 to combine these ASPP blocks roughly. Thereby, we focus primarily on a local field of view in order to detect segment boundaries accurately, but find more deviating and rich features at the same time due to the variable dilation rates. To further improve the baseline model, we consider a wide range of modified loss functions with class weights, and evaluate all three backbone networks.

4.2 Training process of (weakly) supervised models

In the first stage of finetuning, we trained the modified Deeplab V3+ architecture with several pretrained backbones and overall five different loss terms on the complete training dataset. We evaluated them equally to the previous step to determine the optimal loss function for each feature extraction network. After that, we adjusted the hyperparameters related to the cost function to improve the model performance. We then used the separate testing set to independently assess the best performing Deeplab V3+ models among the feature extraction networks (s. chapter 5). The top architecture-loss pairs are then used to train the Deeplab model on the weakly supervised dataset in an iterative process, whereby the training and validation sets are created by the model predictions post-processed with a Potts model. We changed the standard deviations regarding pixel distance and RGB difference slightly compared to the default values listed at Pydense package's Github page (s. footnote 1) to account for the properties of our dataset. In particular, we set s_{xy} and s_{rgb} of the bilateral term to 40 and 10, respectively. This enables predicted corrosion segments to be extended more strictly based on the colour difference to nearby pixels, which is necessary for the corrosion segments onto the specific pile quay wall. We stopped when the IoU scores of the validation set do not improve compared to the last iteration. The initially trained models change the datasets extensively, leading to little overlapping segments. For instance, when the complete weakly supervised dataset is deployed, the first model achieves around 30 % IoU for the corrosion class. Subsequently, the other models benefit from more coherent samples post-processed by a CRF approach, gradually refining the damage segments. This entire procedure restricts the final model from learning self-produced label noise. We provided a detailed visualization of this segmentation approach with respect to the validation and test set in subsection 5.2.

All the models are in general trained using Stochastic Gradient Descent with Nesterov Momentum and L2 regularization, with associated parameters μ and λ set to 0.9 and $1e-5$, respectively, in the Pytorch framework (Paszke et al., 2019). The learning rate is halved when the validation loss does not decrease for three epochs in a row, and Batch Normalization layer momentum starts at 0.2 and gradually decreases to help the non-frozen network section adapt quickly to the task. The explored loss functions include Categorical Cross-Entropy, IoU loss, their combination (compound loss), and the base variants modified by the Focal term or parameter $\gamma = 2$ that shifts the model's attention to hard-to-classify pixels (Lin et al., 2017). Additionally, class weights of 1:3 were added to each error term. We observed significant performance improvements with the Focal IoU loss networks. Further Deeplab models were examined by incrementally increasing the parameter up to 5 for each backbone individually. The aforementioned weakly supervised model was finally trained with ResNet-50 and Focal IoU loss with $\gamma = 4$, and the impact of uncertain ground truth images was investigated by varying their fraction within the dataset between 60 to 100 %. Manually labeled images remained unchanged during each of these training runs to simulate real conditions.

5. EVALUATION AND DISCUSSION

5.1 Supervised models

The performance of the baseline Deeplab V3+ model is evaluated on the dedicated validation dataset after training it with

moderate basic backbones on various distribution- and region-based loss functions. Each image patch, which lacks surrounding context, is processed by the individual segmentation networks and analyzed quantitatively using metrics related to total uncertainty and confusion matrix. In this supervised context, it's evident that VGG-16 performs significantly worse than the other feature extraction networks on average due to the lower output mapping capabilities. However, the default Focal IoU loss variant ($\gamma = 2$) is marginally ahead of the corresponding EfficientNet V2 network. Additionally, we observe that region-based exclusive cost functions exhibit up to one magnitude superior uncertainty scores across all models, indicating their superiority for segmentation tasks with imbalanced datasets. We report only the corrosion class-related F1 and IoU scores in detail in this step of model optimization, as the background detection rates are comparable to those in subsection 4.1 with differences of around 3 %. Given the imbalance of this particular dataset, the total accuracy values can be neglected completely. Overall, the remaining quantities for the different Deeplab models are visualized in Figure 2.



Figure 2. Total uncertainties and corrosion class related metrics of Deeplab V3+ models with respect to validation set consisting of different backbones and trained on various loss functions. The hatched bar sections indicate the difference between IoU and F1 score.

Due to the excellent predictive performance achieved using the Focal IoU loss, we conducted further experiments to optimize the γ parameter for each backbone network separately. The best-performing parameter value was then applied for the first time on the testing set. To obtain complete images instead of smaller patches for visualization or later applications, we used a blending method described in (Chevallier, 2017). This method constructs smooth and continuous segments across patch edges by means of a window function that combines several orientated model predictions. These original sized images were then analyzed using a metric designed to assess agreement with complete corroded segments in the ground truth data. This is particularly helpful because inspection operators are primarily interested in the presence of damaged areas at a certain location, rather than their exact shape and/or size. The resulting scores originate from False Positive and Negative corrosion pixels, and can therefore be understood as supportive for the detection task. To make it more comprehensible: the remaining percentages to 100 represent fully made-up or overlooked damage segments. These metrics are presented in Table 1, alongside common quality measures. In total, we trained five different segmentation network instances by randomly reinitializing the parameters (except the pretrained backbone) to determine the

aleatoric and epistemic uncertainties individually. The total uncertainty is derived by the entropy of the composed Deep Ensemble, while epistemic uncertainty is derived by mutual information, and aleatoric uncertainty is equivalent to their difference. Therefore, the computed result(s) may deviate slightly from the results the single model shown in Figure 2.

Table 1. Results of best Deeplab V3+ segmentation network with respect to test set

Metric	Score
Total uncertainty (aleatoric, epistemic)	0.03492 (0.03491, 1e-5)
Total accuracy [%]	95.7
F1 Score [%] (backg., corr.)	97.7, 71.5
Pixel-wise IoU [%] (backg., corr.)	95.5, 55.6
Corrosion supporting FP and FN [%]	80.6, 86.3

The determined uncertainties are in-line with the other Deep Ensembles in terms of their relative relation. On average, the individual Deeplab models among the frameworks vary by 1e-6 to 1e-5, while the images themselves are responsible for almost the complete amount. This implies that the pretrained backbones, combined with the available but general small dataset, do not hinder the segmentation networks from learning corrosion damage appropriately, as aleatoric uncertainty remains unaffected by more training images. However, this information does not provide a detailed overview of the corrosion predictions in the images themselves. Looking at the other scores, we see that the non-damaged or background areas are classified with an accuracy of 95 % or more. The model struggles with precisely identifying the corrosion areas, but the value is as good as (Katsamenis et al., 2020) found on their available dataset. Due to the domain difference, we conclude that the optimization steps were necessary to achieve this level of performance with the restricted network architecture and overall available resources.

To get insight into the prediction results and their uncertainties in relation to the manually labeled images, we provide colour-coded corrosion segments and greyscale images below for the same scene as shown in Figure 1. In the top image, green, blue and red colors represent True Positive, False Positive and False Negative detections, respectively, while the brightness level indicates regions with higher uncertainty in the bottom images. The individual greyscale values are computed by converting an image-level scaled float value to an 8-bit unsigned integer, ensuring the full spectrum is always occupied despite the significant differences indicated by the tabular values. This example confirms that the trained models usually differ within a narrow buffer around the manually labeled damage segments.

The trained model often misses or wrongly detects corrosion near the areas that we identified as damaged with high confidence manually. Our integrated metrics reflect this by showing values above 80 %. Inspection operators dealing with manual classification on a daily basis can benefit from this information, despite other metrics indicating lower performance. If we also provide them additionally for example with the uncertainty images, they can adjust the segment boundaries using a prospect software package as part of a digitally assisted procedure.

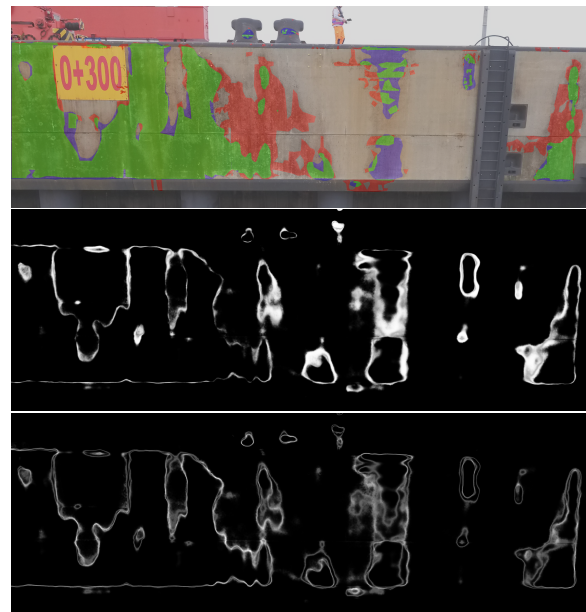


Figure 3. ResNet-50 Deeplab V3+ model's cropped blending prediction with post-processed colour-coding according to corrosion segments from ground truth data (green: True Positive, blue: False Positive, red: False Negative). Aleatoric and epistemic uncertainty calculated from a related five instance Deep Ensemble network.

5.2 Weakly Supervised model

The best-performing segmentation network, Deeplab V3+ ResNet-50, in the supervision context is then demonstrated in our simple designed training pipeline. This pipeline starts with ground truth images generated from a strict Region Growing algorithm applied to the challenging JadeWeserPort dataset. Due to the lack of accurate label images over the complete iterative training process, performance degradation is inevitable. However, the time-consuming labeling effort is significantly reduced. This gives researchers the possibility to obtain a larger dataset, if necessary, in order to improve model's generalization. In the following Figure 4, we show the performance development for different fractions of weakly label images used within the training process.

We draw here for the first time conclusions based on the agreement of blended patch predictions to the test set regarding the individual model training procedures. Especially the initial training phase can be highlighted, because the corresponding Region Growing labeled images achieved only 24.3 and 39.1 % in terms of corrosion IoU and F1 score, respectively. The final model of each training pipeline was determined by the highest IoU damage value of the individual evolved validation set, requiring 3, 2, 3, 3 and 2 iterations in descending order of the weakly label fraction. According to these results, the models iteratively trained on the complete weakly supervised dataset were the most successful, while the optimal Deeplab networks learnt from 70-90 % weak labels had a 5-10 % lower performance in terms of corrosion IoU. The dataset with the largest fraction of manual labels performed worst, with an additional 10 % drop in performance. This outcome was expected, as the strategy of finding the best model relies on monotonically increasing metric scores, which is equivalent to efficient conver-

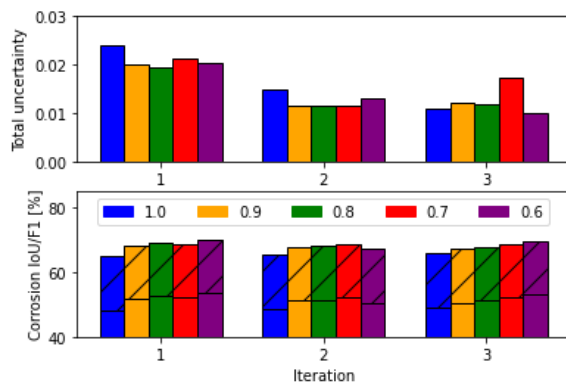


Figure 4. Total uncertainties and corrosion class related metrics of optimal Deeplab V3+ ResNet-50 with respect to test set trained on weakly supervised dataset. Different colours stand for the fraction of deployed weak labels. The hatched bar sections indicate the difference between IoU and F1 score.

gence. When an individual dataset contains inaccurate but obvious corrosion labels, the model can learn the underlying complexity more effectively. However, when a significant amount of the dataset consists of contradictory image pairs in terms of weak labels, there is no fast and straightforward solution. But the iteratively refined images may approximate the true damage segments step by step at different speeds. Hence, we can achieve supervised-like performance after several model trainings when a reasonable amount of accurate labels is available. However, we cannot measure such an advancement in general due to a lack of reference data. Nevertheless, we found a significant improvement between different models (with weak supervised fractions between 0.7 and 1.0) in terms of the labeled test set after a few iterations, as demonstrated in Figure 4. But how much labeling effort can be saved to achieve that? If we continue to calculate with the 50 % example from subsection 3.3, we would reduce the labeling effort for the complete training and validation dataset (70 images in total) by roughly 12 hours. For every accurately produced segmentation ground truth image taking approximately 10 - 15 more minutes, the model gains about 0.15 % IoU corrosion performance within an initial linear growing phase. Afterwards, the training may take more iterations than 3 to convergence, resulting in unpredictable performance capabilities, followed by a period in which a single model recognize weak labels confidently as outliers and take this into account for its parameter optimization automatically. Additionally, the determined total uncertainty over the test data is barely affected for the final models by the relative portion of manual labeled images. Therefore, we suggest that one should only label up to one third of small, imbalanced image datasets precisely to deploy the corresponding model in a short period of time. This will have a noticeable effect within the process chain and guarantees, based on our experiments, at least 90 % of the maximum achievable performance. This gap might be further closed by applying certain image augmentation methods and/or domain-specific loss functions.

In this paragraph, we present a detailed analysis of the performance of our best model on the test set, which provides insights into the potential of our method for real-world damage detection applications. We present the results using the same format as in the supervised setting, with Table 2 and Figure 5. Although most of the quantities in Deeplab model have

Table 2. Results of Deeplab V3+ segmentation network trained on 70 % weakly supervised images with respect to test set

Metric	Score
Total uncertainty (aleatoric, epistemic)	0.02248 (0.02248, 2e-6)
Total accuracy [%]	95.2
F1 Score [%] (backg., corr.)	97.4, 68.5
Pixel-wise IoU [%] (backg., corr.)	94.9, 52.1
Corrosion supporting FP and FN [%]	81.6, 75.0

only slightly changed compared to the fully supervised model, it tends to miss corrosion segments more often, resulting in a higher number of False Negative pixels (75.0 compared to 86.3 %). Conversely, the number of type 1 error for damaged segments decreases. These factors lead to relatively minor changes in the metric scores, but has a more visible impact on the prediction images themselves. Figure 5 illustrates the behaviour explicitly for the centrally located corrosion segments. In contrast, we observe finer, uncertain boundaries around the predicted corrosion segments, especially in aleatoric case, which is reflected in the lower scores in Table 2 compared to the fully supervised results. This difference can be explained by the iteratively trained models that learn from individual refined label images and build up confidence over time.

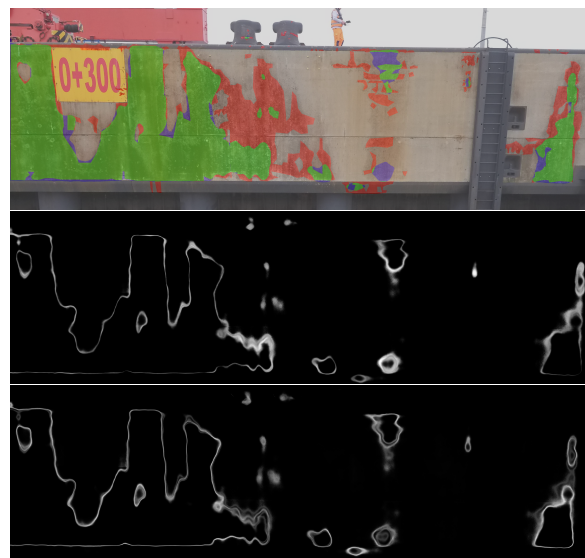


Figure 5. Cropped blending prediction by best ResNet-50 Deeplab V3+ model trained on 70 % weakly supervised images.

Colour-coded in post-processing according to corrosion segments from ground truth data (green: True Positive, blue: False Positive, red: False Negative). Aleatoric and epistemic uncertainty calculated from an affiliated five instance Deep Ensemble network.

6. CONCLUSION

In this study, we developed a modified Deeplab V3+ model for segmentation of corrosion in infrastructure inspections. The model was optimized using 4-fold Cross-Validation on a corrosion dataset, and we found that the combination of ResNet-50 backbone and Focal IoU loss with $\gamma = 4$ (with integrated

class weights of 1:3) achieved optimal results for our imbalanced dataset. We restricted our optimization to moderate large feature extraction networks for efficient training on consumer-level hardware. Our DL model is designed to assist operators visually in future inspections, and we plan to retrain it on further datasets that contain also other types of damage. We also investigated an iterative weakly supervised approach based on Region Growing, which can reduce the manual effort required for ground truth labeling. We found that roughly one third of manual labels are sufficient to achieve around 90 % of the fully supervised detection capabilities. The produced Deeplab V3+ models are part of the automation process of infrastructure inspections to reduce on-site downtime and subjective assessments (for example under stress). In future research, we plan to evaluate the model on a wide range of infrastructure datasets and refine the weakly supervised approach possibly with existing concepts to train a more robust segmentation network. We also plan to provide possibly model predicted uncertainty maps to inspection operators for a justified decision.

Funding This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - NE 1453/5-1 and as part of the Research Training Group i.c.sens [RTG 2159].

REFERENCES

- Adams, R., Bischof, L., 1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641-647.
- Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S., 2020. Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation. *ECCV 2020 Workshop on BioImage Computing*.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation.
- Chevallier, G., 2017. Using a u-net for image segmentation, blending predicted patches smoothly is a must to please the human eye. <https://github.com/-Vooban/Smoothly-Blend-Image-Patches>.
- Duy, L. D., Anh, N. T., Son, N. T., Tung, N. V., Duong, N. B., Khan, M. H. R., 2020. Deep learning in semantic segmentation of rust in images. *Proceedings of the 2020 9th International Conference on Software and Computer Applications, ICSCA 2020*, Association for Computing Machinery, New York, NY, USA, 129-132.
- Hake, F., Lippmann, P., Alkhatib, H., Oettel, V., Neumann, I., 2023a. Automated damage detection for port structures using machine learning algorithms in heightfields.
- Hake, F., Scherff, M., Neumann, I., Alkhatib, H., 2023b. Using semantic segmentation for the damage detection of port and marine infrastructures. Herbert Wichmann Verlag, accepted.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition.
- Hesse, C., Holste, K., Neumann, I., Hake, F., Alkhatib, H., Geist, M., Knaack, L., Scharr, C., 2019. 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadenserkennung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. *Hydrographische Nachrichten*, 113, 26-29.
- Hsu, C.-Y., Hu, R., Xiang, Y., Long, X., Li, Z., 2022. Improving the Deeplabv3+ Model with Attention Mechanisms Applied to Eye Detection and Segmentation. *Mathematics*, 10(15). <https://www.mdpi.com/2227-7390/10/15/2597>.
- Huang, Z., Wang, X., Wang, J., Liu, W., Wang, J., 2018. Weakly-supervised semantic segmentation network with deep seeded region growing. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7014-7023.
- Katsamenis, I., Protopapadakis, E., Doulamis, A., Doulamis, N., Voulodimos, A., 2020. Pixel-level corrosion detection on metal constructions by fusion of deep learning semantic and contour segmentation.
- Krähenbühl, P., Koltun, V., 2012. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. <https://arxiv.org/abs/1210.5644>.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection.
- Marcel, S., Rodriguez, Y., 2010. Torchvision the machine-vision package of torch. *Proceedings of the 18th ACM International Conference on Multimedia*, 1485-1488.
- Nash, W. T., Powell, C. J., Drummond, T., Birbilis, N., 2019. Automated corrosion detection using crowd sourced training for deep learning.
- Nash, W., Zheng, L., Birbilis, N., 2022. Deep learning corrosion detection with confidence. *npj Materials Degradation*, 6.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition.
- Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., Wang, J., 2019. High-resolution representations for labeling pixels and regions.
- Supervisely contributors, 2023. Supervisely. <https://github.com/supervisely/supervisely>.
- Tan, M., Le, Q. V., 2021. EfficientNetV2: Smaller Models and Faster Training. <https://arxiv.org/abs/2104.00298>.
- Tanveer, M., Kim, B., Hong, J., Sim, S.-H., Cho, S., 2022. Comparative Study of Lightweight Deep Semantic Segmentation Models for Concrete Damage Detection. *Applied Sciences*, 12(24). <https://www.mdpi.com/2076-3417/12/24/12786>.
- Zeng, H., Peng, S., Li, D., 2020. Deeplabv3+ semantic segmentation model based on feature cross attention mechanism. *Journal of Physics: Conference Series*, 1678, 012106.

Literaturverzeichnis

- [Abdel-Qader et al. 2006] ABDEL-QADER, Ikhlas ; PASHAIE-RAD, Sara ; ABUDAYYEH, Osama ; YEHA, Sherif: PCA-Based algorithm for unsupervised bridge crack detection. In: *Advances in Engineering Software* 37 (2006), Nr. 12, S. 771–778. – URL <https://www.sciencedirect.com/science/article/pii/S0965997806000809>. – ISSN 0965-9978
- [Akula & Sardana 2019] AKULA, Aparna ; SARDANA, H K.: Deep CNN-based Feature Extractor for Target Recognition in Thermal Images. In: *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019, S. 2370–2375
- [Alamouri et al. 2024] ALAMOURI, Ahmed ; DE ARRIBA LÓPEZ, Vanessa ; ACHANCCARAY DIAZ, Pedro ; BACKHAUS, Jan ; GERKE, Markus: High-resolution data capture and interpretation in support of port infrastructure maintenance. In: *DGPF-Jahrestagung 2024* (2024), S. 269–278
- [Apostolopoulos & Tzani 2022] APOSTOLOPOULOS, Ioannis D. ; TZANI, Mpesiana A.: Industrial object and defect recognition utilizing multilevel feature extraction from industrial scenes with Deep Learning approach. In: *Journal of Ambient Intelligence and Humanized Computing* 14 (2022), Januar, Nr. 8, S. 10263–10276. – URL <http://dx.doi.org/10.1007/s12652-021-03688-7>. – ISSN 1868-5145
- [Barnett & Lewis 1994] BARNETT, Vic ; LEWIS, Toby: *Outliers in statistical data*. Wiley New York, 1994
- [BAW 2018] BAW, Bundesanstalt für Wasserbau: *Kompetenz für die Wasserstraßen – Heute und in Zukunft*. https://www.baw.de/content/publikationen/pressemitteilungen/0/00221_2015.pdf. 2018. – letzter Zugriff am 17.05.2024
- [BAW 2019] BAW, Bundesanstalt für Wasserbau: *Kompetenzzentrum für den Verkehrswasserbau BAW-Strategie 2030*. https://www.baw.de/content/publikationen/www-dokumente-oeffentlich/0/BAW_Kompetenzzentrum_Verkehrswasserbau_Web.pdf. 2019. – letzter Zugriff am 17.05.2024
- [Beckmann et al. 2019] BECKMANN, Gustavo H. ; POLYZOIS, Dimos ; CHA, Young-Jin: Deep learning-based automatic volumetric damage quantification using depth camera. In: *Automation in Construction* 99 (2019), S. 114–124. – URL <https://www.sciencedirect.com/science/article/pii/S0926580518308690>. – ISSN 0926-5805
- [Bernardo & Smith 2009] BERNARDO, José M ; SMITH, Adrian F.: *Bayesian theory*. Bd. 405. John Wiley & Sons, 2009
- [Beshr et al. 2022] BESHAR, Ashraf A. ; HENEASH, Osama G. ; FAWZY, Hossam El-Din ; EL-BANNA, Mona M.: Condition assessment of rigid pavement using terrestrial laser scanner observations. In: *International Journal of Pavement Engineering* 23 (2022), Nr. 12, S. 4248–4259. – URL <https://doi.org/10.1080/10298436.2021.1940180>
- [Besl & McKay 1992] BESL, P.J. ; MCKAY, Neil D.: A method for registration of 3-D shapes. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992), Nr. 2, S. 239–256
- [Bishop 2006] BISHOP, Christopher M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg : Springer-Verlag, 2006. – ISBN 0387310738

- [Blundell et al. 2015] BLUNDELL, Charles ; CORNEBISE, Julien ; KAVUKCUOGLU, Koray ; WIERSTRA, Daan: Weight Uncertainty in Neural Network. In: BACH, Francis (Hrsg.) ; BLEI, David (Hrsg.): *Proceedings of the 32nd International Conference on Machine Learning* Bd. 37. Lille, France : PMLR, 07–09 Jul 2015, S. 1613–1622. – URL <https://proceedings.mlr.press/v37/blundell115.html>
- [BMDV 2014] BMDV, Bundesministerium für Digitales und Verkehr: *Forschungsbericht FE-Nr. 96.980-2011; Verkehrsverflechtungsprognose 2030 sowie Netzumlegung auf die Verkehrsträger Los 2 (Seeverkehrsprognose)*. https://bmdv.bund.de/SharedDocs/DE/Anlage/G/verkehrsverflechtungsprognose-2030-seeverkehr-schlussbericht-los-2.pdf?__blob=publicationFile. 2014. – letzter Zugriff am 13.06.2023
- [BMDV 2016] BMDV, Bundesministerium für Digitales und Verkehr: *Bundesverkehrswegeplan 2030*. https://bmdv.bund.de/SharedDocs/DE/Publikationen/G/bundesverkehrswegeplan-2030-gesamtplan.pdf?__blob=publicationFile. 2016. – letzter Zugriff am 17.05.2024
- [BMDV 2022a] BMDV, Bundesministerium für Digitales und Verkehr: *Implementierungsstrategie BIM-WSV 2030*. https://bmdv.bund.de/SharedDocs/DE/Anlage/DG/implementierungsstrategie-bim-wsv-2030.pdf?__blob=publicationFile. 2022. – letzter Zugriff am 02.06.2024
- [BMDV 2022b] BMDV, Bundesministerium für Digitales und Verkehr: *Zukunftspaket leistungsfähige Autobahnbrücken*. https://bmdv.bund.de/SharedDocs/DE/Anlage/K/presse/zukunftspaket-leistungsfaeheige-autobahnbrueecken.pdf?__blob=publicationFile. 2022. – letzter Zugriff am 17.05.2024
- [BMI 2021] BMI, Bundesministerium des Innern, für Bau und Heimat & BMVG, Bundesministerium der Verteidigung: *Masterplan BIM für Bundesbauten*. https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2021/10/masterplan-bim.pdf;jsessionid=BA100CF59399F0BC4821BB09B107503E.1_cid364?__blob=publicationFile. 2021. – letzter Zugriff am 02.06.2024
- [Breunig et al. 2000] BREUNIG, Markus M. ; KRIEGEL, Hans-Peter ; NG, Raymond T. ; SANDER, Jörg: LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, S. 93–104
- [Brodu & Lague 2012] BRODU, N. ; LAGUE, D.: 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 68 (2012), S. 121–134. – URL <https://www.sciencedirect.com/science/article/pii/S0924271612000330>. – ISSN 0924-2716
- [Bureick et al. 2016] BUREICK, Johannes ; ALKHATIB, Hamza ; NEUMANN, Ingo: Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis. In: *Journal of Applied Geodesy* 10 (2016), Nr. 1, S. 27–35. – URL <https://doi.org/10.1515/jag-2015-0020>
- [Chandola et al. 2009] CHANDOLA, Varun ; BANERJEE, Arindam ; KUMAR, Vipin: Anomaly detection: A survey. In: *ACM Comput. Surv.* 41 (2009), jul, Nr. 3. – URL <https://doi.org/10.1145/1541880.1541882>. – ISSN 0360-0300
- [Chen et al. 2018] CHEN, Liang-Chieh ; ZHU, Yukun ; PAPANDREOU, George ; SCHROFF, Florian ; ADAM, Hartwig: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018, S. 801–818

- [Chen et al. 2012] CHEN, Po-Han ; SHEN, Heng-Kuang ; LEI, Chi-Yang ; CHANG, Luh-Maan: Support-vector-machine-based method for automated steel bridge rust assessment. In: *Automation in Construction* 23 (2012), S. 9–19. – URL <https://www.sciencedirect.com/science/article/pii/S0926580511002263>. – ISSN 0926-5805
- [Chen et al. 2013] CHEN, Shen-En ; LIU, Wanqiu ; BIAN, Haitao ; SMITH, Ben: 3D LiDAR scans for bridge damage evaluations. In: *Forensic Engineering 2012: Gateway to a Safer Tomorrow*. 2013, S. 487–495
- [Chen et al. 2021] CHEN, Xiaolong ; LI, Jian ; HUANG, Shuowen ; CUI, Hao ; LIU, Peirong ; SUN, Quan: An Automatic Concrete Crack-Detection Method Fusing Point Clouds and Images Based on Improved Otsu’s Algorithm. In: *Sensors* 21 (2021), Nr. 5. – URL <https://www.mdpi.com/1424-8220/21/5/1581>. – ISSN 1424-8220
- [Choi & Kim 2005] CHOI, K.Y. ; KIM, S.S.: Morphological analysis and classification of types of surface corrosion damage by digital image processing. In: *Corrosion Science* 47 (2005), Nr. 1, S. 1–15. – ISSN 0010-938X
- [Chollet 2017] CHOLLET, Francois: Xception: Deep Learning With Depthwise Separable Convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, S. 1251–1258
- [Deng et al. 2009] DENG, Jia ; DONG, Wei ; SOCHER, Richard ; LI, Li-Jia ; LI, Kai ; FEI-FEI, Li: Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition Ieee (Veranst.)*, 2009, S. 248–255
- [Destatis 2021] DESTATIS: *Pressemitteilung Nr. 127 vom 16. März 2021, Statistisches Bundesamt (Destatis)*. https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/03/PD21_127_463.html. 2021. – letzter Zugriff am 16.09.2021
- [Diaz et al. 2017] DIAZ, Julianne Alyson I. ; LIGERALDE, Manuel I. ; JOSE, John Anthony C. ; BANDALA, Argel A.: Rust detection using image processing via Matlab. In: *TENCON 2017 - 2017 IEEE Region 10 Conference*, Nov 2017, S. 1327–1331. – ISSN 2159-3450
- [Dreßler 2006] DRESSLER, Matthias: *Datenreduktion von Laserscanneraufnahmen mittels automatisierter statistischer Approximationsverfahren*. 2006. – Diplomarbeit
- [Drixler 1993] DRIXLER, E: *Analyse der Form und Lage von Objekten im Raum*. München, DGK Reihe C, Heft Nr. 409, Dissertation, 1993
- [Dunteman 1989] DUNTEMAN, George H.: *Principal components analysis*. Bd. 69. Sage, 1989
- [Fischler & Bolles 1981] FISCHLER, Martin A. ; BOLLES, Robert C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Commun. ACM* 24 (1981), Juni, Nr. 6, S. 381–395. – URL <https://doi.org/10.1145/358669.358692>. – ISSN 0001-0782
- [Fort et al. 2020] FORT, Stanislav ; HU, Huiyi ; LAKSHMINARAYANAN, Balaji: *Deep Ensembles: A Loss Landscape Perspective*. 2020. – URL <https://arxiv.org/abs/1912.02757>
- [Gal 2016] GAL, Yarin: *Uncertainty in Deep Learning*, University of Cambridge, Dissertation, 2016
- [Gal & Ghahramani 2016] GAL, Yarin ; GHAHRAMANI, Zoubin: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: BALCAN, Maria F. (Hrsg.) ; WEINBERGER, Kilian Q. (Hrsg.): *Proceedings of The 33rd International Conference on Machine Learning* Bd. 48. New York, New York, USA : PMLR, 20–22 Jun 2016, S. 1050–1059. – URL <https://proceedings.mlr.press/v48/gal16.html>

- [Garcia et al. 2021] GARCIA, R. L. ; HAPP, P. N. ; FEITOSA, R. Q.: Large scale semantic segmentation of virtual environments to facilitate corrosion management. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2021* (2021), S. 465–470. – URL <https://isprs-archives.copernicus.org/articles/XLIII-B2-2021/465/2021/>
- [Gatys et al. 2015] GATYS, Leon ; ECKER, Alexander S. ; BETHGE, Matthias: Texture Synthesis Using Convolutional Neural Networks. In: CORTES, C. (Hrsg.) ; LAWRENCE, N. (Hrsg.) ; LEE, D. (Hrsg.) ; SUGIYAMA, M. (Hrsg.) ; GARNETT, R. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 28, Curran Associates, Inc., 2015, S. 262–270. – URL <https://proceedings.neurips.cc/paper/2015/file/a5e00132373a7031000fd987a3c9f87b-Paper.pdf>
- [Ghanta et al. 2011] GHANTA, Sindhu ; KARP, Tanja ; LEE, Sangwook: Wavelet domain detection of rust in steel bridge images. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, S. 1033–1036
- [Gopalakrishnan et al. 2017] GOPALAKRISHNAN, Kasthurirangan ; KHAITAN, Siddhartha K. ; CHOUDHARY, Alok ; AGRAWAL, Ankit: Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. In: *Construction and Building Materials* 157 (2017), S. 322–330. – URL <https://www.sciencedirect.com/science/article/pii/S0950061817319335>. – ISSN 0950-0618
- [Graves 2011] GRAVES, Alex: Practical Variational Inference for Neural Networks. In: SHAWE-TAYLOR, J. (Hrsg.) ; ZEMEL, R. (Hrsg.) ; BARTLETT, P. (Hrsg.) ; PEREIRA, F. (Hrsg.) ; WEINBERGER, K.Q. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 24, Curran Associates, Inc., 2011. – URL https://proceedings.neurips.cc/paper_files/paper/2011/file/7eb3c8be3d411e8ebfab08eba5f49632-Paper.pdf
- [Guldur Erkal & Hajjar 2020] GULDUR ERKAL, Burcu ; HAJJAR, Jerome F.: Using extracted member properties for laser-based surface damage detection and quantification. In: *Structural Control and Health Monitoring* 27 (2020), Nr. 11, S. e2616. – e2616 STC-19-0496.R2
- [Gustafsson et al. 2020] GUSTAFSSON, Fredrik K. ; DANELLJAN, Martin ; SCHON, Thomas B.: Evaluating Scalable Bayesian Deep Learning Methods for Robust Computer Vision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020
- [Haitz et al. 2022] HAITZ, Dennis ; HÜBNER, Patrick ; ULRICH, Markus ; LANDGRAF, Steven ; JUTZI, Boris: Semantic segmentation with small training datasets: a case study for corrosion detection on the surface of industrial objects. In: *Image Processing Forum*, 2022, S. 73–85
- [Hake et al. 2022] HAKE, Frederic ; GÖTTERT, Leonard ; NEUMANN, Ingo ; ALKHATIB, Hamza: Using Machine-Learning for the Damage Detection of Harbour Structures. In: *Remote sensing* 14 (2022), Mai, Nr. 11. – ISSN 2072-4292
- [Hake et al. 2020] HAKE, Frederic ; HERMANN, Matthias ; ALKHATIB, Hamza ; HESSE, Christian ; HOLSTE, Karsten ; UMLAUF, Georg ; KERMARREC, Gael ; NEUMANN, Ingo: Damage Detection for Port Infrastructure by Means of Machine-Learning-Algorithms, 2020. – FIG Working Week 2020
- [Hake et al. 2023a] HAKE, Frederic ; LIPPMANN, Paula ; ALKHATIB, Hamza ; OETTEL, Vincent ; NEUMANN, Ingo: Automated damage detection for port structures using machine learning algorithms in heightfields. In: *Applied Geomatics* 15 (2023), Juni, Nr. 2, S. 349–357

- [Hake et al. 2023b] HAKE, Frederic ; SCHERFF, Marvin ; NEUMANN, Ingo ; ALKHATIB, Hamza: Using Semantic Segmentation for the Damage Detection of Port and Marine Infrastructures. In: WIESER, Andreas (Hrsg.): *Ingenieurvermessung 23*. Germany : Wichmann, VDE, April 2023, S. 93 – 106. – 20. Internationaler Ingenieurvermessungskurs ; Conference date: 10-04-2023 Through 14-04-2023. – ISBN 978-3-87907-734-2
- [Hawkins 1980] HAWKINS, Douglas M.: *Identification of outliers*. Bd. 11. Springer, 1980
- [He et al. 2016] HE, Kaiming ; ZHANG, Xiangyu ; REN, Shaoqing ; SUN, Jian: Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, S. 770–778
- [Helfrick et al. 2009] HELFRICK, Mark N. ; NIEZRECKI, Christopher ; AVITABILE, Peter: Curvature methods of damage detection using digital image correlation. In: *Health Monitoring of Structural and Biological Systems 2009* Bd. 7295 SPIE (Veranst.), 2009, S. 130–141
- [Hesse et al. 2019] HESSE, Christian ; HOLSTE, Karten ; NEUMANN, Ingo ; HAKE, Frederic ; ALKHATIB, Hamza ; GEIST, Michael ; KNAACK, Lisa ; SCHARR, Christian: 3D HydroMapper: Automatisierte 3D-Bauwerksaufnahme und Schadenserkenkung unter Wasser für die Bauwerksinspektion und das Building Information Modelling. In: *Hydrographische Nachrichten* (2019), Nr. 113, S. 26–29
- [Huber & Ronchetti 2009] HUBER, Peter J. ; RONCHETTI, Elvezio: *Robust statistics*. 2009. – URL <https://www.tib.eu/de/suchen/id/TIBKAT%3A1616701536>
- [Jahanshahi & Masri 2013] JAHANSHAHI, Mohammad R. ; MASRI, Sami F.: Parametric Performance Evaluation of Wavelet-Based Corrosion Detection Algorithms for Condition Assessment of Civil Infrastructure Systems. In: *Journal of Computing in Civil Engineering* 27 (2013), Nr. 4, S. 345–357
- [Jiang & Zhang 2020] JIANG, Shang ; ZHANG, Jian: Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. In: *Computer-Aided Civil and Infrastructure Engineering* 35 (2020), Nr. 6, S. 549–564. – URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12519>
- [Jin Lim et al. 2021] JIN LIM, Hyung ; HWANG, Soonkyu ; KIM, Hyeonjin ; SOHN, Hoon: Steel bridge corrosion inspection with combined vision and thermographic images. In: *Structural Health Monitoring* 20 (2021), Nr. 6, S. 3424–3435. – URL <https://doi.org/10.1177/1475921721989407>
- [Kalenjuk et al. 2021] KALENJUK, Slaven ; LIENHART, Werner ; REBHAN, Matthias J.: Processing of mobile laser scanning data for large-scale deformation monitoring of anchored retaining structures along highways. In: *Computer-Aided Civil and Infrastructure Engineering* 36 (2021), Nr. 6, S. 678–694. – URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12656>
- [Kanani 2000] KANANI, Entela: *Robust estimators for geodetic transformations and GIS*, Eidgenössische Technische Hochschule Zürich, Dissertation, 2000
- [Khan et al. 2021] KHAN, Maleika H. ; HUSSEIN SK HEERAH, Mohammad Z. ; BASGEETH, Zuhairah: Automated Detection of Multi-class Vehicle Exterior Damages using Deep Learning. In: *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2021, S. 01–06
- [Khoreva et al. 2017] KHOREVA, Anna ; BENENSON, Rodrigo ; HOSANG, Jan ; HEIN, Matthias ; SCHIELE, Bernt: Simple Does It: Weakly Supervised Instance and Semantic Segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, S. 876–885

- [Kim et al. 2017] KIM, Hyunjun ; LEE, Junhwa ; AHN, Eunjong ; CHO, Soojin ; SHIN, Myoungsu ; SIM, Sung-Han: Concrete Crack Identification Using a UAV Incorporating Hybrid Image Processing. In: *Sensors* 17 (2017), Nr. 9. – URL <https://www.mdpi.com/1424-8220/17/9/2052>. – ISSN 1424-8220
- [Kim & Kim 2020] KIM, Hyunsoo ; KIM, Changwan: Deep-Learning-Based Classification of Point Clouds for Bridge Inspection. In: *Remote Sensing* 12 (2020), Nr. 22. – URL <https://www.mdpi.com/2072-4292/12/22/3757>. – ISSN 2072-4292
- [Kim et al. 2018] KIM, In-Ho ; JEON, Haemin ; BAEK, Seung-Chan ; HONG, Won-Hwa ; JUNG, Hyung-Jo: Application of Crack Identification Techniques for an Aging Concrete Bridge Inspection Using an Unmanned Aerial Vehicle. In: *Sensors* 18 (2018), Nr. 6. – URL <https://www.mdpi.com/1424-8220/18/6/1881>. – ISSN 1424-8220
- [Kim et al. 2015] KIM, Min-Koo ; SOHN, Hoon ; CHANG, Chih-Chen: Localization and Quantification of Concrete Spalling Defects Using Terrestrial Laser Scanning. In: *Journal of Computing in Civil Engineering* 29 (2015), Nr. 6, S. 04014086
- [Krähenbühl & Koltun 2011] KRÄHENBÜHL, Philipp ; KOLTUN, Vladlen: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In: SHAWE-TAYLOR, J. (Hrsg.) ; ZEMEL, R. (Hrsg.) ; BARTLETT, P. (Hrsg.) ; PEREIRA, F. (Hrsg.) ; WEINBERGER, K.Q. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 24, Curran Associates, Inc., 2011. – URL https://proceedings.neurips.cc/paper_files/paper/2011/file/beda24c1e1b46055dff2c39c98fd6fc1-Paper.pdf
- [Lakshminarayanan et al. 2017] LAKSHMINARAYANAN, Balaji ; PRITZEL, Alexander ; BLUNDELL, Charles: Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In: GUYON, I. (Hrsg.) ; LUXBURG, U. V. (Hrsg.) ; BENGIO, S. (Hrsg.) ; WALLACH, H. (Hrsg.) ; FERGUS, R. (Hrsg.) ; VISHWANATHAN, S. (Hrsg.) ; GARNETT, R. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 30, Curran Associates, Inc., 2017. – URL https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf
- [Lee et al. 1997] LEE, S. ; WOLBERG, G. ; SHIN, S.Y.: Scattered data interpolation with multilevel B-splines. In: *IEEE Transactions on Visualization and Computer Graphics* 3 (1997), Nr. 3, S. 228–244
- [Liao & Lee 2016] LIAO, Kuo-Wei ; LEE, Yi-Ting: Detection of rust defects on steel bridge coatings via digital image recognition. In: *Automation in Construction* 71 (2016), S. 294–306. – URL <https://www.sciencedirect.com/science/article/pii/S0926580516301625>. – ISSN 0926-5805
- [Lin et al. 2023] LIN, Siyu ; TANG, Fujian ; DANG, Ji ; LIN, Zhibin: Corrosion Inspection and Evaluation of Roadway Metal Railings with Deep Learning Technique. In: *AMPP CORROSION AMPP* (Veranst.), 2023, S. AMPP–2023
- [Lippmann 2021] LIPPMANN, Paula L.: *Punktwolken-Segmentierung und Anomalien-Detektion mittels Transferlearning und Local-Outlier-Factors*. Hannover, Leibniz Universität Hannover, Masterarbeit (unveröffentlicht). 2021
- [Liu et al. 2020] LIU, Jeremiah ; LIN, Zi ; PADHY, Shreyas ; TRAN, Dustin ; BEDRAX WEISS, Tania ; LAKSHMINARAYANAN, Balaji: Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness. In: LAROCHELLE, H. (Hrsg.) ; RANZATO, M. (Hrsg.) ; HADSELL, R. (Hrsg.) ; BALCAN, M.F. (Hrsg.) ; LIN, H. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 33, Curran Associates, Inc., 2020,

- S. 7498–7512. – URL https://proceedings.neurips.cc/paper_files/paper/2020/file/543e83748234f7cbab21aa0ade66565f-Paper.pdf
- [Liu et al. 2019] LIU, Tingting ; KANG, Kai ; ZHANG, Fen ; NI, Jialiang ; WANG, Tianyun: A corrosion detection algorithm via the random forest model. In: LI, Zhaohui (Hrsg.): *17th International Conference on Optical Communications and Networks (ICOCN2018)* Bd. 11048 International Society for Optics and Photonics (Veranst.), SPIE, 2019, S. 110480B. – URL <https://doi.org/10.1117/12.2518313>
- [Liu et al. 2002] LIU, Zhiwei ; SUANDI, Shahrel A. ; OHASHI, Takeshi ; EJIMA, Toshiaki: Tunnel crack detection and classification system based on image processing. In: *Machine vision applications in industrial inspection X* Bd. 4664 SPIE (Veranst.), 2002, S. 145–152
- [Loquercio et al. 2020] LOQUERCIO, Antonio ; SEGU, Mattia ; SCARAMUZZA, Davide: A General Framework for Uncertainty Estimation in Deep Learning. In: *IEEE Robotics and Automation Letters* 5 (2020), Nr. 2, S. 3153–3160
- [MacKay 1992] MACKAY, David J. C.: *Bayesian methods for adaptive models*, California Institute of Technology, Dissertation, 1992
- [Madrigal et al. 2017] MADRIGAL, Carlos A. ; BRANCH, John W. ; RESTREPO, Alejandro ; MERY, Domingo: A Method for Automatic Surface Inspection Using a Model-Based 3D Descriptor. In: *Sensors* 17 (2017), Nr. 10. – URL <https://www.mdpi.com/1424-8220/17/10/2262>. – ISSN 1424-8220
- [Matthies 2007] MATTHIES, Hermann G.: Quantifying uncertainty: modern computational representation of probability and applications. In: IBRAHIMBEGOVIC, Adnan (Hrsg.) ; KOZAR, Ivica (Hrsg.): *Extreme Man-Made and Natural Hazards in Dynamics of Structures*. Dordrecht : Springer Netherlands, 2007, S. 105–135. – ISBN 978-1-4020-5656-7
- [Meena et al. 2022] MEENA, Gaurav ; MOHBHEY, Krishna K. ; INDIAN, Ajay ; KUMAR, Sunil: Sentiment Analysis from Images using VGG19 based Transfer Learning Approach. In: *Procedia Computer Science* 204 (2022), S. 411–418. – URL <https://www.sciencedirect.com/science/article/pii/S1877050922007888>. – International Conference on Industry Sciences and Computer Science Innovation. – ISSN 1877-0509
- [Megahed et al. 2021] MEGAHEDE, Fadel M. ; CHEN, Ying-Ju ; MEGAHEDE, Aly ; ONG, Yuya ; ALTMAN, Naomi ; KRZYWINSKI, Martin: The class imbalance problem. In: *Nat Methods* 18 (2021), Nr. 11, S. 1270–7
- [Mohammadivojdan et al. 2020] MOHAMMADIVOJDAN, Bahareh ; ALKHATIB, Hamza ; JAHN, Cord-Hinrich ; BROCKMEYER, Marco ; NEUMANN, Ingo: *Surface Based Modeling of Ground Motion Areas in Lower Saxony*. S. 107–123. In: *GeoMonitoring 2020*, 2020
- [Montero et al. 2015] MONTERO, José-María ; FERNÁNDEZ-AVILÉS, Gema ; MATEU, Jorge: *Spatial and spatio-temporal geostatistical modeling and kriging*. Bd. 1. Chichester, UK : John Wiley & Sons, 2015
- [Morgenthal et al. 2019] MORGENTHAL, Guido ; HALLERMANN, Norman ; KERSTEN, Jens ; TARABEN, Jakob ; DEBUS, Paul ; HELMRICH, Marcel ; RODEHORST, Volker: Framework for automated UAS-based structural condition assessment of bridges. In: *Automation in Construction* 97 (2019), S. 77–95. – URL <https://www.sciencedirect.com/science/article/pii/S0926580518305156>. – ISSN 0926-5805
- [Mukhoti et al. 2023] MUKHOTI, Jishnu ; KIRSCH, Andreas ; AMERSFOORT, Joost van ; TORR, Philip H. ; GAL, Yarin: Deep Deterministic Uncertainty: A New Simple Baseline. In: *Proceedings*

- of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, S. 24384–24394
- [Nash et al. 2022] NASH, Will ; ZHENG, Liang ; BIRBILIS, Nick: Deep learning corrosion detection with confidence. In: *npj Materials degradation* 6 (2022), Nr. 1, S. 26
- [Neal 1995] NEAL, Radford M.: *Bayesian Learning for Neural Networks*, University of Toronto, Dissertation, 1995
- [Ovadia et al. 2019] OVADIA, Yaniv ; FERTIG, Emily ; REN, Jie ; NADO, Zachary ; SCULLEY, D. ; NOWOZIN, Sebastian ; DILLON, Joshua ; LAKSHMINARAYANAN, Balaji ; SNOEK, Jasper: Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift. In: WALLACH, H. (Hrsg.) ; LAROCHELLE, H. (Hrsg.) ; BEYGELZIMER, A. (Hrsg.) ; BUC, F. d’Alché (Hrsg.) ; FOX, E. (Hrsg.) ; GARNETT, R. (Hrsg.): *Advances in Neural Information Processing Systems* Bd. 32, Curran Associates, Inc., 2019. – URL https://proceedings.neurips.cc/paper_files/paper/2019/file/8558cb408c1d76621371888657d2eb1d-Paper.pdf
- [Pakrashi et al. 2010] PAKRASHI, Vikram ; SCHOEFS, Franck ; MEMET, Jean B. ; O’CONNOR, Alan: ROC dependent event isolation method for image processing based assessment of corroded harbour structures. In: *Structure and Infrastructure Engineering* 6 (2010), Nr. 3, S. 365–378
- [Pan & Zhang 2022] PAN, Y. ; ZHANG, L.: Dual attention deep learning network for automatic steel surface defect segmentation. In: *Computer-Aided Civil and Infrastructure Engineering* 37 (2022), Nr. 11, S. 1468–1487. – URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12792>
- [Pauls & Yoder 2018] PAULS, Alexander ; YODER, J: Determining optimum drop-out rate for neural networks. In: *Midwest Instructional Computing Symposium (MICS)*, 2018
- [Pedregosa et al. 2011] PEDREGOSA, F. ; VAROQUAUX, G. ; GRAMFORT, A. ; MICHEL, V. ; THIRION, B. ; GRISEL, O. ; BLONDEL, M. ; PRETTENHOFER, P. ; WEISS, R. ; DUBOURG, V. ; VANDERPLAS, J. ; PASSOS, A. ; COURNAPEAU, D. ; BRUCHER, M. ; PERROT, M. ; DUCHESNAY, E.: Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830
- [Piegl & Tiller 2012] PIEGL, Les ; TILLER, Wayne: *The NURBS book*. Berlin/Heidelberg : Springer Science & Business Media, 2012
- [Pont-Tuset et al. 2017] PONT-TUSET, Jordi ; ARBELÁEZ, Pablo ; T. BARRON, Jonathan ; MARQUES, Ferran ; MALIK, Jitendra: Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017), Nr. 1, S. 128–140
- [Prasanna et al. 2016] PRASANNA, Prateek ; DANA, Kristin J. ; GUCUNSKI, Nenad ; BASILY, Basily B. ; LA, Hung M. ; LIM, Ronny S. ; PARVARDEH, Hooman: Automated Crack Detection on Concrete Bridges. In: *IEEE Transactions on Automation Science and Engineering* 13 (2016), Nr. 2, S. 591–599
- [Rabah et al. 2013] RABAH, Mostafa ; ELHATTAB, Ahmed ; FAYAD, Atef: Automatic concrete cracks detection and mapping of terrestrial laser scan data. In: *NRIAG Journal of Astronomy and Geophysics* 2 (2013), Nr. 2, S. 250–255
- [Rasmussen 2004] RASMUSSEN, Carl E.: *Gaussian Processes in Machine Learning*. S. 63–71. In: BOUSQUET, Olivier (Hrsg.) ; LUXBURG, Ulrike von (Hrsg.) ; RÄTSCH, Gunnar (Hrsg.): *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*. Berlin, Heidelberg : Springer

- Berlin Heidelberg, 2004. – URL https://doi.org/10.1007/978-3-540-28650-9_4. – ISBN 978-3-540-28650-9
- [Riveiro et al. 2016] RIVEIRO, B. ; DEJONG, M.J. ; CONDE, B.: Automated processing of large point clouds for structural health monitoring of masonry arch bridges. In: *Automation in Construction* 72 (2016), S. 258–268. – ISSN 0926-5805
- [Scherff et al. 2023] SCHERFF, M. ; HAKE, F. ; ALKHATIB, H.: Adaptation of Deeplab V3+ For Damage Detection on Port Infrastructure Imagery. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-M-1-2023 (2023), S. 301–308. – URL <https://isprs-archives.copernicus.org/articles/XLVIII-M-1-2023/301/2023/>
- [Scherff 2022] SCHERFF, Marvin: *Semantische Segmentierung und Klassifizierung von Bauwerkschäden aus Bildern von Hafeninfrastrukturen*. Hannover, Leibniz Universität Hannover, Masterarbeit (unveröffentlicht). 2022
- [Schischmanow et al. 2022] SCHISCHMANOW, Adrian ; DAHLKE, Dennis ; BAUMBACH, Dirk ; ERNST, Ines ; LINKIEWICZ, Magdalena: Seamless Navigation, 3D Reconstruction, Thermographic and Semantic Mapping for Building Inspection. In: *Sensors* 22 (2022), Nr. 13. – URL <https://www.mdpi.com/1424-8220/22/13/4745>. – ISSN 1424-8220
- [Shorten & Khoshgoftaar 2019] SHORTEN, Connor ; KHOSHGOFTAAR, Taghi M.: A survey on Image Data Augmentation for Deep Learning. In: *Journal of Big Data* 6 (2019), Jul, Nr. 1, S. 60. – URL <https://doi.org/10.1186/s40537-019-0197-0>. – ISSN 2196-1115
- [Sibson 1981] SIBSON, Robin: A brief description of natural neighbour interpolation. In: *Interpreting multivariate data* (1981)
- [Simonyan & Zisserman 2014] SIMONYAN, Karen ; ZISSERMAN, Andrew: *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. – URL <http://arxiv.org/pdf/1409.1556v6>
- [Skidmore 1989] SKIDMORE, Andrew K.: A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. In: *International Journal of Geographical Information System* 3 (1989), Nr. 4, S. 323–334
- [Stegner 1984] STEGNER, Günther: *Helmert-Transformation*. S. 102–107. In: *Geodätische Berechnungsmethoden (Standard-BASIC): Dreiecke, Schnitte, Lagepunktbestimmung, Polygonzüge, Transformationen, Flächen, Freie Stationierung, Punktplott-Routine*. Wiesbaden : Vieweg+Teubner Verlag, 1984. – URL https://doi.org/10.1007/978-3-663-06859-4_20. – ISBN 978-3-663-06859-4
- [Straub 1996] STRAUB, Christian S.: *Recent crustal deformation and strain accumulation in the Marmara Sea region, NW Anatolia, inferred from GPS measurements*, Eidgenössische Technische Hochschule Zürich, Dissertation, 1996
- [Suchocki et al. 2021] SUCHOCKI, Czesław ; BŁASZCZAK-BAK, Wioleta ; JANICKA, Joanna ; DUMALSKI, Andrzej: Detection of defects in building walls using modified OptD method for down-sampling of point clouds. In: *Building Research & Information* 49 (2021), Nr. 2, S. 197–215. – URL <https://doi.org/10.1080/09613218.2020.1729687>
- [Supervisely contributors 2023] SUPERVISELY CONTRIBUTORS: *Supervisely*. <https://github.com/supervisely/supervisely>. 2023
- [Sánchez-Rodríguez et al. 2018] SÁNCHEZ-RODRÍGUEZ, A. ; RIVEIRO, B. ; CONDE, B. ; SOILÁN, M.: Detection of structural faults in piers of masonry arch bridges through automated processing

- of laser scanning data. In: *Structural Control and Health Monitoring* 25 (2018), Nr. 3, S. e2126. – e2126 STC-17-0082.R2
- [Teza et al. 2009] TEZA, Giordano ; GALGARO, Antonio ; MORO, Francesca: Contactless recognition of concrete surface damage from laser scanning and curvature computation. In: *NDT & E International* 42 (2009), Nr. 4, S. 240–249. – ISSN 0963-8695
- [Valdenegro-Toro 2023] VALDENEGRO-TORO, Matias: Sub-Ensembles for Fast Uncertainty Estimation in Neural Networks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2023, S. 4119–4127
- [Van Amersfoort et al. 2020] VAN AMERSFOORT, Joost ; SMITH, Lewis ; TEH, Yee W. ; GAL, Yarin: Uncertainty Estimation Using a Single Deep Deterministic Neural Network. In: III, Hal D. (Hrsg.) ; SINGH, Aarti (Hrsg.): *Proceedings of the 37th International Conference on Machine Learning* Bd. 119, PMLR, 13–18 Jul 2020, S. 9690–9700. – URL <https://proceedings.mlr.press/v119/van-amersfoort20a.html>
- [Wan et al. 2020] WAN, Xiang ; LIU, Lilan ; WANG, Sen ; WANG, Yi: A Transfer Learning Strip Steel Surface Defect Recognition Network Based on VGG19. In: WANG, Yi (Hrsg.) ; MARTINSEN, Kristian (Hrsg.) ; YU, Tao (Hrsg.) ; WANG, Kesheng (Hrsg.): *Advanced manufacturing and automation IX* Bd. 634. Singapore : Springer, 2020, S. 333–341. – ISBN 978-981-15-2340-3
- [Wan et al. 2021] WAN, Xiang ; ZHANG, Xiangyu ; LIU, Lilan: An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets. In: *Applied Sciences* 11 (2021), Nr. 6. – URL <https://www.mdpi.com/2076-3417/11/6/2606>. – ISSN 2076-3417
- [Wang et al. 2023] WANG, Fangqiang ; WANG, Zhigao ; FANG, Xia ; WANG, Mei: DeeplabV3+ Image Processing Based Steel Components Corrosion Analysis of Power Transformation Line Towers. In: *2023 IEEE International Conference on Power Science and Technology (ICPST)*, 2023, S. 992–997
- [Wursthorn et al. 2022] WURSTHORN, K. ; HILLEMANN, M. ; ULRICH, M.: COMPARISON OF UNCERTAINTY QUANTIFICATION METHODS FOR CNN-BASED REGRESSION. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2022* (2022), S. 721–728. – URL <https://isprs-archives.copernicus.org/articles/XLIII-B2-2022/721/2022/>
- [Xia et al. 2022] XIA, Tian ; YANG, Jian ; CHEN, Long: Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning. In: *Automation in Construction* 133 (2022), S. 103992. – URL <https://www.sciencedirect.com/science/article/pii/S092658052100443X>. – ISSN 0926-5805
- [Xu et al. 2019] XU, Yang ; BAO, Yuequan ; CHEN, Jiahui ; ZUO, Wangmeng ; LI, Hui: Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. In: *Structural Health Monitoring* 18 (2019), Nr. 3, S. 653–674
- [Yamaguchi & Hashimoto 2010] YAMAGUCHI, Tomoyuki ; HASHIMOTO, Shuji: Fast crack detection method for large-size concrete surface images using percolation-based image processing. In: *Machine Vision and Applications* 21 (2010), Aug, Nr. 5, S. 797–809. – URL <https://doi.org/10.1007/s00138-009-0189-8>. – ISSN 1432-1769
- [Yang & Xu 2020] YANG, Hao ; XU, Xiangyang: Intelligent crack extraction based on terrestrial laser scanning measurement. In: *Measurement and Control* 53 (2020), Nr. 3-4, S. 416–426

- [Zeng et al. 2023] ZENG, Lu ; ZHANG, Shouming ; WANG, Pengju ; LI, Zuze ; HU, Yongjin ; XIE, Tao: Defect detection algorithm for magnetic particle inspection of aviation ferromagnetic parts based on improved DeepLabv3+. In: *Measurement Science and Technology* 34 (2023), Nr. 6, S. 065401
- [Zhang et al. 2017] ZHANG, Allen ; WANG, Kelvin C. P. ; LI, Baoxian ; YANG, Enhui ; DAI, Xianxing ; PENG, Yi ; FEI, Yue ; LIU, Yang ; LI, Joshua Q. ; CHEN, Cheng: Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. In: *Computer-Aided Civil and Infrastructure Engineering* 32 (2017), Nr. 10, S. 805–819
- [Zhang et al. 2018] ZHANG, Kaige ; CHENG, H. D. ; ZHANG, Boyu: Unified Approach to Pavement Crack and Sealed Crack Detection Using Preclassification Based on Transfer Learning. In: *Journal of Computing in Civil Engineering* 32 (2018), Nr. 2, S. 04018001
- [Zhao et al. 2018] ZHAO, Xuefeng ; LI, Shengyuan ; SU, Hongguo ; ZHOU, Lei ; LOH, Kenneth J.: Image-based comprehensive maintenance and inspection method for bridges using deep learning. In: *Smart Materials, Adaptive Structures and Intelligent Systems* Bd. Volume 2: Mechanics and Behavior of Active Materials; Structural Health Monitoring; Bioinspired Smart Materials and Systems; Energy Harvesting; Emerging Technologies American Society of Mechanical Engineers (Veranst.), 2018, S. V002T05A017
- [Zhao et al. 2022] ZHAO, Y. ; SEO, H. ; CHEN, C.: Displacement analysis of point cloud removed ground collapse effect in SMW by CANUPO machine learning algorithm. In: *Journal of Civil Structural Health Monitoring* 12 (2022), Nr. 2, S. 447–463

Abbildungsverzeichnis

1.1	Übersicht der Originalveröffentlichungen zur kumulativen Dissertation	7
1.2	Ablaufdiagramm des automatischen Schadenserkenntnisprozesses aus 3D-Punktwolken. Ausgehend von der originären 3D-Punktwolke und dem mathematischen Modell der Struktur werden zunächst Punkt zu Objekt Distanzen für jeden Punkt berechnet. Diese Distanzen werden in ein Höhenfeld konvertiert, indem sie gerastert und als Grauwertbild gespeichert werden. Mit einem CNN als generischen Merkmalsgenerator und folgendem Local Outlier Factor Algorithmus können Schäden in dem Höhenfeld erkannt werden.	10
1.3	Ausgeschnittenes RGB-Bild (links) und erstelltes Labelbild (rechts) des Jade-Weser-Ports in Wilhelmshaven. Die Klassen unbeschädigt und Korrosion sind in Schwarz bzw. Weiß dargestellt (Hake et al., 2023b).	10
1.4	Ablaufdiagramm des automatischen Schadenserkenntnisprozesses aus Farbbildern. Ausgehend von der Datenerfassung wird zunächst ein Teildatensatz mit manuellen Labeln für die Optimierung versehen und der Gesamtdatensatz in Trainings- und Testdatensatz aufgeteilt. Nach der Auswahl des Modells wird das Netz trainiert und mit den manuell erzeugten Labeln des Trainingsdatensatzes optimiert. Im letzten Schritt folgt die Evaluierung anhand des Testdatensatzes (Hake et al., 2023b).	12
1.5	3D-Kartierung einer Hafenstruktur über und unter Wasser (Hesse et al., 2019).	13
1.6	Entwickelte Messplattform (Foto: Karsten Holste).	13
1.7	Links ist das in dieser Arbeit verwendete Simulationsprinzip zu sehen: An jeder Sensorposition werden alle Strahlen in vertikalen Inkrementen gemäß den Angaben des Herstellers mit den Ebenen der Spundwand verschnitten. In der Mitte ist die simulierte Punktwolke zu sehen und rechts das entsprechende Labelbild, wobei weiß für beschädigte und schwarz für unbeschädigte Bereiche steht (Hake et al., 2020).	14
1.8	Beispiel eines realen Datensatzes aus dem Stadthafen von Lübeck. Das obere Bild zeigt ein Foto der Kaimauer und das untere Bild die entsprechende Punktwolke (Hake et al., 2022).	15
1.9	Zwei Beispiele für beschädigte Bereiche im realen Datensatz. Beide Bilder zeigen Betonabplatzungen (rot markiert). Das obere Bild zeigt drei kleinere Schäden mit einer Größe von bis zu 50 cm × 50 cm. Das untere Bild zeigt einen größeren Abplatzungsschaden, der 1,5 m breit und 30 cm hoch ist (Hake et al., 2022).	15
1.10	Ausgeschnittenes RGB-Bild des Jade-Weser-Ports in Wilhelmshaven (Hake et al., 2023b).	16
2.1	Reale Daten aus dem Stadthafen Lübeck mit Höhenfeld und Labelbild. Die obere Reihe zeigt die Punktwolke in den drei Klassen unbeschädigt (blau), Schaden (grün) und andere Störobjekte (rot). Die mittlere Reihe zeigt das daraus entstandene Höhenfeld in Grauwerten, wobei hellere Grauwerte für größere Distanzen stehen. In der unteren Reihe ist das binäre Labelbild, welches für die Evaluierung benötigt wird. Dabei entspricht schwarz unbeschädigt und weiß Schaden (Hake et al., 2023a).	21
2.2	Abhängigkeit der Bewertungsmetriken vom Schwellenwert (Hake et al., 2023a).	22
2.3	Abhängigkeit der Bewertungsmetriken vom Schwellenwert für reale Daten (Hake et al., 2023a).	24

2.4	Zwei Beispiele für die Schadenserkenkung anhand des realen Datensatzes aus dem Lübecker Stadthafen. Die oberste Zeile ist jeweils die Punktwolke mit überlagerter Schadenserkenkung. Die Farben Grün, Rot und Gelb kennzeichnen jeweils richtig-positive, falsch-negative und falsch-positive Ergebnisse. Die mittlere Reihe ist jeweils das Höhenfeld und die untere Reihe ist das Labelbild mit überlagerten erkannten Bereichen (Hake et al., 2022).	25
2.5	Beispielbild des JadeWeserPorts, oben das originale Farbbild, in der Mitte die manuell erzeugten Label und unten die Label vom schwach überwachten Region Growing Verfahren (Scherff et al., 2023).	27
2.6	Die korrekte Klassifizierung von Korrosion ist grün eingefärbt, blau sind falsch-positive (klassifizierte Korrosion, wo keine vorhanden ist) und rot falsch-negative (die Korrosion wurde nicht erkannt) (Hake et al., 2023b).	30
2.7	Gesamtunsicherheiten und korrosionsklassenbezogene Metriken von optimalem DeepLabV3+-Modell in Bezug auf den Testsatz, trainiert auf schwach überwachten Daten. Die verschiedenen Farben stehen für den Anteil der verwendeten schwachen Labels. Die schraffierten Balkenabschnitte zeigen den Unterschied zwischen IoU- und F1-Score (Scherff et al., 2023).	31
3.1	Vorgehen zur Generierung der Höhenfelder aus mehreren 3D-Geraden als Referenzgeometrien (Lippmann, 2021).	35
3.2	Vorgehen zur Generierung der Höhenfelder mit quadratischen Ausschnitten der Punktwolken (Lippmann, 2021).	36
3.3	Beispielergebnisse der schwachen Label. Oben Links das Originalbild, oben rechts der Simple Does It Ansatz, unten links der Region-Growing-Ansatz und unten rechtes die händisch erzeugten Label. (Scherff, 2022)	39
3.4	Beispielergebnisse der semantischen Segmentierung zur Erkennung von Roststellen mit Prädiktion der Unsicherheit. Die Ergebnisse enthalten das Originalbild, die dazugehörige semantische Segmentierung und die prädizierte und epistemische Unsicherheiten, die durch die Entropie der Vorhersageverteilung bzw. die mutual Information geschätzt werden. Die Unsicherheiten werden als normierte Grauwerte im Bereich von 0 bis 255 dargestellt.	43

Tabellenverzeichnis

1.1	Parameter der DeepLabV3+ Architektur.	11
2.1	Konfusionsmatrix für Recall = 96,2 % (Schwellwert = -1,5).	23
2.2	Konfusionsmatrix für Recall = 72,6 % (Schwellwert = -1,55).	24
2.3	Ergebnisse der ausgewählten Random Forest Klassifikatoren.	28
2.4	Ergebnisse der ausgewählten DeepLabV3+ Klassifikatoren.	29
2.5	Ergebnisse der Kreuzvalidierung des RGB Random Forest und des ResNet-50 DeepLabV3+ Modells.	30
2.6	Ergebnisse des DeepLabV3+-Segmentierungsnetzwerks, trainiert auf 70 % schwach überwachter Bilder in Bezug auf die Testmenge.	31
3.1	Vergleich der verschiedenen Geometrien zur Höhenfelderstellung.	37

Danksagung

An dieser Stelle sei allen Personen, die zur Fertigstellung dieser Arbeit beigetragen haben, mein aufrichtiger Dank ausgesprochen. Ein besonderer Dank gebührt Herrn Prof. Dr.-Ing. Ingo Neumann und Herrn PD Dr.-Ing. Hamza Alkhatib, die mir die Möglichkeit geboten haben, mich intensiv mit der Thematik auseinanderzusetzen. Insbesondere die zahlreichen konstruktiven und intensiven Diskussionen sind hervorzuheben, da sie einen wesentlichen Beitrag zum Erkenntnisgewinn leisteten. Herrn Prof. Dr.-Ing. Vincent Oettel und Herrn Prof. Dr.-Ing. Harald Sternberg gebührt mein Dank für die Übernahme der Korreferate sowie die wertvollen Anregungen nach der Durchsicht der Arbeit. Herrn Prof. Dr.-Ing. habil. Jürgen Müller spreche ich meinen Dank für die Übernahme des Prüfungsvorsitzes aus.

Bei allen Kolleginnen und Kollegen des Geodätischen Instituts Hannover bedanke ich mich für die angenehme und kollegiale Arbeitsatmosphäre. Die gemeinsame Zeit mit euch am Institut wird mir als sehr positiv in Erinnerung bleiben.

Des Weiteren möchte ich mich bei Dr. Hesse und Partner Ingenieure, dem Fraunhofer Institut für Großstrukturen in der Produktionstechnik, der WKC Hamburg GmbH, der Niedersachsen Ports GmbH & Co. KG und der Wasserstraßen- und Schifffahrtsverwaltung für die Zusammenarbeit bedanken.

Des Weiteren möchte ich den Studierenden meinen Dank aussprechen, die mich durch ihre Abschlussarbeiten und Tätigkeiten als studentische Hilfskräfte während meiner Zeit am Geodätischen Institut unterstützt haben.

Ein besonderes Dankeschön geht schließlich an meine Familie und Freunde, die meinen Weg ermöglichen, begleiten und unterstützen haben.

Lebenslauf

Persönliche Daten

Name: Frederic Hake
Geburstag: 21.01.1992
Geburtsort: Langenhagen

Berufliche Laufbahn

seit 09.2024 ALLSAT GmbH Hannover
08.2018 bis 08.2024 Wissenschaftlicher Mitarbeiter am Geodätischen Institut der Leibniz Universität Hannover
04.2018 bis 06.2018 Dr. Hesse und Partner Ingenieure
01.2018 bis 07.2018 Wissenschaftliche Hilfskraft am Geodätischen Institut der Leibniz Universität Hannover

Ausbildung

10.2015 bis 09.2017 Studium der Geodäsie und Geoinformatik an der Leibniz Universität Hannover, Abschluss: Master of Science
10.2012 bis 09.2015 Studium der Geodäsie und Geoinformatik an der Leibniz Universität Hannover, Abschluss: Bachelor of Science
04.2012 bis 06.2012 Praktikum Allsat GmbH, Hannover
11.2011 bis 01.2012 Praktikum Ingenieurbüro Drecoll, Hannover
06.2011 bis 10.2011 Europäischer Freiwilligendienst im Bystrinski-Naturpark Kamtschatka, Russland
09.1998 bis 05.2011 Freie Waldorfschule Hannover, Abschluss: Abitur