Automatisierte Schadstellenermittlung aus Bildaufnahmen von Bauwerken mit Hilfe von Deep Learning

David Crampen

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme

RWTH Aachen University, Mies-van-der-Rohe Straße 1, 52074 Aachen

E-Mail: crampen@gia.rwth-aachen.de

Abstract: Die Instandhaltung von Bestandsbauwerken ist einer der wichtigsten Aspekte im Gebäudelebenszyklus. Die bei Begehungen durch den Bausachverständigen gesammelten Bilddaten zu Schäden haben das Potenzial den Prozess der Schadensdokumentation zu transformieren. Für die Adaption des vorgestellten digitalen Ansatzes für die Schadensdokumentation ist eine Minimierung des durch die neue Methodik entstehenden Mehraufwands, aufgrund des engen Kostenrahmens, in dem sich Bauleistungen und Maßnahmen zur Instandsetzung bewegen, bedeutsam. Darüber hinaus kann die Digitalisierung von analogen Prozessen zum einen zur Entlastung von Akteuren und zum anderen zur Strukturierung von Abläufen führen. Durch die Nutzung von ohnehin anfallenden Schadstellenbildern können unter Einsatz von maschinellen Lernverfahren zusätzliche Daten generiert und der Prozess der Schadensbewertung unterstützt werden. Im Beitrag wird eine bildbasierte Methode zur Erkennung von Schadstellen an Bauwerken vorgestellt. Zur Schätzung der Fläche der segmentierten Schäden wurde ein Referenzgerät entwickelt, das zur Herstellung eines optischen Maßstabs direkt bei der Aufnahme von Schäden eingesetzt wird. Für die Segmentierung kommt das faltende neuronale Netzwerk Mask-R-CNN zum Einsatz, das für jeden erkannten Schaden Einzelinstanzen generiert, sodass multiple Schäden in einem Bild separat verarbeitet werden können. Der für das Training eingesetzte Datensatz wurde speziell für den Anwendungsfall manuell generiert. Es konnten vielversprechende Ergebnisse erzielt und eine Basis für weiterführende Forschung geschaffen werden.

Keywords: Automation, Neuronale Netzwerke, Instanz-Segmentierung, Bauwerksschäden

1 Einführung

Der Einsatz von künstlicher Intelligenz dringt seit Jahren in nahezu alle Bereiche des Ingenieurwesens vor. Dabei verspricht der Einsatz von künstlicher Intelligenz eine flexible Möglichkeit für das Lösen einer großen Bandbreite von bestehenden Problemen. Bei ausreichender Datenverfügbarkeit stellt Deep Learning einen leistungsfähigen Ansatz für die Automatisierung von verschiedensten Aufgaben dar. Vor allem im Teilbereich „Computer-Vision“ wurden zuletzt Meilensteine in der Zuverlässigkeit und Verarbeitungsgeschwindigkeit von neuronalen Netzwerken zur Bilderkennung erreicht. Dazu kommen im Regelfall faltende neuronale Netzwerke zum Einsatz, die aufgrund ihres zweidimensionalen Schichtaufbaus die Positionen einzelner Bildpixel verarbeiten können. [5]

Bei den Modelltypen wird zwischen Klassifizierungsmodellen, Objekt-Detektionsmodellen wie Fast-RCNN [3] und Faster-R-CNN [6], Modellen zur semantischen Segmentierung und Instanz-Segmentierungsmodellen unterschieden [1]. Bei der Instanz-Segmentierung wird jede Objektklasse auf Pixelbasis in Einzelinstanzen zerlegt. Neben U-Net [7] ist die Mask-R-CNN Architektur [4] eine der erfolgreichsten Architekturen zur Segmentierung von Einzelinstanzen. Die Mask-R-CNN Architektur wird im vorgestellten Ansatz für die Segmentierung von Schäden genutzt.

Im Bauwesen findet die automatisierte Schadenserkennung unter Einsatz von Deep Learning heute vor allem im Infrastrukturbau Anwendung. Die Möglichkeit mit speziell aufgerüsteten Fahrzeugen schnell ganze Straßenzüge aufnehmen zu können, führt zu einer soliden Datenbasis, auf der robuste Modelle trainiert werden können. Generell ist die Risserkennung ein besonders stark beforschter Bereich, zum Beispiel werden in Tang, Mao et al. [9] Risse auf Betonoberflächen wie Dämmen erkannt. Bei der Schadensaufnahme wird, neben der Messung der Rissbreite, in der Regel ein Maßstab in den Schadensbildern aufgenommen, um das ungefähre Schadensausmaß dokumentieren zu können. Eine einfache Variante für einen solchen Maßstab sind Messmarken. Diese werden auf der Wand aufgeklebt und zusammen mit dem Schaden im Bild erfasst. Der Nachteil dieser Methode ist der erhöhte Aufwand für das Befestigen der Marken und die geringe Effizienz einer Marke. Ein besserer Ansatz verwendet einen Entfernungsmesser, der zusammen mit den spezifischen Eigenschaften der eingesetzten Kamera in einen Maßstab umgerechnet werden kann. Dazu muss zwar nichts an der Wand selbst befestigt werden, jedoch erfordert dieser Ansatz die Kopplung bzw. Zuordnung von Entfernungsmesser und Bild, was den Aufwand im post-processing erhöht. Die neu entwickelte Methode stellt einen hybriden Ansatz aus visuellem Maßstab und flexibler Aufnahme aus der Ferne dar. Inspiriert von Chang, Lin et al. [2] wurde ein Referenzgerät aus vier parallelen Lasern entwickelt, das ein quadratisches Objekt auf die Schadensebene projiziert und in der Bildaufnahme platziert wird. Mit diesem Ansatz sind alle notwendigen Informationen in einer Datenquelle (Bild) enthalten, wodurch ein Umformen oder Matching von unterschiedlichen Datenquellen umgangen wird, während der Aufwand für das manuelle Befestigen von Maßstäben vermieden wird.

2 Methoden

Im Folgenden wird die entwickelte Methodik zur Segmentierung und Flächenschätzung von Schadstellen vorgestellt. Dazu wird in 2.1 zunächst der Prototyp des Referenzsystems vorgestellt. In 2.2 wird auf den verwendeten Datensatz eingegangen. In 2.3 wird die Strategie zur Optimierung der Hyperparameter der verwendeten Deep Learning-Architektur beschrieben.

2.1 Referenzsystem

Für die Referenzierung der Schadenstellenbilder, sind die Eigenschaften Flexibilität und Geschwindigkeit von besonderer Bedeutung. Darüber hinaus ist die Minimierung der Kosten zur Erstellung des Referenzsystems von Interesse. [2] lieferten die Inspiration für die Entwicklung des Referenzlasers. Das System besteht aus einem Kunststoffgehäuse, das im 3D-Druckverfahren hergestellt wurde und vier Laserdioden. Dadurch konnten die Herstellungskosten minimiert werden und eine Herstellung nach dem „Rapid-Prototyping-Prinzip“ durchgeführt werden, was gerade in frühen Phasen, in denen Vor- und Nachteile von verschiedenen Lösungen geprüft werden, von Vorteil ist. Das resultierende Gerät ist in Abbildung 1 dargestellt. Durch die I-Form des Gerätes lässt sich ein Smartphone direkt an dem Gerät befestigen, wodurch die relative Position des projizierten Referenzobjektes in den aufgenommenen Bildern weitgehend fixiert werden kann.

Abbildung 1: Referenzgerät zur Erzeugung eines optischen Maßstabs

Die Fläche des Referenzmaßstabs auf der Wand beträgt ca. 150cm². Der Vorteil der parallel angeordneten Laser, ist die Unabhängigkeit vom Abstand zu einer flachen Oberfläche. Dafür müssen die Laser jedoch orthogonal zur Schadensebene ausgerichtet sein. Das Gerät kann über einen externen Schalter ein- und ausgeschaltet werden. Die Sammlung der Daten für den Datensatz wurde mit dem vorgestellten Referenzgerät und einer Smartphone-Kamera durchgeführt.

2.2 Datensatz

Der vollständige Datensatz besteht aus insgesamt 1800 Schadensbildern. Davon stammen ca. 500 Bilder aus dem Internet und 1300 Bilder wurden manuell aufgenommen. Die eigenen Aufnahmen bilden ausschließlich den Außenbereich zu unterschiedlichen Jahreszeiten ab. Die Häufigkeitsverteilung der Objektklassen in den Aufnahmen ist weitestgehend ausgewogen. Die im Datensatz vorhandenen Objektklasssen sind: Riss, Abplatzung, freiliegender Stahl und Referenzobjekt. Letzteres ist auf jedem der 1300 selbst aufgenommenen Bilder bei unterschiedlichen Lichtverhältnissen abgebildet. Die Dauer für die Aufnahme eines Schadens mit dem Referenzobjekt belief sich auf ca. 2-6 Sekunden für ein Bild. Für die Kalibrierung und Validierung der im späteren Verlauf vorgestellten Flächenschätzung wurden 50 Schadstellen manuell aufgemessen. Die zeitlichen Dimensionen des Dokumentationsprozesses mit dem Referenzgerät und die der manuellen Dokumentation zu Kalibrierungszwecken lassen sich vergleichen und zeigen das Potenzial zur Beschleunigung des Prozesses, das durch den Einsatz der vorgestellten Methodik nutzbar gemacht wird, denn das manuelle Vermessen ist von Form und Komplexität der Schadstelle abhängig und nahm in vielen Fällen mehrere Minuten in Anspruch, wohingegen der vorgestellte Ansatz unabhängig vom Schadensausmaß ist.

2.3 Deep Learning-Modell

Mask-R-CNN wurde als Architektur für die Schadenssegmentierung ausgewählt. Es nutz ein FPN (feature pyramid network), das Objekteigenschaften in verschiedenen Auflösungen und unterschiedlichen Größen erlernen kann. Dieser Aspekt unterstützt die Flexibilität der Anwendung, da es zur Erkennung von Objekten keines fixen Abstandes zum Schaden bedarf.

Die „RoI-Align-Layer“ in Mask-R-CNN ermöglicht eine Segmentierung auf Pixelbasis, sie ersetzt die im Vorgänger Faster-R-CNN genutzten RoI-Pooling-Layer und befähigt das Netzwerk erst eine Klassifizierung auf Pixelbasis zu vollziehen. In der Konfiguration von Mask-R-CNN wird ResNet101 als backbone des Netzwerks ausgewählt.

Aufgrund der Zielsetzung der Flächenschätzung ist es von besonderem Interesse, dass das resultierende Modell möglichst präzise Objektmasken erzeugt. Aus diesem Grund wird die im Training relevante Verlustfunktion angepasst, indem der Verlustparameter Validation-mask-loss mit dem Faktor zehn übergewichtet wird. Dies führt zu einer stärkeren Modellanpassung im Falle einer unpräzise prädizierten Instanz-Maske.

Die Optimierung der Hyperparameter wird manuell durchgeführt. Da das Training eines Modells in 100 Epochen auf der verfügbaren Hardware jeweils ca. neun Stunden dauert und Mask-R-CNN die Anpassung von vielen Hyperparametern erlaubt, würde ein Grid-Search-Ansatz zur Optimierung der Hyperparameter den zeitlichen Rahmen sprengen. Die Optimierung folgt einer selbst festgelegten heuristischen Hierarchie mit Baumstruktur, bei der auf jeder Ebene ein Parameter verändert wird.

Als entscheidungsrelevantes Kriterium für die Fortsetzung des Entscheidungsbaums fungiert die mean Average Precision. Diese stellt die durchschnittliche Präzision der Detektion über alle Objektklassen dar. Die mAP wird indirekt über die IoU (Intersection over Union) berechnet, die einen notwendigen Parameter für die Berechnung von Precision und Recall darstellt. Die IoU fungiert als Schwellenwert für die Ähnlichkeit der Objekt Bounding Boxes zwischen Detektionen und Validierungsdaten, über den ein tp (true positive) bestimmt wird. Die mAP kann sowohl auf Basis der Mask-IoU als auch basierend auf der Bounding-Box-IoU bestimmt werden. Generell ist die Bounding-Box-IoU der Standardparameter für den Vergleich verschiedener Modelle, was hauptsächlich in der Verbreitung von Objekt-Detektionsmodellen begründet ist. Modelle zur Instanz-Segmentierung werden jedoch ebenfalls regelmäßig über die mAP der Bounding Box verglichen. Aus diesem Grund wird für den vorliegenden Ansatz ebenfalls die mAP der Bounding Boxes herangezogen.

Für das Training eines Deep Learning Modells im Bereich der Computer-Vision ist es üblich vortrainierte Modellgewichtungen zu verwenden. So kann sehr viel Zeit eingespart werden, weil das Modell lediglich „umtrainiert“ werden muss, statt die Fähigkeit der Bilderkennung von Grund auf zu erlernen. Stufe 1 der Optimierung stellt die Wahl der vortrainierten Gewichtungen dar. Auf Stufe 2 steht die Wahl des Optimizers, der die Hauptkomponente für die Anpassung der Modellgewichtungen nach jeder Iteration durch das Netzwerk übernimmt. In Stufe 3 wird die Lernrate, also das Ausmaß der Anpassung der Gewichtungen optimiert. Die Gradient-Clip Norm auf Stufe 4 stellt einen Schwellenwert für das Ausmaß der Anpassungen in Folge von besonders großen Fehlern innerhalb einer Trainingsiteration dar. Das Lernmomentum wird in Stufe 5 angepasst und beeinflusst die Konvergenz der Lernrate im Lernprozess. Das RoI-Positive-Ratio auf Stufe 6 regelt das Verhältnis von im Training eingesetzten positiven und negativen Samples. In der letzten Stufe 7 werden verschiedene Verfahren der Data Augmentation verglichen, um die Robustheit des Modells zu erhöhen und ein overfitting zu vermeiden, welches bei einem kleinen Datensatz wie dem verwendeten, eine große Gefahr für die Modellperformance darstellt. Overfitting bezeichnet die zu starke Anpassung des Modells an die Trainingsdaten, was zu einer verringerten Generalisierbarkeit und damit auch zu einer schlechten Performance gegenüber neuen Daten führt.

3 Ergebnisse

Im Folgenden werden die Ergebnisse der Untersuchungen zur Modelloptimierung, dann die Ergebnisse der Detektion veranschaulicht und das Ergebnis der Flächenschätzung vorgestellt.

Die Optimierung der Hyperparameter resultierte wie bereits beschrieben in einem Entscheidungsbaum. Als beste Konfiguration stellte sich der Einsatz der COCO-Weights als Basis für das Transferlernen auf Stufe 1 heraus. Die Wahl des Optimizers fiel auf Adam (Adaptive Moment Estimation). Darüber hinaus schnitt die Standardkonfiguration am besten ab. So wurde auf dem eigenen Datensatz eine mAP von 0,46 erreicht.

Für die weitere Optimierung durch den Einsatz von Data Augmentation wurde zunächst geprüft, ob eine Nutzung von Data Augmentation ab Trainingsepoche 1 oder erst in späteren Epochen bessere Ergebnisse erzielt. Das Resultat dieser Prüfung ist eine bessere Modellperformance bei bereits frühem Einsatz von Data Augmentation. Zur weiteren Optimierung wurden verschiedene Data Augmentation Verfahren kombiniert, dazu zählten Rotationen, Spiegeln, Offset, Cutout, Farbänderungen, Gaußfilter und Skalierungen [8]. Die mAP konnte durch den Einsatz von Data Augmentation von 0,46 auf 0,49 gesteigert werden.

3.1 Detektionsergebnisse & Flächenschätzung

Bei der visuellen Begutachtung der prädizierten Schäden wurden verschiedene Modelle miteinander verglichen ein Beispiel für die Ergebnisse von verschiedenen Detektionsmodellen ist in Abbildung 2 dargestellt. Anhand der Ergebnisse zeigt sich, dass der Einsatz von zu viel Data Augmentation das Modell „verwirren“ kann, was dazu führt, dass Objekte wie Schotter o.ä. als Abplatzung segmentiert werden.

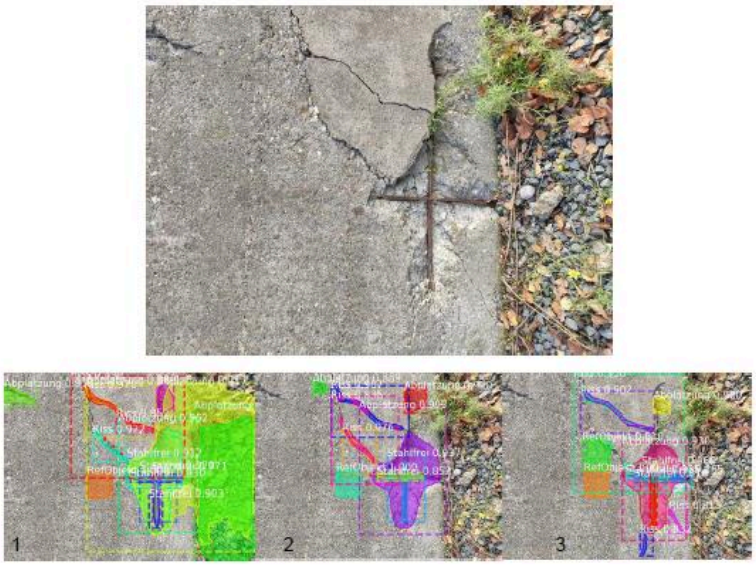

Abbildung 2: Prädizierte Schäden unterschiedlicher Modelle (1 geometrische Data Augmentation, 2 keine Data Augmentation, 3 netzbasierte Data Augmentation

Die Flächenschätzung erfolgt durch die Umrechnung von der Pixelmaske des Referenzobjektes und den realen Maßen des projizierten Quadrates in einen Umrechnungsfaktor, über den die Masken der Schadstellen in reale Maße umgerechnet werden können. Die Kalibrierung der Flächenschätzung gestaltet sich schwierig, da keine Metrik für die Bewertung der Präzision der manuell vermessenen Schäden existiert. Es wurde zunächst angenommen, das die 50 Kalibrierungsdaten korrekt vermessen wurden. Trotz der Kalibrierung verbleibt eine mittlere Abweichung der Fläche von 20% für Risse, 11% für Abplatzungen und 9% für freiliegenden Stahl.

4 Diskussion

Im Folgenden wird zunächst auf den Nutzen und die Kritik an dem entwickelten Workflow eingegangen, bevor auf die Erkenntnisse und Ideen zur Verbesserung eingegangen wird.

Der Mehrwert des entwickelten Workflows liegt in der Beschleunigung und Verbesserung der Schadensdokumentation. Darüber hinaus kann die digitale Darstellung von Schäden in digitalen Bauwerksmodellen die Schadensbewertung verbessern, was positiven Einfluss auf den Instandhaltungsprozess nehmen würde. Der Prozess der Schadensaufnahme kann beschleunigt und seine Ergebnisse digital nutzbar gemacht werden, sofern der Workflow in Zukunft zuverlässig eingesetzt werden kann. Außerdem ist der Ansatz flexibel in Innenräumen einsetzbar. Zusätzlich könnte die Entwicklung von Schäden in Zeiträumen zwischen zwei Aufnahmen analysiert werden.

Um das Verfahren in der Praxis einsetzen zu können, müssen jedoch noch einige Aspekte verbessert werden. Die herausgearbeiteten Kritikpunkte am aktuellen Status liegen in der begrenzten Verfügbarkeit von Daten, was zu großen verbleibenden Unsicherheit des Modells für den Anwendungsfall führt. Zudem setzt sich die Ungenauigkeit der Flächenschätzung aus Ungenauigkeiten der Erkennung des Schadens und zusätzlich Ungenauigkeiten in der Detektionsmaske des Referenzobjektes zusammen. Ein weiterer Kritikpunkt ist die Genauigkeit des entwickelten Gerätes selbst, dieser Kritikpunkt wurde bereits bei den Überlegungen in der frühen Phase der Entwicklung in Kauf genommen, da es sich zunächst nur um einen Prototyp handelt und die Kosten für die Entwicklung möglichst gering sein sollten.

Aus den Kritikpunkten ergibt sich auch das Verbesserungspotenzial für den vorgestellten Ansatz. Ein größerer Datensatz wird die Robustheit des Modells steigern, ein präziseres Referenzgerät wird die Abweichungen der Projektion auf der Schadensebene minimieren. Dazu könnte konkret ein starreres Gehäuse genutzt werden, zudem würde die Möglichkeit zur Nachjustierung der Laser ebenfalls die Genauigkeit des projizierten Objekts steigern. Eine kleinschrittigere Differenzierung der Objektklassen abhängig vom aufgenommenen Material oder die Differenzierung von Rissen nach Rissbreiten könnte potenziell ebenfalls die Detektion verbessern und würde den Informationsgehalt der Dokumentation erhöhen.

5 Fazit & Ausblick

Zusammenfassend lässt sich feststellen, dass der vorgestellte Ansatz ein großes Potenzial für die Vereinfachung der Schadensaufnahme und Dokumentation bieten kann. Dazu muss jedoch die Genauigkeit und Robustheit des entwickelten Models zur Segmentierung der Schäden verbessert werden. Das Übertragen der Ergebnisse der Detektion in digitale Planungsmodelle ist bereits jetzt möglich, im Rahmen des Projektes DigiPark [10] wurde die Möglichkeit zur automatischen Positionierung der Schäden im digitalen Bauwerksmodell geschaffen, was in Kombination mit dem

verbesserten vorgestellten Ansatz zukünftig einen automatischen Workflow möglich machen wird. Der Ansatz ist schnell und vor allem einfach, weshalb eine Adaption in der Praxis, nach der Verbesserung der in 4.2 beschriebenen Aspekte denkbar ist. Zukünftig könnte der Ansatz durch Tiefendaten erweitert werden, was die Bewertung von Schäden in digitalen Modellen möglich machen und damit die generelle Qualität der Schadensbewertung verbessern könnte.

Literatur

[1] Aggarwal, Charu C. (2015): “Data Mining”, Springer International Publishing.

[2] Chang, Wen-Yi; Lin, Franco; Liao, Tai-Shan; Tsai, Whey-fone (2019 - 2019): “Remote Crack Measurement Using Android Camera with Laser-Positioning Technique” in 4th International Conference on Control, Robotics and Cybernetics (CRC) Tokyo, Japan, 9/27/2019 - 9/30/2019: IEEE, S. 196–200.

[3] Girshick, Ross (2015): “Fast R-CNN” in Proc. IEEE Int. Conf. on Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.

[4] He, Kaiming; Gkioxari, Georgia; Dollár, Piotr; Girshick, Ross (2017): “Mask R-CNN” in Proc. IEEE Int. Conf. on Comput. Vis. (ICCV), Oct. 2017, pp. 2961–2969.

[5] O'Shea, Keiron; Nash, Ryan (2015): “An Introduction to Convolutional Neural Networks”, Department of Computer Science, Aberystwyth University, Ceredigion, SY23 3DB 2015.

[6] Ren, Shaoqing; He, Kaiming; Girshick, Ross; Sun, Jian (2015): “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks” in Proc. IEEE Transactions on Pattern Analysis and Machine Inteligence (TPAMI) 2015.

[7] Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015): “U-Net: Convolutional Networks for Biomedical Image Segmentation” in Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015.

[8] Shorten, Connor; Khoshgoftaar; Taghi M. (2019): “A survey on Image Data Augmentation for Deep Learning” in J Big Data 6, 60 2019.

[9] Tang, Jianghong; Mao, Yingchi; Wang, Jing; Wang, Longbao (2019): “Multi-task Enhanced Dam Crack Image Detection Based on Faster R-CNN” in IEEE 4th International Conference on Image, Vision and Computing, July 5-7, 2019.

[10] Blut, Christoph; et. al. (2021): “DigiPark - Digitalisierung in der Bauwerksinstandsetzung” in 7. Kolloquium Erhaltung von Bauwerken : Fachtagung zur Beurteilung, Instandhaltung und Instandsetzung von Bauwerken 2021, pp. 91-100.