

Automatische Extraktion und Aktualisierung von Strassenverkehrsinfrastrukturdaten aus cloudbasierten 3D-Bilddiensten mittels Deep Learning

Mit Mobile Mapping Systemen ist es möglich, Strassenkorridore innert kurzer Zeit digital zu erfassen. Die Interpretation der 3D Bilder mit der Extraktion von strukturierten Daten ist aber mit grossem manuellen Aufwand verbunden. In der vorliegenden Masterarbeit werden zwei Faltungsnetzwerke im Bereich der künstlichen Intelligenz trainiert, um die Aufgabe der Erfassung von Objekten im Strassenbereich zu übernehmen. Die Resultate entsprechen mit der kleinen Anzahl an Trainingsdaten nicht den Erfolgen von anderen Arbeiten. Für bestimmte Klassen kann der entwickelte Algorithmus aber erfolgreich angewendet werden.

Motivation

Objekte der Strasseninfrastruktur werden in GIS verwaltet. Die Nachfrage nach ständig aktuellen Daten führt zu einem hohen Aufwand um die Objekte im Feld oder aus Mobile Mapping Daten manuell zu erfassen. Deep Learning, als Teilbereich der künstlichen Intelligenz, zeigt grosses Potential, Bildbereiche mit gesuchten Objekten automatisch zu detektieren. In Untersuchungen mit 10 Klassen und 50'000 Testbildern werden Klassifikationsergebnisse von bis zu 97 % erreicht (Gastaldi, 2017). In der Masterarbeit sollen erste Untersuchungen für eine automatische Klassifizierung von Infrastrukturen gemacht werden. Zur Verfügung stehen 3D-Bilder und GIS Daten des Kantons Aargaus.

Region Proposal

Um Trainingsdaten für die überwachte Klassifikation zu erhalten, werden die Elemente aus dem GIS in den Bildraum gerechnet. Die berechneten Bildregionen müssen aufgrund verschiedener Faktoren auf ihre Tauglichkeit überprüft werden. Rund 50% der Bildregionen eignen sich aufgrund diverser Gründe für das Training nicht. Mit dem High-Level Modul 'Keras' für die Programmiersprache Python wird das Faltungsnetzwerk aufgesetzt. Für den Vorschlag der Bildregionen wird eine *Faster R-CNN* Architektur nach Ren et al. (2017) gewählt. Als Validierungsdatensatz werden Bildbereiche verwendet, welche mit einem selbsterstellten GUI in den Bildern um die gesuchten Objekte definiert wurden. Tabelle 1 zeigt die erreichten Resultate. Mit dem *RetinaNet*, einer alternativen Netzarchitektur nach Lin et al. (2017) und speziell angepassten Konfigurationsparametern wird versucht, die Resultate für die Detektion von Schächten zu verbessern. Die Resultate sind in der Tabelle 2 und der Abbildung 2 ersichtlich.

	Genauigkeit [%]	Trefferquote [%]
Kandelaber	63.6	85.4
Fussgängerstreifen	100.0	71.9
Einlaufschacht	50.9	48.3
Kontrollschacht	88.2	19.2

Tab. 1: Resultate der Faster R-CNN Architektur

Von der Bildregion zur Welt-Koordinate

Die Zentren der detektierten Regionen werden in den Objektraum gerechnet. Die dazu notwendigen Distanzen werden den Disparitätskarten entnommen, welche den Aufnahmen des Mobile Mappings hinterlegt sind. Da in einer Bildsequenz jedes Objekt mehrmals registriert wird, ergeben sich mehrere Punkte für jedes Objekt. Die Streuung dieser Punkte ist stark abhängig von der Grösse des Objektes in der Wirklichkeit. Diese Punkte können im Objektraum zusätzlich über die Geometrie und Topologie bereinigt werden. Eine Möglichkeit ist beispielsweise die Überprüfung der Höhe gegenüber der Fahrbahn.



Abb. 1: Unterschiedliche Streuung der detektierten Bildregionen im Welt-Koordinatensystem. Die Daten sind noch nicht bereinigt.

Fazit

Aktuelle Resultate der Forschung werden mit den erstellten Algorithmen nicht erreicht. Es stehen dafür zu wenige Trainingsdaten zur Verfügung. Zudem sind nicht alle Objekte für eine automatische Detektion geeignet. Beispielsweise werden Kandelaber häufig mit Hausecken oder Baumstämmen verwechselt. Das Neuronale Netzwerk liefert bei Fussgängerstreifen bereits eine Genauigkeit von 100%. Die Qualität der Resultate kann mit dem RetinaNet zusätzlich gesteigert werden.

	Genauigkeit [%]	Trefferquote [%]
Einlaufschacht	69.7	51.7
Kontrollschacht	54.4	86.2

Tab. 2: Resultate der RetinaNet Architektur



Abb. 2: Beispiele der erkannten Objekte. (a) und (b) mit dem Faster R-CNN. (c) und (d) mit dem RetinaNet.

Gastaldi, X. 2017, Shake-Shake regularization of 3-branch residual. ICLR 2017 submission. URL: <https://openreview.net/pdf?id=HkO-PCmYI>. Stand: 03.01.2018.

Lin, T. et al. 2017, Focal Loss for Dense Object Detection. arXiv preprint, arXiv: 1708.02002.

Ren, S. et al. 2017, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 6.

