



UNIVERSITÄT
LEIPZIG

Universität Leipzig
Fakultät für Mathematik und Informatik

Erkennung von erkrankten Nutzpflanzen anhand von Sentinel-2-Multispektralaufnahmen

Abschlussarbeit zur Erlangung des akademischen Grades
Master of Science (M.Sc.)

vorgelegt von

Simon Hüning

Referent:

Prof. Dr. Martin Middendorf

Zusammenfassung

Hier kommt ein Abstrakt hin.

Inhaltsverzeichnis

1 Einleitung	3
1.1 Über die Arbeit	3
1.2 Motivation	4
2 Methoden	7
2.1 Trainingsdaten	7
2.2 Normalized Difference Vegetation Index	8
2.3 Sentinel-2	9
2.4 Das trainierbare Modell	10
2.4.1 Anforderungen	10
2.4.2 Grundlagen	11
2.4.3 Mask R-CNN	12
2.5 Evaluation des Modells	16
2.5.1 Intersection over Union	17
2.5.2 Precision und Recall	18
2.5.3 Average Precision	19
2.5.4 Mean Average Precision	19
3 Overfitting	21
3.1 Begriffserklärung	21
3.2 Data Augmentation	22
3.3 L2 Regularization	23
3.4 Zusätzliche Methoden	24
4 Konzept und Implementierungsdetails	25
4.1 Konzept	25
4.2 Annotation	26
4.3 Suche nach Sentinelprodukten	28
4.4 Aufbereitung der Sentineldaten	30
4.5 Trainings- und Validierungsdatensatz	32

4.6 Training/Detektion	33
4.6.1 Dataset	34
4.6.2 Config	35
4.6.3 Ablauf	37
5 Ergebnis	43
5.1 Training mit Rohdaten	44
5.2 Datensatzerweiterung durch Rotation	45
5.3 Data Augmentation und Regularization	47
5.4 Ergebnisdiskussion	50
6 Fazit	53
7 Ausblick	55
Literaturverzeichnis	56
Abbildungsverzeichnis	60
Tabellenverzeichnis	62
Erklärung	63

Kapitel 1

Einleitung

1.1 Über die Arbeit

Diese Arbeit beschreibt die Konzipierung und Entwicklung einer Anwendung, die es ermöglichen soll erkrankte Nutzpflanzen wie angebautes Getreide und Gemüse früh zu identifizieren. Des weiteren wird die Performanz der Anwendung überprüft und anschließend die Ergebnisse präsentiert.

Dieses Kapitel gibt einen kurzen Überblick über die Arbeit. Es wird erläutert welche Auswirkungen Krankheiten auf die Agrarwirtschaft haben und in Folge dessen wie sich die Motivation der Thematik daraus bildet.

Kapitel 2 geht auf die Herkunft der Realdaten ein und beschreibt wie Satellitendaten über Felder gewonnen und verarbeitet werden können. Es werden Anforderungen an ein künstliches neuronales Netzwerk festgelegt und erörtert, warum das dort beschriebene Netzwerk ausgesucht wurde. Anschließend wird eine Methode definiert, die es ermöglichen soll die späteren Ergebnisse zu evaluieren.

Overfitting ist eine bekannte Herausforderung im Bereich des maschinellen Lernens. Daher wird in Kapitel 3 erklärt, was Overfitting ist und wie man dem entgegenwirken kann.

Kapitel 4 beschreibt die Konzipierung der Prozessschritte der Anwendung und erläutert wichtige Implementationdetails. Die manuelle Annotation der Trainingsdaten, die Aufbereitung der Satellitendaten und das Training des werden aufgezeigt. Ebenfalls wird die Konfiguration des Netzwerkes erklärt.

In Kapitel 5 werden einige Experimente erklärt, die auf Basis der in Kapitel 3 gezeigten Methoden definiert wurden.

Kapitel 6 diskutiert die Ergebnisse der Experimente und formuliert Erkenntnisse, die daraus gewonnen wurden.

Das siebte Kapitel fasst die Arbeit zusammen und gibt ein Fazit an.

Zuletzt wird in Kapitel 7 ein Ausblick auf weitere Forschungen gegeben.

1.2 Motivation

Krankheitserreger sorgen für hohe Ertragsverluste und verurachen dadurch große wirtschaftliche Schäden und erschwert die Nahrungsproduktion für immer weiter wachsende Weltbevölkerung. Die globalen Ernteeinbußen durch Krankheiten und Schädlinge werden auf etwa 30% geschätzt.[29] Zum Beispiel ist Sorghumhirse die fünft-wichtigste Getreideart in der Welt. Jedoch können allein Befälle von Sorghum-Antragnose, ein durch Pilzbefall verursachtes Pathogen, einen Ertragsausfall von bis zu 50% erwirken.[33, S. 77]

Pestizide und andere Chemikalien ermöglichen die Bekämpfung von Befällen.

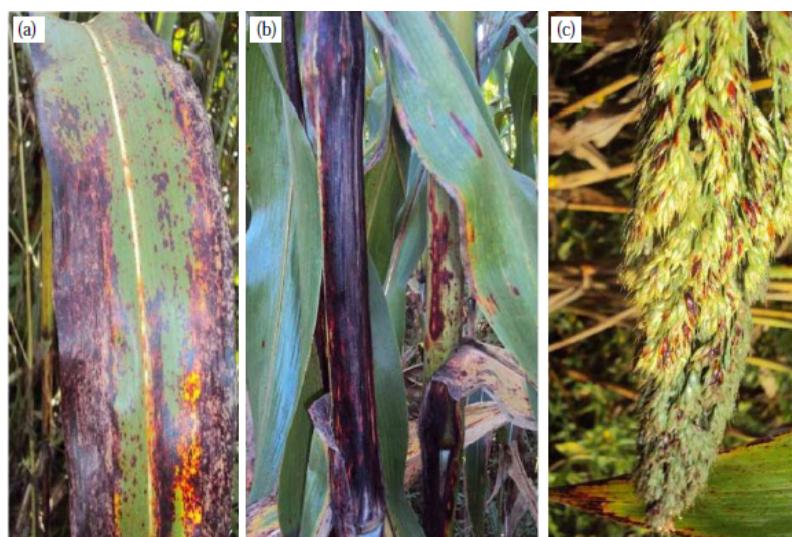


Abbildung 1.1: Sorghum-Anthraknose[30, S. 77]

Jedoch hat ein exzessiver Einsatz von Pestiziden einen negativen Effekt auf

die Umwelt. So sickert es in den Ackerboden und kann das Grundwasser kontaminieren. Daher ist es wichtig, dass Landwirte regelmäßig ihre Felder auf mögliche Schäden untersuchen und diese gezielt eindämmen. Da Landwirte in der Regel großflächige Gebiete betreiben, stellt sich eine manuelle Kontrolle als schwierig dar. Zumal sichtbare Symptome schon Zeichen für einen schweren Befall sein kann.

Bemannte oder unbemannte Fluggeräte können die Überwachung unterstützen und schnelle Einschätzungen über den aktuellen Gesundheitsstatus der Acker geben. So untersuchten Pugh et al. die Fähigkeit einer unbemannten Flugdrohne Anthraknosebefälle in Sorghumfelder einzuschätzen. Sie kamen zu dem Schluss, dass die Methode effektive Analysen zurückgeben kann und durch die Flugeigenschaft eine größere Reichweite hat als traditionelle Fernüberwachungssysteme.[26, S. 861 ff.]

Fluggeräte sind jedoch dadurch eingeschränkt, da sie nur so lange fliegen



Abbildung 1.2: Schematische Ansicht Sentinel-2[15]

können, wie es ihr Energiefüllstand zulässt. Das heißt, dass sie mit Treibstoff versorgt werden müssen, der den Landwirt auch weiter monetär belastet. Raumfahrende Überwachungsinstrumente wie Satelliten, die die Erde in einem hohen Orbit umkreisen, liefern regelmäßige Aufnahmen von nahezu jedem Punkt der Erde. So kann man auf die Aufnahmen der Sentinel-Satelliten des Copernicus-Programms zugreifen und zur eigenen Feldüberwachung nutzen. Chemura, Mutange und Dube zeigten, dass man über Multispektralaufnahmen der Sentinel-2-Plattform unter Laborbedingungen Kaffeerost-Infektionsrate identifizieren kann.[7, S. 877] Dennoch wird erwähnt, dass Feldtesttauglichkeit noch getestet werden muss.[7, S. 859] Diese Arbeit greift diese Idee auf, nutzt dabei aber nicht die Methoden aus dem Paper. Zum einem da Nutzpflan-

zen untersucht werden sollen, die anders als Kaffee im europäischen Raum angebaut werden. Und zum anderen ist ein Ziel dieser Arbeit, ein infiziertes Feld möglichst genau eingrenzen zu können.

Kapitel 2

Methoden

Bevor ein künstliches neuronales Netzwerk trainiert werden kann, müssen Daten vorbereitet werden. In diesem Kapitel wird der Prozess von der Generierung der Trainingsdaten über deren Vorverarbeitung bis hin zum Trainings schritt beschrieben.

2.1 Trainingsdaten



Abbildung 2.1: RGB-Sentinel-2-Aufnahme des zu untersuchenden Ackers. Die infizierten Flächen sind weiß umrandet.

Der infizierte Acker, der die Basis des Datensatzes bildet, befindet sich etwa 15 km nordwestlich von Bologna in Norditalien ($44^{\circ}34'28.92''$ Nord, $11^{\circ}10'21.36''$ Ost). Das Feld hat eine Fläche von $7640,57\text{ m}^2$ und ist mit Sorghum bepflanzt. Mitarbeiter des CREA (Council for Agricultural Research and Economics) haben vor Ort am 12.07.2018 Befäle von Anthracnose und der bakteriellen Streifenkrankheit (med.: Xanthomonas translucens) diagnostiziert. Etwa die Hälfte

der Pflanzen des Feldes sind betroffen. Dabei sind im östlichen Teil des Feldes (Abb. 2.1, innere Markierung) auf einer Fläche von 1043,22 m² von etwa 60 bis 70% der Pflanzen befallen.

2.2 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.[18] Messungen haben ergeben, dass es eine 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.[16]

Eine multispektralen Aufnahme kann mithilfe der Formel

$$NDVI = \frac{Band_{NIR} - Band_{Red}}{Band_{NIR} + Band_{Red}} \quad (2.1)$$

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.[24] 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 bewachsene Feld, so kann man geringe Vegetation ausschließen.

2.3 Sentinel-2

Die Sentinel-2-Satelliten sind eine von sechs Satellitenarten (Sentinel-1 bis -6) des Copernicus-Programms¹, 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 Infrarots 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 2.1).

Bandnummer	Auflösung	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 2.1: Räumliche und spektrale Auflösungen von Sentinel-2A[13]

Besonders wichtig sind die Bänder B4 (Rot) und B8 (Nahe Infrarot). Mit diesen Bändern kann der NDVI (s. Kapitel 2.2) berechnet werden.[12] 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.

Dabei ist es auch wichtig, dass die Satelliten regelmäßige Daten liefern kön-

¹Das 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.[10]

²Im Vergleich hat zum Beispiel der Landsat-8-Satellit, dessen Daten ebenfalls frei verfügbar sind, eine relativ geringe Auflösung von 30 * 30 m.[31]

nen. Durch die gemeinsame Konstellation übertragen die Plattformen alle fünf Tage Daten über einen spezifischen Punkt auf der Erdoberfläche.[14] Damit ist gewährleistet, dass der Feldbesitzer ohne persönliche Inspektion ein bis zweimal in der Woche eine Gesundheitseinschätzung über seine Felder erhält.

2.4 Das trainierbare Modell

In Kapitel 2.2 und 2.3 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.

2.4.1 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.[23] 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

2.4.2 Grundlagen

Vollständig vernetztes neuronales Netz

Künstliche neuronale Netze sind mathematische Modelle, die nach dem Vorbild von biologischen neuronalen Netzen gebildet worden sind. 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).

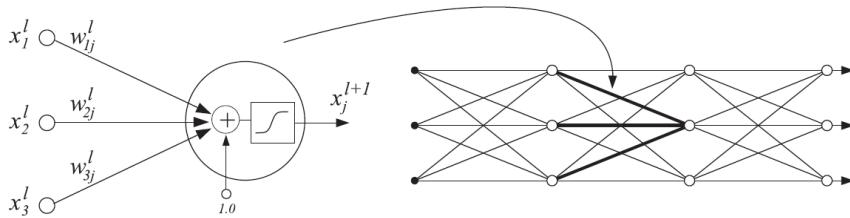


Abbildung 2.2: Künstliches neuronales Netz[32]

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.

$$x_j^{l+1} = f(\sum_i w_{ij}^l x_i^l + w_{bj}^l) \quad (2.2)$$

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.[32]

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,

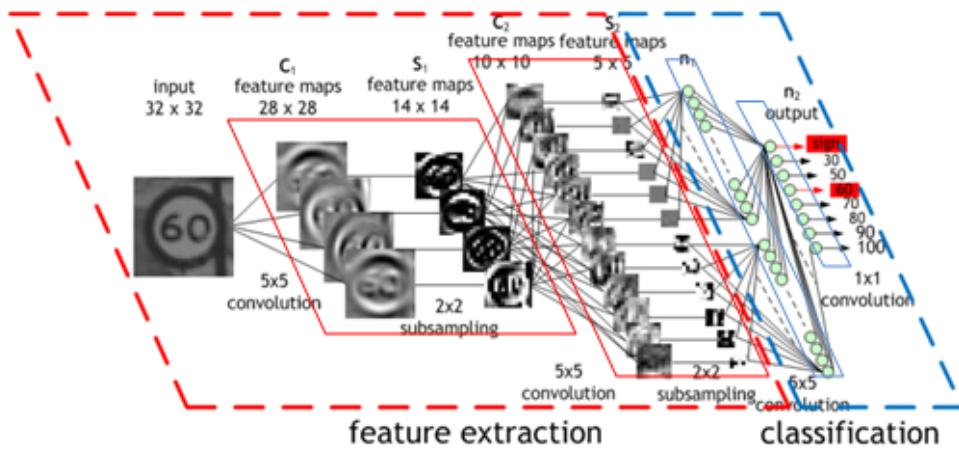


Abbildung 2.3: Architektur eines Convolutional Neural Network[11]

um größere Abstraktionsebenen zu erreichen. Diese Filter sind mathematisch mit Faltungen (engl.: convolutions) zu vergleichen und geben dem Netz den Namen.

Die Dimensionen der Feature Maps werden in einem Poolingschritt³ (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. 2.3 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.

2.4.3 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.

³Es 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.

Einfache Klassifizierung (engl.: classification) ordnet Bilder als Ganzes ei-

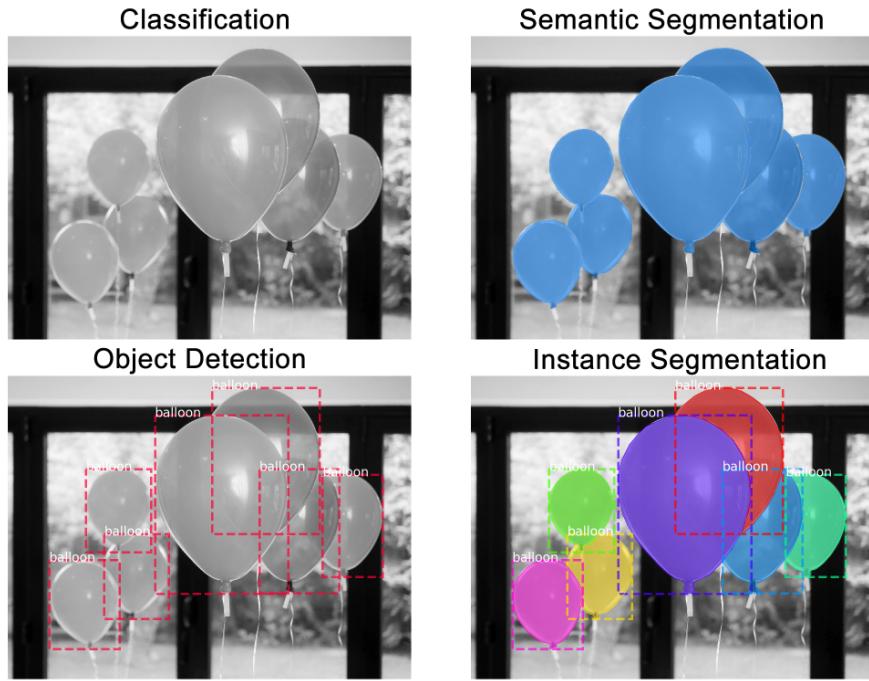


Abbildung 2.4: Unterschied Klassifizierung / semantische Segmentierung / Objekterkennung / Instanzsegmentierung[2]

ner Klasse zu. *Semantische Segmentierung* (engl.: semantic segmentation) beschreibt die Klassifizierung auf Pixelbene. Es wird erkannt zu welcher Klasse eine Menge von Pixeln gehören, aber es wird nicht zwischen einzelnen Objekten unterschieden. *Objekterkennung* (engl.: object detection) entdeckt und lokalsiert 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. 2.4) und ist eine der größten Herausforderungen in der Bildverarbeitung.[17]

Mask R-CNN ist wie Faster R-CNN in zwei Segmente eingeteilt. In dem ersten Segment, dem *Region Proposal Network* (oder auch RPN), werden mehrere Rahmen (engl.: Bounding Boxes) innerhalb eines Bildes vorgeschlagen, die interessante Objekte beinhalten könnten. Das RPN erzeugt Rechtecke - sog. Anker (engl.: Anchors) - von unterschiedlichen Größen und Bildverhältnissen, die sich über die Bildregion verteilen und sich überlappen. Für jeden An-

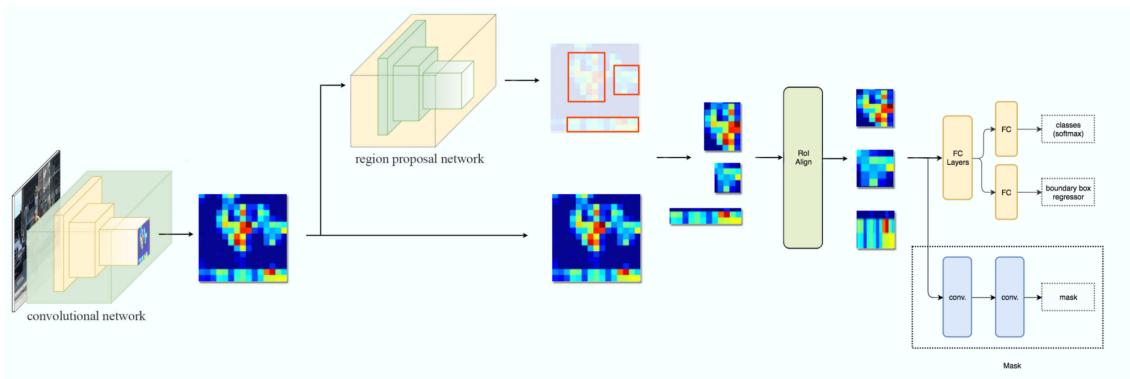


Abbildung 2.5: Mask R-CNN-Architektur[20]

ker wird eine Ankerklasse und eine Bounding-Box-Verfeinerung ausgegeben. Die Klasse unterscheidet Vordergrund und Hintergrund, wobei eine Bounding-Box mit Vordergrundklassifizierung als potentielle Objekterkennung gewertet wird. Ein Anker ist möglicherweise nicht genau über ein Objekt zentriert. Die Verfeinerung ist eine geschätzte Veränderung des Ankers in Position, Höhe und Größe, um besser das Objekt umrahmen zu können. Wenn mehrere Anker sich zu sehr überschneiden, wird der Anker mit der höchsten Wahrscheinlichkeit ein Objekt zu beinhalten übernommen und der restlichen Anker werden verworfen.⁴[2][27] 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.[17]

Im zweiten Segment werden aus den Regionen *Bounding Boxes* (dt.: Rahmen) und Masken generiert und klassifiziert. Die Rahmen haben verschiedene Größen und können Probleme bei der Klassifizierung verursachen. Daher werden die Rahmen auf eine kleine Feature Map gleicher Größe (z.B. $7 * 7$ px) reduziert. Die Autoren von [17] 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.[17]

Die oberen vollständig vernetzten Schichten (*FC Layers* in Abb. 2.5) klassifizieren die Regionen und die Bounding Boxes berechnet. Dieser Zweig ist

⁴Diese Methode wird *Non-max suppression* genannt.

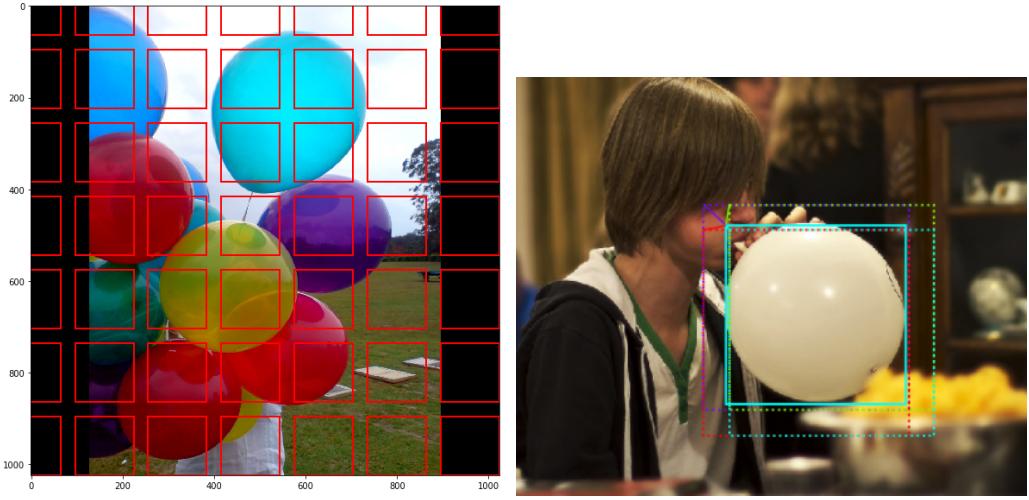


Abbildung 2.6: Links: Vereinfachte Darstellung von Ankern über ein Bild[2] / Rechts: Drei Anker (gepunktet), die das das gleiche Objekt umschließen und die Verfeinerung (durchgezogen), die auf diese angewendet wird, um das Objekt genauer einzugrenzen[2]

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 2.4.2 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.[17]

In [17] wird Mask R-CNN mit den *COCO challenge*-Gewinnern⁵ der Jahre 2015 und 2016 verglichen. Der Vergleich zeigt, dass Mask R-CNN in der Challenge bessere Werte erzielt als die Konkurrenten (s. Tab. Des Weiteren fällt *fully convolutional instance segmentation* (FCIS, dt.: vollständig faltende Instanzsegmentierung) auf, wenn es mit überlappenden Objekten konfrontiert wird.

⁵COCO (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.[8]

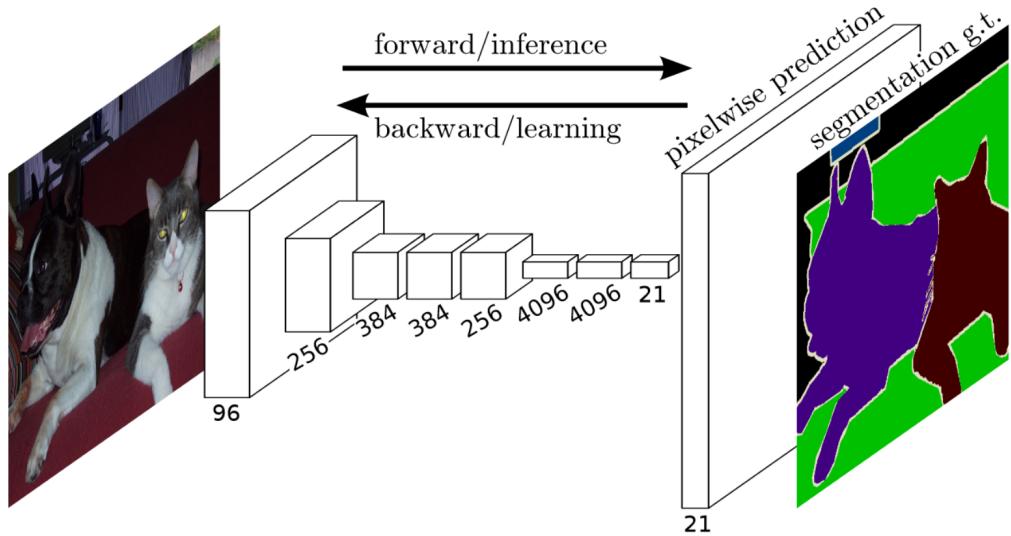


Abbildung 2.7: FCN-Architektur[20]

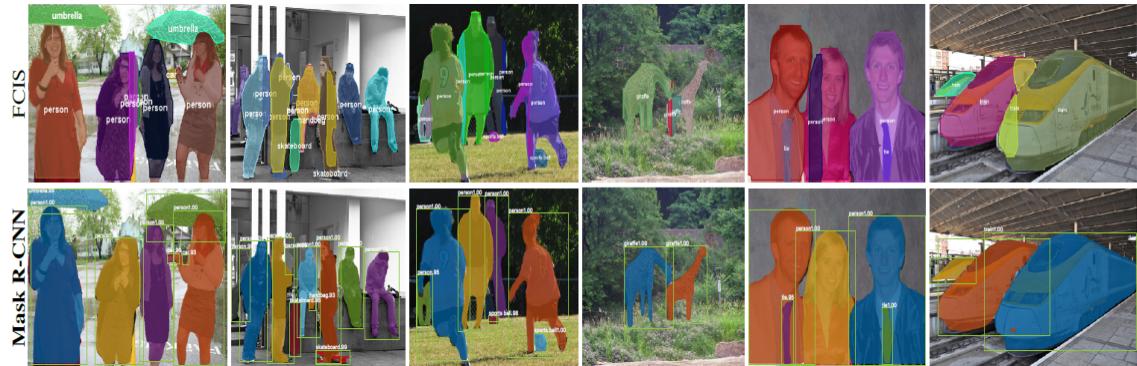


Abbildung 2.8: Bei FCIS entstehen Artefakte, wenn Objekte sich in einem Bild überlappen.[17]

Dort erzeugt es Artefakte, welche durch Mask R-CNN nicht entstehen (s. Abb. 2.8). 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 dieser Arbeit ausgewählt.

2.5 Evaluation des Modells

Jetzt wo gezeigt wurde, welches Modell in dieser Arbeit genutzt wird, fehlt eine Möglichkeit ein trainiertes Modell zu bewerten. *Mean average precision* (oder auch mAP) ist eine Metrik, um die Genauigkeit einer Instanzsegmen-

	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 2.2: 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+++. [17]

tierung zu messen.⁶ Aber bevor die erklärt werden können, muss noch die Begriffe *Precision*, *Recall* und *Intersection over Union* eingegangen werden.

2.5.1 Intersection over Union

Intersection over Union (oder auch IoU, dt: Schnitt über Vereinigung) ist eine wichtige Metrik für die semantische Segmentierung. Sie vergleicht die vorhergesagte Maske mit der Grundwahrheit⁷, um zu messen wie gut die Vorhersage mit der Grundwahrheit übereinstimmt. [21]

$$IoU = \frac{\text{Grundwahrheit} \cap \text{Vorhersage}}{\text{Grundwahrheit} \cup \text{Vorhersage}} \quad (2.3)$$

Die Schnittmenge beinhaltet alle Pixel, die sich in der Grundwahrheit als auch in der vorhergesagten Maske befinden. Pixel, die sich in der Grundwahrheit und in der Vorhersage befinden, werden von der Vereinigung zusammengefasst.

In der semantischen Segmentierung wird für jede Klasse ein unterschiedlicher IoU-Wert berechnet und dann wird der Mittelwert aus diesen Werten ermittelt, um einen globalen Messwert zu haben. In der Instanzsegmentierung wird für jede einzelne Objektinstanz mittels Instanzgrundwahrheit und Instanzvorhersage ein separater IoU-Wert berechnet. Wenn ein bestimmter Grenzwert überschritten wird, gilt diese Instanz als tatsächlich richtige Erkennung. [22]

⁶mAP ist nicht nur auf Instanzsegmentierung limitiert, sondern wird zum Beispiel auch als Metrik in der Objekterkennung genutzt.

⁷Die Grundwahrheit (engl.: ground truth) ist hier die binäre Maske, die die infizierte Fläche repräsentiert.

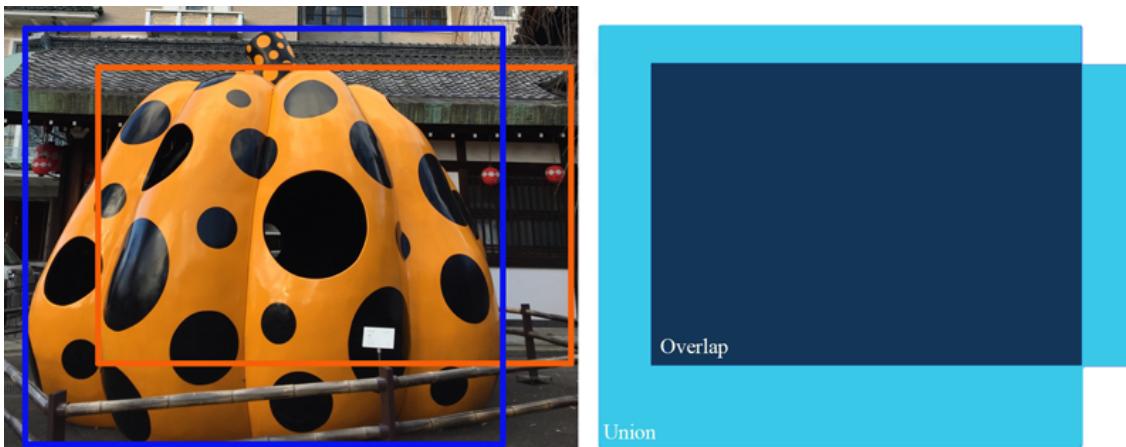


Abbildung 2.9: Beispiel Intersection over Union, Grundwahrheit in blau, Vorhersage in rot[21]

2.5.2 Precision und Recall

Precision (oder auch Falsch-Positiv-Rate) sagt aus mit welcher Wahrscheinlichkeit eine Vorhersage korrekt ist. Diese Metrik wird durch die Formel

$$Precision = \frac{RP}{RP + FP} \quad (2.4)$$

berechnet, wobei *RP* (Richtig-Positiv) die Anzahl der richtigen Erkennungen und *FP* (Falsch-Positiv) die Anzahl der falschen Erkennungen sei.[21] *Precision* ist also der Anteil von tatsächlich richtigen Erkennungen in Relation zu allen Erkennungen. In Bezug auf Instanzsegmentierung wird die Frage beantwortet, wie viele der erkannten Objekte in einem Bild tatsächlich eine passende Grundwahrheitüberschneidung und eine IoU-Grenzwertüberschreitung haben.[22]

Recall (oder auch Falsch-Negativ-Rate) misst die Wahrscheinlichkeit, dass alle tatsächlich wahren Detektionen korrekt erkannt wurden. Diese Metrik wird durch die Formel

$$Precision = \frac{RP}{RP + FN} \quad (2.5)$$

berechnet, wobei *RP* (Richtig-Positiv) die Anzahl der richtigen Erkennungen und *FN* (Falsch-Negativ) die Anzahl der Objekte, die fälschlicherweise nicht erkannt wurden, sei. *Recall* ist also der Anteil von tatsächlich richtigen Erkennungen in Relation zu allen Objekten im Datensatz.[21] In Bezug auf Instanzsegmentierung wird die Frage beantwortet, wie viele der Objekte mit Grundwahrheit in einem Bild als tatsächlich richtig erkannt werden und eine IoU-Grenzwertüberschreitung haben.[22]

2.5.3 Average Precision

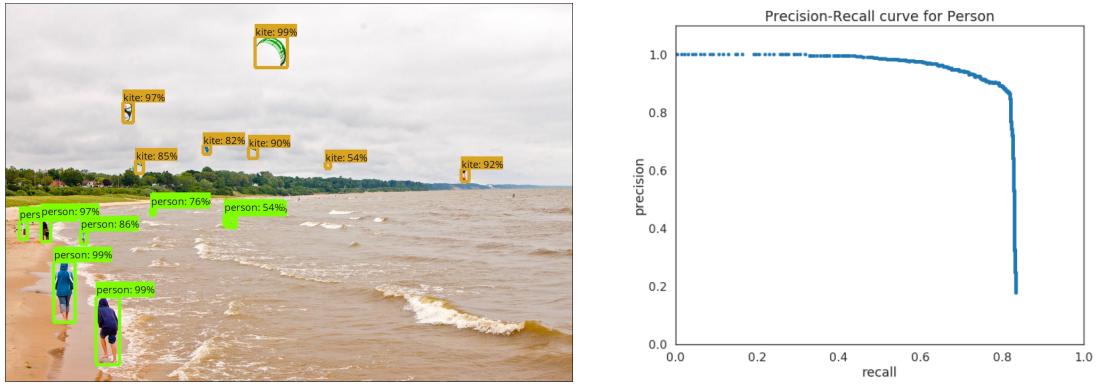


Abbildung 2.10: Links: Beispielbild mit multiplen Detektionen und Klassen[4] Rechts: Beispiel Precision-Recall-Kurve für die Klasse “Person”[19]

Ein *Precision*- und *Recall*-Wert bezieht sich jeweils auf eine detektierte Objektinstanz einer Klasse. Bei mehreren detektierten Objekten einer Klasse in einem Bild können diese in einer Precision-Recall-Kurve visualisiert werden (s. Abb. 2.10). *Average Precision* (oder auch AP) fasst die Form der Kurve zu einem Wert zusammen, indem es den Durchschnitt der *Precision*-Werte an elf *Recall*-Werten $[0, 0.1, \dots, 1]$ berechnet:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} p_{interp}(r) \quad (2.6)$$

Ein *Precision*-Wert p an der *Recall*-Stelle r wird interpoliert, indem der Maximumwert übernommen an der *Recall*-Stelle $\tilde{r} \geq r$ wird:

$$p_{interp}(r) = \max_{\tilde{r}: \tilde{r} \geq r} p(\tilde{r}) \quad (2.7)$$

wobei $p(\tilde{r})$ der *Precision*-Wert p an der *Recall*-Stelle \tilde{r} sei. Die Interpolation reduziert den Einfluss kleiner, lokaler Unebenheiten in der Kurve.[19]

2.5.4 Mean Average Precision

Mean Average Precision ist der Durchschnitt aller *Average Precision*-Werte jeder Klasse in jedem Element eines (Sub-)Datensatzes.⁸ mAP wird zum Beispiel auch in der COCO- oder PASCAL-VOC-Challenge benutzt, um die Resultate der Challenge-Teilnehmer zu bewerten (s. Tabelle 2.2). Aber hier kann es zu

⁸mAP wird oft nur AP genannt.

Unterschieden kommen, wie der *mAP* berechnet wird. So ist es bei der COCO-Challenge der durchschnittliche *mAP* über verschiedene *IoU*-Grenzwerte. Hier wird jeweils ein *mAP* an zehn verschiedenen *IoU*-Werten $[0.5, 0.55, \dots, 0.95]$ berechnet und aus den Ergebnissen wird der Durchschnitt ermittelt.[9] In dieser Arbeit wird stets $IoU = 0.5$ als Grenzwert benutzt, um die Auswertung einfach zu halten.

Kapitel 3

Overfitting

3.1 Begriffserklärung

Genaue Daten über Krankheitsbefäle im Agrarsektor sind rar, da diese in der Regel nicht öffentlich zugänglich sind.⁹ Daher musste mit *Overfitting* gerechnet werden. Das künstliche neurale Netzwerk soll daraufhin trainiert werden, dass es möglichst alle Befäle, 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. Das geschieht, weil sich das Netz auf nicht relevante Datenpunkte konzentriert, die im nur Trainingsdatensatz auftreten, aber nicht die allgemeine Charakteristika der Objekte widerspiegeln.

In Abb. 3.1 ist ein Beispiel wie Overfitting sich auswirken kann. Die linken zwei Bilder 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 über-

⁹Datenschutz kann ein Grund dafür sein.

Genaues
Da-
tum
nö-
tig?

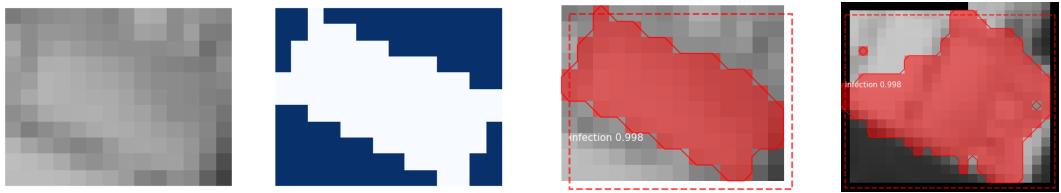


Abbildung 3.1: V.l.n.r. Bild von infizierter Agrarfläche aus Trainingsdatensatz / Binärmaske der infizierten Region, wird gemeinsam mit dem linken Bild zum Training in das KNN gespeist / Selbiges Bild, Ergebnis nach Trainingsdurchlauf, prognostizierte Ergebnisfläche in rot / Bild der selben Fläche, was nicht aus dem Trainingsdatensatz stammt, prognostizierte Ergebnisfläche in rot

gesprungen, was nicht der Realität entspricht. Overfitting ist ein bekanntes Problem im Bereich des maschinellen Lernens und es existieren multiple Methoden, um dem entgegenzuwirken.

3.2 Data Augmentation

Generell ist ein sehr großer Datensatz (≥ 10000 Elemente) für das Training förderlich. Jedoch ist das nicht immer möglich. In diesem Fall kann der Datensatz künstlich durch *Data Augmentation* vergrößert werden. Zu Data Augmentation zählen geringe Veränderungen der Daten - hier Bildmanipulationen. Solche Operationen können unter anderem

- Rotationen,
- Spiegelungen,
- Translationen,
- zufällige Ausschnitte,
- Gauß'sches Rauschen,
- Helligkeitsveränderungen,
- oder Kombinationen davon

beinhalten. Wobei nicht jede Augmentation-Technik für jeden Anwendungsfall sinnvoll ist. Wenn zum Beispiel ein KNN auf Autoerkennung trainiert werden soll, ist es nicht nützlich oder sogar hinderlich die Bilddaten so zu rotieren, dass es von oben nach unten oder umgekehrt zeigt. Des weiteren ist

sinnvoll Data Augmentation einzusetzen, obwohl genügend Daten vorhanden sind. So besteht der Datensatz aus zwei Klassen von Autos, Marke A und Marke B. Die Front der Fahrzeuge der Marke A sind nach rechts ausgerichtet, während die Autos der Marke B nach links ausgerichtet sind. Das neuronale Netz wird diesen markanten Unterschied den Marken zuordnen und wird ein links ausgerichtetes Fahrzeug der Marke A als ein Fahrzeug der Marke B klassifizieren.[5]

Für den Anwendungsfall dieser Arbeit sind Rotationen, Spiegelungen, zufällige Ausschnitte und Translationen valide Optionen zur Vergrößerungen des Datensatzes. Wie in dem vorherigen Beispiel wird das Modell auch Merkmale wie Form, Position im Bild oder Ausrichtung der RoI untersuchen, welche in der Realität keine feste Muster haben und nicht vorhersagbar sind. Damit wird nicht nur der Datensatz vergrößert, sondern auch das Modell generalisiert.

Künstliches Rauschen und Helligkeitsveränderungen sind mit hoher Wahrscheinlichkeit nicht geeignet. Die Bilddaten bestehen aus quantifizierten Werten und die Operationen könnten diese Werte verfälschen und damit das Endergebnisse negativ beeinflussen.

3.3 L2 Regularization

Ein kleiner Datensatz führt zu einem komplexen Modell und komplexere Modelle neigen zu Overfitting. *L2 Regularization* (oder auch *Ridge Regression*) vereinfacht das Modell, indem es hohe Gewichte bestraft und niedrige Gewichte bevorzugt.¹⁰ Hohe Gewichte können mit Anomalien korrelieren, die nur im Trainingsdatensatz auftreten und so wird das Netz Schwierigkeiten haben, fremde Daten richtig zu erkennen.

$$loss + \lambda \sum_{j=1}^p \beta_j^2 \tag{3.1}$$

Ridge Regression addiert einen zusätzlichen Bestrafungsterm (engl.: *penalty term*) zur Verlustfunktion (engl.: *loss function*), wobei *loss* die Verlustfunktion, der letzte Term der Bestrafungsterm und $\lambda > 0$ sei. Hier ist darauf zu achten,

¹⁰Es sei erwähnt, dass es neben L2 Regularization noch L1 Regularization (oder auch Lasso Regression) existiert. Diese Methode wird eingesetzt, um Underfitting zu verhindern.

dass λ sinnvoll gewählt wird. Wenn es zu groß gewählt wird, wird zu viel Gewicht hinzugefügt und das Modell tendiert zum *Underfitting*¹¹.[3][25]

3.4 Zusätzliche Methoden

Wenn ein KNN initial trainiert wird, werden die einzelnen Gewichte zufällig gewählt und dann dem Idealwert angenähert. Dieser Prozess kann zeitlich verkürzt werden indem vor-trainierte Gewichte eingesetzt werden. Je länger ein Modell trainiert werden muss, desto größer ist die Gefahr des *Overfittings*. Darum wird die Mask R-CNN-Implementierung mit Gewichten initialisiert, die auf dem COCO-Datensatz trainiert wurden.

In einem großen Datensatz beeinflussen einzelne Anomalien das Training nicht. Anders können diese Anomalien einen starken Einfluss in einem kleinen Datensatz haben. Eine hohe Anzahl von trainierbaren Parametern (Gewichten) reagieren darauf empfindlicher und es gilt das Modell durch Parameterminimierung resilenter zu machen. Um die Anzahl an Gewichten, die trainiert werden, zu minimieren, kann das *Backbone*¹² vereinfacht werden. Hier werden *ResNet50* und *ResNet101*, welche jeweils 50 und 101 Schichten besitzen, miteinander verglichen. Hierbei sollte *ResNet50* vorteilhafter sein, da es von beiden vorgestellten Architekturen die simplere hat.

Bei einem kleinen Datensatz hilft es die Experimente simpel zu gestalten. Das betrifft auch die Anzahl der Klassen. In dem betroffenen Region wurden zwei verschiedene Krankheiten festgestellt, die als einzelne Klassen deklariert werden können. Um die Datengröße pro Klasse zu erhöhen, werden diese beiden Klassen zu einer Klasse *infection* zusammengefasst.

¹¹Underfitting bezeichnet eine schlechte Performanz des neuronalen Netzes auf sämtliche Daten inkl. dem Trainingsdatensatz.

¹²He et al. bezeichnen die Architektur, die für die Merkmalsextraktion verantwortlich ist, als Backbone. Das Segment, das Klassifizierung, Bounding-Box-Generierung und Maskenerkennung durchführt, wird als *network head* (oder nur *head*) definiert.[17]

Kapitel 4

Konzept und Implementierungsdetails

Das Programmablauf wird in einzelne Schritte unterteilt, auf die in den nächsten Unterkapiteln näher eingegangen werden. Gleichzeitig werden elementare Implementierungsdetails aufgeführt.

4.1 Konzept

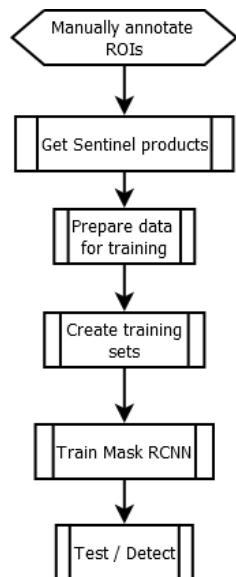


Abbildung 4.1: Gesamtablauf der Anwendung

Zuerst müssen die Daten für das Training bzw. für die Erkennung manuell annotiert werden. Diese Metadaten werden dann genutzt, um automatisch Sentinelprodukte¹³ mittels einer API, die von der Copernicus zur Verfügung gestellt wird, herunterzuladen. Aus den Produkten werden die relevanten Bänder extrahiert und unter anderem die jeweiligen NDVI-Werte berechnet. Nachdem die Produkte für das Training vorbereitet wurden, werden die Daten in ein Trainings- und in ein Validierungsdatensatz aufgeteilt. Der folgende Trainingsprozess basiert auf diesen Datensätzen. Sobald das Training abgeschlossen ist, kann die Performanz des Modells getestet werden.

4.2 Annotation

Zu Beginn werden die Regionen, die entweder für das Training benutzt oder überprüft werden, manuell erfasst. Vorausgesetzte Informationen sind

- Geografische Koordinaten,
- Zeitraum des Befalls und
- Bezeichnung der Infektion.

Als Format dieser Informationen dient *GeoJSON*¹⁴. GeoJSON enthält nicht nur geografische Daten, sondern ist auch um benutzerdefinierte Eigenschaften (*properties*) erweiterbar. Die Annotationen sind also GeoJSON-Features, die ein geografisches Polygon mit Metadaten enthalten.

```
{
  "type": "Feature",
  "properties": {
    "disease": 1,
    "from": "2018-07-12T13:00:00Z-7DAYS",
    "to": "2018-07-12T13:00:00Z+7DAYS"
  },
  "geometry": {
    "type": "Polygon",
    "coordinates": [[[11.171988617177981, 44.574291380353003],
      [11.1726616444942, 44.574017992242283],
      [11.17338129910439, 44.575068359984279],
      [11.171988617177981, 44.574291380353003]]]
  }
}
```

¹³Aufnahmenpakete der Sentinel-Plattformen werden als Produkte bezeichnet.

¹⁴GeoJSON ist eine Erweiterung des JSON-Format und beschreibt geografische Daten und Geometrien. GeoJSON wird durch den RFC7946-Standard definiert.

```

        [11.17273129334275, 44.575299863118993],
        [11.171988617177981, 44.574291380353003]]
    }
}

```

Listing 4.1: Annotation

`properties.disease` enthält die eindeutige, nummerische Repräsentation der Klasse bzw. Krankheit, die in dieser Region enthalten ist. Die Zuordnung der nummerischen Werte und des textuellen Bezeichners werden in einer separaten JSON als Schlüssel-Wert-Paare konfiguiert, wobei der Schlüssel nummerisch und der Wert textuell ist. Hier ist, darauf zu achten, dass der Schlüssel ≥ 1 ist, da 0 der implizite Schlüssel der Mask R-CNN-Implementierung für den Hintergrund ist. Diese Eigenschaft ist nur für das Training von Relevanz.

`properties.from` und `properties.to` sind jeweils Start- und Endzeitpunkt, in dem nach verfügbaren Sentinelprodukten gesucht werden soll. Das Format der jeweiligen Eigenschaften kann eine der folgenden Formen haben¹⁵:

- yyyyMMdd
- yyyy-MM-ddThh:mm:ss.SSSZ (ISO-8601)
- yyyy-MM-ddThh:mm:ssZ
- NOW
- NOW-<n>DAY(S) (oder HOUR(S), MONTH(S), usw.)
- NOW+<n>DAY(S)
- yyyy-MM-ddThh:mm:ssZ-<n>DAY(S)
- NOW/DAY (oder HOUR, MONTH usw.) - Der Wert wird entsprechend (z.B. auf den Tag) gerundet.

Es ist angebracht einen Zeitraum von mehreren Tagen bzw. Wochen zu wählen, da die Sentinel-2-Satelliten keine täglichen Daten liefern und weil eine Infektion typischerweise über einen längeren Zeitraum vorherrscht. Die Zeitspanne ist von der Krankheit abhängig. Hier wurden eine Woche vor und nach dem Aufnahmezeitpunkt genutzt, um nach Produkten zu suchen.

¹⁵Die Formate basieren auf der sentinel-sat-Version 0.12.2.

4.3 Suche nach Sentinelprodukten

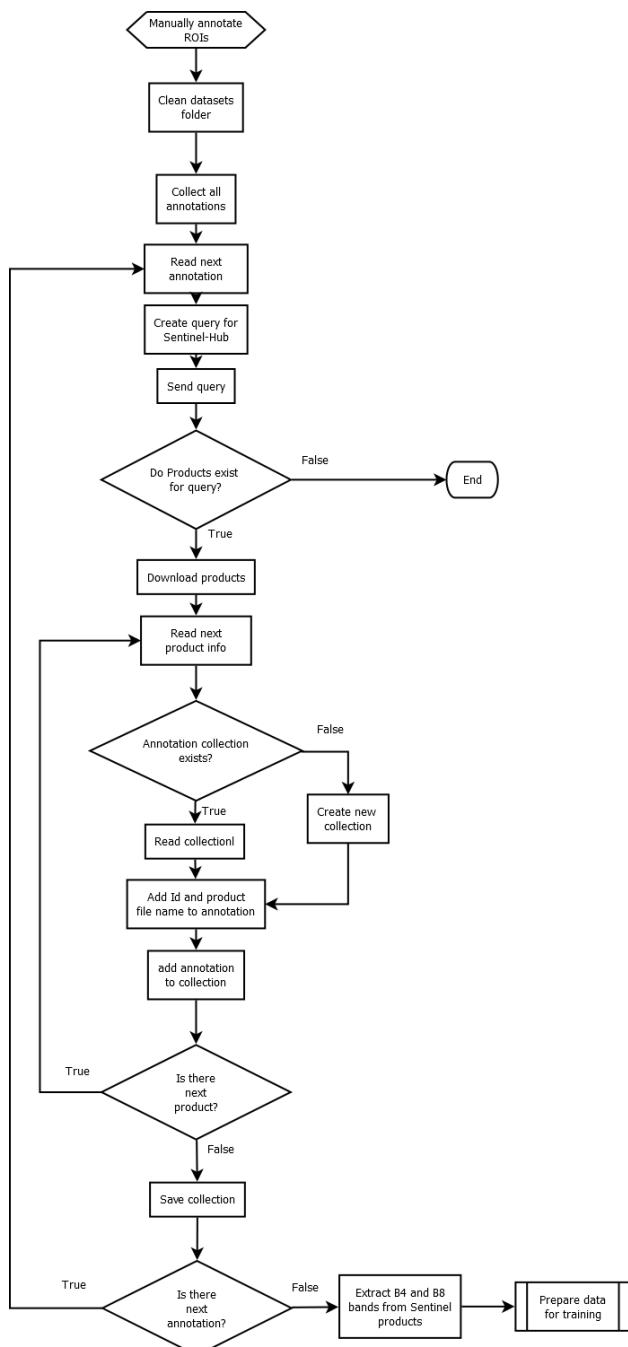


Abbildung 4.2: Ablaufdiagramm Sentineldatenaufbereitung

Der *Copernicus Open Access Hub*¹⁶ ermöglicht freien und offenen Zugriff auf Sentinel-Produkte. Die Daten sind sowohl über eine grafische Oberfläche als auch über eine REST-API verfügbar. Vorausgesetzung für beide Optionen ist

¹⁶<https://scihub.copernicus.eu/>

ein Account, der über die grafische Oberfläche erstellt werden kann.

Die Nutzung der Schnittstelle erfolgt über die Python-Bibliothek `sentinelsat`¹⁷. Bei einer Anfrage müssen die GeoJSON-Dateien in WKT¹⁸ umgewandelt werden, was von der Bibliothek übernommen werden kann. Die WKT-Geometrie wird als *footprint* (dt.: Fußabdruck) bezeichnet. Solang es nicht anders angegeben wird, gibt die Copernicus-API Produkte zurück, die die RoI schneiden. Des Weiteren wird der Plattformname statisch als 'Sentinel-2' definiert, damit keine Produkte von den anderen Sentinelplattformen zurückgegeben werden. Der Suchzeitraum wird aus der jeweiligen GeoJSON-Datei übernommen. Sollten für die Suchanfragen keine Produkte existieren, wird die Anwendung beendet, da es keine Basis gibt, auf der das Netzwerk trainiert werden kann. Eventuell muss bei so einem Fall der Zeitraum erweitert und Prozess wiederholt werden. Bei vorhandenen Produkten lädt das Skript diese herunter. Die Produkte werden in einem komprimierten Format geliefert und enthalten neben zusätzlichen Informationen, Banddaten in separaten Dateien im JPEG2000-Format¹⁹.

Für jedes Produkt, das zur aktuellen Annotation gehört, wird ein neuer Eintrag zu einer *FeatureCollection* hinzugefügt. Für die spätere Entwicklung sind die Annotationsen so leichter zu finden und bearbeitbar. Außerdem bleiben dadurch die Originaldaten unberührt. Dieser Schritt wird für jede vorhandene Annotationsdatei wiederholt. Anschließend werden die Bilddateien für B4 und B8 aus dem Produkt extrahiert.

Insgesamt wurden im angegebenen Zeitraum sechs Produkte gefunden. Jedes Produkt enthält optimale und wolkenfreie Konditionen über dem Zielgebiet zur weiteren Analyse.

¹⁷<https://sentinelsat.readthedocs.io/en/stable>

¹⁸WKT (Well-known text) ist eine Markup-Sprache zur Repräsentation von geometrischen Objekten auf Karten und räumlichen Referenzsystemen.

¹⁹JPEG 2000 genau wie GeoTIFF ist ein Bildformat in dem auch Metadaten abgelegt werden können. So sind Pixel geografischen Koordinaten zuordbar.

4.4 Aufbereitung der Sentineldaten

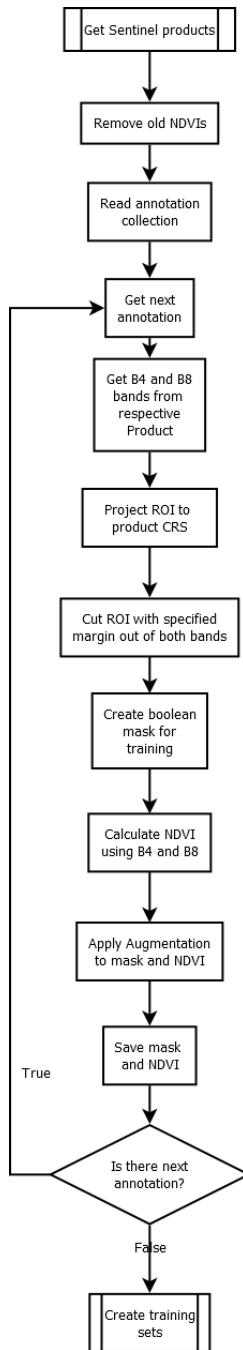


Abbildung 4.3: Ablaufdiagramm der Aufbereitung

Originale Sentinel-2-Aufnahmen haben eine Größe von 10980 * 10980 px und müssen deshalb deutlich verkleinert werden, um die Speicherbelastung zu minimieren. Vor allem da nur kleine Ausschnitte von wenigen Pixeln benötigt werden.

Zu Beginn werden alle Elemente aus der vorher erstellten *FeatureCollection* geladen und nacheinander bearbeitet. Jedes Produkt hat ein eigenes Koordinatenreferenzsystem (engl.: Coordinate Reference System, CRS) und unter Umständen unterscheiden sich die Systeme des Produktes und des Polygons²⁰. Diese müssen gleich sein, um miteinander agieren zu können. Das Produkt-CRS kann aus den extrahierten Bändern gelesen und dann dazu genutzt werden, um die Annotationskoordinaten in eben dieses zu projizieren.

Nachdem abgesichert wurde, dass die RoI in dem Sentinelprodukt enthalten ist, wird aus den Bändern die RoI samt einem vorher definierten Rand ausgeschnitten. Der Rand ist später bei der Data Augmentation hilfreich. Gleichzeitig erstellt die Anwendung eine gleich große binäre Maske, wobei die Elemente der Maske mit den Pixeln der RoI korrespondieren. Die Elemente, die die RoI repräsentieren, enthalten einen wahren Wert.

Nun wird der NDVI aus den B4- und B8-Ausschnitten, wie in Kapitel 2.2 ge-



Abbildung 4.4: V.l.n.r. B4 (RED), B8 (NIR), NDVI

zeigt, berechnet (s. Abb. 4.4). Das Ergebnis wird nun mittels Data Augmentation vervielfältigt. Dazu wird es vier mal um 90° gedreht. Danach wird jedes rotierte Bild horizontal und vertikal gespiegelt. Darauf werden neun mal aus den rotierten und gespiegelten Daten zufällig Bilder in der Größe von 16*16 px ausgeschnitten. Durch diese Operationen vergrößert sich die Grundgesamtheit um den Faktor 108. Jede Operation wird ebenfalls auf die entsprechende Maske angewandt. Dabei wird darauf geachtet, dass die Maske nicht vollständig aus falschen Werten besteht und dass die manipulierten Dateien ebenfalls der *FeatureCollection* hinzugefügt werden.

²⁰Nach RFC 7946 ist das geografische Referenzsystem *World Geodetic System 1984* (WGS 84) das Standardsystem.[6]

4.5 Trainings- und Validierungsdatensatz

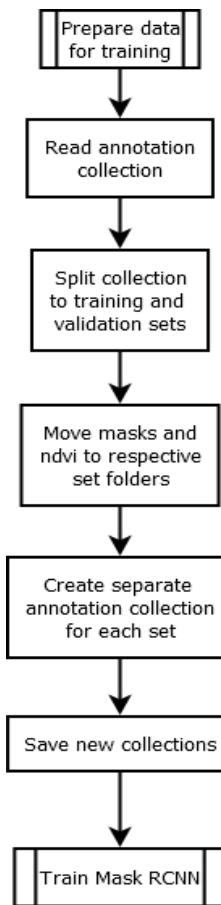


Abbildung 4.5: Ablaufdiagramm der Datensatzaufteilung

Die berechneten NDVI-Bilddateien und deren Masken bilden die Gesamtheit des Datensatzes. Um das Modell trainieren und testen zu können, muss diese Gesamtheit aufgeteilt werden.

- Trainingsdatensatz

Hiermit werden die Gewichte des neuronalen Netzwerkes trainiert.

- Validierungsdatensatz

Dieser Datensatz wird genutzt, um das trainierende Modell wiederholt zu evaluieren. Die Parameter in dem Modell werden auf Basis dieser Evaluation verändert, um das Idealergebnis zu approximieren.

- Testdatensatz

Während der Validierungsdatensatz während des Trainings genutzt wird, ist der Testdatensatz zu Evaluation des fertig trainierten Modells geeig-

net. Das KNN sieht also die Elemente nicht während des Trainingprozesses und repräsentiert „fremde“ Daten. Damit ist der Testdatensatz ein besserer Indikator der Leistung des Modells als der Validierungsdatensatz.



Abbildung 4.6: Darstellung der relativen Datensatzverteilung[28]

In welchem Verhältnis die Datensätze aufgeteilt werden, hat einen Einfluss auf das Training und hängt von der eigentlichen Größe des gesamten Datensatz und von dem trainierten Modell ab. Komplexere Modelle benötigen einen größeren Validierungsdatensatz, während ein Modell mit weniger Parametern mit einem kleineren auskommt.[28] Typischerweise besteht jedoch der Trainingsdatensatz aus dem größten Teil.

Hier wird zunächst die Gesamtmenge zwischen einem unbestimmten Teildatensatz (80% der Gesamtheit) und dem Testdatensatz (20% der Gesamtheit) aufgeteilt. Danach teilt sich der Teildatensatz in den Trainingsdatensatz (80% des Teiles) und in den Validierungsdatensatz (20% des Teiles). Die hier angewandte prozentualen Werte sind häufig angewandte initiale Verhältnisse und können bei Bedarf angepasst werden. Die geteilten Mengen werden unterschiedlichen Orten abgelegt und für jede Sammlung wird jeweils eine neue *FeatureCollection* erzeugt, die die Annotationen der entsprechenden Elemente enthalten.

4.6 Training/Detektion

Mit den vorbereiteten und aufgeteilten Daten kann nun das Netzwerk angeleert werden. In dieser Arbeit wurde eine Implementierung²¹ des kalifornischen Unternehmens Matterport, Inc. erweitert und genutzt. Matterport stellte den Quellcode 2017 unter MIT-Lizenz der Öffentlichkeit zur freien Verfü-

²¹https://github.com/matterport/Mask_RCNN

gung. Das neuronale Netz wurde mittels Python 3, Tensorflow und Keras implementiert. Zum Training eigener Daten müssen die zwei Python-Basisklassen `Dataset` und `Config` erweitert werden.

4.6.1 Dataset

Die Klasse `Dataset` bietet einen Weg, eigene Datensätze in das Modell zu laden, da die Daten in unterschiedlichen Formaten vorkommen können. Für die Erweiterung erbt die neue Klasse `CropDiseaseDataset` von `Dataset` und überschreibt die Methoden `Dataset.load_mask`, `Dataset.load_image` und `Dataset.image_reference`. Zusätzlich wurde die Methode `CropDiseaseDataset.load_crop_disease` implementiert. Eine Instanz von `Dataset` bzw. `CropDiseaseDataset` bildet nicht die gesamte Datenmenge ab, sondern jeweils einen spezifizierten Subdatensatz, die in Kapitel 4.5 definiert wurden. So muss jeder Teil separat instanziert werden.

`CropDiseaseDataset.load_crop_disease` fügt dem Objekt die Klassen- bzw. Bildinformationen wie eindeutige Bezeichnung und Dateipfad durch die internen Methoden `Dataset.add_class` bzw. `Dataset.add_image` hinzu.

`CropDiseaseDataset.load_image` lädt die eigentliche Bilddatei in den Arbeitsspeicher. Die NDVI-Werte wurden als Grauskalenbilder gespeichert und müssen in ein RGB-Format konvertiert werden. Es ist möglich durch Anpassungen des Mask RCNN-Codes, auch Dateien mit mehr oder weniger als drei Farbkanälen zu nutzen. Jedoch verursachte das Fehler, die zum Zeitpunkt der Verschriftlichung nicht gelöst wurde. Da es sich dabei um keinen kritischen Fehler handelt, wurde dieser durch die Konvertierung umgangen. Da die NDVI-Werte zwischen -1 und 1 liegen und Farbwerte aus ganzzahligen Werten bestehen, werden die NDVI-Werte mit 255 multipliziert, bevor sie schließlich hinzugefügt werden.

`CropDiseaseDataset.load_mask` liest die binären Masken aus den Dateien und ordnet sie einer Klasse und einem Bild zu.

Die Methode `CropDiseaseDataset.image_reference` gibt den vollständigen Dateipfad einer Bilddatei zurück.

4.6.2 Config

Die Klasse `Config` ist eine Sammlung von Parametern, die zur Konfiguration des Modells genutzt werden. Diese Parameter können entsprechend angepasst werden und haben jeweils Einfluss auf das Training bzw. auf die Erkennung. Es wird hier nur auf die Parameter eingegangen, die explizit im Rahmen der Entwicklung verändert wurden, um das Ergebnis des Trainings zu verbessern. Die Parameterbezeichnung ist der Name des Parameters, so wie er in `Config` deklariert wurde. Die Erklärung beschreibt den Parameter näher. Die angegebenen Werte sind alle möglichen Werte, die in der Entwicklung genutzt wurden. Welche Werte bei welchem Trainingsdurchlauf genutzt wurden, wird im Kapitel ?? beschrieben.

Parameterbezeichnung: BACKBONE

Erklärung: Die Backbone-Architektur mit der das Modell gebildet wird.

Werte: resnet50

Parameterbezeichnung: DETECTION_MIN_CONFIDENCE

Erklärung: Minimalste Wahrscheinlichkeit einer Detektion, um sich als vorhergesagte Instanz zu qualifizieren. Detektionen mit einer Wahrscheinlichkeit niedriger als dieser Wert werden ignoriert.

Werte: 0

Parameterbezeichnung: GPU_COUNT

Erklärung: Anzahl der Grafikkarten, auf denen das Modell trainiert wird. Wenn die CPU die Berechnungen übernimmt, muss der Parameter den Wert 1 annehmen.

Werte: 1

Parameterbezeichnung: IMAGE_MAX_DIM

Erklärung: Wichtig für IMAGE_RESIZE_MODE. Bildhöhe und -breite werden auf diesen Wert (in px) vergrößert.

Werte: 128

Parameterbezeichnung: IMAGE_RESIZE_MODE

Erklärung: Methode mit der ein Bild vergrößert bzw. verkleinert wird. Die gewählte Option vergrößert die Datensätze auf IMAGE_MAX_DIM*IMAGE_MAX_DIM und füllt das Bild mit 0-Werten, sollte das Ausgangsbild kein Quadrat sein.

Werte: square

Parameterbezeichnung: IMAGES_PER_GPU

Erklärung: Bilder die pro Schritt gleichzeitig für das Training in den Speicher geladen werden. Dieser Wert ist - je nachdem man auf der CPU oder auf der GPU trainiert - abhängig von der Größe der Arbeitsspeichers oder Grafikspeichers. Für die Detektion ist nur ein Bild pro Schritt erlaubt.

Werte: 1, 4

Parameterbezeichnung: LEARNING_RATE

Erklärung: Die Lernrate gibt an, wie stark die Gewichte des Netzwerkes korrigiert werden. Wenn dieser Wert zu klein ist, kann die Konvergenz zum Ideal sehr lange dauern. Ein zu hoher Wert kann das Ideal überschießen oder sogar dafür sorgen, dass die Lernkurve divergiert. He et al. nutzen in ihrer Ausarbeitung einen Wert von 0.02.[17] Dieser sorgt jedoch laut den Entwicklern von Matterport zu einem explosionartigen Anstieg der Gewichte, weswegen sie sich für 0.001 entschieden haben.[1] Hier werden beide Werte untersucht.

Werte: 0.001, 0.2

Parameterbezeichnung: NUM_CLASSES

Erklärung: Die Anzahl der Klassen, die trainiert werden. Hier ist darauf zu achten, dass der Hintergrund als eigene Klasse gesehen wird, auch wenn diese nicht explizit definiert wird. Daher muss die zusätzliche Klasse mit in diesen Wert einberechnet werden.

Werte: 2

Parameterbezeichnung: RPN_ANCHOR_SCALES

Erklärung: Quadratgröße der RPN-Anker. Die Werte sollten kleiner sein als IMAGE_MAX_DIM. Anker, die größer sind, machen keinen Sinn, da Objektinstanzen sich innerhalb der Bildgrenzen befinden.

Werte: (8, 16, 32, 64, 128)

Parameterbezeichnung: STEPS_PER_EPOCH

Erklärung: Anzahl der Trainingsschritte bis eine Epoche abgeschlossen ist. Der Wert ist abhängig von der Größe des Trainingdatensatzes und wie viele Bilder pro Schritt prozessiert werden.

Werte: $\frac{SIZE_{Train}}{IMAGES_PER_GPU}$

Parameterbezeichnung: USE_MINI_MASK

Erklärung: Wenn dieser bool'sche Wert wahr ist, werden Instanzmasken zu einer spezifizierten Größe verkleinert. Hier wird das jedoch deaktiviert, da die Eingangsbilder klein genug sind und diese Operation nicht nötig ist.

Werte: False

Parameterbezeichnung: WEIGHT_DECAY

Erklärung: Einflussgröße der L2 Regulization.

Werte: 0.01, 0.001, 0.0001

Der Konfigurationsparameter BATCH_SIZE wird automatisch aus $GPU_COUNT * IMAGES_PER_GPU$ berechnet. Während IMAGES_PER_GPU angibt, wie viele Bilder pro Rechnereinheit (GPU oder CPU) in das neuronale Netz geladen werden, definiert BATCH_SIZE wie viele Bilder insgesamt pro Trainingsschritt geladen werden.

CropDiseaseConfig erbt von Config und die Werte werden vor jedem erneuteten Trainingsdurchlauf verändert und der Modell-Instanz übergeben.

4.6.3 Ablauf

Die Benutzer können bei Skriptaufruf als Kommandozeilenparameter spezifizieren, ob das neuronale Netz trainiert oder ob ein Bild überprüft werden soll. Für beide Fälle müssen sie spezifizieren, auf welcher Basis das neuronale Netz gebildet werden soll. Hier gibt es drei Optionen für diese Anwendung:

- vortrainierte COCO-Gewichte
- vortrainierte ImageNet-Gewichte
- Fortsetzen eines alten Trainingsdurchlaufs

Wählen die Benutzer eine der ersten beiden Möglichkeiten, werden die vortrainierten Gewichte heruntergeladen. Bei der dritten Option muss eine vorher gespeicherte Datei geladen werden, die die Gewichte des gewünschten Modells enthält. Aus den gewählten Gewichten und dem CropDiseaseConfig-Objekt wird nun eine Mask R-CNN-Instanz gebaut. Dieses Instanz kann eine von zwei Modi

- training
- inference

annehmen. Wobei *training* die Gewichte des Modells für das Training veränderbar macht und *inference* für die Detektion die Gewichte einfriert, da sie während einer Erkennung nicht trainiert werden müssen.

training

Wie in Kapitel 4.6.1 beschrieben, ist für jeden Subdatensatz jeweils ein unterschiedliches `CropDiseaseDataset`-Objekt nötig. In den separaten Objekten werden nun die Daten, Klassen und Masken für das Training, die Validierung und das Testen geladen und vorbereitet.

Das Test-Objekt wird nur indirekt für das Training benutzt. Es hat keinen direkten Einfluss auf den Trainingsverlauf, sondern wird genutzt, um den mAP am Ende jeder Epoche zu berechnen. Hierzu ist eine weitere Mask R-CNN-Instanz im *inference*-Modus notwendig, dessen Gewichte nach jeder Epoche auf die selben Werte des *training*-Modells aktualisiert werden. Die Konfiguration definiert die von `CropDiseaseConfig` abgeleitete Klasse `InferenceConfig`. Anders als die Parameter von `CropDiseaseConfig` werden die Werte nur einmalig festgelegt.

```
class InferenceConfig(CropDiseaseConfig):
    # Nur ein Bild pro Detektion erlauben
    IMAGES_PER_GPU = 1
    GPU_COUNT = 1
    # Jede Erkennung wird angezeigt
    DETECTION_MIN_CONFIDENCE = 0.0
```

Listing 4.2: InferenceConfig

Auf Grund des *inference*-Modells und des Testdatensatzes kann nun am Ende einer Epoche der mAP berechnet werden, mit dessen Hilfe am Ende des Trainings das neuronale Netz evaluiert werden kann. Um Vergleichswerte zu haben, wird dieser für Trainings- und Validierungsdatensatz ebenfalls berechnet.

Bevor das Training starten kann, müssen noch die Anzahl der Epochen und die trainierbaren Schichten angegeben werden. Die Anzahl der Epochen hat ein

Einfluss darauf, wie oft ein Training wiederholt wird. Wird der Wert zu niedrig gewählt, kann es sein, dass die Gewichte nicht ausreichend genug trainiert sind. So riskiert man *Underfitting*. Ist der Wert jedoch zu hoch, werden die Gewichte zu sehr an den Trainingsdatensatz angepasst und ein *Overfitting* ist die Folge. Mit dem *training*-Modus gibt man zwar an, dass die Gewichte variabel sind, aber durch die hier angegebenen Schichten wird genau festgelegt, welche Schichten eingefroren werden und welche nicht. So wird mit `layers='heads'` lediglich der *network head* trainiert, werden alle Schichten mit `layers='all'` trainierbar sind. Bei einem kleinen Datensatz wäre es zum Beispiel sinnvoll, nur die Klassifikatoren zu trainieren, die sich im Netzwerkkopf befinden. Zusätzlich können verschiedene Schichten iterativ trainiert werden. Zum Beispiel wird in den ersten 20 Epochen der *head* trainiert. Anschließend werden alle Schichten für 80 Epochen trainiert.

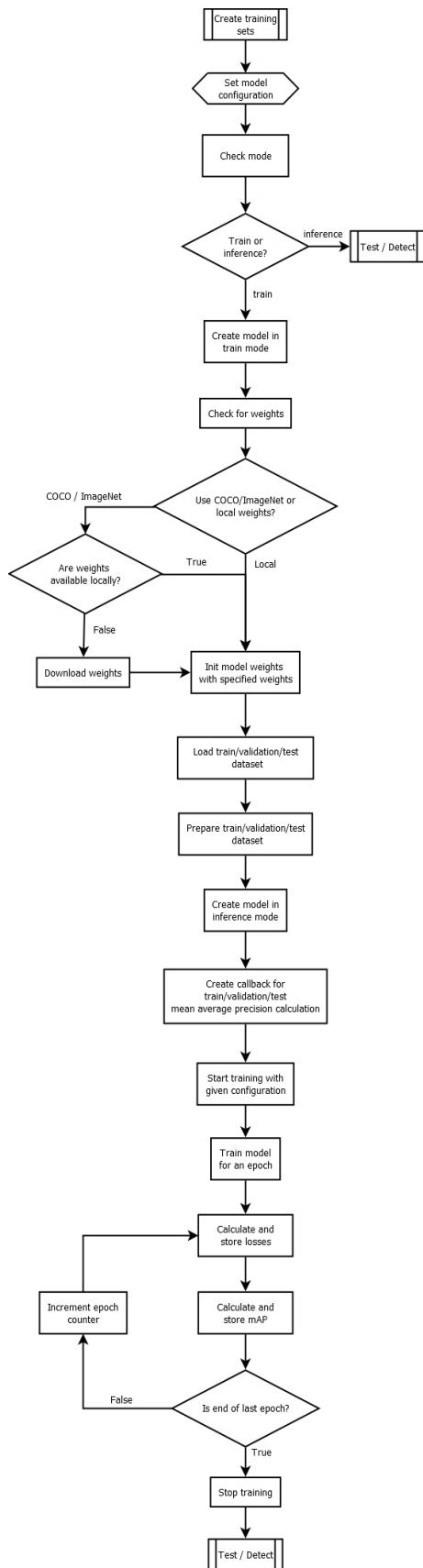


Abbildung 4.7: Ablaufdiagramm des Trainings

inference

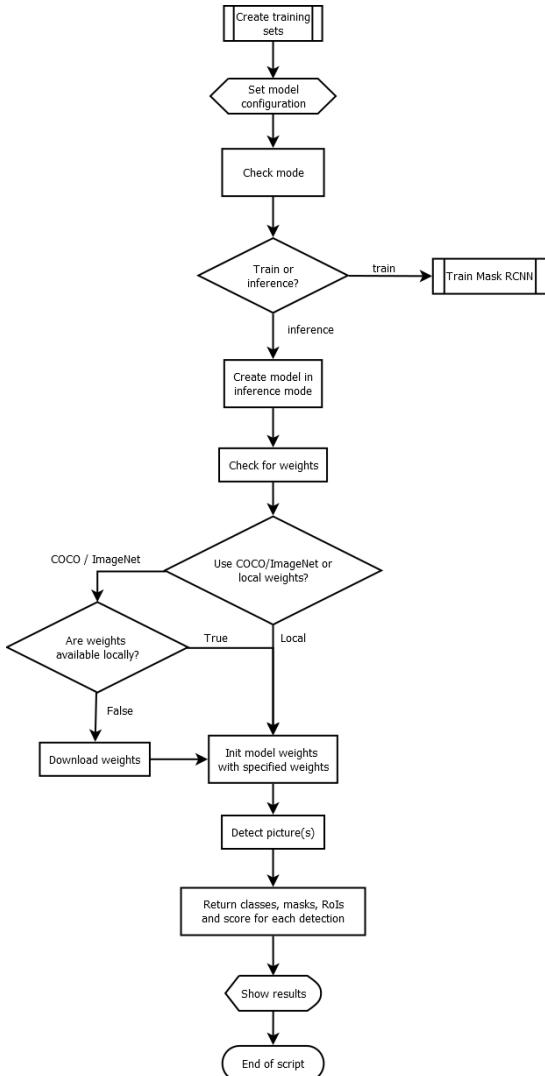


Abbildung 4.8: Ablaufdiagramm der Erkennung

Im *inference*-Mdous wird das Mask R-CNN-Objekt durch `InferenceConfig` konfiguriert. Wenn die angegebenen Gewichte in das Modell geladen wurden, können n Bilder, wobei $n \geq 1$ sei, zur Detektion in das Netzwerk gegeben werden, ohne dass eine spezielle `Dataset`-Instanz vonnöten ist. Es ist wichtig, dass das die NDVI-Werte vor der Detektion zu RGB-Werten konvertiert werden, wie es in Kapitel 4.6.1 beschrieben ist.

Nach der Detektierung mit möglichen Objekten gibt das Modell eine Liste mit n Elementen zurück. Die Elemente dieser Listen bestehen aus m vorhergesagten *RoIs*, Masken und Klassen der jeweiligen Bilder, wobei $m \geq 0$ sei. Diese sind in separaten Listen abgelegt und sind so angeordnet, dass die jeweiligen

Elemente der Listen an Position i zusammengehören, wobei $0 \leq i < m$ sei. So gehören die i -te *RoI*, Klasse und Masken zu der selben vorhergesagten Objektinstanz. Zusätzlich liegt der Instanz ein Wert zwischen 0 und 1 bei, die die Wahrscheinlichkeit der Korrektheit der Vorhersage angibt.

Kapitel 5

Ergebnis

Nach der Beschreibung des Konzept und Verlauf des Python-Skripts geht dieses Kapitel auf die wichtigsten Trainingsdurchläufe ein und diskutiert die Ergebnisse.

Es wurden iterativ Trainingprozeduren durchgeführt mit jeweils unterschiedlichen Konfiguration, wie um vorherigen Kapitel gezeigt, und miteinander verglichen. Die genauen Konfigurationen werden bei den einzelnen Experimenten angeführt. Die Experimente liefen auf einem Rechner mit NVIDIA TITAN Xp 12 GB VRAM Grafikkarte, Intel i7-7800X CPU und 32 GB RAM. Da diese Arbeit zeitlich begrenzt ist, wurden die Experimente auf max. 100 Epochen begrenzt, was die durchschnittliche Laufzeit auf etwa drei bis vier Stunden je Experiment bringt. Diese Beschränkung wirkt ebenfalls Overfitting entgegen, da nicht ausreichend Zeit hat, um sich auf den Trainingsdatensatz „einzugewöhnen“.

Es wurden drei unterschiedliche Datensätze aus den Sentinelprodukten erzeugt.

1. Der originale unaugmentierte Datensatz enthält insgesamt jeweils zwölf Bilder und Masken²² und dient als Basis für das Ausgangsexperiment.
2. Der augmentierte Datensatz (s. Kapitel 3.2) enthält 1291 Dateien und Masken²³. Nachdem die zufällige Ausschnitte produziert wurden, kam es vor, dass einige Masken keinen *RoI*-Anteil enthielten. Diese Masken

²²Trainingsanteil: 7, Validationsanteil: 2, Testanteil: 3

²³Trainingsanteil: 825, Validationsanteil: 207, Testanteil: 259

und auch die zugehörigen Bilddateien wurden verworfen und weicht deswegen von der eigentlichen Größe von 1296 Elementen²⁴ ab.

3. In einem dritten Datensatz wurden Ausschnitte benutzt, deren Bildgrenzen sich an den äußersten Punkten der Masken befinden (s. Abb. 3.1). Dieser Datensatz wurde ausschließlich mittels Rotationen erweitert und enthält 144 Elemente²⁵.

5.1 Training mit Rohdaten

```
class CropDiseaseConfig(Config):  
    BACKBONE = "resnet50"  
    IMAGE_MAX_DIM = 128  
    IMAGE_MIN_DIM = 128  
    IMAGE_RESIZE_MODE = "square"  
    IMAGES_PER_GPU = 4  
    LEARNING_RATE = 0.001  
    NUM_CLASSES = 1 + 1  
    RPN_ANCHOR_SCALES = (8, 16, 32, 64, 128)  
    STEPS_PER_EPOCH = 23  
    USE_MINI_MASK = False
```

Listing 5.1: Konfiguration für Experimente 1

Zuerst wurden ein Modell auf Datensatz 1 und alle Schichten des Modells wurden trainiert. Diese Experimente sollen zeigen, wie ein zu kleiner Datensatz sich auswirken kann.



Abbildung 5.1: *mAP*-Graph von Experiment 1, X-Achse: Epochennummer, Y-Achse: *mAP*-Werte

²⁴Die zwölf ursprünglichen Dateien multipliziert mit dem Data Augmentation-Faktor von 108.

²⁵Trainingsanteil: 92, Validationsanteil: 23, Testanteil: 29

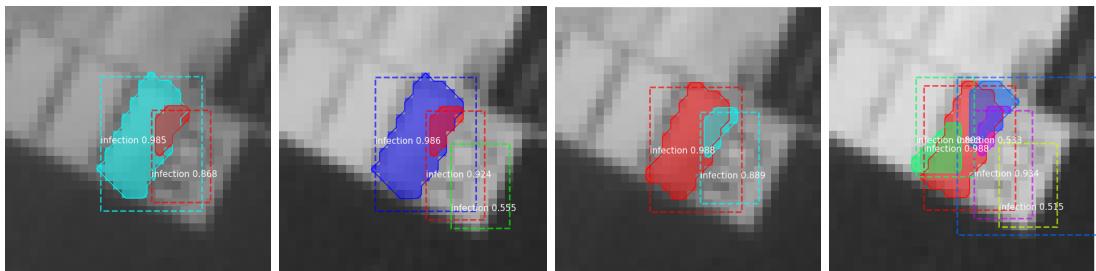


Abbildung 5.2: Beispielvorhersagen anhand der Gewichte von Epoche 55

Man sieht, dass das Modell während des Trainings unterdurchschnittliche Ergebnisse erzielt, was wenig überraschend ist, da der Datensatz sehr klein ist. Da nach jeder Epoche die Gewichte zwischengespeichert werden, können diese einzeln geladen werden. So wurde für eine Detektion ein Modell mit den Gewichten der 55. Epoche initialisiert, da sich hier der *mAP*-Wert auf dem Maximalwert befindet. Die Bilder in Abb. 5.2 entstammen einem fremden Datensatz, ähneln aber in den Grundcharakteristika dem eigentlichen Datensatz 1. Zum Beispiel ist das Zielfeld ähnlich ausgerichtet und zentriert. Man sieht, dass sich die berechneten Masken relativ genau auf das Zielfeld eingrenzen. Jedoch sind die erzeugten Bounding-Boxen und passen nicht zu den entsprechenden Masken, falls eine Maske erkannt wurde. Auch wurden mehrere Objektinstanzen erkannt, wobei eine einzige Instanz detektiert werden sollte. Werden die Bilder in Abb. 5.2 nun vertikal gespiegelt und in das Netzwerk gegeben, erzeugt das Modell keine Masken, obwohl lediglich die Ausrichtung verändert wurde. Das ist ein Hinweis auf Overfitting, da das neuronale Netz minimale Änderungen nicht mehr erkennt.

5.2 Datensatzerweiterung durch Rotation

```
class CropDiseaseConfig(Config):
    BACKBONE = "resnet50"
    IMAGE_MAX_DIM = 128
    IMAGE_MIN_DIM = 128
    IMAGE_RESIZE_MODE = "square"
    IMAGES_PER_GPU = 4
    LEARNING_RATE = 0.001
    NUM_CLASSES = 1 + 1
    RPN_ANCHOR_SCALES = (8, 16, 32, 64, 128)
    STEPS_PER_EPOCH = 206
```

```
USE_MINI_MASK = False
```

Listing 5.2: Konfiguration für Experiment 2

Bei diesem Experiment wurde ein Modell Datensatz 3 und alle Schichten des Modells trainiert. Die Konfiguration bleibt unverändert, jedoch ist hier der augmentierte Datensatz vergleichweise größer.

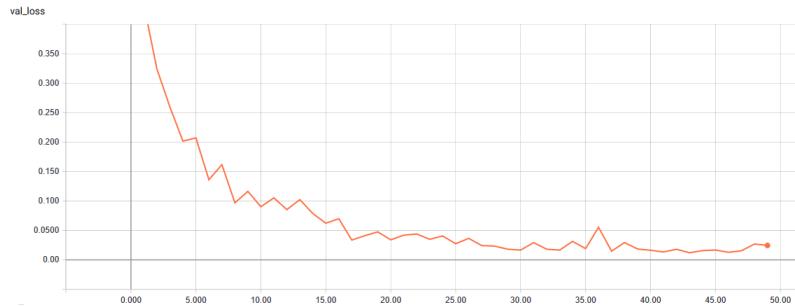


Abbildung 5.3: *loss*-Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: *loss*-Werte

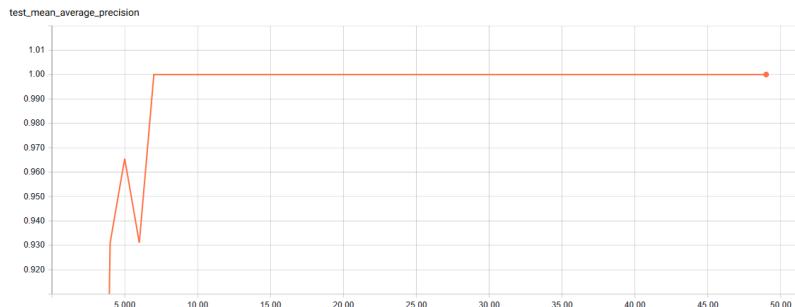


Abbildung 5.4: *mAP*-Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: *mAP*-Werte

Die *mAP*-Kurve konvergiert ab Epoche 7 gegen 1 und verweilt dort für den Rest des Trainings. Dagegen sinken die *loss*-Werte weiterhin und nähern sich ab Epoche 37 0 an. Daher wurde die Gewichte dieser Epoche näher untersucht. Abb. 5.5 zeigt, dass das Modell resistenter gegenüber Rotationen ist. Nichtsdestotrotz reagiert das Modell empfindlich auf Veränderungen wie zum Beispiel ein größerer Ausschnitt oder Translationen (s. unterste Reihe in Abb. 5.5). Daraus ergibt sich, dass das Modell nicht allgemein einsetzbar ist und die Daten durch weitere Augmentationen randomisiert werden müssen.

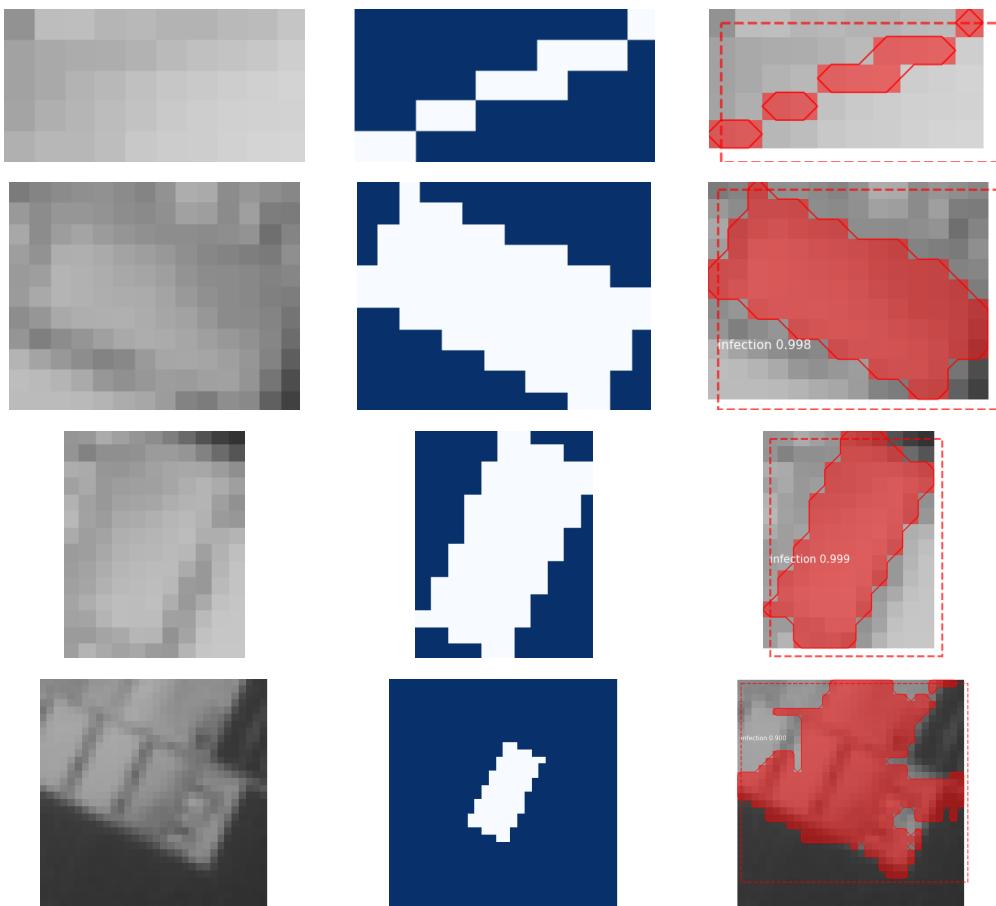


Abbildung 5.5: Beispielvorhersagen anhand der Gewichte von Epoche 37,
links: Ausgangsbild, mitte: Maske, rechts: Detektion

5.3 Data Augmentation und Regularization

```
class CropDiseaseConfig(Config):
    BACKBONE = "resnet50"
    IMAGE_MAX_DIM = 128
    IMAGE_MIN_DIM = 128
    IMAGE_RESIZE_MODE = "square"
    IMAGES_PER_GPU = 4
    LEARNING_RATE = 0.001
    NUM_CLASSES = 1 + 1
    RPN_ANCHOR_SCALES = (8, 16, 32, 64, 128)
    STEPS_PER_EPOCH = 3
    USE_MINI_MASK = False
    WEIGHT_DECAY = 0.0001 # Orange, Dunkelblau, Rot
    WEIGHT_DECAY = 0.001 # Pink
    WEIGHT_DECAY = 0.01 # Hellblau
```

Listing 5.3: Konfiguration für Experiment 3

Diese Sektion vergleicht mehrere Trainingsläufe direkt miteinander. Zur simpleren Kommunikation werden die einzelnen Durchläufe mit den Farben betitelt, wie sie in den Graphen 5.6 und 5.7 (Orange, Rot, Pink, Hell- und Dunkelblau) repräsentiert sind. Die verschiedenen `WEIGHT_DECAY`-Werte in Listing 5.3 sind in den entsprechenden Modellkonfigurationen zu verwenden, wie sie in den Kommentaren benannt sind, wobei der Wert für die orangene, rote und dunkelblaue Konfigurationen implizit in der Basisklasse `Config` definiert ist. Die pinke und hellblaue Konfiguration soll den Einfluss von L2 Regularization zeigen. Alle Modelle bis auf das orangene wurden auf Grundlage von Datensatz 2 trainiert. Aus Gründen, die später erklärt werden, wurde für das orangene Modell ein Datensatz kreiert, das sich auf die kleine Zone innerhalb des Fledes mit der stärkeren Infektionskonzentration beschränkt. Der `head` des dunkelblauen Netzwerks wurde trainiert, während alle Schichten der restlichen Modelle trainiert wurden.

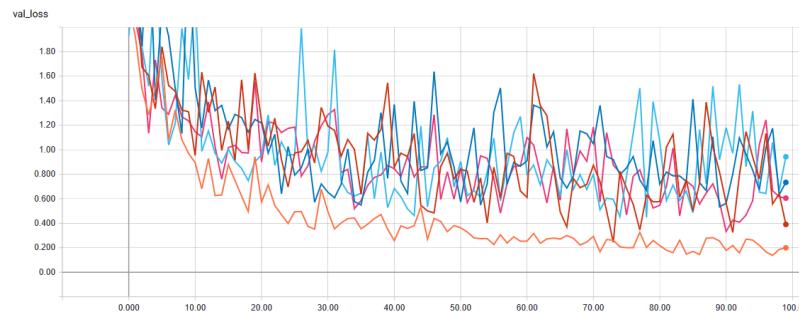


Abbildung 5.6: *loss*-Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: *loss*-Werte



Abbildung 5.7: *mAP*-Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: *mAP*-Werte

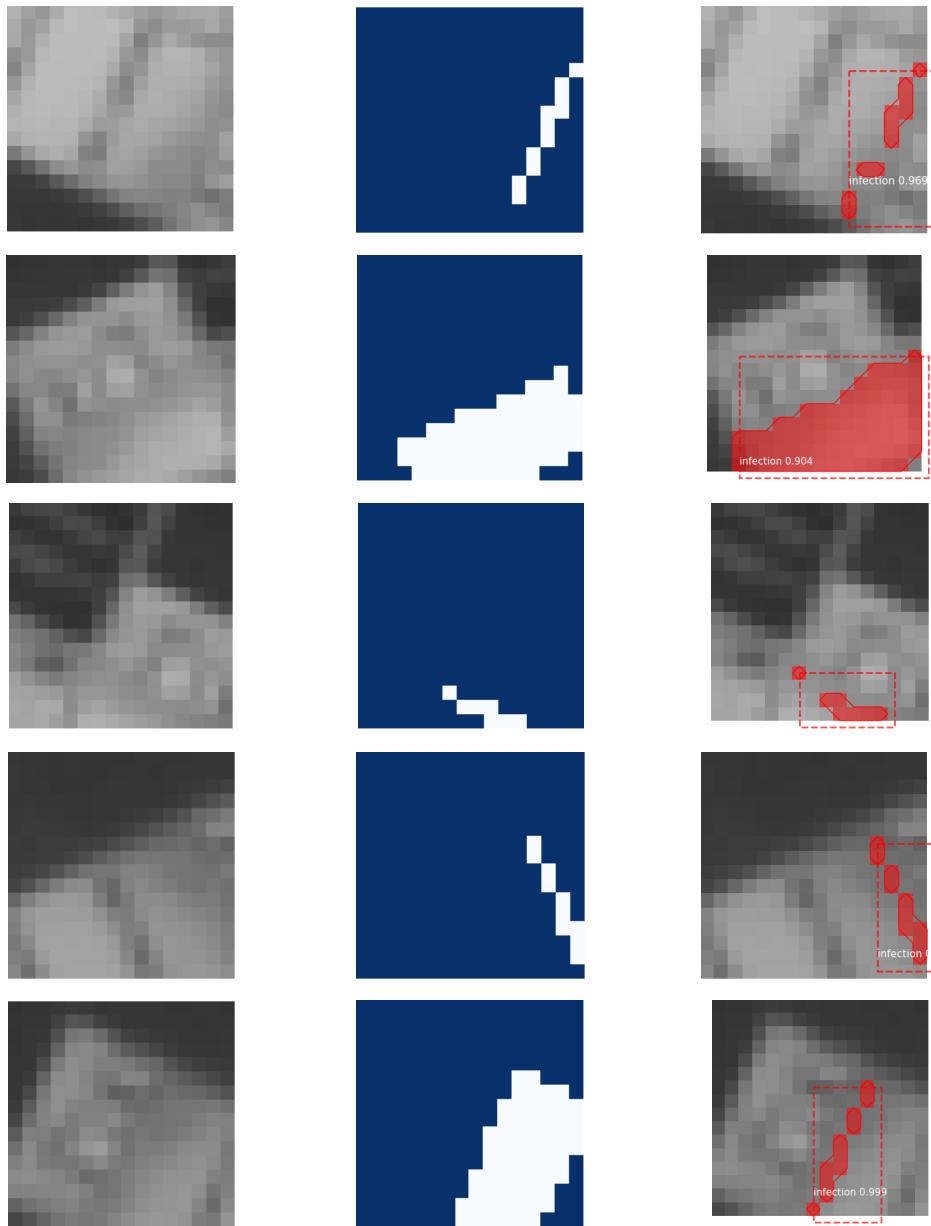


Abbildung 5.8: Beispielvorhersagen verschiedener Modelle, v.l.n.r.: Ausgangsbild, Masken, Vorhersage, v.o.n.u.: Vorhersagen von dunkelblau, rot, hellblau, pink, orange

Für auf Datensatz 3 trainierten Modelle (jeweils Epoche 100) zeichnen sich in Abb. 5.6 starke Schwankungen ab und kein Modell überschreitet $mAP > 0.7$. Die Vorhersagen in den oberen vier Reihen in Abb. 5.8 stimmen zum großen Teil mit den Masken überein. Weitere hier nicht aufgeführte Vorhersagen zeigen ähnliche Ergebnisse. Es kommt vor, dass die kleine Zielregion vorhergesagt wird, obwohl die entsprechende binäre Maske die große *RoI* repräsentiert, so ähnlich wie es in der unteren Reihe zu sehen ist. Dieses Verhalten

erklärt die schlechten *mAP*-Kurven und die Schwankungen. Da die Detektion nicht mit dem erwarteten Ergebnis übereinstimmt und daher der berechnete Fehler höher ist, obwohl die Detektion eigentlich korrekt ist, da sich beide genutzten *RoIs* geografisch schneiden. Durch den Fehler ist die Korrektur der Gewichte entsprechend größer. Um die Annahme zu bekräftigen, wurde ein Modell (orange) auf dem speziellen Datensatz trainiert. Im Vergleich zu den anderen Kurven konvergiert die orangene *loss*-Kurve gegen 0 bzw. die orangene *mAP*-Kurve gegen 1. Wäre mehrere unterschiedliche *RoIs* vorhanden, die sich nicht schneiden, würde der Einfluss sich nicht bemerkbar machen. Was hier jedoch nicht der Fall ist, da beide *RoIs* jeweils etwa die Hälfte des Datensatzes ausmachen.

Es ist zu sehen, dass die *Ridge Regression* (hellblau und pink) keinen merkbaren Einfluss auf das Training haben, weil einen ähnlichen Verlauf wie die dunkelblaue und rote Kurven haben. Im Gegensatz dazu zeigt die *Data Augmentation* positive Ergebnisse (s. Abb. 5.8) und alle Modelle sind deutlich robuster gegenüber manipulierten Bildern. Jedoch schlagen die Detektionen bei größeren Ausschnitten fehl oder geben sinnlose Resultate zurück. Weiterhin sei erwähnt, dass das Modell, dessen *head* trainiert wurde, von allen hier erwähnten Modellen am schlechtesten abschneidet und es zumindest nicht hilfreich ist, nur die oberen Schichten zu trainieren.

5.4 Ergebnisdiskussion

Ergebnisse zeigen, dass Mask R-CNN ein potentielles Werkzeug ist, um Landwirte bei der Kontrolle ihrer Felder zu unterstützen. Die implementierte Anwendung lädt automatisch Sentinelprodukte herunter, extrahiert die gefragten Bildregionen und berechnet daraus einen Gesundheitsindex der Flächen. Ein konfiguriertes Mask R-CNN-Modell analysiert die Indexwerte auf Muster, die mit Krankheiten korrelieren könnten.

Das Ziel der Arbeit eine Vielzahl von Krankheiten zu analysieren und zu identifizieren, konnte nicht erreicht werden, da hier nur eine Region zur Verfügung stand. Aufgrund dessen wurden wie erwartet bei ersten Experimenten Modelle trainiert, die Overfitting aufwiesen. Nichtsdestotrotz half Data Augmentation, die Größe des Datensatzes zu erweitern und Variationen in Form und Ausrichtung der Ausgangsfläche zu erzeugen. Dadurch konnte ein Modell

trainiert werden, dass Resistenz gegenüber Overfitting aufwies. Der Einsatz von Data Augmentation ist ebenfalls sinnvoll sollte kein Overfitting vorliegen. Ausbreitungen von Infektionen sind willkürlich und folgen keinen bestimmten geometrischen Mustern. Data Augmentation hilft durch künstliche Randomisierungen ein Modell darauf vorzubereiten.

Jedoch zeigten die oft bei Overfitting verwandte Ridge Regression keinerlei Auswirkung. Wenn einige ausgewählte Neuronen-Schichten trainiert wurden, verschlechterte sich die allgemeine Performanz. Deswegen sind L2 Regularization und Training ausgewählter Schichten in weiteren Untersuchungen, die nur mit hier untersuchten Felddaten arbeiten, nicht zu empfehlen.

Weiterhin wurde die Vermutung aufgestellt, dass die *RoIs* negative Einflüsse auf die Bewertung des Trainings haben, da eine Region von der zweiten Region umschlossen wird. Ein Modell, dass auf Grundlage einer Region trainiert wurde, erzielt bessere *mAP*-Werte und bestärkt dadurch diese Annahme. Weitere Forschungen sollten den Datensatz, um andere Felddaten erweitern. Durch eine größere Vielfalt können solche Einflüsse ignoriert werden, da der Effekt dadurch entsteht einen minimalen Einfluss bei der Fehlerberechnung hat.

Die geometrische Form der Infektionen (s. Kapitel 2.1) wurden einmalig gemessen und genutzt um die binären Masken für das Mask R-CNN-Training zu erstellen. Das basiert auf der Annahme, dass die Infektion statisch ist und sich nicht ausbreitet, entspricht aber nicht dem realen Zustand. Eine Infektion, sobald sie die ersten Pflanzen in einem Feld befallen hat, breitet sich weiter aus. Dementsprechend ändert sich mit fortschreitender Ausbreitung die Form der infizierten Fläche. Um sicherzustellen, dass das Modell nicht allein auf die Form der binären Masken trainiert wird, sollten optimal Infektionen über einen längeren Zeitraum beobachtet und periodisch die Grenzen des Befalls neu vermessen werden. Da so ein Vorhaben mit einem gewissen Aufwand verbunden ist, ist es unwahrscheinlich, dass multiple Ausbreitungen, die nicht unter Laborbedingungen stattfinden, wie beschrieben dokumentiert werden.

Die NDVI-Werte innerhalb der Masken sind für die Erkennung der Krankheit von Relevanz. Um zu bestätigen, dass das Modell nicht allein auf die Geometrie des Feldes reagiert, da die NDVI, wurden NDVI-Werte des gleichen Feldes

aus Sentinel-2-Produkten berechnet, die im Juli 2017 aufgenommen wurden. Es ist anzunehmen, dass die Nutzpflanzen sich von dem Sorghumfeld aus der Erntesaison 2018 unterscheiden. Ebenfalls ist es unwahrscheinlich, dass besagtes Feld mit den gleichen Krankheiten infiziert ist. Das rote Modell (s. Kapitel 5.3) reagierten negativ auf die NDVI-Daten und gaben keine Detektionen zurück. Wäre die Form für die Detektion ausschlaggebend, würde das Modell hier detektierte Objektinstanzen erzeugen.

Mask R-CNN wurde aufgrund der Performanz und der Fähigkeit zur Instanzsegmentierung ausgewählt. Die Instanzsegmentierung ermöglicht im idealen Fall die Klassifizierung einzelner Infektionen sowie die Eingrenzung und Unterscheidung der Instanzen. Um zuverlässig Krankheiten unterscheiden zu können, ist eine ausreichende Menge an Daten für jede separate Krankheit nötig. Aber wie es schon erwähnt wurde, sind historische Daten über Nutzpflanzenerkrankungen rar. Wenn in zukünftigen Forschungen eine größere Gesamtmenge an Aufzeichnungen zur Verfügung stehen, ist es wahrscheinlich, dass die Teilmenge der einzelnen Klassen gering ausfällt. Folglich ergibt sich daraus wieder die Problematik des Overfittings. In Kapitel 3 wurde deswegen darauf verzichtet, die verfügbaren Daten in Anthraknose und Streifenkrankheit zu unterteilen. Weitere Forschungen sollten diesen Ansatz weiter verfolgen, sollten die Teilmengen der einzelnen Klassen nicht ausreichend sein. Durch die binäre Klassifizierung (infiziert oder nicht infiziert) kann auch der Ansatz der semantischen Segmentierung untersucht werden. Zum Beispiel kann ein FCN-Modell auf die infizierten Flächen trainiert werden. Der Zweig des Mask R-CNN, der für die Maskengenerierung verantwortlich ist, wird mit einer FCN-Architektur gebildet. Die Maskenerkennung wird auf die vorgeschlagenen *RoIs* angewandt, liefert Ergebnisse die vergleichweise genau mit den binären Ausgangsmasken übereinstimmen.

In weiteren Untersuchungen sollte auch das Einwirken von Störfaktoren wie Wolken über der Zielregion untersucht werden. In dem wahrscheinlichen Falle, dass Felder über eine Erntesaison, die mehrere Wochen oder Monate dauern kann, überwacht werden, ist damit zu rechnen, dass die Wetterverhältnisse nicht immer wolkenfreie Aufnahmen zulassen. Sentinel-2-Produkte enthalten neben den Spektralaufnahmen binäre Wolkenmasken, die dazu genutzt werden können Wolken zu lokalisieren. Es gibt neben dem NDVI andere Vegetationsindizes wie den *Enhanced Vegetation Index* (EVI, dt.: verstärkter Vege-

tationsindex), die ebenfalls evaluiert werden können.

Kapitel 6

Fazit

Ziel der Arbeit war es ein heuristisches Modell zu trainieren, das in der Lage ist anhand von Satellitenaufnahmen infizierte Agrarfläche zu analysieren und eventuelle Krankheitsbefäle zu erkennen.

Das Skript, das im Rahmen dieser Abschlussarbeit entwickelt wurde, fasst mehrere Disziplinen zusammen. Deswegen war eine ausführliche Ausarbeitung in Thematiken wie pflanzliche Biologie, Lichtspektren, geografische Referenzsysteme und künstliche neuronale Netze notwendig, um ein tiefgehendes Verständnis für die Problematik und für den Zusammenhang zwischen Lichtreflexion einer Pflanze und deren Stress- bzw. Gesundheitsstatus zu erlangen. Gesunde bzw. gestresste Pflanzen besitzen einen hohen bzw. niedrigen Chlorophyllspiegel und dieser sorgt dafür, dass die Pflanzen im nahen Infrarotbereich stärker bzw. schwächer zurückstrahlen. Diese Reflexionen können über Sentinel-2-Multispektralaufnahmen gemessen werden und daraus lässt sich ein Vitalitätsindikator NDVI berechnen.

Die berechneten NDVI-Werte sollten in einem bewachten Lernverfahren ein neuronales Netz trainieren. Jedoch war früh ersichtlich, dass historische Daten rar sind und dadurch war auch mit Overfitting zu rechnen. Es wurden mehrere Methoden untersucht, um einem Overfitting entgegen zu wirken wie etwa Vergrößerung des Datensatzes durch Data Augmentation oder Hinzufügen eines weiteren Hyperparameters durch Ridge Regression. Um die

Als zu trainierendes KNN wurde die Mask R-CNN-Implementierung ausgewählt, da es unter anderem eine hohe Robustheit aufweist und Objektinstanzen auf Pixelebene klassifizieren kann.

Nachdem Modell und Methoden zur Modellevaluation und gegen Overfitting ausgesucht wurden, wurde ein Prozessablauf ausgearbeitet, der automatisch Sentinelprodukte in einem vorherdefinierten Zeitraum und einer geographischen Umgebung herunterlädt, verarbeitet und das Training bzw. Detektionen ausführt. Darauf wurde der konzipierte Prozess implementiert und das Modell durch anschließende Experimente evaluiert.

Die Experimente zeigten positive Ergebnisse durch Data Augmentation. Die anfangs trainierten Modelle ohne Data Augmentation lieferten bei leichten Veränderungen der bekannten Daten wie Rotationen oder Spiegelungen keine brauchbaren Resultate. Währenddessen sind mit Data Augmentation trainierte Modelle dagegen resistenter. Allerdings zeigte die L2 Regularization keinerlei Auswirkung.

Es ist möglich, mit der Mask R-CNN-Architektur ein Modell trainieren zu können, das zuverlässig die hier untersuchten Krankheiten erkennt und eingrenzen kann. Fremde Pathogene werden von diesem Modell wahrscheinlich nicht erkannt, da jede Krankheit und folglich die Auswirkung auf die Pflanze unterschiedlich ist. Es sind weitere Untersuchungen mit zusätzlichen Daten notwendig, die zum Abschluss der Arbeit nicht verfügbar waren. Zum einen um die in Kapitel 5.3 beschriebenen Schwankungen ausgleichen zu können. Zum anderen um einen Datensatz zu haben, der ein fremdes mit der selben Krankheit infiziertes Feld enthält, der dazu dient die hier erlangten Ergebnisse zu bestätigen.

Kapitel 7

Ausblick

Gegen Ende der Ausarbeitung gaben Mitarbeiter des CREA bekannt, im Laufe des Jahres 2019 weitere Felder auf Infektionen untersuchen zu wollen und diese entsprechend weiterzuleiten. Auf Basis dieser Daten können zukünftig weitere Experimente definiert, durchgeführt und evaluiert werden. Um mehr Satellitendaten in einem Zeitraum zu erhalten, können andere Satelliten wie etwa SPOT-5 angesprochen werden. Wichtig hierbei ist, dass die Satelliten eine gleiche oder höhere räumliche Auflösung besitzen und Aufnahmen im roten und nahen infraroten Bereich machen können.

Literaturverzeichnis

- [1] W. Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017. [Zuletzt besucht: 06.01.2019].
- [2] 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-7c761>, 2018. [Zuletzt besucht: 22.12.2018].
- [3] Anuja Nagpal. L1 and l2 regularization methods. <https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c>, 2017. [Zuletzt besucht: 09.01.2019].
- [4] T. Arlen. Understanding the map evaluation metric for object detection. <https://medium.com/@timothycarlen/understanding-the-map-evaluation-metric-for-object-detection-a07fe6962cf3>, 2018. [Zuletzt besucht: 28.01.2019].
- [5] Bharath Raj. Data augmentation | how to use deep learning when you have limited data - part 2. <https://medium.com/nanonets/how-to-use-deep-learning-when-you-have-limited-data-part-2-data-augmentation-5a2f3a2a2a2>, 2018. [Zuletzt besucht: 09.01.2019].
- [6] H. Butler, M. Daly, A. Doyle, S. Gillies, S. Hagen, and T. Schaub. The GeoJSON Format. RFC 7946, IETF, August 2016.
- [7] A. Chemura, O. Mutanga, and T. Dube. Separability of coffee leaf rust infection levels with machine learning methods at sentinel-2 msi spectral resolutions. *Precision Agriculture*, 18:859–881, 2016.
- [8] C. Consortium. Coco 2018 object detection task. <http://cocodataset.org/#detection-2018>, 2018. [Zuletzt besucht: 06.01.2019].

- [9] C. Consortium. Detection evaluation. <http://cocodataset.org/#detection-eval>, n.d. [Zuletzt besucht: 28.01.2019].
- [10] Copernicus. Copernicus in brief. <https://www.copernicus.eu/en/about-copernicus/copernicus-brief>. [Zuletzt besucht: 15.12.2018].
- [11] N. Corporation. Convolutional neural network (cnn). <https://developer.nvidia.com/discover/convolutional-neural-network>, 2019. [Zuletzt besucht: 04.01.2019].
- [12] 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].
- [13] ESA. Radiometric resolutions. <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>, 2018. [Zuletzt besucht: 20.12.2018].
- [14] ESA. Resolutions. <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions>, 2018. [Zuletzt besucht: 20.12.2018].
- [15] ESA. Satellite description. <https://earth.esa.int/web/sentinel/missions/sentinel-2/satellite-description>, 2018. [Zuletzt besucht: 09.02.2019].
- [16] 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.
- [17] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.
- [18] 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.
- [19] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CoRR*, abs/1611.10012, 2016.

- [20] 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].
- [21] J. Hui. map (mean average precision) for object detection. https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173, 2018. [Zuletzt besucht: 25.01.2019].
- [22] Jeremy Jordan. Evaluating image segmentation models. <https://www.jeremyjordan.me/evaluating-image-segmentation-models/>, 2018. [Zuletzt besucht: 27.01.2019].
- [23] 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.
- [24] NASA. Measuring vegetation (ndvi & evi). https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php, 2000. [Zuletzt besucht: 13.12.2018].
- [25] Prashant Gupta. Regularization in machine learning. <https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a>, 2017. [Zuletzt besucht: 09.01.2019].
- [26] N. A. Pugh, X. Han, S. D. Collins, J. A. Thomasson, D. Cope, A. Chang, J. Jung, T. S. Isakeit, L. K. Prom, G. Carvalho, I. T. Gates, A. Vree, G. C. Bagnall, and W. L. Rooney. Estimation of plant health in a sorghum field infected with anthracnose using a fixed-wing unmanned aerial system. *Journal of Crop Improvement*, 32(6):861–877, 2018.
- [27] 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.
- [28] T. Shah. About train, validation and test sets in machine learning. <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>, 2017. [Zuletzt besucht: 16.01.2019].

- [29] Stefan Parsch. So fatal sind schädlinge und krankheiten für die ernte. <http://www.spiegel.de/wissenschaft/mensch/landwirtschaft-so-fatal-sind-schaedlinge-und-krankheiten-a-1251826.html>, 2019. [Zuletzt besucht: 09.02.2019].
- [30] B. Tsedale, G. Adugna, and F. Lemessa. Distribution and importance of sorghum anthracnose (*colletotrichum sublineolum*) in southwestern and western ethiopia. *Plant Pathology Journal*, 15:75 – 85, 2016.
- [31] 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/algorithms>, n.d. [Zuletzt besucht: 24.12.2018].
- [32] 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.
- [33] F. J. Zeller. Sorghumhirse (*sorghumbicolorl.moench*): Nutzung, genetik, züchtung. *Bodenkultur*, 51:71 – 85, 2000.

Abbildungsverzeichnis

1.1	Sorghum-Anthraknose[30, S. 77]	4
1.2	Schematische Ansicht Sentinel-2[15]	5
2.1	Region of Interest	7
2.2	Künstliches neuronales Netz	11
2.3	CNN	12
2.4	Instanzsegmentierung	13
2.5	Mask R-CNN-Architektur	14
2.6	RPN-Anker	15
2.7	FCN-Architektur	16
2.8	Mask R-CNN vs. FCIS	16
2.9	IoU	18
2.10	Precision-Recall-Kurve	19
3.1	Beispiel Overfitting	22
4.1	Gesamtablauf der Anwendung	25
4.2	Ablaufdiagramm Sentineldatenaufbereitung	28
4.3	Ablaufdiagramm der Aufbereitung	30
4.4	B4 - B8 - NDVI	31
4.5	Ablaufdiagramm der Datensatzaufteilung	32
4.6	Datensatzverteilung	33
4.7	Ablaufdiagramm des Trainings	40
4.8	Ablaufdiagramm der Erkennung	41
5.1	<i>mAP</i> -Graph von Experiment 1, X-Achse: Epochennummer, Y-Achse: <i>mAP</i> -Werte	44
5.2	Beispielvorhersagen Experiment 1	45
5.3	<i>loss</i> -Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: <i>loss</i> -Werte	46

5.4	<i>mAP</i> -Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: <i>mAP</i> -Werte	46
5.5	Beispielvorhersagen Experiment 2	47
5.6	<i>loss</i> -Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: <i>loss</i> -Werte	48
5.7	<i>mAP</i> -Graph von Experiment 2, X-Achse: Epochennummer, Y-Achse: <i>mAP</i> -Werte	48
5.8	Beispielvorhersagen Experiment 2	49

Tabellenverzeichnis

2.1 Räumliche und spektrale Auflösungen von Sentinel-2A[13]	9
2.2 Mask R-CNN im Vergleich	17

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zu widerhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.

Ich versichere, dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Ort:

Datum:

Unterschrift: