

Maschinelle Schadenserkennung auf Basis von Bildaufnahmen

Definition



Die maschinelle Schadenserkennung auf Basis von Bildaufnahmen ist ein Verfahren zur automatischen Analyse von Schäden in Einzelbildaufnahmen, bei der die Schäden in einzelne Typen eingeteilt, durch Rechtecke umgrenzt oder mithilfe von Bildmasken hervorgehoben werden. Das Verfahren ist berührungslos und somit auch zerstörungsfrei und kann überall dort eingesetzt werden, wo ein Objekt fotografisch abgebildet und ausgewertet werden kann.

Anwendungsfall

Anwendungsbereich



- Bildhafte Erfassung und Dokumentation von Schadstellen an Bauwerken, wie bspw. Risse und Abplatzungen bzw. Abblätterungen.
- Manuelle oder (halb)automatische Detektion und *Klassifizierung* visuell identifizierbarer Schäden.
- Bildbasierte Vermessung von erfassten Schäden.

Rahmenbedingungen



- Die Aufnahmesensorik (Kamera) erfordert die Ausleuchtung der zu vermessenden Oberfläche durch natürliches oder künstliches Licht im sichtbaren Wellenlängenbereich.
- Die Objekte sind möglicherweise nur schlecht zugänglich bzw. es herrscht vor Ort ein hohes Gefahrenpotenzial (zum Beispiel Industrieanlagen).
- Vor Ort sind möglichst kurze Messzeiten gewünscht oder erforderlich (zum Beispiel Wasserbauwerke mit Zugangsbeschränkungen aufgrund hochfrequenten Betriebes).

Technischer Hintergrund

Erläuterung der Funktionsweise



- Die zu untersuchende Bauwerksoberfläche wird mit einer Kamera aus einer vorzugsweise senkrechten Aufnahmerichtung zur Oberfläche mit einem Einzelbild fotografiert. Dabei sollte der erfasste Schaden möglichst vollständig im Bild abgebildet werden und in der Tiefenschärfe der Kamera liegen. Die genaue innere Abbildungsgeometrie der Kamera (bspw. Objektivbrennweite und Fokuseinstellung) spielt dabei eine untergeordnete Rolle.
- Für die Erkennung von Schäden wird zunehmend auf sogenannte maschinelle Lernverfahren, vorwiegend Neuronale Netze, zurückgegriffen. Insbesondere bei Bildern werden meist faltungsartige Neuronale Netze (engl. Convolutional Neural Networks, kurz: CNNs) verwendet, die eigenständig Merkmale in den Bildern, die bspw. zur *Klassifizierung* maßgeblich entscheidend sind, erlernen und extrahieren. Je nach Anforderung bzw. Art und Architektur des Netzwerks liefert dieses dann die Ausgabe einer Schadenstyp (bei einer Klassifizierungsaufgabe), eines umgrenzenden Rechtecks (engl. Bounding Box) (bei einer Detektionsaufgabe), einer Bildmaske (bei einer *Segmentierungsaufgabe*) oder auch Kombinationen dieser für die Schäden in den Bildern. Eine Bildmaske ist dabei

eine Matrix, welche beispielsweise die Bereiche eines Zielbildes, die von Interesse sind, markiert. Im Falle der Schadenserkenkung würde eine solche Bildmaske beispielsweise die Bereiche eines Schadensbildes markieren, in denen sich der Schaden befindet.

- Vor dem Produktiveinsatz eines Neuronalen Netzwerks muss das Netz jedoch zunächst mit entsprechenden Trainingsdaten konfrontiert werden, um Muster und Gesetzmäßigkeiten in diesen zu erlernen. Im Falle der Schadenserkenkung umfassen die Trainingsdaten Bilder von Schäden und die entsprechenden Zielausgaben, bestehend aus bspw. den Schadenstypen, um-grenzenden Rechtecken oder auch Bildmasken, je nach Aufgabenstellung. In diesem Zusammenhang spricht man auch von der Trainingsphase des Netzwerks. Bei einer ausreichend hohen Anzahl an Trainingsdaten und -zyklen kann das Netzwerk die Daten schließlich ausreichend gut generalisieren, um für neue, unbekannte Daten zuverlässige Prädiktionen liefern zu können. Dies wird häufig auch als Inferenzphase des Netzwerks bezeichnet.
- Die Auswertung geschieht vollautomatisch. Typischerweise weisen die trainierten Netze jedoch eine gewisse Fehlerrate auf, so dass bei vielen Anwendungen, wie auch bei der Schadenserkenkung, eine nachträgliche manuelle Begutachtung der Ergebnisse erforderlich ist.
- Bei georeferenzierten Aufnahmen bzw. bei bekannter äußerer Orientierung der Kamera ist es zudem möglich, die in den Bildaufnahmen erkannten Schäden räumlich zu verorten und auf diese Weise auch bspw. in ein bestehendes digitales Bauwerksmodell (BIM-Modell) zu überführen.

Mehrwert

Erreichbarkeit / Arbeitssicherheit



- Verfahren ist berührungslos und somit auch zerstörungsfrei
- Kurze Aufenthaltszeiten vor Ort erleichtern die Einhaltung von Vorschriften zur Arbeitssicherheit

Zeitersparnis Inspektion



- Die Zeitersparnis ist sowohl bei den Arbeiten vor Ort als auch – beim Einsatz automatisierter Bildanalyseverfahren – bei der späteren Auswertung gegeben.
- Den größten Zeitumfang beansprucht vorab (meist) einmalig sowohl die Akquise von Trainingsdaten als auch die Konstruktion, Optimierung und das Training des Neuronalen Netzes.
- Es entsteht i.d.R. ein zusätzlicher Zeitaufwand für eine nachträgliche Begutachtung bzw. Validierung der Ergebnisse durch einen menschlichen Operateur.

Kosten



- Erfassungssensorik (gewöhnliche (Smartphone-)Kamera): ca. 500,- EUR (Marktrecherche 2022)
- Dienstleister für die Datenerfassung und/oder die Auswertung der Daten: Kosten in Abhängigkeit von der Größe des Objektes
- Ggf. Personalkosten für nachträgliche manuelle Begutachtung der Ergebnisse

Datenqualität



- Die Aufnahmen dokumentieren neben den geometrischen gleichzeitig auch die bildlich-semantischen Objekteigenschaften.
- Insbesondere die maschinelle Bildanalyse zur automatisierten Schadenserkenkung ist noch eine vergleichsweise junge Methode, die sich noch in der Entwicklung befindet, so dass die Qualität der Detektion und *Klassifizierung* heute noch deutlich variieren kann. Erfahrungen mit selbst trainierten Neuronalen Netzwerken haben bei Betonoberflächen in Parkbauten (vgl. Forschungsprojekt „DigiPark“) vielversprechende Ergebnisse geliefert. Der Algorithmus eines Dienstleisters zur bildbasierten Schadensdetektion in einer Schleuse zeigte dagegen unbefriedigende und für die Praxis unbrauchbare Ergebnisse.

Voraussetzungen

Hardware



- Gewöhnliche (Smartphone-)Kamera (mit ≥ 5 Megapixel)
- Computer mit mittlerer Leistung (z.B. i5-Prozessor, 16 Gigabyte Arbeitsspeicher)

Vorbereitung Untersuchungsobjekt



- Zu vermessenden Oberfläche sollte unverdeckt sein.
- Oberfläche sollte gereinigt sein, damit der Schaden im Bild sichtbar ist und sich deutlich von der Umgebung abhebt.
- Eine diffuse, gleichmäßige Ausleuchtung der zu vermessenden Oberfläche ist erforderlich

Vorbereitung Datenerhebung



- Auswahl der Sensorauflösung (Pixelanzahl) und des Aufnahmeabstands im Hinblick auf die notwendige Detailerkennbarkeit.
- Definition des angestrebten Zieles: *Klassifizierung*, *Detektion* oder *Segmentierung* der Schäden.
- Das Modell des maschinellen Netzwerks muss auf zum Objekt vergleichbaren Daten (z.B. Risse/Schäden auf Betonoberflächen) trainiert worden sein

Umweltbedingungen



- Normales Tageslicht stellt gewöhnlich ausreichende Lichtverhältnisse sicher; ggf. muss künstliche Beleuchtung vorgehalten werden.
- Schattenwürfe auf die zu vermessende Oberfläche sind möglichst zu vermeiden.
- Trockene Witterung nicht zwingend erforderlich...
- Oberflächen sollte eine homogene Textur aufweisen (z.B. Vermeidung von sowohl nassen als auch trockenen Oberflächen)

Erforderliche Genehmigungen



- Genehmigungen sind ggf. für die Zugänglichkeit oder Befliegung des Objektes zur Bildaufnahme erforderlich

Umsetzung

Datenerhebung



- Erstellung der Aufnahmen gemäß den Vorüberlegungen zur Datenerhebung

Datenprozessierung



- *Neuronales Netz* erhält als Eingabe Bildaufnahmen, für welche Prädiktionen vorgenommen werden.
- Ggf. zusätzliche (Vor-)Prozessierung für das Training des Netzwerks erforderlich.

Kompatibilität mit anderen innovativen Methoden



- Das Verfahren kann mit allen photogrammetrischen Verfahren (UAV, terrestrisch) und grundsätzlich auch in Ergänzung zu komplementären Verfahren wie Laserscanning eingesetzt werden.
- Verschiedene Neuronale Netze können kombiniert bzw. komplementär eingesetzt werden, um für unterschiedliche Anwendungsfälle bzw. Anforderungen die Vorteile dieser Netze zu vereinen oder auch um robustere Prädiktionen zu gewährleisten.

Output

Informationen



- Schadentyp
- Koordinaten von umgrenzenden Rechtecken
- *Bildmaske*

- *Bildlich-semantische* Helligkeits- und Farbinformationen

Dateiformate

- Aufnahmen in Standard-Rasterdatenformaten (i.d.R. JPG)
- Klassenangaben: ASCII
- Koordinaten der umgrenzenden Rechtecke: ASCII
- *Bildmasken* in Standard-Rasterdatenformaten (i.d.R. JPG)
- Konvertierbarkeit in beliebige, andere Datenformate generell möglich

Referenzen

Beispielanbieter

- Geodätisches Institut der RWTH Aachen (gia, <http://www.gia.rwth-aachen.de>)
- PALFINGER Structural Inspection GmbH (STRUCINSPECT, <https://strucinspect.com>)
- <https://www.infralytica.com>
- <http://swissinspect.io>

Beispielanwendungen

- Erkennung von Rissen auf Betonoberflächen durch statistische Analyse von Bildhistogrammen (Jutz/Özcan 2022)
- Forschungsprojekt „DigiPark“: Entwicklung eines mobilen Systems zur georeferenzierten Erfassung und Dokumentation von Bauwerksschäden – Pilotprojekt an Parkhäusern (Geodätisches Institut RWTH Aachen 2018-2021)
- Automatisierte Erkennung von Schäden Brückenbauwerken aus Drohnenbildaufnahmen (Fa. STRUCINSPECT)

Quellen

- Jutz L./Özcan B.: Automated Detection of Cracks based on Statistical Analysis of Image Histograms. 33. Bauinformatik Forum, München (2022)
- Sesselmann et al.: Einsatz von Deep Learning zur automatischen *Detektion* und Klassifikation von Fahrbahnschäden aus mobilen LiDAR-Daten. AGIT – Journal für Angewandte Geoinformatik, 5-2019, S. 100-114. Wichmann Verlag
- Eisenbach et al.: Enhancing the Quality of Visual Road Condition Assessment by Deep Learning. World Road Congress 2019, Abu Dhabi, UAE, 2019
- Seichter et al.: How to Improve Deep Learning based Pavement Distress Detection while Minimizing Human Effort. 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)
- Dais et al.: Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. Automation in Construction, 125. 2021
- Perez H./Tah J.: Deep learning smartphone application for real-time detection of defects in buildings. Structural Control and Health Monitoring. 2021. 28, 7.

- Zhan, T., Yin, K., Xiong, J., He, Z., & Wu, S. T. (2020). Augmented reality and virtual reality displays: perspectives and challenges. *Iscience*, 23(8), 101397.

Glossar:

Klassifizierung:	Einteilung eines Objekts zu einer bestimmten Klasse aus einer vordefinierten Menge von Klassen.
Detektion:	Lokalisierung und Klassifizierung (s.o.) von Objekten. In den Bildern werden Objekte gesucht, die durch umgrenzende Rechtecke lokalisiert werden. Zusätzlich wird jedem gefundenen Objekt eine Klasse zugewiesen.
Segmentierung:	Zerlegung eines Ganzen in einzelne Segmente bzw. Abschnitte. Bei Bildern wird eine Segmentierung durchgeführt, indem das Bild in zusammenhängende Bereiche bzw. Regionen unterteilt wird bzw. die Pixel in dem Bild einzelnen Segmenten zugeordnet werden.
Faltendes Neuronales Netz:	Eine spezielle Art des Neuronalen Netzes, welche häufig bei zwei- oder auch dreidimensional und in einer geordneten Reihenfolge vorliegenden Daten verwendet wird. Bei dem „Standardmodell“ eines Neuronalen Netzes (dem Multi-layer Perceptron) durchlaufen die Eingabevariablen das Netz unabhängig voneinander. Insbesondere bei der Bilderkennung spielt die Nachbarschaft von Pixeln jedoch eine entscheidende Rolle. Bei Bildern werden daher aufgrund der zweidimensionalen und geordneten Struktur von Pixeln fast ausschließlich faltende Neuronale Netze verwendet. Diese beinhalten in einem ersten Schritt eine Abfolge von sogenannten Faltungsoperatoren (Bildfaltung (s.u.)) und Pooling (s.u.), um aus einem Zielbild relevante Merkmale zu extrahieren. Diese Merkmale durchlaufen anschließend den restlichen Teil des Netzes, welches der Eingabe schließlich bspw. eine bestimmte Klasse zuordnet.
Bildmaske:	Eine Matrix, welche bestimmten Bereiche in einem Zielbild markiert. Die Matrix ist dabei in der Regel genauso groß wie das Zielbild und jeder Wert in der Matrix korrespondiert mit dem Pixel an der gleichen Position im Zielbild. Bei der Segmentierung (s.o.) von Bildern können Bildmasken eingesetzt werden, um den Pixeln des Zielbildes eine bestimmte Gruppe zuzuweisen (durch die Zuweisung eines Wertes für eine bestimmte Klasse an der entsprechenden Position in der Matrix).
Bildfaltung:	Anwendung einer Filtermatrix auf ein Zielbild. Die Filtermatrix hat dabei oft eine kleine und ungerade Größe (bspw. 3×3, 5×5, ...) und wird mit jedem Pixel des Zielbildes verknüpft. Hierbei wird eine gewichtete Summe des jeweiligen Pixels und seiner Nachbarschaft durchgeführt, was schließlich das entsprechende Pixel im Ergebnisbild darstellt. Das Ergebnisbild wird dabei häufig auch als <i>gefaltetes</i> Bild bezeichnet. Die Werte in der Filtermatrix, oft auch Gewichte genannt, entscheiden darüber, welche Effekte im Ergebnisbild hervorgerufen werden. Dies

kann beispielsweise eine Weichzeichnung, Schärfung oder auch die Extraktion von Merkmalen wie bspw. Kanten sein.

Pooling:

Unterabtastung eines gefalteten Bildes. Hierbei wird das (gefaltete) Bild verkleinert, indem relevante Informationen des Bildes zusammengefasst werden. Auf diese Weise werden überschüssige Informationen im Bild verworfen, was zu einer besseren Generalisierfähigkeit des Neuronalen Netz führt. Außerdem führt die Reduzierung der Bildgröße zu einer Verringerung des Speicherbedarfs und des Rechenaufwands.