

Dissertation

**Detection and Segmentation of Bulk Material Heaps
in Remote Sensing Imagery
using Deep-Learning-based Computer Vision**

Erkennung und Segmentierung von Schüttguthaufwerken
in Fernerkundungsbilddaten
mittels Deep-Learning-basierten Maschinellen Sehens

Submitted by

Andreas Michael Ellinger

Abstract

This thesis proposes a novel approach to detect and segment individual bulk material heaps in remote sensing imagery using deep-learning-based computer vision.

Bulk materials like sand and gravel constitute the second-most used resource on earth (after water) [1]. On the large open storage areas of construction sites, coal mines and material handling facilities, bulk material is usually stored in the form of *heaps* or *stockpiles*. Acquiring the number, position and size of bulk materials heaps is an essential prerequisite for inventory measurement, quantity take-off and process monitoring in the construction, mining and material handling industry.

The approach proposed in this thesis is based on the fact that bulk material heaps do not contain significant overhangs. Therefore, the relevant surfaces of a given 3D scan can be represented as multi-modal 2.5D raster images encoding both surface color and surface topography. This in turn enables the usage of well-established computer vision algorithms for instance segmentation like *Mask R-CNN*.

The proposed method consists of a processing pipeline that (1) derives an occlusion-free 2.5D multi-modal raster image from a 3D true-color surface scan, (2) divides the derived raster image into partially overlapping chunks, (3) sequentially detects and segments bulk material heap instances in each individual chunk using computer vision, (4) merges the resulting instance masks iteratively based on spatial overlap and (5) derives the volume per heap based on the merged instance masks and the 2.5D surface model.

The concept is evaluated using a software prototype and real-world data. The research focus lies on the influence of different data representations and processing configurations. The central research questions are:

(1) If combining color and topography features is beneficial when fine-tuning a model for the detection and segmentation of bulk material heaps, specifically given a convolutional neural network pre-trained on a large-scale open-source data set of RGB images. Also, it is examined to what extent precomputed slope features can support training with small training datasets. A central aspect of these questions is how the size of the domain-specific training set influences the relative performance of different feature combinations and representations.

(2) The influence of the three main parameters that control the chunk-based processing mechanism, specifically the chunk size, the overlap between chunk, and the size of an “inner region” per chunk outside which detected instances are discarded.

The practical validation demonstrates that the proposed method is able to detect and segment individual bulk material heaps in a large variety of scenarios with varying heap shapes and sizes, in different weather and lighting conditions and in a variety of surroundings. Bottom-up orthographic projection provides a simple and efficient method to derive occlusion-free 2.5D raster images of the relevant surfaces from 3D surface scans. Processing the input in overlapping chunks using a sliding-window approach allows to pro-

cess arbitrarily large sites without having to sacrifice resolution. Detected instances can be merged consistently across multiple chunks based on spatial overlap.

Evaluating different data representations and processing configurations led to the following findings:

(1) Given the amount of domain-specific training data available, the highest accuracy is achieved when using only the relative height features as input to the instance segmentation model. Evaluating the results for training sets of different sizes suggest that the benefits of combining color and height features increase with increasing training set size. The overall trend suggests that a combination of color and height features might outperform the height-only scenario once the training set becomes large enough. However, there is not enough training data available to conclusively support or reject this hypothesis.

(2) Regarding manual feature engineering, the results indicate that pre-computed slope features do not improve the detection of individual heaps. However, the results do suggest a slightly improvement regarding the semantic segmentation of heap-covered areas (without distinguishing between individual heaps), particularly if the training set is very small. A possible explanation for this is that while slope features clearly distinguish heap surfaces from the surrounding ground, they may be less suited to distinguish individual heaps that directly border each other.

(3) Regarding key control-parameters of the sliding-window-based processing method, comparing different window sizes indicates that the highest accuracy is achieved when the window size is about four times the average heap size (in terms of diameter) and approximately twice as large as the 95th percentile. Overlap between window positions generally tends to improve the overall accuracy, since the resulting reduction in false negatives outweighs the slight increase in false positives. Considering only predictions inside a pre-defined central region at each position of the sliding window drastically reduces false positives due to missing context at the windows edges.

When evaluated on real-world data from eight different material handling sites (containing 239 heaps in total), using optimized feature representation- and processing configurations, the method achieves a precision of 0.88, a recall of 0.77, a respective F1 score of 0.81 and an average volumetric accuracy among true-positives of 0.88.

While a fully automated monitoring process arguably requires an even higher accuracy, this is primarily a matter of acquiring more training data. The achieved Accuracy is sufficient to deploy the proposed method efficiently in a semi-automatic human-in-the-loop workflow, where the system improves over time without the need for separately collected and manually labeled training data.

Kurzfassung

Diese Arbeit schlägt eine neuartige Methode vor, um Schüttgut-Haufwerke in Fernerkundungs-Bilddaten mittels auf Deep-Learning basierenden Computer-Vision-Algorithmen zu erkennen und zu segmentieren.

Schüttgut-Materialien wie Sand und Kies stellen die nach Wasser weltweit meist genutzte Ressource dar [1]. Auf Baustellen, im Tagebau und auf den weitläufigen Lagerflächen der Baustoffindustrie wird Schüttgut in der Regel in Form von Haufwerken gelagert. Die Anzahl, Position und Größe dieser Haufwerke zu ermitteln ist ein zentraler Bestandteil von Inventurmaßnahmen, Mengenermittlungen und dem Prozess-Monitoring in der Bau-, Tagebau-, und Baustoff-Industrie.

Der in dieser Arbeit vorgeschlagene Ansatz beruht darauf, dass Schüttgut-Haufwerke keine relevanten Überhänge enthalten. Daher können alle relevanten Oberflächen eines 3D Oberflächenscans in einem multi-modalen 2.5D Rasterbild abgebildet werden, was wiederum die Verwendung von etablierten Computer Vision Algorithmen wie *Mask R-CNN* für die Instanz-Segmentierung ermöglicht.

Die vorgeschlagene Methode besteht aus einer Verarbeitungs-Pipeline die (1) ein okklusionsfreies multi-modales 2.5D Oberflächenmodell von einem 3D Echtfarb-Oberflächenscan ableitet, (2) das abgeleitete Rasterbild in partiell überlappende Segmente unterteilt, (3) in diesen Segmenten mittels Computer Vision individuelle Haufwerks-Instanzen erkennt, (4) die resultierenden Instanz-Masken ggf. iterative basierend auf Überlappung zusammenfügt und (5) anhand der zusammengeführten Instanz-Masken und dem 2.5 Oberflächen-Modell das Volumen je Haufwerk ermittelt.

Das Konzept wird anhand eines Software-Prototypen und Echtwelt-Daten validiert und bewertet. Der Forschungsschwerpunkt liegt dabei auf der Auswirkung verschiedener Merkmals-Repräsentationen und Verarbeitungsconfigurationen. Die zentralen Forschungsfragen sind:

(1) Ob und in wie fern eine Kombination von Echtfarb- und Topografie-Merkmalen für das Fine-Tuning eins auf großen Echtfarb-Bilddatensätzen vortrainierten Faltenden Neuronalen Netzes für die Erkennung und Segmentierung von Schüttgut-Haufwerken von Vorteil ist und ob im Voraus berechneter Steigungs-Merkmale das Training bei kleinen Trainingsdatensätzen unterstützen können. Ein zentraler Aspekt dieser Frage besteht darin, inwiefern die Größe des für das Fine-Tuning verwendeten domain-spezifischen Trainingsdatensatzes die relative Performance der unterschiedlichen Merkmals-Kombinationen und -Repräsentationen beeinflusst.

(2) Welchen Einfluss die drei Haupt-Parameter, die die Prozessierung in partiell überlappenden Segmenten kontrollieren, auf das Ergebnis haben; namentlich (a) die Segmentgröße, (b) die Überlappung zwischen Segmenten, sowie (c) die Größe des „inneren Bereichs“ per Segment, außerhalb dessen erkannte Haufwerks-Instanzen nicht berücksichtigt werden.

Die praktische Validierung zeigt, dass die vorgeschlagene Methode in der Lage ist, einzelne Haufwerke in vielfältigen Szenarien mit unterschiedlichen Haufwerks-Formen und -Größen, unter verschiedenen Wetter- und Lichtbedingungen sowie in unterschiedlichen Umgebungen zu erkennen und zu segmentieren. Die invertierte orthographische Projektion bietet eine einfache und effiziente Methode, um okklusionsfreie 2,5D-Rasterbilder der relevanten Oberflächen aus 3D-Oberflächenscans abzuleiten. Die Verarbeitung der Eingabedaten in überlappenden Segmenten mittels eines Gleitfenster-Ansatzes ermöglicht es, beliebig große Areale zu verarbeiten, ohne Auflösung einzubüßen. Erkannte Instanzen können auf Grundlage räumlicher Überlappungen konsistent über mehrere Segmente hinweg zusammengefügt werden.

Die Analyse verschiedener Merkmals-Repräsentationen und Verarbeitungs Konfigurationen führte zu folgenden Erkenntnissen:

(1) Für die verfügbare Menge domain-spezifischer Trainingsdaten wird die höchste Genauigkeit erzielt, wenn ausschließlich die relativen Höhen-Merkmale als Eingangswerte für das Instanz-Segmentierungs-Modell verwendet werden. Eine Auswertung der Ergebnisse für unterschiedlich große Trainingsdatensätze legt nahe, dass der Nutzen einer Kombination von Farb- und Höhen-Merkmalen mit zunehmender Trainingsdatenmenge zunimmt. Der Gesamttrend deutet darauf hin, dass eine Kombination aus Farb- und Höhenmerkmalen das reine Höhen-Szenario übertreffen könnte, wenn der Trainingsdatensatz hinreichend groß ist. Allerdings stehen nicht genügend Trainingsdaten zur Verfügung, um diese Hypothese eindeutig zu stützen oder zu verwerfen.

(2) In Hinblick auf manuelles Feature Engineering weisen die Ergebnisse darauf hin, dass die Verwendung im Voraus abgeleiteter Steigungs-Merkmale die Erkennung individueller Haufwerke nicht verbessert. Die Ergebnisse deuten jedoch auf eine geringfügig verbesserte Erkennung der von Haufwerken bedeckten Gesamtfläche hin (ohne dabei zwischen einzelnen Haufwerken zu unterscheiden), insbesondere wenn der Trainingsdatensatz sehr klein ist. Eine mögliche Erklärung hierfür besteht darin, dass Steigungs-Merkmale Haufwerke zwar klar von dem umgebenden Boden abgrenzen, dabei jedoch weniger geeignet sind um direkt aneinander grenzenden Haufwerke gegeneinander abzugrenzen.

(3) In Bezug auf die Haupt Kontroll-Parameter der Gleitfenster-basierten Verarbeitungsmethode zeigen die Ergebnisse, dass die höchste Genauigkeit erzielt wird, wenn die Fenstergröße etwa dem Vierfachen des durchschnittlichen Haufwerks-Durchmessers und ungefähr dem Doppelten des 95. Perzentils entspricht. Eine Überlappung zwischen den Fenster-Positionen trägt im Allgemeinen zur Verbesserung der Gesamtgenauigkeit bei, da die dadurch erzielte Verringerung der Falsch-Negativ-Erkennungen den leichten Anstieg der Falsch-Positiv-Erkennungen überwiegt. Werden ausschließlich erkannte Instanzen innerhalb eines vordefinierten zentralen Bereichs des Gleitfensters berücksichtigt, reduziert dies die Zahl der Falsch-Positiv-Erkennungen erheblich, die andernfalls durch fehlenden Kontext an den Fenster-rändern entstehen würden.

Die Method erzielt einen Precision-Score von 0.88, einen Recall-Score von 0.77, einen entsprechenden F1-Score von 0.81 und eine auf das Volumen je Haufen bezogene durchschnittliche Genauigkeit von 0.88; ausgewertet auf Daten von acht Standorten (mit ins-

gesamt 239 Haufwerken) und unter Verwendung einer optimierter Merkmals-Repräsentation und Verarbeitungskonfiguration.

Auch wenn ein vollständig automatisierter Monitoring-Prozess wohl eine noch höhere Genauigkeit erfordert, ist dies in erster Linie eine Frage der Verfügbarkeit zusätzlicher Trainingsdaten. Die erzielte Genauigkeit erscheint ausreichend, um die vorgeschlagene Methode effizient in einem halbautomatischen *Human-in-the-Loop*-Workflow einzusetzen, bei dem sich das System im Laufe der Zeit verbessert, ohne dass separat erhobene und manuell annotierte Trainingsdaten benötigt werden.

Keywords: Computer Vision, Instance Segmentation
Remote Sensing, UAV Monitoring
Bulk Material Heaps, Stockpiles
Inventory Measurements, Material Flow Monitoring

Author: Andreas Michael Ellinger

Title of Thesis: Detection and Segmentation of Bulk Material Heaps
in Remote Sensing Imagery
using Deep-Learning-based Computer Vision

Institute: Technische Universität Dresden
Faculty of Civil Engineering
Institute for Construction Informatics

Bibliographic Information: 2025, 275 (324) pages, 177 (220) figures, 18 tables