

# Instanzsegmentierung von stehendem Totholz mittels Mask R-CNN und CIR-Bildern

SEBASTIAN BRIECHLE<sup>1</sup>

*Zusammenfassung: Die Kartierung von stehendem Totholz auf Einzelbaumebene wird – u.a. aufgrund der sich ändernden klimatischen Bedingungen – zu einem immer bedeutenderen Aspekt im Zuge der Waldinventur. Jedoch fehlen bisher leistungsstarke Methoden, die auf Basis von flächenhaft zur Verfügung stehenden Fernerkundungsdaten präzise Ergebnisse auf großen Waldflächen liefern können. In dieser Arbeit wird gezeigt, dass die Deep Learning-basierte Instanzsegmentierung auf diese Problemstellung angepasst werden kann. Hierbei konnte auf Basis von 20 cm CIR-Orthophotos aus dem Nationalpark Bayerischer Wald (530 Bäume/ha) das neuronale Netz Mask R-CNN trainiert werden, das in der Lage ist, einzelne stehende tote Bäume zu detektieren und zudem die Umrisse der Baumkronen zu liefern. Des Weiteren wurde die Generalisierungsfähigkeit des trainierten Modelles anhand von Testdaten mit unterschiedlich starker Ähnlichkeit zum Trainingsdatensatz evaluiert. Generell betrug die Inferenzzeit für 1 ha große Bildkacheln 50 ms (5 s pro km<sup>2</sup>) und untermauert das enorme Potenzial dieser Methode für die großflächige Totholzkartierung in der Praxis.*

## 1 Einleitung

### 1.1 Motivation

Weltweit leiden die Wälder unter sich ändernden klimatischen Bedingungen, starken Winden und einem zunehmenden Schädlingsbefall. Vor diesem Hintergrund hat die flächendeckende Kartierung von abgestorbenen Bäumen mit Hilfe von Fernerkundungsdaten in den letzten Jahren immer mehr an Bedeutung gewonnen. In der Praxis haben großflächige Methoden, die vollautomatisch und zuverlässig Ergebnisse auf der Skala einzelner Bäume liefern können, ein großes Potenzial. Einerseits ist eine genaue quantitative Bewertung der ökologischen Situation von großem Interesse. Andererseits ist es für die von massivem Waldsterben betroffenen Waldbewirtschaftenden und deren Versicherungen wichtig, den wirtschaftlichen Schaden schnell, möglichst genau und vollautomatisch abzuschätzen. Noch fehlt es jedoch an robusten, zuverlässigen und zugleich genauen Methoden zur Totholzkartierung – insbesondere in natürlichen und damit komplexen Wäldern.

### 1.2 Totholzkartierung

Die Literaturrecherche zur Totholzkartierung mit Fernerkundungsdaten zeigt, dass die meisten Ansätze auf Baumebene entweder auf Lidardaten oder Multi-(MS)/Hyperspektralbildern, oder einer Kombination aus beiden, beruhen. Lidar-basierte Methoden zur Totholzkartierung sind in der Regel zweistufig. In einem ersten Schritt erfolgt die Abgrenzung von Einzelbäumen – entweder mittels *canopy height model* (CHM) (PYYSALO & HYYPPÄ 2002) oder punktwolken-basiert (REITBERGER et al. 2009). In einem zweiten Schritt werden die Einzelbäume als „tot“ oder „lebendig“

---

<sup>1</sup> Hochschule München, Fakultät für Geoinformation, Labor für Photogrammetrie und Fernerkundung, Karlstraße 6, D-80333 München, E-Mail: sebastian.briechle@hm.edu

klassifiziert. Hier werden häufig manuell generierte Merkmale und gängige Methoden des maschinellen Lernens (z.B. *random forest*, *support vector machine*) verwendet (KAMINSKA et al. 2018; KRZYSZEK et al. 2020). In letzter Zeit haben tiefe neuronale Netze (DNNs) diese Methoden weitgehend ersetzt (HAMRAZ et al. 2019; BRIECHLE et al. 2020), da sie in der Lage sind, charakteristische Merkmale automatisch aus den Rohdaten zu extrahieren. Im Gegensatz zu Lidar-basierten Methoden machen sich rein bildbasierte Ansätze die Tatsache zunutze, dass sich Totholz insbesondere in *color-infrared* (CIR)-Bildern deutlich von der lebenden Vegetation abhebt. CIR-Orthophotos werden flächendeckend durch die Landesämter bereitgestellt, sie sind vergleichsweise günstig und haben zudem eine hohe geometrische Auflösung (meist 20 cm). Jüngste Veröffentlichungen haben gezeigt, dass *convolutional neural networks* (CNNs) aus dem Computer-Vision-Bereich erfolgreich für Orthophotos aus dem Forstbereich verwendet werden können. SYLVAIN et al. (2019) zeigten beispielsweise das Potenzial von CNNs zur Kartierung des Gesundheitszustandes von Bäumen („lebend“ bzw. „tot“). Unter Verwendung von 20 cm-Luftbildern erzielten die Autoren eine Testgenauigkeit auf Pixelebene von bis zu 94%.

### 1.3 Instanzsegmentierung

Neben DNNs für die Bildklassifizierung, semantische Segmentierung und Objekterkennung sind Modelle, die eine Instanzsegmentierung durchführen, von zunehmendem Interesse. Der größte Vorteil dieser Modelle für forstwirtschaftliche Anwendungen liegt auf der Hand: Sie ermöglichen die Abgrenzung und Klassifizierung einzelner Baumkronen in einer durchgängigen Pipeline. In letzter Zeit wurden einige Arbeiten veröffentlicht, die Mask R-CNN (HE et al. 2017) zur Detektion und Klassifizierung von Einzelbäumen verwendeten. BRAGA et al. (2020) nutzten Mask R-CNN zur Erkennung von Baumkronen in sehr hochauflösenden Satellitenbildern von tropischen Wäldern ( $F_1$ -Score = 0,86). Im selben Jahr setzten CHADWICK et al. (2020) Mask R-CNN auf UAV-basierten RGB-Bildern (GSD = 3 cm) ein ( $F_1$ -Score = 0,91). HAO et al. (2021) trainierten Mask R-CNN zur Erkennung von Baumkronen in hochauflösenden UAV-basierten MS-Bildern einer Plantage in China. Ein Modell mit einer Kombination aus *normalized difference vegetation index* (NDVI) und CHM erzielte die besten Ergebnisse ( $F_1$ -Score = 0,85).

Im Gegensatz zur Klassifizierung von Baumarten wurde über die DL-basierte Instanzsegmentierung zur Totholzkartierung bisher nur sehr wenig berichtet. Daher wird hier nur auf eine Veröffentlichung verwiesen: POLEWSKI et al. (2021) stellten ein zweistufiges Verfahren zur Instanzsegmentierung einzelner umgestürzter Bäume vor. Auf Basis einer semantischen Segmentierung, die mittels U-Net (RONNEBERGER et al. 2015) erstellt wurde, wird ein aktives Konturverfolgungsverfahren genutzt. Unter Verwendung von hochauflösenden CIR-Orthophotos erreichten die Autoren einen  $F_1$ -Score von bis zu 0,87 für Testflächen im Nationalpark Bayerischer Wald (NPBW).

### 1.4 Kernidee

Nach bestem Wissen des Autors wurde Mask R-CNN noch nie zur Kartierung einzelner stehender toter Bäume in einem natürlichen und damit anspruchsvollen Waldgebiet verwendet. Daher besteht der Hauptbeitrag dieser Arbeit darin, das Potenzial der Anwendung von Mask R-CNN auf diesen Praxisfall hinsichtlich Genauigkeit und Flächenleistung zu untersuchen.

## 2 Materialien und Methoden

### 2.1 Untersuchungsgebiet

Das Untersuchungsgebiet für die folgenden Experimente befindet sich im NPBW nahe der deutsch-tschechischen Grenze und umfasst etwa 530 Bäume pro ha. Der NPBW wurde 1970 eingerichtet und die dominierenden Baumarten sind Fichten (*Picea albies*), Rotbuchen (*Fagus sylvatica*), Weißtannen (*Abies alba*) und Lärchen (*Larix*). Darüber hinaus haben natürliche Belastungen – schwere Stürme, Borkenkäferbefall und lange Trockenperioden – zu großen Mengen an stehendem Totholz geführt.

### 2.2 Datensatz und Vorverarbeitung

Im Juni 2016 wurden CIR-Luftbilder mit einer Leica DMC III Kamera (Brennweite 92 mm) aufgenommen (LATIFI et al. 2021). Unter Verwendung eines hochaufgelösten Lidar-basierten digitalen Oberflächenmodelles (DOM) wurden anschließend 20 cm CIR-Orthophotos erstellt. In einem nächsten Schritt wurde das Orthomosaik in Kacheln der Größe  $512 \times 512$  Pixel aufgeteilt. Damit entspricht die Bildgröße in etwa 1 ha.

Anschließend wurden Referenzdaten auf der Grundlage einer visuellen Interpretation manuell annotiert. In der vorliegenden Arbeit wurden abgestorbene Bäume mit Hilfe des Open-Source-Tools LabelMe (RUSSELL et al. 2007) als Kreise unterschiedlicher Größe markiert. Schließlich wurden die Bildkacheln zufällig in Trainings-, Validierungs- und Testdatensätze sortiert (siehe Tabelle 1).

Tab. 1: Übersicht über die Größe des Trainings-, Validierungs- und Testdatensatzes.

| Datensatz  | Anzahl der Bilder | Anzahl der Instanzen | Fläche [km <sup>2</sup> ] |
|------------|-------------------|----------------------|---------------------------|
| train      | 131 (49%)         | 681 (43%)            | 1,37                      |
| validation | 52 (19%)          | 486 (31%)            | 0,55                      |
| test       | 86 (32%)          | 407 (26%)            | 0,90                      |
| $\Sigma$   | 269               | 1574                 | 2,82                      |

### 2.3 Instanzsegmentierung mit Mask R-CNN

Die eigentliche Totholzkartierung wurde mit Detectron2 (WU et al. 2019), einer beliebten PyTorch-basierten Bibliothek durchgeführt. Detectron2 ermöglicht die Einbindung von State-of-the-Art-Technologien in den Workflow und umfasst Modelle wie Faster R-CNN, Mask R-CNN, RetinaNet und DensePose. Um die Detectron2-Pipeline nutzen zu können, muss in einem Vorverarbeitungsschritt ein benutzerdefinierter COCO-Datensatz aus den oben genannten Bildkacheln und Beschriftungen erstellt werden. Dies kann mit Hilfe der Funktion `register_coco_instances()` geschehen.

Als Modell wurde Mask R-CNN mit einem ResNet-50-Backbone zur Merkmalsextraktion und einem Feature Pyramid Network gewählt (`mask_rcnn_R_50_FPN_3x.yaml`), das *bounding boxes* und detaillierte Umrisse einzelner Baumkronen liefert. Tabelle 2 zeigt einige der wichtigsten Hyperparameter und ihre gewählten Werte. Der Code wurde als Jupyter-Notebook auf Google Colab ([colab.research.google.com](https://colab.research.google.com)) unter Verwendung von Cuda 10.1 und PyTorch 1.7, sowie einer NVIDIA® Tesla T4 GPU (2.560 CUDA® Cores, 16 GB GPU-Speicher) ausgeführt.

Tab. 2: Übersicht über die wichtigsten Parameter von Mask R-CNN in Detectron2.

|        | Parameter                             | Bedeutung   | gesetzte Werte   |
|--------|---------------------------------------|---|------------------|
| model  | <i>backbone.freeze_at</i>             | einzufrierende ResNet-50-Blöcke (je größer, desto weniger Parameter werden neu trainiert) | 1, 2, 3          |
|        | <i>anchor_generator.sizes</i>         | Größe der Anker in [pix]  | {8 16 32 64 128} |
|        | <i>anchor_generator.aspect_ratios</i> | Seitenverhältnisse der Anker  | {0,75 1,00 1,33} |
| solver | <i>ims_per_batch</i>                  | Anzahl der Bilder pro Stapel  | 8                |
|        | <i>base_lr</i>                        | initiale Lernrate   | 0,001            |
| eval   | <i>eval_period</i>                    | Intervall für die Validierung   | 50               |

Eines der wichtigsten Ziele beim Training eines neuronalen Netzes ist es, eine Überanpassung (*overfitting*) an die Trainingsdaten zu vermeiden. In der Praxis bedeutet dies, den Trainingsprozess im richtigen Moment zu stoppen. Andernfalls wird das Modell zu sehr für Vorhersagen auf den Trainingsdaten optimiert und verliert seine Generalisierungsfähigkeit auf neuen Daten. Daher wurde ein Abbruchkriterium (*early stopping*) eingeführt, das den *validation loss* überwacht (*patience* = 10), um sicherzustellen, dass das beste Modell gespeichert wird. Um das trainierte Modell so robust wie möglich gegenüber geometrischen und radiometrischen Veränderungen in den Eingangsbildern zu machen und damit die Übertragbarkeit auf neue Datensätze zu erhöhen, wurden die Daten während des Trainings künstlich angereichert (*data augmentation*). Einerseits wurden die Helligkeit, der Kontrast und die Sättigung der Bilder zufällig verändert, um die radiometrischen Eigenschaften zu variieren. Andererseits wurden die Bilder zufälligen geometrischen Transformationen unterzogen (horizontales und vertikales Spiegeln, Drehen und Zuschneiden).

Unter Verwendung des trainierten Modelles wurden Vorhersagen auf unabhängigen Testdatensätzen getroffen und mit den Referenzdaten verglichen. Zur quantitativen Bewertung wurde hierbei der sogenannte AP50-Wert ermittelt. AP50 ist die durchschnittliche Präzision bei einem *intersection over union* (IoU)-Schwellenwert von 50% – eine weit verbreitete Metrik im Bereich der Objekterkennung. Die Metrik wird sowohl für die Erkennung von einzelnen toten Baumobjekten (*AP50\_bbox*), als auch die Trennung von Hintergrund und Objekt (*AP50\_seg*) berechnet. Zur qualitativen Analyse wurden die Vorhersagen auf zufällig ausgewählten Testbildern mit den entsprechenden Eingabebildern überlagert und mit den Referenzdaten verglichen.

## 2.4 Generalisierung

Eine der größten Schwierigkeiten im Zusammenhang mit neuronalen Netzen ist es, einen hohen Grad an Generalisierung und damit eine hohe Übertragbarkeit auf neue Datensätze zu erreichen. Um das trainierte Netz hinsichtlich dieser Fähigkeit beurteilen zu können, wurden zwei weitere CIR-Testdatensätze aus dem NPBW generiert (siehe Tabelle 4). Diese weisen zwar dieselbe Vorverarbeitung und geometrische Auflösung auf wie der bereits angesprochene Trainingsdatensatz (20 cm), unterscheiden sich von diesem jedoch unterschiedlich stark in Bezug auf den Bildinhalt. Die Unterteilung in die Datensätze *test\_similar* und *test\_tough* erfolgte visuell mittels manueller Sortierung. Die anschließende quantitative und qualitative Evaluierung wurde identisch zur Evaluierung des Testdatensatzes durchgeführt (siehe Kapitel 2.3).

### 3 Ergebnisse und Diskussion

#### 3.1 Flächenleistung

Der Erfolg eines automatisierten Ansatzes in der Praxis hängt nicht nur von der Genauigkeit der Vorhersagen ab, sondern vor allem auch von der erreichbaren Flächenleistung. Unter Verwendung eines Jupyter-Notebooks auf Google Colab betrug die Trainingszeit pro Epoche 15 s und dementsprechend 72 min für ein komplettes Training (300 Epochen). Die Inferenzzeit – also die Zeit, die das trainierte Modell für die Vorhersage auf Bildkacheln der Größe von 1 ha benötigt – betrug 50 ms. Dies verdeutlicht die hohe Skalierbarkeit des Ansatzes für die großflächige Kartierung von Totholz: Eine Fläche von 1 km<sup>2</sup> kann in etwa 5 s kartiert werden, die gesamte Fläche des NPBW (248,5 km<sup>2</sup>) in weniger als einer halben Stunde (21 Minuten).

#### 3.2 Quantitative Auswertung

Mit Hilfe der oben vorgestellten Pipeline wurden verschiedene Experimente durchgeführt, um ein optimiertes Mask R-CNN-Modell zu erhalten. Als Auswahlkriterium diente der niedrigste Wert für den *validation loss*, was ein übliches Vorgehen ist. Auf den 86 unabhängigen Testbildern (mit 407 toten Bäumen und einer Fläche von 90 ha) erreichte das beste Modell einen *AP50\_bbox* Wert von 72,6 und einen *AP50\_seg* Wert von 73,0 (siehe Tabelle 3).

Generell ist festzustellen, dass die besten Ergebnisse mit einem Wert für *backbone.freeze\_at* von 2 erzielt wurden, was der Standardeinstellung in Detectron2 entspricht. Darüber hinaus konnten schrittweise Verbesserungen mit einer höheren Intensität der geometrischen (*geom*) und radiometrischen (*radiom*) Datenanreicherung erreicht werden. Der Zusatz *heavy* bedeutet zum einen, dass hier die Intervalle der Parameter für die Datenanreicherung größer als in den anderen Experimenten eingestellt wurden – außerdem wurden die Eingabebilder während des Trainings zufällig zugeschnitten.

Tab. 3: Ergebnis der Instanzsegmentierung auf dem Testdatensatz (86 Kacheln, 407 tote Bäume) für eine unterschiedlich hohe Anzahl an geschätzten Parametern im Zuge der Merkmalsextraktion und eine unterschiedlich starke Intensität der Datenanreicherung während des Trainings; bestes Ergebnis in fett.

| <b>backbone.freeze_at</b> | <b>Datenanreicherung</b> | <b>AP50_bbox</b> | <b>AP50_seg</b> |
|---------------------------|--------------------------|------------------|-----------------|
| 1                         | geom                     | 67,0             | 65,9            |
| 1                         | geom, radiom             | 68,4             | 69,4            |
| 2 (default)               | geom                     | 66,8             | 69,0            |
| 2 (default)               | geom, radiom             | 68,6             | 71,5            |
| 2 (default)               | geom_heavy, radiom       | <b>72,6</b>      | <b>73,0</b>     |
| 3                         | geom                     | 61,1             | 59,1            |
| 3                         | geom, radiom             | 61,5             | 60,3            |

#### 3.3 Qualitative Auswertung

Zur qualitativen Analyse des besten Modelles zeigt Abbildung 1 beispielhafte Bildkacheln (links), Referenzdaten (Mitte) und Vorhersagen (rechts). Offensichtlich wurden die meisten abgestorbenen Bäume von Mask R-CNN erfolgreich erkannt und umrissen. Obwohl kreisförmige Masken für die Annotationen verwendet wurden, war das DNN in der Lage, die Umrisse der einzelnen Bäume zu lernen. Die visuelle Interpretation zeigt auch, dass die Größe der Baumkronen in den

meisten Fällen korrekt geschätzt werden konnte. Diese Tatsache eröffnet die Möglichkeit, in zukünftigen Experimenten Kronendurchmesser aus den Polygonflächen abzuschätzen.

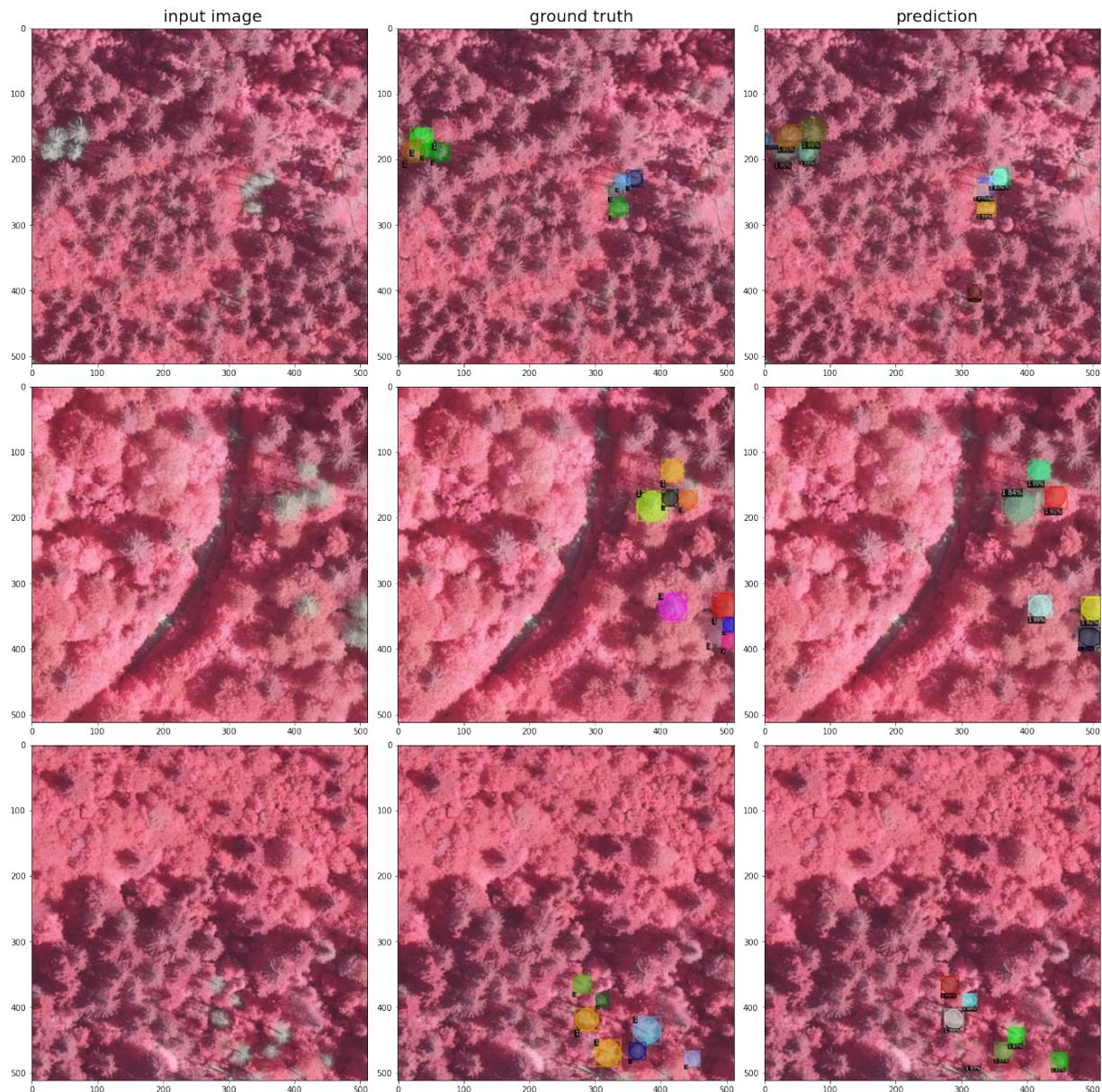


Abb. 1: Beispiele für eine erfolgreiche Totholzkartierung in verschiedenen Szenarien unter Verwendung von Mask R-CNN; Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

Dennoch zeigen einige Beispiele von Fehlklassifikationen, dass der vorgestellte Ansatz noch Verbesserungspotenzial hat. In einzelnen Fällen produzierte das DNN falsch-positive Ergebnisse bei teilweise beschatteten Straßen, bei denen das CIR-Bild geometrische und radiometrische Strukturen aufweist, die denen eines toten Baumes sehr ähnlich sind (Abbildung 2, obere Reihe). Hier könnte die nachträgliche Filterung mittels Wegepolygonen bzw. die Integration der relativen Höheninformation (nDOM) Abhilfe schaffen.

Darüber hinaus traten vor allem an den Bildrändern vermeintliche falsch-positive Fehlklassifizierungen auf (Abbildung 2, untere Reihe). Hierbei handelt es sich um offensichtliche Fehler in den Referenzdaten – das DNN traf eigentlich eine korrekte Vorhersage. Eine Lösung für das Problem der abgeschnittenen Objekte an den Kachelrändern und der daraus resultierenden Fehler bei der Erkennung liegt auf der Hand: Bei der Erstellung der Eingangsdaten sollte eine ausreichend große Überlappung zwischen benachbarten Kacheln sichergestellt werden. Anschließend könnten die einzelnen abgeschnittenen Polygone aufgrund der geringeren Fläche im Vergleich zum komplett erfassten Einzelbaum in einem Nachbearbeitungsschritt gefiltert werden.

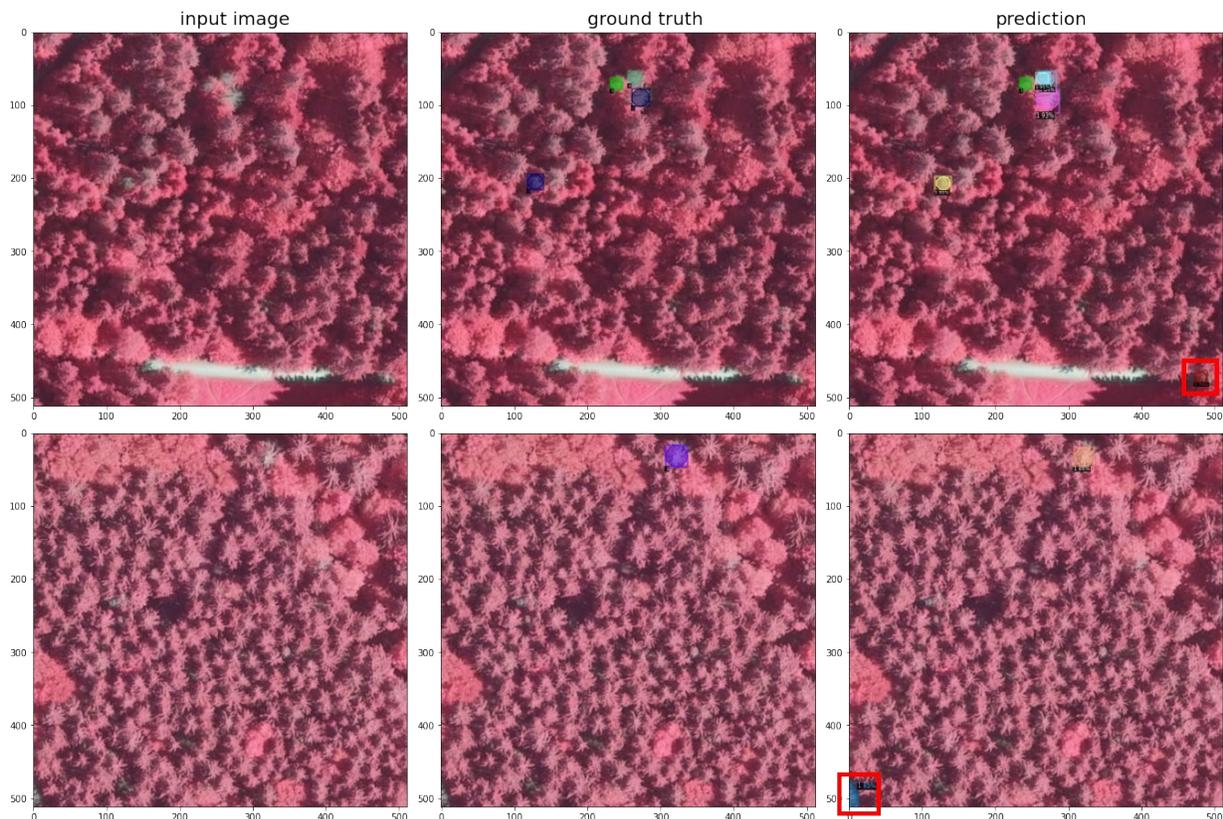


Abb. 2: Auftretende Fehlklassifizierungen aufgrund (i) einer teilweise schattierten Straße (obere Reihe) und (ii) eines fehlenden Labels in den Referenzdaten am Bildrand (untere Reihe); rotes Quadrat markiert Fehlklassifizierungen; Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

### 3.4 Übertragbarkeit

Die Fähigkeit des trainierten Modelles zur Generalisierung – und damit zur Übertragbarkeit auf neue Datensätze – wurde anhand von zwei weiteren Testdatensätzen aus dem NPBW überprüft, deren Bildinhalte eine unterschiedlich starke Ähnlichkeit mit denen des Trainingsdatensatzes aufweisen (siehe Kapitel 2.4). Dabei wurde das Modell ausgewählt, das auf den ursprünglichen Testdaten die besten Ergebnisse lieferte ( $AP50_{bbox} = 72,6$ ;  $AP50_{segm} = 73,0$ ; siehe Tabelle 3). Wie zu erwarten war, zeigt die quantitative Auswertung ein unterschiedliches Bild für die beiden Datensätze *test\_similar* und *test\_tough*.

Auf Szenen mit einer Komplexität, die mit der des Trainingsdatensatzes vergleichbar ist (*test\_similar*), konnten zufriedenstellende Ergebnisse erzielt werden (siehe Tabelle 4 und Abbildung 3

oben). Hier waren die Metriken für die Detektion lediglich um 5,2 (*AP50\_bbox*) bzw. für die Maskierung um 3,6 (*AP50\_seg*) niedriger als auf dem 90 ha großen Testdatensatz.

Auf dem Testdatensatz mit von den Trainingsdaten deutlich abweichendem Bildinhalt (*test\_tough*; siehe Abbildung 3 unten) zeigt sich, dass das trainierte neuronale Netz Schwierigkeiten hat, korrekte Vorhersagen im Rahmen der Totholzerkennung zu treffen (siehe Tabelle 4). Die entsprechenden Metriken liegen hier ca. 16 bis 18 Punkte unter dem Ergebnis, das auf dem Datensatz *test\_similar* erzielt wurde. Die Ergebnisse zeigen also deutlich, dass in zukünftigen Experimenten die Diversität des Trainingsdatensatzes hinsichtlich der Bildinhalte erhöht werden sollte, um die Übertragbarkeit zu steigern.

Tab. 4: Ergebnisse hinsichtlich der Fähigkeit zur Generalisierung des trainierten Modelles auf Datensätze mit unterschiedlich stark von den Trainingsdaten abweichenden Bildinhalten.

| Datensatz           | Anzahl der Bilder | Anzahl der Instanzen | Fläche [ha] | AP50_bbox | AP50_seg |
|---------------------|-------------------|----------------------|-------------|-----------|----------|
| <i>test_similar</i> | 16                | 142                  | 17          | 67,4      | 69,4     |
| <i>test_tough</i>   | 37                | 489                  | 39          | 49,5      | 52,6     |

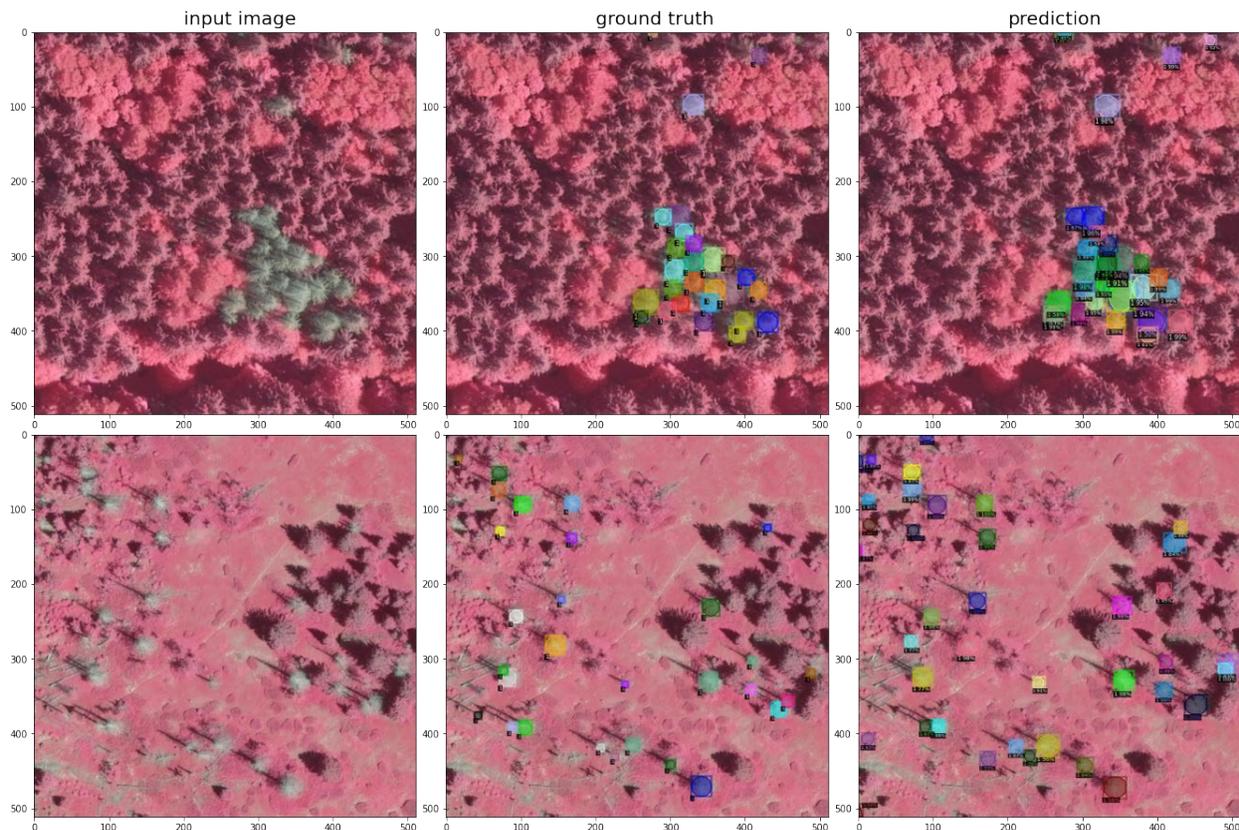


Abb. 3: Ergebnis der Instanzsegmentierung auf je einem repräsentativen Beispielbild der Datensätze *test\_similar* (1. Zeile) und *test\_tough* (2. Zeile); Eingabebild (links), Referenzdaten (Mitte), Vorhersage (rechts).

## 4 Fazit und Ausblick

In dieser Arbeit wurde Mask R-CNN – ein tiefes neuronales Netz zur Instanzsegmentierung – zur Totholzkartierung auf Einzelbaumebene verwendet. Die Wirksamkeit des vorgestellten Ansatzes wurde anhand eines 20 cm CIR-Orthomosaiks aus einem natürlichen Waldgebiet (530 Bäume/ha) demonstriert. Das beste Modell erreichte auf unabhängigen Testdaten einen *AP50\_bbox* Wert von 72,6 und einen *AP50\_segm* Wert von 73,0. Insgesamt bestätigen die Ergebnisse das große Potenzial für die Kartierung einzelner toter Bäume mit Mask R-CNN. Außerdem war die Leistung von Mask R-CNN in Detectron2 praxistauglich: Auf einer NVIDIA® Tesla T4 GPU (16 GB) betrug die Trainingszeit für den vorgestellten Datensatz 72 Minuten (300 Epochen) und die Inferenzzeit für 1 ha große Bildkacheln von 50 ms. Damit ist der vorgestellte Ansatz prinzipiell für die großflächige Kartierung von Totholz geeignet: Ein Gebiet von 1 km<sup>2</sup> kann in 5 s kartiert werden, der komplette NPBW (248,5 km<sup>2</sup>) in 21 min. Somit können flächendeckende und zuverlässige Karten für die Endnutzer erstellt werden, die für Anwendungen wie die automatisierte Waldinventur und Monitoringprojekte von großer Bedeutung sind. Für die Zukunft sind die folgenden Verbesserungen geplant:

- Generierung verbesserter, erweiterter und vielfältigerer Referenzdaten, indem man sich auf professionelle Felddaten und/oder Annotationen von Forstexpert:innen stützt;
- Integration von Wegedaten und/oder der relativen Höheninformation (nDOM), um die Anzahl der Fehlklassifizierungen im Postprocessing zu reduzieren und damit die Ergebnisse zu verbessern;
- Durchführung detaillierter Studien zur Übertragbarkeit auf weitere Untersuchungsgebiete (u.a. Frankenwald, Schwarzwald, Harz) und Eingangsdaten wie z. B. CIR-Orthophotos der Landesämter oder hochauflösende Satellitenbilder;

Grundsätzlich könnte die vorgestellte Methode auch auf die kombinierte Klassifizierung von Totholz und einzelnen Baumarten ausgeweitet werden. Allerdings sollte aufgrund der hohen Komplexität dieser Aufgabe eine höhere Anzahl von MS-Kanälen und ein Lidar-basiertes nDOM in die Pipeline aufgenommen werden.

## 5 Literaturverzeichnis

- BRAGA, G., PERIPATO, V., DALAGNOL, R., FERREIRA, M.P., TARABALKA, Y., ARAGÃO, L., DE CAMPOS VELHO, H.F., SHIGUEMORI, E.H. & WAGNER, F.H., 2020: Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sensing*, **12**, 1288, <https://doi.org/10.3390/rs12081288>.
- BRIECHLE, S., KRZYSZEK, P. & VOSSELMAN, G., 2020: Classification of tree species and standing dead trees by fusing UAV-based lidar data and multispectral imagery in the 3D deep neural network PointNet++. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, **V-2-2020**, 203-210, <https://doi.org/10.5194/isprs-annals-V-2-2020-203-2020>.
- CHADWICK, A.J., GOODBODY, T., COOPS, N., HERVIEUX, A., BATER, C.W., MARTENS, L.A., WHITE, B. & RÖESER, D., 2020: Automatic Delineation and Height Measurement of Regenerating Conifer Crowns under Leaf-Off Conditions Using UAV Imagery. *Remote Sensing*, **12**, <https://doi.org/10.3390/rs12244104>.

- HAMRAZ, H., JACOBS, N., CONTRERAS, M. & CLARK, C., 2019: Deep learning for conifer/deciduous classification of airborne lidar 3D point clouds representing individual trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, **158**, 219-230, <https://doi.org/10.1016/j.isprsjprs.2019.10.011>.
- HAO, Z., LIN, L., POST, C.J., MIKHAILOVA, E.A., LI, M., CHEN, Y., YU, K. & LIU, J., 2021: Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 112-123, <https://doi.org/10.1016/j.isprsjprs.2021.06.003>.
- HE, K., GKIOXARI, G., DOLLAR, P. & GIRSHICK, R., 2017: Mask R-CNN. In: *Proceedings of CVPR*, 2980-2988, <https://doi.org/10.1109/ICCV.2017.322>.
- KAMINSKA, A., LISIEWICZ, M., STERENCZAK, K., KRASZEWSKI, B. & SADKOWSKI, R., 2018: Species-related single dead tree detection using multi-temporal ALS data and CIR imagery. *Remote Sensing of the Environment*, **219**, 31-43, <https://doi.org/10.1016/j.rse.2018.10.005>.
- KRZYSZEK, P., SEREBRYANYK, A., SCHNÖRR, C., ČERVENKA, J. & HEURICH, M., 2020: Large-scale mapping of tree species and dead trees in Šumava National Park and Bavarian Forest National Park using lidar and multispectral imagery. *Remote Sensing*, **12**, <https://doi.org/10.3390/rs12040661>.
- LATIFI, H., HOLZWARTH, S., SKIDMORE, A., BRÚNA, J., ČERVENKA, J., DARVISHZADEH, R., HAIS, M., HEIDEN, U., HOMOLOVÁ, L., KRZYSZEK, P., SCHNEIDER, T., STARÝ, M., WANG, T., MÜLLER, J. & HEURICH, M., 2021: A laboratory for conceiving Essential Biodiversity Variables (EBVs)—The ‘Data pool initiative for the Bohemian Forest Ecosystem’. *Methods in ecology and evolution*, **12**(11), 2073-2083, <https://doi.org/10.1111/2041-210X.13695>.
- POLEWSKI, P., SHELTON, J., YAO, W. & HEURICH, M., 2021: Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 297-313, <https://doi.org/10.1016/j.isprsjprs.2021.06.016>.
- PYYSALO, U. & HYYPPÄ, H., 2002: Reconstructing tree crowns from laser scanner data for feature extraction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **34**.
- REITBERGER, J., SCHNÖRR, C., KRZYSZEK, P. & STILLA, U., 2009: 3D segmentation of single trees exploiting full waveform lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 561-574, <https://doi.org/10.1016/j.isprsjprs.2009.04.002>.
- RONNEBERGER O., FISCHER P. & BROX T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, **9351**, Springer, Cham, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- RUSSELL, B., TORRALBA, A., MURPHY, K. & FREEMAN, W.T., 2007: LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 2007.
- SYLVAIN, J.-D., DROLET, G. & BROWN, N., 2019: Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS Journal of Photogrammetry and Remote Sensing*, **156**, 14-26, <https://doi.org/10.1016/j.isprsjprs.2019.07.010>.
- WU, Y., KIRILLOV, A., MASSA, F., LO, W.-Y. & GIRSHICK, R., 2019: Detectron2. <https://github.com/facebookresearch/detectron2>.