



FOM Hochschule für Oekonomie & Management

Hochschulzentrum Münster

Hausarbeit

im Studiengang Big Data & Business Analytics

zur Erlangung des Grades eines

Master of Science (M. Sc.)

über das Thema

**Semantische Segmentierung von Satellitenbildern auf Basis neuronaler
Netzwerke**

von

Fiete Ostkamp, Verena Rakers und Artur Gergert

Betreuer : Dipl. Ing. Mustafa Er

Matrikelnummer : 557851, 536491 , 562394

Abgabedatum : 13. November 2021

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
Symbolverzeichnis	VI
1 Einleitung	1
1.1 Einordnung in den Kontext	1
1.2 Anwendungsfelder	1
1.2.1 Google, Here, Apple Maps, OSM	1
1.2.2 Humanitäre Hilfe	1
1.2.3 Wissenschaftliche Flächenbeobachtung (Regenwald)	1
2 Grundlagen	2
2.1 Grundlagen von [Semantic Segmentation]	2
2.2 Satellitenbilder als Datengrundlage	2
2.3 Neuronale Netzwerkarchitekturen zu Segmentierung von Satellitenbildern .	5
2.4 Jaccard Score	7
3 Praktische Umsetzung	8
4 Kritische Betrachtung/Fazit	8
4.1 In Bezug auf Mehrwert	8
4.2 Ausblick auf produktiven Einsatz	8
Anhang	9
Literaturverzeichnis	10

Abbildungsverzeichnis

1	Elektromagnetisches Spektrum	3
2	RGB vs. 16-Band Bild	3
3	Vergleich multispektraler und hyperspektraler Bildaufnahmen	4
4	(a) Datenwürfel eines multispektralen Bildes, (b) Spektrum des Pixels $P(i, j)$	4
5	U-Net	7

Tabellenverzeichnis

Abkürzungsverzeichnis

CNN	Convolutional Neural Network
FCN	Fully Convolutional Neural Network

Symbolverzeichnis

1 Einleitung

1.1 Einordnung in den Kontext

- Thema
- Kontext

1.2 Anwendungsfelder

- Zielgruppe
- Mehrwert

1.2.1 Google, Here, Apple Maps, OSM

1.2.2 Humanitäre Hilfe

1.2.3 Wissenschaftliche Flächenbeobachtung (Regenwald)

2 Grundlagen

2.1 Grundlagen von [Semantic Segmentation]

2.2 Satellitenbilder als Datengrundlage

Die Datenquellen für die semantische Segmentierung erzeugen zum Großteil große Satellitensysteme aus dem Weltraum heraus. Von dortaus können schnell und kostengünstig Daten über große Gebietsflächen gesammelt werden.¹ Diese Fernerkundungssysteme sind in der Lage durch Lichtstrahlung Informationen von Objekten aus unterschiedlichen Dimensionen heraus zu sammeln. Satelliten sind so in der Lage Bilder eines selben Objektes oder einer Perspektive zu generieren, welches sich durch die Strahlung auf unterschiedlichen Wellenlängenbändern unterscheidet. Wie in Abbildung 1 dargestellt nimmt das menschliche Auge lediglich einen kleinen Bereich des elektromagnetischen Spektrums wahr, welcher sich aufteilt in einen roten, einen grünen und einen blauen Bereich. Bilder, die von der Farbgebung so aussehen, wie das menschliche Auge das abgebildete Objekt auch in der Natur wahrnimmt, sind mithilfe des roten, grünen und blauen Wellenbereichs erzeugt worden. Aus diesem Grund werden diese Bilder auch häufig RGB-Bilder genannt.² Bilder aus anderen Spektralbereichen, die das menschliche Auge nicht wahrnehmen kann, enthalten jedoch weitreichende Informationen zur Identifikation diverser Objekte aus der Landwirtschaft, Lebensmittelproduktion, städtische sowie außerstädtische Gebiete, Öl- und Mineraler Exploration etc.³ Abbildung 2 zeigt exemplarisch ein RGB-Bild mit einem Bild auf Basis von sechzehn Spektralbändern.⁴ Ein weiterer Grund dafür, dass für die Datengewinnung in Form derartiger Bildaufnahmen auf Fernerkundungssysteme zurückgegriffen wird, ist die synoptische und ganzheitliche Sicht auf die Erde. Von der Position aus dem All können so kostengünstig Daten unterschiedlichster Positionen der Erde erzeugt werden.⁵ Um aus dieser Position Bilder zu generieren sind Satellitensysteme mit Sensoren ausgestattet. Die Sensoren werden häufig unterschieden in

- Multispektrale Sensoren
- Hyperspektrale Sensoren

Multispektralsensoren sind in einer parallelen Anordnung am Satellitensystem angebracht und messen häufig zwischen drei und sechs Spektralbänder im sichtbaren bis mittleren

¹ Landgrebe, D., 1997, S. 2.

² \leavevmode{\color {red}Hier muss noch eine Quelle hin}.

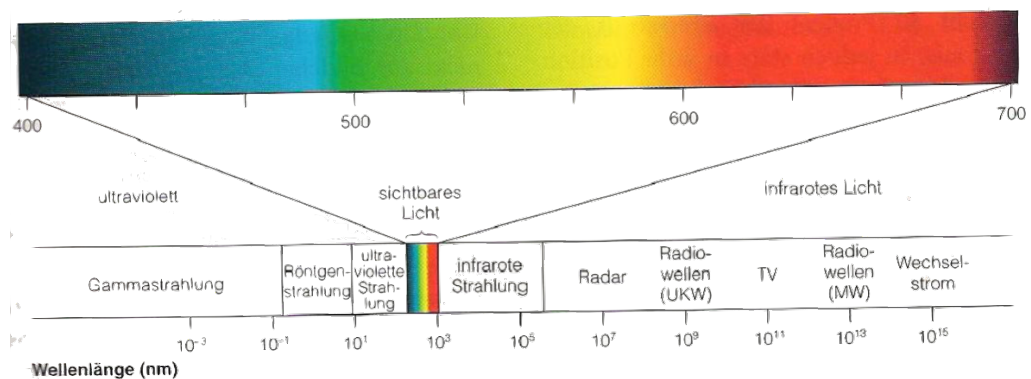
³ Vgl. Landgrebe, D., 1997, S. 2.

⁴ Hier noch die 16 Spektralbänder ergänzen

⁵ Vgl. Landgrebe, D., 1997, S. 2.

Infrarotbereich des elektromagnetischen Spektrums, während hyperspektrale Fernerkundungssensoren in der Lage sind viele, sehr schmale zusammenhängende Spektralbänder im sichtbaren, nahen und mittleren und thermischen Infrarotbereich des elektromagnetischen Spektrums zu erfassen.⁶ Generel wird im Bereich zwischen zwei und zehn Spektralbändern noch von multispektralen Systemen gesprochen, während alle Bilder, die Informationen aus mehr als zehn Spektralbändern enthalten, von einem hyperspektralen System erzeugt worden sind.⁷ Abbildung 3 stellt multispektrale und hyperspektrale Bilder vergleichend gegenüber. Um die Bildinformationen zu speichern, müssen drei Dimensionen für jedes Pixel gespeichert werden. In der Abbildung 4 wird der dreidimensionale Datenwürfel $I(x, y, \lambda)$ illustriert. Die Koordinaten x, y beinhalten die räumlichen Informationen des Bildes und die dritte Dimension λ speichert die Daten des Spektralbandes mit der Dichte I .⁸

Abbildung 1: Elektromagnetisches Spektrum



Quelle: Ditzinger, T., 2013, S. 7

Abbildung 2: RGB vs. 16-Band Bild

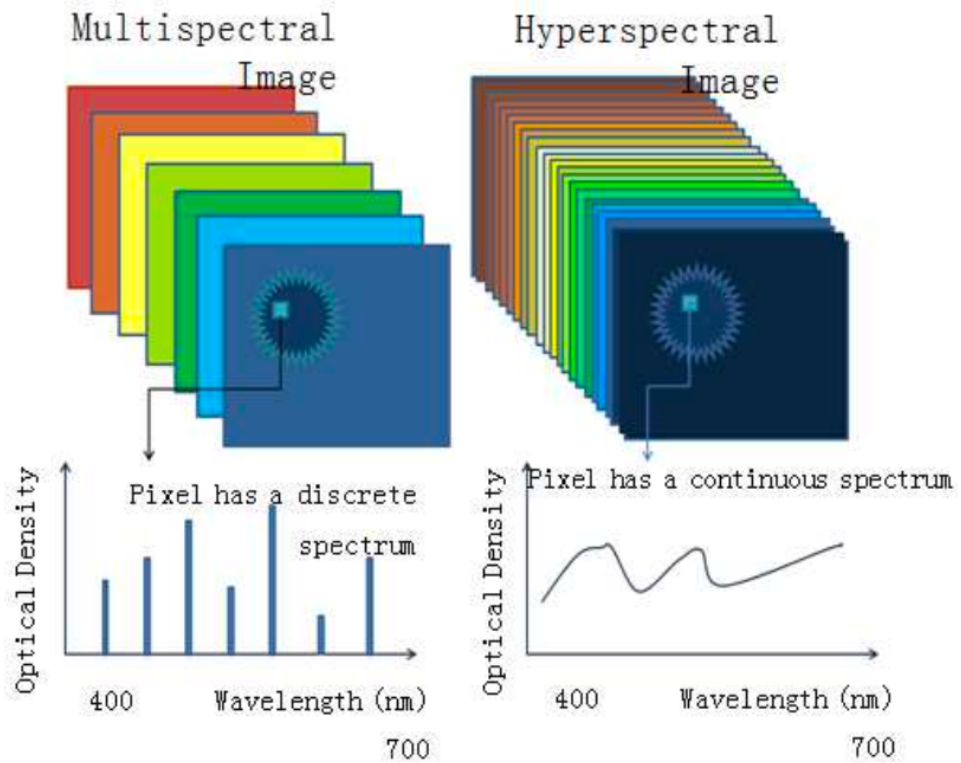
Hier muss noch ein Beispielbild rein...vermutlich aus unserem Datensatz

⁶ Vgl. Govender, M., Chetty, K., Bulcock, H., 2007, S. 1.

⁷ Vgl. Ibraheem, I., 2015, S. 2.

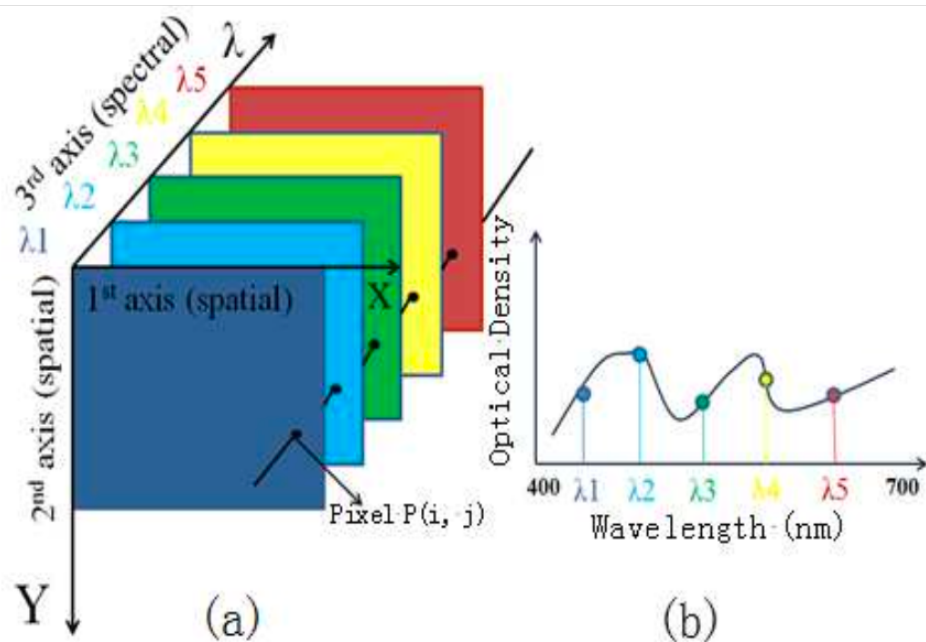
⁸ Vgl. ebd., S. 2.

Abbildung 3: Vergleich multispektraler und hyperspektraler Bildaufnahmen



Quelle: Ibraheem, I., 2015, S. 2

Abbildung 4: (a) Datenwürfel eines multispektralen Bildes, (b) Spektrum des Pixels $P(i, j)$



Quelle: ebd., S. 3

Die Informationen werden zeilenweise oder bandweise gespeichert.⁹ Bei dem Zeilenweisen speichern der Bilder wird ein $M * N$ -Bild mit K -Bändern enthält die Bilddatei $M * K$ Zeilen und N Spalten. Die ersten K Zeilen der Bilddatei entsprechen dabei der ersten Pixelreihe der Aufnahme, die nächsten K Zeilen der zweiten Pixelreihe, usw. Bei der bandweisen Speicherung der Bilder werden die gesamten Bildinformationen je Band nacheinander gespeichert. Es wird also jeweils eine $M * N$ für das erste Band untereinander geschrieben, dann die $M * N$ -Matrix für das zweite Band, bis zum K -ten Band.

Gespeichert werden die Dateien häufig in Formaten die *.mat oder *.tif. Bei diesen Formaten ist die Möglichkeit gegeben zusätzlich zu den Bildinformationen geographische Informationen wie beispielsweise Lagedaten in Form von Koordinaten zu speichern.

2.3 Neuronale Netzwerkarchitekturen zu Segmentierung von Satellitenbildern

Für die maschinelle Verarbeitung von Bildern kommen häufig neuronale Netzwerkstrukturen zum Einsatz, da mit solchen in der Vergangenheit bahnbrechende Ergebnisse auf diesem Gebiet erzielt werden konnte.¹⁰ Die spezielle Art der Netzwerkarchitektur für solche Anwendungsgebiete sind die Convolutional Neural Network (CNN)s. Die Idee bei CNNs ist es, einen oder meist mehrere rechteckige „Filter“ über ein Bild schieben, was im mathematischen Sinne einer Faltung bzw. einer Convolution entspricht. Ziel ist es die Gewichte der Filter so optimal zu trainieren, dass jeder Filter jeweils ein bestimmtes Merkmal eines Bildes erkennen kann. Je mehr Filter das neuronale Netzwerk also hat, desto mehr Merkmale kann es extrahieren und damit komplexere Muster lernen. Der Filter wird wie oben bereits erwähnt von den Gewichten repräsentiert, die zu trainieren sind. Um die Rechenanforderungen zu reduzieren, wird die Größe des Filters im Laufe eines Netzwerks in der Regel kleiner, während ihre Anzahl jedoch steigt, sodass Merkmale auf granularerer Ebene gelernt werden können.

Der ursprüngliche Zweck der CNN-Architektur ist es einem Bild eine Klasse zuzuordnen. Wenn Bilder mehrere Klassen enthalten, muss es zusätzlich ermöglicht werden die Größe und Lokalisierung der jeweiligen Klassen innerhalb des Bildes zu erhalten. Gleichzeitig muss das Netzwerk tief genug sein, um die einzelnen Klassen zu „lernen“, damit es zwischen den Klassen unterscheiden kann. Es kann also nicht die reine downsampling Architektur verwendet werden, wie sie es im klassischen CNN der Fall ist, sondern die Informationen an welcher Stelle welche Objektklasse lokalisiert sind, müssen ebenfalls

⁹ Vgl. Upadhyay, P., Gupta, S., 2012, S. 2.

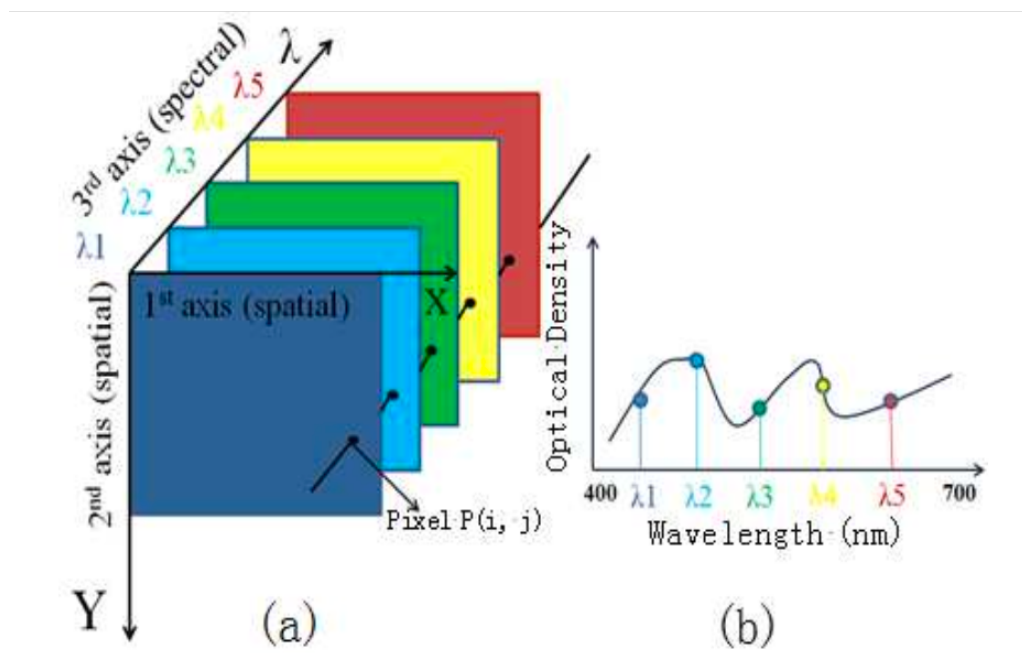
¹⁰ Vgl. Pritt, M., Chern, G., 2020, S. 1.

gegeben sein. Das soeben geschilderte Problem wird durch die Idee des Fully Convolutional Neural Network (FCN) weitestgehend gelöst. Während bei einem reinen CNN die erste Schicht der Größe des Bildes entsprechen muss, da es fest mit dem Eingabebild verbunden ist, wird bei einem FCN das Modell ab der ersten Schicht bereits faltbar gemacht, sodass das Modell auch gleichzeitig unabhängig von der Größe der Inputdaten ist und damit mehr Flexibilität gewährleistet. Des Weiteren werden durch eine vollständig verbundene Schicht globale Bildinformationen verarbeitet, was sich generell für eine Klassifizierungsaufgabe gut eignet. Wenn es jedoch darum geht das Bild zu segmentieren sind die kleineren Faltungsschichten, die von vorn herein über das Bild gleiten können mehr von Vorteil. Zusammenfassend werden bei FCNs also die voll verknüpften Schichten der CNNs durch Faltungsschichten ersetzt. Abbildung ?? verdeutlicht die Funktionsweise eines CNN, welches ein reines Katzenbild der Klasse „Cat“ zuordnet, während das FCN dazu in der Lage die Klasse „Cat“ in einem Bild mit mehreren Klassen zu lokalisieren. Ebenfalls ist in der Abbildung ?? ersichtlich, dass die Auflösung der Heatmap nicht der des Eingabebildes entspricht. Der nächste Schritt ist demnach die grobe Merkmalszuordnung möglichst in die Ursprungsauflösung zurückzuübersetzen. Dieser Schritt ist in der Literatur oft als „learned upsampling“ bezeichnet. Nachdem beim downsampling die Größe der Schichten immer kleiner geworden sind, um eine gewisse Detailtiefe beim Lernen zu trainieren, ist das Vorgehen beim upsampling genau andersherum, um die Merkmalskarte auf die Ursprungsgröße des Ausgangsbildes zu bringen. Während beim downsampling die Filter über die Eingabedaten gleiten und Punktprodukte an jeder Position berechnen und jeweils einen Ausgabewert weiterleiten, wird jeder Filterwert beim upsampling mit einem Eingabepixel multipliziert, über dem der Filter positioniert ist. Die Ergebnisse werden anschließend auf die Ausgabe-Merkmalskarte projiziert. Filterprojektionen, die sich in der Ausgabe überschneiden werden in der regel addiert. Abbildung ?? veranschaulicht die Vorgehensweise beim upsampling. Sowohl im downsampling als auch im upsampling werden also die Filter vom neuronalen Netz trainiert. Im Ergebnis wird also die grobe Ausgabe wieder in Pixel übersetzt. Die Ergebnisse dabei sind jedoch ungenau und nicht trennscharf, da nur das Hinzufügen einer Upsamplingschicht allein den großen Schritt von der Detailtiefe des Downsamplings hin zum Ausgangsformat nicht bewältigen kann. Die Trennschärfe wird durch die sogenannte „Skip-Layer-Fusion“ gelöst. Hier werden durch Skip-Verbindungen, die eine Fusion zwischen nicht benachbarten Schichten darstellen, die Informationen genutzt, um den räumlichen Kontext, der beim detailreichen Lernen einzelner Klassen verloren geht, mehr zu berücksichtigen bzw zu übertragen. Dadurch soll das Spannungsverhältnis zwischen Detailtiefe und Lokalisierung versucht werden zu lösen. In der Abbildung ?? wird das gesamte Vorgehen einmal veranschaulicht. Hier ist das downsampling vom Ausgangsbild bis zur letzten Faltungsschicht, die nur noch eine Dimen-

sion hat dargestellt. Der letzte Schritt ist das upsampling, welches in der Abbildung in der obersten Reihe zu aller erst ohne Skip-Layer Fusion dargestellt ist. Dort ist zu erkennen, dass die Projektion der Merkmalskarte grob funktioniert. In der zweiten Zeile werden die Informationen aus der vorangehenden pool4 Schicht mit berücksichtigt, welches im Ergebnis einen höheren Detailierungsgrad aufweist. In der dritten Zeile werden die pool3 und die pool4 Schicht berücksichtigt, welche die Segmentierungskarte noch genauer machen.

Eine Weiterentwicklung des FCNs ist das U-Net. Dieses Modell soll die oftmals kritisierte Ungenauigkeit der FCNs an den Segmentierungsgrenzen beheben bzw. optimieren. Die Architektur der U-Net sieht aus wie der Buchstabe „U“ und ist im Folgenden exemplarisch dargestellt.

Abbildung 5: U-Net



Hier muss noch das U-Net rein

2.4 Jaccard Score

3 Praktische Umsetzung

4 Kritische Betrachtung/Fazit

4.1 In Bezug auf Mehrwert

4.2 Ausblick auf produktiven Einsatz

Anhang

Anhang 1: Beispielanhang

Dieser Abschnitt dient nur dazu zu demonstrieren, wie ein Anhang aufgebaut sein kann.







Anhang 1.1: Weitere Gliederungsebene

Auch eine zweite Gliederungsebene ist möglich.

Anhang 2: Bilder

Auch mit Bildern. Diese tauchen nicht im Abbildungsverzeichnis auf.

Abbildung 6: Beispielbild

Name	Änderungsdatum	Typ	Größe
 abbildungen	29.08.2013 01:25	Dateiordner	
 kapitel	29.08.2013 00:55	Dateiordner	
 literatur	31.08.2013 18:17	Dateiordner	
 skripte	01.09.2013 00:10	Dateiordner	
 compile.bat	31.08.2013 20:11	Windows-Batchda...	1 KB
 thesis_main.tex	01.09.2013 00:25	LaTeX Document	5 KB

Literaturverzeichnis

Ditzinger, Thomas (2013): Illusionen des Sehens, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, [Zugriff: 2021-10-16]

Govender, M., Chetty, K., Bulcock, H. (2007): A Review of Hyperspectral Remote Sensing and Its Application in Vegetation and Water Resource Studies, in: Water SA, 33 (2007), Nr. 2, [Zugriff: 2021-10-16]

Ibraheem, Issa (2015): Early Detection of Melanoma Using Multispectral Imaging and Artificial Intelligence Techniques, in: American Journal of Biomedical and Life Sciences. Special Issue: Spectral Imaging for Medical Diagnosis "Modern Tool for Molecular Imaging", Vol. 3 (2015), S. 29–33

Landgrebe, David (1997): On Information Extraction Principles for Hyperspectral Data, in: (1997)

Upadhyay, Pragati, Gupta, Sudha (2012): Introduction To Satellite Imaging Technology And Creating Images Using Raw Data Obtained From Landsat Satellite, in: 1 (2012), S. 41–45

Internetquellen

Pritt, Mark, Chern, Gary (2020): Satellite Image Classification with Deep Learning, arXiv: 2010.06497 [cs], <<http://arxiv.org/abs/2010.06497>> (2020-10-13) [Zugriff: 2021-10-13]

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe. Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde/Prüfungsstelle vorgelegen hat. Ich erkläre mich damit **einverstanden/nicht einverstanden**, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird. Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Münster, 13.11.2021

(Ort, Datum)

A handwritten signature in black ink, consisting of a large, stylized 'H' followed by a series of loops and a final flourish.

(Eigenhändige Unterschrift)