KATHOLISCHE UNIVERSITÄT
EICHSTÄTT-INGOLSTADT

Mathematisch-Geographische Fakultät

Deutsches Zentrum
für Luft- und Raumfahrt
German Aerospace Center

Earth Observation Center
Deutsches Fernerkundungsdatenzentrum
Abteilung Georisiken und zivile Sicherheit
Team Stadt und Gesellschaft

CNN-basierte semantische Segmentierung von Gebäudetypen mittels hochaufgelöster Luftbilder und normalisierten digitalen Oberflächenmodellen

CNN-based semantic segmentation of building types using high-resolution aerial imagery and normalized digital surface models

Masterarbeit

Abgabedatum: 11.01.2024

Verfasser

Moritz Hertrich

Mail: Moritz.Hertrich@stud.ku.de

Matr. Nr.: 277992

Studiengang:

M. Sc. Geographie: Umweltprozesse und Naturgefahren

Gutachter der KU

Apl. Prof. Dr. Tobias Heckmann

Betreuerin am DLR

Dorothee Stiller

Zusammenfassung

Informationen über den Typ eines Gebäudes mittels Fernerkundungsdaten zu ermitteln ist für viele Anwendungen von essentieller Bedeutung. Hierzu zählt insbesondere die Risikoanalyse im Naturgefahrenbereich, in welcher der Wert eines gefährdeten Gebäudes häufig durch den Gebäudetyp bedingt ist. Fernerkundungsdaten als Grundlage für die Erkennung dieser Gebäudetypen zu verwenden ist vorteilhaft, da diese meist flächendeckender und aktueller vorliegen, als im Feld manuell kartierte Gebäude. Zudem ist mit einer Methode zur automatischen Ableitung von Gebäudetypen der Arbeitsaufwand erheblich reduziert.

In dieser Arbeit werden mehrere Gebäudetypen, definiert nach ihrer Form und ihrer Nutzung, gebildet und es wird getestet, wie gut diese mit Hilfe der RGB-Kanäle sowie dem nahen Infrarot von hochaufgelösten Luftbildern und einem normalisierten digitalen Oberflächenmodell (nDOM) erkennbar sind. Als Methode werden hierfür aktuelle Convolutional Neural Networks (CNNs) mit einer Encoder-Decoder Struktur verwendet. In einem zweiten Schritt wird zudem systematisch getestet, wie unterschiedliche Einstellungen der wichtigsten Parameter die Erkennung der Gebäudetypen verändern. Diese Parameter umfassen Eigenschaften der Daten (Untersuchungsgebiete, Auflösungen, Datenkanäle, Anzahl der Daten, Auswahl spezifischer Daten) und der CNNs (Modellarchitekturen, Epochenanzahl, vortrainierte Modelle, Zufallszahlen in dem Modell).

Die Ergebnisse zeigen, dass von den Gebäudetypen der Form besonders Reihenhäuser, Hauptgebäude und Nebengebäude am besten erkennbar sind. Von den nach ihrer Nutzung definierten Gebäudetypen sind Wohngebäude und Nicht-Wohngebäude am besten erfassbar, was insbesondere für die angesprochene Risikoanalyse von Vorteil ist. Die Analyse der Ergebnisse der Parameter-Tests zeigt, dass das Höhenmodell die Gebäudetyperkennung stark verbessert und das nahe Infrarot nur ohne das nDOM von Vorteil ist. Im Bereich der Parameter des Modells ist die Modellarchitektur von *FPN* am besten für die vorliegende Aufgabe geeignet und die Ergebnisse lassen sich durch die Verwendung von vortrainierten Gewichten innerhalb des Modells deutlich verbessern.

Abstract

Determining information about the type of a building using remote sensing data is essential for many applications. This includes, in particular, risk analysis in the field of natural hazards, where the value of a building at risk is often determined by the type of the building. Using remote sensing data as a basis for recognizing these building types is advantageous, as it is usually more comprehensive and up-to-date than manually mapped buildings in the field. In addition, a method for the automatic derivation of building types considerably reduces the work involved.

In this work, several building types, defined according to their shape and use, are formed and it is tested how well they can be recognized with the help of the RGB channels and the near infrared of high-resolution aerial images and a normalized digital surface model (nDSM). Current Convolutional Neural Networks (CNN) with an encoder-decoder structure are used as the method for this. In a second step, it is also systematically tested how different settings of the most important parameters change the recognition of the building types. These parameters include properties of the data (study areas, resolutions, data channels, number of data, selection of specific data parts) and the CNNs (model architecture, number of epochs, pre-trained models, random numbers in model).

The results show that of the building types by form, terraced houses, main buildings and outbuildings are the most recognizable. Of the building types defined by use, residential and non-residential buildings are the most recognizable, which is particularly advantageous for the risk analysis mentioned above. The analysis of the results of the parameter tests show that the elevation model greatly improves building type recognition and that the near infrared is only advantageous without the nDSM. In terms of model parameters, the *FPN* model architecture is best suited to the task at hand and the results can be significantly improved by using pre-trained weights within the model.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

CNN	Convolutional Neural Network
DL	Deep Learning
DLR	Deutsches Zentrum für Luft und Raumfahrt e.V.
DOM	Digitales Oberflächenmodell
DOP	Digitales Orthophoto
FCN	Fully Convolutional Network
GPU	Graphics Processing Unit
IoU	Intersection over Union
LiDAR	Light Detection and Ranging
LoD	Level of Detail
ML	Maschinelles Lernen
nDOM	Normalisiertes Digitales Oberflächenmodell
NIR	Nahes Infrarot
OA	Overall Accuracy
o. A.	Ohne Angabe
RGB	Rot, Grün, Blau
TDOP	True Digital Orthophotos
UInt16	Unsigned Integer 16

1 Einleitung

Gebäude sind eines der wichtigsten Elemente von bebauter Landschaft. Als solche weisen sie zahlreiche Funktionen auf, deren Erkennung eine essentielle Aufgabe in zahlreichen Themengebieten ist. Ein Beispiel hierfür ist die Risikoanalyse im Naturgefahrenbereich, in welcher das Risiko aus dem Produkt der Gefahr, der Vulnerabilität und dem Wert des gefährdeten Objekts gebildet wird (VAN WESTEN et al. 2006). Im Fall von Naturgefahren, wie Lawinen oder Hochwasserereignissen, sind in dieser Formel häufig Gebäude das gefährdete Objekt, dessen Wert sich je nach vorhandener Gebäudefunktion stark verändert. Einer Scheune ist beispielsweise in den allermeisten Fällen ein anderer Wert zuzuweisen als einem Mehrfamilienhaus. Folglich spielt die Einteilung des Gebäudes in unterschiedliche Gebäudetypen eine große Rolle. Ebenso wichtig ist der Gebäudetyp unter anderem in der Erfassung amtlicher Katasterdaten und dem damit verbundenen Schließen von Informationslücken bei einem unvollständigen Gebäudedatensatz (HECHT et al. 2015), in der Stadtplanung für die Analyse des Gebäudebestandes (LU et al. 2014) oder in der Energiemodellierung für die Berechnung des Energiebedarfs (STRELTSOV et al. 2020).

Informationen über den Gebäudetyp können zwar im Feld manuell gesammelt werden, dies ist jedoch mit einem großen Arbeits- und Zeitaufwand verbunden (BANDAM et al. 2022, 46). Zudem sind solche Daten oft nicht einheitlich flächendeckend vorhanden, veraltet oder nicht frei verfügbar (HECHT et al. 2015). Diese Nachteile sind lösbar durch die Verwendung von Fernerkundungsdaten, welche häufig für eine große Fläche einheitlich vorliegen und im Fall von Satellitendaten oder Luftbildern zumeist ein definiertes Aufnahmeintervall besitzen. Besteht ein Algorithmus zur automatisierten Klassifizierung von Gebäuden mit diesen Datengrundlagen, so ist auch der Arbeitsaufwand zur Gebäudeklassifizierung geringer.

Durch die stetig anwachsende Rechenleistung in den letzten Jahren weisen besonders Deep-Learning (DL) Algorithmen für Aufgaben der Szenen-Klassifizierung, Objekterkennung oder Segmentierung von Fernerkundungsdaten starke Erfolge auf (MA et al. 2019). Deep Learning-Methoden sind ein Teilbereich des maschinellen Lernens beziehungsweise der künstlichen Intelligenz und basieren auf künstlichen neuronalen Netzen (KETKAR & MOOLAYIL 2021). Diese neuronalen Netze wiederum können komplexe Probleme lösen, bei denen eine explizite Modellierung schwierig oder nicht möglich ist. Insbesondere für die Verarbeitung von Bilddaten werden dabei sogenannte Convolutional Neural Networks (CNNs) verwendet. Gerade auch für die semantische Segmentierung im urbanen Raum mittels Fernerkundungsdaten zeigen solche CNNs sehr vielversprechende Ergebnisse (NEUPANE et al. 2021).

In der vorliegenden Arbeit werden solche CNNs zur semantischen Segmentierung von unterschiedlichen Gebäudetypen mittels hochaufgelöster Luftbilder und normalisierten digitalen Oberflächenmodellen (nDOM) verwendet. Ziel der Arbeit ist es hierbei, zum einen zu erarbeiten, welche Gebäudetypen, wie gut erkannt werden können und zum anderen wird aufgezeigt, wie ausgewählte Parameter der Daten und des Modells die Klassifizierung der Gebäude verändern.
In den folgenden Kapiteln von Kapitel 1 wird der aktuelle Forschungsstand zum Thema der vorliegenden Arbeit analysiert und das Ziel dieser Arbeit wird näher dargestellt. In Kapitel 2 wird anschließend auf CNNs und deren Eignung für die Analyse von Bilddaten eingegangen. Kapitel 3 stellt die verwendeten Daten sowie die Methodik zum Erreichen des Ziels der Arbeit dar. Im Anschluss werden die Ergebnisse in Kapitel 4 dargestellt, welche in Kapitel 5 diskutiert werden.

1.1 Einordnung der Erkennung von Gebäudetypen im wissenschaftlichen Kontext

Im folgenden Kapitel wird der aktuelle Forschungsstand zur semantischen Segmentierung von Gebäudetypen mittels Fernerkundungsdaten dargestellt. Hierbei wird zunächst auf die zumeist verwendeten Datengrundlagen sowie die analysierten Gebäudetypen eingegangen (Kapitel 1.1.1). Hierauf folgt die Erkennung von Gebäudetypen mit traditionellen Methoden (Kapitel 1.1.2) und im Anschluss wird der aktuelle Forschungsstand zur Gebäudetypsegmentierung mit CNNs dargestellt (Kapitel 1.1.3). Auf dieser Wissensbasis baut die vorliegende Arbeit auf und es wird verdeutlicht, in welchem Bereich des Themengebiets es bislang nicht ausreichend Forschung gibt.

1.1.1 Datengrundlagen und Schwierigkeiten der Gebäudetypenerkennung in der Fernerkundung

Der Großteil der vorhandenen Literatur zu Gebäuden im Fernerkundungskontext beschäftigt sich mit der Gebäudedetektion, also der binären Unterscheidung in Gebäude oder Nicht-Gebäude (LI J. et al. 2022). Wie eingangs erwähnt, spielt in vielen Themengebieten, wie der Energiebedarfsmodellierung oder dem Naturgefahrenmanagement, insbesondere der Typ des Gebäudes eine wesentliche Rolle (STRELTSOV et al. 2020; BHUYAN et al. 2022).
In Studien zur Gebäudetypklassifizierung werden Gebäude nach bestimmten Charakteristika in unterschiedliche Klassen klassifiziert. Diese Charakteristika sind zumeist entweder der Gebäudemorphologie (HECHT et al. 2015; WURM et al. 2016) oder der Gebäudefunktion (BELGIU

et al. 2014; FAN et al. 2014; DU et al. 2015; CAO & QIU 2018; HOFFMANN et al. 2019; STRELTSOV et al. 2020) zuzuordnen. Beispiele für eine Einteilung nach der Morphologie bzw. der Geometrie sind die Klassen Reihenhaus, Doppelhaus oder Blockrandbebauung. Funktionelle Gebäudetypen betreffen die Nutzung der Gebäude, wie beispielsweise Wohngebäude, Wirtschaftsgebäude oder Industriegebäude. Weitere durchgeführte Unterteilungen von Gebäuden in Klassen betreffen die Dachform der Gebäude (CHEN et al. 2019; WANG et al. 2022) oder deren Höhe (HUANG et al. 2021).

Als Datengrundlage für die Erkennung unterschiedlicher Gebäudetypen in der Fernerkundung werden zumeist sehr hochaufgelöste Satelliten- oder Luftbilder, LiDAR-Daten bzw. daraus berechnete nDOMs verwendet oder, wie in der vorliegenden Arbeit, eine Kombination aus beiden Datenquellen (HUANG et al. 2017).

Die sogenannte Segmentierung dieser Daten ist in der Fernerkundung ein gängiges Vorgehen zur Verarbeitung solcher Rasterdaten. Bei der Segmentierung wird ein Bild in mehrere Segmente unterteilt, indem benachbarte Pixel mit ähnlichen Merkmalswerten (Helligkeit, Textur, Farbe etc.) gruppiert werden (MINAEE et al. 2021). Werden den einzelnen Pixeln zusätzlich Kategorien zugeordnet, so wird von semantischer Segmentierung gesprochen (GUO et al. 2018). CNNs können in diesem Zusammenhang für die semantische Segmentierung verwendet werden oder wie bei ZHAO et al. (2017) nur für die Segmentierung ohne semantische Zusammenhänge.

In komplexen urbanen Gebieten ist die Segmentierung von Gebäudetypen in multispektralen Fernerkundungsdaten nur mit Hilfe von typischen Bildmerkmalen sehr herausfordernd. Dies kann damit begründet werden, dass unterschiedliche Gebäudetypen oft ähnliche spektrale Eigenschaften besitzen können und dieselben Gebäudetypen sich wiederum in ihrem spektralen Erscheinungsbild unterscheiden. Ursache hierfür ist, dass die Verwendung der Baumaterialien oft nicht in Verbindung mit dem Gebäudetyp stehen. Darüber hinaus machen Kühltürme, Pavillons, Verzierungen etc. auf Dächern das Erscheinungsbild von Gebäuden in einem Bild mit hoher räumlicher Auflösung sehr komplex (YANG H. et al. 2018). Zudem weist die „Hintergrund“ Klasse, also alle Pixel ohne Gebäude, äußerst unterschiedliche spektrale Eigenschaften auf (Vegetation, Straßen, Autos etc.).

Wie bereits angesprochen, können unterstützend Höheninformationen verwendet werden, welche die Gebäudetyperkennung vereinfachen. So klassifizierten LU et al. (2014) Gebäude in Einfamilienhäuser, Mehrfamilienhäuser und Nichtwohngebäude nur unter Verwendung von *Light Detection and Ranging* (LiDAR) Daten und den hieraus errechenbaren Informationen über die Form und Höhe der Gebäude.

In verschiedenen Studien werden neben Fernerkundungsdaten auch zusätzliche Datengrundlagen für die Gebäudetyperkennung verwendet. Dazu gehören *Street View* Bilder, die häufig von *Google Street View* (GSV) (CAO & QIU 2018; HOFFMANN et al. 2019) stammen sowie eine Vielzahl von geolokalisierten Daten, zumeist aus Open Street Map (BANDAM et al. 2022; BHUYAN et al. 2022). Mit diesen zusätzlichen Informationen lässt sich die Gebäudetypdetektion häufig verbessern, da mehr Informationen vorliegen, welche alleinig auf Basis der aus dem Nadir aufgenommenen Fernerkundungsdaten nicht erkennbar sind. Diese Daten besitzen jedoch Nachteile, die bei Fernerkundungsdaten nicht vorhanden sind. Dies ist beispielsweise die regional sehr unterschiedliche Qualität und Quantität der Open Street Map Daten (HOFFMANN et al. 2019). GSV-Daten wiederum sind nur in begrenzter Menge kostenfrei verfügbar (GOOGLE LLC 2023) und besonders im ländlichen Raum nicht flächendeckend vorhanden.

1.1.2 Semantische Segmentierung von Gebäudetypen mit traditionellen Methoden des maschinellen Lernens

In dem Großteil der vorhandenen Studien zur Erkennung von Gebäudetypen auf Fernerkundungsdaten werden datenbasierte Methoden des maschinellen Lernens (*machine learning*, ML) verwendet, wobei auch regelbasierte Klassifizierungsalgorithmen Anwendung finden.
Regelbasierte Ansätze wie bei FAN et al. (2014), LU et al. (2014) oder HUSSAIN & SHAN (2016) basieren auf einer Reihe von Regeln in Bezug auf verschiedene Informationen, beispielsweise zur Gebäudetypologie oder zum Baujahr des Gebäudes, mit deren Hilfe Gebäudetypen klassifiziert werden. Nach KIM et al. (2022) sind diese Ansätze zwar intuitiv und meist leicht anwendbar, benötigen jedoch eine vollständige Datengrundlage zur Klassifizierung, wie sie häufig nicht vorhanden ist. Beachtet werden muss außerdem, dass das vorhandene Regelwerk nicht zu komplex werden sollte und der Anwender ausreichendes Wissen über das Fachgebiet besitzt. Durch datenbasierte Ansätze lassen sich diese Probleme teils lösen. Einen Vergleich führen HECHT et al. (2013) durch, indem sie mehrere Klassifizierungsmethoden zur Erkennung unterschiedlicher morphologischer Gebäudetypen miteinander vergleichen. Hierzu zählen klassische Algorithmen wie *k-nearest neighbor* oder Klassifikations- und Regressionsbäume, aber auch ML-Methoden. Die Analyse der Genauigkeitsmetriken zeigt, dass die ML-Methoden bessere Werte ergeben, was darauf zurückzuführen sein könnte, dass diese vorhandene Muster in den Daten besser erkennen können.
Zu den traditionellen überwachten Klassifizierungs-ML-Methoden zählt insbesondere der Random-Forest Algorithmus, welcher sich in den letzten Jahren zu einem der am häufigsten angewandten Ansätze für die automatisierte Erfassung bebauter Gebiete entwickelt hat (BEL-

GIU et al. 2014; DU et al. 2015; HECHT et al. 2015; HUANG et al. 2017; KIM et al. 2022). Ebenso verwendet werden in diesem Bereich Support Vector Machines (LU et al. 2014; HUANG et al. 2017) oder lineare Diskriminanzfunktionen (WURM et al. 2016). Zu beachten ist, dass die Anwendbarkeit von datenbasierten ML-Ansätzen in Gebieten mit starken regionalen Abhängigkeiten jedoch eingeschränkt ist. Das jeweilige Modell ist also nur in Regionen mit ähnlichen Gebäudestrukturen- und typen gut anwendbar, da es mit diesen trainiert wurde (HECHT et al. 2015).

Im Gegensatz zu den überwachten Klassifizierungsverfahren verwenden JOCHEM et al. (2021) und JOCHEM & TATEM (2021) unüberwachte Machine-Learning Verfahren, um auf der Grundlage der räumlichen Muster von Gebäudegrundrissen unterschiedliche Siedlungstypen zu gruppieren. Da in der vorliegenden Arbeit die zu detektierenden Klassen bereits festgelegt sind, wird in dieser Arbeit eine Methode des überwachten maschinellen Lernens, ein neuronales Netz angewandt.

XIE & ZHOU (2017) verwenden ein neuronales Netz mit Backpropagation um traditionelle Wohnhäuser mit Innenhof, mehrstöckige Wohngebäude und Hochhäuser zu differenzieren. Besitzen solche neuronalen Netze mehr als zwei verborgene Schichten (engl. *hidden Layer*), so werden sie als tiefe Netzwerke (*Deep Learning Networks*) betrachtet.

1.1.3 Semantische Segmentierung von Gebäudetypen mittels CNNs

Bei der Nutzung von Fernerkundungsdaten wie RGB-Bildern ist die semantische Segmentierung von Gebäudetypen ein klassisches *Computer Vision* Problem. Im Vergleich zu traditionellen Bildsegmentierungsmodellen hat in diesem Bereich der Bildsegmentierung in den letzten Jahren eine neue Generation von Deep Learning-Modellen bemerkenswerte Leistungsverbesserungen gezeigt und in gängigen Benchmarktests oft die höchsten Genauigkeitsmetriken erreicht (YUAN et al. 2020; MINAEE et al. 2021). Infolgedessen und dank der zunehmenden Verfügbarkeit von großen Datensätzen sowie zunehmenden Rechenkapazitäten werden DL-Methoden auch in der Fernerkundung immer häufiger eingesetzt (ZHANG et al. 2016; ZHU et al. 2017; BALL et al. 2017).

Für die semantische Segmentierung sind vor allem CNNs eine der am häufigsten verwendeten Methoden, welche die meisten Algorithmen der visuellen Erkennung übertreffen, da sie unter anderem automatisch relevante kontextbezogene Merkmale erkennen können (ZHANG et al. 2016; YANG H. et al. 2018). Nachteilig sind jedoch der große benötigte Trainingsdatensatz sowie die benötigte Rechenleistung (YUAN et al. 2020). Auf den grundlegenden Aufbau der in der vorliegenden Arbeit verwendeten CNNs wird in Kapitel 2 eingegangen.

Im Zusammenhang mit Gebäuden werden CNNs in der Fernerkundung besonders für generelle Landnutzungsklassifikationen (ZHANG et al. 2018; YANG C. et al. 2018; BHOSLE & MUSANDE 2019) oder die Klassifikation von Dachformen (ALIDOOST & AREFI 2018; CHEN et al. 2019; BUYUKDEMIRCIOGLU et al. 2021; WANG et al. 2022) eingesetzt.

Seit ein paar Jahren finden CNNs auch in der semantischen Segmentierung von Gebäudetypen Anwendung. Die im Folgenden analysierten Studien hierzu sind in Tabelle 1 zusammengefasst. Auf einen Vergleich der Genauigkeitsmetriken wird in der Diskussion in Kapitel 5 genauer eingegangen.

Tabelle 1: Analysierte Studien zur semantischen Segmentierung von Gebäudetypen mittels CNNs, „Trennung“ bezieht sich auf die mögliche methodische Trennung von Segmentierung und Klassifizierung der Gebäude.

Autor	Klassen	Algorithmus	Trennung	Daten (Herkunft, Auflösung)	Gütemaße
HOFFMANN et al. (2021)	Wohngebäude, Nicht-Wohngebäude	Bayesian FCN (U-Net ähnlich)	nein	Luftbilder (Google Maps, ~ 0,6 – 2,4 m)	IoU je Klasse: 0,036 –0,706
DROIN et al. (2020)	Wohngebäude, Nicht-Wohngebäude, Nebengebäude	FCN-vgg19	nein	Luftbilder (o. A., 0,4 m)	OA: 0,93 % Kappa: 0,73
DIMASSI et al. (2021)	Wohngebäude, Nicht-Wohngebäude	RexNet und 4 andere CNNs	ja	Satellitenbilder (Mapbox, ~ 0,5 m)	OA: 0,93 – 0,95 %
STRELTSOV et al. (2020)	Wohngebäude, Wirtschaftsgebäude	Segmentierung: U-Net Klassifizierung: ResNet-152	ja	Luftbilder (USGS, 0,3 m)	Segmentierung: mIoU 0,76 und 0,81 Klassifizierung: IoU 0,99 und 0,74
HOFFMANN et al. (2019)	Wirtschaftsgebäude, Wohngebäude, öffentliche Gebäude, Industriegebäude	Inceptionv3, VGG16 und diverse Fusionsstrategien	nein	Luftbilder (BingMaps, ~ 0,3 – 2 m) + GSV	Nur Luftbilder: F1: 0,51 – 0,66 Luftbilder + GSV: F1: 0,68 – 0,73
CAO & QIU (2018)	Ein- und Zweifamilienhäuser, begehbare Mehrfamilienhäuser, Mehrfamilienhäuser mit Aufzug, gemischte Wohn- und Geschäftsgebäude, Geschäfts- und Bürogebäude, Industrie- und Produktionsgebäude, Verkehrs- und Versorgungseinrichtungen, öffentliche Einrichtungen und Institutionen, + 3 Landnutzungsklassen	Basiert auf SegNet und Teilen von VGG16	nein	Luftbilder (BingMap, 0,3 m) + GSV	Nur Luftbilder: mIoU: 0,47, Genauigkeit: 0,78 Luftbilder + GSV: mIoU: 0,48, Genauigkeit: 0,78
PAN et al. (2020)	Alte Gebäude, alte Fabriken, Gebäude mit Eisendach, neue Gebäude	U-Net	nein	Satellitenbilder (Worldview-2, 2 m durch pan-sharpening 0,5 m)	OA: > 0,83 %, IoU: > 0,86, F1: 0,63 – 0,88
HUANG et al. (2021)	Niedrige Wohngebäude, mittelhohe Wohngebäude, hohe Wohngebäude, Industriegebäude, + 5 andere Klassen (Bodenbedeckung + Schatten)	M2-3-DCNN	nein	Satellitenbilder (Ziyuan 3, 2,1 und 3,5 m)	Genauigkeit der Klassen: 0,679 – 968

HOFFMANN et al. (2021) verwenden ein bayessches neuronales Netz mit einer U-Net (RONNEBERGER et al. 2015) ähnlichen Architektur zusammen mit einem Luftbild zur Erkennung von Gebäuden und deren Einteilung in die Klassen Wohngebäude und Nicht-Wohngebäude, wobei sie je nach Untersuchungsgebiet zu variierenden Modellleistungen kommen. Das besondere an dem verwendeten Netz ist die Möglichkeit, Unsicherheitswerte für jeden Pixel anzugeben und in das Modell mit einzubeziehen, wobei in der Studie bei einer niedrigen Auflösung (ca. 2,4 m) generell hohe Unsicherheitswerte angegeben werden und bei einer höheren Auflösung (ca. 0,6 m) besonders Wohngebäude am sichersten detektiert werden.

DROIN et al. (2020) verwenden die Klassen Wohngebäude, Nicht-Wohngebäude und Nebengebäude für die Erkennung mittels der DL-Architektur FCN-vgg19. Sie stellen fest, dass dieser Ansatz erste vielversprechende Ergebnisse zeigt und insbesondere die Hinzunahme von Höheninformationen zu den RGB-Daten in zukünftigen Studien analysiert werden sollte.

Wie HOFFMANN et al. (2021) klassifizieren auch DIMASSI et al. (2021) die Gebäudetypen Wohngebäude und Nicht-Wohngebäude nur mittels RGB-Bilder. Sie verwenden jedoch im Gegensatz zu HOFFMANN et al. (2021) einen in Kapitel 1.1.1 angesprochenen zweistufigen Ansatz, indem erst eine semantische Segmentierung der binären Unterscheidung zwischen Gebäude und keinem Gebäude angewandt wird, wonach im Anschluss die erhaltenen Gebäudeumrisse klassifiziert werden. Hierfür vergleichen sie fünf DL-Modelle wobei mit RexNet (HAN et al. 2021) die besten Genauigkeitsmetriken erreicht werden.

Einen ebenso wie bei DIMASSI et al. (2021) vorhandenen zwei-stufigen Ansatz wählen STRELTSOV et al. (2020). Mittels U-Net segmentieren sie Gebäude, die anschließend mit ResNet-152 als Wohngebäude oder Wirtschaftsgebäude klassifiziert werden. Hierzu wird nicht jedes Pixel einzeln klassifiziert, sondern jedes einzelne Gebäude wird mit dem Kontext in dem es sich befindet, klassifiziert.

Wie anfangs angesprochen ist die Verbindung von Luftbildern mit GSV-Bildern eine zahlreich angewandte Methode zur Erkennung von Gebäudetypen. HOFFMANN et al. (2019) und CAO & QIU (2018) verwenden in diesem Zusammenhang CNNs zur semantischen Segmentierung von Gebäudetypen auf ihren Luftbildern und testen die Klassifizierung mit und ohne GSV-Bilder. Hierzu wählen Erstere Inception3 (SZEGEDY et al. 2016) sowie VGG16 (SIMONYAN & ZISSERMAN 2015) und Letztere SegNet (BADRINARAYANAN et al. 2017). Die von ihnen detektierten Gebäudetypen sind in Tabelle 1 ersichtlich. In beiden Studien werden außerdem mit dem ImageNet-Datensatz vortrainierte Gewichte für das CNN verwendet. Die ImageNet Datenbank beinhält mehrere Millionen von RGB-Bildern mit zugewiesenen Klassen (DENG et al. 2009). Anzumerken ist, dass durch die Hinzunahme von GSV-Bildern bei CAO & QIU

(2018) keine Verbesserung der Genauigkeitsmetriken erreicht wird, bei HOFFMANN et al. (2019) jedoch schon.
Im Gegensatz zu den bisher genannten Studien verwenden PAN et al. (2020) nicht nur RGB-Bilder, sondern acht multispektrale Kanäle. Die von ihnen analysierten Gebäudetypen sind alte Gebäude, alte Fabriken, Gebäude mit Eisendach und neue Gebäude. Durch einen Vergleich mit dem Random Forest Algorithmus zeigen sie zudem die im Vergleich zu diesem Algorithmus bessere Eignung des U-Net CNNs für die vorliegende Aufgabe der Gebäudetyp-Klassifikation.
Auch HUANG et al. (2021) verwenden für ihre Detektion der Gebäudetypen niedriges, mittelhohes, hohes Gebäude sowie Industriegebäude mehrere Kanäle. Die Besonderheit ihrer Studie liegt darin, dass sie diese Daten aus unterschiedlichen Blickwinkeln verwenden und so ohne die Verwendung von LiDAR-Daten eine dreidimensionale Struktur der Gebäude verwenden können.

1.2 Ziel der vorliegenden Arbeit

Wie anfangs in Kapitel 1 erläutert, ist die CNN-basierte semantische Segmentierung unterschiedlicher Gebäudetypen mittels Fernerkundungsdaten eine wichtige Aufgabe in vielen Fachbereichen. In der vorhandenen Literatur zu diesem Thema werden meist nach ihrer Nutzung definierte Gebäudetypen verwendet. Nach ihrer Form definierte Gebäudetypen weisen wie bei PAN et al. (2020) meist nur sehr spezielle Klassen auf. In der vorhandenen Literatur fehlt ein systematischer Vergleich zwischen beiden Klassengruppen sowie eine Übersicht, welche der vielen möglichen Gebäudetypen durch CNNs erkennbar bzw. nicht-erkennbar sind. Die zweite Lücke in dem aktuellen Forschungsstand zur Gebäudeklassifikation mittels Fernerkundungsdaten stellt das Fehlen eines umfassenden Tests unterschiedlicher, gängiger Daten- und Modellparameter dar, welche das Ergebnis beeinflussen können. Hierzu zählt unter anderem die bisher nicht erfolgte Hinzunahme eines Höhenmodells zu dem Luftbild für die semantische Segmentierung unterschiedlicher Gebäudetypen.
In der vorliegenden Arbeit sollen diese Forschungslücken geschlossen werden, indem **zwei zentrale Forschungsfragen** beantwortet werden sollen. In einer ersten Testreihe werden mehrere Klassengruppen der Gebäudeform und der Gebäudenutzung gebildet und anschließend deren Erkennung mittels aktueller CNNs getestet. Die erste Forschungsfrage, welche in dieser Arbeit beantwortet werden soll, lautet demnach:

1. **Wie gut lassen sich mittels ausgewählter CNNs unterschiedliche, nach ihrer Form und nach ihrer Nutzung definierte Gebäudetypen auf Grundlage eines Luftbildes und nDOMs erkennen?**

In einer zweiten Testreihe werden mehrere Einstellungen unterschiedlicher Daten- und Modell-Parameter verändert, um deren Auswirkung auf das Modellergebnis zu testen. Hierzu zählt die Hinzunahme eines normalisierten digitalen Oberflächenmodells (nDOM) zu dem Luftbild. Die zweite Forschungsfrage hierzu lautet:

2. **Welche Auswirkungen zeigen ausgewählte Modell- und Daten-Parameter auf die Gebäudetypenerkennung?**

Für diese Forschungsfrage werden mehrere Parameter untersucht, deren Einfluss auf die Modellleistung als am größten angenommen wird. Die getesteten Daten-Parameter umfassen hierbei folgende Aspekte:

1. Variierende Untersuchungsgebiete mit unterschiedlichen Verteilungen der einzelnen Gebäudetypen
2. Unterschiedliche Auflösungen der Daten
3. Mehrere Kombinationen aus den Datenkanälen RGB, NIR und nDOM

Hinzu kommen die Modell-Parameter, welche ausschließlich das verwendete CNN betreffen. Diese Parameter umfassen:

1. Unterschiedliche Modellarchitekturen
2. Variierende Anzahl an Modellepochen
3. Auf andere Daten vortrainierte Modelle

Zuletzt werden für unterschiedliche Schritte der Methodik in der Datenvorverarbeitung sowie innerhalb des CNNs unterschiedliche Zufallsvariablen gewählt, um den Einfluss des Zufalls auf das Modellergebnis zu erkennen.

Ergänzend kann erwähnt werden, dass neben den zwei zentralen Forschungsfragen auch mehrere konkrete Unterfragen gebildet werden können. Diese Unterfragen sollen dazu beitragen, die zwei Hauptforschungsfragen zu beantworten, indem unterschiedliche Teilbereiche der Hauptfragen näher betrachtet werden. Eine Übersicht stellt Tabelle 2 dar.

Tabelle 2 Zentrale Forschungsfragen der vorliegenden Arbeit mit dazugehörigen Unterfragen.

Zentrale Forschungsfrage	Unterfragen
1. Wie gut lassen sich mittels ausgewählter CNNs unterschiedliche, nach ihrer Form und nach ihrer Nutzung definierte, Gebäudetypen auf Grundlage eines Luftbildes und nDOMs erkennen?	- Wie genau sind nach ihrer Form definierte Gebäudetypen erkennbar? - Welche nach ihrer Nutzung definierten Gebäudetypen sind erfassbar?
2. Welche Auswirkungen zeigen ausgewählte Modell- und Daten-Parameter auf die Gebäudetypenerkennung?	- Welchen Einfluss hat der Zufall in Form gewählter Zufallsvariablen auf das Modellergebnis? **Daten-Parameter** - Ist ein Zusammenhang zwischen den Test- und Trainingsdaten und der Erkennung der einzelnen Gebäudetypen vorhanden? - Wie gut funktioniert die Gebäudeklassifizierung auf Basis des unverbesserten amtlichen Gebäudedatenbestands? - Wie verändert die Verwendung von unterschiedlichen Auflösungen die Erkennung der Gebäudetypen? - Welche Datenkanäle sind für die Erkennung der Gebäudetypen am wichtigsten? - Hat die Auswahl eines Teils der Trainingsdaten eine Auswirkung auf die Erkennung der Gebäudetypen? **Modell-Parameter** - Welche der ausgewählten Modellarchitekturen führt zu dem besten Ergebnis? - Nach welcher Trainingsepoche sind die besten Ergebnisse feststellbar? - Wie verändert die Verwendung von vortrainierten Gewichten das Modellergebnis?

2 Convolutional Neural Networks (CNNs)

Wie in Kapitel 1.1.3 erläutert, zeigen Deep-Learning Architekturen in Aufgaben der semantischen Segmentierung aktuell die besten Ergebnisse im Vergleich zu traditionellen Methoden des maschinellen Lernens. Im Bereich des Deep-Learning wiederum zählen CNNs zu den erfolgreichsten und am häufigsten angewandten Algorithmen in der Bildsegmentierung. Im folgenden Kapitel wird ein Überblick über die Funktionsweise von CNNs zur semantischen Bildsegmentierung gegeben, indem zuerst generell auf neuronale Netze (NN) eingegangen wird. Im Anschluss werden CNNs erläutert und es wird auf den Spezialfall der Anwendung zur semantischen Segmentierung eingegangen. Im Anschluss wird der Nutzen sowie die generelle Funktionsweise der wichtigsten Modellparameter erklärt. Dabei wird insbesondere auf solche Parameter eingegangen, welche für ein Verständnis der in dieser Arbeit angewandten Experimente wichtig sind. Parameter, deren spezifische Einstellungen im Methodenkapitel 3.3.2 angegeben werden, sind in Kapitel 2.2 fett gedruckt markiert.

2.1 Neuronale Netze und Layertypen eines CNNs

Neuronale Netze (NN) bestehen grundsätzlich aus einer Vielzahl von miteinander verbundenen Recheneinheiten (sogenannte Neuronen), welche verschiedene Merkmale aus den Eingangsdaten erlernen, um hierdurch die Ausgabe eines Modells zu optimieren. Sie ahmen die Art und Weise nach, in der biologische Neuronen einander Signale senden. (KETKAR & MOOLAYIL 2021).

Diese Neuronen sind in verschiedenen Layern angeordnet, welche unterschiedliche Funktionen besitzen. Die verschiedenen Arten und Funktionen von Layern ermöglichen es dem Modell, die angesprochenen Merkmale aus den Eingabebildern zu extrahieren und diese miteinander zu kombinieren, um dem Ziel der Anwendung näher zu kommen. Ein normales bzw. künstliches Neuronales Netz (*Artificial Neural Network*, ANN) besitzt normalerweise drei Layertypen, welche in einem CNN weiter untergliedert werden. Diese grundlegenden Layer sind ein Eingabelayer, mehrere Zwischenlayer (*Hidden Layer*) sowie ein Ausgabelayer. In Abbildung 1 sind die Neuronen als Kreise dargestellt, welche mit den Neuronen der nachfolgenden Layer verbunden sind.

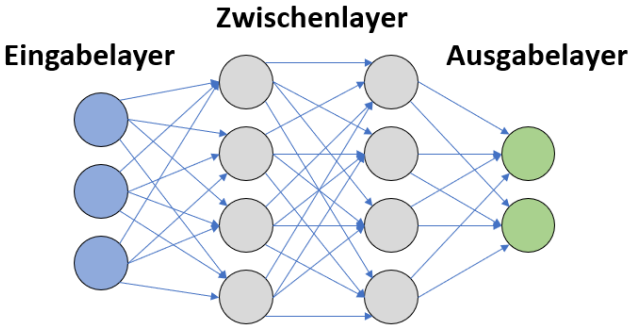

Abbildung 1 Künstliches neuronales Netz mit den drei grundlegenden Layertypen.

Der Eingabelayer lädt die Daten, welche typischerweise die Form eines mehrdimensionalen Vektors haben. Im vorliegenden Fall sind dies die Kanäle des Luftbildes sowie des Höhenmodells. Für jedes Pixel jedes Datenkanals existiert hier ein Neuron.

In den anschließenden *Hidden Layers* werden Entscheidungen, im vorliegenden Fall zur Klassifizierung, unter Berücksichtigung des Ergebnisses des jeweils vorhergehenden Layers getroffen und es wird analysiert, wie eine stochastische Veränderung des Vorgehens in diesen Layern die Ausgabe beeinträchtigt oder verbessert, was als Lernprozess bezeichnet wird. Das Lernen selbst findet hierbei in zu den einzelnen Neuronen gehörigen Gewichten statt. Diese werden mit jeder Veränderung an die Daten angepasst und stellen so sinngemäß das Gedächtnis des Modells dar. Ist eine Vielzahl von *Hidden Layers* hintereinander angeordnet, so wird dies als Deep-Learning bezeichnet und es handelt sich um ein *Deep Neural Network* (DNN) (O'SHEA & NASH 2015).

Das Ergebnis des Modells wird im Ausgabelayer als Wahrscheinlichkeitswert jeder Klasse für jeden Bildpixel angegeben. Dafür wird eine logistische Funktion auf die Ausgabe des letzten *Hidden Layer* angewandt. Hierdurch werden die errechneten Ausgabe-Logits in eine Wahrscheinlichkeitsverteilung transformiert.

In einem klassischen DNN sind alle Neuronen für jeden Pixel der einzelnen Layer miteinander verbunden (*fully connected*). In einem CNN ist dies jedoch nicht der Fall, da anstelle eines *fully connected layers* eine Faltungsoperation (*convolution*) verwendet wird, durch welche jedes Neuron nur mit einer kleinen Region, dem sog. rezeptiven Feld, des vorangegangenen Layers verbunden ist (KETKAR & MOOLAYIL 2021). Dies führt zu einer wesentlich geringeren Anzahl an Gewichten, als bei vollständig verknüpften neuronalen Netzen wie den erwähnten DNNs und stellt den größten, den Rechenaufwand betreffenden, Vorteil von CNNs dar (MINAEE et al. 2021). Diese Verknüpfungen geschehen in dem sog. *Convolutional Layer*, welcher zusammen mit drei weiteren Layern die wichtigsten *Hidden Layers* eines klassischen CNNs darstellt (O'SHEA & NASH 2015):

1. Der *Convolutional Layer* wendet einen Filter auf die Eingabedaten an, um Merkmale und Muster herauszufiltern.
2. Mit dem *Activation Layer*, wird Nichtlinearität eingebaut, was das Erlernen von komplexen Beziehungen in den Daten ermöglicht.
3. Anschließend reduziert der *Pooling Layer* durch ein Downsampling die Bildgröße, um den Rechenaufwand zu verringern.
4. Der *Fully Connected Layer* führt die letztendliche Vorhersage bzw. Klassifikation auf Basis der vorherigen Layer durch.

Im Folgenden werden diese Layertypen genauer vorgestellt.

Convolutional Layer

Der *Convolutional Layer* ist Hauptbestandteil eines CNNs. In diesem werden lernfähige Filter, also mit veränderbaren Gewichten, auf einen Teil der Daten des Input Layers angewandt. In den Gewichten der Neuronen dieser Filter sind hauptsächlich einzelne Muster bzw. Merkmale abgespeichert, welche der jeweilige Filter in den analysierten Daten erkennen kann. Zum Beispiel kann ein Filter, dessen Gewichte aus diskreten Zahlen bestehen, die das Muster eines Daches bilden, das Vorhandensein eines Dachs erkennen (LOPEZ PINAYA et al. 2020). Die Größe der Filter ist kleiner als die der Eingabedaten, welche durch den Filter schrittweise abgetastet werden. Im Fall des in der vorliegenden Arbeit unter anderem verwendeten FPN-Netzwerks mit ResNet50 als Encoder beträgt die Größe der Filter beispielsweise je nach Layer zwischen 1*1 und 7*7 Pixel je Eingangskanal (R, G, B, NIR, nDOM). Die Schrittweite (*stride*) des Filters liegt zwischen einem und drei Pixel. Während der *Convolution*, also der schrittweisen Anwendung der Filter, wird das Skalarprodukt aus den Gewichten der Neuronen des Filters und dem jeweiligen Teil der Eingabedaten berechnet. Das Ergebnis der Skalarprodukte wird anschließend als *Feature Map* bezeichnet. Innerhalb derselben *Feature Map* hat jedes Neuron die gleichen Gewichte, basierend auf der Annahme, dass ein Muster an jeder Position auftreten kann (SAAD et al. 2017). Diese *Feature Maps* werden dabei immer komplexer und größer, je „tiefer“ das CNN wird. Von weniger komplexen Merkmalen (*Features*) wie Linien oder Ecken werden zunehmend auch komplexere Merkmale wie Formen und Texturen erlernt. Durch die Anwendung mehrerer Filter für die Erfassung unterschiedlicher Muster entstehen mehrere *Feature Maps*, welche zusammengeschichtet die Ausgabe des *Convolutional Layers* bilden.

Activation Layer

Auf die Ausgabe des Convolutional Layers wird in dem Activation Layer eine nicht-lineare Aktivierungsfunktion angewandt. Dies bedeutet, dass deren Funktionsverlauf nicht proportional zur x-Variablen verläuft. Hierdurch können auch nicht-lineare Abhängigkeiten erkannt werden (HAN et al. 2021). In CNNs wird häufig die simple *rectified linear unit* (ReLU)-Funktion verwendet, welche Werte kleiner als Null auf Null setzt und alle anderen Werte beibehält. Mathematisch ist die Funktion also: $f(x) = \max(0, x)$.

Pooling Layer

Der *Pooling Layer* wird in das Modell eingebaut, um die Größe der Eingabekachel zu verringern. Dies erhöht die Geschwindigkeit des Modells, vermindert den benötigten Speicherplatz und verringert die Gefahr der Überanpassung (*overfitting*) (LI Z. et al. 2022). Letzteres bedeutet eine zu gute Anpassung an die Trainingsdaten, was zu schlechteren Ergebnissen bei der Anwendung auf separaten Testdaten führt. Durchgeführt wird das *Pooling*, indem ein Filter über die Daten läuft und eine mathematische Operation (in den hier verwendeten Modellen das Übernehmen des Maximums) auf die Daten anwendet. Bei einer Filtergröße von 2*2 wird aus einem 32*32 Pixel großen Raster so ein 16*16 Pixel großes Raster, wobei jeweils der maximale Pixelwert von 2*2, also 4 Pixeln, in dem neuen Raster vorhanden ist. Der *Pooling Layer* entspricht also einem *Downsampling* der Daten.

Fully-Connected Layer

Wie bereits angesprochen, ist in dem *Fully-Connected Layer* jedes Neuron mit den Neuronen des vorherigen Layers verbunden. Dieser Layer fasst so die in den vorherigen *Convolution-* und *Pooling Layern* erlernten Merkmale zusammen und kombiniert diese, um eine Entscheidung zur Klassifikation zu treffen. Um in diesem Layer verarbeitet zu werden, wird die *Feature Map* in einen 1D-Vektor transformiert (LOPEZ PINAYA et al. 2020). Diese Form hat auch die Ausgabe des Layers. Von Vorteil ist dies insbesondere für die effiziente Verarbeitung von sequentiellen Daten, wie beispielsweise Zeitreihen.

2.2 CNN-Architektur zur semantischen Segmentierung und Transfer Learning

In klassischen CNNs für Klassifikationsaufgaben sind die letzten Layer des CNNs meist *Fully-Connected Layer mit 1D-Vektoren als Ausgabe*. Da für eine semantische Segmentierung jedoch ein 2D- Raster mit Wahrscheinlichkeitswerten je Klasse je Pixel und nicht nur ein 1D-Vektor mit Wahrscheinlichkeiten je Klasse benötigt wird, werden in CNNs zur semantischen

Segmentierung oft keine *Fully-Connected Layer* verwendet, sondern diese werden durch Layer zum Upsampling (sowie teilweise durch *Convolutional Layer)* ersetzt. Daher entspricht die Größe des Ausgabelayers der Größe des Eingabelayers. Die Ersetzung des *Fully-Connected Layers* ist außerdem nötig, weil der letzte *Convolutional Layer* nach mehreren *Pooling Layern* meist eine kleine *Feature Map* erstellt, die zwar für die Klassifikation des gesamten Bildes herangezogen werden kann, für eine Klassifizierung jedes einzelnen Pixels in Form einer semantischen Segmentierung jedoch zu ungenau ist (BAHETI et al. 2019). Die mit zusätzlichen *Convolutional Layern* sowie Layern zum Upsampling gebildete Architektur kann in einen **Encoder** und einen **Decoder** Teil unterschieden werden. Diese Architektur ist an dem Beispiel eines FCN-Netzwerks in Abbildung 2 dargestellt.

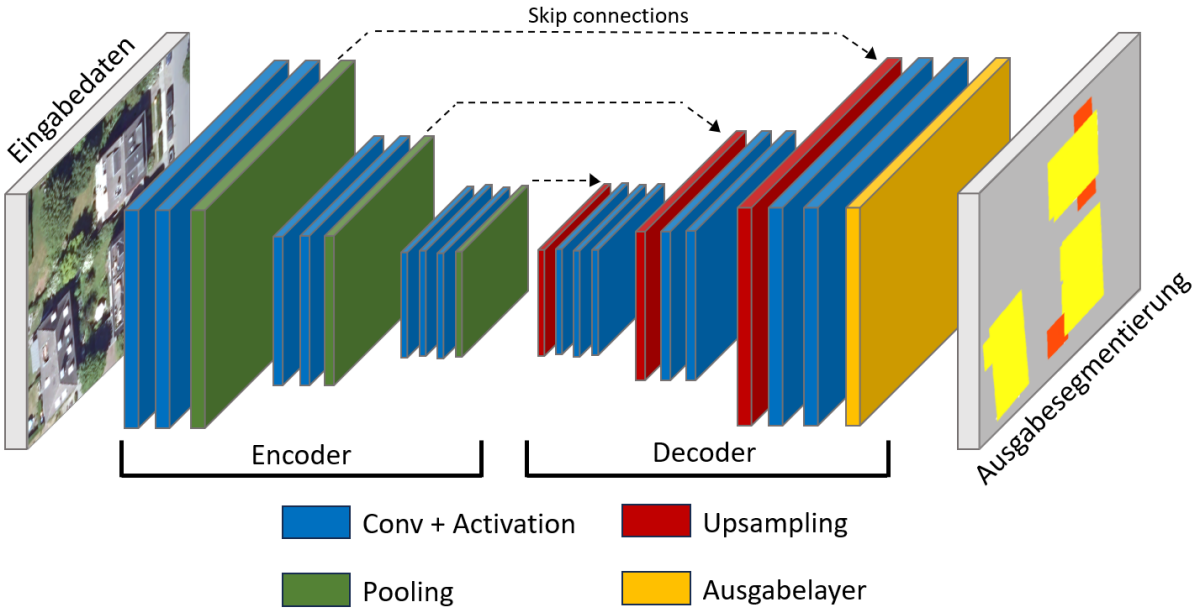

Abbildung 2 Beispielhafte Darstellung eines CNNs mit Encoder-Decoder FCN Architektur zur semantischen Segmentierung (nach: PENG et al. 2019).

Aus dem Eingabebild wird, wie erläutert, im Encoder-Abschnitt eine *Feature Map* mit geringerer Größe erstellt. Dies geschieht durch die Anwendung von mehreren der vorgestellten Layern, welche zu einer pyramidenartigen Anordnung mit immer niedrigeren Auflösungen führen. Dies ermöglicht es, Merkmale aus unterschiedlich großem Kontext zu erkennen und ist in Abbildung 2 gut an der unterschiedlichen Größe der einzelnen Layer erkennbar. Im Decoder-Abschnitt wird die Größe anschließend wieder an die des Eingabebildes angepasst und der Ausgabelayer erzeugt durch eine logistische **Aktivierungsfunktion** aus der Ausgabe des letzten Layers (in Form von Logits) das finale Segmentierungsraster in Form von Wahrscheinlichkeiten je Klasse. Der Encoder-Teil extrahiert also Merkmale aus dem Eingangsbild und der Decoder-Teil erzeugt das Ausgaberaster. In Abbildung 2 sind zusätzlich Skip-Verbindungen ersichtlich. Diese ermöglichen es dem Netzwerk, eine oder mehrere Schichten

im Encoder und Decoder zu umgehen, wodurch das Netzwerk mehr räumliche Informationen über das Eingangsbild behalten kann (HE et al. 2016; QI et al. 2019).

Für die Anzahl, Anordnung und Kombinierung der vorgestellten Layer eines CNNs gibt es keine fest definierte Struktur sondern eine Vielzahl unterschiedlicher Möglichkeiten, welche unterschiedliche **CNN-Architekturen** bilden. Somit stellt Abbildung 2 ein anschauliches Beispiel für ein CNN zur semantischen Segmentierung dar, es existieren jedoch zahlreiche Variationen.

Nach einem Durchlauf aller Layer des Modells, genannt *forward pass*, wird das Ergebnis des Modells durch eine **Verlustfunktion** evaluiert. Diese Funktion misst die Diskrepanz zwischen den Modellausgaben und den tatsächlichen Labels (LI Z. et al. 2022).

Das Ziel ist es, das Ergebnis dieser Funktion so weit wie möglich zu minimieren. Hierbei wird nicht nur die Passgenauigkeit zwischen Modellausgabe und Label berücksichtigt, sondern zumeist auch die Größe der Gewichte. Dadurch wird versucht, das Modell dazu zu zwingen, einfachere Gewichtskonfigurationen zu bevorzugen, was einer Überanpassung an die Trainingsdaten vorbeugen kann. Ausgeführt wird dies durch einen **Optimierungsalgorithmus**, welcher ermittelt, wie die Gewichte der Neuronen des Netzwerks verändert werden sollten, um das Ergebnis der Verlustfunktion im nächsten Durchlauf zu minimieren (KETKAR & MOOLAYIL 2021). Der Optimierungsalgorithmus berücksichtigt sowohl die Minimierung der Verlustfunktion als auch die Größe der Gewichte. Der Parameter **Decay** steuert hierbei die Stärke der Regularisierung, die auf die Gewichte angewendet wird. Je größer der *Decay*, desto eher werden größere Gewichte minimiert, was einer Überanpassung an die Trainingsdaten entgegen wirken kann (SCHINDLER et al. 2020).

Ein weiterer der Gewichtsoptimierung zugeordneter Parameter ist die **Lernrate**. Diese bestimmt die Schrittgröße, mit der die Gewichte des Netzwerks während des Trainings aktualisiert werden. Sie beeinflusst, wie stark und schnell das Modell lernt sich an die Daten anzupassen und die Verlustfunktion zu minimieren. Eine passende Wahl der Lernrate ist wichtig, da eine zu hohe Lernrate zum Überspringen der optimalen Werte führen kann und eine zu niedrige Lernrate zu langsamen Lernen führen kann (MAFENI MASE et al. 2020).

In jedem Durchlauf werden also die Gewichte der Neuronen verändert, was das „Lernen“ des Modells darstellt. Das CNN kann die vorhandene Aufgabe durch diese Struktur lösen, ohne direkt durch spezielle Regeln an die Aufgabe angepasst zu sein.

Da die Anzahl der Datenkacheln häufig so groß ist, dass der vorhandene Rechenspeicher nicht ausreicht, um sie gemeinsam durch das Modell zu prozessieren, werden die Daten in einzelne Teilmengen, genannt ***Batches***, unterteilt. Diese Teilmengen durchlaufen nacheinander einzeln

die erläuterten Stadien des Modells. Sind alle Teilmengen einmal prozessiert worden, so wird dies eine **Epoche** genannt. Das Prozessieren in einzelnen Batches verringert den benötigten Speicherplatz und führt dazu, dass das Modell schneller trainiert wird, da die Gewichte häufiger aktualisiert werden (KETKAR & MOOLAYIL 2021). Würde keine Teilmenge gebildet, so würden die Gewichte nur einmal je Epoche aktualisiert. Nach einer Epoche wird das Modell zudem mittels einer Genauigkeitsmetrik validiert. Meist handelt es sich hier um die mittlere IoU, die in Kapitel 3.3.4 erläutert wird.

In der ersten Epoche besitzen diese Gewichte normalerweise zufällige Werte. Allerdings kann im Modell auch sogenanntes **Transfer Learning** verwendet werden. Hierbei werden die Gewichte eines Modells verwendet, welches für eine andere Aufgabe bereits vortrainiert wurde. Das hierbei erlangte „Wissen" kann auf eine andere Aufgabe angewandt werden. Bei einem ausreichend großen Trainingsdatensatz können hierzu beispielsweise die Gewichte des ersten *Convolutional Layers* des vortrainierten Modells übernommen werden und die restlichen Gewichte werden anschließend mit den Trainingsdaten trainiert, wobei auch eine Vielzahl an anderen Methoden zur Übernahme von Gewichten eines vortrainierten Modells bestehen (MINAEE et al. 2021). Transfer Learning verschafft dem Modell bei einer ähnlichen Aufgabe also einen Startvorteil gegenüber dem Training von Grund auf (GUPTA et al. 2022).

3 Verwendete Daten und Methodik

Im folgenden Kapitel werden die verwendeten Daten sowie die in dieser Arbeit angewandte Methodik zur Beantwortung der in Kapitel 1.2 gestellten Forschungsfragen erläutert. In Abbildung 3 ist ein Überblick über dieses Kapitel dargestellt, wobei die Nummerierung der Unterkapitel in kursiver Schrift mit angegeben ist.

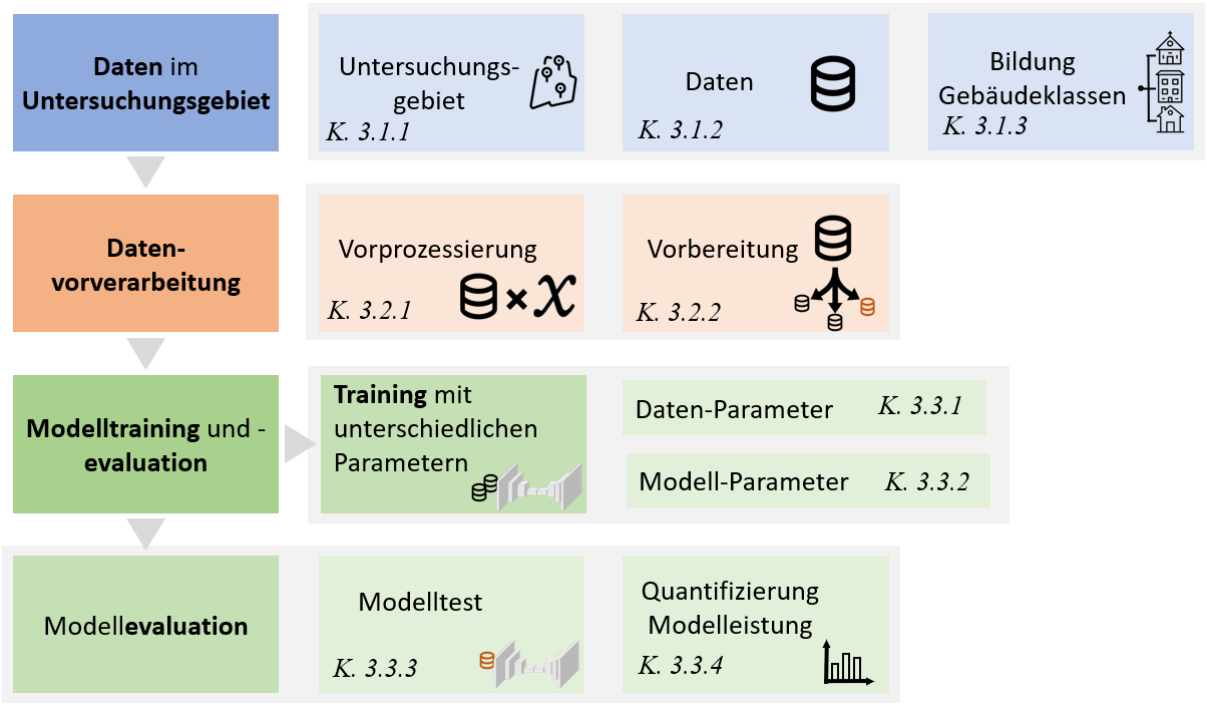

Abbildung 3 Übersicht über die einzelnen Teilbereiche der angewandten Methodik.

Zu Beginn werden die Daten erläutert. Darauf folgt die Methodik zur Prozessierung der Daten. Diese umfasst zunächst die Vorverarbeitung der Daten für das Modell. Anschließend wird das Modell mit unterschiedlichen Parametern angewandt und am Ende der Prozessierung wird die jeweilige Modellleistung evaluiert. So kann der Einfluss unterschiedlicher Parameter auf das Modell analysiert werden.

Um vergleichbare und vor allem reproduzierbare Ergebnisse zu erhalten, wird die Methodik deterministisch gewählt. Für vorhandene Zufallsvariablen werden dementsprechend Pseudozufallszahlen mit demselben Initialwert (*seed*) verwendet. Die verwendete Methodik erzeugt also bei gleicher Eingabe und bei Ausführung auf gleicher Soft- und Hardware die gleiche Ausgabe.

Bis auf die Änderung der Datenauflösung sowie für Kartenvisualisierungen wird, soweit nicht anders angegeben, für die komplette Methodik Python (Version 3.10.12) verwendet. Die Datenauflösung wird teils in SAGA GIS (Version 9.1.0) angepasst und Kartenvisualisierungen werden in QGIS (Version 3.30.2) erstellt. Als Grafikprozessor (GPU), auf welcher die CNN-

Modelle rechnen, werden eine NVIDIA Quadro RTX 4000 mit 8 GB Speichervolumen sowie eine NVIDIA RTX A4000 mit 16 GB Speichervolumen verwendet.

3.1 Untersuchungsgebiet, Daten und verwendete Gebäudetypen

Das folgende Kapitel gibt zunächst einen Überblick über das Untersuchungsgebiet (Kapitel 3.1.1) sowie die verwendeten Daten (Kapitel 3.1.2). In Kapitel 3.1.3 werden anschließend die in dieser Arbeit verwendeten Gebäudetypen genauer erläutert.

3.1.1 Darstellung des Untersuchungsgebiets

Als Untersuchungsgebiet werden sechs Teilgebiete in Nordrhein-Westfalen (NRW) gewählt. Für die Erkennung unterschiedlicher Gebäudetypen mittels Fernerkundungsdaten ist NRW von Vorteil, da die verwendeten Daten hier frei verfügbar sind und die benötigte Auflösung besitzen. Zudem ist NRW das bevölkerungsreichste Bundesland Deutschlands, was zu einer Vielzahl an benötigten Gebäuden führt (BPB 2020). Um von jedem Gebäudetyp ausreichend Gebäude für die Modellanwendung zur Verfügung zu haben, beinhalten die ausgewählten Teilgebiete alle eine hohe Zahl an Gebäuden mit unterschiedlichen Gebäudetypen. Einen Überblick über die ausgewählten Teilgebiete gibt Abbildung 4.

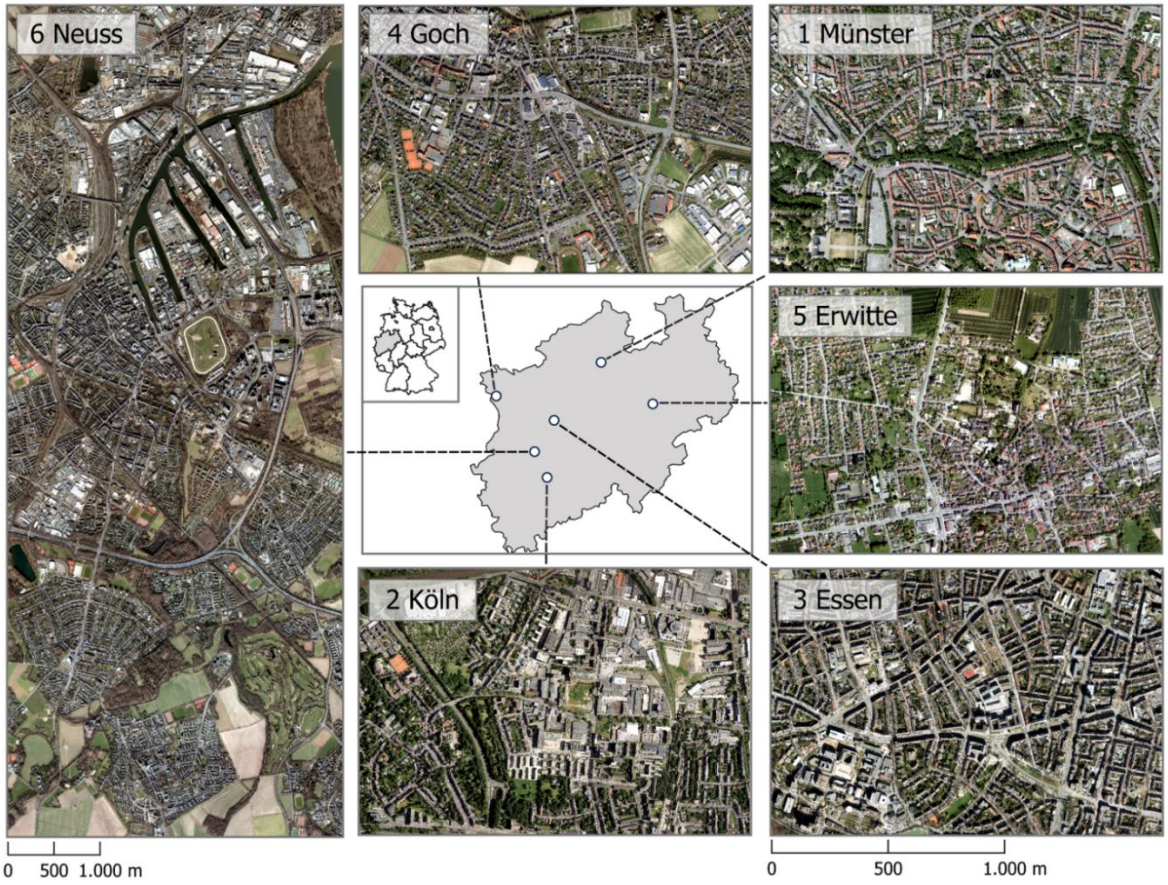

Abbildung 4 Überblick über die sechs verwendeten Untersuchungsgebiete in NRW.

Bis auf das sechste Teilgebiet der Stadt Neuss umfassen alle fünf Teilgebiete eine Fläche von 1,9 km^2. Für Neuss wird eine größere Fläche von 32,6 km^2 gewählt, da mit der Region Neuss getestet werden soll, ob die verwendeten Daten in ausreichender Menge auch ohne eine spätere Teilung mit Überlappungsbereichen sowie ohne manuelle Bearbeitung als Trainingsdaten nutzbar sind. Die Hauptgebiete, in welchen der Großteil der in dieser Studie erarbeiteten Tests durchgeführt wird, umfassen jedoch die Regionen 1 bis 5.

Die ersten drei Regionen beinhalten hierfür urbanen Raum:

- Region 1 befindet sich im Stadtbezirk Münster-Mitte und umfasst den nördlichen Teil der historisch geprägten und eng bebauten Altstadt von Münster sowie einen Teil des durch dichte Bebauung gekennzeichneten Stadtviertels „Kreuzviertel“ (TEMLITZ 2007).
- Region 2 befindet sich im Stadtteil Müngersdorf am westlichen Rand Kölns und umfasst einen Großteil des Gewerbegebiets „Technologiepark Köln“ mit vielen Bürogebäuden sowie Wohngebiete mit überwiegend Reihenhäusern im Süden und Mehrfamilienhäusern in Westen.
- Als weitere urbane Region befindet sich Region 3 im Stadtteil Holsterhausen in Essen. Die Region ist primär als Wohngebiet mit klassischer Blockrandbebauung zu charakterisieren, unterbrochen durch das Universitätsklinikum Essen im Süden der ausgewählten Region.

Mit Region 4 und 5 sind als Kontrast zu den großen Agglomerationsräumen zwei deutlich weniger dicht bebaute, semi-urbane Regionen mit dem Charakter von Kleinstädten vertreten:

- Region 4 umfasst mit dem südwestlichen Teil der Stadt Goch zahlreiche Einfamilienhäuser und Doppelhaushälften. Am östlichen Rand ist zudem ein kleines Industriegebiet vertreten und im Norden ein Teil des Stadtzentrums.
- Region 5 in der Kleinstadt Erwitte ergänzt den Datensatz mit einem ländlichen Gebiet und weist mit zahlreichen Einfamilienhäusern sowie einem kleinen Ortskern eine rurale Struktur auf.

Schließlich ist als Region 6 ein wesentlicher Teil der Großstadt Neuss vertreten. Sie auf Grund ihrer Repräsentativität für ganz Deutschland gewählt und umfasst Industriegebiete, Häfen, dichte Wohnbebauung im Zentrum sowie rurale Vororte im Süden.

3.1.2 Verwendete Daten

Die Datengrundlage für die vorliegende Arbeit bilden Luftbilder, ein nDOM sowie 3D-Gebäudemodelle des Untersuchungsgebiets, welche vom Land NRW frei zur Verfügung ge-

stellt werden (IT.NRW 2023). Anzumerken ist, dass die Zeitpunkte der Erstellung der drei Datengrundlagen nicht genau übereinstimmen (DOP: 2019, nDOM: 2017 – 2021, Gebäudemodell 2021).

Die verwendeten Luftbilder wurden mittels eines bildbasierten Oberflächenmodells entzerrt, wodurch es sich um sogenannte *True Digital Orthophotos* (TDOPs) handelt. Sie sind verzerrungsfrei, georeferenziert und es ist nahezu jeder Punkt der Oberfläche ohne Schatteneffekte dargestellt (LIU et al. 2018). Diese DOPs umfassen die vier spektralen Kanäle rot, grün, blau (RGB) und das nahe Infrarot (NIR). Letzteres erweitert den verwendeten sichtbaren Teil des elektromagnetischen RGB-Spektrums (750 – 400 nm) um große Wellenlängen im Bereich bis zu 3000 nm (WHETSEL 1968). Diese DOPs liegen in einer Auflösung von 10 cm vor.

Das nDOM ist ein Differenzmodell aus einem Digitalen Oberflächenmodell (DOM) sowie einem Digitalen Geländemodell (DGM). Das hier verwendete DOM wurde durch Bildkorrelation von orientieren digitalen Luftbildern erstellt und bildet die Geländeoberfläche mit den sich darauf befindenden Objekten wie Vegetation und Bebauung ab. Im Gegensatz dazu wurde das DGM auf Basis von Daten einer LiDAR-Befliegung erarbeitet und repräsentiert das Gelände ohne sich darauf befindende Objekte (GEOBASIS NRW 2022). Mit einer Auflösung von 50 cm stellt das aus diesen zwei Höhenmodellen gewonnene nDOM die relative Höhe der Objekte über der Geländeoberfläche dar.

Die amtlichen 3D-Gebäudemodelle stellen die erfassten Gebäude dreidimensional dar. In der verwendeten Detaillierungsstufe, *Level of Detail* (LoD), 2 sind die Gebäude durch einen Quader mit der Gebäudehöhe auf dem Grundriss aus dem amtlichen Liegenschaftskataster sowie einer standardisierten Dachform repräsentiert. Für die Verwendung im zweidimensionalen Raum werden die Gebäudemodelle in die Ebene projiziert, so dass die Fläche des Gebäudes mit der Sicht aus dem Nadir übereinstimmt. Jedes Gebäude besitzt in dem amtlichen Ausgangsdatensatz zudem mehrere Attribute wie die Gebäudefunktion oder die Gebäudehöhe. Die Lagegenauigkeit des Vektordatensatzes liegt bei wenigen Zentimetern und die Höhengenauigkeit bei bis zu 5 m (BEZIRKSREGIERUNG KÖLN 2017). Da unter anderem durch die unterschiedlichen Aufnahmezeitpunkte der verwendeten Daten zwischen dem DOP, dem nDOM und den LoD2-Gebäudedaten Unterschiede im Gebäudebestand vorliegen, wurden die Gebäudeformen manuell angepasst. Diese Anpassungen umfassen das Hinzufügen von fehlenden Gebäuden, die Entfernung von nicht vorhandenen Gebäuden sowie die Anpassung der Gebäudeform bei differierenden Formen. Diese angepassten Daten wurden vom DLR (Deutsches Zentrum für Luft und Raumfahrt e.V.) zur Verfügung gestellt und sind für Region 1 bis 5 vorhanden. Verwendet werden dieselben Gebäudepolygone wie bei STILLER et al. (2023).

Für Region 6 werden nur die originalen LoD-Daten verwendet um, wie im vorherigen Kapitel angesprochen, zu testen, wie ausschlaggebend die manuelle Gebäudeanpassung für das Training des Modells ist.

3.1.3 Bildung und Erläuterung der verwendeten Gebäudeklassen

Wie in Kapitel 1.1.1 erläutert, werden Gebäudetypen in der vorhandenen Literatur zumeist nach der Gebäudefunktion oder der Gebäudemorphologie eingeteilt. In der vorliegenden Arbeit werden beide Einteilungsmöglichkeiten angewandt, indem zwei Klassengruppen zur Nutzung bzw. Funktion der Gebäude und vier Klassengruppen zur Morphologie bzw. Form der Gebäude gebildet werden. Dieses Klassifizierungsschema mit den verwendeten sechs Klassengruppen ist in Abbildung 5 ersichtlich.

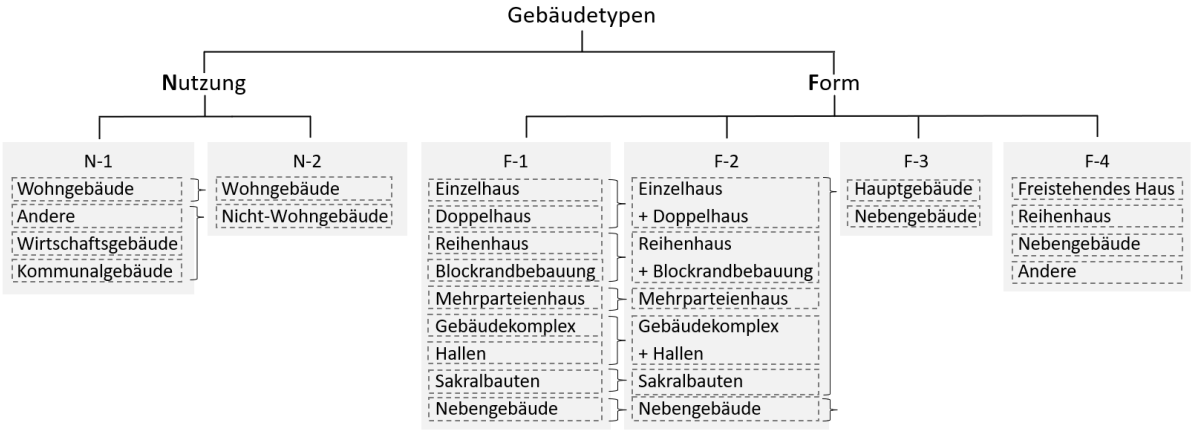

Abbildung 5 Klassifizierungsschema der analysierten Gebäudetypen nach ihrer Nutzung (N-) und ihrer Form (F-).

Aus den Klassengruppen N-1 sowie F-1 ergeben sich jeweils die darauffolgenden Klassengruppen N-2 bzw. F-2 und F-3. In den drei Ausgangsklassengruppen (N-1, F-1 und F-2) sind diejenigen Gebäudetypen, die in einer anderen Klassengruppe eine neue Klasse ergeben, in Abbildung 5 mit einer geschweiften Klammer gekennzeichnet. Jede Klasse ist strichliert umrandet.

Jeder Gebäudetyp besitzt eine unterschiedliche Häufigkeit in den fünf Hauptuntersuchungsgebieten. Zur Veranschaulichung dieses Klassenungleichgewichts sind in Abbildung 6 die summierten Flächen der Pixel der Gebäudetypen je Region angegeben. Da Region 6 eine deutlich größere Ausdehnung besitzt, nur für einen Parameter-Test verwendet wird und hier nicht alle Daten zu allen Klassengruppen vorhanden sind, wird diese Region nicht mit angegeben.

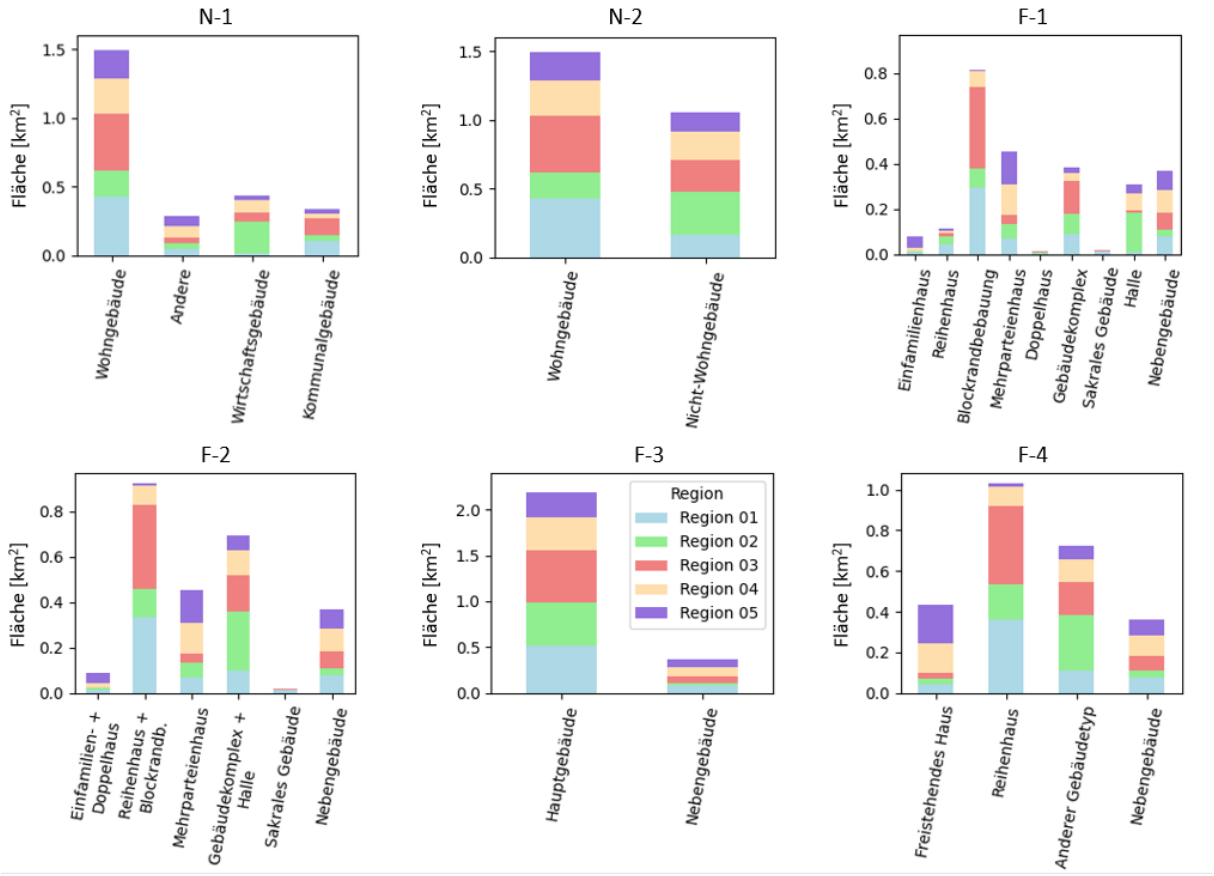

Abbildung 6 Summierte Fläche der einzelnen Gebäudetypen der sechs Klassengruppen für die Untersuchungsgebiete der Regionen 1-5.

Nutzung

Die vier Gebäudetypen zur Gebäudenutzung in Klassengruppe N-1 werden so gewählt, dass eine Klassifizierung durch CNNs, entgegen der Tatsache, dass die Gebäudefunktion häufig nicht mit äußeren Erscheinungsmerkmalen zusammenhängt, grundsätzlich möglich erscheint. Diese Annahme beruht auf dem, auch in der Architektur häufig angewandten, bekannten Designprinzip „Form Follows Function“, nach dem sich die Form eines Gebäudes vor allem aus dem Zweck des Gebäudes ergeben sollte (SUNDBORG et al. 2019). Um eine zu hohe Komplexität der Klassengruppen zu vermeiden, werden vier generalisierte Klassen gewählt und es wird so auf eine detailliertere Einteilung verzichtet. Sind in einem Gebäude mehrere Nutzungstypen vorhanden, so wird meist diejenige Nutzung angegeben, die den überwiegenden Teil des Gebäudes beansprucht.

Der Gebäudetyp Wohngebäude dieser Klassengruppe dient zur Unterbringung von Personen, wohingegen die restlichen Gebäudeklassen eine andere Funktion erfüllen. Wirtschaftsgebäude dienen dem Wirtschaften und umfassen beispielsweise Lagerhallen, Werkstätten oder Supermärkte. Kommunalgebäude werden im Gegensatz hierzu meist von der Kommunalverwaltung, also der öffentlichen Verwaltung, geleitet und verwaltet. Hinzu kommen zu diesem

Begriff Gebäude der sozialen Infrastruktur wie Krankenhäuser oder Seniorenheime. Weitere Beispiele für die Klasse der Kommunalgebäude sind Museen oder Schulen. Alle weiteren Gebäudetypen, auch solche ohne festgestellte Nutzung, werden unter der Klasse „Andere“ geführt. HOFFMANN et al. (2019) wählten eine ähnliche Einteilung, wobei bei ihnen die Klasse „Andere“ „Industriegebäude“ sind. Sie geben ebenso an, dass sich diese Klassifizierung besser für die Detektion mittels Methoden des maschinellen Lernens eignet, als bestehende komplexere Klassifikationsschemata und dass sie für die Gewinnung von weiteren Daten, wie den soziodemografischen Parametern Bevölkerungsdichte oder dem Einkommen in Städten, genutzt werden kann. Verallgemeinert wird diese Klassengruppe in der Klassengruppe N-2, indem alle Gebäudetypen außer Wohngebäude zur Klasse Nicht-Wohngebäude zusammengefasst werden. Gerade im Naturgefahrenmanagement können mit dieser Unterscheidung wichtige Aussagen zum Schadensausmaß getroffen werden (BHUYAN et al. 2022). Die Nutzungsklassen der Ausgangsklassengruppe N-1 werden primär aus dem im vorherigen Kapitel angesprochenen Attribut der Gebäudefunktion der verwendeten 3D-Gebäudemodelle gebildet, wobei die Gebäudefunktionsklassen der Gebäudemodelle zusammengefasst werden. Wo diese Information nicht vorhanden ist, werden den Gebäuden manuell eine passende Klasse zugeteilt. Ist dies nicht möglich, so wird die Klasse „Andere“ angewandt. Das komplette Schema, welche der Klassen der LoD-Daten jeweils welche Klassen der Klassengruppe N-2 bilden, ist in Anhang 1 angegeben. In der Klasse der Wirtschaftsgebäude der Klassengruppe N-1 sind so beispielsweise Gebäude der Klassen Fabrik, Gewächshaus oder Werkstatt zu finden. Da die Gebäudemodelle für alle Untersuchungsgebiete vorliegen, sind die Nutzungsklassen in allen sechs Untersuchungsgebieten vorhanden.

Form

Der Klassifizierung nach der Nutzung der Gebäude steht die Klassifizierung nach der äußeren Form der Gebäude gegenüber. Klassengruppe F-1 beinhält hierfür neun Gebäudetypen. Ein Beispiel jedes Gebäudetyps in Klassengruppe F-1 ist in Abbildung 7 ersichtlich.

Abbildung 7 Beispielhafte Gebäude der Gebäudetypen der Form-Klassengruppe F-1. a: Einzelhaus, b: Doppelhaus, c: Reihenhaus, d: Blockrandbebauung, e: Mehrparteienhaus, f: Gebäudekomplex, g: Sakralgebäude , h: Halle, i: Nebengebäude (GOOGLE EARTH 2023).

Die Einteilung orientiert sich hier an gängigen Klassifikationsschemata, wie dem des Instituts für Wohnen und Umwelt, welches die hier aufgeführten Gebäudetypen Einfamilienhaus, Reihenhaus und Mehrfamilienhaus zur Abschätzung des Gebäudeenergiebedarfs verwendet (LOGA et al. 2015).

Wie bei BANDAM et al. (2022) sind in der vorliegenden Einteilung Einfamilienhäuser frei stehende Gebäude ohne Abtrennung zu anderen Gebäuden. Doppelhäuser sind Gebäude, welche mit einem anderen Gebäude verbunden sind und Reihenhäuser sind mit zwei Häusern verbunden. Ein Mehrparteien- bzw. Mehrfamilienhaus besteht aus mehreren, auf mehrere Geschosse verteilten Wohnungen. Dieser Unterschied zum Reihenhaus ist gut in dem Vergleich von Abbildung 7c und Abbildung 7e zu erkennen. Blockrandbebauung bezeichnet eine Bebauung von Gebäuden um einen gemeinsamen Kern, meist einen Innenhof, herum (SUNDBORG et al. 2019). Bilden mehrere Gebäude eine solche Gruppierung, so sind sie als Blockrandbebauung klassifiziert. In anderen Definitionen ist keine klare Grenze zwischen Blockrandbebauung und Mehrfamilienhäusern vorhanden, da mehrere Mehrparteienhäuser in geschlossener Bauweise hier eine Blockrandbebauung bilden (WURM et al. 2021). Hallen sind Gebäude mit meist klar erkennbarer rechteckiger Bauweise. Sakralbauten umfassen alle Arten von religiösen Gebäuden, überwiegend Kirchen. Die letzte Klasse der Nebengebäude wird durch sekundäre Gebäude gekennzeichnet, die meist eine kleine Grundfläche besitzen und von einem Hauptgebäude abgetrennt sind. Beispiele hierfür sind Schuppen, Schrebergartenhütten oder Garagen.

Die Attribute zur Klassengruppe F-1 wurden vom DLR zur Verfügung gestellt. Die ersten fünf Klassen dieser Klassengruppe wurden manuell auf Basis des DOPs sowie von Google Earth erstellt, indem jedem Gebäude eine der Klassen zugewiesen wurde. Die Klassen Halle, Sakralbau und Nebengebäude wurden primär, wie die Klassen zur Gebäudenutzung, aus dem Attribut der Gebäudefunktion der 3D-Gebäudemodelle gebildet und sind hierdurch auch der Nutzungsart zuzuordnen. Durch die meist signifikante Form einer Halle, einer Kirche oder eines Schuppens wurden diese Gebäudetypen jedoch der Form-Klassengruppe zugeordnet. In einem zweiten Schritt wurden Gebäude dieser drei Klassen manuell überarbeitet und ggf. angepasst.

Für die Klassen der Klassengruppe F-2 wird als Basis die Konfusionsmatrix der Klassengruppe F-1 herangezogen. Hiermit kann analysiert werden, welche Klassen das Modell häufig miteinander verwechselt. So werden primär Klassen zusammengelegt, die das Modell nicht gut unterscheiden kann.

Klassengruppe F-4 basiert zwar auf Klassengruppe F-1, hier werden die benötigten Gebäudetypen jedoch manuell angepasst, um den hier verwendeten Klassen zugeordnet zu werden. Freistehende Gebäude umfassen alle Gebäude, die rein optisch eine Gebäudeeinheit bilden und Reihenhäuser sind durch unterschiedliche aneinandergebaute Gebäudestrukturen gekennzeichnet. Nebengebäude sind dieselben wie in Klassengruppe F-1 und die Klasse der „anderen“ Gebäude umfasst Klassen wie Hallen und Gebäudekomplexe. Besonders die Klasse der Mehrparteienhäuser der Klassengruppe F-1 wird für die Gruppe F-4 in freistehende- und Reihenhäuser aufgeteilt.

Die Daten zum Gebäudetyp, definiert nach der Form, liegen nur für die fünf Hauptuntersuchungsgebiete vor und nicht für Region 6. Somit werden für die Form-Klassen nur Region 1 bis 5 verwendet.

3.2 Datenvorprozessierung und -vorbereitung

Nach NEUPANE et al. (2021) und YU et al. (2021) kann die Datenvorverarbeitung vor der Übergabe an das Modell unterteilt werden in die Datenvorprozessierung, welche die Eigenschaften der Daten auf Pixel- oder spektraler Ebene verändert, und die Datenvorbereitung, welche Methoden zur Vorbereitung von Labels sowie die Unterteilung in Trainings-, Test- und Validierungsdatensatz anwendet.

3.2.1 Datenvorprozessierung

Um eine einheitliche Datengrundlage zu erhalten, wird die ursprüngliche Auflösung des nDOMs von 50 cm mit der Resampling-Methode der B-Spline Interpolation nach LEE et al. (1997) auf 10 cm verrechnet. Mit derselben Methode werden außerdem das DOP sowie das nDOM in den Auflösungen 10 cm, 20 cm und 50 cm erstellt. Die Gebäudedaten werden von dem ursprünglichen Vektorformat in das Rasterformat mit einer Auflösung von 10 cm überführt. Für gröbere Auflösungen der Gebäudedaten wird der Modalwert der Rasterzellen verwendet. Wie in Kapitel 3.1.3 beschrieben, werden grundsätzlich nur die Daten der Klassengruppen N-1, F-1 und F-4 als Raster benötigt, da aus N-1 und F-1 die restlichen Klassengruppen gebildet werden können. Für alle Rasterdatensätze mit unterschiedlichen Auflösungen wird je Region dieselbe Datenvorprozessierung angewandt, deren Ablauf in Abbildung 8 dargestellt ist.

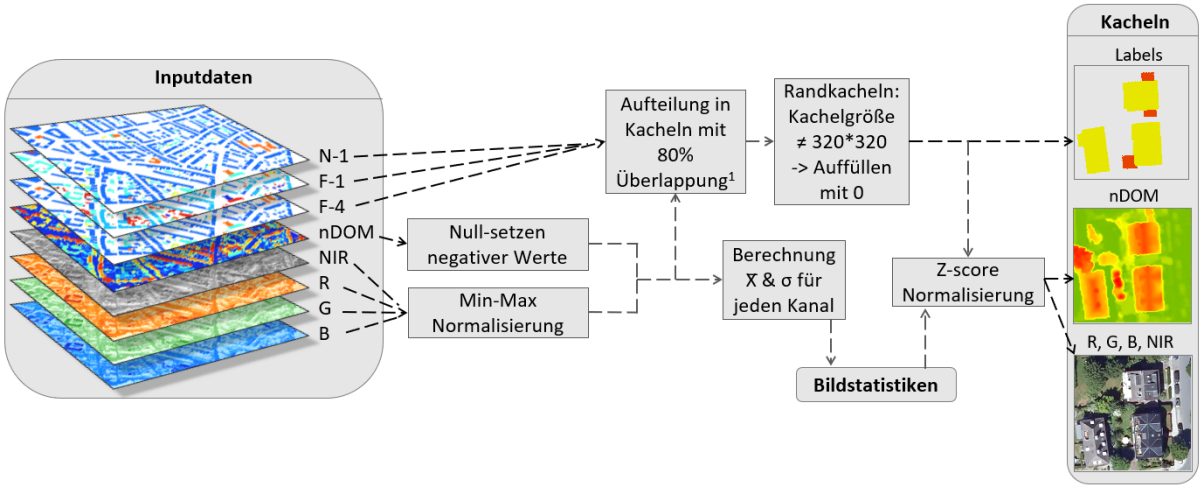

Abbildung 8 Workflow der Datenvorprozessierung für jede Region, $\bar{X}$ = Mittelwert, σ = Standardabweichung, [1] Region 6 wird ohne Überlappung geteilt.

Der folgende Text ist nach den einzelnen Abläufen der Abbildung 8 gegliedert.

Min-Max Normalisierung

Für die Bildkanäle (R, G, B, NIR) steht zu Beginn die Datennormalisierung in Form einer Skalierung zwischen einem festgelegten Minimum und Maximum (Min-Max Normalisierung). Im vorliegenden Fall wird hierfür der Bereich zwischen 0 und 65535 gewählt, was dem Wertebereich des vorzeichenlosen 16-Bit Integer Datentyp (UInt16) entspricht, welcher für alle Eingangsdaten einheitlich gewählt wird. Grund hierfür ist, dass die Höhenangaben des nDOMs so ohne Nachkommastelle in Zentimetern angeben werden können (maximaler Höhenwert: 8100 cm).

Die Formel der Min-Max Normalisierung lautet nach SINGH & SINGH (2020):

$$x'_{i,n} = \frac{x_{i,n} - min(x_i)}{max(x_i) - min(x_i)}(nMax - nMin) + nMin$$

wobei *min* und *max* den minimalen bzw. maximalen Wert des i-ten Merkmals bezeichnen. Die unteren und oberen Grenzen für die Neuskalierung der Daten werden mit *nMin* (hier 0) bzw. *nMax* (hier 65535) bezeichnet. Daten-Normalisierung mittels Min-Max Normalisierung stellt eine häufig verwendete Technik der Datenvorprozessierung im Bereich des maschinellen Lernens dar, da hiermit die Pixelintensität geändert wird, was zu einem größeren Kontrast der Daten führen kann. Zudem erleichtert die Normalisierung dem Modell das Lernen der Beziehungen zwischen den einzelnen Daten, was in neuronalen Netzen das Modelltraining beschleunigen und verbessern kann (RAJU et al. 2020; HAN et al. 2023). Für das nDOM wird die Normalisierung nicht angewandt, um einen Verlust der originalen Informationen des Höhenunterschieds zu vermeiden.

Null-setzen negativer Werte

Stattdessen werden in dem nDOM vorhandene negative Werte auf null gesetzt. Solche negative Werte könnten aus den unterschiedlichen Erfassungszeitpunkten des DOMs sowie des DGMs stammen oder aus den in Kapitel 3.1.2 dargelegten unterschiedlichen Erstellungsmethoden dieser zwei Ausgangshöhenmodelle resultieren.

Aufteilung in Kacheln mit 80% Überlappung

Da der Speicherplatz des Grafikprozessors (GPU) begrenzt ist, können CNNs in vielen Szenarien nicht direkt das gesamte Bild in Originalgröße als Eingabe verwenden, insbesondere nicht bei Megapixelrastern wie im vorliegenden Fall (AN et al. 2020). Da ein zu großes Downsampling aufgrund des damit verbundenen Informationsverlusts keine Option darstellt, werden die Eingangsraster in mehrere kleine Kacheln mit einer Größe von 320*320 Pixeln unterteilt. Bei einer Auflösung von 10 cm können klassische Einfamilienhäuser mit dieser Größe komplett in einer Kachel dargestellt werden. Eine solche Unterteilung in Kacheln wird in 61 von 71 durch NEUPANE et al. (2021) analysierten Studien zur semantischen Segmentierung von Objekten im urbanen Raum mittels Deep-Learning Methoden in der Fernerkundung angewandt. Im Zuge der Kachelung wird eine Überlappung der einzelnen Kacheln von 80 % in x- und y-Richtung angewandt. Hierdurch vergrößert sich zum einen der vorhandene Datensatz zum Training und Test des Modells. Zum anderen können so Randeffekte bei dem Test des Modells umgangen werden, welche dadurch entstehen können, dass das Modell am Rand der Kacheln zu wenig Umgebungsinformationen besitzt, um den Gebäudetyp korrekt zu klas-

sifizieren. Dies wird in Kapitel 3.3.3 näher erläutert. Eine Ausnahme stellt Region 6 dar, in welcher keine Überlappung angewandt wird. Ursache hierfür ist die größere Fläche der Region und dass diese nicht als Testgebiet verwendet wird.

Nach Abschluss der Unterteilung in Kacheln unter Anwendung der Überlappung ergibt sich für Region 1 – 5 jeweils eine Kachelanzahl von 46.898 statt 1.908 ohne Überlappung und für Region 6 auf Grund des größeren Gebiets aber der fehlenden Verwendung von Überlappungen eine Kachelanzahl von 32.148.

Randkacheln: Auffüllen mit Null-Werten

Für Kacheln am Rand der Untersuchungsregionen, deren Größe nicht exakt 320*320 beträgt, werden die restlichen Pixel mit Null-Werten in allen Kanälen aufgefüllt. Dies ist nötig, wenn die Größe der Region nicht durch 64 teilbar ist (320 * 20 % (Pixel ohne Überlappung) = 64).

Berechnung $\bar{X}$ & σ für jeden Kanal

Wie in Abbildung 8 ersichtlich, erfolgt nach der Min-Max Normalisierung sowie dem Null-Setzen negativer nDOM-Werte die Berechnung von Bildstatistiken für alle Kanäle zur späteren Z-score Normalisierung. Hierfür wird für jeden Kanal (R, G, B, NIR, nDOM) je Region der Mittelwert sowie die Standardabweichung aller Werte berechnet. Um einen Wert für alle Regionen zu erhalten, wird der Mittelwert der zuvor erhaltenen Mittelwerte sowie der Standardabweichung gebildet. Die Anwendung dieser zwei berechneten Variablen auf den Datensatz erfolgt nach der Aufteilung der Datensätze in kleinere Kacheln.

Z-score Normalisierung

Nach der Kachelung der Daten werden diese mit den zwei zuvor gewonnenen Variablen des Mittelwerts $\bar{X}$ sowie der Standardabweichung σ je Kanal mit der folgenden Formel normalisiert: $x'_{i,n} = \frac{x_{i,n} - \bar{X}_i}{\sigma_i}$. Diese Umskalierung führt zu einem Mittelwert der Daten von Null und einer Standardabweichung von eins. Ausreißer haben in den Daten hiernach einen deutlich kleineren Einfluss auf das Modell. Teils wird diese Methode auch als Standardisierung bezeichnet, wobei die Begriffe Normalisierung und Standardisierung in diesem Zusammenhang in der Literatur häufig synonym verwendet werden (HAN et al. 2023). Nach SINGH & SINGH (2020) führt diese Methode im Vergleich zu 13 anderen Normalisierungsmethoden zu den besten Ergebnissen bei dem von ihnen verwendeten *Machine Learning* Modell.

3.2.2 Datenvorbereitung

Nach der Datenvorprozessierung werden die in Kacheln unterteilten Daten anschließend in Test-, Validierungs- und Trainingsdaten aufgegliedert (Datenvorbereitung). Der Ablauf dazu ist in Abbildung 9 ersichtlich. Nach den einzelnen Schritten dieser Abbildung ist der folgende Text gegliedert.

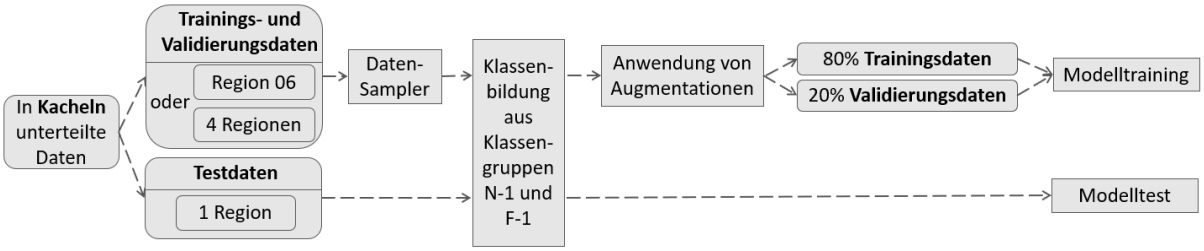

Abbildung 9 Workflow der Datenvorbereitung zur Anwendung im Modell.

Auswahl Trainings- und Validierungsdaten

Zu Beginn steht die Auswahl von Daten zum Test des Modells. Damit die Ergebnisse dieses Tests repräsentativ sind, darf das Modell nicht mit diesen trainiert worden sein. Um außerdem eine Bewertung treffen zu können, wie gut das Modell in einer räumlich von dem Trainingsgebiet getrennten Region anwendbar ist, werden die Kacheln einer kompletten Region nur als Testdaten verwendet.

Daten-Sampler

Zweiter Schritt ist die Auswahl einer Teilmenge der Daten für das Modelltraining. In Abbildung 9 geschieht dies im sog. Daten-Sampler. Diese Datenauswahl erfolgt, wenn nicht alle Kacheln verwendet werden sollen. Ist dies der Fall, so werden zufällig 25.000 Kacheln aus den Kacheln aller Regionen ausgewählt sofern in diesen Kacheln mindestens 500 Pixel zu irgendeinem Gebäudetyp gehören, also keine Hintergrundpixel sind. Die Auflösung spielt hierbei keine Rolle. Eine beispielhafte zufällige Auswahl von 1000 Kacheln mit mindestens 500 Gebäudepixeln ist in Abbildung 10 ersichtlich.

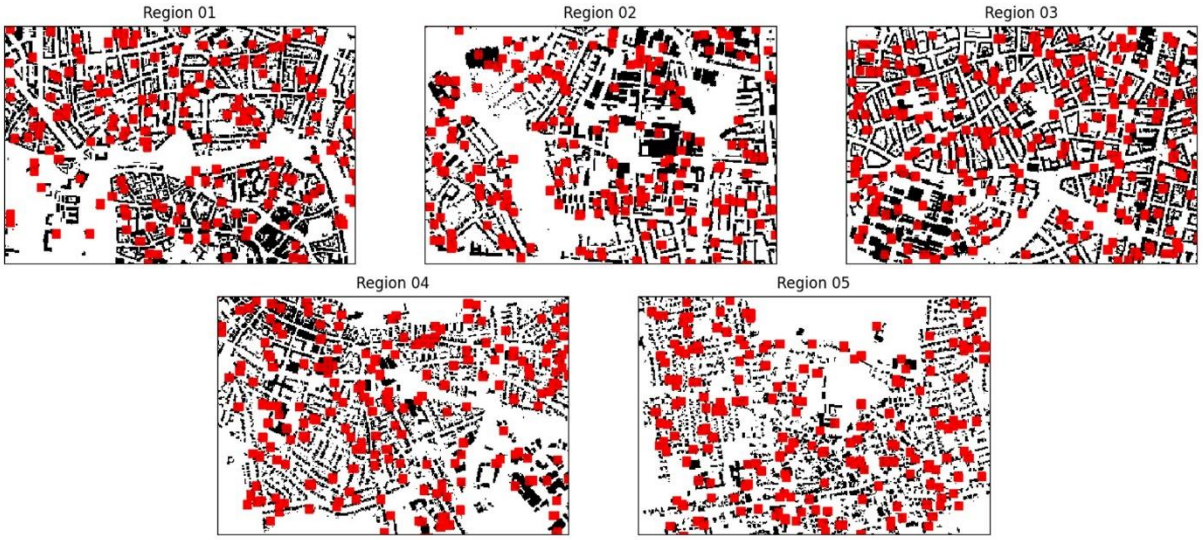

Abbildung 10 Exemplarische Darstellung von 1000 durch den Daten-Sampler ausgewählte Kacheln (rot), welche jeweils mindestens 500 Gebäudepixel beinhalten, auf allen Gebäuden (schwarz).

Erkennbar ist, dass in Regionen ohne Gebäude keine Kacheln ausgewählt werden und dass durch die Zufallsauswahl die verwendeten Kacheln über alle Regionen gleichmäßig verteilt sind. Mit dieser Datenauswahl wird folglich sichergestellt, dass auf den ausgewählten Kacheln, mit welchen das Modell trainiert wird, ausreichend Gebäudepixel vorhanden sind. Für den Modelltest werden alle Kacheln der Testregion verwendet.

Klassenbildung

Da in den unterteilten Kacheln die drei Labelraster der Klassengruppen F-1, F-4 und N-1 vorhanden sind, wird nach der Datenauswahl die benötigte Klassengruppe aus den zwei Basisgruppen F-1 und N-1 gebildet bzw. nur eine davon ausgewählt. Dies geschieht analog zu den in Kapitel 3.1.3 vorgestellten Klassengruppen.

Anwendung von Augmentationen

Anschließend werden auf die ausgewählten Trainings- und Validierungsdaten Augmentationen angewandt. Diese werden nach NEUPANE et al. (2021) von der Mehrheit der Studien zur Erkennung von Objekten im urbanen Raum mittels Deep-Learning verwendet, da sich hiermit der Trainingsdatensatz vergrößern lässt und sich die Leistung des gesamten Lernverfahrens des Modells steigern lässt. Ursache für letzteres ist, dass die Generalisierung des Modells, also die Leistung bei der Anwendung auf unbekannte Daten wie den Testdatensatz, verbessert wird, wenn es an unterschiedliche Szenarien der Gebäude angepasst ist. Traditionelle Augmentationstechniken umfassen unter anderem die hier angewandten Methoden. Diese werden ausgewählt, da sie relativ einfach zu implementieren sind und die Generalisierung des Modells verbessern können (SHORTEN & KHOSHGOFTAAR 2019). Im Speziellen sind dies die ho-

rizontale Spiegelung, die vertikale Spiegelung sowie die Gruppe aus Verschiebung, Skalierung und Rotation der Bilder, welche jeweils mit einer Wahrscheinlichkeit von 50 % auf alle Kacheln angewandt werden. Die Wahrscheinlichkeit, dass auf eine Kachel keine Augmentation angewandt wird liegt demnach bei 12,5 %. Die jeweilige Augmentation wird dabei für alle Kanäle sowie das Label der Originalkachel angewandt. In Abbildung 11 sind die angewandten Methoden an einem Beispielbild dargestellt.

Abbildung 11 Beispielhafte Darstellung der angewandten Daten Augmentierungs-Techniken.

Aufteilung in Trainings- und Validierungsdaten

Zuletzt steht bei der Datenvorbereitung die zufällige Aufteilung in Trainings- und Validierungsdaten. Hierfür wird ein Verhältnis von 4:1 angewandt, was zusammen mit dem Verhältnis 9:1 das zumeist verwendete Verhältnis in Studien zur Erkennung von Objekten im urbanen Raum mittels Deep-Learning darstellt (NEUPANE et al. 2021).

3.3 Modelltraining mit unterschiedlichen Parametern und Leistungsevaluation

Um den Einfluss eines Parameters auf das Modellergebnis zu bewerten, werden mehrere Modelltrainings durchgeführt, wobei die restlichen Parameter jeweils identisch belassen werden. Alle Tests eines Parameters werden als ein Experiment bezeichnet. Die durchgeführten Experimente umfassen einerseits Parameter des Modells bzw. CNNs (Kapitel 3.3.2) wie z. B. die Epochenanzahl und andererseits Parameter der Daten (Kapitel 3.3.1) wie z. B. die Datenauflösung.

Wird die Einstellung eines Parameters geändert, so wirft dies die Frage auf, welche Standard-Einstellungen für die restlichen Parameter getroffen werden sollten. Die Wahl der Standard-Einstellung je Parameter wird im Folgenden jeweils begründet und wird überwiegend für alle Experimente beibehalten, auch wenn sich bei dem Test eines Parameters herausstellt, dass eine andere als die Standard-Einstellung das beste Ergebnis liefert. Dies ist nötig, um Vergleichbarkeit zwischen den einzelnen Modellanwendungen und Experimenten zu gewährleisten. Wird eine andere als die Standard-Einstellung verwendet, so ist dies angegeben.

Um nach dem Modelltraining die Leistung des Modells auf einem unabhängigen Testdatensatz zu bewerten, wird das trainierte Modell im Modelltest auf eine separate Region angewandt. Dieses Vorgehen wird später in Kapitel 3.3.3 erläutert, worauf in Kapitel 3.3.4 dargestellt wird, mit welchen Metriken das Modellergebnis bewertet wird.

3.3.1 Getestete Daten-Parameter

Um den Einfluss unterschiedlicher Daten auf die Erkennung unterschiedlicher Gebäudetypen zu erfassen, werden sieben Experimente mit unterschiedlichen Eingangsdaten in das Modell durchgeführt, wobei für die restlichen Parameter die Standard-Einstellungen beibehalten werden. Diese Experimente betreffen demnach Daten-Parameter, welche zum Teil bereits im Daten-Kapitel 3.1 angesprochen wurden. In Tabelle 3 sind die Daten-Parameter mit den getesteten Einstellungen aufgeführt. Ausgewählt werden diese Parameter in der Annahme, dass sie den größten Einfluss auf die semantische Segmentierung von unterschiedlichen Gebäudetypen haben. Damit wird versucht herauszufinden, welche Eingangsdaten die Entscheidung des Modells wie stark beeinflussen. Zudem soll getestet werden, ob das Modell mit seiner Parametrisierung für eine deutschlandweite Anwendung in Frage kommt.

Tabelle 3 Getestete Daten-Parameter mit den jeweiligen Einstellungen. Standard-Einstellungen sind fett gedruckt markiert.

Parameter	Gewählte Einstellung
Gebäudeklassengruppen	N-1, N-2, F-1, **F-2**, F-3
Untersuchungsgebiete (UGs)	Region 1-5: Test in einem UG und Training in jeweils Restlichen, **Standard: Testregion ist Region 1** Region 6: Training in Region 6 und Test in Region 1
Auflösung	**10**, 20, 50 cm
Datenkanäle	RGB, RGB+NIR, RGB+nDOM, **RGB+NIR+nDOM**, nDOM
Anzahl Daten	**25.000 Kacheln** und alle Kacheln
Datenauswahl	**Mit** und ohne **Daten-Sampler**
Seed-Werte	5 verschiedene Zufallszahlen getestet zur Auswahl der Trainingsdaten und Anwendung der Augmentationen, **Standard**: **keine veränderte Zufallszahl**

Wie in dieser Tabelle ersichtlich, werden alle gebildeten Gebäudeklassengruppen getestet, um eine Aussage über die Güte der Erkennung der unterschiedlichen Gebäudetypen nach Form und Nutzung treffen zu können. Der Test der unterschiedlichen Klassengruppen wird den Da-

ten-Parametern zugeordnet, da es sich auch bei den Gebäudetypen um unterschiedliche Daten handelt, welche für das Modelltraining und den Modelltest verwendet werden. Für die meisten Experimente wird als Standard-Einstellung die Klassengruppe F-2 gewählt, da Gebäudetypen der Form auf Basis von Fernerkundungsdaten vermutlich besser erkennbar sind als Klassentypen der Nutzung. Zudem könnte die Klassengruppe F-1 zu detailliert sein, um von dem Modell erkannt zu werden (TAUBENBÖCK et al. 2013).

Um feststellen zu können, ob sich die Gebäudetyperkennung von Region zu Region unterscheidet, werden bis auf Region 6 alle Regionen als Testregion verwendet. Als Trainingsdaten werden die jeweils verbliebenen Regionen ausgewählt. Als Standard Testregion wird Region 1 gewählt, da diese mit ihrem innerstädtischen Bereich in den verschiedenen Klassengruppen alle Gebäudetypen am besten abdeckt (vgl. Abbildung 6). Die Auswahl der Trainingsdaten erfolgt hier standardmäßig aus den restlichen Regionen 1-4. Region 6 wird separat als Trainingsregion getestet, um die Gebäudetypenerkennung mit dem unverbesserten amtlichen Gebäudebestand zu testen.

Um die Unterfrage zu beantworten, ob ein Zusammenhang zwischen der Häufigkeit der Gebäudetypen in den Test- und Trainingsdaten und der Erkennung der einzelnen Gebäudetypen besteht, wird eine Korrelationsanalyse durchgeführt. Hierzu wird der Pearson-Korrelationskoeffizient zwischen den Ergebnis-Werten je Klasse je Test-Region und der Anzahl der Pixel der jeweiligen Klasse in der jeweiligen Testregion sowie in den jeweiligen Trainingsregionen berechnet.

Als weiteres Experiment werden verschiedene Auflösungen von 10, 20 und 50 cm getestet. Ursächlich hierfür ist, dass die Bilddaten mit 10 cm Auflösung für das Untersuchungsgebiet vorliegen. In einer Auflösung von 20 cm sind diese für ganz Deutschland vorhanden und ein Test mit dieser Auflösung lässt so Rückschlüsse darüber zu, wie gut das Modell auf ganz Deutschland angewendet werden könnte (BKG 2023). Die Auflösung von 50 cm ist für Satellitenbilder häufig verfügbar und hat bei PAN et al. (2020) für die Erkennung unterschiedlicher Gebäudetypen bereits gute Ergebnis-Metriken ergeben. Als Standardauflösung wird die höchste verfügbare Auflösung von 10 cm gewählt, da mit dieser am meisten Details erkennbar sind. Zudem geht so durch ein Downsampling keine Information verloren, da nur die Auflösung des nDOMs von 50 auf 10 cm verändert werden muss. Innerhalb des CNNs wird die Auflösung durch mehrere Pooling Layer zudem bereits vermindert.

Wie bei der höchsten Auflösung liegt die höchste Informationsdichte ebenso bei der Wahl von allen verfügbaren Eingangskanälen vor und es werden bei den meisten Experimenten alle vorhandenen Datenkanäle verwendet. Zu den RGB-Kanälen kommt also das nahe Infrarot

(NIR) hinzu, da hiermit Vegetation besser erkannt werden kann und die Hinzunahme dieses Kanals bei ZHANG et al. (2018) zu einer Verbesserung der Unterscheidung in Gebäude und Nicht-Gebäude führte. Hinzu kommt außerdem das nDOM, da dieses die Gebäudeerkennung ebenso stark verbessern kann (LI J. et al. 2022). Getestet werden auch andere, in Tabelle 3 ersichtliche Kanalkombinationen, um herauszufinden, wie bedeutend welche Kanäle für die Klassifikation durch das Modell sind.

Außerdem soll evaluiert werden, ob die Auswahl bestimmter Trainingsdaten eine starke Auswirkung auf das Modell hat. Hierzu wird evaluiert, ob die Verwendung aller Trainingsdatenkacheln (ca. 180.000) das Modell verbessert, da es mehr Daten „sieht“, aus denen es unterschiedliche Repräsentationen der Gebäudetypen erkennen kann. Für die restlichen Experimente wird eine Standard-Anzahl von 25.000 Kacheln verwendet, da hiermit bereits nahezu alle Gebäude des gesamten UGs mindestens einmal in einer Kachel vorhanden sind.

Ursache hierfür ist unter anderem die Verwendung des in Kapitel 3.2.2 vorgestellten, in fast allen Experimenten verwendeten Daten-Samplers, welcher nur Kacheln mit mindestens 500 Gebäudepixeln auswählt. Dessen Nützlichkeit wird in einem Experiment evaluiert.

Der letzte getestete Parameter betrifft ebenso die Auswahl der Trainingsdaten sowie die Anwendung von Augmentationen. Hierfür werden unterschiedliche Zufallszahlen getestet was zu einer anderen akzidentellen Auswahl der 25.000 Kacheln aus dem gesamten Datensatz und einer zufälligen Anwendung von Augmentationen führt. Standardmäßig wird hierbei immer dieselbe Zufallszahl als *Seed* verwendet.

3.3.2 Modelltraining und getestete Modell-Parameter

Für das Modelltraining wird die GPU (Grafikkarte) verwendet. Ursache hierfür ist die Möglichkeit der parallelen Prozessierung auf mehreren Kernen, was zu einer im Vergleich zur CPU (Prozessor) höheren Rechengeschwindigkeit führt (LI J. et al. 2022). Das verwendete Deep-Learning Framework, also die genutzte Programmbibliothek für das Ausführen der CNNs in Python, ist PyTorch (PASZKE et al. 2019). Für die verwendeten CNNs selbst wird die Programmbibliothek *segmentation-models-pytorch* von IAKUBOVSKII (2019) in der Version 0.3.3 verwendet.

Wie in Kapitel 2 erläutert, gibt es mehrere CNN-Parameter, welche die Leistung des CNN beeinflussen können. In Kapitel 2.2 werden diese erläutert und in diesem Kapitel wird auf ihre spezifischen Einstellungen eingegangen. Wie bei den Daten-Parametern werden auch für ausgewählte Modell-Parameter unterschiedliche Einstellungen getestet und es werden Standard-Einstellungen für die restlichen Parameter festgelegt. Im Folgenden wird diese Wahl begrün-

det und es wird auch auf die Einstellung der restlichen Modell-Parameter eingegangen, welche nicht getestet werden. Soweit nicht anders angegeben, werden für die Tests der Daten-Parameter die Standard-Einstellungen der Modell-Parameter verwendet. Analog zu den Experimenten der Daten-Parameter sind die Modell-Parameter mit ihren jeweiligen Einstellungen in Tabelle 4 dargestellt. Parameter, für die die Auswirkung unterschiedlicher Einstellungen getestet werden, sind mit einem (T) markiert.

Tabelle 4 Verwendete CNN-Parameter mit den jeweils gewählten Einstellungen. Für mit (T) markierte Parameter werden mehrere Einstellungen getestet wobei fett markierte Einstellungen die Standard-Einstellungen darstellen.

Parameter	Gewählte Einstellung
Datentyp während Training	16-Bit und 32-Bit für Unet++ und MANet, 32-Bit für Unet und FPN
Batch Größe	16
Aktivierungsfunktion	Softmax2d
Verlustfunktion	Kreuz-Entropie mit Gewichten, Kreuz-Entropie ohne Gewichte (bei Verwendung aller Daten)
Optimierungsalgorithmus	AdamW
Decay	0.05
Lernrate	Angepasst an Epoche und Ergebnis der Verlustfunktion
Encoder	ResNet50
Decoder/CNN-Architekturen (T)	Unet**, Unet++**, FPN, MANet
Epochenanzahl (T)	10, 20, 30, 40, **50**, 70, 200
Vortrainierte Gewichte (T)	Wenn ja vortrainiert auf ImageNet, **Standard: nicht vortrainiert**
Seed-Werte (T)	7 verschiedene getestet, **Standard: deterministisches Modell**

Als **Datentyp** während des Trainings wird für die zwei Decoder Unet++ sowie MANet eine Mischung aus 32-Bit und 16-Bit Gleitkommazahlen (*float*) verwendet, wobei für die wichtigsten Operationen des Modells 16-Bit verwendet werden. Für die Decoder FPN und Unet werden ausschließlich 32-Bit Gleitkommazahlen verwendet. Die Verwendung von 16-Bit kann die Trainingsgeschwindigkeit erhöhen, führt jedoch im Vergleich zur Verwendung von 32-Bit zu einer niedrigeren Genauigkeit (LIGHTNING.AI 2023). Im vorliegenden Fall muss für Unet++ sowie MANet zumeist 16-Bit verwendet werden, da der vorhandene Speicher der GPU andernfalls zu klein für die Berechnungen ist. Die Ursache hierfür könnte die höhere Anzahl an

trainierbaren Gewichten bei Unet++ und MANet darstellen (ZHOU et al. 2018; STILLER et al. 2023).

Als **Batch-**Größe werden 16 Kacheln je Iteration durch das Modell verwendet, was bei dem vorhandenen GPU-Speicher und der Kachelgröße eine ausreichend schnelle Rechendauer und gleichzeitig eine repräsentative Auswahl der Trainingsdaten bedeutet.

Die Wahl der **Aktivierungsfunktion** für den Ausgabelayer hängt typischerweise von dem Vorhersagetyp ab. Für Aufgaben mit mehreren Klassen wird hierbei meist die Softmax-Funktion verwendet, da hiermit eine Wahrscheinlichkeit für jede Klasse berechnet werden kann (LI Z. et al. 2022). Die Wahrscheinlichkeit aller Klassen ergibt anschließend den Wert 1. Für die semantische Segmentierung kommt die räumliche Komponente hinzu, da die Funktion auf 2D-Eingabetensoren angewandt wird, was zu Wahrscheinlichkeitswerten je Pixel führt.

Verwendet wird die Softmax-Aktivierungsfunktion meist mit der Kreuz-Entropie-**Verlustfunktion** (KETKAR & MOOLAYIL 2021). Auch diese Funktion wird für Klassifizierungsaufgaben häufig verwendet und HE et al. (2016) stellten fest, dass sich mit dieser Verlustfunktion und dem in der vorliegenden Studie verwendeten Encoder ResNet gute Ergebnisse erzielen lassen. Die Kreuz-Entropie-Verlustfunktion bewertet den Unterschied zwischen der Wahrscheinlichkeitsverteilung aus dem Modelltraining und der tatsächlichen Verteilung. Hierbei wird die vorhergesagte Wahrscheinlichkeit mit dem tatsächlichen Ausgabewert (0 oder 1) in jeder Klasse verglichen und ein Strafwert wird auf Grundlage des Abstands zwischen diesen Werten berechnet (LI Z. et al. 2022). Wie in Kapitel 2.2 beschrieben, soll dieser Strafwert möglichst verringert werden. Im vorliegenden Fall wird in fast allen Experimenten die Kreuz-Entropie-Funktion mit Gewichten (nicht die Gewichte des CNNs) verwendet, wodurch seltene Klassen stärker in der Verlustfunktion gewichtet werden. Hierfür wird zunächst eine Analyse der Häufigkeitsverteilung jeder Klasse in den Trainingsdaten durchgeführt (ähnlich wie in Abbildung 6, nur für die verwendeten Trainingsdaten). Anschließend wird das Gewicht jeder Klasse folgendermaßen berechnet: $\frac{1}{Anzahl\ Pixel\ Klasse} \times \frac{Pixelanzahl\ gesamt}{Anzahl\ aller\ Klassen}$ und mit dem Strafmaß verrechnet. Somit werden Klassen mit einer geringeren Häufigkeit stärker gewichtet und dem Klassen-Ungleichgewicht kann etwas entgegengewirkt werden. Nur für Experimente mit allen Trainingsdaten wird auf Grund der erreichten Rechenspeichergrenze bei der Gewichtsberechnung die Kreuz-Entropie-Funktion ohne Gewichte verwendet.

Als **Optimierungsalgorithmus** wird AdamW (LOSHCHILOV & HUTTER 2017) verwendet, da hiermit direkt ein **Decay** der Lernrate von 0.05 implementiert werden kann. Zudem vereint

dieser Optimierer nach LI Z. et al. (2022) und KETKAR & MOOLAYIL (2021) Vorteile von drei anderen Optimierungsalgorithmen. Der Wert 0.05 wird gewählt, da er im Vergleich zur Studie zu AdamW von LOSHCHILOV & HUTTER (2017) einen eher hohen Wert darstellt, was zu einer guten Generalisierung des Modells führen soll.

Für die **Lernrate** wird eine Anpassung an die jeweilige Epoche und das Ergebnis der Verlustfunktion implementiert. An die Verlustfunktion ist die Lernrate angepasst, indem sie um den Faktor zehn reduziert wird, wenn das Ergebnis der Verlustfunktion sich über 10 Epochen nicht weiter verringert. So soll das Modell sich der optimalen Verlustfunktion in kleineren Schritten nähern und nicht über das Ziel hinausgehen, wenn keine Verbesserung des Modells vorhanden ist. Mit der Epoche ist die Lernrate verknüpft, da die Lernrate gebildet wird, indem sie mit der initialen Lernrate (0,0001) je Epoche mit einem anderen Faktor multipliziert wird. Für die ersten vier Epochen bildet sich dieser Faktor mit folgender Formel: $0{,}1^{5-Epoche}$ und für die restlichen Epochen mit $0{,}95^{Epoche}$. So ist gewährleistet, dass das Modell mit einer größeren Lernrate (Lernrate steigt bis zur vierten Epoche) anfangs schnell dazulernt, um sich anschließend dem Optimum in kleineren Schritten (Lernrate sinkt ab der fünften Epoche) anzunähern (LI Z. et al. 2022).

Als **Encoder** wird ResNet50 (HE et al. 2016) gewählt, da STILLER et al. (2023) für die Gebäudesegmentierung die Encoder-Decoder Architektur von ResNet50 und FPN als effizienteste Kombination feststellen, welche eine gute Modellleistung und gleichzeitig eine kurze Trainingsdauer aufzeigt.

Als Modellarchitektur bzw. **Decoder** werden Unet (RONNEBERGER et al. 2015), Unet++ (ZHOU et al. 2018), FPN (LIN et al. 2017) und MANet (FAN et al. 2020) getestet. FPN wird ausgewählt, da dieses in der Studie von STILLER et al. (2023) bei der Erkennung von Gebäuden auf Luftbildern die beste Effizienz zeigt. Zudem wird die Gebäudetypenerkennung mit Unet durchgeführt, da ROBINSON et al. (2022) hiermit gute Präzisions- und Sensitivitätswerte bei der Erkennung des Gebäudetyps der Geflügelfarmen auf Luftbildern feststellen. Zudem verwenden generell die meisten Studien, welche unterschiedliche Decoder vergleichen, Unet (NEUPANE et al. 2021; AMIRGAN et al. 2022; STILLER et al. 2023). Dies macht es zu einer Art Benchmark-Model im CNN-Bereich. Unet++ wird getestet und als Standard-Decoder ausgewählt, da dieses in der Studie von ZHOU et al. (2018) eine bessere Leistung zeigt, als Unet. Außerdem wird MAnet verwendet, da AMIRGAN et al. (2022) in ihrem Decoder-Vergleich zur Erkennung von Gebäuden mittels Luftbildern für diesen Decoder die besten Metriken erzielen.

Als Standard-Decoder wird nur für den Test der Verwendung aller Kacheln von Unet++ abgewichen und Unet gewählt, da mit Unet++ der Rechenspeicher des Rechners zu klein ist. Ursache hierfür ist, dass Unet weniger Gewichte beinhält, als Unet++.

Als Standard-**Epochen**-Anzahl werden 50 Epochen festgelegt, wobei Epochenanzahlen zwischen 10 und 200 getestet werden. Diese Epochenanzahl führt zu einer Trainingsdauer von, je nach Parametern, 14 bis 40 Stunden. Auf Grund der begrenzten Rechenressourcen und der Vielzahl von ausgeführten Experimenten stellt dies eine auf der vorhandenen Recheninfrastruktur durchführbare Länge dar. Grundsätzlich sollte jedoch die Evaluation der Modellleistung die Epochenanzahl bedingen (LI Z. et al. 2022). Als Modell für den Modelltest wird nicht das Modell der 50. Epoche ausgewählt, sondern das Modell, das mit dem Validierungsdatensatz während des Trainings den höchsten IoU-Wert erreicht. Hiermit wird einer Überanpassung entgegengewirkt, was in Abbildung 12 dargestellt ist. Der englische Fachbegriff in der Literatur hierfür lautet *early stopping* (BALL et al. 2017).

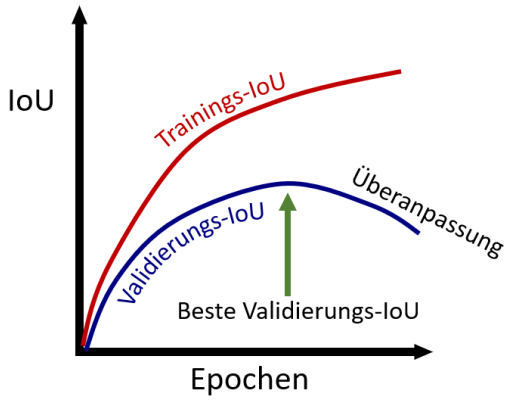

Abbildung 12 Exemplarische Darstellung einer Überanpassung des Modells ab einer bestimmten Epochenanzahl.

Nur für den Test mit allen Trainingskacheln wird eine Epochenanzahl von 19 Epochen verwendet, da das Training mit 50 Epochen ca. 14 Tage in Anspruch nehmen würde.

Der vorletzte getestete Modell-Parameter betrifft **vortrainierte Gewichte**, da solche die Effizienz und Leistung eines Modells verbessern können (GUPTA et al. 2022). Standardmäßig wird das Modell von Grund auf mit den Trainingsdaten trainiert. Um jedoch zu ermitteln, wie sich die Gebäudetyperkennung durch die Übernahme von Gewichten des ersten Convolutional Layers eines auf den in Kapitel 1.1.3 angesprochenen ImageNet-Datensatz vortrainierten Netzwerks verändert, wird dies als Experiment durchgeführt. Hierbei wird also analog zu der in Kapitel 2.2 vorgestellten Methode vorgegangen. In diesem Fall werden vier Modellausführungen für Klassengruppe F-2 und drei Ausführungen für die Klassengruppe N-2 getestet und verglichen:

1. Keine vortrainierten Gewichte werden verwendet (Zufällige Werte der Gewichte zu Beginn).
2. Nur für die RGB-Kanäle werden die jeweiligen Gewichte der RGB-Kanäle aus dem vortrainierten Modell übernommen. Für die restlichen Kanalgewichte werden dieselben zufälligen Gewichtswerte übernommen wie in Ausführung 1.
3. Die Gewichte der RGB-Kanäle werden wie in Ausführung 2 gewählt. Für den NIR-Kanal wird das vortrainierte Gewicht des roten Kanals und für das nDOM das vortrainierte Gewicht des grünen Kanals verwendet.
4. (Nur für Klassengruppe F-2) Die Gewichte der RGB-Kanäle und des nDOMs werden wie in Ausführung 4 gewählt und für den NIR-Kanal wird das Gewicht des grünen Kanals verwendet.

Von Interesse ist hier, ob auf unterschiedliche Datenkanäle vortrainierte Gewichte die Modellleistung verändern, wenn sie für andere Datenkanäle verwendet werden. Dies umfasst insbesondere die Verwendung von auf RGB-Bilder vortrainierte Gewichte für Höhenmodelle.

Der letzte Modell-Test betrifft die in Kapitel 3 angesprochenen Zufallszahlen, also ***seed*-Werte**. Standardmäßig wird hier für alle Algorithmen derselbe Wert verwendet. Um jedoch den Einfluss des Zufalls auf die Leistung des Modells zu erkennen, werden sieben Zufallswerte gewählt und die Modelle hiermit ausgeführt. Von Bedeutung ist dieses Experiment deswegen, weil nach FELLICIOUS et al. (2020) durch die Ausführung eines Modells mit mehreren Zufallszahlen seine Robustheit gegenüber dem Zufall besser eingeschätzt werden kann. Hierbei verwenden Algorithmen innerhalb des Modells für Zufallszahlen also unterschiedliche Werte. Ein Beispiel hierfür sind die initialen Werte der Gewichte bei einem nicht-vortrainierten Modell. Hierbei handelt es sich nur um Zufallszahlen für das CNN, nicht jedoch um die Zufallszahlen für die Trainingsdatenbearbeitung.

3.3.3 Modelltest

Um die Leistung des trainierten Modells in Form von bestimmten Metriken mit unabhängigen Daten zu bestimmen, ist ein Test des Modells essentiell. Der Ablauf dieses Modelltests ist in Abbildung 13 zusammengefasst.

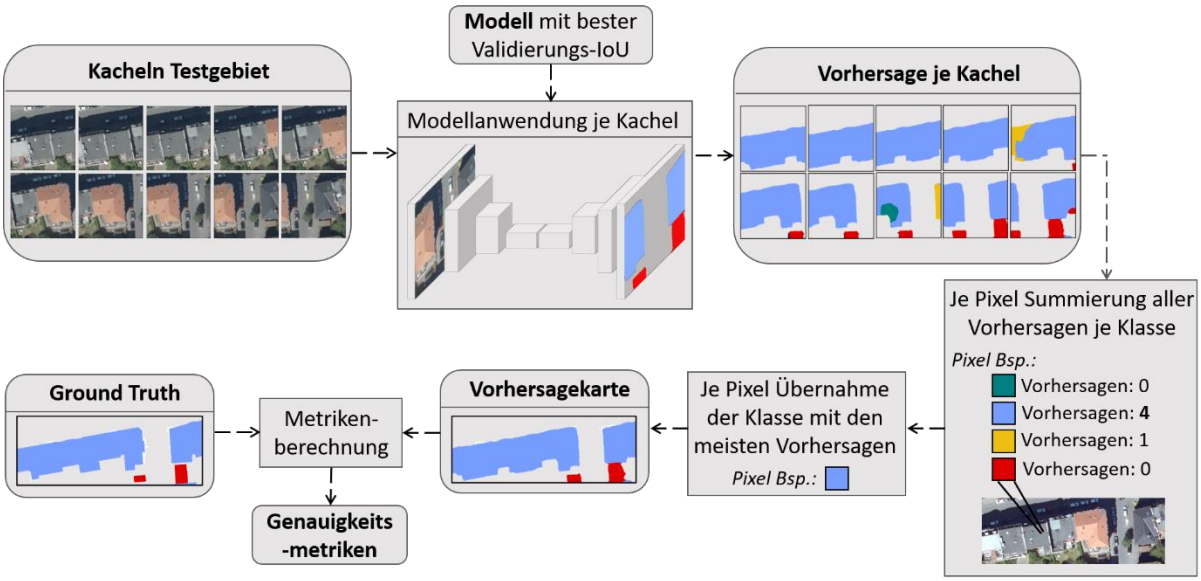

Abbildung 13 Workflow des Modelltests.

Wie in Kapitel 3.2.2 beschrieben, wird für den Modelltest eine komplette, von den Trainingsregionen räumlich abgegrenzte, Region verwendet. Auf die einzelnen Kacheln dieser Region wird das zuvor trainierte Modell angewendet. Im vorliegenden Fall wird das Modell der Epoche genommen, dass auf den Validierungsdatensatz während des Modelldurchlaufs die höchsten IoU-Werte erreicht hat. Es wird also nicht zwingend das Modell der letzten Epoche angewandt. Nach der Anwendung auf den Testdatensatz existiert für jede der 46.898 Kacheln der Testregion je Pixel eine Vorhersage des Gebäudetyps oder der Klasse „Hintergrund“. Da die einzelnen Kacheln sich gegenseitig überlappen, ist der nächste Arbeitsschritt die Zusammenfassung der Vorhersagen der einzelnen Kacheln je Pixel. Dies ist nötig, da das Modell auf unterschiedlichen Kacheln zwar dasselbe Pixel bzw. anschaulicher dasselbe Gebäude, aber unterschiedlichen Kontext hierzu „sieht“. Hieraus folgt, dass die Klassifizierung jedes Pixels nicht zwangsläufig auf jeder Kachel dieselbe ist.

Zur Zusammenfassung aller Vorhersagen je Pixel werden zunächst je Pixel alle Vorhersagen je Klasse aufsummiert. In dem Beispielpixel in Abbildung 13 wird viermal die blaue Klasse (Reihenhäuser) und einmal die orange Klasse (anderer Gebäudetyp) vorhergesagt. Im Anschluss hieran wird für die letztendliche Vorhersagekarte je Pixel die Klasse übernommen, welche die meisten Vorhersagen besitzt. Für das Beispielpixel ist dies die blaue Klasse („Reihenhaus“).

Grund für die Wahl dieses Vorgehens ist, dass so unterschiedliche Kontextinformationen in die Erstellung der Vorhersagekarte einfließen, da es auf Grund von Rechenspeichereinschränkungen nicht möglich ist, die komplette Testregionen auf einmal durch das Modell zu klassifizieren.
Die so erstellte Vorhersagekarte wird anschließend mit den vorhandenen Labels, also der *Ground Truth*, verglichen, indem mehrere Metriken berechnet werden.

3.3.4 Quantifizierung der Modellleistung

Um die Güte eines Modells in der Vorhersage von Klassen zu bewerten, ist eine Quantifizierung der Modellleistung wesentlich. Metriken für diese Quantifizierung dienen dem Vergleich mit anderen Methoden, geben Einsicht in die Stärken und Schwächen des Modells und zeigen auf, wie gut das Modell für reale Anwendungen geeignet ist (MAXWELL et al. 2021a). Um aussagekräftige Informationen zu diesen Vorteilen zu erhalten, ist es nach MAXWELL et al. (2021b) wichtig, mehrere Metriken zu berechnen und anzugeben, insbesondere bei einem vorhandenen großen Klassenungleichgewicht wie im vorliegenden Fall. In ihren Literaturstudien zur Objekterkennung bzw. -klassifizierung mittels Deep-Learning in der Fernerkundung identifizierten NEUPANE et al. (2021) und MAXWELL et al. (2021a) unabhängig voneinander die gleichen am häufigsten verwendeten Metriken, welche in der vorliegenden Arbeit verwendet werden: Gesamtgenauigkeit (*Overall Accuracy*, OA) und Kappa für alle Klassen sowie Präzision, Sensitivität, F1-Score und Intersection over Union (IoU) für jede Klasse.
Diese Metriken werden wie folgt berechnet:

$$OA = \frac{Anzahl\ korrekt\ klassifizierter\ Pixel}{Gesamtanzahl\ aller\ Pixel}$$

$$Kappa = \frac{OA - erwartete\ Übereinstimmung}{1 - erwartete\ Übereinstimmung}$$

$$Präzision = \frac{TP}{TP\ +\ FP}$$

$$Sensitivität = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2\ *\ TP}{2\ *\ TP\ +\ FP\ +FN}\ oder\ \frac{2\ *\ Präzision\ *\ Sensitivität}{Präzision\ +\ Sensitivität}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

Die Gesamtgenauigkeit (OA) ist der Anteil der korrekt klassifizierten Pixel, geteilt durch die Anzahl aller Pixel. Bei großen Klassenungleichgewichten muss die OA jedoch kritisch gesehen werden, da Klassen mit einer niedrigen Auftretenshäufigkeit die OA nur gering beeinflus-

sen (MAXWELL et al. 2021b). Kappa berücksichtigt zusätzlich den Zufallsfaktor und ist ein Maß für den Erfolg der Klassifizierung im Vergleich zu dem, was durch zufällige Übereinstimmung zwischen Klassifikation und Referenz erreicht werden könnte. Der Wert der „erwarteten Übereinstimmung“ in der oben angeführten Formel ist die hypothetische Wahrscheinlichkeit einer zufälligen Übereinstimmung. Ein Kappa-Wert von 1 gibt hierbei eine perfekte Übereinstimmung an (CONGALTON & GREEN 2019).

Für eine binäre Klassifikation, wie sie bei der Analyse von nur einer Klasse vorliegt, lassen sich in einer Konfusionsmatrix die wahr-positiven (*true positive*, TP), wahr-negativen (*true negative*, TN), falsch-positiven (*false negative*, FN) und falsch-negativen (*false negative*, FN) Zustände der klassifizierten Pixel unterscheiden. Mit diesen lassen sich die angegebenen Metriken je Klasse berechnen.

Die Präzision, im traditionellen Fernerkundungskontext auch User's Accuracy, ist hierbei der Anteil der richtig klassifizierten Pixel an der Anzahl aller Pixel, die dieser Klasse zugeordnet werden. Die Sensitivität hingegen, im traditionellen Fernerkundungskontext auch Producer's Accuracy, ist der Anteil der richtig klassifizierten Pixel an der Anzahl aller Pixel dieser Klasse.

Eine Kombination aus den beiden genannten Metriken stellt der F1-Score da. Dieser wird häufig als das harmonische Mittel über Präzision und Sensitivität bezeichnet und ist auch als Dice Koeffizient bekannt (THARWAT 2021). Die letzte verwendete Metrik stellt der IoU, auch genannt Jaccard-Index, dar. Dieser ist ein Verhältnis zwischen der Schnittmenge der Referenz- und der klassifizierten Pixel und der Vereinigung der beiden Gruppen. Wie in der angeführten Formel hierfür erkennbar, ist seine Berechnung ähnlich zu der des F1-Score, wodurch beide Metriken nicht linear korreliert sind (MAXWELL et al. 2021a).

Für einen Überblick über die Ergebnisse der einzelnen Experimente wird zumeist die IoU je Klasse sowie die OA über alle Klassen verwendet, wobei in Einzelfällen für eine bessere Übersichtlichkeit nur der IoU-Wert angegeben wird. Der Fokus liegt auf der Metrik der IoU, da die Ergebnisse je Klasse essentiell für die Beantwortung der Forschungsfragen dieser Arbeit sind. Beachtet werden muss hierbei, dass die IoU ein eher kritisches Maß ist, wodurch hohe IoU-Werte schwieriger zu erreichen sind, als beispielsweise bei dem F1 Score. Für die wichtigsten Experimente zur Beantwortung der ersten Forschungsfrage werden alle aufgeführten Metriken angegeben, was eine bessere Interpretation mit anderen Studien ermöglicht, auch wenn der F1-Score und der IoU-Wert redundante Informationen zeigen (MAXWELL et al. 2021a).

4 Ergebnisse

Im folgenden Kapitel werden die mit der in Kapitel 3 vorgestellten Methodik erarbeiteten Ergebnisse zur Beantwortung der in Kapitel 1.2 genannten Forschungsfragen dargestellt. Die Unterteilung erfolgt, wie im Methodik-Kapitel, in die Darstellung der Ergebnisse zu getesteten Daten-Parametern und zu getesteten Modell-Parametern. Eine Übersicht über die durchgeführten Experimente mit der jeweiligen Kapitelnummer ist in Abbildung 14 dargestellt. Die getesteten *Seed*-Werte betreffen sowohl das Modell wie auch die Datenvorverarbeitung. Für eine bessere Übersichtlichkeit werden die Ergebnisse beider *Seed*-Tests in demselben Kapitel dargestellt.

Daten-Parameter	Klassengruppen *K. 4.1.1*	UGs *K. 4.1.2*	Auflösungen Daten *K. 4.1.3*
	Datenkanäle *K. 4.1.4*	Anzahl Daten *K. 4.1.5*	Datenauswahl *K. 4.1.6*
Modell-Parameter	Decoder *K. 4.2.1*	Epochenanzahl *K.4.2.2*	
	Vortrainierte Mod. *K. 4.2.3*	Seed-Werte *K. 4.2.4*	

Abbildung 14 Übersicht über die getesteten Parameter mit Nummer des jeweiligen Ergebnis-Kapitels.

Für eine Übersicht, bei welchem Parameter die größte Spannweite der OA für die jeweils getesteten Einstellungen zu finden ist, lässt sich Abbildung 15 analysieren. Dargestellt ist hier je Parameter die Differenz aus der höchsten OA und der niedrigsten OA, welche mit unterschiedlichen Einstellungen dieses Parameters erreicht wurde. Bis auf den Parameter der Klassengruppen wurden für diese Abbildung nur die Tests je Parameter verwendet, die mit Klassengruppe F-2 durchgeführt werden.

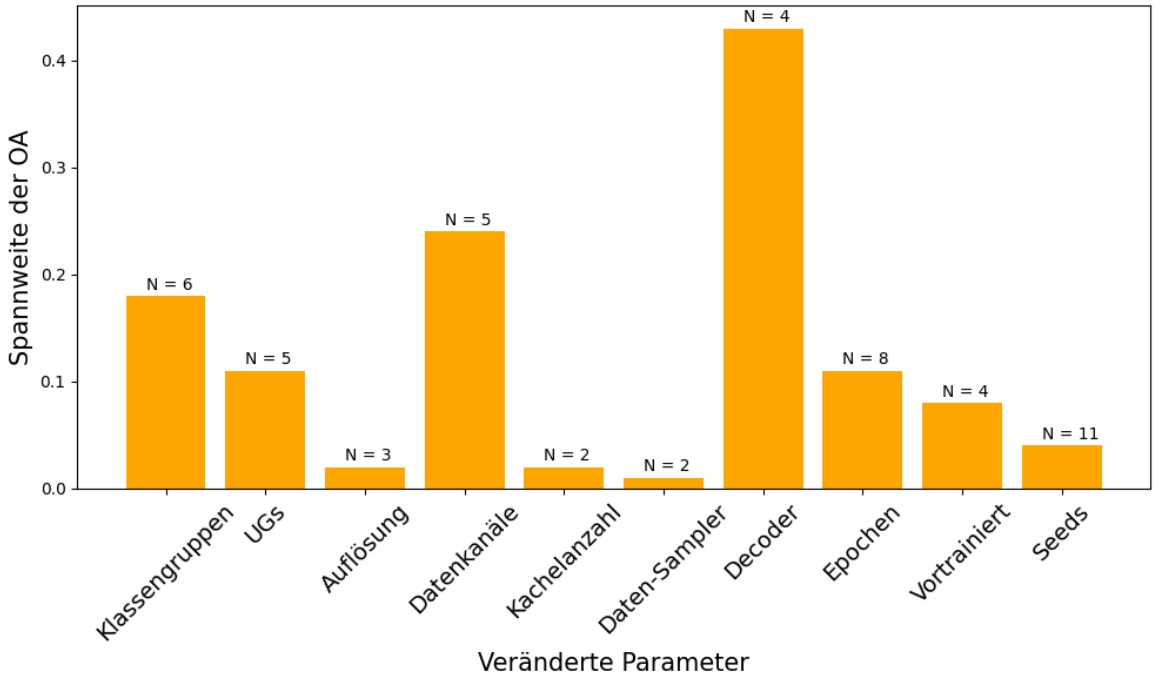

Abbildung 15 Spannweite der OA für alle getesteten Einstellungen je Parameter für Klassengruppe F-2 (ausgenommen des Parameters der Klassengruppen). N = Anzahl der Modelltrainings je Parameter.

Ersichtlich ist, dass die Spannweite der OA für den Decoder am größten ist, gefolgt von den verwendeten Datenkanälen und den verwendeten Klassengruppen. Die Auflösung, Kachelanzahl, Verwendung des Daten-Samplers und unterschiedliche Zufallszahlen (*Seeds*) weisen im Gegensatz hierzu nur relativ geringe Spannweiten auf.

4.1 Einfluss der Daten-Parameter

Im folgenden Kapitel werden die Ergebnisse der unterschiedlichen Parameter-Einstellungen für die Trainings- und Test-Daten dargestellt.

4.1.1 Semantische Segmentierung der analysierten Gebäudeklassen

Zunächst wird die Genauigkeit des Modells hinsichtlich seiner korrekten Identifizierung der unterschiedlichen Klassengruppen analysiert. In Tabelle 5 sind dazu die OAs und Kappa-Werte jeder Klassengruppe angegeben, wobei Klassengruppe F-3 die höchste OA sowie den höchsten Kappa-Wert aufweist, gefolgt von den Klassengruppen F-4 und N-1. Die Klassengruppe N-1 ist nach der OA genauso gut erkennbar wie die Klassengruppe F-4, welche dieselbe Anzahl an Klassen besitzt. Bei einem Vergleich der OAs lässt sich außerdem feststellen, dass die Klassengruppen der Nutzung bessere OAs aufweisen, als die zwei ersten Klassengruppen der Form. Die OA variiert zwischen allen Klassengruppen mit einer Spannweite von insgesamt 18 Prozent und der Kappa-Wert um 31.

Tabelle 5 Genauigkeiten (Overall Accuracy) und Kappa-Werte je Klassengruppe.

Klassegruppe	F-1	F-2	F-3	F-4	N-1	N-2
OA [%]	0,76	0,8	0,94	0,87	0,87	0,81
Kappa	0,56	0,62	0,87	0,75	0,73	0,63

Für einen Vergleich der Erkennung von unterschiedlichen Gebäudetypen sind nachfolgend in Abbildung 16 bis Abbildung 22 die in Kapitel 3.3.4 angegebenen Metriken zur Quantifizierung der Modellleistung je Klassengruppe für jeden Gebäudetyp angegeben.

In Abbildung 16 lässt sich für Klassengruppe F-1 ein maximaler IoU-Wert von 0,33 bei dem Gebäudetyp der Nebengebäude sowie 0,32 bei der Blockrandbebauung feststellen. Alle anderen IoU-Werte liegen unter einem Wert von 0,3 wobei sakrale Gebäude und Doppelhäuser nicht erkannt werden. Auffallend ist die hohe Präzision der Erkennung von Blockrandbebauung. Die Sensitivität variiert zwischen den Klassen deutlich weniger als die Präzision.

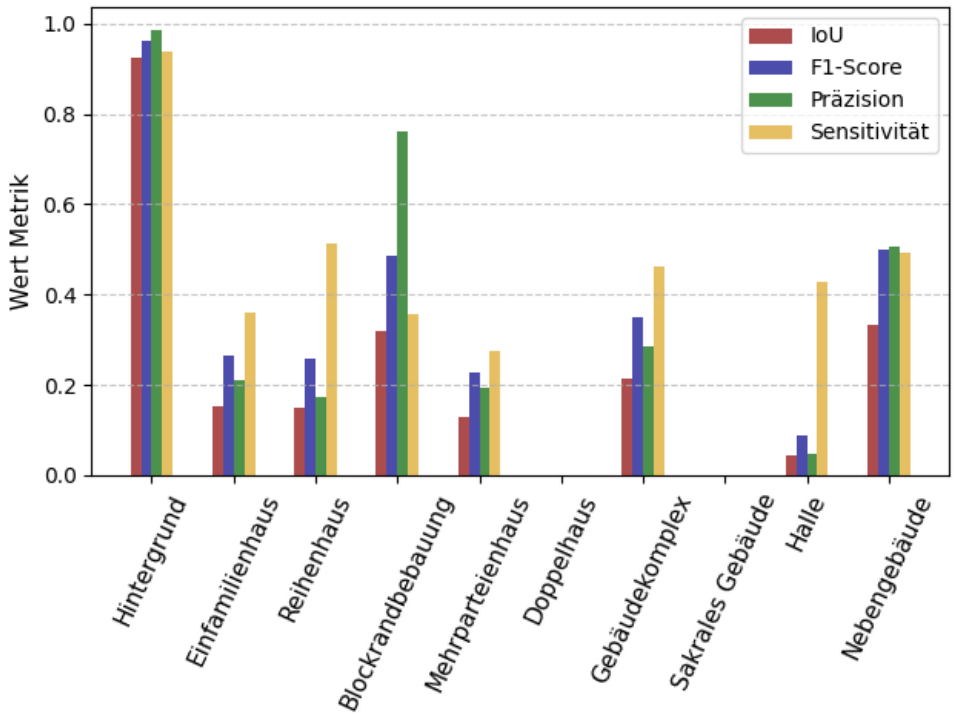

Abbildung 16 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe F-1.

In Abbildung 17 sind die Metriken für Klassengruppe F-2 ersichtlich, welche aus der Klassengruppe F-1 gebildet wird. Die besten IoU-Werte von 0,43 und 0,34 weisen auch hier die Gebäudetypen Reihenhaus inkl. Blockrandbebauung und die Nebengebäude auf. Die Präzision ist auch hier bei der Klasse der Reihenhäuser inkl. Blockrandbebauung am höchsten. Im Vergleich mit Klassengruppe F-1 besitzen alle aus mehreren Gebäudetypen zusammengesetzte Klassen höhere IoU-Werte, als die einzelnen Klassen. Sind dieselben Klassen in beiden

Klassengruppen vorhanden, so sind sich die Metriken bis auf die sakralen Gebäude sehr ähnlich. Dies gilt also für die Nebengebäude sowie die Mehrparteienhäuser.

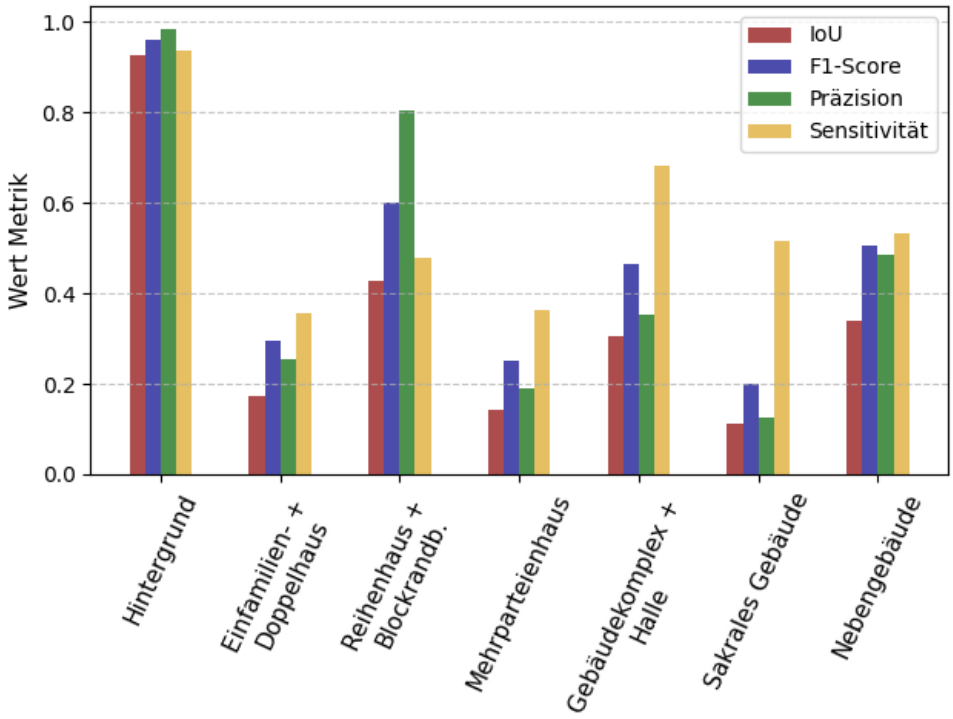

Abbildung 17 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe F-2.

Werden zusätzlich die Metriken der Klassengruppe F-3 in Abbildung 18 mit analysiert, so ist erkennbar, dass die Erkennung der Nebengebäude in allen vier Metriken um ca. 0,04 erhöht ist. Das Zusammenfügen aller anderen Klassen in die Klasse Hauptgebäude führt hier zu einem IoU-Wert von 0,88 für diese Klasse.

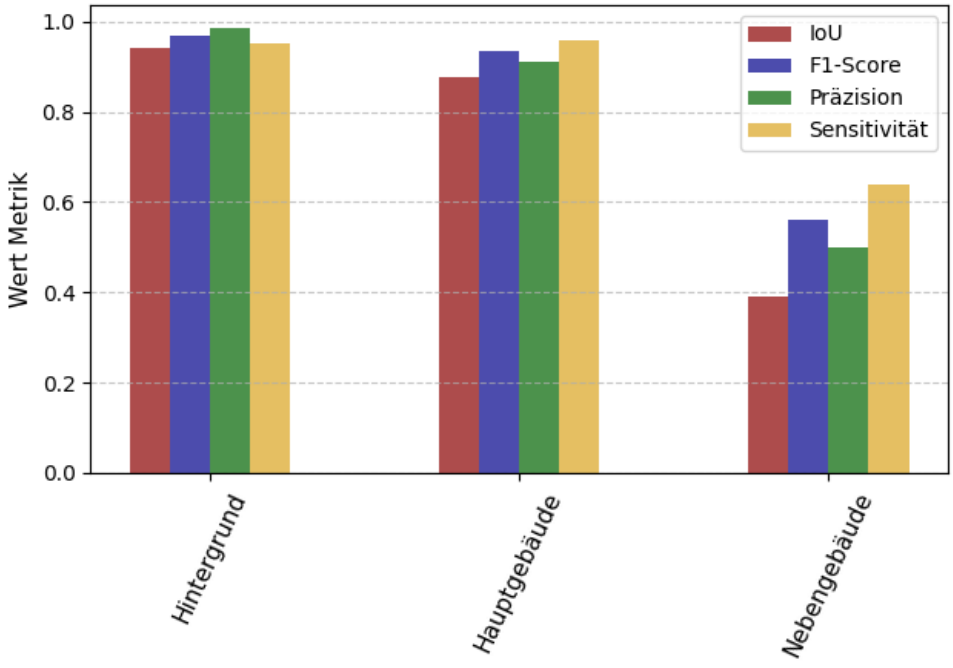

Abbildung 18 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe F-3.

In Abbildung 19 sind die Metriken je Klasse für Klassengruppe F-4 ersichtlich. Am besten erkennbar sind hier Reihenhäuser mit einer IoU von 0,66, was einen um 0,2 höheren Wert als in Klassengruppe F-2 darstellt. Für die Klasse der Nebengebäude lassen sich ähnliche Werte wie in den anderen Form-Klassengruppen feststellen. Außerdem wird die Klasse der anderen Gebäudetypen nach ihrer IoU nur minimal besser erkannt als die Klasse der Nebengebäude.

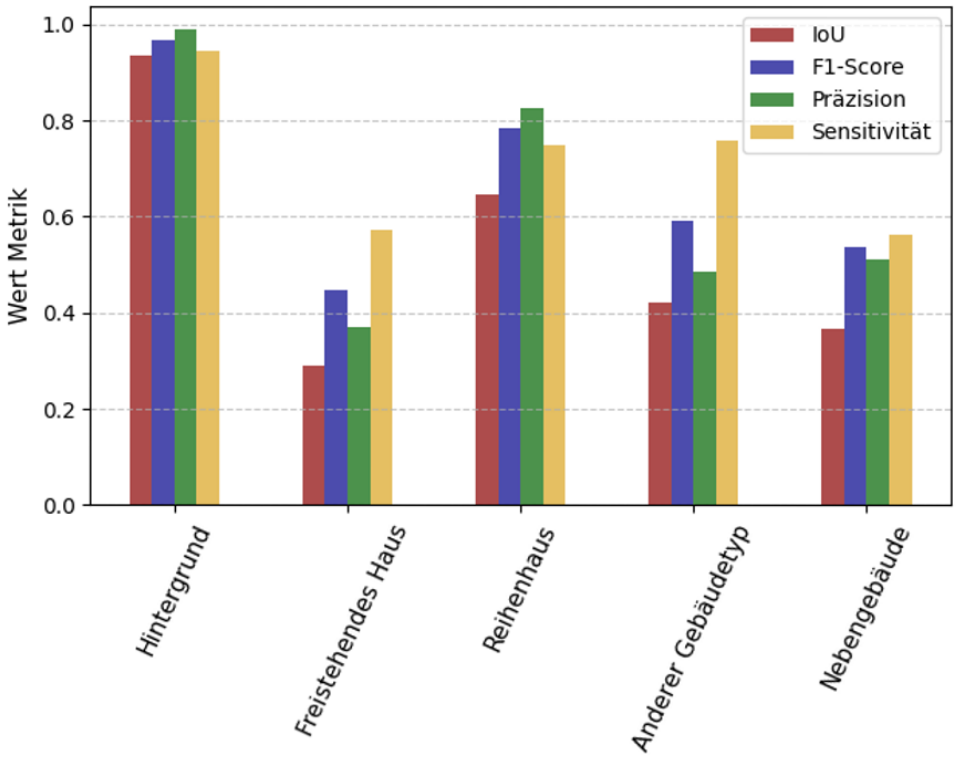

Abbildung 19 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe F-4.

Ein Ausschnitt aus der Vorhersagekarte für diese Klassengruppe ist in Abbildung 20 dargestellt. Für einen qualitativen Vergleich ist unterhalb zudem die Karte der tatsächlichen Gebäudetypen angegeben. Ersichtlich ist grundsätzlich die gute Erkennung der Gebäude selbst, was auch in den guten Metriken für die Hintergrundklasse (also nicht mit Gebäuden bebaute Flächen) deutlich wird. Werden die einzelnen Gebäudetypen betrachtet, so ist eine gute Erkennung der roten Nebengebäude feststellbar. Schuppen und Carports etc., meist entlang von freistehenden Gebäuden, sind also rein qualitativ gut durch das Modell erkennbar. Ebenso verhält es sich mit freistehenden Gebäuden, wenn diese nicht in der Nähe von anderen Gebäuden stehen, was einer klassischen Situation im ländlichen Raum bzw. einer Vorstadtsiedlung entspricht. Hier kann das Modell freistehende Gebäude gut erkennen. Im städtischen Raum, wie in dem hier dargestellten Ausschnitt aus Region 1, ist jedoch eine enge Bebauung vorhanden, was die Gebäudetypenerkennung erschwert. Hierdurch werden z. B. die Gebäude südwestlich der Kirche als Reihenhaus erkannt, obwohl diese freistehend sind.

Abbildung 20 Ausschnitt aus der Vorhersagekarte und der tatsächlichen Karte der Gebäudetypen für Klassengruppe F-4.

In Abbildung 21 sind die Metriken für die erste Klassengruppe der Gebäudenutzung, N-1, dargestellt. Erkennbar ist, dass Wohngebäude am besten erkannt werden und Wirtschaftsgebäude nur sehr schlecht. Kommunalgebäude und Gebäude der Klasse „Andere“ werden bei der Einbeziehung aller vier Metriken ca. gleich gut erfasst. Hervorzuheben ist die im Vergleich zu den restlichen Klassen hohe Präzision bei der Klasse der Wohngebäude.

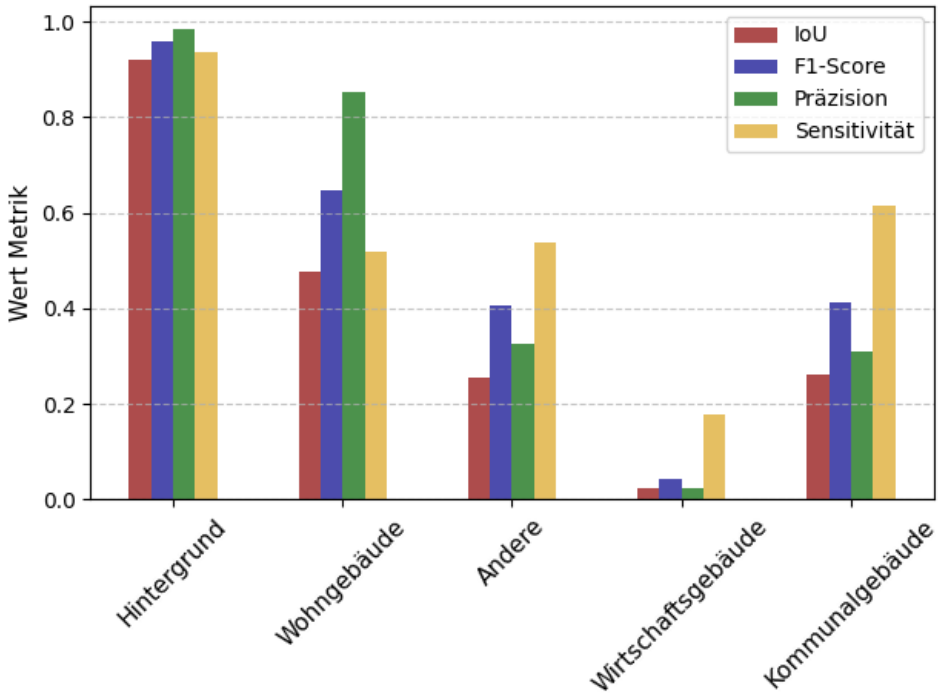

Abbildung 21 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe N-1.

Werden die Klassen der Klassengruppe N-1 zusammengefasst, so bildet dies die Klassengruppe N-2. Deren Metriken zur Analyse der Modellleistung sind in Abbildung 22 ersichtlich. Erkennbar ist, dass die Metriken der Wohngebäude in der Klassengruppe N-2 etwas über denen der Klassengruppe N-1 liegen und Nicht-Wohngebäude als Klasse besser erkannt werden, als die einzelnen Klassen in Klassengruppe N-2.

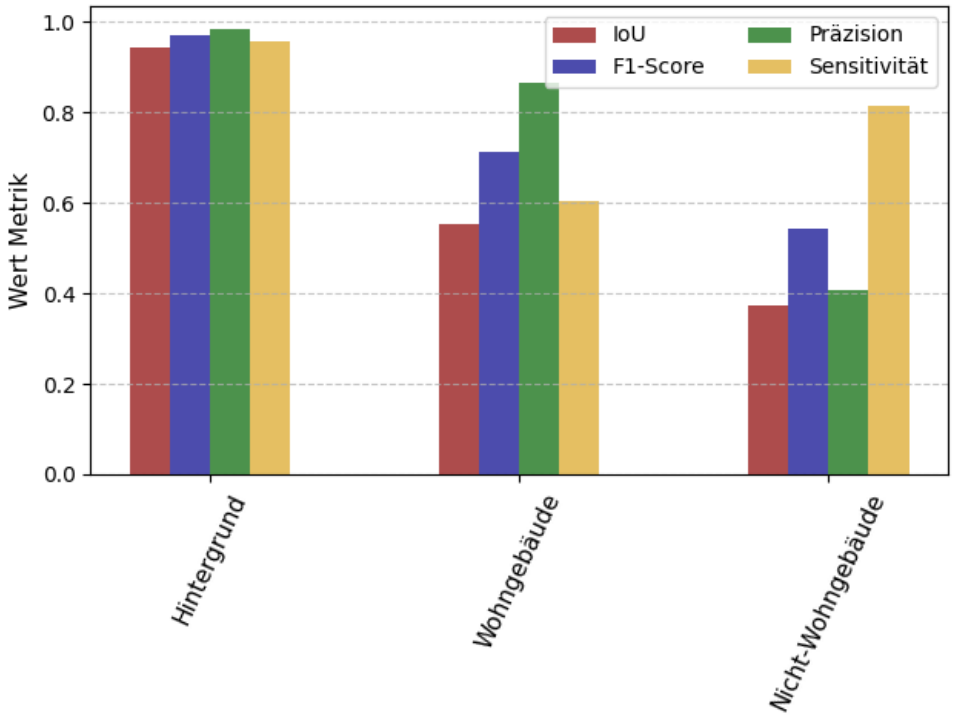

Abbildung 22 Metriken zur Evaluation der Modell-Leistung für die Klassengruppe N-2.

4.1.2 Untersuchungsgebiete für das Modelltraining und den Modelltest

Um zu erfassen, welche Rolle die Wahl des Trainings- und Testgebiets spielt, wurden alle fünf Hauptuntersuchungsregionen in verschiedenen Modelltrainings als Testregionen verwendet. Als Trainingsdaten wurden jeweils 25.000 Kacheln zufällig aus den restlichen vier Regionen ausgewählt. Das Ergebnis für alle Regionen ist in Abbildung 23 ersichtlich.

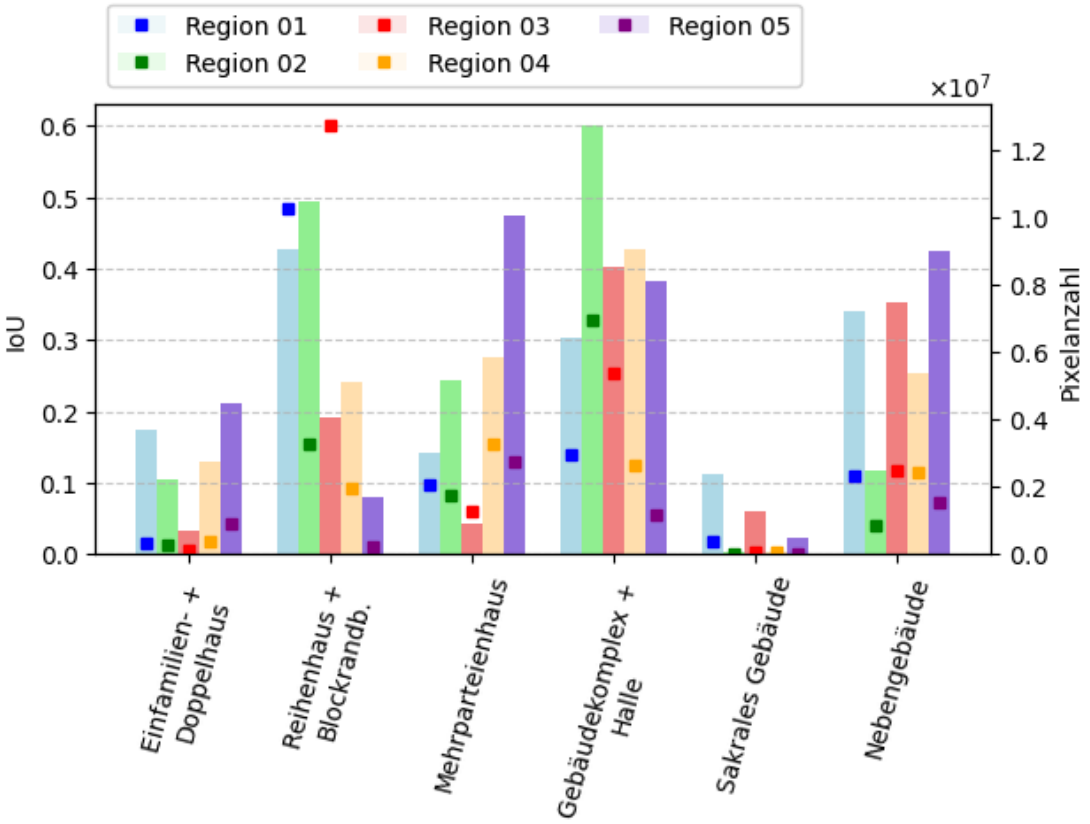

Abbildung 23 IoU der Klassen je Testregion (Balken) und jeweilige Pixelanzahl (kleines Quadrat).

Als Balkendiagramm sind hier die IoU-Werte je Klasse und Testregion dargestellt. Die Anzahl der Pixel der einzelnen Klassen je Testregion ist als farbiges Quadrat dargestellt und zeigt damit dasselbe wie bei Klassengruppe F-2 in Abbildung 6. Ursache für den Einbezug der Klassenhäufigkeit ist die mögliche Verbindung zwischen Klassenerkennung und Klassenhäufigkeit, jeweils im Trainings- und Testdatensatz. Zu sehen ist, dass die Erkennung der einzelnen Klassen in unterschiedlichen Regionen teils stark variiert, wobei für bestimmte Klassen wie die Klasse Einfamilien- und Doppelhäuser oder die Klasse der sakralen Gebäude sichtbar ist, dass diese in allen Regionen nur unzureichend erkannt werden.

Eine Kombination der IoU-Werte je Klasse mit den Pixelhäufigkeiten dieser Klasse in der Testregion und den jeweiligen Trainingsregionen mittels einer Korrelationsanalyse, führt zu den in Tabelle 6 angegebenen Werten.

Tabelle 6 Pearson Korrelationskoeffizient und p-Wert aus der Korrelationsanalyse zwischen den IoU-Werten je Klasse je Test-Region und der Anzahl der Pixel der jeweiligen Klasse in der jeweiligen Testregion sowie in den jeweiligen Trainingsregionen.

	Korrelationskoeffizient	p-Wert
Testdaten	0,69	$2,0 * 10^{-5}$
Trainingsdaten	0,80	$1,4 * 10^{-7}$

Aus dem hohen Pearson Korrelationskoeffizient kann die geschlussfolgert werden, dass zwischen den IoU-Werten der Klassen und der jeweiligen Häufigkeit dieser Pixel in der Testregion sowie in den Trainingsregionen nach COHEN (1988) ein starker Zusammenhang besteht. Der Zusammenhang zwischen der erreichten Modellleistung und der Häufigkeit der Klassen in den Trainingsregionen ist hierbei größer als der Zusammenhang mit der Klassenhäufigkeit in der Testregion. Ersichtlich ist durch den unter 0,05 liegenden p-Wert zudem, dass der Zusammenhang in beiden Fällen signifikant ist.

Um herauszufinden, wie gut die Erkennung der Nutzungs-Gebäudetypen ohne die manuelle Datenaufbereitung funktioniert, wurde Region 6 (Neuss) hinzugezogen. Hier wurden die Gebäudepolygone nicht an das Luftbild angepasst, sondern sie wurden zusammen mit den Gebäudefunktionen ausschließlich aus dem amtlichen 3D-Gebäudedatensatz übernommen. In Abbildung 24 sind die IoU-Werte der Gebäudetypen der Klassengruppe N-1 für das Training in Region 2-5 sowie das Training in Region 6 dargestellt. Erkennbar ist, dass das Training in Region 6 nicht zwingend zu einem schlechteren Klassifizierungsergebnis führt, da die Klasse der Wohngebäude um mehr als 20 % besser erkannt wird, als bei dem Training mit Kacheln aus Region 2-5. Die restlichen Klassen werden etwas schlechter erkannt. Auch in der OA wird die verbesserte Erkennung der Wohngebäude sichtbar. Die OA liegt bei einem Training mit den Daten aus Region 6 um 5 % über der OA des Trainings mit den Daten der Regionen 2-5.

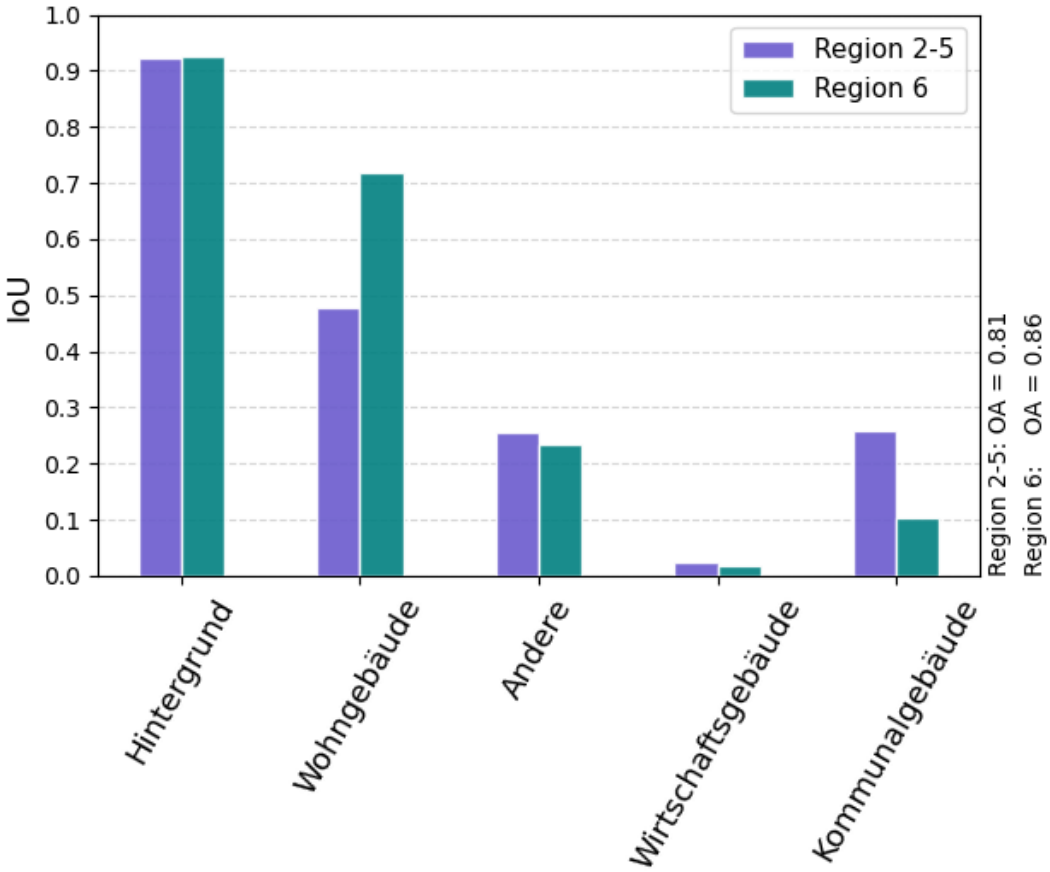

Abbildung 24 IoU-Werte je Klasse der Klassengruppe N-1 für das Modelltraining in Region 2-5 und in Region 6.

4.1.3 Vergleich der räumlichen Auflösungen

Um der Frage nachzugehen, wie sich die Gebäudetypenerkennung mit unterschiedlichen Auflösungen verändert, wurden die Auflösungen 10, 20 und 50 cm getestet. Die OA mit einer Auflösung von 10 cm und 20 cm beträgt 80 %, und mit 50 cm Auflösung wird eine OA von 79 % erreicht. Das Ergebnis der IoU je Klasse ist in Abbildung 25 dargestellt.

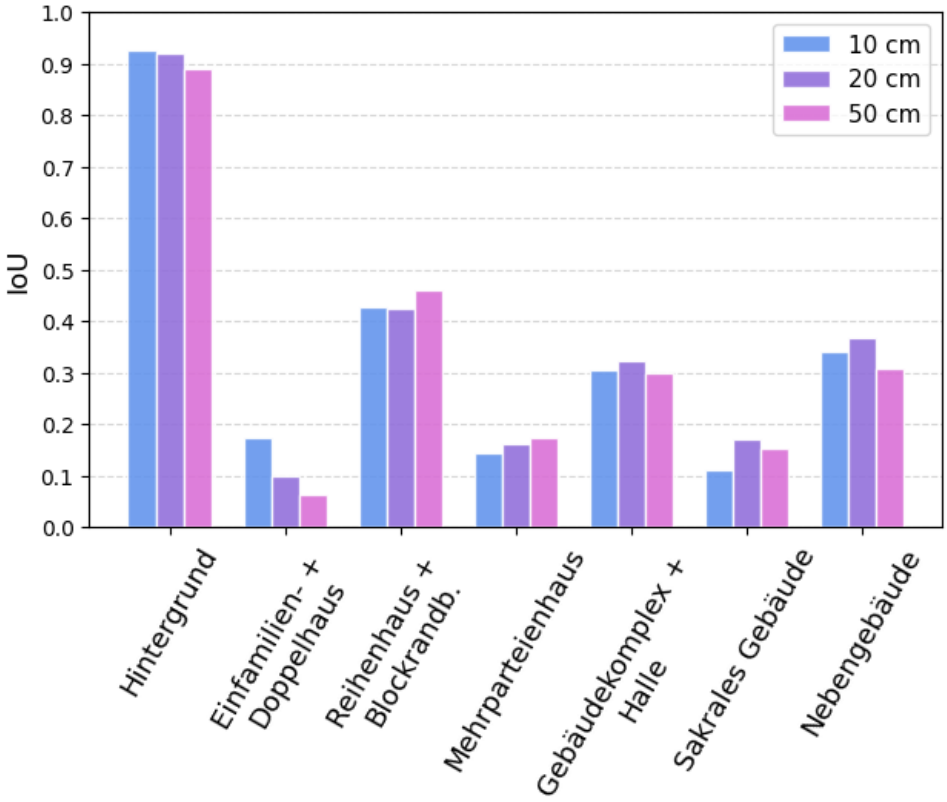

Abbildung 25 IoU-Werte je Klasse für die Auflösungen 10, 20 und 50 cm.

Grundsätzlich lässt sich hier festhalten, dass die Spannweite der IoU je Klasse eher gering ist. Für die Klasse Einfamilien- und Doppelhäuser ist erkennbar, dass eine höhere Auflösung bessere Ergebnisse erzielt. Die Klasse Reihenhäuser und Blockrandbebauung sowie die Klasse der Mehrparteienhäuser sind bei einer Auflösung von 50 cm am besten erkennbar. Für die restlichen Klassen werden mit der Auflösung von 20 cm die beste IoU-Werte erreicht. Ein eindeutiges Muster für alle Gebäudetypen ist demnach nicht feststellbar.

4.1.4 Verwendete Datenkanäle

Um den Einfluss der einzelnen Datenkanäle auf die semantische Segmentierung der Gebäudetypen abschätzen zu können, wurden fünf Modelle mit unterschiedlichen Datengrundlagen trainiert und getestet. Die OAs der Modelle sind in Tabelle 7 angegeben.

Tabelle 7 OA der Modelle, trainiert mit unterschiedlichen Datenkanälen.

Datenkanäle	RGB + NIR + nDOM	RGB + nDOM	RGB + NIR	RGB	nDOM
OA [%]	0,8	0,85	0,69	0,61	0,82

Auffallend ist besonders die niedrigere OA ohne die Verwendung eins nDOMs. Ein ähnliches Ergebnis zeigt sich auch bei der Analyse der IoUs je Klasse in Abbildung 26.

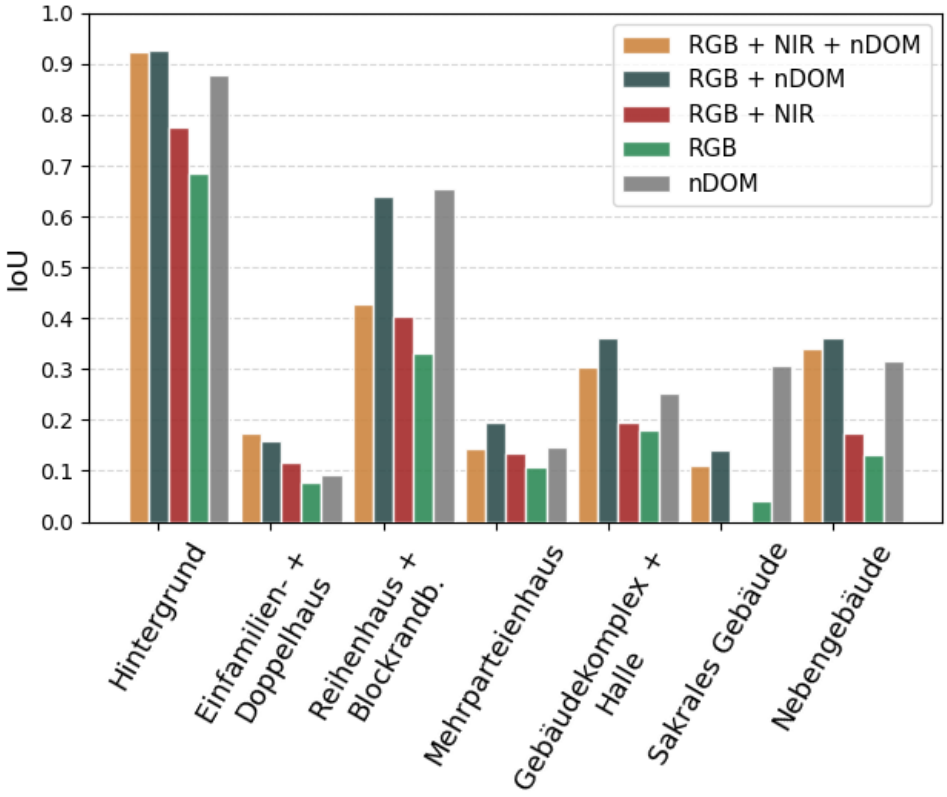

Abbildung 26 Einfluss unterschiedlicher verwendeter Datenkanäle in Form der IoU je Klasse.

Hier ist ersichtlich, dass die Hinzunahme des Höhenmodells (+ nDOM) die Erkennung jeder Klasse verbessert. Die Verwendung der Datenkanäle RGB + NIR und nur RGB liefert somit fast immer die niedrigsten IoUs. Eine Ausnahme stellt hier die Klasse Einfamilien- und Doppelhaus dar. Die Verwendung des NIR-Kanals ist ohne das nDOM von Vorteil, da hiermit wiederum in fast allen Klassen höhere IoU-Werte erreicht werden, als bei der Verwendung der RGB-Kanäle ohne NIR. Werden die Kanalkombinationen mit nDOM verglichen, so fällt auf, dass das NIR hier meist zu einer Verschlechterung der Gebäudetypenerkennung führt.

4.1.5 Anzahl der Trainingsdaten

Da für die meisten Experimente aus allen Trainingskacheln 25.000 ausgewählt wurden, wurde getestet, wie die Verwendung aller Trainingskacheln das Modellergebnis verändert. Dieses Ergebnis ist in Abbildung 27 dargestellt. Wie in Kapitel 3.3 begründet, wurde von den Standard-Einstellungen der Parameter abgewichen indem der Decoder Unet, die Verlustfunktion Kreuz-Entropie ohne Gewichte und eine Epochenanzahl von 19 verwendet wurden. Für die in Abbildung 27 ebenfalls dargestellte Referenz wurden dieselben Einstellungen gewählt wobei 50 statt 19 Epochen verwendet wurden. Ersichtlich ist, dass die Verwendung aller Trainingskacheln das Ergebnis nicht in besonderem Maße verändert. Dies zeigt sich auch in der OA, welche für beide Tests 0,86 % beträgt. Für drei der erkannten Klassen verschlechtert die Ver-

wendung aller Daten die Gebäudetyperkennung und für die Klasse der Reihenhäuser + Blockrandbebauung verbessert sich die Erkennung leicht. Die Klassen der Einfamilien- und Doppelhäuser sowie der sakralen Gebäude werden hingegen mit beiden Datensätzen nicht erkannt.

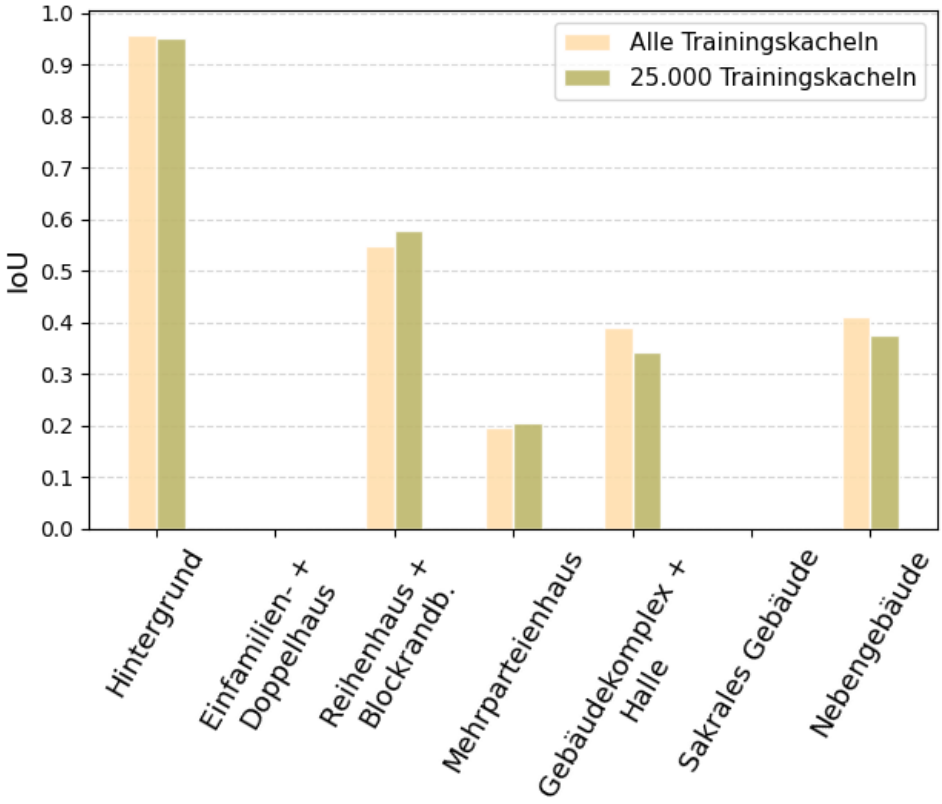

Abbildung 27 Test-Ergebnis der Verwendung aller Trainingskacheln (ca. 180.000) mit Referenz (25.000 Kacheln).

4.1.6 Auswahl unterschiedlicher Trainingskacheln

Um die Nützlichkeit des in Kapitel 3.2.2 vorgestellten Daten-Samplers zu testen, wurde ein Modelltraining mit durch den Daten-Sampler ausgewählten Daten sowie eines mit zufällig ausgewählten Daten durchgeführt. Das Ergebnis ist in Abbildung 28 dargestellt. Augenfällig ist, dass die Auswahl von bestimmten Trainingskacheln die IoU-Werte nur bei den Klassen der Mehrparteienhäuser und der Nebengebäude nicht erhöht. Sakrale Gebäude werden durch den Daten-Sampler teils erkannt, was ohne Daten-Sampler nicht der Fall ist. Die OA ist bei beiden Varianten fast dieselbe und liegt ohne Daten-Sampler um 1 % niedriger als mit Sampler.

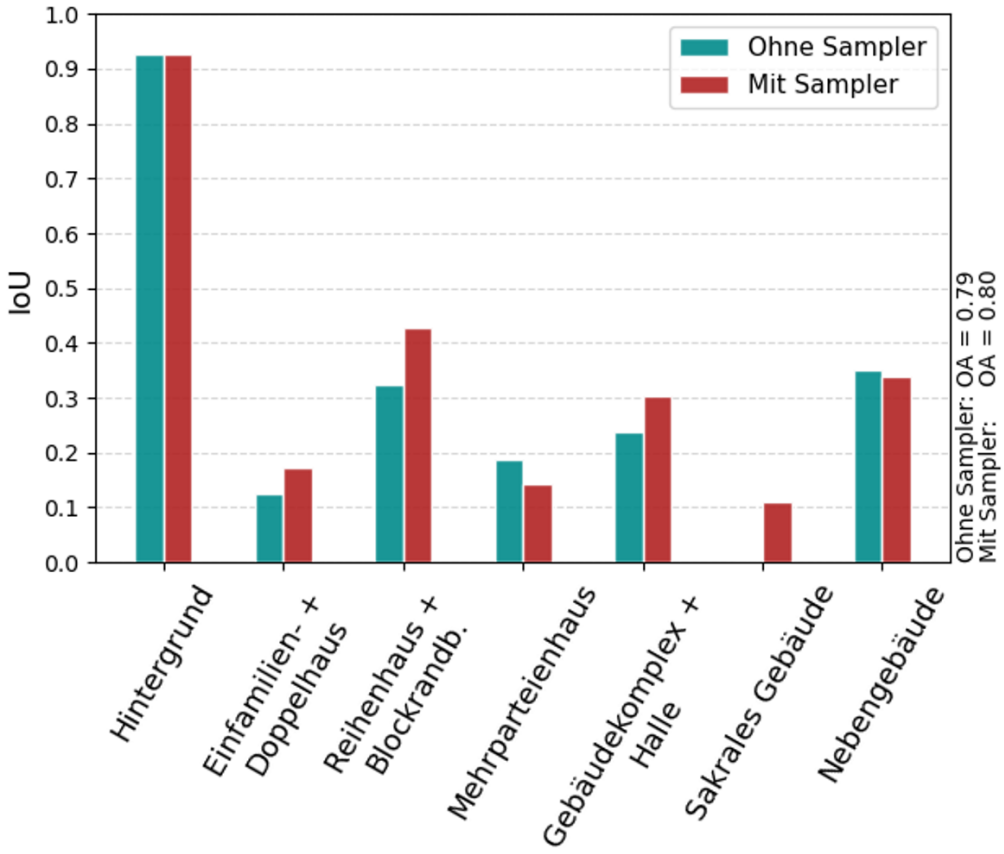

Abbildung 28 Ergebnis der Klassifizierung mit Daten-Sampler und ohne Daten-Sampler.

4.2 Einfluss der Modellparameter

Da auch die Modellparameter einen Einfluss auf die Ergebnisse des Modells haben, wurden diese gezielt verändert, um deren Auswirkung auf die semantische Segmentierung unterschiedlicher Gebäudetypen zu analysieren. Im Folgenden werden die Ergebnisse dieser Parameter-Tests vorgestellt.

4.2.1 Modellarchitekturen

Die OAs der Tests mit den Decodern Unet++, Unet, FPN und MAnet sind in Tabelle 8 aufgelistet. Den höchsten Wert zeigt hier FPN, gefolgt von Unet und Unet++. Mit MAnet lässt sich nur eine OA von 0,4 % erreichen.

Tabelle 8 OA der Tests mit unterschiedlichen Decodern.

Decoder	Unet++	Unet	FPN	MAnet
OA [%]	0,8	0,82	0,83	0,4

Die geringere Eignung von MAnet zeigt sich auch bei der Analyse der IoU-Werte je Klasse in Abbildung 29.

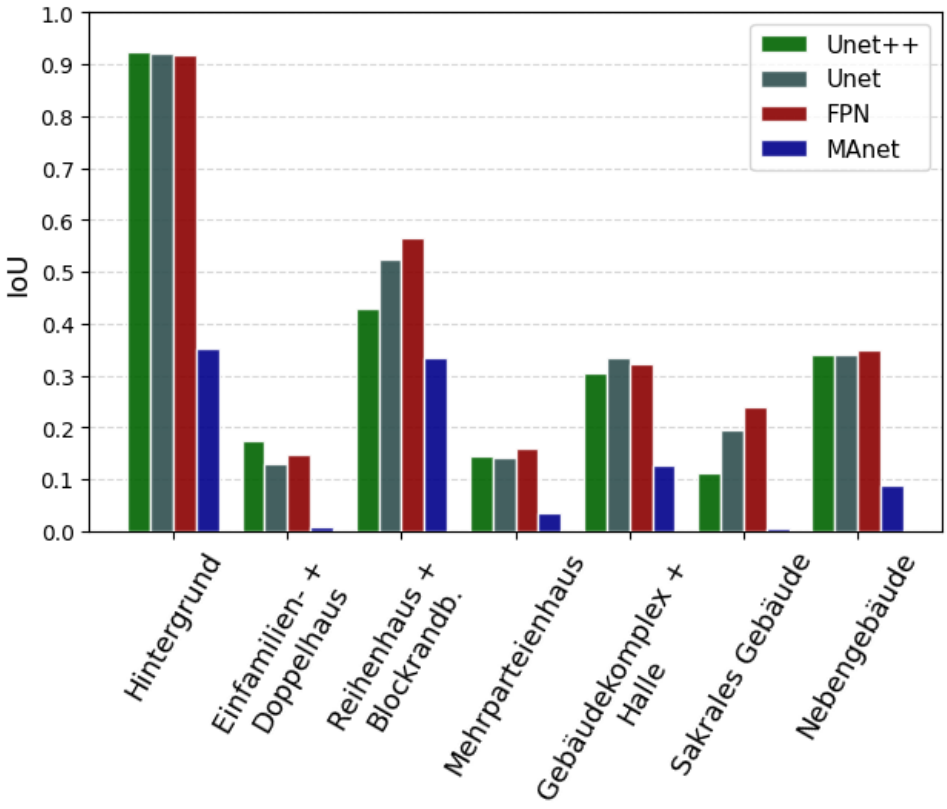

Abbildung 29 Ergebnis der Tests mit vier unterschiedlichen Decodern.

Mit MAnet sind nur maximale Werte von 0,35 zu erreichen. Deutlich bessere Werte zeigen hier die restlichen Decoder, wobei FPN in vier der insgesamt sechs Gebäudeklassen die besten Werte aufweist. Unet++ ist nur in der Klasse der Einfamilien- und Doppelhäuser am besten geeignet und die größte Differenz zwischen den drei Decodern ist in der Klasse der Reihenhäuser + Blockrandbebauung ersichtlich.

4.2.2 Anzahl der Trainings-Epochen

Um festzustellen, nach welcher Epochenanzahl das beste Ergebnis erzielt wird, wurde dasselbe Modell mit acht verschiedenen Anzahlen an Epochen trainiert. Die erreichten IoU-Werte mit den Testdaten je Klasse für jede Epochenanzahl sind in Abbildung 30 dargestellt. Da, wie in Kapitel 3.3.2 erläutert, jedoch nicht das Modell der letzten Epoche verwendet wurde, sondern das Modell, das die beste Validierungs-IoU aufweist, ist die tatsächlich verwendete Epochenanzahl in der Legende in Klammern angegeben. Hierbei ist bereits ersichtlich, dass die tatsächliche Epochenanzahl bis zur 70. Epoche zumeist fast der maximalen Epochenanzahl entspricht. Nur bei einem Training von 200 Epochen zeigt das Modell mit der 66. und damit deutlich niedrigeren Epoche die beste Validierungs-IoU und wurde dementsprechend für den Test dieses Modells verwendet. Werden die IoU-Werte je Klasse betrachtet, so zeigt sich, dass kein linearer Anstieg mit steigender Epochenanzahl vorhanden ist. Ein grundsätzlicher

Trend zu höheren Werten bei höheren Epochenzahlen ist jedoch ersichtlich und so zeigt das Modell der 66. Epoche in vier Klassen die höchsten IoU-Werte bei dem Modelltest. Diese Epoche zeigt mit 82 % zudem die beste OA.

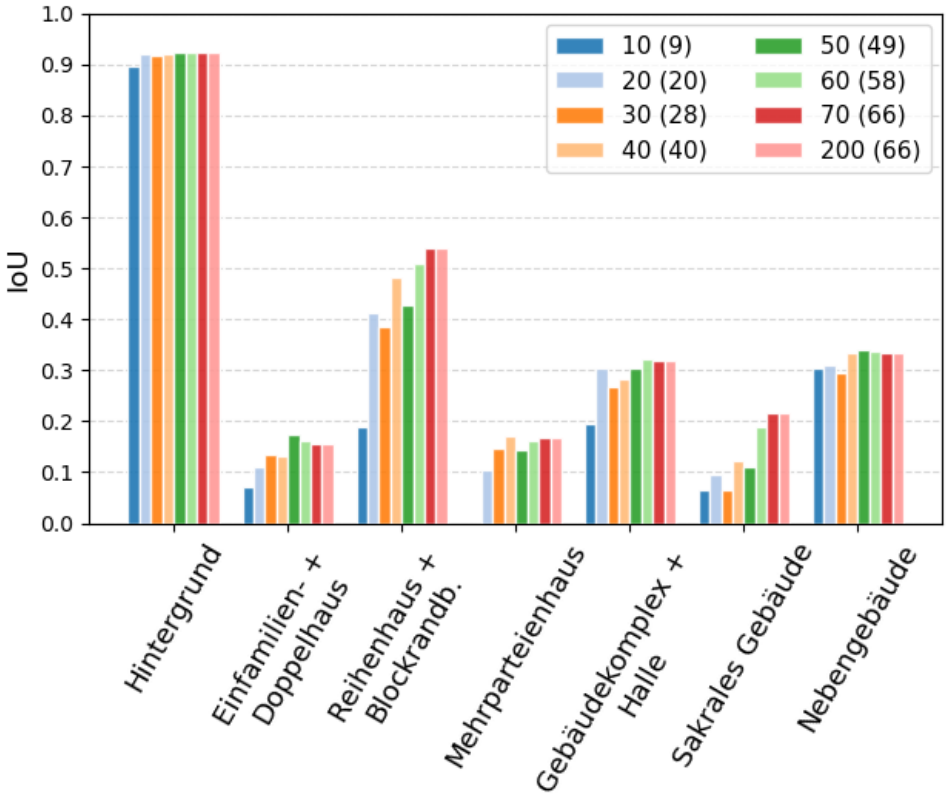

Abbildung 30 IoU-Werte des Modelltests je Klasse für Modelle, trainiert mit acht unterschiedlichen Epochen, in Klammern ist die tatsächliche Epochenanzahl angegeben.

Die beste Validierungs-IoU der 66. Epoche bei einer gesamten Epochenanzahl von 200 Epochen zeigt sich auch bei Analyse der Validierungs-IoU je Epoche in Abbildung 31.

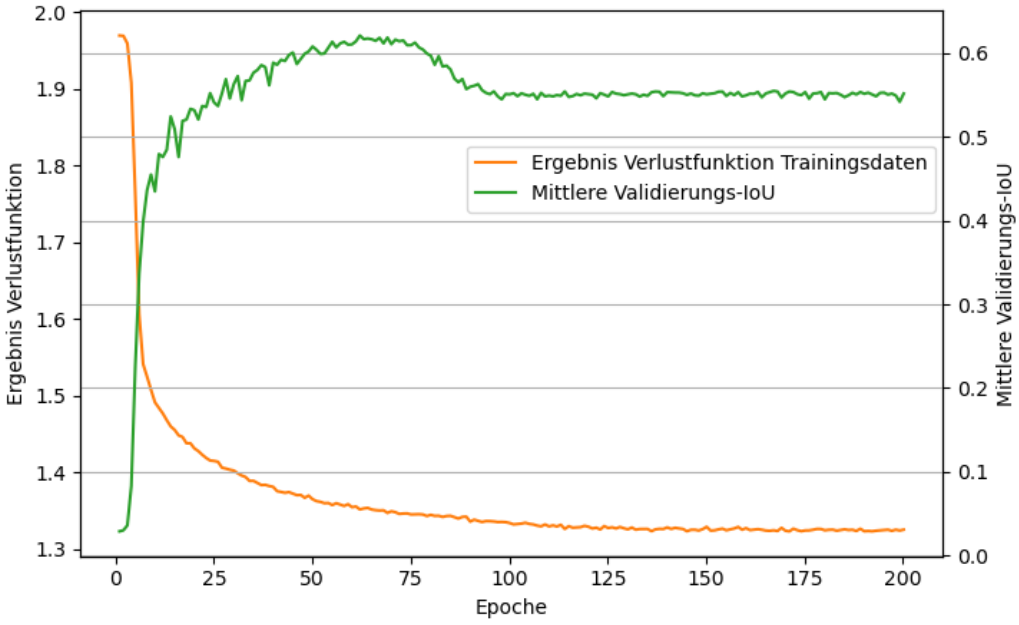

Abbildung 31 Ergebnis der Verlustfunktion und mittlere IoU der Modellvalidierung je Epoche bei einer gesamten Epochenanzahl von 200 Epochen.

Hier ist ersichtlich, dass die Validierungs-IoU (Mittelwert über alle Klassen) bis zur 66. Epoche ansteigt und anschließend wieder sinkt. Ab der 100. Epoche zeigt sich ein Plateau. Dargestellt ist in der Abbildung zudem das Ergebnis der Verlustfunktion, welche zu Beginn stark abfällt, um anschließend mit einer deutlich kleiner werdenden Steigung niedriger zu werden.

4.2.3 Ergebnisse des vortrainierten Modells

Die getesteten Varianten zur Verwendung von vortrainierten Gewichten wurden in Kapitel 3.3.2 erläutert. In Abbildung 32 sind die Ergebnisse dieser Tests sowie als Referenz das Ergebnis des nicht vortrainierten Modells für die Klassengruppe F-2 dargestellt. Bis auf die Klasse der Einfamilien- und Doppelhäuser werden alle Klassen mit vortrainierten Gewichten besser erkannt als für den Fall, bei dem die Gewichte von Anfang an mit den Trainingsdaten trainiert werden. In diesen Klassen ist die Erkennung zudem verbessert, wenn für alle Layer vortrainierte Gewichte verwendet werden und nicht nur für die RGB-Kanäle. Werden die zwei Modelle verglichen, in denen die vortrainierten Gewichte des roten und grünen Kanals für den Kanal des nahen Infrarots verwendet werden, so kann festgestellt werden, dass die Erkennung unter Verwendung der Gewichte des roten Kanals hier höhere IoU-Werte für diese Klassen liefert. In Abbildung 32 steht RGBRG für die Verwendung der Gewichte des roten Kanals für das NIR und RGBGG für die Verwendung der Gewichte des grünen Kanals für das NIR.

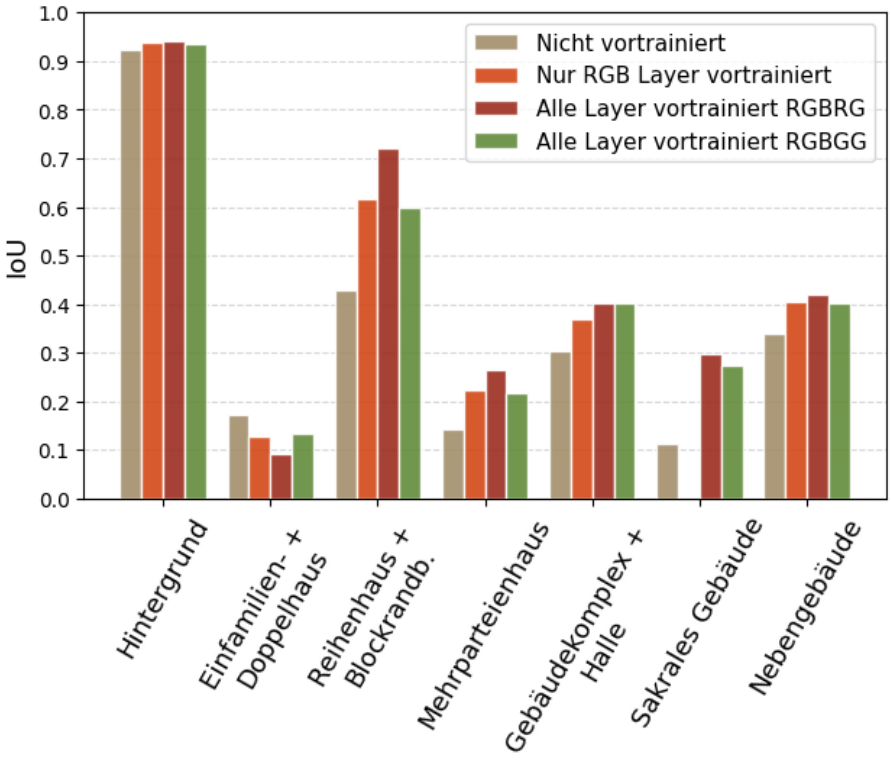

Abbildung 32 IoU-Werte je Klasse der Klassengruppe F-2 unter der Verwendung von vortrainierten Gewichten für unterschiedliche Layer.

Wie für Klassengruppe F-2 wurden vortrainierte Gewichte auch für Klassengruppe N-2 verwendet. In dem Ergebnis hierzu in Abbildung 33 ist auch für Klassengruppe N-2 ersichtlich, dass die vortrainierten Gewichte für alle Klassen die Erkennung derselben erhöhen. Besonders auffallend ist die große Verbesserung der IoU-Werte bei der Verwendung von vortrainierten Gewichten für die RGB-Kanäle. Werden auch für das NIR sowie das nDOM vortrainierte Gewichte verwendet, so steigert dies die IoU-Werte weiter. Dies geschieht jedoch nicht in dem Ausmaß, wie bei dem Vergleich der Ergebnisse der Verwendung von nicht-vortrainierten Gewichten und dem Ergebnis mit vortrainierten Gewichten für die RGB-Kanäle.

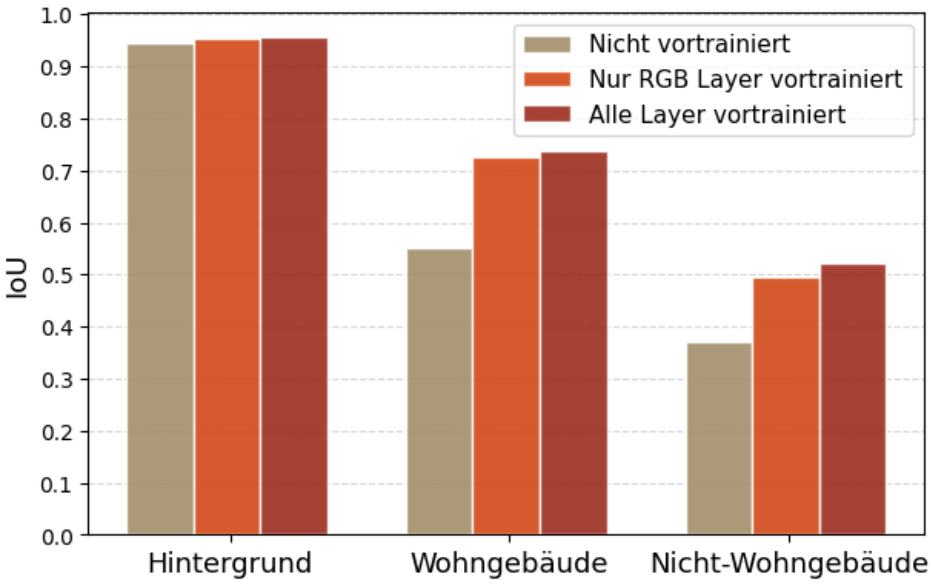

Abbildung 33 IoU-Werte je Klasse der Klassengruppe N-2 unter der Verwendung von vortrainierten Gewichten für unterschiedliche Layer.

4.2.4 Einfluss der Zufallsvariablen

Um den Einfluss des Zufalls auf das Modellergebnis zu testen, wurden unterschiedliche Zufallswerte (*Seeds*) für die Datenvorverarbeitung (5 Werte) und das Modelltraining (7 Werte) getestet. Die OA befinden sich für die Modell-*Seeds* zwischen 0,8 % und 0,83 % und für die Daten-*Seeds* zwischen 0,79 % und 0,81 %. Das Ergebnis in Form der IoU je Klasse ist in Abbildung 34 ersichtlich.

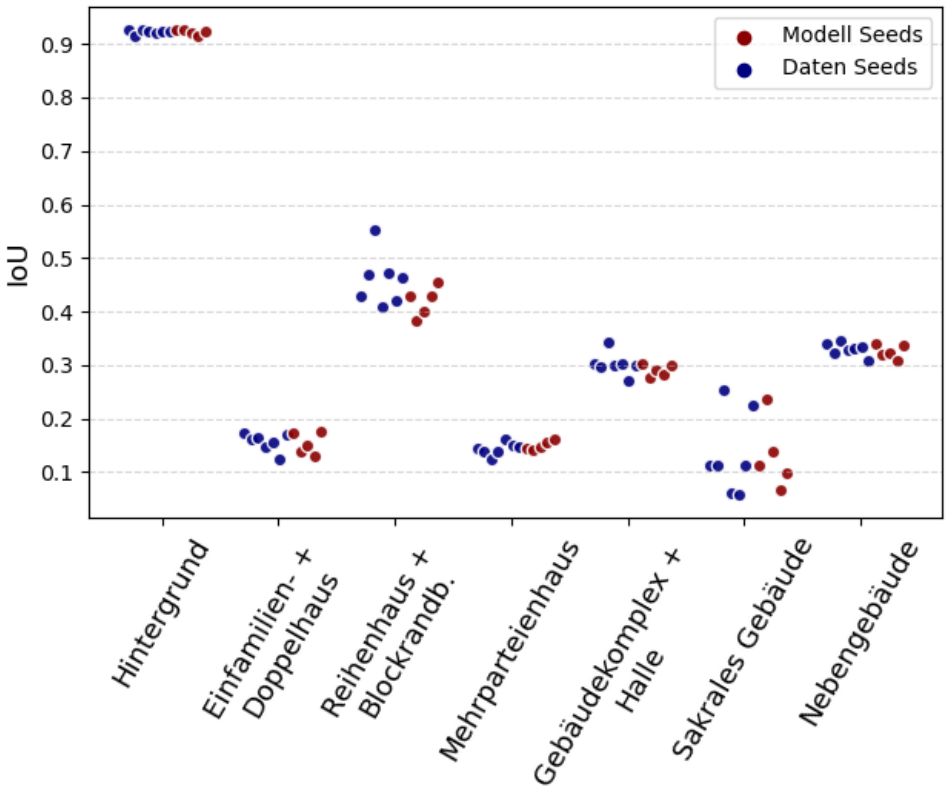

Abbildung 34 IoU-Werte je Klasse für unterschiedliche *Seed*-Werte, aufgeteilt in *Seed*-Werte im CNN-Modell und in der Datenvorverarbeitung.

Diese zeigt, dass die Streuung bei den Modell- und Daten-*Seeds* je Klasse ähnlich groß ist. Zwischen den einzelnen Klassen sind jeweils Differenzen in der Spannweite der IoU-Werte sichtbar. Am größten ist die Spannweite bei der Klasse der sakralen Gebäude, gefolgt von der Klasse der Reihenhäuser + Blockrandbebauung. In den restlichen Klassen beträgt die Spannweite der IoU maximal 10.

5 Diskussion

Im folgenden Kapitel werden die Ergebnisse aus Kapitel 4 sowie die angewendete Methodik dieser Arbeit diskutiert, indem die Ergebnisse mehrerer Tests eines Parameters miteinander verglichen werden und ein Bezug zur Literatur hergestellt wird. Der Aufbau des Kapitels ist ähnlich zu dem des Ergebniskapitels, indem zuerst in Kapitel 5.1 auf Ergebnisse der Daten-Parameter eingegangen wird. Hierbei wird außerdem die gewählte Methodik bei der Datenvorverarbeitung kritisch betrachtet. Anschließend werden in Kapitel 5.2 die Wahl der nicht veränderten Modellparameter sowie die Ergebnisse der Tests der Modell-Parameter diskutiert. Tabelle 9 bietet eine Übersicht über die zwei zentralen Forschungsfragen dieser Arbeit, sowie jeweils die wichtigsten Erkenntnisse aus den dazu durchgeführten Experimentreihen aus Kapitel 4.

Tabelle 9 Zusammenfassung der wichtigsten Ergebnisse je zentraler Forschungsfrage.

Wie gut lassen sich mittels ausgewählter CNNs unterschiedliche, nach ihrer Form und nach ihrer Nutzung definierte Gebäudetypen auf Grundlage eines Luftbildes und nDOMs erkennen? • Nach ihrer Nutzung definierte Gebäudetypen sind bei ausgewählten Klassen nicht zwingend schwieriger zu erkennen als nach ihrer Form definierte Gebäudetypen • Form: Reihenhäuser, Hauptgebäude und Nebengebäude sind am besten erkennbar • Nutzung: Wohngebäude und Nicht-Wohngebäude sind am besten differenzierbar • Das Zusammenfassen von (passenden) Klassen verbessert die Erkennung zumeist • Dieselbe Klasse in unterschiedlichen Klassengruppen erreicht meist ähnliche Genauigkeiten
Welche Auswirkungen zeigen ausgewählte Modell- und Daten-Parameter auf die Gebäudetypenerkennung? • Die Ergebnisse je Klasse korrelieren mit der Häufigkeit der Klasse in den Trainingsdaten stärker als mit der Häufigkeit in den Testdaten • Für die Nutzungsklassen kann der amtl. Gebäudedatensatz als Trainingsdatensatz verwendet werden • Verwendung eines nDOMs verbessert Gebäudetypenerkennung stark, NIR nur vorteilig ohne nDOM • FPN ist am besten als Decoder geeignet • Die optimale Epochenanzahl liegt bei ~ 60 Epochen • Fast alle Klassen werden mit vortrainierten Gewichten besser erkannt

Um den Einfluss aller getesteten Parameter vergleichend zu betrachten wurde im Ergebniskapitel die Spannweite der OA aller Modelltrainings je Parameter der Klassengruppe F-2 sowie aller Klassengruppen für den Parameter der Klassengruppen angegeben (Abbildung 15). Die hieraus gefolgerte Aussage, dass die Wahl des Decoders den größten Einfluss auf die Modellleistung besitzt, ist nur bedingt generalisierbar, da dies explizit nur für die in der vorliegenden Arbeit getesteten Einstellungen gilt. Um den Einfluss zu quantifizieren, wäre zudem eine Darstellung der Verteilung, wie beispielsweise mittels Boxplots, passend. Durch die teils niedrige Anzahl von teilweise nur zwei Tests je Parameter, ist dies jedoch keine passende Art der Visualisierung. Beachtet werden muss zudem, dass nur die Veränderung der Modellleistung, nicht aber die Modellverbesserung, sichtbar ist. Zuletzt ist zu bedenken, dass nur die Spannweite der OA dargestellt wird, welche besonders von der häufigen Hintergrundklasse abhängt. Trotz dieser Einschränkungen stellt das Ergebnis in Abbildung 15 einen wichtigen Teil dieser Arbeit dar, da hieraus auch Schlussfolgerungen gezogen werden können, bei welchen Parametern für zukünftige Tests am ehesten eine große Veränderung der Modellleistung zu erwarten ist.

5.1 Einfluss der Daten-Parameter

In diesem Kapitel wird die Methodik der Datenvorverarbeitung kritisch betrachtet und die Ergebnisse der Parameter-Tests werden diskutiert.

5.1.1 Kritische Betrachtung der Methodik der Datenvorverarbeitung

Im Folgenden werden vier Aspekte der Datenvorverarbeitung kritisch betrachtet.

Im Bereich der Datenvorprozessierung in Kapitel 3.2.1 lässt sich festhalten, dass ein Überschneidungsanteil der einzelnen Kacheln von 80 % einen im Vergleich zur Literatur hohen Wert darstellt, da in den meisten Studien der semantischen Segmentierung mittels Deep-Learning im Fernerkundungsbereich ein Wert von 50% verwendet wird (NEUPANE et al. 2021). Nötig ist die Überlappung in der vorliegenden Arbeit, um den Trainingsdatensatz für das Modell zu erhöhen. Die hier verwendete Gesamtanzahl von 25.000 Kacheln kann jedoch auch mit einem geringeren Überschneidungswert bereits erreicht werden. Für den Testdatensatz ist die hohe Anzahl an sich überschneidenden Pixeln jedoch wünschenswert, da so die Genauigkeit erhöht werden kann, wenn derselbe Pixel in unterschiedlichen Kontexten betrachtet wird (AN et al. 2020).

Zur verwendeten Datenvorprozessierung kann außerdem angemerkt werden, dass die Höhenwerte des nDOMs nicht nach ihrem Minimum und Maximum normalisiert wurden, um das

Modell mit den ursprünglichen Differenzen der Höhenwerte zu trainieren. Werden alle Daten normalisiert, so kann dies nach HAN et al. (2023) die Lernphase von neuronalen Netzen beschleunigen. Wichtig ist jedoch vor allem die Anwendung einer Normalisierungstechnik und da die für Methoden des maschinellen Lernens meist am besten geeignete Z-Wert Normalisierung (SINGH & SINGH 2020) auf alle Datenkanäle angewendet wurde, sollte die fehlende Min.-Max. Normalisierung keine zu großen Auswirkungen haben.

Zur Datenvorbereitung in Kapitel 3.2.2 können zwei Aspekte kritisch hinterfragt werden. Der Erste betrifft die Aufteilung der Daten in Trainings- und Validierungsdaten. Durch den vorhandenen Überschneidungsbereich ist der Großteil der Pixel bzw. Gebäude in mehreren Kacheln vorhanden. Da die Auswahl der Trainings- und Validierungsdaten aus allen geteilten Kacheln zufällig erfolgte, kann dasselbe Pixel bzw. Gebäude unter Umständen in beiden Datensätzen vorhanden sein, obwohl es sich nicht um dieselbe Kachel handelt. Da es sich hier jedoch nur um die Modellvalidierung und nicht den Modelltest handelt, ist dies nicht als methodisch inkorrekt zu betrachten. Wichtig ist während des Tests besonders, dass das Modell nie mit den Gebäuden des Modelltests trainiert wurde (MAXWELL et al. 2021b). Durch die Auswahl eines kompletten, räumlich abgegrenzten Untersuchungsgebiets, ist dies im Rahmen dieser Arbeit vollumfänglich gewährleistet.

Der zweite Aspekt der Datenvorbereitung ist die Anwendung von Augmentationen. Typischerweise werden diese angewandt, um den Datensatz für das Modelltraining zu vergrößern. Für kleine Datensätze kann so die Modellleistung um bis zu 20 % gesteigert werden (MINAEE et al. 2021). Im vorliegenden Fall sind jedoch durch den großen Überschneidungsbereich der Kacheln ausreichend Kacheln vorhanden und es besteht für eine Auswahl von 25.000 Kacheln kein Grund zur Datenerweiterung. Da die Daten teilweise dasselbe Gebäude zeigen, kann die Anwendung von Augmentationen hierauf wiederum einen Vorteil darstellen. Um wie bei SHORTEN & KHOSHGOFTAAR (2019) mit dem Vorgehen in vielen Studien übereinzustimmen, sollten diese in diesem Fall jedoch auf alle Kacheln angewandt werden, welche dasselbe Gebäude zeigen und nicht wie in der vorliegenden Arbeit nur per Zufall mit einer Wahrscheinlichkeit von 87,5 % für alle Kacheln. Der Nutzen der Augmentationen könnte in einem weiteren Experiment analysiert werden.

5.1.2 Semantische Segmentierung der analysierten Gebäudeklassen

In der vorliegenden Arbeit wird mit der Verwendung von mehreren Gebäudetypen der Nutzung und der Form eine Wissenslücke in aktuellen Studien zu diesem Thema geschlossen. Hierzu zählt das grundsätzliche Ergebnis, dass nach ihrer Nutzung definierte Gebäudetypen

bei ausgewählten Klassen nicht zwingend schwieriger zu erkennen sind, als nach ihrer Form definierte Gebäudetypen. Dieses lässt sich auch kritisch betrachten. Wichtig ist hierbei, dass dieses Ergebnis nur für ausgewählte Klassen gilt, nicht verallgemeinert werden kann und es teils auf die Metrik zur Quantifizierung der Modellleistung ankommt. So wird die Klassengruppe N-1 nach ihrer OA genauso gut erkannt wie die Klassengruppen F-4. Beide Klassengruppen beinhalten vier Gebäudetypen. Bei einem Vergleich der IoUs je Klasse ist jedoch ersichtlich, dass die Klassen der Form im Mittel deutlich bessere Werte aufweisen als die Klassen der Nutzung. Dies ist ein Hinweis darauf, dass für einen Vergleich kompletter Klassengruppen die Metrik der mittleren IoU über alle Klassen ebenfalls passend wäre. Die grundsätzliche Variation der OA um 18 % zwischen den Klassengruppen folgt jedoch nicht nur aus der unterschiedlich guten Erkennung der Hintergrundklasse, da diese in allen Klassengruppen ähnlich gut erkannt wird. Dies ist ein Hinweis auf die gute Differenzierung zwischen Gebäude und Nicht-Gebäude. Die einzelnen Klassen beeinflussen also auch die OA, so dass diese als Vergleichsmaß genutzt werden kann. Bei einem Vergleich der Klassengruppen F-3 (Haupt- und Nebengebäude) und N-2 (Wohn- und Nicht-Wohngebäude) fällt auf, dass Nebengebäude und Nicht-Wohngebäude nach ihrer IoU ähnlich gut erfasst werden. Hauptgebäude hingegen werden deutlich besser erkannt als Wohngebäude. Hier werden also die Formklassen insgesamt besser erkannt. Bei einem Vergleich von N-2 und F-1 ist wiederum feststellbar, dass Wohn- und Nicht-Wohngebäude besser erkannt werden als jede Klasse der Form.

Das zweite, die Klassengruppen der Form und Nutzung betreffende Ergebnis, ist, dass das Zusammenfassen von (passenden) Klassen deren Erkennung meist verbessert. Dies ist für alle Klassen der Klassengruppe F-2 der Fall. Nachvollziehbar ist dies, da die Klassen der Klassengruppe F-2 auf Basis der Konfusionsmatrix von F-1 gebildet wurden. So wurden Klassen zusammengelegt, die das Modell typischerweise miteinander verwechselt. Zudem weisen diese zusammengefassten Klassen meist ein ähnliches Erscheinungsbild auf. Auch für die Nutzungs-Klassengruppe verbessert das Zusammenfassen der Klassen von N-1 in die Unterscheidung zwischen Wohn- und Nicht-Wohngebäuden die Metriken der Klassengruppe.

Die Klasse der Wohngebäude ist hierbei in der Klassengruppe N-2 etwas besser erkennbar, was grundsätzlich mit der Annahme übereinstimmt, dass dieselbe Klasse in mehreren Klassengruppen ähnlich gut erkennbar ist. Feststellbar ist dies auch für gleiche Klassen in den Form-Klassengruppen, wie z. B. den Nebengebäuden. Dass sich das Ergebnis leicht verbessert, könnte mit dem geringeren Klassenungleichgewicht sowie der verringerten Modellkomplexität zusammenhängen.

Klassengruppen der Form

Wie für alle Klassengruppen gilt auch für die Formklassengruppen, dass die OA jeweils stark mit den Metriken zur Erkennung der einzelnen Klassen in Abbildung 16 bis Abbildung 22 zusammenhängt. So erreichen die zwei Klassen für Klassengruppe F-3 im Vergleich mit den restlichen Form-Klassengruppen die höchsten IoUs und die höchste OA. Am zweithöchsten sind die IoU-Werte der Klassengruppe F-4 und ebenso die OA dieser Klassengruppe. Das Ergebnis, dass Reihenhäuser sowie Hauptgebäude und Nebengebäude am besten erkennbar sind, ist insofern nachvollziehbar, da diese Klassen besonders durch ihre benötigte Fläche gekennzeichnet sind. Diese ist bei Reihenhäusern zumeist deutlich größer als bei anderen Gebäudetypen. Nebengebäude wiederum besitzen eine besonders kleine Fläche und zumeist auch Höhe. Die gute Erkennung von Hauptgebäuden in Klassengruppe F-3 wiederum könnte auch hiermit zusammenhängen, da Gebäude grundsätzlich gut erkannt werden und diese nur von den Nebengebäuden unterschieden werden müssen.

Bei dem Vergleich der qualitativen Darstellung des Ergebnisses der Klassengruppe F-4 in Abbildung 20 und der quantitativen Beurteilung in Abbildung 19 fällt auf, dass eine leichte Differenz besteht, welche Klasse am besten erkannt wird. Quantitativ werden Reihenhäuser am besten erkannt und qualitativ wirken freistehende Gebäude als am besten erkennbar. Diese Differenz könnte auf das Klassenungleichgewicht zurückzuführen sein, durch welches deutlich mehr Reihenhäuser als freistehende Gebäude in der Testregion vorhanden sind. Es wird jedoch deutlich, wie wichtig auch die rein optische, qualitative Beurteilung des Ergebnisses für die Interpretation desselben ist.

Bei einem Vergleich der Klassengruppen F-1 und F-2 fällt auf, dass besonders die zusammengesetzte Klasse aus Reihenhäusern und Blockrandbebauung bessere Metriken als die einzelnen Klassen erreicht. Dies ist zum einen auf den Fakt zurück zu führen, dass die meisten Blockrandbebauungen Reihenhäuser beinhalten, was die Erkennung der Reihenhäuser in Gruppe F-2 vereinfacht. Wird also in Klassengruppe F-2 nur noch versucht, Reihenhäuser, unabhängig davon, ob als Blockrandbebauung oder nicht, zu detektieren, so gelingt dies besser. Der zweite Punkt, wieso Blockrandbebauung schlechter erkannt wird ist, dass diese Gebäudeform am besten erkennbar ist, wenn sie komplett sichtbar ist. Auf einer Kachel von 320*320 Pixeln ist dies jedoch nicht möglich. Erst ab einer Auflösung von 30 cm könnte ein typischer Gebäudeblock von 100*100 m komplett in eine Kachel passen und selbst dann ist nicht sicher gewährleistet, dass dies der Fall ist. Bei einer Auflösung von 50 cm ist die Wahrscheinlichkeit erhöht, dass der gesamte Gebäudeblock sichtbar ist und so wird Blockrandbebauung in dieser Arbeit bei einer Auflösung von 50 cm besser erkannt, als bei 10 oder 20 cm.

Für Reihenhäuser hingegen ist es nicht so wichtig, dass sich das komplette Reihenhaus in einer Kachel befindet, da bereits bei einem kleineren Ausschnitt meist erkennbar ist, ob es sich um mehrere Gebäude bzw. ein „langes“ Reihenhaus handelt oder nicht.

Klassengruppen der Nutzung

Das Ergebnis, dass bei den Klassengruppen der Nutzung Wohn- und Nicht-Wohngebäude am besten erfassbar sind, ist insbesondere für eine Risikoanalyse im Naturgefahrenbereich von großem Interesse. Das trainierte Modell könnte hierfür beispielsweise auf eine potentiell durch Hochwasser betroffene Fläche angewandt werden und mit einer hinreichenden Genauigkeit könnte abgeschätzt werden, wie viele Wohngebäude mit potentiellen Bewohnern von dem Hochwasser betroffen sind. Die Genauigkeit der Erkennung reicht jedoch vermutlich nicht aus, um gebäudespezifische Aussagen zu treffen, auf deren Basis sicherheitsrelevante Maßnahmen, wie beispielsweise zum Schutz einzelner Gebäude, getroffen werden.

Bei der Analyse der Ergebnisse der Nutzungs-Klassengruppe N-1 fällt besonders auf, dass Wirtschaftsgebäude fast nicht erkennbar sind. Zurückzuführen ist dies vermutlich darauf, dass in der städtischen Testregion Wirtschaftsgebäude nicht als gut erkennbare Industriegebäude oder Hallen vorhanden sind, sondern in ihrem äußeren Erscheinungsbild keine erkennbaren Unterscheidungsmerkmale zu den restlichen Gebäudetypen aufweisen. Die hier vorhandenen Wirtschaftsgebäude sind also beispielsweise Bürogebäude oder kleinere Supermärkte in ehemaligen Wohngebäuden und so auch nur schwer allein aufgrund ihrer Morphologie bzw. äußeren Erscheinung unterscheidbar. Insbesondere im innerstädtischen Bereich kann die Nutzung höchst komplex und divers sein und so sind Wirtschaftsgebäude nur schwer von Wohngebäuden zu differenzieren. Hinzu kommt, dass in der Test-Region nur sehr wenige Wirtschaftsgebäude zur Modellvalidierung vorhanden sind (vgl. Abbildung 6).

Die Ergebnisse der Nutzungs-Klassen lassen sich besser als die Form-Klassen in vorhandener Literatur zur semantischen Segmentierung von unterschiedlichen Gebäudetypen mittels Deep-Learning Verfahren einordnen, da diese Klassenart häufiger verwendet wird (vgl. Tabelle 1). Beachtet werden muss jedoch, dass in diesen Studien kein Höhenmodell verwendet wird und die Ergebnisse stark von der verwendeten Datengrundlage abhängen.

HOFFMANN et al. (2021) erreichen beispielsweise mit einer U-Net ähnlichen Architektur zusammen mit einem Luftbild mit ca. 0,6 m Auflösung für die Klasse Wohngebäude für unterschiedliche Untersuchungsgebiete IoU-Werte von 0,45 – 0,71 und für Nicht-Wohngebäude IoU-Werte von 0,34 – 0,54.

Die in der vorliegenden Arbeit ermittelten IoU-Werte in Abbildung 22 mit einer OA von 0,81 % befinden sich in diesem Wertebereich. Beachtet werden muss hierbei, dass dies nicht

die höchsten für diese Klassengruppen erreichbaren Werte darstellt. Werden beispielsweise die Ergebnisse für Klassengruppe N-2 verwendet, welche unter der Verwendung von vortrainierten Gewichten erreicht wurden, so sind höhere IoU-Werte feststellbar. Für Wohngebäude sind dies 0,74 und für Nicht-Wohngebäude 0, 52 (vgl. Abbildung 33). Die OA beträgt hier 0,92. Diese Werte sind bereits leicht besser sowie im oberen Bereich der IoU-Werte von HOFFMANN et al. (2021). Wie für Klassengruppe F-2 gezeigt, könnten auch für Klassengruppe N-2 zudem mit unterschiedlichen Untersuchungsgebieten bessere IoU-Werte erreicht werden.
Auch DIMASSI et al. (2021) klassifizieren diese zwei Gebäudetypen nur mittels RGB-Bildern, wobei sie einen zweistufigen Ansatz wählen (erst Segmentierung, dann Klassifizierung der Gebäude). Sie erreichten OA-Werte von 0,93-0,95 %. Ursache für diesen hohen Wert ohne die Verwendung eines Höhenmodells könnte sein, dass in dem verwendeten Trainingsdatensatz Nicht-Wohngebäude nur gut erkennbare Gebäudetypen enthalten wie Moscheen, Einkaufzentren oder Industrieanlagen. Besonders im innerstädtischen Bereich schwer zu erkennende Gebäude wie Supermärkte oder Bürogebäude sind in der Definition der Nicht-Wohngebäude bei DIMASSI et al. (2021) jedoch nicht enthalten. Im Rahmen dieser Arbeit wurde dies nicht berücksichtigt, sondern es wurde ein realitätsnahes Szenario mit herausfordernden Testgebieten gewählt, was die niedrigere Modellleistung erklären kann.
Vergleichbar sind die Ergebnisse der Klassengruppe N-2 zuletzt mit den Ergebnissen von DROIN et al. (2020). Diese erreichen eine OA von 0,93 % für die Klassen Wohngebäude, Nicht-Wohngebäude und Nebengebäude. Diese Klassenkombination wird in der vorliegenden Arbeit nicht getestet. Ersichtlich ist in den Ergebnissen der vorliegenden Arbeit, dass die höchste OA bei der Klasse der Haupt- und Nebengebäude erreicht wird. Dass DROIN et al. (2020) mit der Klasse der Nebengebäude etwas höhere OA-Werte erreichen als im vorliegenden Fall die Klassengruppe N-2 (also ohne Nebengebäude), könnte ein Hinweis darauf sein, dass Nebengebäude grundsätzlich gut differenzierbar sind.
Allgemein lässt sich festhalten, dass der Vergleich mit anderen Studien nicht trivial ist, da viele unterschiedliche Parameter-Einstellungen zu anderen Ergebnisse führen können und in der vorhandenen Literatur variierende Methoden auf unterschiedliche Daten angewandt werden. Für einen guten Vergleich wäre ein Referenzdatensatz nötig, auf den die einzelnen Methoden angewandt werden. Für die binäre Unterscheidung in Gebäude und kein Gebäude existieren hierfür bereits Datensätze wie der Vaihingen- und Potsdam-Datensatz (ROTTENSTEINER et al. 2012) und für die Erkennung unterschiedliche Dachformen existiert beispielsweise der Datensatz des IEEE GRSS Data Fusion Contest (PERSELLO et al. 2023). Hinzu kommt die Vielfalt an Metriken zur Analyse der Modellleistung, was einen Vergleich er-

schwert. Um dies zu erleichtern werden für die vorliegende Arbeit mehrere Metriken berechnet und für die wichtigsten Ergebnisse angegeben.

5.1.3 Auswahl und Test unterschiedlicher Trainingsdaten

Untersuchungsgebiete

Werden die verwendeten Untersuchungsgebiete betrachtet, so lässt sich festhalten, dass unterschiedliche Arten besiedelter Gebiete in NRW für die Methodik verwendet werden. Diese Diversität ist besonders wichtig, um abschätzen zu können, wie gut das hier trainierte Modell auf ganz Deutschland anwendbar wäre. Grundsätzlich sollte dies möglich sein, da die ausgewählten Gebiete in NRW als repräsentativ für Deutschland betrachtet werden. In Regionen außerhalb von Deutschland mit anderen Siedlungsmustern, stellen diese das Modell vor Herausforderungen, da diese Siedlungsstrukturen in den Trainingsdaten nicht vorhanden sind. Für diesen Fall müsste das Modell neu trainiert werden oder es könnten die in NRW vortrainierten Gewichte als Ausgangsbasis für ein neues Training verwendet werden. Die Modellleistung der gewählten Gebäudetypen in NRW wäre hierfür nicht direkt vergleichbar.
Wird analysiert, wie unterschiedliche Untersuchungsgebiete das Modellergebnis beeinflussen, so ist eine teils große Variation erkennbar. Ursache hierfür könnte die Inhomogenität der Untersuchungsregionen sein. Ob jedoch die unterschiedlichen Trainingsdaten oder die unterschiedlichen Testdaten für die Differenz in der Modellleistung verantwortlich sind, ist mit der Methode, eine Region als Testregion und alle anderen Regionen als Trainingsregion zu verwenden, nicht identifizierbar. Einen Hinweis hierauf liefert die Korrelationsanalyse, durch welche eine höhere Korrelation zwischen der Anzahl der Gebäudetypen zu den Ergebnissen der Trainings- als zu den Testdaten herausgefunden wurde. Dies ist als vorteilhaft einzustufen, da dasselbe Modell auf unterschiedlichen Testdaten möglichst dasselbe Ergebnis liefern sollte, während die Trainingsdaten das Modell stark verändern können.
Ergänzend ist anzumerken, dass die gerechnete Korrelation nicht genau die Korrelation der verwendeten Daten darstellt, da für das Modelltraining 25.000 zufällig ausgewählte Kacheln verwendet wurden, für die Korrelationsanalyse jedoch die Häufigkeiten der Pixel je Gebäudetyp in einer kompletten Region verwendet wurden. Da die 25.000 Kacheln jedoch eine repräsentative Auswahl der Daten sein sollten, ist das Ergebnis der Korrelationsanalyse verwendbar.
Ein weiteres, die Untersuchungsgebiete betreffendes Ergebnis, stellt die mögliche Verwendung des amtlichen Gebäudedatensatzes ohne manuelle Verbesserung als Trainingsdaten dar. Das hiermit eine ähnliche Modellleistung wie mit den verbesserten Daten erreichbar ist, ist als

sehr vorteilhaft einzustufen, da so ein großer Mehraufwand in der Datenvorverarbeitung nicht nötig ist. Für die Klasse der Wohngebäude verbessert sich die Modellleistung stark, was auf die unterschiedliche Häufigkeit dieser Klasse in den Trainingsdaten zurück zu führen sein könnte. Auf die grundsätzliche Erkennung der Gebäudeumrisse hat die Verwendung von nicht angepassten Daten vermutlich nur einen geringen Einfluss, da die Hintergrundklasse für beide Fälle fast genau gleich gut erkannt wird.

Auflösung

Die getesteten Auflösungen verändern nach Abbildung 15 die Modellleistung nur in einem verhältnismäßig geringen Maß. Dies ist als positiv zu beurteilen, da die Modelle so mit einer ähnlichen Modellleistung auf andere Regionen angewandt werden können, in denen eine Auflösung von 10 cm für das Luftbild nicht vorliegt. Wie in Kapitel 3.3.1 erläutert, ist beispielsweise für ganz Deutschland ein Luftbild mit 20 cm Auflösung vorhanden und die Auflösung von 50 cm befindet sich bereits in einem Bereich, der von vielen Satellitendaten erreicht wird. HOFFMANN et al. (2021) vergleichen die Erkennung der Gebäudetypen Wohngebäude und Nicht-Wohngebäude sowie der Hintergrundklasse mittels CNNs und den zwei unterschiedlichen Auflösungen von ca. 0,6 m und ca. 2,4 m. Sie stellen eine bessere Erkennung der Gebäudetypen bei höher aufgelösten Daten, als bei niedrig aufgelösten Daten fest. In der vorliegenden Arbeit ist dies nicht der Fall, was an den anderen hierfür verwendeten Gebäudetypen liegen könnte, deren Erkennungsmerkmale in unterschiedlichen Auflösungen unterschiedlich gut erfassbar sind. HOFFMANN et al. (2021) stellen zudem eine bessere Erkennung der Hintergrundklasse für niedrigere Auflösungen fest. In der vorliegenden Arbeit wird der Hintergrund mit sinkender Auflösung schlechter erkannt, was mit dem Fehlen von Kontextinformationen zusammenhängen könnte. Die Differenz zu dem Ergebnis von HOFFMANN et al. (2021) könnte zudem an dem Fehlen der Verwendung eines Höhenmodells in ihrer Studie liegen.

Das Ergebnis, dass die Klassen der Reihenhäuser inkl. Blockrandbebauung und Mehrparteienhäuser mit 50 cm Auflösung am besten erkennbar sind, lässt den Schluss zu, dass größere Gebäude mit niedrigeren Auflösungen besser erkennbar sind. Die mögliche Ursache im Zusammenhang mit dem jeweils sichtbaren Teil des Gebäudes auf einer Kachel wurde in Kapitel 5.1.2 bereits diskutiert. Hinzu kommt, dass bei geringeren Auflösungen mehr Kontext in Form von anderen Gebäuden je Kachel sichtbar ist. Das Modell bezieht also für die Klassifikation des Pixels eines Gebäudes auch die Gebäude um diesen Pixel herum mit in die Entscheidung ein. Durch die vorhandene positive räumliche Autokorrelation kann es sein, dass dies vorteilhaft für die Modellleistung ist. Hiermit ist nach TOBLER (1970) gemeint, dass

räumlich nähere Dinge stärker als entfernte Dinge miteinander zusammenhängen, wobei alles mit allem zusammenhängt. Im vorliegenden Fall bedeutet dies, dass beispielsweise Einfamilienhäuser sich zumeist in Gegenden mit anderen Einfamilienhäusern befinden. Diese anderen Einfamilienhäuser werden von dem Modell bei einer niedrigen Auflösung und dem damit verbundenen größeren Gebiet mit mehr Kontextinformationen je Kacheln mitbeachtet.

Datenkanäle

Grundsätzlich lässt sich die in Kapitel 3.1.2 beschriebene Datenherkunft betreffend feststellen, dass die verwendeten Daten nicht exakt das gleiche Erstellungsdatum aufweisen. Die hierdurch entstandenen Differenzen in dem Gebäudebestand des Gebäudemodells zu dem DOP wurde durch die manuelle Bearbeitung der Gebäudepolygone gelöst. Zwischen dem DOP und dem nDOM wiederum liegen jedoch bis zu zwei Jahre Unterschied, in welchen sich der Gebäudebestand verändern konnte. Es können also z. B. Gebäude existieren, welche im nDOM vorhanden sind und im DOP nicht. In dem Gebäudebestand würde das Gebäude nicht existieren. Da die fünf Haupt-Untersuchungsgebiete eine Fläche von insg. 9,5 km^2 ausmachen, sollte die Anzahl der Flächen, auf denen ein Unterschied zwischen DOP und nDOM vorhanden ist, durch die große Zahl der Fläche ohne Änderung ausgeglichen werden und keine große Rolle spielen.

Das Ergebnis, dass das NIR mit den RGB-Kanälen und ohne nDOM die Erkennung der Gebäude verbessert, ist nachvollziehbar, da der NIR-Kanal in der Fernerkundung zumeist zur Vegetationserkennung oder -klassifizierung verwendet wird. Dies unterstützt somit die Abgrenzung der Vegetation von Gebäuden (PROTOPAPADAKIS et al. 2021). Bei ZHANG et al. (2018) verbessert das NIR dementsprechend die Landnutzungsklassifikation mit der Klasse der Gebäude. Dass in der vorliegenden Arbeit jeder Gebäudetyp mit dem NIR besser erfasst wird, als nur mit den RGB-Kanälen, könnte daran liegen, dass die meisten Gebäude hiermit besser erfassbar sind, unabhängig von ihrem Typ. Für eine Analyse, ob das NIR explizit auch die Differenzierung der einzelnen Gebäudetypen voneinander unterstützt, müsste die Klassifizierung getrennt von der Segmentierung der Gebäude durchgeführt werden. Einen Hinweis darauf, dass das NIR nur bei der Abgrenzung der Gebäude von Vorteil ist, nicht aber bei ihrer Klassifizierung, ist das Ergebnis, dass das NIR mit dem nDOM zu einer geringeren Modellleistung führt, als ohne NIR und mit nDOM. Grund hierfür ist die These, dass das nDOM die Abgrenzung der Gebäude bereits stark verbessert und das NIR hierzu keinen Mehrwert liefern kann, wie dies ohne das nDOM der Fall ist.

Dass das nDOM die Gebäudeerkennung verbessert ist nachvollziehbar, da hiermit eine klare Abgrenzung der Gebäude von ihrer Umgebung möglich ist. Die verbesserte Erkennung unter-

schiedlicher Gebäudetypen wiederum zeigt sich besonders bei der Klasse der Nebengebäude. Ursache hierfür könnte sein, dass Nebengebäude wie Schuppen, Carporte oder Gartenhäuser zumeist eine deutlich geringere Höhe als andere Gebäude aufweisen. Grundsätzlich könnte eine Trennung der Verwendung der Luftbilder und der Höhenmodelle in dem CNN das Ergebnis verbessern, da das direkte Zusammenführen dieser Daten vor dem CNN die Variabilität unterschiedlicher Datenquellen ignoriert (LUO et al. 2021). Hiermit ist gemeint, dass wie beispielsweise bei BITTNER et al. (2018) die Luftbilddaten und das Höhenmodell unabhängig voneinander in zwei Encodern verarbeitet werden und die Ergebnisse erst durch ein Zusammenführen im Decoder erreicht werden.

Zu den Ergebnissen ohne nDOM lässt sich abschließend sagen, dass die IoU-Werte der Klassen so zwar niedriger sind als mit nDOM, jedoch für ausgewählte Klassen je nach Anwendungsfall eine ausreichende Güte aufweisen. Mit Ausnahme der Klassen für Nebengebäude und Gebäudekomplexe inkl. Hallen zeigen die IoU-Werte ohne Verwendung des nDOMs meist einen maximalen Rückgang um bis zu 10 im Vergleich zu den IoU-Werten mit nDOM. Somit könnte das Modell auch ohne nDOM eine hinreichend gute Klassifizierungsgenauigkeit erreichen. Dies ist besonders von Interesse, wenn die in der vorliegenden Arbeit verwendete Methodik in Regionen angewandt werden soll, in denen kein hochaufgelöstes Höhenmodell zur Verfügung steht. Zumindest für die in NRW meist verbreiteten Bebauungsformen wurde dies in dieser Arbeit gezeigt und die Ergebnisse deuten darauf hin, dass dies auch mit anderen Bebauungsformen möglich wäre.

Anzahl der Trainingsdaten

Für den Test, ob die Verwendung aller Trainingskacheln ein anderes Ergebnis zeigt, wurde keine große Veränderung gegenüber der Verwendung von 25.000 Kacheln festgestellt. Dies lässt sich damit begründen, dass in den 25.000 Kacheln bereits fast alle Gebäude vorhanden sind. Im Falle der Verwendung von allen Kacheln zeigen die hinzukommenden Kacheln also dieselben Gebäude, nur an anderen Stellen der Kacheln, durch unterschiedliche Augmentationen und unterschiedliche Überlappungsbereiche.

Wichtig ist bei der Interpretation des Ergebnisses hierzu, dass es nicht direkt mit den anderen Ergebnissen vergleichbar ist, da ein anderer Decoder, eine andere Verlustfunktion und eine andere Epochenanzahl verwendet wurden. Die Epochenanzahl ist zudem innerhalb des Experiments nicht konsistent. Wie in Kapitel 4.2.2 gezeigt, weisen die Epochenanzahlen 20 und 50 Epochen jedoch vergleichbare Ergebnisse auf, sodass dies keine Einschränkung der Aussagekraft dieses Experiments bedeuten sollte. Auffallend ist für das Ergebnis des Tests unterschiedlicher Anzahlen an Trainingsdaten, dass Einfamilien- inkl. Doppelhäuser und sakrale

Gebäude mit beiden Datenanzahlen nicht erkannt werden. Da diese Klassen in dem Experiment unterschiedlicher Decoder jedoch zumindest teilweise erkannt werden, ist die Verlustfunktion ohne Gewichtung der Klassen die Ursache hierfür. Dies ist plausibel, da die entsprechenden Klassen nach Abbildung 6 am seltensten in den Daten vorhanden sind. Durch die Verwendung einer gewichteten Verlustfunktion erhalten diese seltenen Klassen ein stärkeres Gewicht im Modell und können dementsprechend besser erkannt werden (MAXWELL et al. 2021a).

Auswahl der Trainingskacheln

Das Ergebnis des Tests mit und ohne Daten-Sampler zeigt, dass die Auswahl von Kacheln mit einer Mindestmenge an Gebäudepixeln sinnvoll ist. Dass sich die Erkennung der Hintergrundklasse nicht verändert, zeigt zudem, dass eine höhere Anzahl an Kacheln ohne Gebäude in den Trainingsdaten nicht zu einer besseren Differenzierung zwischen Gebäude- und Nicht-Gebäude führt. Das Modell „sieht“ also auch mit Daten-Sampler genügend Flächen ohne Gebäude um deren Merkmale zu erlernen. Kritisch anzumerken ist, dass durch die gewählte Vorgehensweise unabhängig von der Auflösung mindestens 500 Pixel einem Gebäude zugeordnet sein müssen. Je höher also die Auflösung, desto größer auch die mindestens vorhandene Gebäudefläche. Bei höheren Auflösungen könnte hier auch eine Mindestfläche statt Mindestpixelanzahl gewählt werden. Zudem könnte eine Mindestmenge von Pixeln jeder Gebäudeklasse gewählt werden, so dass das Klassenungleichgewicht reduziert wird. Indirekt wird dies durch die verwendete Verlustfunktion mit Gewichten berücksichtigt.

5.2 Einfluss der Modellparameter

Zur generellen Wahl der Modellparameter ist es wichtig anzumerken, dass diese Wahl in Abhängigkeit von den spezifischen Anforderungen des verwendeten Datensatzes und der Architektur des Modells getroffen wurde, um die bestmöglichen Ergebnisse zu erzielen. Im Idealfall wird für die meisten der vorhandenen Hyperparameter des Modells die beste Einstellung gesucht. Alle möglichen Einstellungen aller Parameter zu testen ist jedoch sehr schnell sehr rechenintensiv (BOCHINSKI et al. 2017). Hierfür könnten automatisierte Tuningalgorithmen eine Lösung darstellen (ASZEMI & DOMINIC 2019). Für die Beantwortung der Forschungsfrage dieser Arbeit ist es jedoch von Bedeutung einzelne Einstellungen der gewählten Parameter zu testen. Auf diese Weise wurden die drei wichtigsten Modellparameter manuell optimiert. Grundsätzlich könnte insbesondere die Wahl eines anderen Encoders oder Optimie-

rungsalgorithmus das Ergebnis noch verbessern. Auch die Lernrate sowie die Batch-Größe stellen häufig optimierte Parameter dar (ASZEMI & DOMINIC 2019).

Zum generellen Vorgehen ist zudem zu hinterfragen, dass für alle Modelltrainings *early stopping* verwendet wurde. Grundsätzlich wird diese Methode, nicht das Modell einer bestimmten Epoche zu verwenden, sondern das Modell mit der besten Validierungs-IoU oder dem niedrigsten Ergebnis der Verlustfunktion, sehr häufig in Studien mit DL-Methoden verwendet, wie beispielsweise bei den Studien im Gebäudekontext von HOFFMANN et al. (2021), WURM et al. (2021), LI et al. (2019) oder STILLER et al. (2023). Im vorliegenden Fall kann hiermit zwar eine Überanpassung des Modells an die Trainingsdaten vermindert werden, es wird jedoch der Parameter der Epochenanzahl indirekt umgangen, indem nicht für alle Modelltrainings das Modell derselben Epoche verwendet wird. Da dieses Vorgehen für alle Modelltrainings verwendet wird, sind die Ergebnisse aber als vergleichbar zu betrachten.

Zuletzt ist zur generell verwendeten Methodik anzumerken, dass der Umgang mit mehreren Kacheln der Modellvorhersage im Rahmen der Quantifizierung der Modellleistung auch anders möglich ist. Durch das Vorgehen, den am häufigsten vorhergesagten Gebäudetyp je Pixel zu verwenden, ist es möglich, dass ein Gebäude in unterschiedliche Gebäudetypen aufgeteilt wird, was je nach Gebäudetypen theoretisch möglich ist, in der Praxis jedoch zumeist nicht der Fall ist. Lösbar wäre dies durch eine Klassifizierung der einzelnen Gebäude wie bei STRELTSOV et al. (2020). Dies setzt voraus, dass zuvor eine Segmentierung der einzelnen Gebäude stattgefunden hat, so dass jedes Gebäude einzeln klassifiziert werden kann. Zudem könnte die Methode der Instanz-Segmentierung verwendet werden, um direkt einzelne Gebäude voneinander getrennt zu detektieren und einer Klasse zuzuweisen (MINAEE et al. 2021). Eine weitere Lösung wäre die Verwendung des Gebäudetyps, welcher am häufigsten je Gebäude vorhergesagt wird, wobei auch hier eine vorausgehende Segmentierung der Gebäude notwendig ist.

Modellarchitekturen

Das Ergebnis, dass der Decoder FPN für die Erkennung unterschiedlicher Gebäudetypen die beste Modellleistung erreicht, ist identisch mit dem Ergebnis von STILLER et al. (2023), dass FPN für die Segmentierung von Gebäuden unter der Verwendung der Kanäle des DOPs sowie des nDOMs mit denselben Daten wie in dieser Arbeit am besten geeignet ist. Unter der Verwendung derselben Daten aber mit unterschiedlichen Aufgaben zeigt also dasselbe Modell die beste Modellleistung. Hierzu passt auch das Ergebnis, dass MAnet in der vorliegenden Arbeit die niedrigste Modellleistung aufzeigt, was entgegengesetzt zu dem Decoder-Vergleich von AMIRGAN et al. (2022) ist. In ihrer Studie werden unterschiedliche Decoder für die Erkennung

von Gebäuden auf Luftbildern mit einem anderen Datensatz getestet und sie erreichen die beste Modellleistung mit MAnet. Wichtig ist hier, dass andere Daten verwendet werden, als in der vorliegenden Arbeit und bei STILLER et al. (2023). Grundsätzlich kann bei einem Vergleich mit anderen Studien also festgestellt werden, dass je nach Datengrundlage unterschiedliche Decoder das beste Ergebnis zeigen.
Im Vergleich mit STILLER et al. (2023) lässt sich zudem feststellen, dass die Differenz der Modellleistung zwischen unterschiedlichen Decodern bei einer Betrachtung mehrerer Gebäudetypen größer ist, als bei einer Betrachtung der Gebäude als einzige Klasse. Außerdem lässt sich durch das Ergebnis, dass Unet++ keine besseren Ergebnisse zeigt als Unet, festhalten, dass mehr trainierbare Gewichte in dem Modell nicht zwingend zu einem besseren Ergebnis führen.

Epochenanzahl

Die Epochenanzahl ist nach ASZEMI & DOMINIC (2019) und LI Z. et al. (2022) einer der wichtigsten Hyperparameter eines CNNs da hiermit die Über- und Unteranpassung an die Trainingsdaten gesteuert werden kann. Das Ergebnis, dass um die 60. Epoche die besten IoU-Werte erreichbar sind zeigt, dass die Standard-Epochenanzahl von 50 Epochen für die restlichen Experimente hierzu gut passt. Auch mit der 30. Epoche sind bereits ähnlich gute Werte erreichbar, was gerade im Kontext der „grünen KI“ (SCHWARTZ et al. 2020) für einen reduzierten Rechenaufwand von Interesse ist. Eine übermäßig große Epochenanzahl wie eine Anzahl von 200 Epochen bringt dementsprechend keinen Mehrwert.
Das Ergebnis aus Abbildung 31, dass die Validierungs-IoU bis zur 66. Epoche ansteigt und anschließend zu einem Plateau hin abfällt, während das Ergebnis der Trainings-Verlustfunktion weiter sinkt, deutet darauf hin, dass ab der 66. Epoche eine Überanpassung an die Trainingsdaten stattfindet. Das Modell lernt hier also Muster in den Trainingsdaten, welche idealerweise als Hintergrundrauschen und nicht als valide Muster erlernt werden (KETKAR & MOOLAYIL 2021). Grundsätzlich könnte auch eine zu hohe Lernrate Ursache für eine mit zunehmender Epochenzahl abnehmende Genauigkeit des Modells sein, da die Modellgewichte sich bei einer hohen Lernrate stark ändern und daher das lokale Minimum der Verlustfunktion überspringen können, an dem die Genauigkeit am höchsten wäre. Da die Genauigkeit in Form der IoU jedoch erst nach der 66. Epoche sinkt und die Lernrate sich automatisch anpasst (vgl. Kapitel 3.3.2), handelt es sich sehr wahrscheinlich um eine Überanpassung an die Daten.

Vortrainierte Modelle

Von den acht in Tabelle 1 aufgeführten Studien zur Erkennung von Gebäudetypen auf Fernerkundungsdaten verwenden vier vortrainierte Gewichte für ihre DL-Modelle (CAO & QIU 2018; HOFFMANN et al. 2019; STRELTSOV et al. 2020; DIMASSI et al. 2021). Auch in der vorliegenden Arbeit werden diese verwendet und erhöhen die Modellleistung, da fast alle Klassen mit vortrainierten Gewichten besser erkannt werden. Typischerweise wird diese Methode angewandt, wenn zu wenig Trainingsbilder und -labels zur Verfügung stehen (NEUPANE et al. 2021). Im vorliegenden Fall ist dies für spezielle Gebäudetypen der Fall (vgl. Häufigkeitsverteilung der Gebäudetypen in Abbildung 6). Dem steht entgegen, dass sich die Erkennung der Gebäudetypen auch bei einer Verwendung aller Trainingskacheln nicht stark verbessert, was auf ausreichend Daten hinweist. Jedoch führt die Verwendung aller Trainingskacheln wie bereits erläutert nicht zu einer Hinzunahme von „neuen" Gebäuden, sondern primär zu einer Verwendung von unterschiedlichen Verortungen und Darstellungen eines Gebäudes in mehreren Kacheln. Der vorhandene Datensatz könnte also noch vergrößert bzw. verändert werden, so dass das Klassenungleichgewicht reduziert wird, wie dies im vorherigen Kapitel bei der Diskussion der Auswahl der Trainingskacheln beschrieben wurde.

Die Idee, wieso für eine gänzlich andere Aufgabe vortrainierte Gewichte zu besseren Ergebnissen führen, ist, dass durch diese Elemente einer niedrigen Merkmalsebene erkannt werden können. Hiermit sind beispielsweise Kanten oder zusammenhängende Muster gemeint, welche auf den meisten Bildern vorhanden sind (GUPTA et al. 2022). In dem Modell der vorliegenden Arbeit geschieht dies also unter anderem in dem ersten *Convolutional Layer*, da in diesem die vortrainierten Gewichte angewandt werden.

Von Interesse ist zudem, dass mit den vortrainierten Gewichten des roten Kanals für das NIR bessere IoU-Werte erreicht werden können als mit den Gewichten des grünen Kanals. Ursache hierfür könnte sein, dass die spektralen Eigenschaften des nahen Infrarots zu denen des roten Kanals ähnlicher sind als zu denen des grünen Lichts. Der Wellenlängenbereich des NIR ist zudem näher an dem des roten Kanals als dem des grünen Kanals.

Zufallsvariablen

Für robuste Schlussfolgerungen sind nach BOUTHILLIER et al. (2021) mehrere Modelldurchläufe notwendig. Hierfür wurden in der vorliegenden Arbeit unterschiedliche *Seeds* gewählt, wodurch anschließend eine Spannweite der OA von 3 % für die *Seeds*, verwendet in Algorithmen des Modells, und auch für die *Seeds* der Datenvorverarbeitung erreicht wurde. Dies ist eine ähnliche Spannweite wie bei FELLICIOUS et al. (2020), welche unterschiedliche *Seeds*

für die Gewichtsinitialisierung sowie den Optimierungsalgorithmus testen. Sie verwenden den CIFAR-10 Datensatz, in welchem es um die Bildklassifizierung geht und geben eine Spannweite der Genauigkeit für fünf Modelle von bis 2,7 bis 5,1 % an.

Werden die Ergebnisse je Gebäudetyp analysiert, so fällt auf, dass die Streuung der *Seed*-Werte der Daten und der Modelle bei denselben Klassen am größten ist, besonders bei der Klasse der sakralen Gebäude Dies ist insofern erstaunlich, da die *Seeds* der Daten und der Modelle an sehr verschiedenen Stellen der Datenverarbeitung zum Einsatz kommen. Also bei der Datenanalyse im Modell wie auch bei der Datenauswahl zuvor. Dies lässt den Schluss zu, dass die Erkennung der sakralen Gebäude grundsätzlich stark vom Zufall abhängt. Für die restlichen Klassen ist es als positiv zu betrachten, dass der Zufall keinen zu großen Einfluss auf das Modellergebnis hat. Der Modell-*Seed* ist nach FELLICIOUS et al. (2020) besonders bei der initialen Wahl der Werte der Gewichte des Modells von Bedeutung. Diese werden bei dem ersten Durchlauf des Modells zufällig gewählt. Hier sollte sich die Spannweite der Ergebnisse verringern, wenn vortrainierte Gewichte verwendet werden, da die initialen Gewichtswerte dann nicht zufällig gewählt, sondern übernommen werden.

Die Daten-*Seeds* könnten zusätzlich in *Seeds* für die Trainingsdatenauswahl und die Augmentationen unterscheiden werden. Welches der beiden Elemente für die unterschiedlichen Ergebnisse der Daten-*Seeds* verantwortlich ist, könnte so erkannt werden. Ganz grundsätzlich wäre eine größere Anzahl an *Seed*-Tests hilfreich, wobei die in dieser Arbeit gezeigten Ergebnisse bereits einen guten Einblick in den Einfluss des Zufalls auf die Modellergebnisse geben.

6 Fazit

Die semantische Segmentierung von Gebäudetypen mittels flächig und aktuell verfügbarer Fernerkundungsdaten ist wichtig für eine Vielzahl von Anwendungen. Hierzu fehlt in aktueller Literatur ein Überblick, welche Gebäudetypen der Nutzung und der Form mit aktuellen DL-Architekturen wie gut erkennbar sind. Zudem existiert kein umfassender Test um den Einfluss mehrerer Parameter auf die Erkennung unterschiedlicher Gebäudetypen zu analysieren. Dementsprechend zeigt diese Arbeit, wie gut die Erfassung unterschiedlicher Gebäudetypen möglich ist und welchen Einfluss verschiedene Parameter-Einstellungen auf die Erkennung haben.

Um die erste Forschungslücke zu schließen, wurden mehrere Klassengruppen gebildet, welche typische Gebäudetypen in Deutschland beinhalten und es wurde analysiert, wie die semantische Segmentierung dieser Gebäudetypen funktioniert. Grundsätzlich variiert die Modellleistung je Gebäudetyp stark, wobei für Gebäudetypen der Form Reihenhäuser sowie Haupt- und Nebengebäude am besten erfassbar sind. Von den Gebäudetypen der Nutzung sind Wohn- und Nicht-Wohngebäude am besten differenzierbar, was besonders für eine Anwendung in der Risikoanalyse von Naturgefahren vorteilhaft ist. Einschränkend soll hier angemerkt werden, dass je nach Anwendungsfall die mögliche Genauigkeit der Erkennung berücksichtigt werden muss.

Um die zweite Forschungslücke zu schließen, wurden mehrere Parameter der Daten sowie des DL-Modells ausgewählt und systematische Tests unterschiedlicher Einstellungen je Parameter wurden durchgeführt. Hierbei zeigte sich, dass die passenden Trainingsdaten von großer Bedeutung sind, wobei für Gebäudetypen der Nutzung der amtliche Gebäudedatensatz verwendet werden kann. Zudem verbessert die Hinzunahme eines nDOMs zu RGB-Daten die semantische Segmentierung stark wobei das nahe Infrarot nur ohne das nDOM zu besseren Ergebnissen führt. Das CNN-Modell betreffend ist als Decoder FPN am besten geeignet und die Verwendung von vortrainierten Gewichten führt zu einer verbesserten Erkennung von nahezu allen Klassen. Insbesondere für das nDOM sowie das NIR erhöht die Verwendung von Gewichten, vortrainiert auf einen RGB-Datensatz die Modellleistung.

Je nach Anwendungsfall kann die in dieser Arbeit entwickelte Methodik in Verbindung mit den optimalen Parametereinstellungen, die in dieser Studie erarbeitet wurden, für die Identifizierung spezifischer Gebäudetypen verwendet werden. In dieser Arbeit wurden hierfür nur Fernerkundungsdaten betrachtet. Durch eine Hinzunahme von zusätzlichen semantischen Informationen, wie beispielsweise Informationen aus Open Street Map, könnte diese Klassifizierung noch verfeinert werden. Von besonderem Interesse ist auch die Anwendung in

Regionen mit begrenzter Datenverfügbarkeit. Dies umfasst auf zonaler Ebene die Anwendung außerhalb NRWs, um zu testen, wie gut die Methodik deutschlandweit funktioniert, mit einer geringeren Auflösung des DOPs und des nDOMs. Zusätzlich besteht auf globaler Ebene Interesse daran, die Anwendbarkeit der Methodik in Regionen ohne vorhandenes nDOM und mit hochaufgelösten Satellitendaten zu untersuchen. So kann herausgefunden werden, wie gut die semantische Segmentierung unterschiedlicher Gebäudetypen in Regionen mit begrenzten Datensätzen funktioniert wobei die vorliegende Arbeit bereits Anhaltspunkte hierzu aufzeigt.

Literaturverzeichnis

ALIDOOST, F. und AREFI, H. (2018): A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image. In: PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 86 (5-6), 235–248.

AMIRGAN, B., AWAD, B., ERER, I. und MUSAOĞLU, N. (2022): A Comparative Study for Building Segmentation in Remote Sensing Images Using Deep Networks: Cscrs Istanbul Building Dataset and Results. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-2-2022, 1–6.

AN, Y., YE, Q., GUO, J. und DONG, R. (2020): Overlap Training to Mitigate Inconsistencies Caused by Image Tiling in CNNs. In: Bramer, M. und Ellis, R. (Hrsg.): Artificial Intelligence XXXVII 12498. Cham: Springer International Publishing, 35–48.

ASZEMI, N. M. und DOMINIC, P. (2019): Hyperparameter Optimization in Convolutional Neural Network using Genetic Algorithms. In: International Journal of Advanced Computer Science and Applications, 10 (6), 269–278.

BADRINARAYANAN, V., KENDALL, A. und CIPOLLA, R. (2017): SegNet: a Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. In: IEEE transactions on pattern analysis and machine intelligence, 39 (12), 2481–2495.

BAHETI, B., GAJRE, S. und TALBAR, S. (2019): Semantic Scene Understanding in Unstructured Environment with Deep Convolutional Neural Network. In: TENCON 2019-2019 IEEE Region 10 Conference. IEEE, 790–795.

BALL, J. E., ANDERSON, D. T. und CHAN, C. S. (2017): Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. In: Journal of Applied Remote Sensing, 11 (04), 1–55.

BANDAM, A., BUSARI, E., SYRANIDOU, C., LINSSEN, J. und STOLTEN, D. (2022): Classification of Building Types in Germany: A Data-Driven Modeling Approach. In: Data, 7 (4), 45–67.

BELGIU, M., TOMLJENOVIC, I., LAMPOLTSHAMMER, T., BLASCHKE, T. und HÖFLE, B. (2014): Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data. In: Remote Sensing, 6 (2), 1347–1366.

BEZIRKSREGIERUNG KÖLN (2017): Datenformatbeschreibung 3D-Gebäudemodell LoD1 Deutschland (LoD1-DE). Online unter: https://www.adv-online.de/icc/exteng/binarywriterservlet?imgUid=1c3607bb-6a25-3a51-afc2-2f77072e13d6&uBasVariant=11111111-1111-1111-1111-111111111111 (19.09.2023).

BHOSLE, K. und MUSANDE, V. (2019): Evaluation of Deep Learning CNN Model for Land Use Land Cover Classification and Crop Identification Using Hyperspectral Remote Sensing Images. In: Journal of the Indian Society of Remote Sensing, 47 (11), 1949–1958.

BHUYAN, K., VAN WESTEN, C., WANG, J. und MEENA, S. R. (2022): Mapping and characterising buildings for flood exposure analysis using open-source data and artificial intelligence. In: Natural Hazards, 119 (2), 1–31.

BITTNER, K., ADAM, F., CUI, S., KORNER, M. und REINARTZ, P. (2018): Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11 (8), 2615–2629.

BOCHINSKI, E., SENST, T. und SIKORA, T. (2017): Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms. In: 2017 IEEE international conference on image processing (ICIP), 3924–3928.

BOUTHILLIER, X., DELAUNAY, P., BRONZI, M., TROFIMOV, A., NICHYPORUK, B., SZETO, J., SEPAHVAND, N. M., RAFF, E., MADAN, K., VOLETI, V., KAHOU, S. E., MICHALSKI, V., ARBEL, T., PAL, C., VAROQUAUX, G. und VINCENT, P. (2021): Accounting for variance in machine learning benchmarks. In: Proceedings of Machine Learning and Systems, 3, 747–769.

BUNDESAMT FÜR KARTOGRAPHIE UND GEODÄSIE (BKG) (2023): WMS Digitale Orthophotos Bodenauflösung 20 cm. Online unter: https://gdz.bkg.bund.de/index.php/default/webdienste/digitale-orthophotos/wms-digitale-orthophotos-bodenauflosung-20-cm-wms-dop.html (17.11.2023).

BUNDESZENTRALE FÜR POLITISCHE BILDUNG (BPB) (2020): Bevölkerung nach Bundesländern. Online unter: https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61535/bevoelkerung-nach-bundeslaendern/ (18.09.2023).

BUYUKDEMIRCIOGLU, M., CAN, R. und KOCAMAN, S. (2021): Deep learning based roof type classification using very high resolution aerial imagery. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2021, 55–60.

CAO, R. und QIU, G. (2018): Urban Land Use Classification Based on Aerial and Ground Images. In: International Conference on Content-Based Multimedia Indexing (CBMI), 1–6.

CHEN, Q., WANG, L., WU, Y., WU, G., GUO, Z. und WASLANDER, S. L. (2019): Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings. In: ISPRS Journal of Photogrammetry and Remote Sensing, 147, 42–55.

COHEN, J. (1988): Statistical power analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates.

CONGALTON, R. G. und GREEN, K. (2019): Assessing the Accuracy of Remotely Sensed Data*:* Principles and Practices. Boca Raton: CRC Press.

DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K. und FEI-FEI, L. (2009): ImageNet: A large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, 248–255.

DIMASSI, M., SAMHAT, A. E., ZARAKET, M., HAIDAR, J., SHUKOR, M. und GHANDOUR, A. J. (2021): Buildings Classification using Very High Resolution Satellite Imagery. arXiv preprint: 2111.14650.

DROIN, A., WURM, M. und SULZER, W. (2020): Semantic labelling of building types. A comparison of two approaches using Random Forest and Deep Learning. In: Publikationen der DGPF, 29, 527–538.

DU, S., ZHANG, F. und ZHANG, X. (2015): Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. In: ISPRS Journal of Photogrammetry and Remote Sensing, 105, 107–119.

FAN, H., ZIPF, A. und FU, Q. (2014): Estimation of Building Types on OpenStreetMap Based on Urban Morphology Analysis. In: Connecting a digital Europe through location and place, 19–35.

FAN, T., WANG, G., LI, Y. und WANG, H. (2020): MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. In: IEEE Access, 8, 179656–179665.

FELLICIOUS, C., WEISSGERBER, T. und GRANITZER, M. (2020): Effects of Random Seeds on the Accuracy of Convolutional Neural Networks. In: Nicosia, G., Ojha, V., La Malfa, E., Jansen, G., Sciacca, V., Pardalos, P., Giuffrida, G., und Umeton, R. (Hrsg.): Machine Learning, Optimization, and Data Science. Cham: Springer International Publishing, 93–102.

GEOBASIS NRW (2022): Normalisiertes Digitales Oberflächenmodell 50 NW. Online unter: https://www.geoportal.nrw/?activetab=map#/datasets/iso/aa8f6bf6-1e2e-45bb-9b56-88b47e49cb2c (19.08.2023).

GOOGLE EARTH (2023): Google Earth. Online unter: https://earth.google.com/ (25.09.2023).

GOOGLE LLC (2023): Street View Static API – Nutzung und Abrechnung. Online unter: https://developers.google.com/maps/documentation/streetview/usage-and-billing?hl=de (20.12.2023).

GUO, Y., LIU, Y., GEORGIOU, T. und LEW, M. S. (2018): A review of semantic segmentation using deep neural networks. In: International Journal of Multimedia Information Retrieval, 7 (2), 87–93.

GUPTA, J., PATHAK, S. und KUMAR, G. (2022): Deep Learning (CNN) and Transfer Learning: A Review. In: Journal of Physics: Conference Series, 2273 (1), 12029–12039.

HAN, D., YUN, S., HEO, B. und YOO, Y. (2021): Rethinking Channel Dimensions for Efficient Model Design. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 732–741.

HAN, J. und PEI, J. und Hanghang TONG (2023): Data Mining*:* Concepts and Techniques. Cambridge: Morgan Kaufmann Publishers.

HE, K., ZHANG, X., REN, S. und SUN, J. (2016): Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

HECHT, R., HEROLD, H., MEINEL, G. und BUCHROITHNER, M. (2013): Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning. In: 26th international cartographic conference, 1–18.

HECHT, R., MEINEL, G. und BUCHROITHNER, M. (2015): Automatic identification of building types based on topographic databases – a comparison of different data sources. In: International Journal of Cartography, 1 (1), 18–31.

HOFFMANN, E. J., ALI, M. und ZHU, X. X. (2021): Zooming into Uncertainties: Towards Fusing Multi Zoom Level Imagery for Urban Land Use Segmentation. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2090–2093.

HOFFMANN, E. J., WANG, Y., WERNER, M., KANG, J. und ZHU, X. X. (2019): Model Fusion for Building Type Classification from Aerial and Street View Images. In: Remote Sensing, 11 (11), 1259.

HUANG, X., LI, S., LI, J., JIA, X., LI, J., ZHU, X. X. und BENEDIKTSSON, J. A. (2021): A Multispectral and Multiangle 3-D Convolutional Neural Network for the Classification of ZY-3 Satellite Images Over Urban Areas. In: IEEE Transactions on Geoscience and Remote Sensing, 59 (12), 10266–10285.

HUANG, Y., ZHUO, L., TAO, H., SHI, Q. und LIU, K. (2017): A Novel Building Type Classification Scheme Based on Integrated LiDAR and High-Resolution Images. In: Remote Sensing, 9 (7), 679–702.

HUSSAIN, E. und SHAN, J. (2016): Urban building extraction through object-based image classification assisted by digital surface model and zoning map. In: International Journal of Image and Data Fusion, 7 (1), 63–82.

IAKUBOVSKII, P. (2019): Segmentation Models Pytorch. Online unter: https://github.com/qubvel/segmentation_models.pytorch (15.11.2023).

JOCHEM, W. C., LEASURE, D. R., PANNELL, O., CHAMBERLAIN, H. R., JONES, P. und TATEM, A. J. (2021): Classifying settlement types from multi-scale spatial patterns of building footprints. In: Environment and Planning B: Urban Analytics and City Science, 48 (5), 1161–1179.

JOCHEM, W. C. und TATEM, A. J. (2021): Tools for Mapping Multi-Scale Settlement Patterns of Building Footprints: an Introduction to the R Package Foot. In: PloS one, 16 (2), e0247535.

KETKAR, N. und MOOLAYIL, J. (2021): Deep Learning with Python. Berkeley, CA: Apress.

KIM, J., HATZIS, J. J., KLOCKOW, K. und CAMPBELL, P. A. (2022): Building classification using random forest to develop a geodatabase for Probabilistic Hazard Information (PHI). In: Natural Hazards Review, 32 (3).

LANDESBETRIEB INFORMATION UND TECHNIK NORDRHEIN-WESTFALEN (IT.NRW) (2023): OpenGeodata.NRW. Online unter: https://www.opengeodata.nrw.de/produkte/geobasis/ (18.09.2023).

LEE, S., WOLBERG, G. und SHIN, S. Y. (1997): Scattered data interpolation with multilevel B-splines. In: IEEE Transactions on Visualization and Computer Graphics, 3 (3), 228–244.

LI J., HUANG X., TU L., ZHANG T. und WANG L. (2022): A review of building detection from very high resolution optical remote sensing images. In: GIScience & Remote Sensing, 59 (1), 1199–1225.

LI, W., HE, C., FANG, J., ZHENG, J., FU, H. und LE YU (2019): Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data. In: Remote Sensing, 11 (4), 403.

LI Z., LIU F., YANG W., PENG S. und ZHOU J. (2022): A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. In: IEEE transactions on neural networks and learning systems, 33 (12), 6999–7019.

LIGHTNING.AI (2023): N-Bit precision (Basic). Online unter: https://lightning.ai/docs/pytorch/stable/common/precision_basic.html (14.11.2023).

LIN, T.-Y., DOLLAR, P., GIRSHICK, R., HE, K., HARIHARAN, B. und BELONGIE, S. (2017): Feature Pyramid Networks for Object Detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2017, 2117–2125.

LIU, Y., ZHENG, X., AI, G., ZHANG, Y. und ZUO, Y. (2018): Generating a High-Precision True Digital Orthophoto Map Based on UAV Images. In: ISPRS International Journal of Geo-Information, 7 (9), 333.

LOGA, T. und STEIN, B., Nikolaus DIEFENBACH und Rolf BORN (2015): Deutsche Wohngebäudetypologie*:* Beispielhafte Maßnahmen zur Verbesserung der Energieeffizienz von typischen Wohngebäuden. Darmstadt: IWU (22.09.2023).

LOPEZ PINAYA, W. H., VIEIRA, S., GARCIA-DIAS, R. und MECHELLI, A. (2020): Convolutional neural networks. In: Mechelli, A. und Sandra Vieira (Hrsg.): Machine Learning. Cambridge, UK: Academic Press, 173–191.

LOSHCHILOV, I. und HUTTER, F. (2017): Decoupled Weight Decay Regularization. arXiv preprint: 1711.05101.

LU, Z., IM, J., RHEE, J. und HODGSON, M. (2014): Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. In: Landscape and Urban Planning, 130, 134–148.

LUO, L., LI, P. und YAN, X. (2021): Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review. In: Energies, 14 (23), 7982.

MA, L., LIU, Y., ZHANG, X., YE, Y., YIN, G. und JOHNSON, B. A. (2019): Deep learning in remote sensing applications: A meta-analysis and review. In: ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166–177.

MAFENI MASE, J., CHAPMAN, P., FIGUEREDO, G. P. und TORRES TORRES, M. (2020): Benchmarking Deep Learning Models for Driver Distraction Detection. In: Nicosia, G., Ojha, V., La Malfa, E., Jansen, G., Sciacca, V., Pardalos, P., Giuffrida, G., und Umeton, R. (Hrsg.): Machine Learning, Optimization, and Data Science. Vol. 12566. Cham: Springer International Publishing, 103–117.

MAXWELL, A. E., WARNER, T. A. und GUILLÉN, L. A. (2021a): Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. In: Remote Sensing, 13 (13), 2450.

MAXWELL, A. E., WARNER, T. A. und GUILLÉN, L. A. (2021b): Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. In: Remote Sensing, 13 (13), 2591.

MINAEE, S., BOYKOV, Y., PORIKLI, F., PLAZA, A., KEHTARNAVAZ, N. und TERZOPOULOS, D. (2021): Image Segmentation Using Deep Learning: A Survey. In: IEEE transactions on pattern analysis and machine intelligence, 44 (7), 3523–3542.

NEUPANE, B., HORANONT, T. und ARYAL, J. (2021): Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis. In: Remote Sensing, 13 (4), 808.

NICOSIA, G., OJHA, V., LA MALFA, E., JANSEN, G., SCIACCA, V., PARDALOS, P., GIUFFRIDA, G., und UMETON, R., Hrsg. (2020): Machine Learning, Optimization, and Data Science*:* 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part II. Cham: Springer International Publishing.

O'SHEA, K. und NASH, R. (2015): An Introduction to Convolutional Neural Networks. arXiv preprint: 1511.08458.

PAN, Z., XU, J., GUO, Y., HU, Y. und WANG, G. (2020): Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. In: Remote Sensing, 12 (10), 1574.

PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J. und CHANAN, G. e. a. (2019): PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: al Information Processing Systems, 32, 8024–8035.

PENG, D., ZHANG, Y. und GUAN, H. (2019): End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. In: Remote Sensing, 11 (11), 1382.

PERSELLO, C., HÄNSCH, R., VIVONE, G., CHEN, K., YAN, Z., TANG, D., HUANG, H., SCHMITT, M. und SUN, X. (2023): 2023 IEEE GRSS Data Fusion Contest: Large-Scale Fine-Grained Building Classification for Semantic Urban Reconstruction [Technical Committees]. In: IEEE Geoscience and Remote Sensing Magazine, 11 (1), 94–97.

PROTOPAPADAKIS, E., DOULAMIS, A., DOULAMIS, N. und MALTEZOS, E. (2021): Stacked Autoencoders Driven by Semi-Supervised Learning for Building Extraction from near Infrared Remote Sensing Imagery. In: Remote Sensing, 13 (3), 371.

QI, F., LIN, C., SHI, G. und LI, H. (2019): A Convolutional Encoder-Decoder Network With Skip Connections for Saliency Prediction. In: IEEE Access, 7, 60428–60438.

RAJU, V. N. G., LAKSHMI, K. P., JAIN, V. M., KALIDINDI, A. und PADMA, V. (2020): Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. In: Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 729–735.

ROBINSON, C., CHUGG, B., ANDERSON, B., FERRES, J. M. L. und HO, D. E. (2022): Mapping Industrial Poultry Operations at Scale With Deep Learning and Aerial Imagery. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 7458–7471.

RONNEBERGER, O., FISCHER, P. und BROX, T. (2015): U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., und Frangi, A. (Hrsg.): Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. Cham: Springer, 234–241.

ROTTENSTEINER, F., SOHN, G., JUNG, J., GERKE, M., BAILLARD, C., BENITEZ, S. und BREITKOPF, U. (2012): THE ISPRS BENCHMARK ON URBAN OBJECT CLASSIFICATION AND 3D BUILDING RECONSTRUCTION. In: ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, I-3, 293–298.

SAAD, A., MOHAMMED, A. T. und SAAD, A.-Z. (2017): Understanding of a Convolutional Neural Network. In: 2017 international conference on engineering and technology (ICET). Ieee, 1–6.

SCHINDLER, G., ROTH, W., PERNKOPF, F. und FRÖNING, H. (2020): Parameterized Structured Pruning for Deep Neural Networks. In: Nicosia, G., Ojha, V., La Malfa, E., Jansen, G., Sciacca, V., Pardalos, P., Giuffrida, G., und Umeton, R. (Hrsg.): Machine Learning, Optimization, and Data Science. Vol. 12566. Cham: Springer International Publishing, 16–27.

SCHWARTZ, R., DODGE, J., SMITH, N. A. und ETZIONI, O. (2020): Green AI. In: Communications of the ACM, 63 (12), 54–63.

SHORTEN, C. und KHOSHGOFTAAR, T. M. (2019): A survey on Image Data Augmentation for Deep Learning. In: Journal of Big Data, 6 (1).

SIMONYAN, K. und ZISSERMAN, A. (2015): Very Deep Convolutional Networks for Large-Scale Image Recognition. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, Conference Track Proceedings.

SINGH, D. und SINGH, B. (2020): Investigating the impact of data normalization on classification performance. In: Applied Soft Computing, 97, 105524–105546.

STILLER, D., STARK, T., STROBL, V., LEUPOLD, M., WURM, M. und TAUBENBÖCK, H. (2023): Efficiency of CNNs for Building Extraction: Comparative Analysis of Performance and Time. In: 2023 Joint Urban Remote Sensing Event (JURSE), 1–4.

STRELTSOV, A., MALOF, J. M., HUANG, B. und BRADBURY, K. (2020): Estimating residential building energy consumption using overhead imagery. In: Applied Energy, 280, 116018.

SUNDBORG, B., SZYBINSKA MATUSIAK, B. und ARBAB, S. (2019): Perimeter blocks in different forms – aspects of daylight and view. In: IOP Conference Series: Earth and Environmental Science, 323 (1), 12153.

SZEGEDY, C., VANHOUCKE, V., IOFFE, S., SHLENS, J. und WOJNA, Z. (2016): Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2818–2826.

TAUBENBÖCK, H., KLOTZ, M., WURM, M., SCHMIEDER, J., WAGNER, B., WOOSTER, M., ESCH, T. und DECH, S. (2013): Delineation of Central Business Districts in mega city regions using remotely sensed data. In: Remote Sensing of Environment, 136, 386–401.

TEMLITZ, K. (2007): Westfalen regional*:* Aktuelle Themen, Wissenswertes und Medien (= Siedlung und Landschaft in Westfalen, 35). Münster: Aschendorff.

THARWAT, A. (2021): Classification assessment methods. In: Applied Computing and Informatics, 17 (1), 168–192.

TOBLER, W. R. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region. In: Economic Geography, 46, 234–240.

VAN WESTEN, C. J., VAN ASCH, T. und SOETERS, R. (2006): Landslide hazard and risk zonation—why is it still so difficult?. In: Bulletin of Engineering Geology and the Environment, 65 (2), 167–184.

WANG, Y., LI, S., TENG, F., LIN, Y., WANG, M. und CAI, H. (2022): Improved Mask R-CNN for Rural Building Roof Type Recognition from UAV High-Resolution Images: A Case Study in Hunan Province, China. In: Remote Sensing, 14 (2), 265.

WHETSEL, K. B. (1968): Near-Infrared Spectrophotometry. In: Applied Spectroscopy Reviews, 2 (1), 1–67.

WURM, M., DROIN, A., STARK, T., GEIß, C., SULZER, W. und TAUBENBÖCK, H. (2021): Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling. In: International Journal of Geo-Information, 10 (1), 23.

WURM, M., SCHMITT, A. und TAUBENBOCK, H. (2016): Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9 (5), 1901–1912.

XIE, J. und ZHOU, J. (2017): Classification of Urban Building Type from High Spatial Resolution Remote Sensing Imagery Using Extended MRS and Soft BP Network. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10 (8), 3515–3528.

YANG C., ROTTENSTEINER F. und HEIPKE C. (2018): Classification of land cover and land use based on convolutional neural networks. In: ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, IV-3, 251–258.

YANG H., YUAN J., LUNGA D., LAVERDIERE M., ROSE A. und BHADURI B. (2018): Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11 (8), 2600–2614.

YU, D., XUEYING WANG UND HULIN WU (2021): EHR Data Pre-Processing and Preparation. In: Wu, H., Yamal, J.-M., Yaseen, A., und Maroufy, V. (Hrsg.): Statistics and Machine Learning Methods for EHR Data. Boca Raton: CRC Press Taylor & Francis Group, 111–147.

YUAN, Q., SHEN, H., LI, T., LI, Z., LI, S., JIANG, Y., XU, H., TAN, W., YANG, Q., WANG, J., GAO, J. und ZHANG, L. (2020): Deep learning in environmental remote sensing: Achievements and challenges. In: Remote Sensing of Environment, 241, 111716.

ZHANG, L., ZHANG, L. und DU, B. (2016): Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. In: IEEE Geoscience and Remote Sensing Magazine, 4 (2), 22–40.

ZHANG, P., KE, Y., ZHANG, Z., WANG, M., LI, P. und ZHANG, S. (2018): Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. In: Sensors, 18 (11).

ZHAO, W., DU, S., WANG, Q. und EMERY, W. J. (2017): Contextually guided very-high-resolution imagery classification with semantic segments. In: ISPRS Journal of Photogrammetry and Remote Sensing, 132, 48–60.

ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHSH, N. und LIANG, J. (2018): UNet++: a Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J. M. R. et al. (Hrsg.): Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Vol. 11045. Cham: Springer, 3–11.

ZHU, X. X., TUIA, D., MOU, L., XIA, G.-S., ZHANG, L., XU, F. und FRAUNDORFER, F. (2017): Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. In: IEEE Geoscience and Remote Sensing Magazine, 5 (4), 8–36.

Anhang

Anhang 1 Klassifikationsschema für Bildung der Klassengruppe N-1 aus den verwendeten LoD-Daten

Wohngebäude

Bezeichner LoD-Daten	Funktionscode LoD-Daten	Bezeichner LoD-Daten	Funktionscode LoD-Daten
Wohngebäude	1000	Gemischt genutztes Gebäude mit Wohnen	1100
Wohnhaus	1010	Land- und forstwirtschaftliches Wohngebäude	1210
Wochenendhaus	1312	Bauernhaus	1221
Wohngebäude mit Gemeinbedarf	1110	Gebäude für Gewerbe und Industrie mit Wohnen	2320
Wohngebäude mit Gewerbe und Industrie	1130	Gebäude für Handel und Dienstleistung mit Wohnen	2310
Wohngebäude mit Handel und Dienstleistungen	1120	Kaserne	3073
Wohn- und Betriebsgebäude	1131	Schwesternwohnheim	1023
Wohn- und Bürogebäude	1122	Seniorenheim	1022
Wohn- und Geschäftsgebäude	1123	Studenten-, Schülerwohnheim	1024
Wohn- und Verwaltungsgebäude	1121	Wohnheim	1020
Wohn- und Wirtschaftsgebäude	1222		

Andere

Gartenhaus	1313	Gebäude zur Gasversorgung	2570
Schuppen	2723	Gebäude zur Versorgung	2500
Carport	1611	Gebäude an unterirdischen Leitungen	2560
Garage	2463	Gebäude zur Entsorgung	2600
Gebäude zum Parken	2460	Müllbunker	2621
Tiefgarage	2465	Gebäude im Stadion	3230
Silo	1201	Zuschauertribüne, überdacht	1431
Treibhaus, Gewächshaus	2740	Gebäude im Freibad	3222
Treibhaus	2741	Überdachung	1610
Schutzhütte	3281	Schornstein, Schlot, Esse	1290
Schutzbunker	3074	Nach Quellenlage nicht zu spezifizieren	9998
Umformer	1400	Toilette	2612
Umspannwerk	2522		

Wirtschaftsgebäude

Gebäude für Land- und Forstwirtschaft	2700	Parkhaus	2461
Land- und forstwirtschaftliches Betriebsgebäude	2720	Parkdeck	2462
Scheune	2721	Einkaufszentrum	2052
Scheune und Stall	2726	Gebäude für Handel und Dienstleistungen	2010
Stall	2724	Geschäftsgebäude	2050
Betriebsgebäude	2112	Kaufhaus	2051
Fabrik	2111	Kiosk	2055
Gebäude für Gewerbe und Industrie	2100	Laden	2054
Produktionsgebäude	2110	Speditionsgebäude	2150
Werkstatt	2120	Fahrzeughalle	2464
Wirtschaftsgebäude	2729	Tankstelle	2130
Gebäude für Wirtschaft oder Gewerbe	2000	Waschstraße, Waschanlage, Waschhalle	2131
Gebäude für Vorratshaltung	2140	Gebäude für Forschungszwecke	2160
Lagerhalle, Lagerschuppen, Lagerhaus	2143	Campingplatzgebäude	2074
Speichergebäude	2142	Hotel, Motel, Pension	2071
Gebäude für betriebliche Sozialeinrichtung	2180	Gaststätte, Restaurant	2081
Bürogebäude	2020	Kantine	2083
Versicherung	2040	Gebäude für Bewirtung	2080
Kreditinstitut	2030		

Kommunalgebäude

Ärztehaus, Poliklinik	3053	Sport-, Turnhalle	3211
Krankenhaus	3051	Badegebäude	3220
Gebäude für Gesundheitswesen	3050	Hallenbad	3221
Feuerwehr	3072	Kegel-, Bowlinghalle	2093
Gebäude für Sicherheit und Ordnung	3070	Bezirksregierung	3018
Justizvollzugsanstalt	3075	Botschaft, Konsulat	3016
Polizei	3071	Gericht	3015
Allgemeinbildende Schule	3021	Kreisverwaltung	3017
Berufsbildende Schule	3022	Rathaus	3012
Forschungsinstitut	3024	Verwaltungsgebäude	3010
Gebäude für Bildung und Forschung	3020	Zollamt	3014
Hochschulgebäude (Fachhochschule, Universität)	3023	Gotteshaus	3045
Kinderkrippe, Kindergarten, Kindertages-stätte	3065	Kapelle	3043
Bibliothek, Bücherei	3037	Kirche	3041
Gebäude für soziale Zwecke	3060	Gebäude für religiöse Zwecke	3040
Wartehalle	2412	Gemeindehaus	3044
Gebäude für Fernmeldewesen	2540	Friedhofsgebäude	3080
Gebäude zur Elektrizitätsversorgung	2520	Trauerhalle	3081
Gebäude zur Energieversorgung	2501	Gebäude für kulturelle Zwecke	3030
Gebäude zur Freizeitgestaltung	1310	Museum	3034
Jugendfreizeitheim	3061	Burg, Festung	3038
Seniorenfreizeitstätte	3063	Schloss	3031
Kino	2092	Gebäude im botanischen Garten	3270
Theater, Oper	3032	Gewächshaus (Botanik)	3272
Gebäude für Sportzwecke	3210	Gebäude für öffentliche Zwecke	3000
Gebäude zum Sportplatz	3212		

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbstständig und ohne Hilfe Dritter verfasst habe. Bei der Masterarbeit wurden keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Alle den angegebenen Quellen entnommenen wörtlichen oder sinngemäßen Inhalte wurden von mir entsprechend kenntlich gemacht.

_____ _____

Ort, Datum Unterschrift