Kapitel 1

Konzept

1.1 Normalized Difference Vegetation Index

Es gibt eine starke Korrelation zwischen dem physiologischen Status einer Pflanze und deren Chlorophyllgehalt. Faktoren wie Krankheit, Dürre oder Umweltverschmutzung haben einen negativen Einfluss auf den Chlorophyllspiegel.[10] Messungen haben ergeben, dass es einen Verbindung zwischen dem Reflexionsgrad im nahen Infrarotbereich und im Rotbereich und dem Chlorophyllgehalt gibt. Das heißt, dass eine gesunde, adulte Pflanze im nahen Infrarotbereich stärker reflektiert als zum Beispiel eine pathologisch veränderte Pflanze. Jedoch bleibt die Reflexion im roten Lichtspektrum in beiden Fällen vergleichsweise schwach. Andere vegetationsfreie Oberflächen wie Acker, Straßen oder Wasser strahlen auch im nahen Infrarotbereich schwach zurück. Dadurch ergibt sich eine zerstörungsfreie Methode, mit einer Multispektralkamera die Vitalität („Grünheit“) einer oder mehrerer Pflanzen zu bestimmen.[8]

Eine multispektralen Aufnahme kann mithilfe der Formel

$$NDVI = \frac{Band_{NIR} - Band_{Red}}{Band_{NIR} + Band_{Red}} \tag{1.1}$$

dazu genutzt werden, den *Normalized Difference Vegetation Index* (NDVI) zu berechnen. Wobei $Band_{NIR}$ der nahe Infrarotbereich (Near Infrared) und $Band_{RED}$ der sichtbare rote Bereich des elektromagnetischen Spektrums ist. Der NDVI gibt quantifizierte Werte im Bereich von -1 bis 1 zurück. Dabei deuten Werte, die kleiner als 0 sind, auf Wasserobflächen hin. 0 bedeutet keine Vegetation. Bei Werte nahe 0 handelt es sich um spärliche oder ungesunde Vegetation. Das bedeutet je näher ein Wert an 1 ist, desto dichter bewachsen und gesünder ist die beobachtete Vegetationsfläche.[13] Dass bei einem niedrigen, positiven NDVI nicht unterschieden

werden kann, ob eine Fläche kaum bewachsen ist oder ungesunde Vegetation besitzt, kann hier vernachlässigt werden. Das Gebiet, das in dieser Arbeit untersucht wird, ist ein bewachsenes Feld, so kann man geringe Vegetation ausschließen.

1.2 Sentinel-2

Die Sentinel-2-Satelliten sind eine von sechs Satellitenarten (Sentinel-1 bis -6) des Copernicus-Programms[1], die zur Erdbeobachtung in einen 786 km hohen sonnensynchronen Orbit gebracht wurden. Die Instrumente der Sentinel-2-Satelliten können Aufnahmen in Bereichen des roten und nahen Infrarot- bis hin zum Kurzwelleninfrarotspektrums. Die Aufnahmen haben Gesamtgröße von $100 * 100$ km und je nach Band eine von Auflösung von 10m, 20m oder 60m (s. Tabelle 1.1).

Bandnummer	Räumliche Auflösung	Mittlere Wellenlänge (nm)	Bandbreite (nm)
B1	60	443,9	27
B2	10	496,6	98
B3	10	560	45
B4	10	664,5	38
B5	20	703,9	19
B6	20	740,2	18
B7	20	782,5	28
B8	10	835,1	145
B8a	20	864,8	33
B9	60	945	26
B10	60	1373,5	75
B11	20	1613,7	143
B12	20	2202,4	242

Tabelle 1.1: Räumliche und spektrale Auflösungen von Sentinel-2A[6]

Besonders wichtig sind die Bänder B4 (Rot) und B8 (Nahes Infrarot). Mit diesen Bändern kann der NDVI (s. Kapitel 1.1) berechnet werden.[5] Die Sentinel-2-Satelliten bieten mit $10 * 10$ m pro Pixel eine hohe räumliche Auflösung.[2] Diese Eigenschaft ist wichtig, um eine mögliche Infizierung genau eingrenzen zu können.

[1]Das Copernicus-Programm wurde von der Europäischen Union zur Erdbeobachtung ins Leben gerufen. Die gesammelten Daten werden für wissenschaftliche, wirtschaftliche und private Anwendungszwecke zur Verfügung gestellt.[3]

[2]Im Vergleich hat zum Beispiel der Landsat-8-Satellit, dessen Daten ebenfalls frei verfügbar sind, eine relativ geringe Auflösung von $30 * 30$ m.[15]

Dabei ist es auch wichtig, dass die Satelliten regelmäßige Daten liefern können. Durch die gemeinsame Konstellation übertragen die Plattformen alle fünf Tage Daten über einen spezifischen Punkt auf der Erdoberfläche.[7] Damit ist gewährleistet, dass der Feldbesitzer ohne persönliche Inspektion ein bis zweimal in der Woche eine Gesundheitseinschätzung über seine Felder erhält.

1.3 Mask R-CNN

In Kapitel 1.1 und 1.2 wurde erklärt wie Daten über die möglichen Erkrankungen geliefert und verarbeitet werden können. Auf den zugrunde liegenden Bilddaten soll nun ein künstliches neuronales Netzwerk (KNN) trainiert werden. In diesem Kapitel wird darauf eingegangen, welche Anforderungen an das KNN gestellt werden, warum das Titel gebende Netz ausgewählt wurde und wie dieses funktioniert.

1.3.1 Anforderungen

Das KNN muss in der Lage sein, wahrscheinliche Krankheiten in der zu untersuchenden Agrarfläche möglichst genau eingrenzen und klassifizieren zu können. Das ist besonders wichtig, wenn ein Feld von multiplen Krankheiten betroffen ist.

Es ist damit zu rechnen, dass Daten unter bewölkten Bedingungen aufgenommen werden. Nach starken Niederschlägen können Acker teils oder gänzlich überflutet sein.[12] Das sorgt selbst unter wolkenfreien Bedingungen für einen niedrigen NDVI, obwohl die Nutzpflanzen gesund sind. Das neuronale Netz muss mit solchen „Ausreißern“ umgehen können.

Daraus ergeben sich folgende Kriterien für das neuronale Netzwerk:

- Erkennung auf Pixelebene
- Robustheit
- Hohe Genauigkeit

1.3.2 Grundlagen

Vollständig vernetztes neuronales Netz

Künstliche neuronale Netze sind nach dem Vorbild von biologischen neuronalen Netzen gebildet worden. So ist ein KNN ebenfalls eine Verbindung von künstlichen Neuronen. Diese Neuronen sind in Schichten angeordnet und jede die Neuronen einer Schicht sind mit den Neuronen nächsten bzw. letzten Schicht verbunden. Zwischen der ersten und der letzten sog. Ausgangsschicht existieren n versteckte Schichten (engl.: hidden layers).

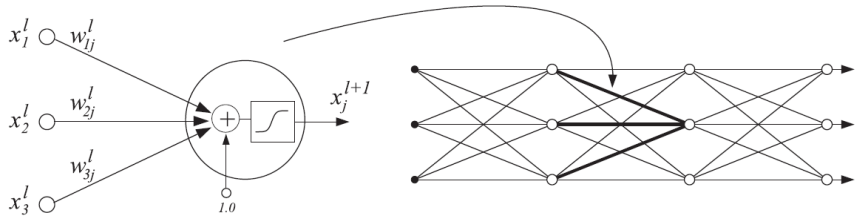

Abbildung 1.1: Künstliches neuronales Netz[16]

Ein Neuron besitzt mehrere Eingangsverbindungen (Gewichte) und ein Ausgangsneuron. Ob ein Neuron „feuert“, wird durch eine lineare oder nicht-lineare Aktivierungsfunktion bestimmt. Die Eingangsgewichte sind veränderbare Werte, die je nach Höhe einen starken oder niedrigen Einfluss auf die Aktivierungsfunktion haben.

$$x_j^{l+1} = f(\sum_i w_{ij}^l x_i^l + w_{bj}^l) \tag{1.2}$$

beschreibt das Neuron j in Schicht $l+1$, wobei

- w_{ij}^l die Gewichte sind, die Neuron i in Schicht l mit Neuron j verbinden.
- w_{bj}^l der Biasterm des j-ten Neurons in Schicht l ist.
- f die Aktivierungsfunktion ist.[16]

Convolutional Neural Networks

Convolutional Neural Networks (CNN, dt.: faltendes neuronales Netzwerk) sind Kategorien von neuronalen Netzen, die besonders in der *Computer Vision* Anwendung finden. In der ersten Schicht werden mehrere Merkmale (engl.: features) durch Filter extrahiert und in separate sog. *Feature Maps* abgelegt, um größere Abstraktionsebenen zu erreichen. Diese Filter sind mathematisch mit Faltungen (engl.: convolutions) zu vergleichen und geben dem Netz den Namen.

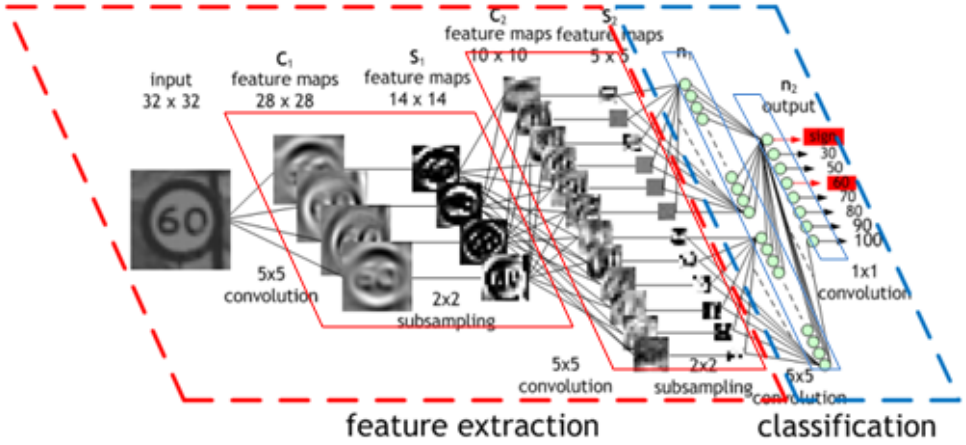

Abbildung 1.2: Architektur eines Convolutional Neural Network[4]

Die Dimensionen der Feature Maps werden in einem Poolingschritt[3] (oder auch *subsampling*) reduziert. Dadurch bleiben nur relevante Informationen erhalten und das CNN wird bis zu einem gewissen Grad robust gegenüber Translationen und Rotationen. In der Regel werden die Faltungen und das das Pooling zwei Mal durchgeführt, wie es in Abb. 1.2 abgebildet ist.

Nach der Merkmalextraktion werden die Feature Maps zur Klassifikation in eine eindimensionale Schichten geglättet. Die folgenden Schichten bis zur Ausgangsschicht sind vollständig vernetzt.

1.3.3 Mask R-CNN

Im Rahmen dieser Arbeit wird das *Mask Region-based Convolutional Neural Network* untersucht. Mask R-CNN ist eine von Facebook AI Research (FAIR) entwickelte Erweiterung des *Faster R-CNN* und kann verschiedene Instanzen einer Klasse in einem Bild von einander trennen. Dazu muss zuerst die Begriffe der Instanzsegmentierung definiert werden.

Einfache Klassifizierung (engl.: classification) ordnet Bilder als Ganzes einer Klasse zu. *Semantische Segmentierung* (engl.: semantic segmentation) beschreibt die Klassifizierung auf Pixelebene. Es wird erkannt zu welcher Klasse eine Menge von Pixeln

[3]Es gibt verschiedene Arten von Pooling (Max, Average, Sum, ...). Dabei wird die $m * m$ px große Feature Map in sich angrenzende $n * n$ px große Felder eingeteilt ($n < m$). Im Falle von Max-Pooling wird der höchste Wert aus dem Feld übernommen.

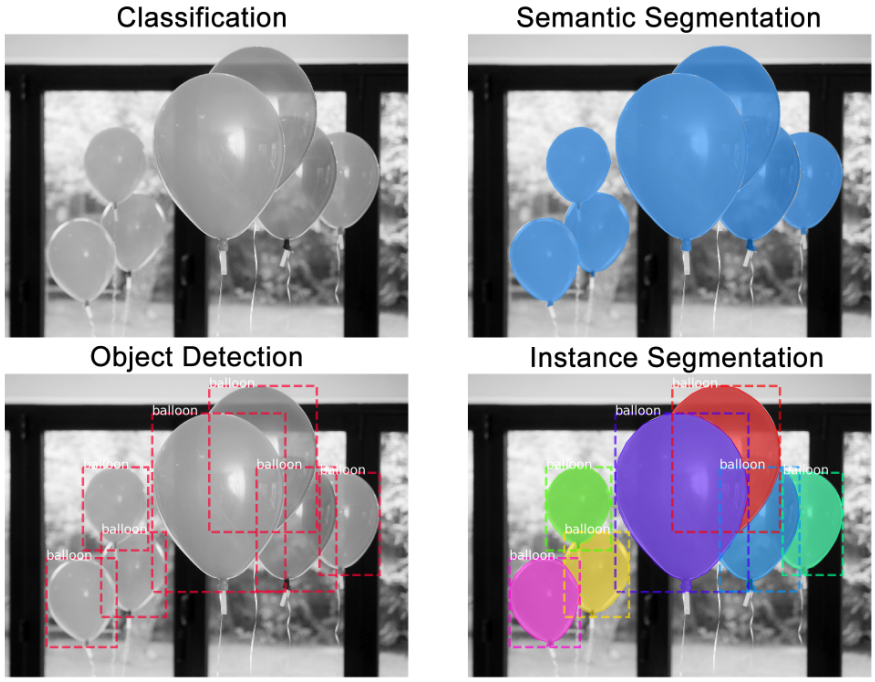

Abbildung 1.3: Unterschied Klassifizierung / semantische Segmentierung / Objekterkennung / Instanzsegmentierung[1]

gehören, aber es wird nicht zwischen einzelnen Objekten unterschieden. *Objekterkennung* (engl.: object detection) entdeckt und lokalisiert unterschiedliche Objekte, indem es eine Bounding Box um jedes erkannte Objekt zieht. Jedoch fehlt hier die pixelgenaue Abgrenzung einzelner Objektinstanzen. Instanzsegmentierung (engl.: instance segmentation) kombiniert *Objekterkennung* und *semantische Segmentierung* und ist so in der Lage zwischen einzelnen Objekten zu unterscheiden und ihnen entsprechende Pixel zuzuordnen (s. Abb. 1.3) und ist eine der größten Herausforderungen in der Bildverarbeitung.[9]

Mask R-CNN ist wie Faster R-CNN in zwei Segmente eingeteilt. In dem ersten Segment, dem *Region Proposal Network* (RPN, dt.: Region vorschlagendes Netzwerk), werden mehrere Rahmen (Bounding Boxes) innerhalb eines Bildes vorgeschlagen, die interessante Objekte beinhalten könnten.[14] Die vorgeschlagene Regionen, die einzeln von CNNs bewertet werden, ist der Kernansatz von R-CNN. Das RPN wurde identisch von Faster R-CNN für Mask R-CNN übernommen.[9]

Im zweiten Segment werden aus den Regionen *Bounding Boxes* (dt.: Rahmen) und Masken generiert und klassifiziert. Die Rahmen haben verschiedene Größen und

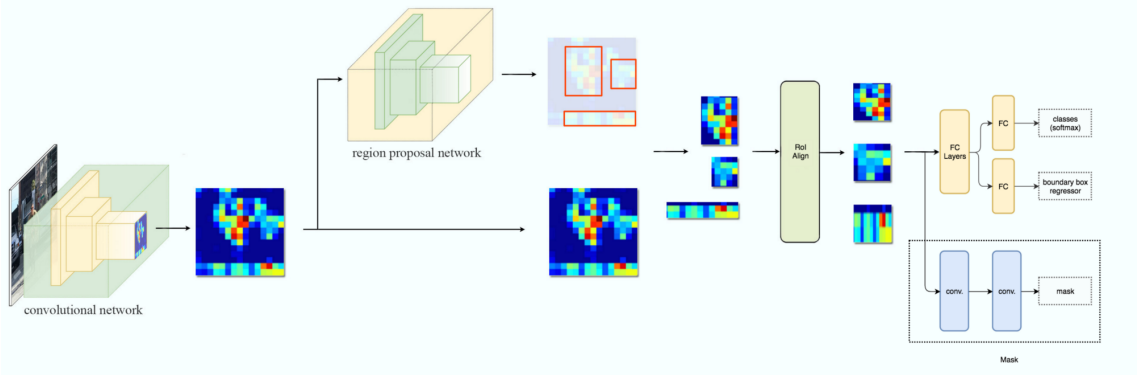

Abbildung 1.4: Mask R-CNN-Architektur[11]

können Probleme bei der Klassifizierung verursachen. Daher werden die Rahmen auf eine kleine Feature Map gleicher Größe (z.B. $7 * 7$ px) reduziert. Die Authoren von [9] schlagen eine Methode namens *RoI-Align* vor, bei der Proben aus der Feature Map entnommen werden und eine bilineare Interpolation angewendet wird. In dem bei Faster R-CNN angewandten Verfahren *RoI-Pooling* entstehen durch Quantisierung Informationsverluste und räumliche Abweichungen zwischen Bounding Box und Feature Map, was negative Auswirkungen auf die Maskengenerierung haben kann.[9]

Die oberen vollständig vernetzten Schichten (*FC Layers* in Abb. 1.4) klassifizieren die Regionen und die Bounding Boxes berechnet. Dieser Zweig ist für die Objekterkennung wichtig und noch mit Faster R-CNN gemeinsam.

Gleichzeitig werden in einem parallelen Zweig je Bounding Box k $m * n$ große Masken zur semantischen Segmentierung erzeugt, wobei k die Anzahl der Klassen ist. Anders als in dem ersten Zweig des zweiten Segmentes werden die Masken durch *fully convolutional networks* (FCN, dt.: vollständig faltende Netzwerke) prognostiziert. Diese bestehen nur aus faltenden Schichten, wie sie in Kapitel 1.3.2 beschrieben sind. Eine Maske ist eine räumliche Kodierung eines Objektes und daher ist es wichtig räumliche Informationen beizubehalten. Diese können durch die Pixel-zu-Pixel-Übereinstimmung extrahiert werden, welche sonst durch vollständig vernetzter Schichten verloren gehen. Diese geben einen Vektor ohne räumliche Dimensionen aus.[9]

FCN und wie wird Maske erzeugt?

In [9] wird Mask R-CNN mit den *COCO challenge*-Gewinnern[4] der Jahre 2015

Definition Back-

[4]COCO (Common Objects in Context, dt.: Gewöhnliche Objekte im Kontext) enthält einen Datensatz von über 200000 Bildern in über 80 Kategorien. Der Datensatz ist eine oft genutzte Basis, um Objekterkennungstechniken zu evaluieren und zu bewerten.[2]

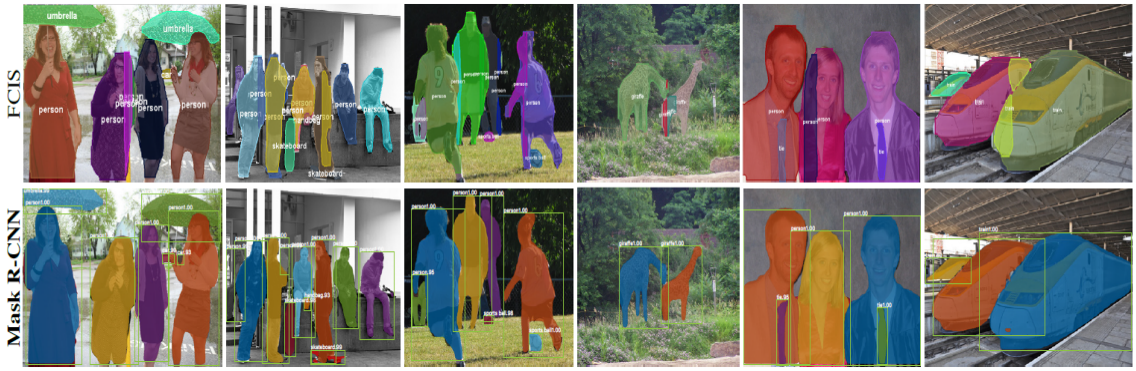

Abbildung 1.5: Bei FCIS entstehen Artefakte, wenn Objekte sich in einem Bild überlappen.[9]

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
MNC [10]	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [26] +OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ [26] +OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	**37.1**	**60.0**	**39.4**	**16.9**	**39.9**	**53.5**

Tabelle 1.2: Instance segmentation *mask* AP auf COCO *test-dev*. MNC und FCIS sind Sieger der COCO 2015 und 2016 Challenge. Mask R-CNN erzielt deutlich bessere Ergebnisse als die komplexere FCIS+++.[9]

und 2016 verglichen. Der Vergleich zeigt, dass Mask R-CNN in der Challenge bessere Werte erzielt als die Konkurenten (s. Tab. Ḋesweiteren fällt *fully convolutional instance segmentation* (FCIS, dt.: vollständig faltende Instanzsegmentierung) auf, wenn es mit überlappenden Objekten konfrontiert wird. Dort erzeugt es Artefakte, welche durch Mask R-CNN nicht entstehen (s. Abb. 1.5). Durch diese Gegenüberstellungen wird gezeigt, dass Mask R-CNN alle aufgeführten Anforderungen erzielt. Es erkennt Klasseninstanzen auf Pixelebene und weist eine hohe Robustheit auf. Auch die Genauigkeit hebt sich beim direkten Vergleich ab. Aus diesen Gründen wurde Mask R-CNN im Rahmen diese Arbeit ausgewählt.

Kapitel 2

Overfitting

2.1 Begriffserklärung

Genaue Daten über Krankheitsbefälle im Agrarsektor sind rar, da diese in der Regel nicht öffentlich zugänglich sind.[1] Daher musste mit *Overfitting* gerechnet werden. Das künstliche neurale Netzwerk soll daraufhin trainiert werden, dass es möglichst alle Befälle, die untersucht werden, erkennt. Dafür wird es im ersten Schritt mit einem Trainingsdatensatz trainiert. Im folgenden Schritt mit einem kleineren Validierungsdatensatz überprüft, wie gut das Netz trainiert wird. Overfitting tritt auf, wenn das Netz auf die Daten aus dem Trainingsdatensatz mit sehr hoher Erfolgsquote erkennt, jedoch vergleichsweise schlechte Ergebnisse bei der Validierung bzw. bei unbekannten Daten erzielt.

In Abb. 2.1 ist ein Beispiel wie Overfitting sich auswirken kann. Die linken zwei Bil-

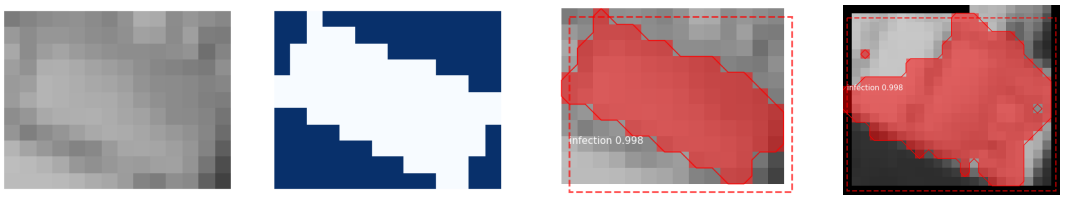

Abbildung 2.1: V.l.n.r. Bild von infizierter Agrarfläche aus Trainingsdatensatz / Binärmaske der inifizierten Region, wird gemeinsam mit dem linken Bild zum Training in das KNN gespeist / Selbiges Bild, Ergebnis nach Trainingsdurchlauf, prognostizierte Ergebnisfläche in rot / Bild der selben Fläche, was nicht aus dem Trainingsdatensatz stammt, prognostizierte Ergebnisfläche in rot

Erwähnen, dass Mask RCNN Bild und Mas-

[1] Datenschutz kann ein Grund dafür sein.

der sind ein exemplarischer Auszug aus dem Trainingsdatensatz. Einmal eine visuelle Repräsentation der NDVI-Werte der infizierten Agrarfläche und die Binärmaske, welche die infizierte Fläche markiert. Das selbe Bild wurde nach einem erfolgreichen Trainingsdurchlauf der Mask R-CNN-Implementierung übergeben und es hat den erkrankten Bereich nahezu perfekt erkannt. Das vierte Bild zeigt zentriert das selbe Feld. Jedoch ist der Ausschnitt größer, rotiert und die Aufnahme stammt von einem anderen Datum. Der Prognose zur Folge ist die Infizierung auf die benachbarten Felder übergesprungen, was nicht der Wahrheit entspricht. Overfitting ist ein bekanntes Problem im Bereich des maschinellen Lernens und es existieren multiple Methoden, um dem entgegenzuwirken.

Genaues Datum nötig?

Literaturverzeichnis

[1] W. Abdulla. Splash of color: Instance segmentation with mask r-cnn and tensorflow. https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238 2018. [Zuletzt besucht: 22.12.2018].

[2] C. Consortium. Coco 2018 object detection task. http://cocodataset.org/#detection-2018, 2018. [Zuletzt besucht: 06.01.2019].

[3] Copernicus. Copernicus in brief. https://www.copernicus.eu/en/about-copernicus/copernicus-brief. [Zuletzt besucht: 15.12.2018].

[4] N. Corporation. Convolutional neural network (cnn). https://developer.nvidia.com/discover/convolutional-neural-network, 2019. [Zuletzt besucht: 04.01.2019].

[5] ESA. Level-2a algorithm overview. https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm, 2018. [Zuletzt besucht: 20.12.2018].

[6] ESA. Radiometric resolutions. https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric, 2018. [Zuletzt besucht: 20.12.2018].

[7] ESA. Resolutions. https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions, 2018. [Zuletzt besucht: 20.12.2018].

[8] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3):271 – 282, 2003.

[9] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.

[10] G. A. F. Hendry, J. D. HOUGHTON, and S. B. BROWN. The degradation of chlorophyll — a biological enigma. *New Phytologist*, 107(2):255–302, 1987.

[11] J. Hui. Image segmentation with mask r-cnn. `https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272`, 2018. [Zuletzt besucht: 06.01.2019].

[12] C. Mattupalli, C. A. Moffet, K. N. Shah, and C. A. Young. Supervised classification of rgb aerial imagery to evaluate the impact of a root rot disease. *Remote Sensing*, 10(6), 2018.

[13] NASA. Measuring vegetation (ndvi & evi). `https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php`, 2000. [Zuletzt besucht: 13.12.2018].

[14] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[15] U.S. Department of the Interior. What are the band designations for the landsat satellites? `https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm`, n.d. [Zuletzt besucht: 24.12.2018].

[16] J. Verrelst, J. Muñoz, L. Alonso, J. Delegido, J. P. Rivera, G. Camps-Valls, and J. Moreno. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and -3. *Remote Sensing of Environment*, 118:127 – 139, 2012.