



Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Interfakultären Fachbereich für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

**„Ableitung von Gebäudeobjekten mittels objektbasierter
Klassifikation auf Basis von Digitalen Orthophotos und
Satellitendaten“**

vorgelegt von

Dipl.-Ing. (FH) Marius Rokus
106693, UNIGIS MSc Jahrgang 2020

Betreuer:

Univ.-Prof. Mag. Dr. Thomas Blaschke

Zur Erlangung des Grades

„Master of Science – MSc“

Karlsruhe, 09.05.2022

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig verfasst und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel verwendet und die den verwendeten Quellen und Hilfsmittel wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit wurde noch keiner anderen Prüfungskommission in dieser oder einer ähnlichen Form vorgelegt. Sie wurde bisher auch nicht veröffentlicht.

Karlsruhe, 09.05.2022

Marius Rokus

Danksagung

An dieser Stelle möchte ich mich bei den Personen bedanken, die mich bei der Ausarbeitung der Masterarbeit unterstützt und motiviert haben.

Zuerst gebührt mein Dank Herrn Prof. Dr. Thomas Blaschke für die fachliche Begleitung und Beurteilung meiner Masterarbeit. Für die unkomplizierte und angenehme Betreuung möchte ich mich herzlichst bedanken.

Zusätzlich möchte ich mich auch beim gesamten UNIGIS-Team für die professionelle Organisation des Studiums, die interessanten Module und die schnelle Beantwortung jeglicher Fragen bedanken.

Danken möchte ich auch meinem Arbeitgeber dem Landesamt für Geoinformation und Landentwicklung Baden-Württemberg und dem Vermessungsamt des Landkreises Heilbronn, sowie meinen Kolleginnen und Kollegen des Referats 53 für die Hilfe bei fachlichen Fragestellungen.

Ein besonderer Dank gilt dabei meiner Kollegin, Frau Maryse Hillemann. Ihre Unterstützung bei der Ausarbeitung des Themas, Ihre hilfreichen Anregungen und Ihre konstruktive Kritik waren mir eine sehr große Hilfe.

Mein größter Dank geht an meine Frau Helena, die mich den ganzen Weg des Studiums emotional und moralisch unterstützt hat, alle Höhen und Tiefen mit mir gemeistert und mir, trotz aller schwierigen Rahmenbedingungen, stets den Rücken freigehalten hat.

Ohne dich stünde ich nicht hier.

Zusammenfassung

Die Entwicklung von automatischen Verfahren zur Detektion von Landschaftsveränderungen aus Fernerkundungsdaten hat in der Landesvermessung in den letzten Jahren an Interesse gewonnen. Die möglichen Anwendungsfelder sollen nun verstärkt untersucht werden. Hinweise über die Ersterfassungen von topographischen Objekten oder die Nacherfassung von veränderten, bereits bestehenden Objekten, wie den Gebäuden des Liegenschaftskatasters, erfolgen oftmals weiterhin mit traditionellen Mitteln wie Informationen aus Baugenehmigungsverfahren oder über Ortsbegehungen.

Flugzeuggestützte digitale Bildaufnahmen sind seit mehreren Jahren fester Bestandteil der Geobasisdatenerfassung in den Landesvermessungen der Bundesrepublik Deutschland. Mit dem neuen Qualitätsstandard TrueDOP wird diese Entwicklung fortgesetzt und liefert ab 2023 für Baden-Württemberg landesweit und flächendeckend 4-Kanal-Rasterdaten der Erdoberfläche ohne Verkippeffekt in einer lagerichtigen orthogonalen Darstellung mit einer Bodenauflösung von 20 cm. Auch Daten von Erdbeobachtungssatelliten sind durch ihre stetige Weiterentwicklung bezüglich der geometrischen Auflösung zu einer nennenswerten Alternative zu den etablierten hochauflösenden flugzeuggestützten Bildaufnahmen geworden. In dieser Thesis sind besonders die satellitengestützten Bildaufnahmen der Mission *PlanetScope* mit 8 Spektralkanälen, einer Pixelgröße von 3 m und einer täglichen Wiederkehrzeit hervorzuheben.

Die vorliegende Arbeit befasst sich mit der Frage, ob sich die beiden untersuchten Rasterdatensätze der flugzeug- und satellitengestützten Trägersysteme für eine Detektion von Gebäudeobjekten des amtlichen Liegenschaftskatasters eignen.

Mithilfe von Deep Learning-Methoden und der Anwendung von objektbasierter Bildanalyse werden die Rasterdaten klassifiziert und hinsichtlich ihrer Eignung für die Erkennung von Gebäudeobjekten untersucht und bewertet.

Für die Durchführung wird ein Trainingsgebiet mit einer Größe von 30 km² definiert, welches eine vielfältige und heterogene Verteilung von Gebäudeobjekten besitzt. Diese aus dem amtlichen ALKIS-Datensatz stammenden, eindeutigen Gebäudeobjekte dienen nach der Datenbereinigung als Trainingsdatensatz für das Neuronale Netz.

Zur Validierung wird ein Kontrolldatensatz in einem vom Trainingsgebiet räumlich getrennten Bereich innerhalb des befliegenen Interessensgebietes ausgewählt. Als *Ground-Truth*-Daten stehen wiederum die bereinigten amtlichen ALKIS-Gebäudedaten zur Verfügung. Ergänzend liegen Geodaten zu laufenden Baugenehmigungsverfahren vor, die als zusätzliche Kontrollmöglichkeit für bisher noch nicht eingemessenen, da noch im Bau befindlichen Objekten dienen.

Es werden drei Klassen definiert. Dabei gliedert sich die Aufteilung in zwei Zielklassen und eine Nicht-Zielklasse: (1) die erste Zielklasse umfasst kleine Objekte mit ei-

nem Flächeninhalt kleiner als 20 m²; (2) die zweite Zielklasse sind alle Gebäudeobjekte größer als 20 m²; (3) die Nicht-Zielklasse definiert alle restlichen Landschaftselemente, die keinem Gebäudeobjekt zugeordnet werden können, wie etwa Vegetation, Acker, Tiefbau, etc.

Nach der Datenbereinigung stehen 20.000 unikale Gebäudeobjekte, 5.000 Kleinstobjekte und 25.000 Nicht-Gebäudeobjekte zur Verfügung.

Die beiden Rasterdatensätze werden jeweils mit einem *Convolutional Neural Network* (CNN) trainiert und klassifiziert und die Ergebnisse anschließend mithilfe des Kontrollgebietes quantifiziert. Um eine Überanpassung des Trainingsdatensatzes zu vermeiden und die Qualität des Netzes zu verbessern, wird eine *Data Augmentation* angewendet. Dadurch können die Trainingsdaten jeder Klasse um den Faktor 8 bzw. 16 für die TrueDOP-Daten und Faktor 16 für die Satellitendaten erweitert werden.

Um die optimale CNN-Architektur und die bestmöglichen Parametereinstellungen zu finden, werden heuristische Methoden angewendet und der daraus erkennbare erfolgversprechendste Lösungsweg mit iterativen Schritten ermittelt und weiterverfolgt, um eine Verbesserung der Erkennungsqualität zu erreichen. Die dabei erreichten Wahrscheinlichkeiten für das Vorhandensein der gesuchten Klassen werden anschließend mithilfe von objektbasierter Bildanalyse und wissensbasierter Klassifikation verarbeitet.

In einem nächsten Schritt werden die Erkennungsqualitäten auf Basis der TrueDOP- und der Satellitendaten in Form einer Genauigkeitsbewertung quantitativ beurteilt, im Kontext der Fragestellung gegenübergestellt und eine Aussage darüber getroffen, welcher Datensatz sich hinsichtlich einer Detektion von fehlenden, also nicht erfassten Gebäudeobjekten eignet.

Dabei zeigte sich im Verlauf der Untersuchung, dass für die zu erkennende Klasse (1) „*Kleine Objekte*“, im Fall der TrueDOP-Daten keine Wahrscheinlichkeiten ermittelt werden konnten. Das durchgeführte Training des CNN erreichte keine zufriedenstellende Ableitung und Merkmalsextraktion von Objekten mit einer geringen Flächengröße, wenn diese einer eigenen individuellen Klasse zugeordnet sind. Die Gesamtgenauigkeit dieser Untersuchung befand sich nur im Bereich von 80 %.

Eine Lösung stellt die Integration der Klasse (1) in die Hauptklasse (2) dar. Nachdem die Umstellung auf eine binäre Klassifikation mit einer Zielklasse und einer Nicht-Zielklasse erfolgt, erreicht die Gesamtgenauigkeit einen Wert von 91,6 % bis 93,6 % für die TrueDOP-Daten. Dieser Wert konnte durch eine Verfeinerung mittels objektbasierter Bildanalyse auf 94,6 % erhöht werden.

Bei der Untersuchung der Daten der Satellitenmission *PlanetScope* wird eine Gesamtgenauigkeit im Bereich von 77,7 % bis 88,4 % für die ausgewählten Parametereinstellungen erzielt. Durch die Verfeinerung der anschließenden objektbasierten Bildanaly-

se kann die Gesamtgenauigkeit auf hohem Niveau gehalten werden. In Folge der Anpassung reduziert sich die Gesamtgenauigkeit zwar leicht auf 84,5 %, aber das Ergebnis der Erkennung erscheint visuell differenzierter und erreicht in Folge eine höhere Güte durch die bessere Separation zwischen Gebäudeobjekten und sonstigen Infrastrukturen in dicht besiedelten Wohnbereichen.

Auffälligkeiten der ersten Untersuchungen der *PlanetScope*-Daten zeigten sich in einer signifikanten Überklassifikation des Kontrollgebietes. Diese konnte im weiteren Verlauf der Untersuchung durch Verlängerung der Trainingszeiten reduziert werden, welche allerdings in einer Reduktion der Gesamtgenauigkeit resultieren. Bezüglich der *PlanetScope*-Daten lässt sich schlussfolgern, dass der zeitliche Aspekt und der Aufwand für die Suche der optimalen Parameterwahl und den Trainingszeiten eine höhere Bedeutung einnehmen, um ein hochwertiges Klassifikationsergebnis zu erreichen.

Als Fazit der Untersuchungen der Thesis kann festgehalten werden, dass die Klassifikation von ALKIS-Gebäudeumringen aus TrueDOP grundsätzlich eine sehr hohe Genauigkeit erreicht, wenn viel Speicherplatz und eine große Rechenleistung zur Verfügung stehen und lange Berechnungszeiten in Kauf genommen werden können. Die Untersuchungen der Thesis wurden in einem räumlich beschränkten Gebiet von 30 km² durchgeführt. Für eine flächendeckende landkreisweite Realisierung bedarf es weiterer Analysen, insbesondere bezogen auf die benötigten IT-Komponenten.

Die Klassifikation von ALKIS-Gebäudeumringen aus *PlanetScope*-Daten erreicht eine annähernd vergleichbare Genauigkeit, allerdings mit Einschränkungen: Anbauten, Überdachungen und andere kleine Objekte sind in der vorhandenen Ausprägung der Auflösung und Bandanzahl nicht eindeutig klassifizierbar. Die geringe Pixelgröße von Objekten ermöglicht eine höhere Anzahl an Trainingsdaten, was in einem robusteren trainierten Netz resultieren kann, allerdings auf Kosten der Trainingszeiten. Die erreichten Ergebnisse weisen auf eine erkennbare Überrepräsentation der Gebäude im Kontrollgebiet hin. Vorteilhaft wirkte sich eine anschließende objektbasierte Verfeinerung der Klassifikation auf die visuelle Einschätzung der Ableitung der Gebäudeobjekte aus. Die Erkennbarkeit von Einzelobjekten ist möglich, setzt aber Trainingsdaten und -parameter in hoher Qualität voraus.

Abstract

The development of automatic methods for the detection of landscape changes from remote sensing data has gained interest in State Survey in recent years. The potential fields of application are now to be investigated more intensively. Information about the initial registration of topographic objects or the subsequent registration of already existing objects, such as buildings of the real estate cadastre, is often still obtained by traditional means, such as information from building permit procedures or via local inspections.

Airborne digital imagery has been an integral part of geospatial data acquisition in State Offices of the Federal Republic of Germany for several years. With the new quality standard TrueDOP, this development will be continued and, from 2023, will provide Baden-Württemberg with 4-band images without tilting effect in a positionally correct orthogonal representation with a ground sampling distance of 20 cm. Data from earth observation satellites have also become a significant alternative to the established high-resolution airborne imagery due to their continuous development in terms of geometric resolution. The satellite-based images of the *PlanetScope* mission with 8 spectral bands, a pixel size of 3 m and a daily revisit time are particularly noteworthy.

The present work deals with the question of whether the two investigated raster data sets of the airborne-based and satellite-based systems are suitable for the detection of building objects of the official real estate cadastre.

With the help of deep learning methods and object-based image analysis, the raster data sets are classified and evaluated regarding their suitability for the detection of building objects.

For the implementation, a training area with the size of 30 km² is defined, which has a diverse and heterogeneous distribution of building objects. These unique building objects, which originate from the Authoritative Real Estate Cadastre Informations System (ALKIS) data set, serve as the training data set for the neural network after data cleaning.

For validation, a control data set is selected in an area spatially separated from the training area within the area of interest. The official building data from the area is used as ground truth data. In addition, geodata on ongoing building permit procedures are available, which serve as another control possibility for objects that have not yet been officially added to ALKIS because they are still under construction.

Three classes are defined. The data is divided into 2 target classes and one non-target class: (1) the first target class includes small objects with an area smaller than 20 m²; (2) the second target class are building objects larger than 20 m²; (3) the non-target

class defines all remaining landscape elements that cannot be assigned to a building class, such as vegetation, cropland, roads, etc.

After data cleaning, 20,000 unique building objects, 5,000 unique small objects, and 25,000 unique non-building objects are available.

The two raster data sets TrueDOP and *PlanetScope* are each trained and classified using a Convolutional Neural Network (CNN), and the results are quantified using the control area. To avoid overfitting and to improve the quality of the network, data augmentation is applied. This allows the training data of each class to be augmented by a factor of 8 or 16 for TrueDOP data and a factor of 16 for satellite data.

To find the optimal CNN architecture and the best possible parameter settings, heuristic methods are applied, and the most promising solution pursued with iterative steps to achieve improved detection quality. The resulting probabilities for the presence of the searched classes are then further processed using object-based image analysis and knowledge-based classification.

In a next step, the detection qualities based on TrueDOP, and satellite data are quantitatively assessed in the form of an accuracy evaluation and compared in the context of the question which data set is best suited for the detection of missing, i.e. not recorded, building objects.

During the study it became clear that for the class (1) small objects no probability values could be determined in the case of TrueDOP data. The performed training of the CNN did not achieve a satisfactory feature extraction of small sized objects. The overall accuracy of this investigation was only in the range of 80 %.

A solution is the integration of class (1) into the main class (2). After switching to a binary classification with a target class and a non-target class, the overall accuracy reached a value of 91.6 % to 93.6 % for the TrueDOP data. This value could be increased to 94.6 % by refinement using object-based image analysis.

When examining the data from the *PlanetScope* satellite mission, an overall accuracy in the range of 77.7 % to 88.4 % is achieved. By refining using object-based image analysis, the overall accuracy can be maintained at a high level. As a result of the adjustment, the overall accuracy is slightly reduced to 84.5 %, but the detection results appear visually more differentiated and subsequently achieve a higher quality due to the better separation between building objects and other infrastructures in densely populated residential areas.

Anomalies of the first investigations of the *PlanetScope* data revealed a significant over-classification of the control area. This could be reduced in the further course of the study by extending the training times, which however resulted in a reduction of the overall accuracy. Concerning the *PlanetScope* data it can be concluded that the

time aspect and investigations regarding the optimal parameter selection and the training times are of high importance to achieve a high quality classification result.

In summary, the study concludes that the classification of ALKIS building perimeters from TrueDOP achieves a very high accuracy, if a lot of hard disk space and a large processing power are available and long computation times can be accepted. The investigations were accomplished in a spatially limited area of 30 km². For a county-wide implementation, further analyses are required, especially with respect to the required IT components.

The classification of ALKIS buildings from *PlanetScope* data achieves an approximately comparable accuracy, but with limitations: building extension, roofings and other small objects cannot be unambiguously classified in the existing specification of resolution and number of bands. The small pixel size allows for a higher number of training data, which can result in a more robust trained network, but at the expense of training times. The obtained results indicate a detectable overrepresentation of buildings in the control area. A subsequent object-based refinement of the classification had a beneficial effect on the visual assessment of the detection of the building objects. The detectability of single objects is possible but requires training data of high quality.

Abkürzungsverzeichnis

AdV	<i>Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland</i>
ALB	<i>Automatisiertes Liegenschaftsbuch</i>
ALK	<i>Automatisierte Liegenschaftskarte</i>
ALKIS	<i>Amtliches Liegenschaftskatasterinformationssystem</i>
ANN	<i>Artificial Neuronal Networks</i>
bDOM	<i>bildbasierte digitale Oberflächenmodell</i>
BW	<i>Baden-Württemberg</i>
CNN	<i>Convolutional Neural Network</i>
DGM	<i>Digitales Geländemodell</i>
DIM	<i>Dense Image Matching</i>
DLM	<i>Digitales Landschaftsmodell</i>
DOP	<i>Digitale Orthophotos</i>
DT	<i>Decision Tree</i>
ESP	<i>Estimation of Scale Parameter</i>
EVI	<i>Enhance Vegetation Index</i>
FN	<i>false negative</i>
FNEA	<i>Fractal Net Evolution Approach</i>
FP	<i>false positive</i>
GDI	<i>Geodateninfrastruktur</i>
GRVI	<i>Green-Red-Vegetation Index</i>
GT-Daten	<i>Ground-Truth-Daten</i>
LGL BW	<i>Landesamt für Geoinformation und Landentwicklung Baden-Württemberg</i>
LiDAR	<i>Light Detection and Ranging</i>
LK	<i>Landkreis</i>
LN/LB	<i>Landnutzung/Landbedeckung</i>
LV	<i>Lokale Varianz</i>
MRS	<i>Multiresolution Segmentation</i>
NDBI	<i>Normalized Difference Built-Up Index</i>
NDRR	<i>Normalized Difference RedEdge Red Index</i>
NDSI	<i>Normalized Difference Soil Index</i>
nDSM	<i>Normalized Digital Surface Model</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NDWI	<i>Normalized Difference Water Index</i>

NIR	<i>Nahes Infrarot</i>
OBIA	<i>Object-based image analysis</i>
ÖbVI	<i>Öffentlich bestellter Vermessungsingenieure</i>
ROC	<i>Veränderungsrate (Rate of Change)</i>
SAVI	<i>Soil Adjusted Vegetation Index</i>
SI	<i>Shadow Index</i>
SK	<i>Stadtkreis</i>
SP	<i>Scale Parameter</i>
TN	<i>true negative</i>
TP	<i>true positive</i>
UAV	<i>Unmanned Aerial Vehicle</i>
WV-BI	<i>WorldView Built-Up Index</i>

Abbildungsverzeichnis

Abbildung 1 - Lage des Landkreises Heilbronn in BW.....	8
Abbildung 2: Darstellung der Bildmitten der DOP20-Befliegung nach Befliegungstagen für den Landkreis Heilbronn.....	9
Abbildung 3: Erfassungsbereiche (Footprints) der flugzeug- und satellitengestützten Befliegungen.....	10
Abbildung 4: Zuwachs der ALKIS-Gebäudeobjekte in BW von 2019 bis 2021.....	12
Abbildung 5: Zuwachs der ALKIS -Gebäudeobjekte im SK und LK Heilbronn zwischen 2019 und 2021.....	13
Abbildung 6 - Verteilung der Flächengrößen der Gebäudeobjekte im Stadt- und Landkreis Heilbronn.....	14
Abbildung 7 - potenzielle Kontrollflächen der Klassifikation innerhalb des Untersuchungsgebietes.....	15
Abbildung 8 - Darstellung der Intraklassen-Heterogenität innerhalb der Klasse „Gebäudeobjekte“ a) gewerbliche Objekte b) Mischbebauung c) Einzelhausbebauung.....	16
Abbildung 9 - Beispieldarstellung Digitales Orthophoto in Baden-Württemberg.....	17
Abbildung 10 - Vergleich der Unterschiede des Verkippungseffektes bei DOP (links) und TrueDOP (rechts).....	18
Abbildung 11 - Geringe Kantenschärfe an Dachtraufen bei TrueDOP.....	19
Abbildung 12 - fehlerhafte Bildinformation von Ackerflächen im TrueDOP.....	19
Abbildung 13 – Bildfehler im TrueDOP: a) Sich bewegende Objekte erscheinen transparent b) Datenlücken im bDOM wegen schlechter Korrelation führen zu Artefakten in Gewässern.....	20
Abbildung 14 - Beispieldarstellung der vorliegenden PlanetScope-Satellitendaten.....	22
Abbildung 15 – Überstrahlungseffekt heller Objekte (links) und grüne Farbreflexion (rechts).....	24
Abbildung 16 – Einordnung von Deep Learning.....	27
Abbildung 17 - Aufbau eines einfachen ANN.....	28
Abbildung 18 - Funktionsweise der Faltungsoperation (nach Yamashita et al. (2018)).....	30
Abbildung 19 - Funktionsweise des Max-Pooling-Filters (nach Yamashita et al. (2018)).....	30
Abbildung 20 - Funktionsweise für die ersten beiden Reihen mit dem Hyperparameter Stride 2.....	31
Abbildung 21 - Übersicht der Bearbeitungsschritte für die praktische Umsetzung der Aufgabenstellung.....	38
Abbildung 22 - Übersicht der Anzahl an Gebäudeobjekten je km ²	39
Abbildung 23 - Darstellung des Kontrollgebietes (weißer Rahmen) innerhalb des LK Heilbronn.....	40
Abbildung 24 - Übersicht des Kontrollgebietes in der originären Auflösung für DOP- (links) und PlanetScope-Daten (rechts).....	41
Abbildung 25 - Beispiele von potenziellen Fehlerquellen der ALKIS-Umränge für den Einsatz zur Gebäudeobjektdetektion a) fehlendes Gebäude b) Gebäudeeinmessung stimmt nicht mit Bildaufnahme überein c) Baum verdeckt Gebäudeobjekt d) Gebäudeeinmessung größer als das tatsächliche Gebäude.....	42
Abbildung 26 - Workflow für die halbautomatische Kontrolle der ALKIS-Daten für das Trainingsgebiet.....	42
Abbildung 27 - Kombination der ALKIS-Daten (rote Umrandung) und der im Bau befindlichen Objekte (blaue Umrandung) in der DOP-Darstellung.....	44
Abbildung 28 – Workflow für die Erstellung der Trainingsdaten.....	44
Abbildung 29 - Trainingsdaten für die Zielklasse Nicht-Gebäude (grün) für DOP (links) und SAT (rechts). ALKIS-Gebäudeobjekte sind in orange dargestellt.....	48
Abbildung 30 - Beispieldarstellung der angewandten Data Augmentation mithilfe von Rotation, Spiegelung und Helligkeitsanpassung.....	49
Abbildung 31 - visuelle Repräsentation der angewendeten Schleife für die Rotation der Training Samples des DOP-Datensatzes.....	50
Abbildung 32 - Duplikate bei horizontaler und vertikaler Spiegelung nach vorheriger Drehung.....	51
Abbildung 33 – FME-Workbench zur Spiegelung der Training Samples für SAT-Daten.....	51
Abbildung 34 – FME-Workbench zur Helligkeitsanpassung der Training Samples für DOP.....	52
Abbildung 35 – Bisher nicht eingemessenes Neubaugebiet in Ellhofen.....	54

<i>Abbildung 36 - Darstellung der manuell erfassten Gebäudeobjekte gemäß Bauerfassungskataster (blau) im Vergleich des noch im Bau befindlichen Objektes in der Bildmitte</i>	54
<i>Abbildung 37 - Beispiele der GT-Klasse "Nicht-Gebäude" (blau)</i>	55
<i>Abbildung 38 - Darstellung der Funktionsweise der unter eCognition verwendeten CNN-Architektur mit den entsprechenden Parametern für die DOP-Daten und der anschließenden objektbasierten Klassifikation (nach Ghorbanzadeh et al. (2021))</i>	56
<i>Abbildung 39 - Darstellung der Funktionsweise der unter eCognition verwendeten CNN-Architektur für SAT und objektbasierter Klassifikation (nach Ghorbanzadeh et al. (2021))</i>	57
<i>Abbildung 40 - Darstellung der Abschätzung der Segmentierungsparametern für DOP mit der Varianz (rot) und der Veränderungsrate (blau)</i>	59
<i>Abbildung 41 - Darstellung der Abschätzung der Segmentierungsparametern für PlanetScope-Daten mit der Varianz (rot) und der Veränderungsrate (blau)</i>	60
<i>Abbildung 42- Beispiel der Segmentierung mit MRS und einem SP von 10 für DOP (a-d) und 7 für PlanetScope (e-h)</i>	61
<i>Abbildung 43 - Darstellung der DOP Index-Berechnung mit a) nDSM b) NDVI c) NDWI und d) NDSI und SAT-Index-Berechnung e) NDRR f) SI und g) WV-BI</i>	64
<i>Abbildung 44 – Überblick der Darstellung der Heatmaps nach der Berechnung des trainierten CNN für DOP für die 3 Klassen a) Gebäudeobjekte b) Kleine Objekte c) Nicht-Gebäude. Helle Pixel entsprechen dabei jeweils einer hohen Zugehörigkeitswahrscheinlichkeit.</i>	66
<i>Abbildung 45 - Darstellung CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen) im Vergleich zur Ground-Truth-Klasse Gebäudeobjekte (rote Gebäudeumringe)</i>	67
<i>Abbildung 46 - Darstellung der CNN-Klassifizierung der Zielklasse "Kleine Objekte" (lila Flächen) im Vergleich zur Ground-Truth-Klasse (rote Gebäudeumringe)</i>	68
<i>Abbildung 47 - Darstellung der CNN-Klassifikation der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Flächen) im Vergleich zu den GroundTruth-Klassen (rote Umringe)</i>	68
<i>Abbildung 48 - Darstellung der Erkennungsqualität für das Neubaugebiet (a-c) und als Vergrößerung der drei Klassen d) Gebäudeobjekt e) kleine Objekte f) Nicht-Gebäude</i>	69
<i>Abbildung 49 - Überblick der Darstellung der Heatmaps nach Anpassung der Trainingsparameter für DOP (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit für die 3 Klassen a) Gebäudeobjekte b) Kleine Objekte c) Nicht-Gebäude</i>	71
<i>Abbildung 50 – Schwierigkeiten der Differenzierbarkeit der Klasse "Kleine Objekte". a) eindeutige Charakteristik am Beispiel Garagenreihe b) nicht eindeutige Charakteristik am Beispiel Anbauten und Carports</i>	73
<i>Abbildung 51 - Überblick der Darstellung der Heatmaps nach Anwendung des trainierten CNN für DOP für 2 Klassen a) Gebäudeobjekte b) Nicht-Gebäude c) - d) Vergrößerung der Zugehörigkeitswahrscheinlichkeit zu „Gebäudeobjekte“</i>	74
<i>Abbildung 52 – Beispiel 1: CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche) im Vergleich zur Ground-Truth-Klasse „Gebäudeobjekte“ (rote Gebäudeumringe) und „Nicht-Gebäude“ (rote regelmäßige Quadrate)</i>	74
<i>Abbildung 53 – Beispiel 2: Darstellung CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche) im Vergleich zur Ground-Truth-Klasse „Gebäudeobjekte“ (rote Gebäudeumringe) und „Nicht-Gebäude“ (rote regelmäßige Quadrate)</i> ...75	75
<i>Abbildung 54 - Überblick der Darstellung der Heatmaps nach der Berechnung des trainierten CNN für DOP mit angepassten Parametern für die 2 Klassen a) „Gebäudeobjekte“ b) „Nicht-Gebäude“</i>	77
<i>Abbildung 55 - Beispiel der CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen a-c) und der Nicht-Zielklasse „Nicht-Gebäude“ (graue Fläche d-f) im Vergleich zur Ground-Truth-Klasse (rote Umringe)</i>	78

<i>Abbildung 56 – Klassifikation der Gebäudeobjekte (blau) mit den Kontrolldaten der Zielklasse (rote Gebäudeumrandung) und Nicht-Gebäude (rote quadratische Flächen) für das Neubaugebiet in Weinsberg (a) und der Vergrößerung der Darstellung mit dargestellten Segmentierungsobjekten (b – d)</i>	79
<i>Abbildung 57 – Satellitendaten: Überblick der Darstellung der Heatmaps nach Anwenden des trainierten CNN mit Standardparametern (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit) für die 2 Klassen a) Gebäudeobjekte b) Nicht-Gebäude c) und d) Vergrößerung der Zugehörigkeitswahrscheinlichkeit zur Klasse „Gebäudeobjekte“</i>	82
<i>Abbildung 58 - Beispiele aus der CNN-Klassifizierung der Satellitendaten für die Zielklasse "Gebäudeobjekte" (blaue Flächen – linke Spalte), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche – rechte Spalte) im Vergleich zur Ground-Truth-Klasse Gebäudeobjekte (rote Gebäudeumringe) und Nicht-Gebäude (rote regelmäßige Quadrate)</i>	83
<i>Abbildung 59 – Satellitendaten: Überblick der Darstellung der Heatmaps nach der Berechnung des trainierten CNN mit angepassten Parametern (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit der 2 Klassen) a) Gebäudeobjekte b) Nicht-Gebäude c) und d) Vergrößerung der Wahrscheinlichkeit „Gebäudeobjekte“</i>	84
<i>Abbildung 60 - Beispiele aus der CNN-Klassifizierung der Satellitendaten für die Zielklasse "Gebäudeobjekte" (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche) im Vergleich zur Ground-Truth-Klasse Gebäudeobjekte (rote Gebäudeumringe) und Nicht-Gebäude (rote regelmäßige Quadrate)</i>	86
<i>Abbildung 61 - Vergleich der Segmentierung der Gebäudeobjekte vor (linke Seite) und nach (rechte Seite) der OBIA-Verfeinerung für DOP</i>	89
<i>Abbildung 62 - Vergleich der Segmentierung der Gebäudeobjekte vor (linke Spalte) und nach (recht Spalte) der OBIA-Verfeinerung für SAT</i>	90

Tabellenverzeichnis

<i>Tabelle 1 - Vergleich des Aufnahmedatums und -zeitpunktes der zu untersuchenden Rasterdatensätze</i>	11
<i>Tabelle 2 - Prozentuale Veränderung der Gebäudeobjekte für BW im Vergleich zum Vorjahr</i>	12
<i>Tabelle 3 - Prozentuale Veränderung der Gebäudeobjekte innerhalb SK/LK Heilbronn</i>	13
<i>Tabelle 4 - Gebäudeobjekte aufgeschlüsselt nach den 10 am meisten vorkommenden Gebäudefunktion im SK/LK Heilbronn</i>	14
<i>Tabelle 5 - Datenformate Digitale Orthophotos</i>	17
<i>Tabelle 6 - Dateninformationen PlanetScope (SuperDove)</i>	23
<i>Tabelle 7 - Bandbreite der Spektralkanäle der PlanetScope-Aufnahmen</i>	23
<i>Tabelle 8 - Schematische Darstellung einer Konfusionsmatrix</i>	35
<i>Tabelle 9 - Anzahl der noch zur Verfügung stehenden Gebäudeobjekte im Trainingsgebiet nach jedem Bereinigungsschritt</i>	43
<i>Tabelle 10 - Auflistung der fünf meistvorkommenden Gebäudeobjekte nach Objektart innerhalb des Untersuchungsgebietes</i>	46
<i>Tabelle 11 - Darstellung der zur Verfügung stehenden Samples im Trainingsgebiet für DOP und SAT (gerundet)</i>	47
<i>Tabelle 12 - Vergleich der Samples aus dem DOP- und dem SAT-Datensatz</i>	48
<i>Tabelle 13 - Anzahl der verwendeten Training Samples bevor und nach der Data Augmentation für DOP und SAT mit dem Transformationsmultiplikator</i>	53
<i>Tabelle 14 - Vergleich der Kennzahlen der Training Samples für DOP- und SAT-Daten</i>	53
<i>Tabelle 15 - Anzahl der GT-Daten pro Klasse</i>	55
<i>Tabelle 16 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 3-Klassen</i>	70
<i>Tabelle 17 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 3-Klassen (angepasste Trainingsparameter)</i>	72
<i>Tabelle 18 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen und den laufzeitoptimierten Parametern</i>	76
<i>Tabelle 19 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen (qualitätsoptimiert)</i>	80
<i>Tabelle 20 - Zusammenfassung der Auswartergebnisse für die Ableitung von Gebäudeobjekten aus DOP</i>	80
<i>Tabelle 21 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten mit 2-Klassen und den laufzeitoptimierten Parametern</i>	83
<i>Tabelle 22 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten der 2-Klassen mit qualitätsoptimierten Parametern</i>	86
<i>Tabelle 23 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen nach wissensbasierter Bildanalyse</i>	89
<i>Tabelle 24 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten mit 2-Klassen nach wissensbasierten Bildanalyse</i>	91

Inhaltsverzeichnis

Eigenständigkeitserklärung	II
Danksagung	III
Zusammenfassung	IV
Abstract	VII
Abkürzungsverzeichnis	X
Abbildungsverzeichnis	XII
Tabellenverzeichnis	XV
Inhaltsverzeichnis	XVI
1 Einleitung	1
1.1 Ausgangslage und Motivation.....	2
1.2 Stand der Forschung	4
1.3 Zielsetzung und Gliederung.....	6
2 Datengrundlage	8
2.1 Interessensgebiet.....	8
2.2 Daten.....	11
2.2.1 Gebäudedatenbestand aus ALKIS	11
2.2.2 Digitale Orthophotos	16
2.2.3 Satellitendaten	21
3 Methodische Grundlagen	25
3.1 Objektbasierte Herangehensweise	25
3.2 Deep Learning-Ansatz.....	27
3.3 Bildsegmentierung.....	32
3.4 Bewertungsmetrik	34
4 Praktische Umsetzung	37
4.1 Auswahl Trainings- und Kontrollgebiet	38
4.1.1 Trainingsgebiet.....	38
4.1.2 Kontrollgebiet	40
4.2 Bereinigung der Referenz- und Kontrolldaten	41
4.3 Aufbau der Trainingsdaten	44
4.4 Kontrolldaten (Ground Truth).....	53
4.5 CNN-Architektur	56
4.6 Segmentierung und Klassifikation.....	58
4.6.1 Abschätzung der Segmentierungsparameter für DOP und SAT	58
4.6.2 Wissensbasierte OBIA-Verfeinerung.....	61
5 Ergebnisse	65

5.1	Evaluierung der Klassifikationsgüte für DOP-Daten.....	65
5.1.1	Klassifikation mit 3-Klassen.....	65
5.1.2	Binäre Klassifikation.....	73
5.2	Evaluierung der Klassifikationsgüte für <i>PlanetScope</i> -Daten	81
5.3	OBIA-Verfeinerung der Klassifikation	87
5.3.1	Wissensbasierte Bildanalyse der TrueDOP-Daten	87
5.3.2	Wissensbasierte Bildanalyse der <i>PlanetScope</i> -Daten.....	89
6	Diskussion	92
7	Schlussfolgerung und Ausblick	96
	Literaturverzeichnis	XCIX
	Glossar	CXII
	Anhang	CXIII

1 Einleitung

Künstliche topographische Objekte, wie Gebäude und Bauwerke, gehören zu den wahrnehmbarsten Elementen der Landschaft. Gemeinsam mit den natürlichen Objekten, wie beispielsweise der Vegetation oder den Gewässern, bilden sie einen Teil der Landbedeckungsklassen, die die Erdoberfläche physisch und biologisch bedecken (AdV-Projektgruppe LB/LN 2020).

Durch das Bevölkerungswachstum und die wirtschaftliche, sowie industrielle Entwicklung, insbesondere in Industrieländern, ist die Zahl der Veränderungen in der Landbedeckung/Landnutzung (LB/LN) gestiegen (Talukdar et al. 2020). Kenntnisse über LB/LN sind für eine Reihe von Fachbereichen, die auf der Verwendung von Erdberechnungen beruhen, essentiell (Talukdar et al. 2020). Dazu zählen etwa Stadt- und Regionalplanung, Folgenabschätzung und Umweltauswirkungen, sowie Umwelt- oder Katastrophenmonitoring (Hashem und Balakrishnan 2015; Nguyen und Liou 2019; Dao und Liou 2015).

Eine automatische Erkennung bzw. ein hoher Automatisierungsgrad zur Ableitung dieser topographischen Objekte mithilfe von Fernerkundungsdaten ist seit vielen Jahren ein fester Bestandteil der Geowissenschaften und das ausgesprochene Ziel der Vermessungsverwaltungen (AdV-Projektgruppe LB/LN 2020). Eine besonders herausfordernde Aufgabe ist die automatische Erkennung und präzise Lokalisierung der Gebäudeobjekte (Vakalopoulou et al. 2015; Alidoost und Arefi 2018).

Die Möglichkeit die Erdoberfläche mithilfe von Fernerkundungsaufnahmen, wie Luftbildern oder Satellitenaufnahmen auszuwerten, hat zu einer erheblichen Kosten- und Zeitreduktion im Vergleich zu herkömmlichen Felduntersuchungen geführt (Carraza-García et al. 2019). Durch rasante Entwicklungen in der Luft- und Raumfahrt, sowie der Sensor- und Computertechnik in den letzten Jahrzehnten, werden sehr hochauflösende Fernerkundungsaufnahmen zunehmend als Basis zur Klassifizierung der Landschaftselemente genutzt (Hu et al. 2015).

Viele gesetzlich vorgeschriebene Verfahrensabläufe bei der Gebäudeeinemessung, wie beispielsweise die Ankündigung von Vermessungsarbeiten, die Anfertigung des Fortführungsrisse und der Fertigungsaussage oder die Einreichung der Vermessungsschrift (Kriesten 2017), sind mit einem hohen Personal- und Zeitaufwand verbunden, und folglich kostenintensiv. Eine Möglichkeit der Reduktion weiterer Zeitaufwände besteht in einer Automatisierung der Detektion und Feststellung von potenziell einmessungspflichtigen Objekten des Liegenschaftskatasters auf der Erdoberfläche. Der technologische Wandel der Digitalisierung rückt unter anderem die Frage in den Vordergrund, ob eine Anpassung teils manueller, nicht gesetzlich vorgeschriebener Arbeitsabläufe für eine valide und gesicherte Erkennung von Gebäudeobjekten ermöglicht werden kann, welche bisher auf einen hohen Personalaufwand vertraute.

Zusätzlich dazu, ermöglichen die Veränderungen der letzten Jahrzehnte im Bereich der Luftbildauswertung, die in die Entwicklung effizienter Algorithmen zur automatischen oder halbautomatischen Erkennung und Extraktion von Objekten geflossen sind, neue Strategien und Methoden, um die herausfordernde Aufgabe der Gebäudeerkennung zu meistern (Cui et al. 2008).

Die Gebäudeerkennung und -extraktion aus Luftbildern oder Satellitenaufnahmen ist seit vielen Jahren ein vielbeachtetes Aufgabenfeld, mit einer großen Auswahl an verschiedenen Methoden (Feng und Zhao 2009; Ghanea et al. 2016).

Eine dieser Methoden ist die objektbasierte Bildanalyse (object-based image analysis (OBIA)), die sich durch viele Anwendungen in den letzten Jahren bewährte (Blaschke et al. 2014b). Die objektbasierten Analysemethoden können die Genauigkeit der Landbedeckungsklassifizierung bei hoch bis sehr hochauflösenden Aufnahmen wirksam verbessern (Zhang et al. 2020). Grundsätzlich lässt sich OBIA als Methode verstehen, ein segmentiertes Objekt einer bestimmten Kategorie zuzuweisen, die bei der Objekterkennung Anwendung findet, um ein oder mehrere bestimmte Objekte der Erdoberfläche (künstliche und natürliche) in Fernerkundungsaufnahmen zu lokalisieren und entsprechend vorherzusagen (Audebert et al. 2017).

In den letzten Jahren sind besonders die Deep Learning-Algorithmen hervorzuheben, die bemerkenswerte Ergebnisse, nicht nur bei der Gebäudeerkennung, geliefert haben (Luo et al. 2021). Trotz der vielversprechenden Ergebnisse bleibt allerdings noch Raum für Verbesserung, etwa im Bereich der Nutzung von objektbasierter, anstelle der pixelbasierter Klassifikation (Tavakkoli Piralilou et al. 2019).

1.1 Ausgangslage und Motivation

In der Landesvermessung in Baden-Württemberg (BW) sollen Ansätze erarbeitet werden, die es ermöglichen, mit einem geringen Personalaufwand Veränderungen der Landschaft zu erkennen. In diesem Zusammenhang sind auch die Gebäudeobjekte zu nennen, da diese einer stetigen Veränderung unterliegen. Als Teil des Liegenschaftskatasters werden Gebäudeobjekte, wie Wohnhäuser, Garagen oder Überdachungen, in den von den Vermessungs- und Katasterverwaltungen geführten öffentlichen Register, landesweit nachgewiesen, dargestellt und beschrieben (AdV 2022). Neben den Flurstücken sind die Gebäudeobjekte ein Produkt des Liegenschaftskatasters und finden sich zusammengeführt im Amtlichen Liegenschaftskatasterinformationssystem (ALKIS). Diese Daten bilden die Grundlage für die Geodateninfrastruktur (GDI) in Deutschland. Bei den ALKIS-Gebäudedaten handelt es sich um einen dynamischen Datenbestand, der in BW jährlich um circa 50.000 Objekte¹ wächst. Diese hohe Änderungsrate erfor-

¹ Eigene Recherche durch ALKIS-Abrufe für BW der Jahre 2019, 2020 und 2021

dert für die Fortführung und Dokumentation des Liegenschaftskatasters einen hohen Personalaufwand der zuständigen Katasterverwaltungen.

Dabei wird die Erfassung der Änderungen von den unteren Vermessungsbehörden, sowie von Öffentlich bestellten Vermessungsingenieuren (ÖbVI) durchgeführt (LGL BW 2022d). Das Landesamt für Geoinformation und Landentwicklung (LGL BW) ist verantwortlich, die hohen und landesweiten Qualitätsstandards für das Liegenschaftskataster sicherzustellen, sowie für die Einhaltung der fachspezifischen Vorschriften bei der Führung des Liegenschaftskatasters (LGL BW 2022d).

Eine weitere amtliche Aufgabe des LGL BW ist die turnusmäßige multispektrale Befliegung der Landesfläche von BW. Aktuell wird dabei jährlich ein Drittel der Landesfläche von BW mit Luftbildern in einer Bodenauflösung von 20 cm mit belaubter Vegetation abgedeckt. Diese hochauflösenden multispektralen Aufnahmen verfügen über eine große Menge an Bodeninformationen auf Pixelebene und werden in verschiedenartigen Anwendungsbereichen der Endnutzer verwendet, wie etwa in der Stadtkartierung, der Landwirtschaft oder für das Waldmonitoring (Zhang et al. 2018). Die aus den Luftbildern abgeleiteten Digitalen Orthophotos (DOP) der flugzeuggestützten Befliegung zählen zu den Standardprodukten der Fernerkundung in der Vermessungsverwaltung und dienen als Grundlage für weitere Produkte, wie beispielsweise dem Digitalen Landschaftsmodell (DLM) (LGL BW 2022).

Zusätzlich zu den flugzeuggestützten Standardprodukten untersucht das LGL BW auch die Potenziale weiterer Fernerkundungsdaten. Die raschen Entwicklungen in der satellitengestützten Erdbeobachtung haben die Verfügbarkeit von hochauflösenden Aufnahmen von Satellitendaten erhöht und für viele Bereiche verfügbar gemacht und innovative Kooperationen hervorgebracht (BMI 2017). In jüngster Vergangenheit hat vor allem das europäische Copernicus-Programm mit seinen Sentinel-Satelliten die Aufmerksamkeit der Verwaltungen auf die Daten der Erdbeobachtungssatelliten gelenkt. Diese öffentlichen, offenen und multitemporalen Bild- und Radardaten ermöglichen es, zusätzlich zu der kleinräumigen Datenebene der *Unmanned Aerial Vehicle* (UAV) und der mittlräumigen Ebene durch die flugzeuggestützte Erfassung, nun auch eine weitere Datenebene der großflächigen, landkreisübergreifenden Aufnahmen abzudecken und nutzbar zu machen. Im Rahmen der Digitalisierungsstrategie *digital@bw* mit dem Projekt „Kompetenzfeld Satellitendaten“ hat es sich das Land BW zur Aufgabe gemacht, die Potenziale und die Nutzbarmachung von kostenfreien und -pflichtigen Satellitendaten unterschiedlicher Auflösung für die öffentliche Verwaltung zu eruieren und dadurch einer breiten Nutzergruppe wie Institutionen, Bürgern oder der Wissenschaft das Nutzungspotenzial für eine Vielzahl von Anwendungsfeldern aufzuzeigen (Wild-Pfeiffer et al. 2019). Die Spannweite der geometrischen Auflösung von Satellitenmissionen reicht hierbei von einer mittleren Auflösung im Bereich von 10 m bis zu sehr hochauflösenden Aufnahmen weniger Dezimeter (ESA 2022).

Den Rasterdatensätzen der flugzeuggestützten Befliegung stehen somit die Daten der satellitengestützten Trägersysteme gegenüber. Dabei spielen die Daten kommerzieller Anbieter mit Aufnahmen im Bereich von 1 bis 3 m eine hervorzuhebende Rolle, da hier eine Kombination aus hoher Aufnahmequalität, großflächiger Erfassung und einem moderaten Kostenfaktor zu erwarten ist.

All diese technologischen Entwicklungen in der Computer Vision und die großen Erfolge von Deep Learning in der Bilderkennung, sowie die große Verfügbarkeit von sehr hochauflösenden Bildaufnahmen im Dezimeterbereich bei flugzeug- und satellitengestützten Trägersystemen ermöglichen neue Herangehensweisen für bisher konventionelle Melde- und Erfassungsmethoden in der öffentlichen Verwaltung.

Auf Basis dieser Ausgangslage soll in Baden-Württemberg mithilfe von Deep Learning mit den Rasterdaten aus den flugzeug- und satellitengestützten Trägersystemen die Möglichkeit der Bestimmung und Ergänzung fehlender, also nicht erfasster Objekte des amtlichen Gebäudedatenbestandes eruiert werden und eine Aussage darüber getroffen werden, welche der beiden zugrundeliegenden Rasterdatensätze eine höhere Erkennungsqualität bieten.

1.2 Stand der Forschung

Das Thema der Gebäudeerkennung und -extraktion weist umfangreiche Literatur und eine Vielzahl an verschiedenen Ansätzen und Möglichkeiten auf, um Gebäude und andere künstliche Objekte aus Fernerkundungsdaten abzuleiten. Heutzutage sind sehr hochauflösende Aufnahmen aufgrund der schnellen Entwicklung fortschrittlicher Technologien in der Fernerkundung im großen Umfang verfügbar (Zhang et al. 2018). In den letzten zwei Jahrzehnten wurden hochauflösende flugzeug- und satellitengestützte Aufnahmen etwa für Anwendungen mit Bezug zu statischer Landbedeckung (Walton et al. 2008; Lang et al. 2008) oder der automatisierten Extraktion von Gebäudeumrissen (Jin und Davis 2005) verwendet. Trotz der vielfältigen Ansätze zur automatischen Detektion und Extraktion von Gebäudeobjekten und ihrer Lage aus Fernerkundungsdaten handelt es sich um eine herausfordernde Aufgabe (Alidoost und Arefi 2018).

Die Wissenschaft beschäftigt sich seit längerer Zeit mit der Detektion von Gebäuden und anderen künstlichen Objekten mithilfe von passiven Sensoren, wie Rasterdaten aus Luft- oder Satellitenbildern; dabei wurden Einzelaufnahmen (Lin und Nevatia 1998), multiple Aufnahmen (Fradkin et al. 2001) oder Objektschatten (Sirmacek und Unsalan 2008) verwendet. Weitere Entwicklungen ermöglichen Objektdetektionen durch die Anwendung von aktiven Sensoren wie Radar- (Ferro et al. 2013) oder Laserdaten (Verma et al. 2006). Auch Publikationen, die hochauflösende Bilddaten mit Hilfsdaten wie Digitalen Höhenmodellen kombinieren (Qin und Fang 2014), oder aus-

schließlich auf Basis von Digitalen Höhenmodellen ableiten (Zhou et al. 2013; Weidner und Förstner 1995), zeigen Erfolge.

Dabei nutzen die Verfasser beispielsweise Methodiken der unüberwachten Klassifikation, wie den *fuzzy K-means*-Algorithmus (Bezdek 1981), um aus hochauflösenden Satellitendaten und LiDAR-Daten Gebäude und Bäume abzuleiten (Tamés-Noriega et al. 2015) oder Verfahren der multivarianten Statistik, wie die Hauptkomponentenanalyse (Aytekin et al. 2012).

Auch überwachte Klassifikationsverfahren, wie *Random Forest* (Breiman 2001), *Support Vector Machine* (Chang und Lin 2011; Cortes und Vapnik 1995) oder *Decision Tree* (Kamiński et al. 2018; Safavian und Landgrebe 1991) finden Anwendung in der Gebäudedetektion und -extraktion.

Hermosilla et al. (2011) vergleichen die automatische Erkennung eines schwellenbasierten Ansatzes mit dem einer überwachten Klassifizierung mithilfe eines Entscheidungsbaumes (*Decision Tree* (DT)) (Hermosilla et al. 2011).

Auch Bemühungen mit modernen Klassifikationsalgorithmen in Form von Neuronalen Netzen zur Ableitung von Objekten wurden unternommen, etwa mittels einer großen hierarchischen Bilddatenbank als Trainingsdatensatz für die Objekterkennung und Bildklassifikation (Deng et al. 2009). In jüngster Vergangenheit verfolgen viele Ansätze den Weg der modernen Klassifikationsalgorithmen wie *Convolutional Neural Networks* (CNN) (vgl. Kapitel 4.5). Dabei gibt es allgemeine Ansätze, wie von Krizhevsky et al. (2017), die Neuronale Netze mit 1.000 verschiedenen Klassen trainieren. Die Entwicklungen umfassen aber auch spezielle CNN-Ansätze, beispielsweise zur automatischen Gebäudeerkennung mit Rasterdaten aus Luftbildern (Yuan 2016), zur Gebäudeschadensermittlung nach Überflutungsereignissen (Fujita et al. 2017); zur Gebäudeerkennung aus Google Earth-Aufnahmen (Zhang et al. 2016) oder die Erkennung von 3-dimensionalen Gebäudemodellen durch die Kombination von Rasterdaten aus Luftbildern mit LiDAR für einen wissensbasierten Ansatz (Alidoost und Arefi 2016).

Werden ausschließlich Publikationen der letzten Jahre betrachtet, zeigt sich die automatische Gebäudeerkennung aus Fernerkundungsaufnahmen weiterhin als ein vielbeachteter Forschungsbereich. Insbesondere bei einer Detektion aus traditionellen Orthophotos (Chen et al. 2021; Shao et al. 2020) oder aus Satellitendaten (Hu et al. 2021; Ma et al. 2019), als Fusion von flugzeug- und satellitengestützten Aufnahmen (Ji et al. 2019) oder mit dem Fokus auf der Extraktion von genauen Gebäudeumringen (Shao et al. 2020; Zhu et al. 2021; Schuegraf und Bittner 2019). Die Publikationen zeigen vielversprechende Ergebnisse in der Erkennung von Gebäudeobjekten und der Extraktion von Gebäudeumringen, weisen aber auch darauf hin, dass es auch Herausforderungen

birgt. So bedeuten Gebäudeobjekte eine große Heterogenität innerhalb der Klasse und oftmals unklare Objektgrenzen (Ghorbanzadeh et al. 2021).

Viele der oben genannten Studien verwenden einen *Machine-Learning*-Ansatz und Nutzen für die Umsetzung Deep Learning-Techniken wie CNN. Heutzutage haben CNN und andere Deep Learning-Algorithmen bemerkenswerte Leistungen in der automatischen Erkennung von Objekten erreicht (Alidoost und Arefi 2018). Große Verbesserungen in der Genauigkeit bei der automatischen Gebäudeerkennung konnten mit CNN erreicht werden (Boonpook et al. 2018; Chen et al. 2020).

CNN basieren auf den biologischen mehrschichtigen Architekturen neuronaler Netze (Jin et al. 2019). Diese Netze bestehen aus mehreren nichtlinearen Schichten, deren Bearbeitungsweise als eine neue Methoden einer intelligenten Mustererkennung angesehen wird und eine wichtige neue Richtung für Bildverarbeitung in der Fernerkundung bedeuten (Zhao et al. 2015). Während CNN gemeinhin als leistungsfähige und vielversprechende Lösung für die Klassifizierung gesehen werden, gibt es verschiedene Lösungswege bei der Segmentierung (Längkvist et al. 2016). Bei der Klassifizierung aus Fernerkundungsaufnahmen besteht der hauptsächliche Prozess aus der Merkmalsextraktion und der Klassifizierung basierend auf Bildmerkmalen (Jin et al. 2019).

Norman et al. (2021) haben gezeigt, dass die Ergebnisse aus *Machine-Learning*-Algorithmen wie SVM und DT durch den nachgelagerten Einsatz von Klassifizierungstechniken wie OBIA verbessert werden. Längkvist et al. (2016) beschreiben, dass CNN für eine pixelweise Klassifikation verwendet und dann mithilfe von Segmentierungsverfahren verbessert werden können. Auch weitere Studien bestätigen es, so zeigt die Anwendung von OBIA in Niedrigflughöhen mit UAVs (Hairie et al. 2018) oder in der Detektion von Erdbeben (Ghorbanzadeh et al. 2019), dass die Ergebnisse verbessert, respektive die Deep Learning-Limitation entschärft werden können.

1.3 Zielsetzung und Gliederung

Der Fokus dieser Masterthesis liegt auf der Untersuchung verschiedener Rasterdatensätze, die als Basis für die Erkennung von Objekten des ALKIS-Gebäudedatenbestandes eingesetzt werden können. Die geometrisch hochauflösenden Digitalen Orthophotos im Qualitätsstandard TrueDOP werden den zeitlich besser aufgelösten Satellitendaten von *PlanetScope* gegenübergestellt und beide im Kontext der Fragestellung beurteilt. Das Ziel der Arbeit ist die Aussage hinsichtlich der Eignung der Datensätze für die Detektion von fehlenden, also nicht erfasster Gebäudeobjekte im ALKIS-Datenbestand. Als entscheidende Metriken zählen dabei neben der Genauigkeit der Erkennung auch Kriterien, wie die Laufzeiten oder der Speicherplatzbedarf.

Die Beantwortung der Frage ist von besonderem Interesse, weil eine automatisierte Gebäudedetektion mit einer hohen Genauigkeit in einem festgelegten Wiederkehrzyklus eine Zeitersparnis und Reduzierung des Personalaufwands bedeuten kann.

Die Arbeit gliedert sich in folgende Teilbereiche: (1) Der nachfolgende Abschnitt beschreibt die Auswahl des Interessensgebietes und die für die Thesis zur Verfügung stehenden Daten. (2) Darauf folgt ein Abschnitt zu den methodischen Grundlagen, dabei wird die objektbasierte Herangehensweise erläutert, gefolgt vom Deep Learning-Ansatz und der Bildsegmentierung. Abschließend beschreibt dieses Kapitel die Bewertungsmetrik zum Vergleich der Ergebnisse der untersuchten Daten. (3) Der Abschnitt „Praktische Umsetzung“ konzentriert sich auf die angewendeten Werkzeuge und den Workflow von der Auswahl des Trainings- und Kontrollgebietes, bis zum Aufbau des Trainingsdatensatzes und der Architektur des Neuronalen Netzes, hin zum bestmöglichen Modell zur objektbasierten Merkmalsextraktion von Gebäudeobjekten, jeweils an die untersuchten Rasterdaten angepasst. (4) Die Beschreibung und Darstellung der erreichten Ergebnisse der flugzeug- und satellitengestützten Daten folgt in Kapitel 5 und die Arbeit schließt mit (5) einer Ergebnisdiskussion und einem Ausblick.

2 Datengrundlage

2.1 Interessensgebiet

Das in dieser Masterarbeit behandelte Interessensgebiet ist der Landkreis Heilbronn. Der Landkreis (LK) befindet sich nördlich der Landeshauptstadt Stuttgart in BW und zählt rund 348.000 Einwohner (LRA Heilbronn 2021). Die Stadt Heilbronn liegt innerhalb des LK Heilbronn. Zusammen mit dem Stadtkreis (SK) umfasst die Fläche in Summe 1.200 km², circa 100 km² entfallen dabei auf den SK².

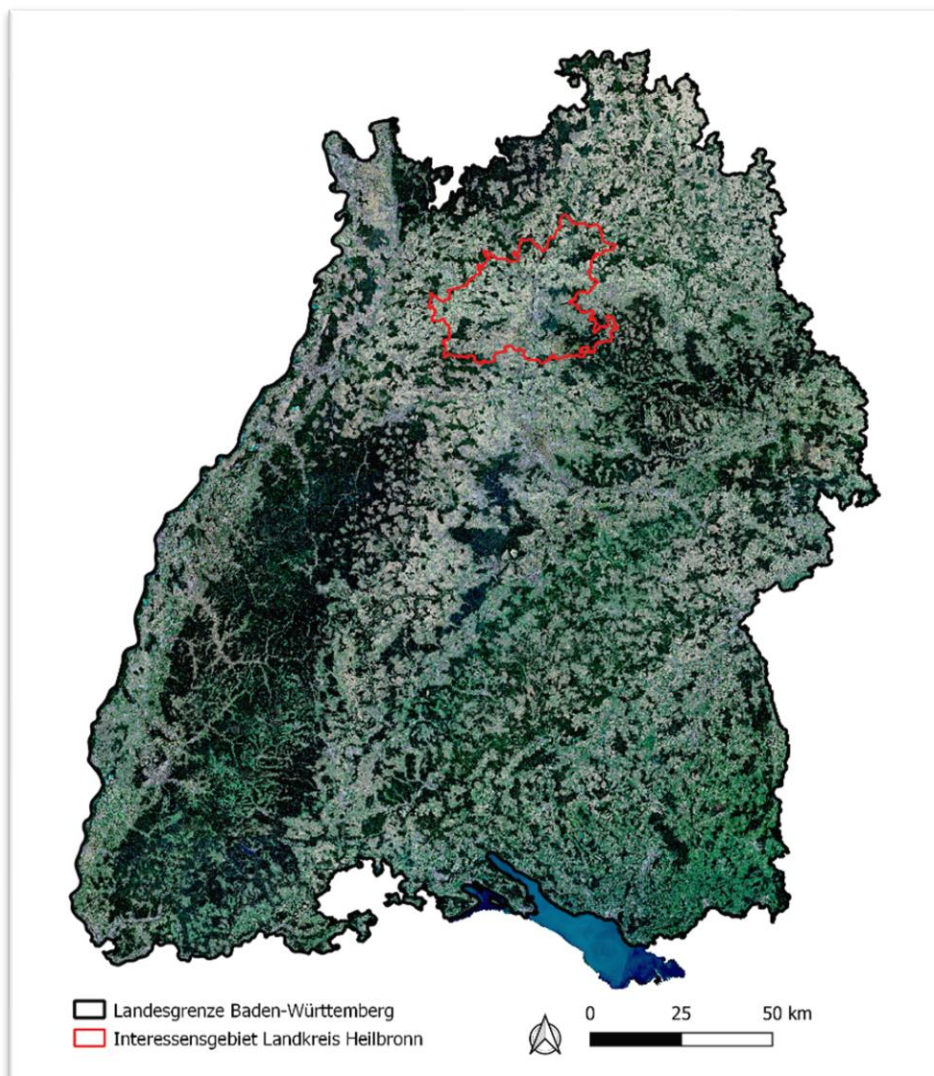


Abbildung 1 - Lage des Landkreises Heilbronn in BW

² eigene Recherche amtlicher Daten

Definition des Untersuchungsgebietes

Die Luftbildbefliegung in BW unterliegt aktuell einem zyklischen 3-Jahres-Rhythmus, das heißt jedes Jahr wird wechselnd ein Drittel der Landesfläche digital erfasst.

Damit das Ziel möglichst vergleichbarer Datensätze erreicht werden kann, ist es notwendig, dass der Bildflug und der Satellitenüberflug zeitlich möglichst eng zusammenliegen.

Im Sommer 2021 fand im Zuge der DOP20-Befliegung die Erfassung mit Luftbildern für einen Teil des Landkreises statt. Dabei wurden Teile der Landkreisfläche etappenweise an drei Tagen vom 31.05.21 bis 02.06.21 mit Luftbildern erfasst.

Die folgende Abbildung zeigt die Bildmitten des erfassten Befliegungsgebietes je Befliegungstag. Die räumlich ausgedehnteste Erfassung fand dabei am 01.06.21 (blaue Punkte) statt und deckte einen Großteil der Stadt Heilbronn (schwarze Umrandung innerhalb des LK), sowie den westlichen Teil des LK Heilbronn ab.

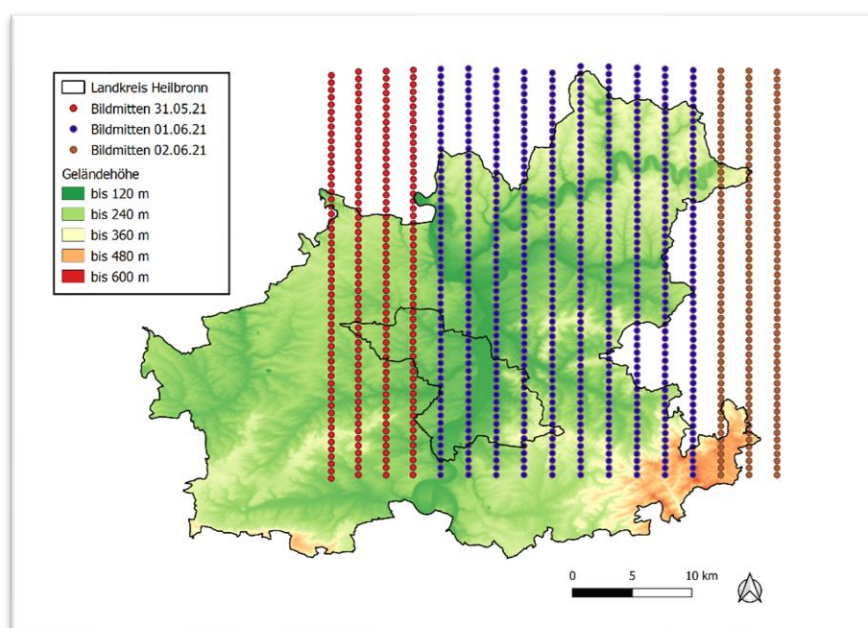


Abbildung 2: Darstellung der Bildmitten der DOP20-Befliegung nach Befliegungstagen für den Landkreis Heilbronn

Um eine möglichst hohe flächige Vergleichbarkeit der untersuchten Rasterdaten zu erreichen, orientierte sich die Suche der zeitlich naheliegenden Satellitenaufnahmen am Zeitpunkt des Bildfluges vom 01.06.21. Die untersuchten Satellitendaten von *PlanetScope* (Kap. 2.2.3) haben den Vorteil einer hohen Wiederkehrzeit von einem Tag und einem räumlich vergleichsweise breiten Aufnahmestreifen.

Am gleichen Tag (01.06.21) fand ein nahezu wolkenfreier Überflug der *PlanetScope*-Satelliten über dem Landkreis Heilbronn statt. Abbildung 3 zeigt den Überlappungsbereich der beiden untersuchten Rasterdaten: Zu sehen sind die beiden Erfassungsbereiche der Datensätze innerhalb der Landkreisfläche. Die Erfassungsbereiche umfassen

zusätzlich noch Teile außerhalb der Landkreisgrenze, die aber für die weitere Untersuchung nicht relevant sind. Die *PlanetScope*-Daten decken innerhalb des Landkreises eine Fläche von ungefähr 1.048 km² ab (blaue Schraffierung). Das entspricht einem Flächenanteil des SK/LK Heilbronn von fast 96 %. Die Luftbildbefliegung vom 01.06.21 umfasst innerhalb des Landkreises eine Fläche von 673 km², ein Anteil von 56% der Landkreisfläche (rote Schraffierung).

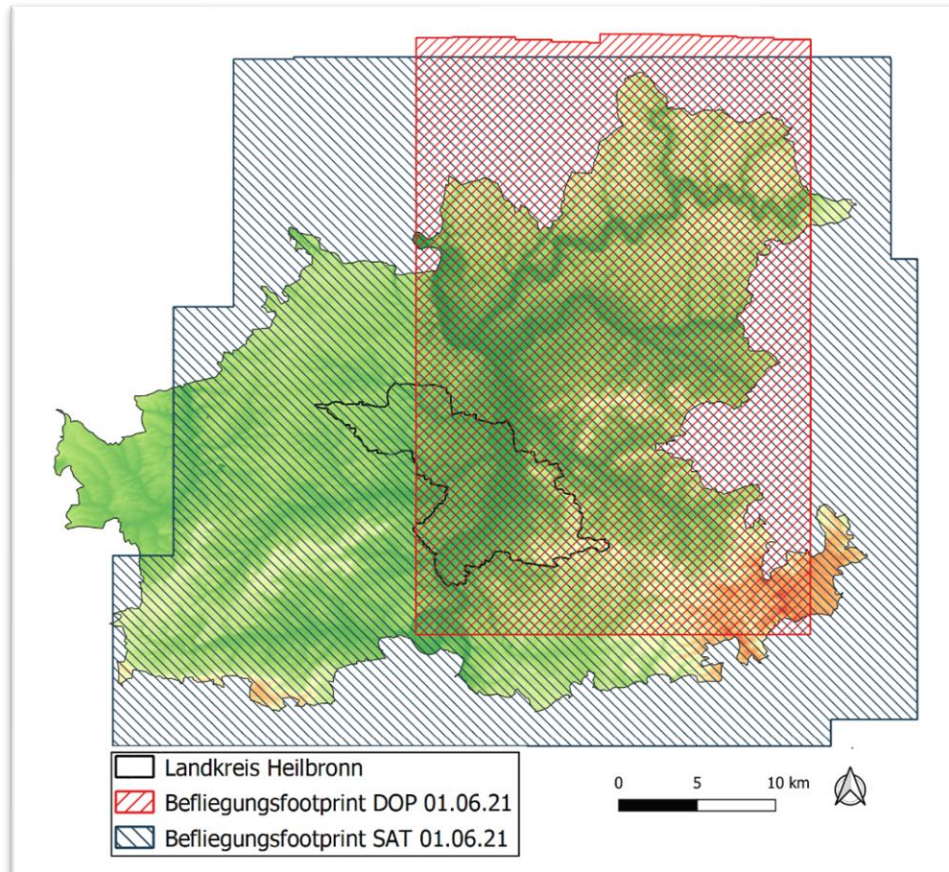


Abbildung 3: Erfassungsbereiche (Footprints) der flugzeug- und satellitengestützten Befliegungen

Somit ergibt sich aus der Überlappung der Abdeckung der Luftbildbefliegung und des Satellitenüberflugs die maximal mögliche Ausdehnungsfläche des Untersuchungsgebietes. Die Aufnahmefläche des Satellitenüberflugs beinhaltet die Befliegungsfläche der Luftbildbefliegung vom 01.06.21 vollständig (bis auf eine kleine Fläche im Norden außerhalb des Interessensgebietes). Somit existiert ein tagesgleicher Rasterdatensatz aus Satellitendaten und Luftbildern in der gemeinsamen Überlappungsfläche.

Tabelle 1 - Vergleich des Aufnahmedatums und -zeitpunktes der zu untersuchenden Rasterdatensätze

	flugzeuggestützt	satellitengestützt
Aufnahmedatum	01.06.21	01.06.21
Aufnahmezeitpunkt	13:20 – 16:00 Uhr	10:15 Uhr
abgedeckte Landkreisfläche (Interessensgebiet)	672,9 km ²	1.147,3 km ²

Die Fläche bildet das Untersuchungsgebiet für diese Thesis und beinhaltet insgesamt 168.446 von insgesamt 274.256 ALKIS-Gebäudeobjekten (2021) im LK Heilbronn³ (siehe Kapitel 2.2.1).

2.2 Daten

Das folgende Kapitel beschreibt die Vektordaten, die in der Untersuchung als Referenz- und Kontrolldaten zum Einsatz kommen und die Rasterdatensätze, die als Basis für die Klassifikation dienen.

2.2.1 Gebäudedatenbestand aus ALKIS

Für die Bearbeitung der Thesis standen die amtlichen Hausumringe aus ALKIS zur Verfügung. Dabei handelt es sich um Objekte mit georeferenzierten Polygonen, die die Gebäude- und Bauwerkgrundrisse des Liegenschaftskatasters beschreiben (LGL BW 2022c). Diese Daten wurden sowohl als Trainingsdaten für das CNN genutzt, als auch als Kontrolldaten, um die Erkennungsqualität zu verifizieren. Im Folgenden werden die Daten als Gebäudeobjekte bezeichnet.

Mit Beginn der Migration des Automatisierten Liegenschaftsbuches (ALB) und der Automatisierten Liegenschaftskarte (ALK) zu ALKIS Mitte des Jahres 2011 wurde die Basis für eine ganzheitliche und objektstrukturierte Modellierung für bundesweit einheitliche Geobasisdaten geschaffen (LGL BW 2011). Laut LGL BW (2011) waren die Harmonisierung der Daten von ALKIS und dem Amtlichen Topographisch-Kartographischen Informationssystem (ATKIS), die Integration der Sach- und Geometriedaten von ALB und ALK und die Verwendung von Standards und Normen (ISO, OGC) die Ziele der Migration.

³ Abruf der ALKIS-Daten (eigene Recherche)

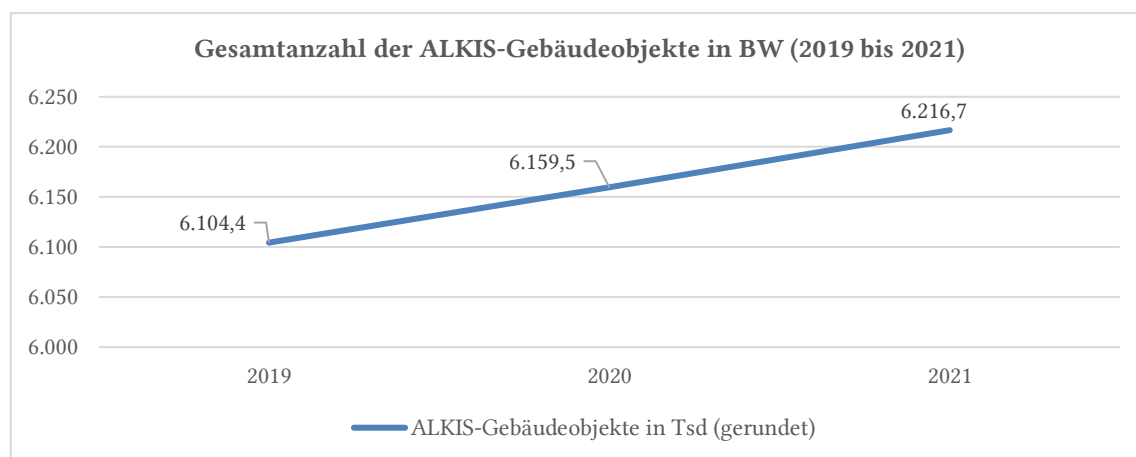


Abbildung 4: Zuwachs der ALKIS-Gebäudeobjekte in BW von 2019 bis 2021

Der Gebäudedatenbestand aus ALKIS umfasst für BW im Jahr 2021 insgesamt 6.216.727 unterschiedliche flächenhafte Gebäudeobjekte. Im Vergleich zum Jahr 2020 ist das ein Anstieg um 57.279 Objekte von damals 6.159.448 Gebäudeobjekten (siehe Abbildung 4). In den letzten drei Jahren ist ein stetiger Zuwachs der Gesamtanzahl der ALKIS-Gebäudeobjekte in BW von jährlich ungefähr 0,9 % zu beobachten.

Tabelle 2 - Prozentuale Veränderung der Gebäudeobjekte für BW im Vergleich zum Vorjahr

Jahr	Gebäudeobjekte	abs. Veränderung zum Vorjahr	% Veränderung zum Vorjahr
2019	6.104.395	k.A.	k.A.
2020	6.159.448	+55.053	0,9 %
2021	6.216.727	+57.279	0,9 %

Werden die Gebäudeobjekte isoliert für das Interessensgebiet betrachtet, zeigt sich ein vergleichbares Bild. Das bedeutet, dass es im SK/LK Heilbronn jährlich zu einem Anstieg in der Gesamtzahl der Gebäudeobjekte kommt. Laut dem Datenabruf für das Jahr 2021 beträgt die Gesamtanzahl der Gebäudeobjekte innerhalb der Landkreisfläche inklusive der Stadtfläche Heilbronn 274.256 Objekte (siehe Abbildung 5).

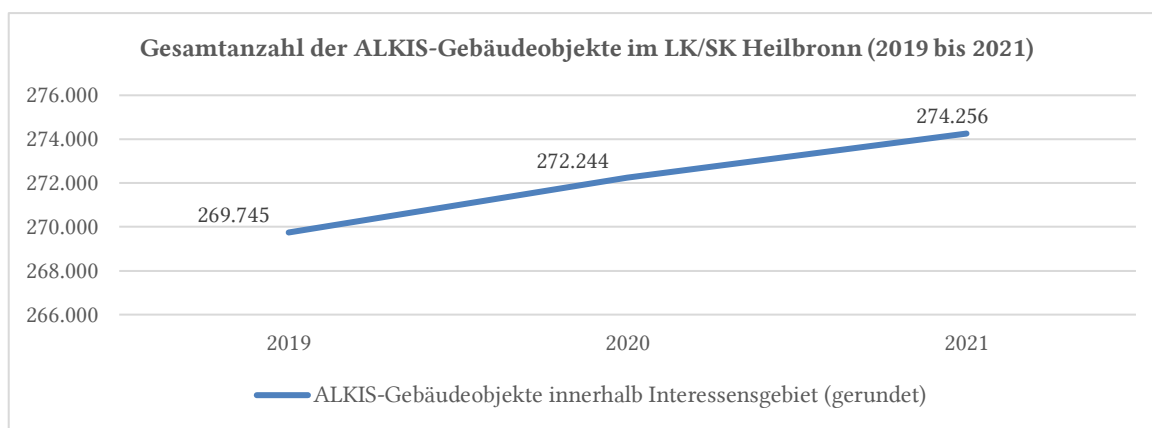


Abbildung 5: Zuwachs der ALKIS -Gebäudeobjekte im SK und LK Heilbronn zwischen 2019 und 2021

Abbildung 5 zeigt den Zuwachs der Gesamtzahl der ALKIS-Gebäudeobjekte für den LK und SK Heilbronn. Es zeigt sich eine Zunahme um 0,74 % im Vergleich von 2020 zu 2021, respektive um 0,93 % von 2019 zu 2020.

Tabelle 3 - Prozentuale Veränderung der Gebäudeobjekte innerhalb SK/LK Heilbronn

Jahr	Gebäudeobjekte	abs. Veränderung zum Vorjahr	% Veränderung zum Vorjahr
2019	269.742	k.A.	k.A.
2020	272.244	+2.502	0,93 %
2021	274.256	+2.021	0,74 %

Werden die Gebäudeobjekte aufgeschlüsselt nach ihrer Gebäudefunktion betrachtet (siehe Tabelle 4), zeigt sich, dass die Objektart *Wohnhaus* den größten Anteil im SK/LK-Heilbronn mit 112.844 Objekten besitzt. 2021 vergrößerte sich der Datensatz um 569 Objekte. Auf dem zweiten und dritten Platz befinden sich die Objekte *Garage* und *Schuppen*. Bei diesen beiden Objekten wuchs der Datensatz für das Jahr 2021 um 1.171 Objekte, respektive 17 Objekte.

Tabelle 4 - Gebäudeobjekte aufgeschlüsselt nach den 10 am meisten vorkommenden Gebäudefunktion im SK/LK Heilbronn

Gebäudfunktion (GFK)	2019	2020	2021
Wohnhaus (1010)	111.555	112.275 (+720)	112.844 (+569)
Garage (2463)	93.924	95.557 (+1.633)	96.728 (+1.171)
Schuppen (2723)	16.043	16.133 (+90)	16.150 (+17)
Überdachung (1610)	6.763	6.837 (+74)	6.932 (+95)
Wirtschaftsgebäude (2729)	6.751	6.781 (+30)	6.783 (+2)
Betriebsgebäude (2112)	5.931	5.956 (+25)	5.978 (+22)
Scheune (2721)	5.511	5.462 (-49)	5.407 (-55)
Wohn-/Geschäftshaus (1123)	3.585	3.594 (+9)	3.593 (-1)
Stall (2724)	2.840	2.761 (-79)	2.709 (-52)
Gebäude für Vorratshaltung (2140)	2.534	2.558 (+24)	2.583 (+25)

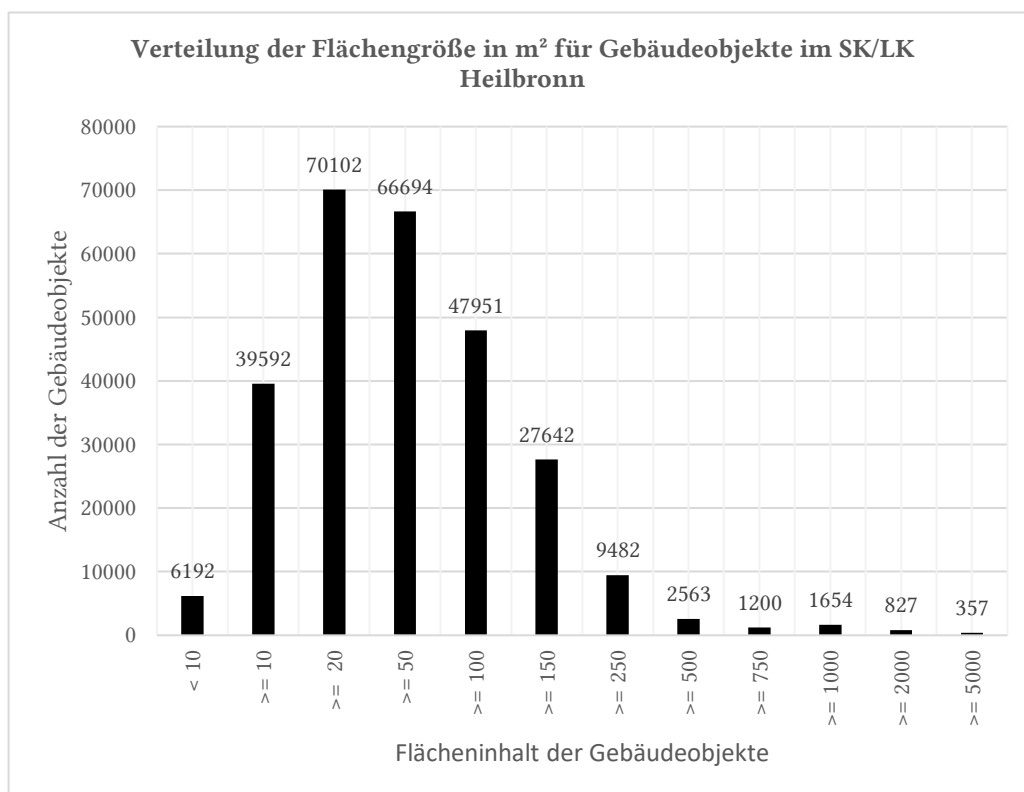


Abbildung 6 - Verteilung der Flächengrößen der Gebäudeobjekte im Stadt- und Landkreis Heilbronn

Wird die Verteilung der Gebäudeobjekte innerhalb des SK/LK Heilbronn betrachtet, zeigt sich, dass 6.192 Objekte einen Flächeninhalt von kleiner als 10 m² besitzen (vgl. Abbildung 6). Somit existiert ein beträchtlicher Anteil an Objekten innerhalb des Interessensgebietes, der im Satellitendatensatz (geometrische Auflösung: 3 m) nur durch 1x1 Pixel im Bild abgebildet wird.

Abbildung 6 zeigt weiter, dass eine hohe Zahl an Objekten kleiner als 20 m² vorhanden sind. Knapp 39.600 Objekte haben einen Flächeninhalt zwischen 10 m² bis unter 20 m². Das sind etwa 15 % der Gebäudeobjekte im Stadt- und Landkreis.

Das Untersuchungsgebiet, bestehend aus Trainings- und Kontrollgebiet, soll die vorherrschenden Charakteristiken und topographischen Gegebenheiten repräsentativ wiedergeben. Neben einer großen Anzahl an verschiedenartigen Gebäudeobjekten sollen auch sonstige landschaftliche Flächen beinhaltet sein. Das Stadtgebiet Heilbronn wird in dieser Thesis lediglich als Referenz- und Trainingsgebiet verwendet. Für das Verfahren zur Ableitung der Gebäudeobjekte wurden seitens des Vermessungsamtes im Landratsamt Heilbronn potenzielle Kontrollgebiete innerhalb des LK Heilbronn vorgeschlagen, welche eine besondere Bedeutung besitzen und zum Zeitpunkt der Datenerfassung eine präzente Veränderungsdynamik aufweisen.

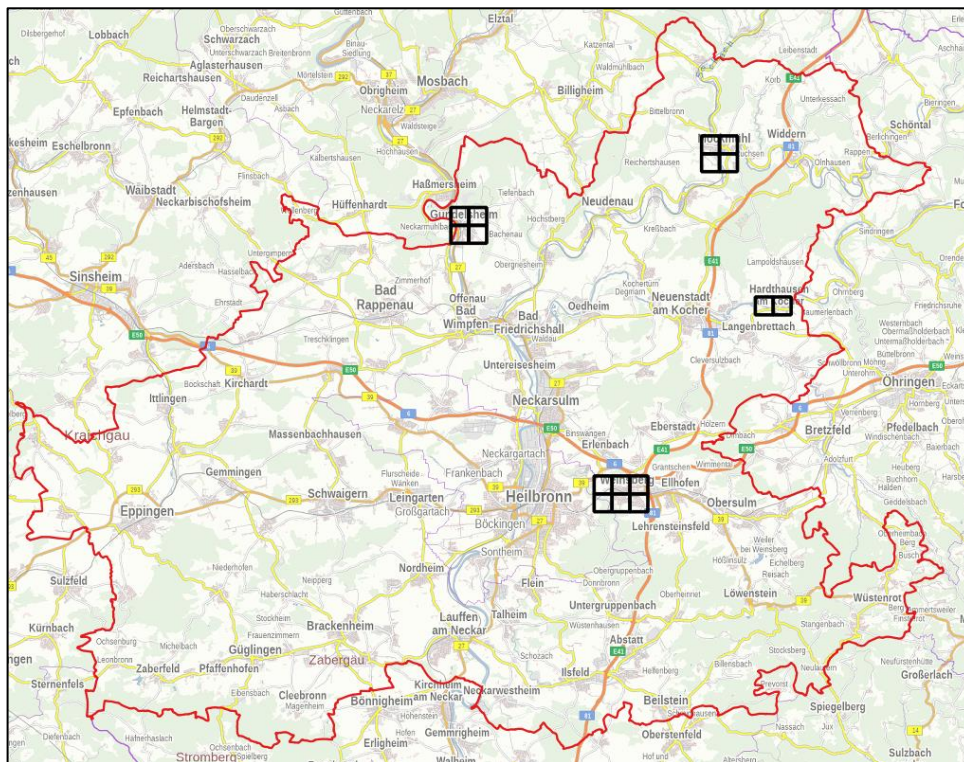


Abbildung 7 - potenzielle Kontrollflächen der Klassifikation innerhalb des Untersuchungsgebietes

Wie eingangs beschrieben, sind Gebäudeobjekte sehr unterschiedlich in der Ausprägung von Form, Größe und Farbe und weisen eine starke Heterogenität innerhalb der Klasse auf (Ghorbanzadeh et al. 2021). Abbildung 8 zeigt diese Unterschiedlichkeit der ALKIS-Gebäudeobjekte am Beispiel der hochauflösenden Orthophotos. Bildbeispiel a) zeigt gewerbliche Objekte innerhalb des Untersuchungsgebietes. Dabei handelt es sich hauptsächlich um großflächige Objekte. Bildbeispiel b) zeigt eine Mischbesiedlung aus kleinen bis mittelgroßen Objekten, sowie unterschiedlicher Dachfärbung. Bildbeispiel c) priorisiert vor allem Einzelhausbebauung mit Kleinstobjekten wie Garagen und Schuppen.



Abbildung 8 - Darstellung der Intra-Klassen-Heterogenität innerhalb der Klasse „Gebäudeobjekte“ a) gewerbliche Objekte b) Mischbebauung c) Einzelhausbebauung

2.2.2 Digitale Orthophotos

Bei den Digitalen Orthophotos (DOP) handelt es sich um verzerrungsfreie Rasterdaten, die maßstabstreu die Erdoberfläche in einer photographischen Abbildung darstellen (AdV 2019c). Laut dem Produktstandard der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) muss die Farbtiefe mindestens 8 Bit aufweisen. Die geometrische Genauigkeit bei DOP20 weist eine Standardabweichung der Lagekoordinaten von $\pm 0,4$ m auf. Die spektrale Ausprägung kann dabei von 1-Kanal panchromatisch über 3-Kanal-Echtfarbbilder bis zu 4-Kanal-Multispektralbildern (Rot/Grün/Blau/NIR) reichen (AdV 2019c).

Die DOP gehören zu den Standardbilddaten der amtlichen Landesvermessungen. Momentan umfasst die von LGL BW beauftragte Befliegung jährlich ein Drittel der Landesfläche von BW in einer Bodenauflösung von 20 cm mit einer Farbtiefe von 8 Bit und als 4-Kanal-Multispektralbild (LGL BW 2022b). Das entspricht einer Fläche von ungefähr $12.000 \text{ km}^2/\text{Jahr}$.

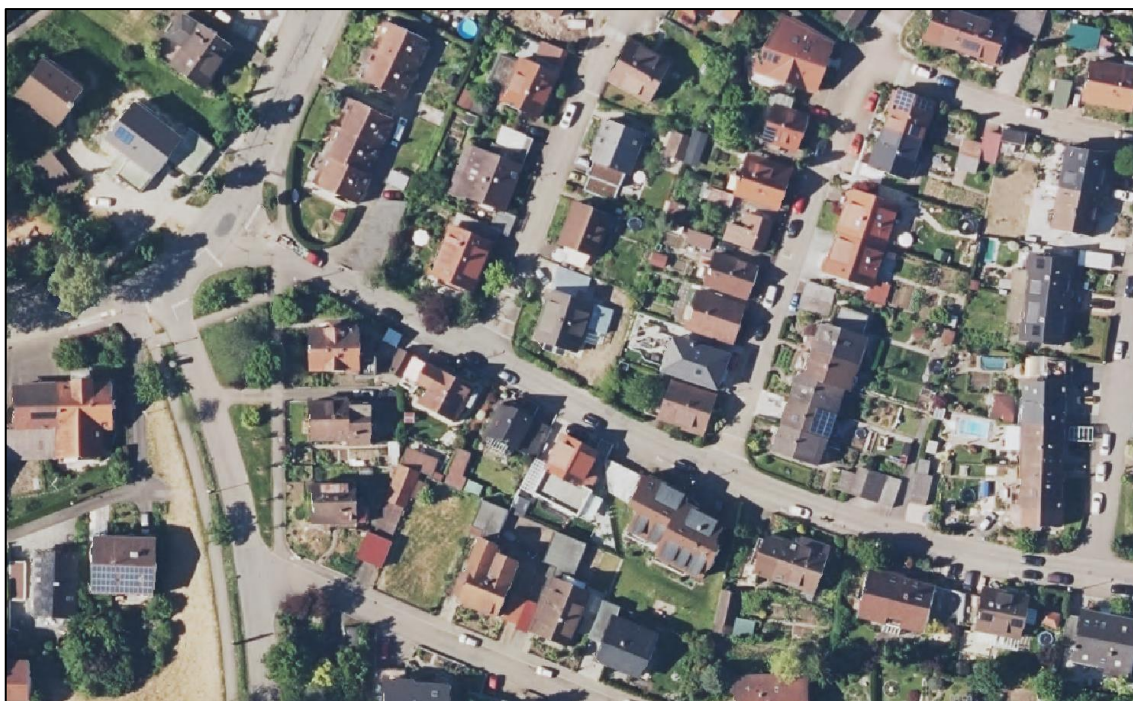


Abbildung 9 - Beispieldarstellung Digitales Orthophoto in Baden-Württemberg

Die DOP zeichnen sich besonders durch eine geringe atmosphärische Störung aus und werden in einer Flughöhe von etwa 4.000 Meter (DOP20) durchgeführt⁴. Da eine Flugfreigabe eines Befliegungsgebietes nur bei optimalen Wetterbedingungen erteilt wird, spielt die Problematik der Wolkenbedeckung für diesen Datensatz eine untergeordnete Rolle.

Tabelle 5 - Datenformate Digitale Orthophotos

Zusammenfassung DOP	
Spektralkanäle	Red, Green, Blue, NIR
Geometrische Auflösung	0,2 m
Farbtiefe	8 Bit
Standardabweichung	0,4 m
Georeferenzierung	ETRS89_UTM32
Datenformat	TIF

Der Qualitätsstandard „TrueDOP“

Im Jahr 2017 wurde von den Mitgliedsverwaltungen der AdV beschlossen, dass ein bundesweit flächendeckender Datenbestand in der Qualitätsstufe „TrueDOP“ bis spätestens Anfang 2023 vorliegen soll (AdV 2018). TrueDOP gelten als verzerrungs- und kippeffektfreie Rasterdaten, die maßstabstreu die Erdoberfläche und die darauf befindlichen Objekte in einer photographischen Abbildung darstellen (AdV 2019b).

⁴ interne Auskunft

Bei der klassischen Orthorektifizierung wird nur das Geländere relief entfernt, wodurch Objekte mit einer gewissen Höhe, die sich auf der Erdoberfläche befinden, wie etwa Gebäude oder Brücken, lagefalsch entzerrt werden (Chen et al. 2021). Im Unterschied zur klassischen Orthorektifizierung der DOP, für die die Ableitung neben den orientierten Luftbildern das Digitale Geländemodell (DGM) verwendet wird, kommt bei der Erstellung der TrueDOP das bildbasierte digitale Oberflächenmodell (bDOM), generiert durch das *Dense Image Matching* (DIM), zum Einsatz. Durch die Verwendung eines Oberflächenmodells kann der Verkippungseffekt von Objekten, die eine Höhe über Grund aufweisen, entfernt werden (AdV 2019c; AdV 2019b).

Da das Produkt TrueDOP originär auf den Luftbildern basiert, liegt die geometrische Auflösung ebenfalls bei 20 cm. Diese Daten sind in BW in einer Quadratmeter-Kachelung mit einer Auflösung von 5.000 x 5.000 Pixel verfügbar. Das Produkt hat laut Leistungsbeschreibung der AdV eine radiometrische Auflösung von mindestens 12 Bit/Kanal bei 4 Spektralkanälen (Red, Green, Blue, NIR) mit einer Mindestüberlappung in der Längsüberdeckung von mindestens 80 % und in der Querüberdeckung von mindestens 30 %, im urbanen Raum von 60 % (AdV 2019a).

Als Ergebnis der stattfindenden Befliegungskampagne des Landes BW werden die TrueDOP zurzeit in einem wiederholenden 3-Jahres-Zyklus erzeugt. Die Befliegung für das Interessensgebiet fand am 01.06.21 (vgl. Abschnitt 2.1) mit einer Längs- und Querüberlappung von 60/80 % statt. Die dabei entstandenen Luftbilder werden anschließend in mehreren Prozessierungsschritten mit der Software SURE der Fa. nFrames zu TrueDOP generiert (M. Rothermel et al. 2012). Die vorliegende Arbeit verwendet für die Untersuchung die 4-Kanal-Multispektralbilder (RGBI) mit einer Bodenauf- lösung von 20 cm.

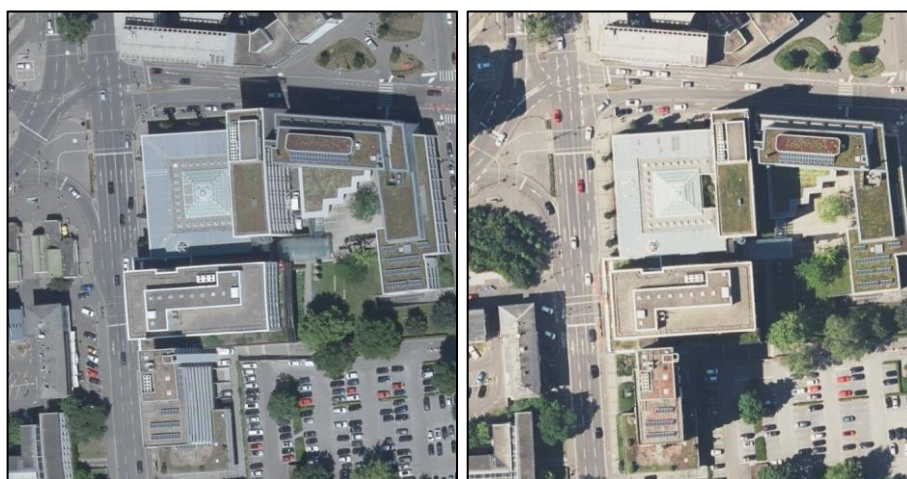


Abbildung 10 - Vergleich der Unterschiede des Verkippungseffektes bei DOP (links) und TrueDOP (rechts)

Abbildung 10 verdeutlicht die Unterschiede der Qualität bezüglich des Kippeffekts bei den traditionellen DOP und den TrueDOP. Auf der linken Aufnahme ist die Fassade des Hochhauses auf dem Orthophoto von 2018 durch den Kippeffekt zu erkennen,

wobei die TrueDOP-Aufnahme von 2021 eine Darstellung ohne Verkippungseffekt zeigt. Die TrueDOP-Aufnahmen erlauben die grundrisstreue Darstellung des Gebäudegrundrisses aus ALKIS und der darüberliegenden Dachfläche (Baltrusch 2016), wodurch eine Steigerung des Informationsgehalts erreicht wird, da sichttote Bereiche durch einen fehlenden Verkippungseffekt minimiert werden (GEObasis.nrw 2020). Untersuchungen, wie etwa von Chen et al. (2021) zeigen, dass die Extraktion von Gebäudeobjekten aus Digitalen Orthophotos oft durch die Herausforderungen des Kippeffektes erschwert wird und die umklappfreien TrueDOP-Aufnahmen einen optimaleren Datensatz für diese Aufgabenstellung darstellen.

Allerdings bringt die neue Qualitätstufe TrueDOP nicht nur die eben genannten Vorteile einer Reduktion des Umklappeffektes mit sich, sondern auch Nachteile durch eine geringe Kantenschärfe an Objekten, die durch einen Höhengsprung entsteht (AdV 2019b). Besonders deutlich ist dieser Umstand an Gebäudeobjekten zu erkennen, wie Abbildung 11 zeigt. Diese Ungenauigkeiten können durch die unzureichende Kantenschärfe des bDOM des DIM-Prozesses erklärt werden (AdV 2019b).

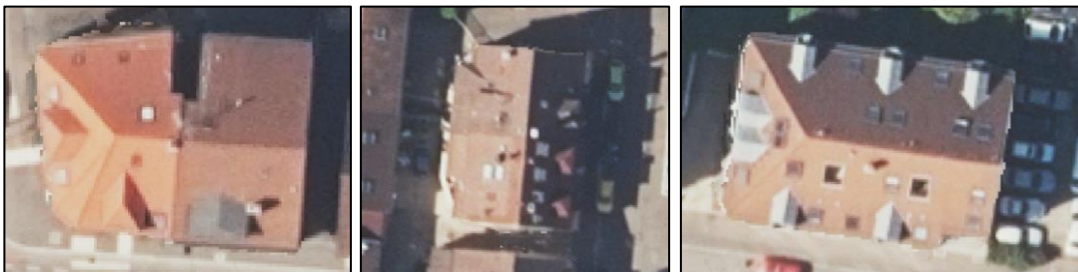


Abbildung 11 - Geringe Kantenschärfe an Dachtraufen bei TrueDOP

Zudem können weitere Nachteile auftreten, wie beispielsweise fehlende Informationen durch wiederholende Strukturen, etwa bei Ackerfurchen (AdV 2019b). Für das Untersuchungsgebiet wurden nur sehr vereinzelt Bildfehler in Ackerflächen erkannt. Auf die Größe des Untersuchungsgebietes bezogen und aufgrund der Tatsache, dass Ackerflächen nur Gegenstand der Nicht-Zielklasse sind, kann der Fehler vernachlässigt werden.



Abbildung 12 - fehlerhafte Bildinformation von Ackerflächen im TrueDOP

Weitere Fehler in den TrueDOP-Daten sind die transparente Darstellung von sich bewegenden Objekten sowie Datenlücken in Gewässern (AdV 2019b). Abbildung 13

zeigt die beiden Fehler im TrueDOP-Datensatz des Untersuchungsgebietes. Da es sich bei der gesuchten Zielklasse um stationäre Objekte handelt und diese sich nicht innerhalb von Gewässern befinden, sind beide potenziellen Fehlerquellen ohne Relevanz für die Untersuchungen dieser Arbeit.

Im Folgenden werden die zu untersuchenden Digitalen Orthophotos der Qualitätsstufe TrueDOP der Einfachheit halber als DOP-Daten bezeichnet.



Abbildung 13 – Bildfehler im TrueDOP: a) Sich bewegende Objekte erscheinen transparent b) Datenlücken im bDOM wegen schlechter Korrelation führen zu Artefakten in Gewässern

2.2.3 Satellitendaten

Die heutige satellitengestützte Fernerkundung begann 1972 mit dem Start des Landsat-1-Satelliten (Blaschke et al. 2014b), dem ersten zivilen Erkundungssatelliten. Seit jeher offenbaren die Satellitendaten eine besondere Charakteristik durch wiederholende Aufnahmen. Damit stellen Satellitendaten eine kostengünstige und zuverlässige Datenquelle dar, die konsistente Zeitreihen bieten kann (Peña-Barragán et al. 2011). Durch diese Zeitreihen lassen sich unter anderem auch die Entwicklungen in urbanen Räumen analysieren und Gebäudeobjekte extrahieren (Hu et al. 2021; Tamés-Noriega et al. 2015; Jin und Davis 2005; Aytekin et al. 2012).

Seit den Anfängen mit Landsat-1 wurden über die Jahre viele Erdbeobachtungssatelliten unterschiedlicher räumlicher, zeitlicher oder spektraler Auflösung gestartet und betrieben. Dazu zählen etwa: Landsat 2-9 (1982-2021), SPOT-1-7 (1986-2014), IKONOS (1999), MODIS-Terra (2000), Quickbird (2001) und WorldView-1-4 (2008-2016).

Die bildgebenden Satellitenmissionen der Erdbeobachtung der letzten 20 Jahre verfügen neben sehr hochauflösenden Satellitensystemen, wie WorldView-3 (Pixelgröße: 0,31 m panchromatisch) oder QuickBird (Pixelgröße: 0,65 m panchromatisch), auch Satelliten mit einer hohen spektralen Auflösung wie MODIS-Terra (36 Bänder). Viele Erdbeobachtungssatelliten zeichnen sich zudem durch eine hohe Wiederholrate aus. Dabei reicht die Bandbreite von weniger als einem Tag (SkySat), täglich (PlanetScope, Rapideye, SPOT-6/-7) bis zu mehreren Tagen (3 Tage: DEIMOS-1, 5 Tage: Sentinel-2).

PlanetScope-Daten

Der Fokus der Untersuchung zur Ableitung der Gebäudeobjekte aus ALKIS wird auf die Satellitendaten *PlanetScope* der dritten Generation (auch SuperDove genannt) von PlanetLabs gelegt. Die SuperDoves operieren in einer Höhe von 475 km in einem sonnensynchronen Orbit und erfassen die Erdoberfläche mit multispektralen Aufnahmen in einer Pixelgröße von 3 m mit einer täglichen Wiederkehrrate (PlanetLabs 2022). Abbildung 14 zeigt eine Beispieldarstellung der *PlanetScope*-Satellitendaten als Echtfarbenbild, die für diese Untersuchung zur Verfügung stehen.



Abbildung 14 - Beispieldarstellung der vorliegenden *PlanetScope*-Satellitendaten

Tabelle 6 fasst die wichtigsten Dateninformationen zusammen. Die verwendeten *PlanetScope*-Daten zeichnen sich durch eine hohe Farbtiefe (16 Bit) in 8 Spektralkanälen aus. Die Konstellation kann 200 Millionen km² pro Tag mit einer Bodenauflösung (GSD) von durchschnittlich 4,1 m erfassen (PlanetLabs 2022). Aus der Datenbeschreibung geht weiter hervor, dass die Aufnahmen schließlich orthorektifiziert in einer Pixelgröße von 3 m zur Verfügung stehen. Die Daten liegen im TIF-Format in WGS84 vor, werden aber für die Bearbeitung der Thesis in das amtliche Koordinatenreferenzsystem ETRS89/UTM transformiert. Aus den Metadaten der Aufnahmen geht hervor, dass zum Zeitpunkt der Erfassung 0 % Wolkenbedeckung und Dunst vorherrschten (vgl. Anhang 1).

Tabelle 6 - Dateninformationen *PlanetScope* (SuperDove)

Zusammenfassung PlanetScope	
Instrument	PSB.SD (SuperDove)
Spektralkanäle	Coastal Blue Blue Green I Green II Yellow Red Red-Edge NIR
Geometrisch GSD/Pixel Size	4,1 m/3 m
Farbtiefe	16 Bit
Wiederholrate	Täglich bei Nadir
Bilderfassung	200 Millionen km ² /Tag
Georeferenzierung	WGS84
Datenformat	TIF

Neben den bereits genannten Vorteilen der *PlanetScope*-Daten, wie der spektralen Vielfalt durch 8 Kanäle (vgl. Tabelle 7), der großen Flächendeckung (vgl. Tabelle 6), sowie dem kontinuierlich hohen Aktualisierungsgrad, weisen die Aufnahmen auch Nachteile auf. Dazu zählt stellenweise eine merkliche Überbelichtung (Abbildung 15). Helle Objekte, wie Industriegebäude, weisen eine teilweise sichtbare Überstrahlung auf (Beispiel a), außerdem treten Farbreflexionen auf, die eine Auswertung des Bildmaterials an dieser Stelle nicht möglich machen (Beispiel b).

Tabelle 7 - Bandbreite der Spektralkanäle der *PlanetScope*-Aufnahmen

Spektralband	Bandbreite (nm)
Coastal Blue	431-452
Blue	465-515
Green I	513-549
Green II	547-583
Yellow	600-620
Red	650-680
Red-Edge	697-713
NIR	845-885

Im Vergleich zu den DOP-Daten weisen die Satellitendaten zwar eine geringere Bodenauflösung auf, besitzen aber vergleichsweise mehr spektrale Kanäle, was als großer Vorteil angesehen werden kann. Jede weitere spektrale Information, die über die Standard-RGBI-Kanäle hinausgeht, verbessert die Unterscheidungsmöglichkeiten zwischen den Objekten (Vakalopoulou et al. 2015).

Die Entscheidung zur Wahl der *PlanetScope*-Daten für diese Thesis begründet sich insbesondere darin, dass die Daten durch die hohe Wiederkehrzeit grundsätzlich täglich zur Verfügung stehen und dadurch eine Detektion von Gebäudeobjekte in einer hohen Aktualität durchgeführt werden kann. Ein weiterer interessanter Aspekt für die Wahl der *PlanetScope*-Daten ist die Frage, ob eine höhere geometrische Auflösung der DOP-Daten eine Voraussetzung für die Detektion von Gebäuden ist oder ob eine mittlere Auflösung in Kombination mit einer größeren Anzahl an Spektralkanälen auch zu einem adäquaten Ergebnis führen kann. Für diese Untersuchung wird also eine geometrische Komponente einer zeitlichen Komponente gegenübergestellt.

Im Folgenden werden die zu untersuchenden Satellitendaten *PlanetScope* der Einfachheit halber als SAT-Daten bezeichnet.

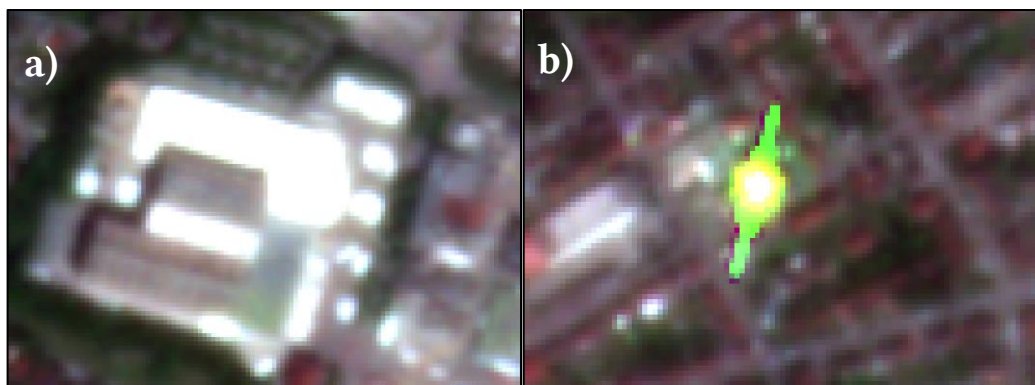


Abbildung 15 – Überstrahlungseffekt heller Objekte (links) und grüne Farbreflexion (rechts)

3 Methodische Grundlagen

Im folgenden Kapitel werden die Methoden aufgeführt, die im Anschluss für die Umsetzung der Fragestellung angewendet werden. Dabei wird zuerst die objektbasierte Herangehensweise beschrieben, gefolgt von Erläuterungen zum genutzten Klassifikator. Den Abschluss bilden die Beschreibung der Bildsegmentierung und der Methoden zur Ergebnisbewertung.

3.1 Objektbasierte Herangehensweise

In der Geoinformatik werden digitale Geodaten, also räumliche Informationen auf zwei Arten dargestellt, in Form von Vektor- oder als Rasterdateien (Lange 2020). Beim Vektordatenmodell wird die Welt in Punkte, Linien und Flächen abstrahiert (Bill 2016). Im Gegensatz dazu, zählt beim Rasterdatenmodell, etwa aus einer Luftaufnahme, das Pixel als kleinste Messeinheit zur visuellen Repräsentation der erfassten Informationen und stellen diese Bildpixel als gleichmäßiges quadratisches Gitter dar (Lange 2020). Diese Bildpixel stellen die räumlichen Informationen als gleichmäßiges quadratisches Gitter dar (Lange 2020).

Dieser pixelbasierte Ansatz spielte viele Jahren eine zentrale Rolle in der Bildklassifikation von Fernerkundungsdaten, allerdings stellten Blaschke und Strobl (2001) fest, dass, obwohl stetige Verbesserungen bezogen auf die Auflösung und die Datenverfügbarkeit über die Jahre verzeichnet werden konnten, die Bildanalyse weiterhin auf den Konzepten der pro-pixel Analyse aus den 70er Jahren beruhte. Blaschke et al. (2004) beobachteten eine Zunahme an Anwendungen, die „über das Pixel hinausgehen“. Als Grundlage dieses neuen Ansatzes dient die Bildsegmentierung, welche den Grundstein der objektbasierten Bildanalyse darstellt (Hay und Castilla 2008).

Der Hauptunterschied der beiden Herangehensweisen ist die kleinste Bearbeitungseinheit der Bildaufnahme. Bei einer pixel-basierten Klassifikation wird jedes einzelne Pixel klassifiziert; ein Pixel ist normalerweise die kleinste Entität einer Bildaufnahme (Blaschke et al. 2014b). Bei einer objektbasierten Klassifikation hingegen werden Bildpixel zu spektralen homogenen Bildobjekten aggregiert und das individuelle Objekt klassifiziert (Liu und Xia 2010). Eine Gruppe von Pixeln mit einer ähnlichen spektralen und räumlichen Eigenschaft werden dabei als Objekt definiert (Zhang et al. 2020).

In den folgenden Jahren wurden diese Bildsegmentierungsmethoden populärer (Blaschke 2010). Die Entwicklung führte schließlich dazu, dass die objektbasierte Bildanalyse laut Blaschke et al. (2014) zur neuen Standardmethode bei der Bildanalyse erklärt wurde. Sie führten an, dass die Vielfalt der verwendbaren georäumlichen Daten, die über reine Bilddaten hinausgehen, und die steigende Notwendigkeit für multidisziplinäre Forschung zu einem einzigartigen Merkmal der objektbasierten Bildanalyse werden könne (Blaschke et al. 2014b). Die Vorteile gegenüber pixelweiser Klassifi-

kation in hochauflösenden Bildern liegen laut Jozdani et al. (2019) unter anderem darin, dass ein einzelnes Pixel in einem hochauflösenden Bild bekanntlich nur einen kleinen Teil eines Landschaftselements repräsentiert. Das führt dazu, dass bei der pixelweisen Bearbeitung die Vielfältigkeit der verschiedenen LN/LB-Typen nicht richtig modelliert werden kann (Jozdani et al. 2019). Anders ausgedrückt, führen Jozdani et al. (2019) fort, hindert das Fehlen zusätzlicher spektraler, räumlicher und textueller Informationen die pixelweise Klassifikation daran, einzelne Pixel der richtigen Klasse zuzuordnen. Beim pixelbasierten Ansatz werden in hochauflösenden Bildern keine Nachbarschaftspixel beachtet, welche ein Teil der selben Landbedeckungsklasse sein können (Hossain und Chen 2019). Hay et al. (2001) bezeichnen ein Bildobjekt als eine grundlegende Einheit innerhalb eines Bildes, dabei besitzt jede dieser Pixelgruppen ähnliche digitale Werte, die aber eigene Größe, Form und geographische Beziehungen besitzen. Daraus lässt sich schließen, dass Objekte innerhalb einzelner Regionen eine höhere Homogenität aufweisen, als zwischen ihnen und ihren Nachbarn (Yu et al. 2006). Ein weiterer Vorteil der Verwendung von Bildobjekten ist die Reduktion der spektralen Variation, die den sogenannten *Salt-and-Pepper*-Effekt verhindert, welcher grundsätzlich bei einem pixelbasierten Klassifikationsverfahren auftritt (Liu und Xia 2010; Yu et al. 2006). Zur Steigerung der Klassifizierungsgenauigkeit können im objektbasierten Ansatz zusätzlich eine große Anzahl von Merkmalen, die sowohl räumliche als auch textur- und kontextbezogene Eigenschaften von Objekten charakterisieren, als ergänzende Informationen zu den spektralen Beobachtungen abgeleitet werden, welche die Klassifizierungsgenauigkeit verbessern (Guo et al. 2007).

Die objektbasierte Bildanalyse baut auf Segmentierung, Kantenerkennung, Merkmalsextraktion und Klassifikationskonzepten auf, welche seit vielen Jahrzehnten in der Fernerkundung etabliert sind (Blaschke 2010). Obwohl OBIA den *Salt-and-Pepper*-Effekt verhindern und zu einer höheren Genauigkeit als die pixelbasierten Methoden führt (Blaschke und Strobl 2001), bleiben Fehlklassifikation in komplexen städtischen Gebieten nach wie vor bestehen (Chen et al. 2019). Trotzdem attestieren verschiedene Publikationen der letzten Jahre der objektbasierten Klassifikation (insbesondere bei hochauflösenden Aufnahmen) durchgehend bessere Ergebnisse als der traditionellen pixelbasierten Herangehensweise (Blaschke et al. 2014b; Jozdani et al. 2019; Myint et al. 2011; Blaschke 2010; Yu et al. 2006). Dies gilt insbesondere auch für die Gebäudeextraktion (Liu und Xia 2010).

OBIA hat im letzten Jahrzehnt, sowohl in der Forschung als auch in der Industrie, große Beachtung erlangt und gehört als etablierte Methode zu den besten und am weitverbreitetsten Techniken im Bereich der Fernerkundung für die Extraktion von Objekten aus hochauflösenden Aufnahmen (El-naggar 2018).

Somit lässt sich im Fall, der in dieser Thesis verwendeten hochauflösenden DOP-Aufnahmen und Satellitenaufnahmen sagen, dass eine objektbasierte Herangehens-

weise der pixelbasierten vorzuziehen ist, da davon ausgegangen werden kann, dass dadurch Ergebnisse höherer Güte erzielt werden können.

OBLA besteht normalerweise aus zwei Hauptphasen: (1) Bildsegmentierung und (2) Merkmalsextraktion und Klassifikation (Hossain und Chen 2019).

Die Bildsegmentierung wird im Kapitel 3.3 erläutert. Der zweite Schritt wird in dieser Thesis mit einem Deep Learning-Ansatz gelöst, welcher nachfolgend theoretisch beschrieben wird. Die Ergebnisse werden ausführlich ab Kapitel 5 erläutert und diskutiert.

3.2 Deep Learning-Ansatz

Als Deep Learning wird ein Unterbereich des Maschinellen Lernens bezeichnet, was wiederum der Künstlichen Intelligenz zugeordnet wird (Sarker 2021). Beim Maschinellen Lernen bilden Computeralgorithmen die menschliche Intelligenz nach, indem sie aus der Umgebung lernen (El Naqa und Murphy 2015).

Deep Learning beschreibt einen Teilbereich der Künstlichen Neuronalen Netze (*Artificial Neuronal Networks* (ANN)), denen es Rechenmodellen, die aus mehreren Verarbeitungsschichten bestehen, ermöglicht, Repräsentationen von Daten mit mehreren Abstraktionsebenen ohne menschliches Eingreifen selbst zu erlernen (LeCun et al. 2015).

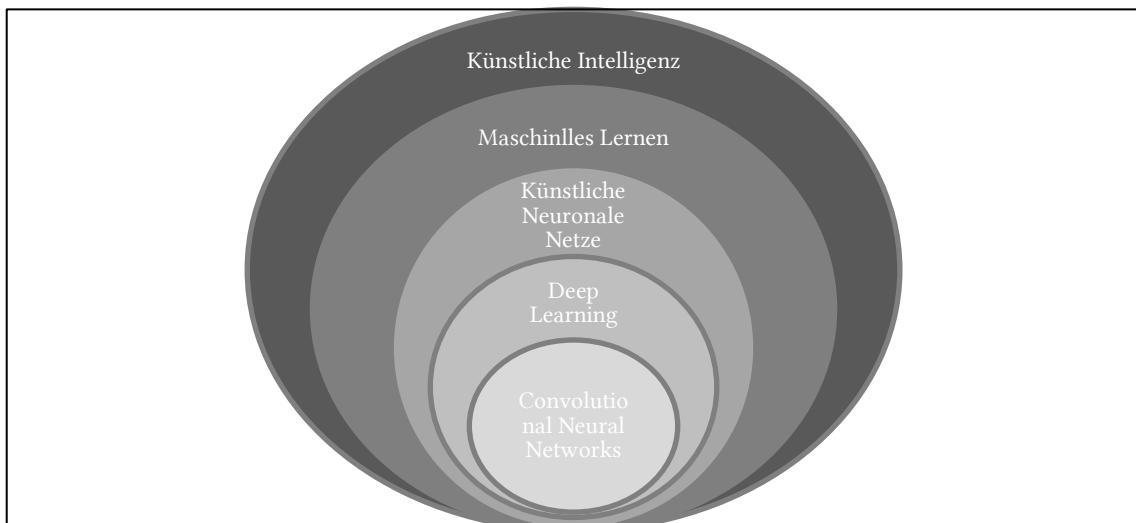


Abbildung 16 – Einordnung von Deep Learning

In den letzten Jahren haben Deep Learning-Ansätze große Bereiche des Maschinellen Lernens revolutioniert, wie etwa die Bildklassifikation, die Videoverarbeitung und die Spracherkennung (Liu et al. 2020). Die jüngsten Fortschritte auf diesem Gebiet haben gezeigt, dass damit vereinzelt sogar menschliche Fähigkeiten übertroffen werden, wenn es hochkomplexe Aufgaben zu lösen galt (Zhu et al. 2017). Audebert et al. (2017) sprechen davon, dass Deep Learning-Methoden einen neuen Weg darstellen, um alte Probleme in der Fernerkundung zu lösen.

Deep Learning basiert auf der hierarchischen Lernrepräsentation von Merkmalen unter Verwendung einer tiefen Struktur aus verschiedenen verborgenen Schichten (*Hidden Layers*) (Deng und Dong 2014). Laut Kattenborn et al. (2021) befinden sich innerhalb dieser Schichten Neuronen, die durch Gewichtungen und *Biases* miteinander verbunden sind. Die erste Schicht bei einem ANN ist die Eingabeschicht, welche die Inputdaten aufnimmt, die letzte Schicht ist die Ausgabeschicht, die die Vorhersage der Klassifizierung ausgibt, dazwischen befinden sich die verborgenen Schichten, die den Merkmalsraum der Eingabedaten umwandeln, sodass sie mit der Ausgabe übereinstimmen (Kattenborn et al. 2021).

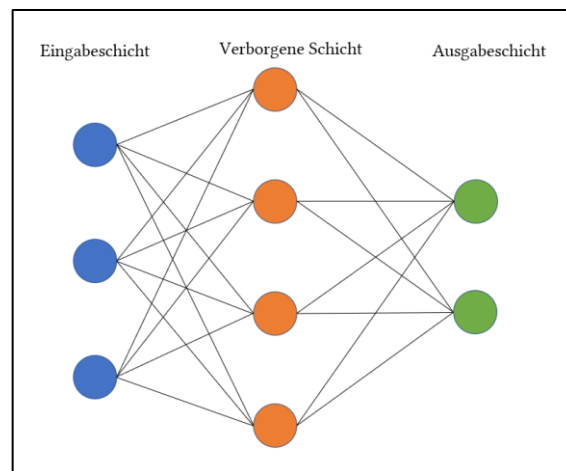


Abbildung 17 - Aufbau eines einfachen ANN

Als CNN wird ein mehrschichtiges Künstliches Neuronales Netz mit Faltungsfiler bezeichnet, welches eigenständig Merkmale direkt aus den Eingangsdaten extrahieren kann (Sarker 2021). Es handelt sich dabei um eine Variante des Deep Learning-Ansatzes, der von biologischen Prozessen inspiriert wurde (LeCun und Bengio 1995) und dessen Ursprung auf den Versuch zurückgeht, die Mustererkennung des menschlichen Gehirns nachzubilden (Fukushima und Miyake 1982). Das erste Mal von LeCun et al. (1989) für die Erkennung von Handschriften vorgestellt, gehören CNN mittlerweile zu den bekanntesten Deep Learning-Ansätzen (Albawi et al. 2017).

Wie jedes typische Modell eines neuronalen Netzes besitzt ein CNN Neuronen, die innerhalb von Schichten miteinander verbunden sind (vgl. Abbildung 17). Zusätzlich besitzen CNN mindestens eine Faltungsschicht (*Convolution*) innerhalb der verborgenen Schichten, welche die Erkennung räumlicher Muster ermöglicht (Kattenborn et al. 2021).

Diese spezifische Netzwerk-Architektur ist für die Prozessierung von Informationen konzipiert, die in Form von Arrays vorliegen, wie etwa die einzelnen RGB-Kanäle eines Farbbildes (LeCun et al. 2015). Aus diesen Eingangsdatensätzen erlernt das CNN automatisch und adaptiv räumliche Hierarchien von Merkmalen (Yamashita et al. 2018). Laut LeCun et al. (2015) basiert die Funktionsweise auf Repräsentationslernen:

Dadurch gelingt es einer Maschine auf Basis von Rohdaten automatisiert die für die Erkennung oder Klassifizierung erforderlichen Repräsentationen abzuleiten. Bei einer Deep Learning-Repräsentations-Lernmethode mit mehreren Repräsentationsebenen, die durch die Zusammenstellung einfacher, aber nicht linearer Module erreicht werden, wird eine Umwandlung einer Repräsentation zu einer höheren, etwas abstrakteren Ebene durchgeführt (LeCun et al. 2015). CNN konnten in den letzten 15 Jahren dadurch bemerkenswerte Ergebnisse in der Bild- (Girshick et al. 2013; Ciresan et al. 2012), Sprach- (Hinton et al. 2012; Karpathy et al. 2014) und Mustererkennung (Albawi et al. 2017) erzielen. Im Vergleich zu ANN ist die verringerte Anzahl an Parametern bei CNN ein großer Vorteil (Albawi et al. 2017) (vgl. Abschnitt 4.5).

Im Vergleich zu anderen Klassifizierungsalgorithmen (wie etwa *random forest* oder *support vector machine*) benötigen CNN keine Extraktion und händische Auswahl der Merkmale (Liu und Abd-Elrahman 2018). Audebert et al. (2017) schreiben, dass seit den frühen Anwendungen zur Straßenerkennung um 2010, die CNN seitdem erfolgreich bei der Klassifikation und Bestimmung von Objekten in Luftbildern verwendet wurden.

Ein CNN ist ein mathematisches Konstrukt und besteht typischerweise aus drei Arten von Schichten: *Convolution*, *Pooling* und *Fully Connected Layers* (Yamashita et al. 2018; LeCun et al. 2015). Diese werden im Folgenden genauer beschrieben.

Faltung (*Convolution*):

Die Faltung ist eine spezielle lineare Operation, die für die Merkmalsextraktion verwendet wird. Dabei wird ein kleines Zahlenfeld, der sogenannte *Kernel*, auf die Eingabe angewendet (Yamashita et al. 2018). Nach Kattenborn et al. (2021) befinden sich verschiedene optimierbare Kernel in der Faltungsschicht (*convolution layer*), deren Aufgabe es ist, die Eingabe bzw. die vorangegangenen verborgenen Schichten zu transformieren. Sie führen in ihrer Publikation weiter aus, dass die sich daraus ergebenden Transformationen darauf abzielen, Muster zu erkennen. Diese entscheidenden Muster werden iterativ durch die Faltung gelernt, was im Wesentlichen das Gleiten des Filters über die Schicht und die Neuberechnung der Einheit ausgehend von dem Filter und den Werten der betrachteten Schicht bedeutet (Kattenborn et al. 2021). LeCun et al. (2015) bezeichnen die Einheiten einer Faltungsschicht als *feature map*. Innerhalb einer Schicht ist jede Einheit mit den *feature maps* der vorangegangenen Schicht mit einer Gewichtung verbunden.

Abbildung 18 visualisiert die eben vorgestellte Funktionsweise der *Convolution*-Schicht. In diesem Beispiel wird ein 3 x 3 Kernel auf die Eingabe angewandt und das Produkt zwischen jedem Element des Kernels und der Eingabe berechnet, welche wiederum aufsummiert werden, um so den Wert an der entsprechenden Position der *feature map* zu erhalten (Yamashita et al. 2018).

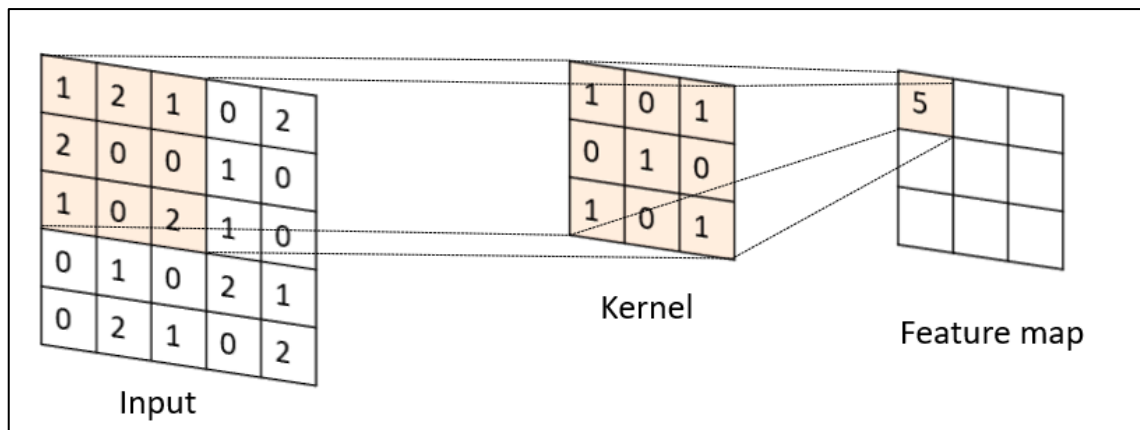


Abbildung 18 - Funktionsweise der Faltungsoperation (nach Yamashita et al. (2018))

Zusammenlegung (Pooling):

Albawi et al. (2017) beschreiben, dass die Hauptidee des *Poolings* das *Downsampling* zur Reduktion der Komplexität für weitere Ebenen ist. Bei Anwendung im Bereich der Bildbearbeitung kann es als eine Maßnahme zur Verringerung der Auflösung angesehen werden (Albawi et al. 2017). In ihrer Publikation bezeichnen Albawi et al. (2017) das Max-Pooling als die gängigste Methode; dabei wird das Bild in Rechtecke unterteilt und anschließend nur der Maximalwert des Innenraums jedes Unterbereiches zurückgegeben. Der bekannteste Ansatz beim Max-Pooling ist dabei der 2x2-Block.

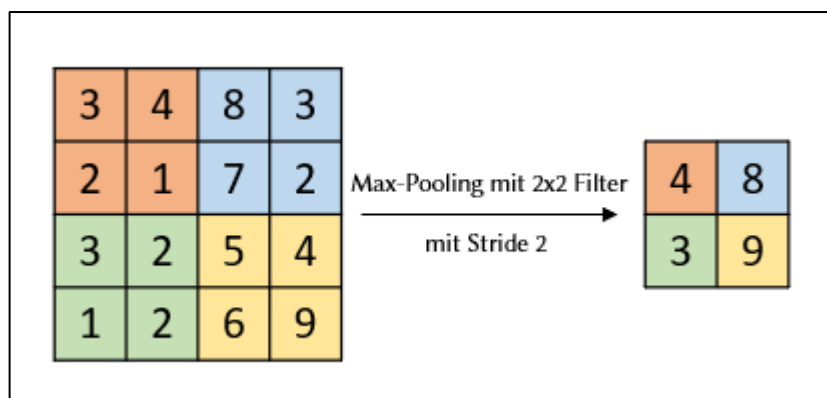


Abbildung 19 - Funktionsweise des Max-Pooling-Filters (nach Yamashita et al. (2018))

Die Funktionsweise beim Max-Pooling lässt sich mithilfe der Abbildung 19 wie folgt erklären: Ein 4x4-Input soll ein Max-Pooling mit einem 2x2-Filter erfahren. Der 4x4-Input wird in vier Unterbereiche aufgeteilt und der 2x2-Output beinhaltet die Maximalwerte aus den jeweiligen korrespondierenden Unterbereichen. Dabei können noch Hyperparameter wie der *Stride* angepasst werden (Abbildung 20), der oftmals einen Wert von 2 hat. Der *Stride* gibt die Schrittweite des gleitenden Fensters an.

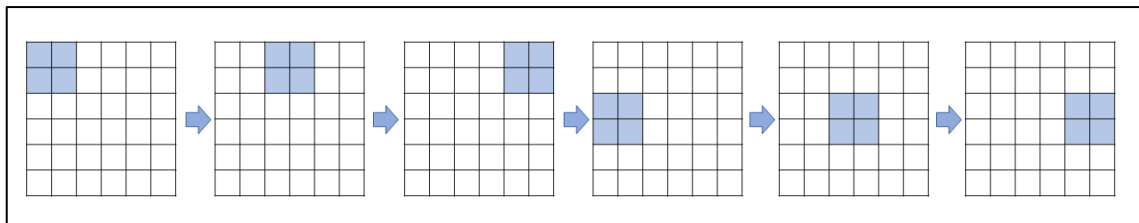


Abbildung 20 - Funktionsweise für die ersten beiden Reihen mit dem Hyperparameter Stride 2

Fully Connected Layer

Als letzte Schicht findet sich in einem CNN die vollständig verbundene Schicht (*fully connected layer*). Diese Schicht hat eine Einheit für jede Klasse, die das Netz vorher sagt, und eine dieser Einheiten enthält Informationen aller vorherigen Schichten (Trimble 2020).

Herausforderungen des Deep Learning-Ansatzes:

Herausforderungen in der Erkennungsgenauigkeit liegen in der großen Bandbreite von Variationen innerhalb der Klassen und in der großen Anzahl von Objektkategorien (Liu et al. 2020). Bei den Variationen innerhalb einer Klasse wird zwischen zwei Typen unterschieden: Faktoren, die ein Objekt durch verschiedene Variationen bezüglich Größe, Farbe, Form, Material und Textur aufweist, sowie die Variationen die durch Bildaufnahmebedingungen entstehen, wie die Tageszeit, die Wetterverhältnisse, die Aufnahmebeleuchtung oder spezifische Kameraeigenschaften (Liu et al. 2020). Allgemein betrachtet, wird die CNN-basierte Klassifikationsmethode auf Pixelbasis ausgeführt und die Klassifizierungsergebnisse können in den Übergangsbereichen zwischen verschiedenen Klassen durch Mischpixel verschlechtert werden (Jin et al. 2019). Diese Schwierigkeiten sind durch eine Anpassung der Trainingsmenge oder durch die Vergrößerung der Modelltiefe des CNN überwindbar (Jin et al. 2019; Sameen und Pradhan 2019). Durch die Einführung weiterer *Hidden Layer* und Knoten kann ebenfalls eine marginale Verbesserung herbeigeführt werden (Lin et al. 2016). Auch das Anwenden einer *Data Augmentation* stellt eine mögliche Lösung zur Minimierung der Auswirkungen der zuvor genannten Herausforderungen dar (Radovic et al. 2017).

Training des Netzes:

Die grundsätzliche Arbeitsweise eines CNN basiert auf der Verfahrensweise einer überwachten Klassifizierung. Bevor das Neuronale Netz eine Klassifizierung durchführen kann, muss zuerst ein Modell generiert werden, welches mithilfe von Referenzdaten trainiert wird. Als Training eines Neuronalen Netzes wird der Prozess bezeichnet, bei dem optimierbare Kernels in den Faltungsschichten und die Gewichtungen in den

vollständig verbundenen Schichten gefunden werden (Yamashita et al. 2018). Ziel ist dabei die Unterschiede zwischen der Ausgangsprognose und gesicherten gelabelten Referenzdaten des Trainingsdatensatzes zu minimieren (Yamashita et al. 2018). Als gelabelte Referenzdaten werden gesicherte Bildinformationen zu den gesuchten Klassen bezeichnet auf Basis derer das Neuronale Netz die Gewichtungen für eine spezielle Aufgabe optimiert (Kattenborn et al. 2021), wie beispielsweise der Erkennung von Gebäudeobjekten.

Der Trainingsprozess in Neuronalen Netzen nutzt die Methode des Backpropagation-Algorithmus (Yamashita et al. 2018), dabei wird das statistische Gradientenverfahren zur Optimierung der Verlustfunktion angewendet, bei dem die Gewichte jeder Schicht des Netzes iterativ aktualisiert werden, solange bis die Fehlerrate des Netzes das gewünschte Mindestmaß enthält (Jozdani et al. 2019).

Gemäß Yamashita et al. (2018) misst die Verlustfunktion durch Vorwärtspropagation (*Forward-Propagation*) die Kompatibilität zwischen der Ausgangsprognose des Netzes und den gegebenen gelabelten Referenzdaten. Das Gradientenverfahren wird als Optimierungs-Algorithmus genutzt, der iterativ die Trainingsparameter, wie die Gewichtungen anpasst, um die Verlustfunktion zu minimieren (Yamashita et al. 2018).

3.3 Bildsegmentierung

Wie bereits im Abschnitt 3.1 „Objektbasierte Herangehensweise“ beschrieben, ist bei der Verwendung einer objektbasierten Bildanalyse die Bildsegmentierung von entscheidender Bedeutung (Blaschke et al. 2008; Hossain und Chen 2019). Die Genauigkeit der objektbasierten Merkmalsextraktion und Klassifikation hängt von der Qualität der Bildsegmentierung ab (Su und Zhang 2017), somit hat die Segmentierung eine direkte Auswirkung auf das Endergebnis (Blaschke et al. 2004).

Eine Segmentierung ist eine Aufteilung eines Arrays von Messungen auf der Grundlage der Homogenität (Blaschke et al. 2014b). Es unterteilt ein Bild - oder beliebige Raster- und Punktdaten - in räumlich kontinuierliche, zusammenhängende und homogene Bereiche, die als „Segmente“ bezeichnet werden (Blaschke et al. 2014b). Dieser Segmentierungsschritt, bei dem Pixelgruppen zu semantisch vergleichbaren, sich nicht überlappenden Objekten zusammengefügt werden, ist der kleinste ansprechbare Baustein bei OBIA und ermöglichen die zusätzliche Extraktion von spektralen, räumlichen und textur-basierten Merkmalen, die für eine Klassifikation verwendet werden können (Foody et al. 2006; Jozdani et al. 2019).

Nach Blaschke et al. (2014b) kann eine Segmentierung in drei Methoden kategorisiert werden: (a) die pixelbasierte (Mardia und Hainsworth 1988), (b) die kantenbasierte (Perona und Malik 1990) und (c) die regionenbasierte Methode (Beveridge et al. 1989).

Laut Hossain und Chen 2019 kann ergänzend je nach Literatur auch eine (d) hybride Methode (Haris et al. 1998) verwendet werden.

Hossain und Chen 2019 führen weiter aus, dass die pixelbasierte Herangehensweise aus der Bildung von Schwellenwerten und der dadurch resultierenden Segmentierung im Merkmalsraum besteht (Schiewe 2002).

Zur kantenbasierten Methode schreiben Hossain und Chen 2019, dass zuerst Kanten identifiziert und diese anschließend mit Algorithmen zu Umrissen geschlossen werden müssen. Dabei wird angenommen, dass sich die Pixeleigenschaften an Kanten abrupt verändern (Shih und Cheng 2004). Eine der bekanntesten kantenbasierten Segmentierungsmethoden ist die *Watershed Transformation* (Vincent und Soille 1991). Dabei wird dem Algorithmus allerdings sowohl eine kantenbasierte als auch eine regionenbasierte Segmentierung zugesprochen (Meinel und Neubert 2004).

Im Gegensatz dazu, werden bei der regionenbasierten Segmentierung zuerst Objekte bestimmt und anschließend die Objektgrenzen gebildet (Ohta et al. 1980; Pong et al. 1984). In ihrer Publikation beschreiben Hossain und Chen 2019 die Vorgehensweise wie folgt: Ein Objekt beginnt im Zentrum und erweitert sich nach außen, bis es die Objektgrenzen erreicht (Yu-Jin Zhang 2006).

Nicht nur die Wahl der Segmentierungsmethode an sich, ist ausschlaggebend für das Ergebnis, sondern auch die Wahl der entsprechend zu setzenden Parameterwerte (Zhang et al. 2020). Die Bestimmung der optimalen Parameter ist oftmals nur mit heuristischen Mitteln wie „*trial and error*“ zu erreichen (Hay et al. 2003). Bei ungeeigneten Parametern können Über- oder Unter-Segmentierungen der Objekte entstehen (Möller et al. 2007).

Eine Über-Segmentierung teilt ein semantisches Objekt in mehrere kleinere Bildobjekte auf, eine Unter-Segmentierung wiederum tritt auf, wenn verschiedene semantische Objekte in ein großes Bildobjekt gruppiert werden (Liu und Xia 2010). Eine Über-Segmentierung ist einer Unter-Segmentierung vorzuziehen (Hay und Castilla 2008; Gao et al. 2011). Liu und Abd-Elrahman (2018) führen an, dass die Verwendung von zu kleinen Segmentierungskennzahlen zu kleineren und wahrscheinlich zu homogeneren Objekten führt. Bei sehr kleinen Objektgrößen werden allerdings die eindeutige Trennung zwischen den Klassen verringert und die Berechnungszeiten erhöht (Liu und Abd-Elrahman 2018; Zhang et al. 2020). Eine zu große Segmentierungskennzahl wiederum führt zu vielen großen Objekten, die dazu neigen verschiedene Klassen miteinander zu vermischen (Liu und Abd-Elrahman 2018). Bei einer Über-Segmentierung kann jedes Pixel innerhalb eines übersegmentierten Objektes noch zur richtigen Klasse zugeordnet werden, da sich verschiedene semantische Objekte nicht vermischen (Marpu et al. 2010). Es kann theoretisch weiterhin eine Genauigkeit von 100 % erreicht werden (Liu und Xia 2010).

Die Software, die in dieser Thesis für die Segmentierung und auch für die Klassifikation verwendet wird, ist eCognition Developer 10.2 der Fa. Trimble. Bezogen auf die Bildsegmentierung wurde die Methodik der *Multiresolution Segmentation* (MRS) (Batz und Schäpe 2000) aus den entsprechend zur Verfügung stehenden Möglichkeiten ausgewählt und auf die DOP- und die SAT-Daten angewendet.

Multiresolution Segmentation

Die *Multiresolution Segmentation* ist eine weitverbreitete Methode zur Segmentierung von Fernerkundungsaufnahmen (Belgiu und Drăguț 2014). Der Algorithmus basiert auf dem *Fractal Net Evolution Approach* (FNEA) und ist eingebettet in eCognition Developer der Fa. Trimble. Die Software stellte bei ihrer Markteinführung im Jahr 2000 als erste kommerziell verfügbare Software für objektbasierte Bildanalysen einen wichtigen Wendepunkt dar (Benz et al. 2004; Blaschke 2010; Blaschke et al. 2015).

Laut Liu und Xia (2010) handelt es sich bei MRS um einen regionenbasierten und aufwärtsgerichteten Ansatz, der kleine Bildobjekte nacheinander in größere Bildobjekte zusammenfasst. Der Segmentierungsprozess beginnt dabei auf Einzelpixel-Ebene die zu Objekten anwachsen, so lange der Grad der Heterogenität unter einem vom Nutzer definierten Schwellwert (*scale parameter*) liegt (Liu und Xia 2010). Der *scale parameter* (SP) ist, neben den Parametern *shape* und *compactness*, einer von drei manuell zu setzenden Vorgaben, die aber die größten Auswirkungen auf den Prozess der Segmentierung haben (El-naggar 2018). Die Genauigkeit des Klassifizierungsprozesses ist vom Aufteilungsprozess eines Bildes abhängig, und dieser hängt von der Größe der Bildobjekte ab, die vom SP angegeben werden (Drăguț et al. 2014). Durch den Parameter *shape* wird entweder der Einfluss der Farbe oder der Form priorisiert, *compactness* kann als das Produkt zwischen der Breite und der Länge des jeweiligen Objektes gesehen werden (El-naggar 2018). Der Parameter *shape* passt dabei die Gewichtung der spektralen Homogenität gegenüber der Form des Objekts an, während der Parameter *compactness* die Form zwischen glatten Rändern und kompakten Kanten definiert (Myint et al. 2011).

3.4 Bewertungsmetrik

Bei Messungen der Präzision von Klassifizierungsalgorithmen hat sich die Genauigkeitsanalyse (*Accuracy Assessment*) etabliert (Congalton 2019) und dient als Standardwerkzeug bei der Klassifikation von Fernerkundungsdaten. Für die Genauigkeitsanalyse eines Klassifizierungsergebnisses wird erfahrungsgemäß die Konfusionsmatrix (Kohavi und Provost 1998) verwendet. Es handelt sich dabei um eine einfache Kreuztabelle von Referenz- und Kontrolldaten der kartierten Klassen. Diese enthält Informationen über die tatsächliche und die vorhergesagte Klassifikation für einzelne Objekte (Shultz et al. 2010).

Eine Konfusionsmatrix beinhaltet auch weitere Informationen wie die Hersteller-, die Nutzer- und die Gesamtgenauigkeit und wird auch als Fehlermatrix bezeichnet (Ahmad und Quegan 2012).

Dabei fallen die Ergebnisse der Klassifizierung in vier Klassen:

1. True positive (TP): Der vorhergesagte und der tatsächliche Wert stimmen überein. Das CNN erkennt mit einer hohen Wahrscheinlichkeit ein Gebäudeobjekt und die Referenz gibt an der gleichen Stelle auch ein Gebäude an.
2. True negative (TN): Der vorhergesagte und der tatsächliche Wert stimmen überein. CNN erkennt ein Nicht-Gebäude und laut Referenz befindet sich tatsächlich kein Gebäude in dieser Stelle.
3. False positive (FP): Der vorhergesagte und der tatsächliche Wert stimmen nicht überein. Die Vorhersage des CNN erkennt ein Gebäudeobjekt, laut Referenz befindet sich aber kein Gebäude an dieser Stelle.
4. False negative (FN): Der vorhergesagte und der tatsächliche Wert stimmen nicht überein. Die Vorhersage des CNN erkennt ein Nicht-Gebäudeobjekt, in Realität befindet sich ein Gebäudeobjekt an dieser Stelle.

Tabelle 8 – Schematische Darstellung einer Konfusionsmatrix

		Validierung	
		Gebäude	Nicht-Gebäude
Klassifikation	Gebäude	TP	FP
	Nicht-Gebäude	FN	TN

Gemäß Oczipka (2007) lassen sich die weiteren Genauigkeitskennzahlen folgendermaßen ableiten:

$$a) \text{ Gesamtgenauigkeit} = \frac{\text{Gesamtanzahl der korrekt klassifizierten Pixel}}{\text{Gesamtanzahl aller Referenzpixel}} \times 100$$

Die Gesamtgenauigkeit (*Overall Accuracy*) ist ein Maß für das Gesamtverhalten des Klassifikators. Sie beinhaltet die Nutzer- und Herstellergenauigkeiten für das gesamte Bild. Der Wert zeigt den Gesamtprozentsatz der korrekt klassifizierten Pixel an (Jog und Dixit 2016).

$$\text{b) } \textit{Nutzergenauigkeit} = \frac{\textit{Anzahl der richtig klassifizierten Pixel pro Klasse}}{\textit{Gesamtanzahl aller klassifizierten Pixel pro Klasse}} \times 100$$

Die Nutzergenauigkeit (*User's Accuracy*) stellt das Verhältnis der Anzahl an richtig klassifizierten Pixel in einer Klasse zur Gesamtanzahl der Pixel, die in dieser Klasse klassifiziert wurden, dar (Zhang et al. 1995). Sie gibt somit pro Klasse an, welcher Prozentsatz der vorhergesagten Objekte tatsächlich dieser Klasse angehört.

$$\text{c) } \textit{Herstellergenauigkeit} = \frac{\textit{Anzahl der richtig klassifizierten Pixel pro Klasse}}{\textit{Gesamtanzahl der Pixel der Referenzklasse}} \times 100$$

Die Herstellergenauigkeit (*Producer's Accuracy*) ist das Verhältnis der Anzahl der richtig klassifizierten Pixel einer Klasse zur Gesamtanzahl der Referenzpixel dieser Klasse (Zhang et al. 1995). Im Wesentlichen sagt sie aus, welcher Prozentsatz der Referenzdaten vom Klassifikator richtig zugeordnet wurde.

4 Praktische Umsetzung

Das folgende Kapitel beschreibt die praktische Umsetzung der methodischen Grundlagen aus dem vorherigen Kapitel im Hinblick auf eine Ableitung von Gebäudeobjekten aus den zuvor beschriebenen Rasterdaten.

Für die praktische Umsetzung stehen die beiden zu untersuchenden Rasterdatensätze, sowie der Vektordatensatz des amtlichen Gebäudedatenbestandes aus ALKIS zur Verfügung. Nach einer Vorverarbeitung, in der die ALKIS-Gebäudedaten aufbereitet und selektiert werden, folgt der Prozess des Aufbaus der Trainings- und Kontrolldatensätze für die zu untersuchenden DOP- und SAT-Daten.

Die bereinigten ALKIS-Gebäudedaten dienen zusammen mit den jeweiligen Rasterdaten dazu die *Training Samples* zu generieren. Für jedes ALKIS-Objekt wird ein geometrischer Mittelpunkt gebildet, der das Zentrum des *Samples* bildet. Auf Grundlage einer festgelegten *Samples*-Größe werden die Referenzdaten für die DOP- und SAT-Daten generiert. Die Größe des *Samples*-Fensters ist abhängig von der zugrundeliegenden Pixelgröße der jeweiligen Rasterdaten. Innerhalb der Rasterdatensätze bleibt die Größe des *Samples*-Fensters unverändert, unabhängig von der Größe des betreffenden Gebäudeobjektes.

Dieser Ansatz verfolgt den Zweck, dass jedes ALKIS-Gebäudeobjekt, unabhängig von seiner Größe, nur durch ein *Sample* im Trainingsdatensatz repräsentiert wird. Der Grund liegt in der verwendeten Software: Wird der vollständige Gebäudeumring aus ALKIS als Grundlage für die Generierung der *Training Samples* verwendet und ist dieser größer als die Ausdehnung der *Samples*-Fenster, generiert die Software auch mehrere *Samples*-Einheiten desselben Objektes. Dadurch folgt eine Überrepräsentation und somit eine Priorisierung von großflächigen Objekten im Referenzdatensatz. Das Ziel dieser Maßnahme ist es, eine möglichst gleichmäßige Verteilung der Referenzdaten auf große und kleine Gebäude sicherzustellen.

Nachdem die *Samples* generiert wurden, wird das CNN auf deren Grundlage für die DOP- und SAT-Daten trainiert. Dieses trainierte Modell wird anschließend auf ein für das Modell bisher ungesehene Kontrollgebiet angewendet. Das Modell berechnet in diesem Gebiet für jedes Pixel eine Wahrscheinlichkeit für das Vorhandensein eines Gebäudeobjektes.

Nachdem das trainierte Modell für DOP- und SAT-Daten auf das Kontrollgebiet angewendet wurde, folgt eine Segmentierung dieses Kontrollgebietes. Dabei werden die Bildobjekte mithilfe von Wahrscheinlichkeitswerten der trainierten Objektklassen gebildet. Vor der Validierung, dem letzten Schritt, wird eine regelbasierte OBIA basierend auf Expertenwissen zur Optimierung der Ergebnisse aus dem CNN durchgeführt. Abbildung 21 visualisiert die einzelnen Schritte des beschriebenen Workflows.

Für die Durchführung des Trainings des CNN mit anschließender Klassifizierung des Kontrollgebietes steht eine Workstation mit folgenden Spezifikationen zur Verfügung:

- Intel Xeon CPU E5-2667 @3,2 GHz
- Arbeitsspeicher 64 GB
- Grafikkarte Quadro M4000
- Windows 10 64 Bit

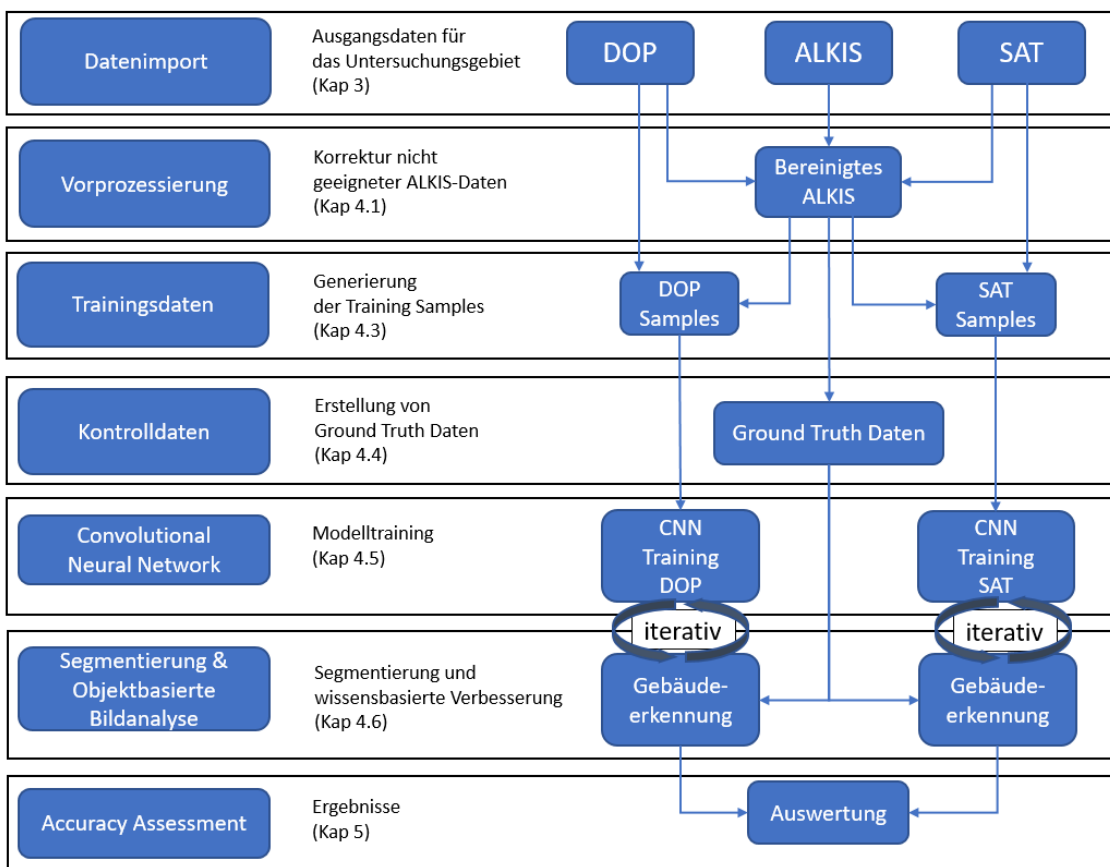


Abbildung 21 - Übersicht der Bearbeitungsschritte für die praktische Umsetzung der Aufgabenstellung

4.1 Auswahl Trainings- und Kontrollgebiet

4.1.1 Trainingsgebiet

Der maximale Bereich des Untersuchungsgebietes, welcher sich aus dem Überlappungsbereich der Satellitendaten und der Luftbildbefliegung ergibt, umfasst eine Fläche von ungefähr 673 km². Mit Erhöhung der räumlichen Auflösung von Bilddaten, steigt auch die Datengröße signifikant an, wodurch die Bildsegmentierung eine große Menge an Ressourcen und Berechnungszeit benötigt (Zhang et al. 2020). Dieser Umstand und die limitierten Hardware-Ressourcen machen die Ableitung von Gebäudeobjekten aus DOP-Daten für das gesamte zur Verfügung stehende Untersuchungsge-

biet unmöglich. Nach der erforderlichen Größenreduktion steht für die Bearbeitung der Thesis eine Fläche von 30 km² als Bearbeitungsgebiet bzw. Trainingsgebiet zur Verfügung. Diese gewählte Flächengröße erlaubt einen Kompromiss zwischen der Rechenzeit und einer breiten Variabilität an zu untersuchenden Gebäudeobjekten.

Um eine hinreichende Anzahl an Gebäudeobjekten innerhalb dieses 30 km² großen Bearbeitungsgebietes zu erreichen, wird eine dichtbesiedelte Fläche ausgewählt. Dafür wird ein Gitterraster mit einer Kantenlänge von 1 km über das Interessensgebiet gelegt und die Anzahl der Gebäudeobjekte innerhalb dieser Quadratkilometer miteinander verglichen. In Abbildung 22 erscheinen Kacheln mit einer hohen Dichte an Gebäudeobjekten in einem dunklen Grünton, Flächen mit einer sehr geringen Anzahl an Gebäudeobjekten sind weiß.

Anhand dieser Auswertung kann das potenzielle Bearbeitungsgebiet einerseits mit einer hohen Quantität an verschiedenen Gebäudeobjekten gewählt werden, gleichzeitig sind aber auch ausreichend Objekte der Nicht-Zielklasse auszuwählen. Aus diesem Grund muss ein Trainingsgebiet bestimmt werden, welches auch Flächen mit einer sehr geringen Anzahl an Gebäudeobjekten beinhaltet.

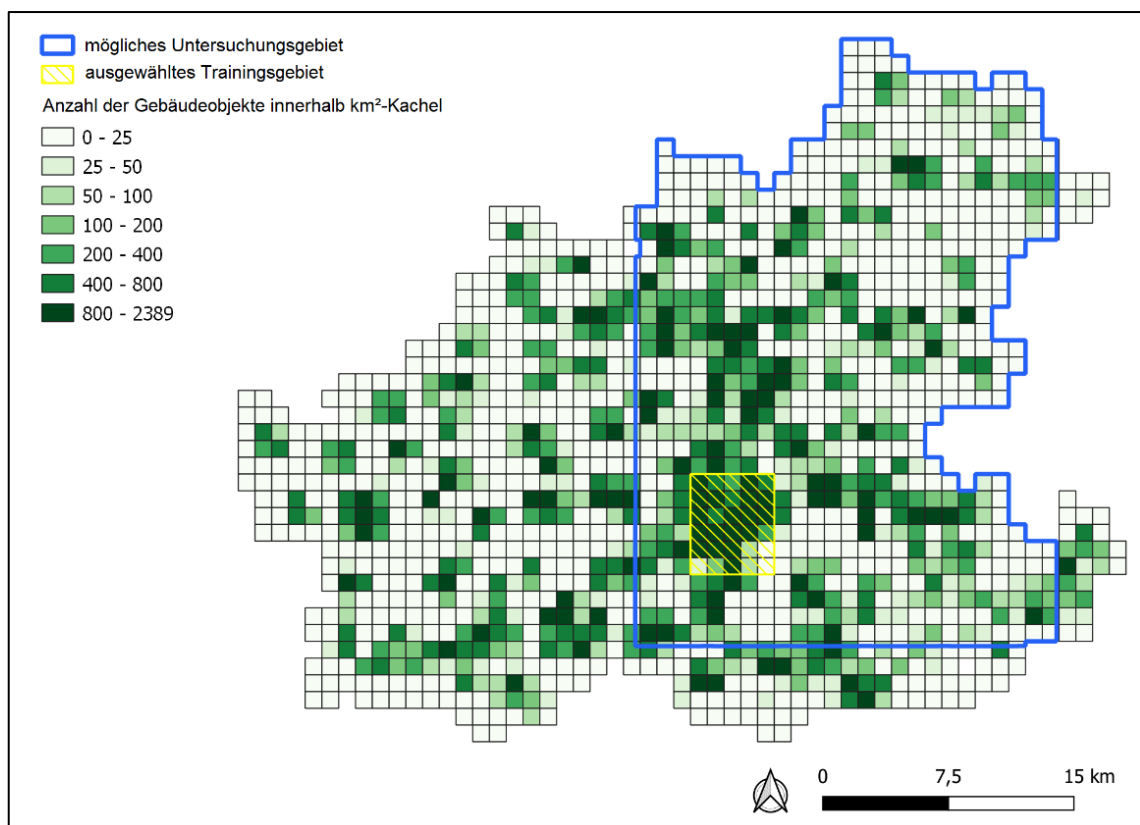


Abbildung 22 - Übersicht der Anzahl an Gebäudeobjekten je km²

Das in Abbildung 22 dargestellte Trainingsgebiet (gelbe Schraffierung) wurde ausgewählt, da es sich durch diese geforderten Charakteristiken auszeichnet. In diesem Ge-

biet des LK/SK Heilbronn verfügen 18 der 30 vorhandenen Kacheln über die höchste Dichte an Gebäudeobjekten aus ALKIS. Für weitere 6 Kacheln liegt die Anzahl an Gebäudeobjekten pro km² im Mittelfeld (zwischen 100 bis 400 unikale Gebäudeobjekte). Die restlichen 6 Kacheln weisen weniger als 50 verschiedene Gebäudeobjekte pro km² auf.

Entsprechend verfügt das ausgewählte Trainingsgebiet insgesamt über 28.330 Gebäudeobjekte (unbereinigt).

4.1.2 Kontrollgebiet

Für das Kontrollgebiet wurde eine Fläche innerhalb des Landkreises Heilbronn definiert, welche sich durch eine große Zahl an nicht eingemessenen Gebäudeobjekten auszeichnet. Dadurch soll die Erkennungsqualität der entwickelten Methodik, zumindest für einen kleinen Teilbereich, auch ohne statistische Berechnungen visuell bewertet werden können. Abbildung 23 zeigt die Lage des Kontrollgebietes innerhalb des LK Heilbronn.



Abbildung 23 - Darstellung des Kontrollgebietes (weißer Rahmen) innerhalb des LK Heilbronn
(rote Umrandung)

Abbildung 24 präsentiert das Kontrollgebiet in einer vergrößerten Darstellung für die DOP- und SAT-Daten in einer RGB-Darstellung. Es handelt sich um einen Teil der Gemeinde Weinsberg und Ellhofen mit einer Gesamtgröße von 6 km².

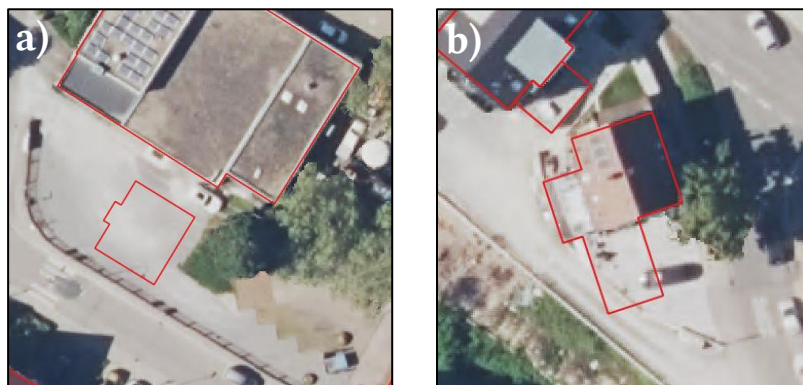


Abbildung 24 - Übersicht des Kontrollgebietes in der originären Auflösung für DOP- (links) und *PlanetScope*-Daten (rechts)

4.2 Bereinigung der Referenz- und Kontrolldaten

Als Referenz- und Kontrolldaten dienen die in ALKIS erfassten, amtlichen Daten der Gebäudeobjekte (vgl. Abschnitt 2.2.1). Die Aufnahme neuer Gebäudedaten findet laufend statt, allerdings erfolgt der Datenabruf der aktualisierten ALKIS-Daten jährlich an einem Stichtag im Frühjahr. Dadurch können Diskrepanzen zwischen der visuellen Bilddarstellung und der Gebäudeeinmessung entstehen. Abbildung 25 zeigt Beispiele solch einer unterschiedlichen Datenlage im ALKIS-Gebäudedatenbestand und in den Digitalen Orthophotos.

Diese potenzielle Fehlerquelle erfordert eine erneute Kontrolle, sowohl für die Trainingsdaten (Abschnitt 4.1.1), als auch für die Kontrolldaten (Abschnitt 4.1.2), um einen gesicherten und konsistenten Datensatz zu garantieren.



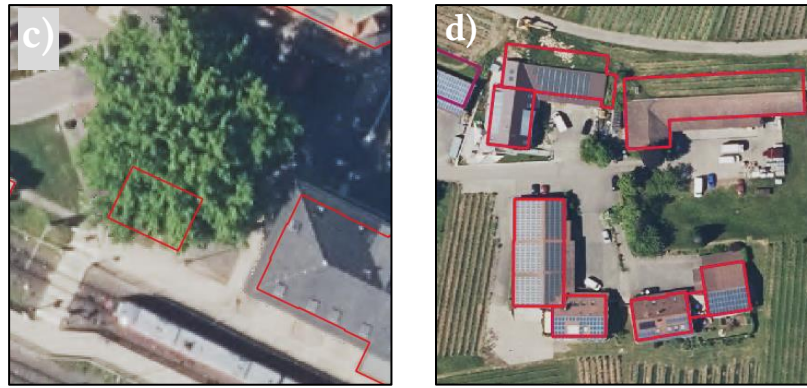


Abbildung 25 - Beispiele von potenziellen Fehlerquellen der ALKIS-Umränge für den Einsatz zur Gebäudeobjektdektion a) fehlendes Gebäude b) Gebäudeeinmessung stimmt nicht mit Bildaufnahme überein c) Baum verdeckt Gebäudeobjekt d) Gebäudeeinmessung größer als das tatsächliche Gebäude

Die Beseitigung solch fehlerhafter Daten aus dem Referenzdatensatz erfolgt zweistufig. Im ersten Schritt wird ein halbautomatisches Verfahren auf das Bearbeitungsgebiet angewendet, um fehlerbehaftete Umränge der Gebäudeobjekte zu erkennen und zu entfernen. Anschließend erfolgt eine zusätzliche manuelle Kontrolle, um sicherzustellen, dass Mängel, die von der automatischen Kontrolle übersehen wurden, abschließend aus dem Datensatz entfernt werden.

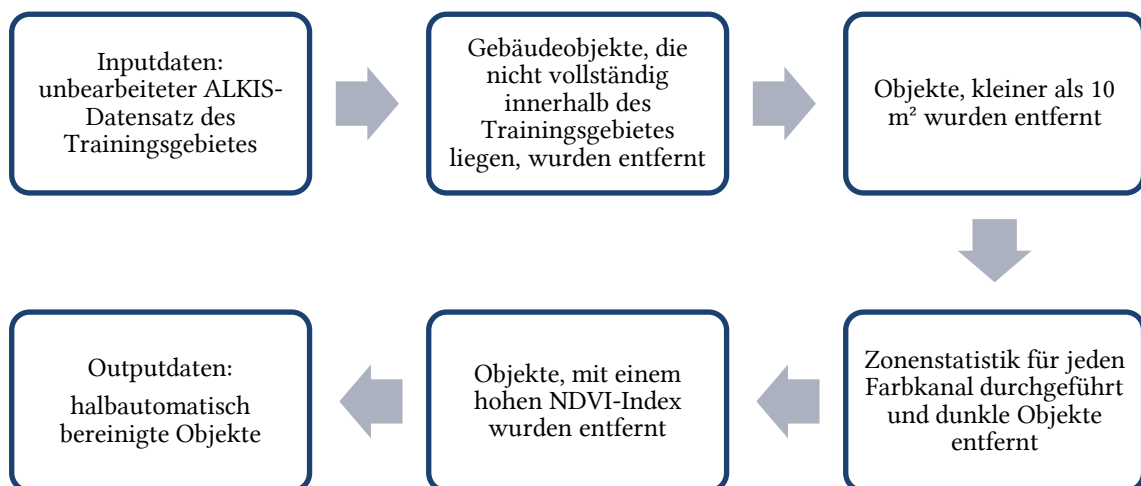


Abbildung 26 - Workflow für die halbautomatische Kontrolle der ALKIS-Daten für das Trainingsgebiet

Abbildung 26 zeigt den Workflow für die halbautomatische Datenkontrolle der ALKIS-Gebäudeobjekte: Der unbearbeitete Datensatz aus ALKIS für das Trainingsgebiet dient als Inputdaten. In einem ersten Schritt werden Objekte im Randbereich entfernt, deren Umrang sich nicht vollständig innerhalb des Trainingsgebietes befindet. Im

zweiten Schritt werden Gebäudeobjekte jeglicher Art entfernt, die unter der Erkennungsschwelle der *PlanetScope*-Daten liegen. Diese wurde auf einen Flächeninhalt von 10 m² definiert, was ungefähr der Abmessung eines 1 x 1 Pixels der *PlanetScope*-Daten entspricht (Pixelgröße: 3 m). Dadurch reduziert sich der Trainingsdatensatz um 841 Objekte. Im dritten Schritt wird innerhalb jedes Objektumrings eine Zonenstatistik für den Rot- und NIR-Kanal durchgeführt. Dabei werden die folgenden Statistiktypen berechnet: der Median (*median*), die Standardabweichung (*std*), der Bereich (*zonal range*), das Minimum (*minimum*), das Maximum (*maximum*), die Mehrheit (*majority*) und die Minderheit (*minority*), der in den jeweiligen Objekten befindlichen Farbwerte. Anschließend werden dunkle Objekte, deren Standardabweichung im roten Kanal < 10 sowie deren Median im roten Kanal < 50 liegt, entfernt. Zusätzlich werden Objekte, deren Schwellwert für die *zonal range* des roten Kanals < 50 und die *majority* im roten Kanal < 50 aus dem Datensatz entfernt. Als letzten Schritt werden die Objekte mit einem hohen NDVI-Wert entfernt. Weitere Schwellwerte können je nach Bedarf angewendet werden, wurden aber für diese Thesis nicht genutzt.

Tabelle 9 – Anzahl der noch zur Verfügung stehenden Gebäudeobjekte im Trainingsgebiet nach jedem Bereinigungsschritt

Bereinigungsschritt	Anzahl der vorhandenen Gebäudeobjekte (Anzahl der entfernten Objekte)
Inputdaten	28.333
Entfernen der Randobjekte	27.943 (-390)
Entfernen der Objekte kleiner als 10 m²	27.102 (-841)
Entfernen von dunklen Objekten	26.931 (-171)
Entfernen von begrüntem Objekten	26.598 (-333)
Durch halbautomatische Kontrolle entfernte Objekte: 1.735	
Zusätzliche manuelle Kontrolle	26.103
Durch manuelle Kontrolle entfernte Objekte: 495	

Es verbleiben nach der Bereinigung des Trainingsgebietes insgesamt 26.103 Gebäudeobjekte.

Der zeitliche Aufwand verhindert eine Kontrolle aller im Untersuchungsgebiet befindlichen Gebäudeobjekte. Aus diesem Grund wird der oben aufgeführte Schritt nur isoliert für das Trainingsgebiet und analog dazu zusätzlich für das im Kapitel 4.1.2 aufgeführten Kontrollgebiet durchgeführt. Der Unterschied bei den Kontrolldaten ist die zusätzliche manuelle Erfassung, der sich noch im laufenden Baugenehmigungsverfahren

ren befindlichen, baulich aber bereits abgeschlossenen Gebäudeobjekte zu den bereits eingemessenen Gebäudeobjekten im Kontrollgebiet (vgl. Abbildung 27). Von ursprünglich 4.290 unbereinigten Gebäudeobjekten innerhalb des Kontrollgebietes stehen nach der halbautomatischen und manuellen Bereinigung noch insgesamt 4.014 Gebäudeobjekte als Kontrollobjekte zur Verfügung.



Abbildung 27 - Kombination der ALKIS-Daten (rote Umrandung) und der im Bau befindlichen Objekte (blaue Umrandung) in der DOP-Darstellung

4.3 Aufbau der Trainingsdaten

Wie bereits in Kapitel 3.2 beschrieben, benötigt der Algorithmus bei einer überwachten Klassifizierung Trainingsdaten, um daraus relevante Merkmale ableiten zu können.

Der folgende Abschnitt 4.3 beschreibt den Workflow zur Generierung der Trainingsdaten für die Klassen:

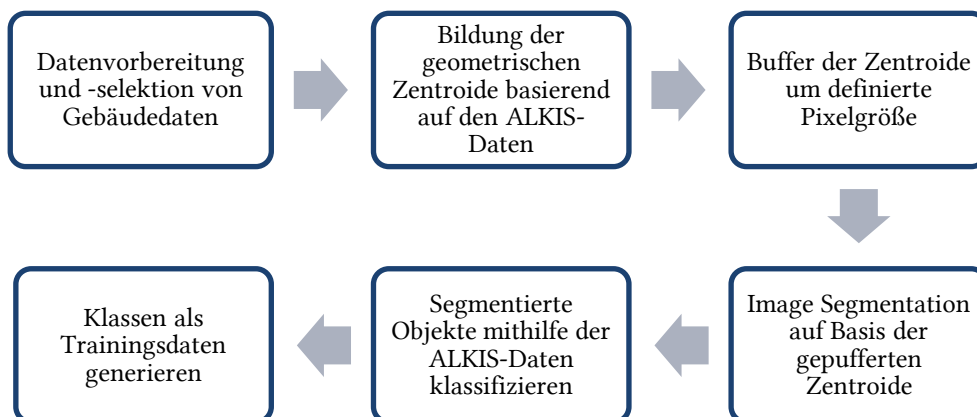


Abbildung 28 – Workflow für die Erstellung der Trainingsdaten

Abbildung 28 verdeutlicht den angewandten Workflow für die Generierung der Trainingsdaten. Es folgt eine Kurzzusammenfassung, die im Laufe des Abschnitts ausführlich erläutert wird: Zuerst wird der bereinigte ALKIS-Datensatz des vorherigen Abschnitts 4.2 genutzt und die geometrischen Zentroide der jeweiligen Objekte bestimmt. Auf Grundlage dieser Zentroide wird mit einer festgelegten Pixelgröße (6 Pixel bei DOP und 1 Pixel bei SAT) eine Fläche durch eine Pufferfunktion generiert und darauf eine *Image Segmentation* erstellt. Diese Schritte ermöglichen es eCognition die segmentierten *image objects* mithilfe einer thematischen Karte zu klassifizieren. Als Ergebnis befinden sich nun über das Trainingsgebiet verteilt klassifizierte Bildsegmente der zu trainierenden Klassen.

Das Trainingsgebiet beinhaltet insgesamt etwas mehr als 26.000 bereinigte Gebäudeobjekte auf einer Fläche von 30 km² (vgl. Abschnitt 4.2). Die Erstellung einer großen Menge an Trainingsdaten ist sehr kosten- und zeitintensiv (Tuia et al. 2011; Liu et al. 2020). Laut Blaschke et al. (2014) sind Objektklassen mit einer einzigartigen spektralen Signatur relativ „trivial“ zu klassifizieren, bei Objektklassen die ähnliche spektrale Signaturen als andere Klassen aufweisen, erweist sich eine eindeutige Klassenzuordnung oft als schwierig. Eine große Herausforderung bei der Nutzung von hochauflösenden Bildern für Anwendungen wie der Landbedeckungskartierung ist der hohe Grad an Komplexität städtischer Merkmale, bezogen auf spektrale, räumliche und strukturelle Eigenschaften (Blaschke 2010). In diese Gruppe fallen auch Gebäudeobjekte, deren spektrale und räumliche Eigenschaften, die innerhalb einer städtischen Fläche stark voneinander abweichen können (Myint et al. 2011). Außerdem zeichnen sich Gebäudeobjekte durch verschiedene Dachformen, -texturen und -farben aus, was zu einer deutlichen Varianz innerhalb der Klasse führen wird (Chen et al. 2019). Gleichzeitig führt die massenhafte Erstellung von zahlreichen *Samples*-Daten nicht automatisch zu besseren Ergebnissen, da verschiedene Nachbarschaftspixel dieselben Informationen beinhalten und es dadurch zu redundanten Daten kommt (Tuia et al. 2011). Graf und Wegenkittl (2012) weisen darauf hin, dass im Fall der überwachten Klassifizierung, keine allgemeingültigen Regeln existiert, wie viele Datensätze pro Klasse notwendig sind, damit zwei oder mehrere Klassen separierbar sind, da auch Faktoren wie die beschreibenden Merkmale, die Komplexität der Klassengrenzen und der eingesetzte Klassifikator wichtig sind. Bei der Erkennung von spezifischen Zielobjekten in einem Fernerkundungsdatensatz spielen auch die Größe und die Ausprägung eine wichtige Rolle. Besteht ein Fokus auf größeren Objekten wird ein größeres *Sample Patch*-Fenster benötigt, das allerdings bei kleinen Objekten zu Herausforderungen führen kann (Liu et al. 2019). Somit ist das Ziel eine *Sample Patch*-Größe zu wählen, die es ermöglicht, sowohl kleine Zielobjekte wie Schuppen und Gartenhäuser, wie auch großflächige Objekte zu klassifizieren.

Für die Bestimmung der optimalen Größe des *Sample Patch*-Fensters wird innerhalb des Untersuchungsgebietes der Flächeninhalt der jeweiligen ALKIS-Objekte berechnet und daraus der Mittelwert, Median und die Standardabweichung gebildet.

Tabelle 10 - Auflistung der fünf meistvorkommenden Gebäudeobjekte nach Objektart innerhalb des Untersuchungsgebietes

GFK	Art des Gebäudeobjekts	Gesamtanzahl	Flächeninhalt (m ²)		
			mean	median	stddev
31001_1010	Wohnhaus	66.564	118,5	105,6	63,7
31001_2463	Garage	56.313	31,5	24,7	26,7
31001_2723	Schuppen	8.378	32,5	21,9	36,2
51009_1610	Überdachung	4.232	86,6	28,2	234,2
31001_2112	Betriebsgebäude	3.520	777,2	200	2.500,5

Tabelle 10 zeigt für die Objektart Wohnhaus im Untersuchungsgebiet eine durchschnittliche Fläche von 118,5 m² mit einer Standardabweichung von 63,7 m² an. Der Median liegt mit 105,6 m² nahe dem Mittelwert. Die zweit- und drittgrößten Objektklassen Garage und Schuppen haben einen Mittelwert von 31,5 m², respektive 32,5 m² und besitzen eine Standardabweichung von 26,7 m² bzw. 36,2 m², was für eine große Streuung des Flächeninhalts der Objekte spricht. Besonders hervorzuheben ist, dass der halbe Datensatz der Objektart Garage einen kleineren Flächeninhalt als 26,7 m² aufweist. Somit ist ein nicht unmerklicher Teil der Gebäudeobjekte im Größenbereich von ungefähr 20 m².

Ghorbanzadeh et al. (2021) schreiben in Bezug auf die Detektion von Unterkünften in Flüchtlingscamps, dass bei der Wahl einer zu großen *Sample Patch*-Größe, das CNN dazu verleitet werden kann, ein größeres Gebiet aus mehreren zusammenhängenden Zielobjekten mit geringen Abständen als ein einziges Wohngebäude zu klassifizieren. Mit der Bildung einer zweiten Zielklasse für Objekte kleiner als 20 m² wird deswegen in dieser Thesis der Versuch unternommen, diesen Umstand im Fall der hochauflösenden DOP-Daten zu umgehen. Zusätzlich liegt ein besonderes Interesse des Vermessungsamtes Heilbronn an der Erkennung von kleinen Strukturen, wie Anbauten, Garagen und Überdachungen neben der Detektion von Hauptgebäuden. Die Topologie der Verteilung der ALKIS-Objekte zeichnet ein wechselndes Bild zwischen der Zielklasse „Gebäudeobjekte“, mit Wohngebäude und Industrien und der zweiten Zielklasse „Kleine Objekte“ mit den Überdachungen, Garagen und Carports.

Somit wird für das Training des Modells für die DOP-Daten zusätzlich zur Zielklasse „Gebäudeobjekte“ und der Nicht-Zielklasse „Nicht-Gebäude“, eine Ziel-Klasse „Kleine Objekte“ für Gebäude kleiner als 20 m² definiert.

Für die zu untersuchenden Satellitendaten wird aufgrund der geringeren geometrischen Auflösung die Trennung der beiden Zielklassen nicht durchgeführt und nur die Zielklasse „Gebäudeobjekte“ und die Nicht-Zielklasse „Nicht-Gebäude“ definiert. Die Objekte mit einer Fläche kleiner als 20 m² finden sich im Fall der SAT-Daten in der gemeinsamen Zielklasse „Gebäudeobjekte“. Dadurch erhöht sich nun auch die Anzahl der für den Aufbau der Trainingsdaten zur Verfügung stehenden *Samples* für die Klasse (vgl. Tabelle 11).

Die Nicht-Zielklasse beinhaltet dabei jegliche Landschaftselemente, die nicht zu den Gebäudeobjekten aus ALKIS gezählt werden, etwa Objekte der Landbedeckungen Gewässer, Vegetation oder Infrastruktur (Tiefbau) und Landwirtschaft.

Tabelle 11 – Darstellung der zur Verfügung stehenden *Samples* im Trainingsgebiet für DOP und SAT (gerundet)

Name der Klasse	Anzahl an Samples	
	DOP	SAT
Gebäudeobjekte	20.000	25.000
Kleine Objekte	5.000	
Nicht-Gebäude	25.000	25.000

Für die Generierung der Trainingsdaten für die DOP wurde eine *Sample Patch*-Größe von 128x128 Pixel gewählt. Das entspricht einer geometrischen Abmessung von 25,6 m Kantenlänge (bei 0,2 m/px). So kann ein quadratisches Objekt mit einer theoretischen Dachfläche von ungefähr 655 m² rechnerisch vollständig mit einem *Sample* erfasst werden. Damit kann sichergestellt werden, dass der Großteil der unter Tabelle 10 aufgeführten Gebäudeobjekte vollständig abgebildet wird.

Bei den Satellitendaten wird ein ähnlicher Ansatz verfolgt. Unter Berücksichtigung der Pixelgröße von 3 m wird in diesem Fall eine *Sample Patch*-Größe von 8x8 Pixel gewählt. Das entspricht im Falle der *PlanetScope*-Daten einer Kantenlänge von 24 m (bei 3 m/px) und ermöglicht die vollständige Abdeckung von Dachflächen der satellitengestützten Bilddaten mit einer theoretischen Fläche von 576 m².

Neben den auf Basis des ALKIS-Datensatzes erstellten Trainingsdaten zu den Klassen „Gebäudeobjekte“ und „Kleine Objekte“ werden die Trainingsdaten der Nicht-Zielklasse gesondert abgeleitet. Die Überlegung dabei ist, dass jedem Bildpixel, welches nicht einer der gesuchten Zielklassen zugeordnet wird, ein potenzielles *Sample* für die Nicht-Zielklasse ist. Für die Bildung der *Samples* der Nicht-Zielklasse wird ein regelmäßiges Gitterraster (128x128 Pixel für DOP und 8x8 Pixel für SAT) über das Trainingsgebiet generiert. In Abhängigkeit der aufbereitenden ALKIS-Gebäudeobjekte werden ausschließlich die Kacheln des Gitterrasters beibehalten, die keine Überschneidung mit Gebäudeobjekten aufweisen.

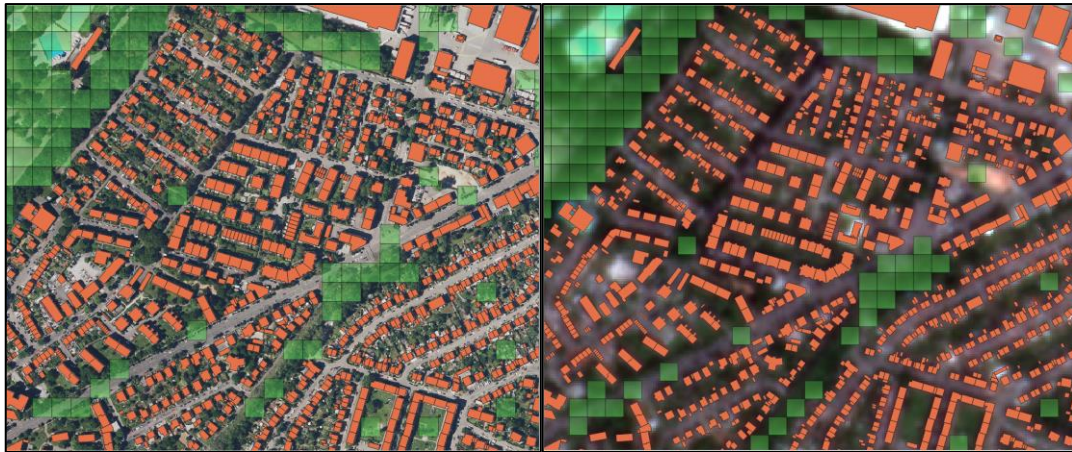


Abbildung 29 - Trainingsdaten für die Zielklasse Nicht-Gebäude (grün) für DOP (links) und SAT (rechts). ALKIS-Gebäudeobjekte sind in orange dargestellt.

Tabelle 12 zeigt die *Samples* nach der Erstellung in eCognition für DOP und SAT, wobei hier jeweils die RGB-Echtfarbendarstellungen gewählt wurden. Der Prozess der Erstellung der *Samples* nutzt grundsätzlich alle verfügbaren Bildkanäle des jeweiligen Rasterdatensatzes: 4 Kanäle für DOP und 8 Kanäle für SAT. Weitere Beispiele des *Samples*-Ordners befinden sich im Anhang.

Tabelle 12 - Vergleich der *Samples* aus dem DOP- und dem SAT-Datensatz

	DOP (128 x 128 px)	SAT (8 x 8 px)
Gebäudeobjekte		
Kleine Objekte		
Nicht-Gebäudeobjekte		

Diese entwickelte Methodik entspricht in den Grundzügen der Methode von Vakalopoulou et al. (2015), in der ein binäres Problem in Form von Gebäude oder Nicht-Gebäude zu lösen ist. Die *Sample Patches* der Gebäude werden mithilfe von Referenzdaten zentriert auf den Mittelpunkt eines Gebäudes gebildet. Die Erstellung der *Sample Patches* der Nicht-Gebäude erfolgt rein zufällig.

Data Augmentation

Um komplexe Netzwerkstrukturen zu entwickeln, wird eine große Menge an Trainingsdaten benötigt, damit die höchste Klassifizierungsgenauigkeit erreicht werden kann (Scott et al. 2017). Eine effektive Methode, um die Anzahl an Trainingsdaten zu

erhöhen, ist die *Data Augmentation*. Dabei wird die Menge an Trainingsdaten ausschließlich mit Informationen aus den Trainingsdaten selbst erhöht (Perez und Wang 2017). Eine *Data Augmentation* vermeidet eine Überanpassung (*Overfitting*) des Modells (Radovic et al. 2017) und führt zu einer höheren Erkennungsqualität (Shorten und Khoshgoftaar 2019). Die als *Traditional Data Augmentation* beschriebene Methode umfasst verschiedene Transformationen der Trainingsdaten, etwa die Rotation der Bilder mit verschiedenen Winkeln (90° , 180° , 270°), die horizontale oder vertikale Spiegelung, die Farbanpassung in Form von Helligkeitswerten, die Skalierung und Streckung der Bilder mit verschiedenen Faktoren oder die Verzerrung der Bilder mit verschiedenen Winkeln oder einem Faktor (Han et al. 2018). Im Bereich der Fernerkundung sind solche Transformationen prinzipiell anwendbar, da beispielsweise eine Drehung eines Bildobjektes aus einer orthogonalen Luftaufnahme keinen Einfluss auf die Bildorientierung und die Perspektive der zu beobachtenden Klasse hat, während im Gegensatz dazu, beispielsweise aufgenommene Bilder auf Bodenhöhe nicht ohne weiteres transformiert werden können (Beispiel: der Himmel ist oben) (Scott et al. 2017).



Abbildung 30 - Beispieldarstellung der angewandten *Data Augmentation* mithilfe von Rotation, Spiegelung und Helligkeitsanpassung

Um eine Überanpassung zu vermeiden, werden für die Untersuchung der Thesis drei Transformationen auf die *Training Samples* der Daten der DOP- und *PlanetScope*-Daten angewendet: Rotation, Spiegelung und eine Helligkeitsanpassung.

a) Rotation

Eine Rotation der *Samples* konnte innerhalb der Software eCognition direkt während der Generierung der *Samples* angewendet werden. Abbildung 31 visualisiert den eingesetzten *Loop-Workflow*: Während des ersten Durchganges wird die voreingestellte Menge der *Samples* je Klasse erstellt. Nachdem dieser Durchgang mit der Erstellung der *Samples* abgeschlossen ist, wird das Trainingsgebiet gedreht (hier: 90° im Uhrzeigersinn für DOP und 45° für SAT) und die Generierung der *Samples* beginnt erneut. Dieser Prozess wiederholt sich für alle angegebenen Drehwinkel des Arrays. Im Fall der Thesis sind das neben der nicht rotierten Ausgangsvariante mit 0°, zusätzlich die Drehwinkel 90°, 180° und 270° für DOP und 45°, 90°, 135°, 180°, 225°, 270° und 315° für die SAT-Daten. Wegen der geringeren Datengröße der *Samples* für SAT, ist es möglich mehr Transformationsfaktoren zu verwenden, ohne dabei die Datenmenge der hochauflösenden DOP-Daten zu erreichen. Weitere Durchläufe der Schleife können bei Bedarf erstellt werden. Nachdem die Drehwinkel im Array durchlaufen sind, endet die Schleife.

Mit dem gesetzten Parameter des Arrays konnte im ersten Schritt eine Vervielfachung für die DOP-Daten bzw. eine Verachtfachung für die SAT-Daten der ursprünglichen angegebenen *Training Samples* erreicht werden.

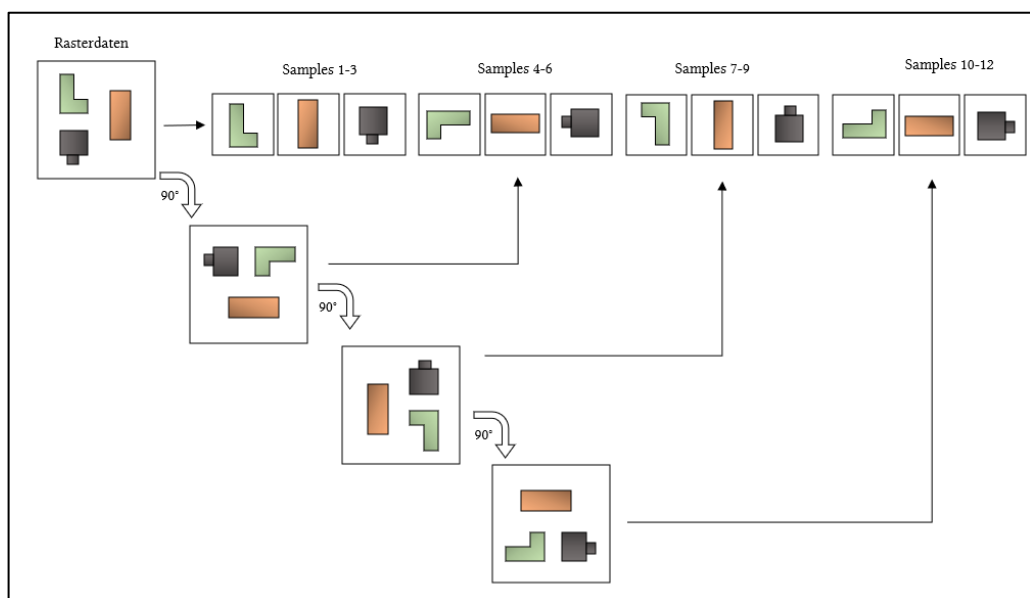


Abbildung 31 - visuelle Repräsentation der angewendeten Schleife für die Rotation der *Training Samples* des DOP-Datensatzes

eCognition bietet bei der Erstellung der *Samples* neben dem *raw*-Format auch den Export von *tif*-Dateien der erstellten *Sample Patches* an (Trimble 2020). Dadurch ist es möglich, die Daten auch außerhalb der Softwareumgebung weiterzuverarbeiten. Für die nächste Transformation wird diese Möglichkeit genutzt. Die Spiegelung und die Helligkeitsanpassungen der *Training Samples* müssen außerhalb von eCognition erfol-

gen, da keine derartigen Funktionen in der Software implementiert sind. Aus diesem Grund wurden entsprechende Workflows in FME entwickelt.

b) Spiegelung

Auf den bereits vierfach bzw. achtfach vergrößerten Datensätzen wird nun zusätzlich eine horizontale Spiegelung durchgeführt, was eine weitere Verdopplung der *Training Samples* zur Folge hat. Eine zusätzliche vertikale Spiegelung der Datensätze findet nicht statt, da die Kombination einer horizontalen und vertikalen Spiegelung im Anschluss an eine mehrfache Rotation zu Duplikaten führt und keinen Mehrgewinn bringt. Abbildung 32 visualisiert diesen Umstand schematisch.

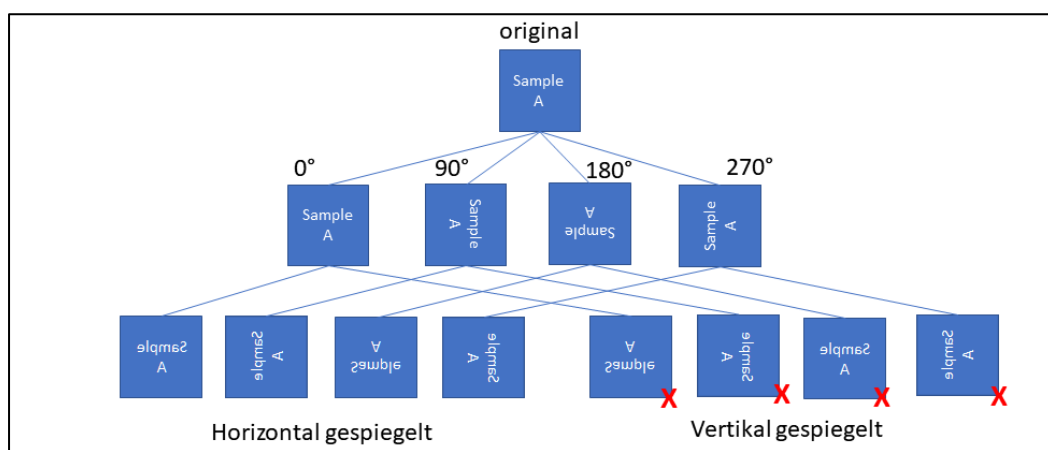


Abbildung 32 - Duplikate bei horizontaler und vertikaler Spiegelung nach vorheriger Drehung

Bezogen auf die Rechenzeit benötigt der Transformationsprozess für die Erweiterung der *Samples* um einen Faktor 2 ungefähr 12-15 Stunden bei den DOP-Daten, respektive Faktor 3 Stunden bei den Satellitendaten. Der Workflow prozessiert jeweils ein *Sample* und überschreibt dabei das Original mit einer unveränderten und einer horizontal gespiegelten Variante. Dadurch wird eine fortlaufende Nummerierung ohne Dopplung der *Samples* sichergestellt.

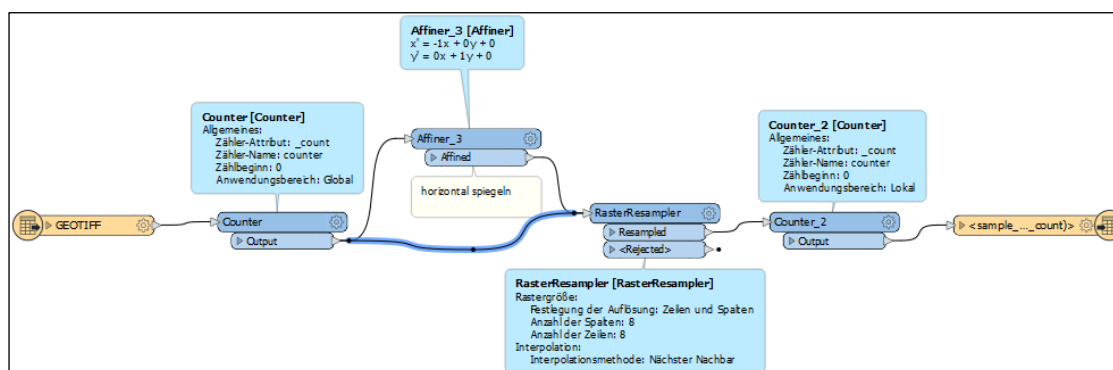


Abbildung 33 – FME-Workbench zur Spiegelung der *Training Samples* für SAT-Daten

c) Helligkeitsveränderung

Der FME-Workbench für die Helligkeitsanpassung verändert den Farbwert jeden Kanals eines multispektralen Bildes, um einen festgelegten Wert. In der Thesis wurden Helligkeitsveränderungen von +5 %, +10 % und -5 % in Abhängigkeit des ursprünglichen Helligkeitswertes des Eingangsbildes getestet. Für die SAT-Daten wird die dritte Transformation nicht angewendet, da die Anpassung des Farbwertes bei Aufnahmen mit 8-Kanälen in FME nicht umsetzbar war.

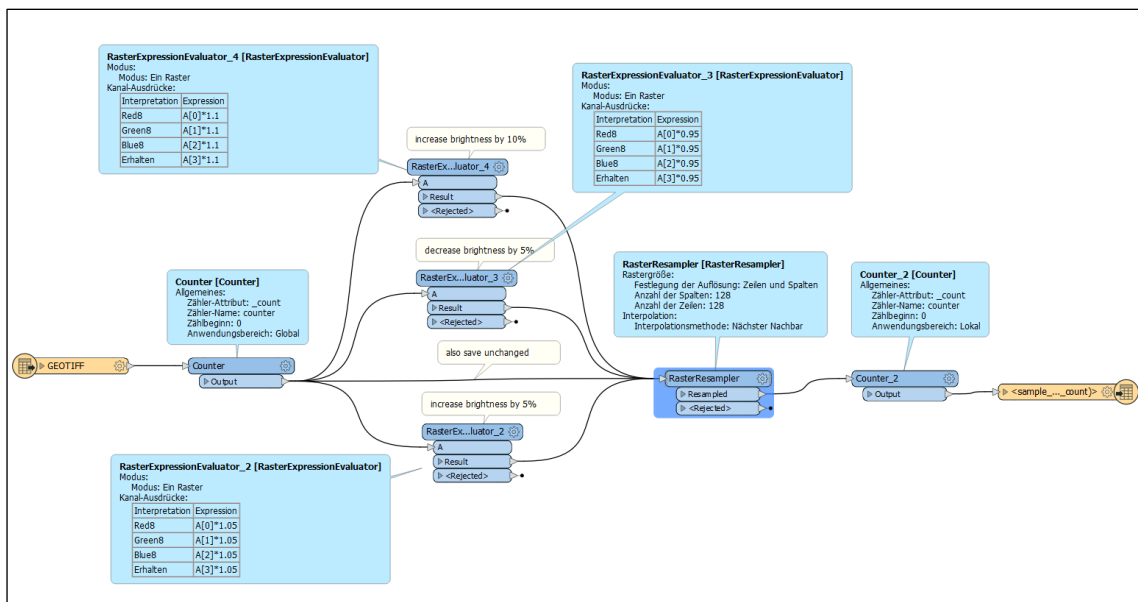


Abbildung 34 – FME-Workbench zur Helligkeitsanpassung der *Training Samples* für DOP

Die *Data Augmentation* wurde für den DOP-Datensatz mit den zwei Transformationen Rotation und Spiegelung für die Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“ durchgeführt, sowie einer zusätzlichen Transformation „Helligkeitsanpassung“ für die dritte Klasse „Kleine Objekte“ umfasst. Die Helligkeitsanpassung wurde nur für die *Samples* der Klasse „Kleine Objekte“ durchgeführt, um eine gleichverteilte Anzahl an *Samples* in den drei Klassen zu gewährleisten. Mit der gleichen Gesamtanzahl der erweiterten *Samples* soll sichergestellt werden, dass innerhalb eines Trainingsdatensatzes keine Klasse eine vorteilhafte Gewichtung erfährt und andere Klassen dadurch in den Trainingsdurchläufen weniger berücksichtigt werden.

Die Transformationen der beiden Klassen für Satellitendaten umfassen die Rotation und die Spiegelung. Die geringere Auflösung der Satellitendaten und die damit verbundene kleinere Datengröße der *Samples* ermöglicht einen erweiterten Array des Drehwinkels von 45°, was zu einem höheren Multiplikator führt. (vgl. Tabelle 13).

Tabelle 13 – Anzahl der verwendeten *Training Samples* bevor und nach der *Data Augmentation* für DOP und SAT mit dem Transformationsmultiplikator

	Klasse	Anzahl Ausgangslage	Rotation	Spiegelung	Helligkeitsanpassung	Anzahl Ergebnis
DOP	Gebäudeobjekte	20.000	x4	x2	n/a	160.000
	Kleine Objekte	5.000	x4	x2	x4	160.000
	Nicht-Gebäude	20.000	x4	x2	n/a	160.000
SAT	Gebäudeobjekte	25.000	x8	x2	n/a	400.000
	Nicht-Gebäude	25.000	x8	x2	n/a	400.000

Um eine gleiche Gesamtanzahl an Samples für jede Klasse innerhalb eines untersuchten Rasterdatensatzes zu erreichen, wurde die Menge der Eingangssamples entsprechend angepasst. Es stehen somit jeweils 160.000 „Gebäudeobjekte“, „Kleine Objekte“ und „Nicht-Gebäude“ für die DOP- und jeweils 400.000 „Gebäudeobjekte“ und „Nicht-Gebäude“ für die SAT-Daten für das Training des Neuronalen Netzes zur Verfügung.

Tabelle 14 - Vergleich der Kennzahlen der *Training Samples* für DOP- und SAT-Daten

	DOP	SAT
Samples Patch-Größe	128 x 128 Pixel	8 x 8 Pixel
Reelle Abmessung der Samples	25,6 x 25,6 m	24 x 24 m
Datengröße eines Samples	70 KByte	2 KByte
Generierung der Samples	3 h	0,5 h
Transformationsdauer der Data Augmentation	12 h pro Klasse	3 h pro Klasse
Speicherplatz	10,7 GByte pro Klasse (bei 160.000 Samples)	0,8 GByte pro Klasse (bei 400.000 Samples)
Vorteile	Höhere Genauigkeit	Höhere Zahl an Samples, schneller generiert

4.4 Kontrolldaten (Ground Truth)

Wie in Abschnitt 4.1.2 beschrieben, handelt es sich bei dem Kontrollgebiet um einen Teil der Gemeinden Weinsberg und Ellhofen mit einer Größe von 6 km². Neben den amtlich eingemessenen Gebäudeobjekten aus ALKIS, die sich innerhalb des Kontrollgebietes befinden, werden auch noch im laufenden Baugenehmigungsverfahren befindliche Gebäudeobjekte manuell dem *Ground-Truth*-Daten (GT-Daten) hinzugefügt. Abbildung 35 zeigt einen Ausschnitt von baulich abgeschlossenen Gebäudeobjekten, für die bisher keine Gebäudeeinmessung durchgeführt wurde. Zum Vergleich: Die rot

umrandeten Gebäudeobjekte im unteren rechten Bildbereich visualisieren amtlich erfasste Gebäudeobjekte aus dem ALKIS-Datenbestand.



Abbildung 35 – Bisher nicht eingemessenes Neubaugebiet in Ellhofen

Bevor die manuell zu erfassenden Objekte in den GT-Datensatz aufgenommen werden, wird die jeweilige Bauphase einer visuellen Prüfung unterzogen, um sicherzustellen, dass diese bereits abgeschlossen ist. Aktuell im Bau befindliche Gebäudeobjekte werden nicht im CNN trainiert und müssen folglich nicht detektiert werden.



Abbildung 36 - Darstellung der manuell erfassten Gebäudeobjekte gemäß Bauerfassungskataster (blau) im Vergleich des noch im Bau befindlichen Objektes in der Bildmitte

Die Erstellung der GT-Daten für die Klasse „Nicht-Gebäude“ wird auf dieselbe Weise umgesetzt, wie in Kapitel 4.2 beschrieben. Es wird allerdings nicht jede generierte Fläche des Gitterrasters für die Klasse „Nicht-Gebäude“ nach der Bereinigung als GT-

Daten verwendet. Stattdessen werden zufällige Flächen ausgewählt (20 % der maximalen Anzahl).

Um eine Vergleichbarkeit zwischen den *Ground Truth*-Daten der DOP- und der SAT-Daten sicherzustellen, werden die Kacheln des GT-Datenbestandes der Nicht-Zielklasse der DOP (128x128 Pixel), zur Generierung der Nicht-Zielklasse für die SAT-Daten (8x8 Pixel) genutzt. Dabei wird das geographische Zentrum jeder Kachel des DOP-Gitterrasters gebildet, auf dessen Basis anschließend ein quadratischer Buffer mit 8x8 Pixel generiert wird. Dieses neu erstellte Gitterraster stellt die GT-Daten der Nicht-Zielklasse für die SAT-Daten dar. Abbildung 37 zeigt Beispiele der Klasse „Nicht-Gebäude“ für die DOP-Daten, die per zufälligem Auswahlverfahren bestimmt wurden.



Abbildung 37 - Beispiele der GT-Klasse "Nicht-Gebäude" (blau)

Der GT-Datensatz besteht, analog zum Trainingsdatensatz der DOP, aus den Klassen „Gebäudeobjekte“, „Nicht-Gebäude“ und „Kleine Objekte“. In Summe sind das für das Kontrollgebiet 5.196 Objekte. Tabelle 15 zeigt die Klassen mit der jeweiligen Anzahl der Kontrollobjekte.

Tabelle 15 - Anzahl der GT-Daten pro Klasse

Klasse	Anzahl Ground-Truth-Daten
Gebäudeobjekte	3.387
Kleine Objekte	627
Nicht-Gebäude	1.182

4.5 CNN-Architektur

Die für diese Thesis verwendete Deep Learning-Technologie von eCognition basiert auf der Google TensorFlow-Library (Trimble 2020). Für das Training des Modells der flugzeuggestützten Daten wurden die Trainingsdaten der drei Klassen mit jeweils 160.000 *Samples* in einem ersten Schritt gemischt, damit nach der *Data Augmentation* zufällige *Samples* ausgewählt werden.

Ein wichtiger Trainingsparameter ist die *batch size*. Bei einer *batch size* von 1.000 werden pro Trainingsschritt 1.000 *Samples* genutzt. Im Fall der DOP-Daten mit 160.000 *Samples* sind somit 160 Trainingsschritte notwendig, um einen vollständigen Trainingsdurchlauf aller *Samples* zu erreichen (was einer Epoche entspricht). Den Trainingsdatensatz in einzelne *Batches* zu trennen, ermöglicht ein iteratives Training des Modells, so lange, bis alle *Samples* gesehen wurden (Kattenborn et al. 2021). Allgemein gesprochen ist es unwahrscheinlich, dass ein CNN nach einer Epoche bereits seine optimale Konfiguration erreicht, da in einem *batch*-Durchlauf vermutlich nicht alle nützlichen Merkmale abgebildet sind (Kattenborn et al. 2021).

Die Anzahl der Trainingsschritte wirkt sich signifikant auf die Dauer des Trainings aus (Trimble 2020). Im Falle eines Testlaufs wird abhängig von der Menge an *Samples* und der Zahl an Trainingsschritten eine Trainingslaufzeit von 1.800 bis 2.100 Minuten (oder 30-35 Stunden) benötigt. Da die verwendete Workstation das Netz nicht dediziert trainieren konnte, sondern parallel für andere Arbeiten verwendet wurde, kann die Laufzeit nur annähernd als Zeitraum angegeben werden. Ein weiterer Trainingsparameter, der gesetzt werden muss, ist die *learning rate*. Diese bestimmt wie stark die Gewichte bei jedem Trainingsschritt angepasst werden können (Trimble 2020) und wurde auf der Standardeinstellung 0,006 belassen.

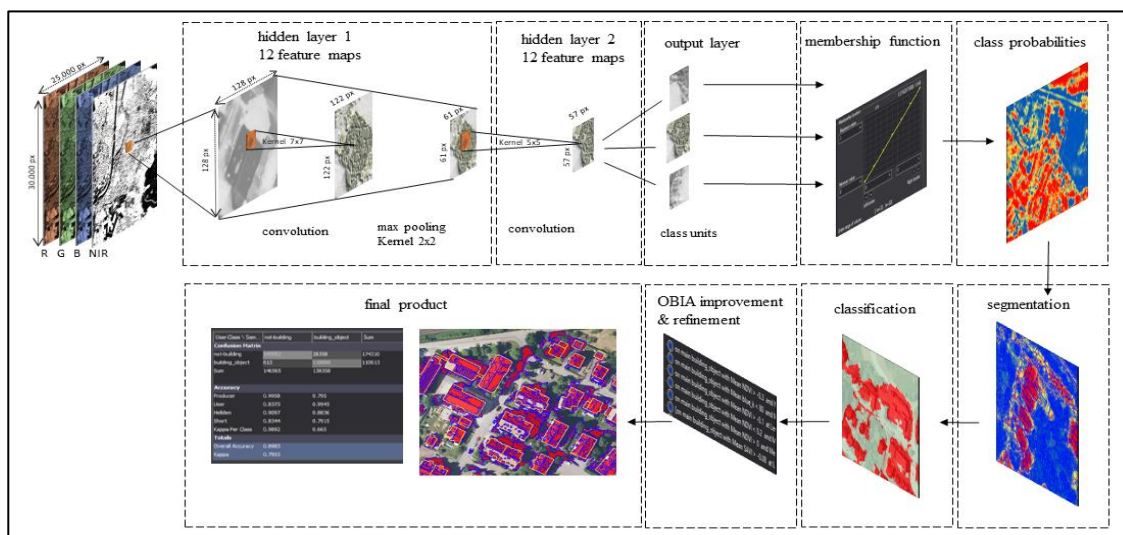


Abbildung 38 - Darstellung der Funktionsweise der unter eCognition verwendeten CNN-Architektur mit den entsprechenden Parametern für die DOP-Daten und der anschließenden objektbasierten Klassifikation (nach Ghorbanzadeh et al. (2021))

Der allgemeine Aufbau des CNN, mit dem das beschriebene Training für die DOP durchgeführt wird, ist in Abbildung 38 dargestellt. Es handelt sich dabei um eine Architektur mit zwei verborgenen Schichten. Auf die bereits beschriebenen 128x128 Pixel *Samples* der DOP-Daten wird in der *Convolution*-Schicht ein *Kernel* der Größe 7x7 Pixel angewendet und 12 *Feature Maps* gebildet. Durch ein *Pooling* wird die Größe reduziert. In der nächsten Schicht wird erneut ein *Kernel*, diesmal mit 5x5 Pixel auf das reduzierte *Sample* angewendet und wiederum 12 *Feature Maps* gebildet, die dabei extrahierten Merkmale werden in der Ausgabeschicht auf 3 Klassen verteilt. Die geringe *Kernel*-Größe ist in der Rechenzeit begründet. Größere *Kernel*-Größen erfordern höhere Rechenzeiten, die in Fall der Thesis oftmals in einem Programmabsturz resultierten. Der Grund für diese Abstürze wird in der begrenzten Speicherplatzkapazität der Workstation gesehen.

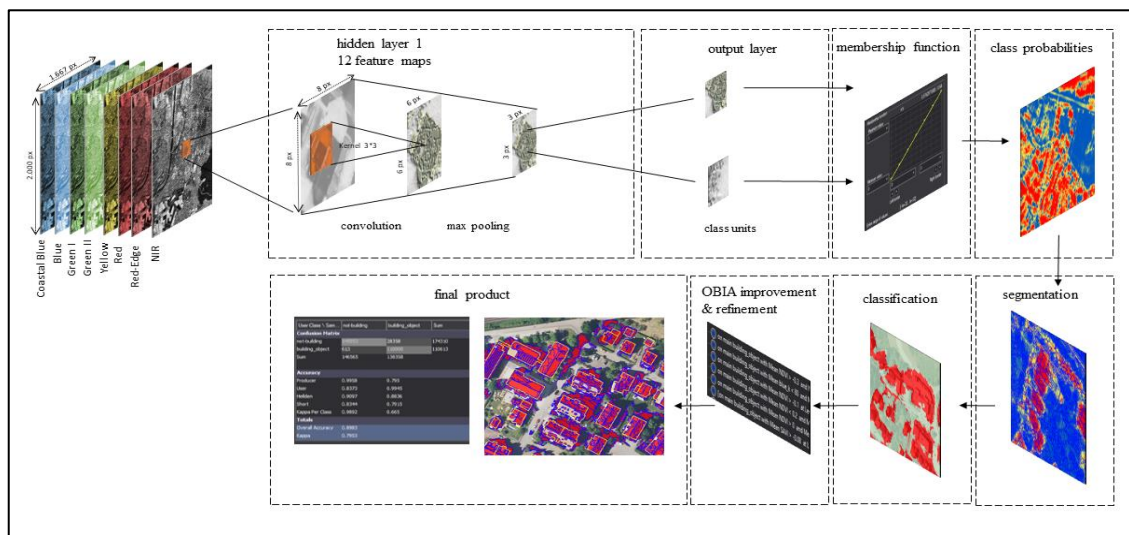


Abbildung 39 - Darstellung der Funktionsweise der unter eCognition verwendeten CNN-Architektur für SAT und objektbasierter Klassifikation (nach Ghorbanzadeh et al. (2021))

Für die Satellitendaten wird eine veränderte CNN-Architektur gewählt, da die *Samples* hier eine deutlich geringere Größe haben. Es wird ein CNN mit einem *Hidden Layer* verwendet und in einer Variante auf ein *max-pooling* verzichtet. Die Faltung wird mit einer *Kernel*-Größe von 2x2 bzw. 3x3 Pixel durchgeführt und 10 bzw. 12 *Feature Maps* gebildet. Außerdem werden auch für das Training verschiedene Varianten durchgeführt, etwa mit einer *batch size* von 200 bzw. 500 und einer Anzahl an Trainingsschritten von 6.000 bzw. 10.000.

Als Ergebnis wird für jedes Pixel eine Vorhersage des Wahrscheinlichkeitswertes der Zugehörigkeit zu den betrachteten Klassen generiert (Ghorbanzadeh et al. 2021) und von der Software als *Heatmap* ausgegeben. Anschließend erfolgt die Bildsegmentierung mittels MRS (Baatz und Schäpe 2000) zu Bildobjekten (*image objects*) basierend auf den Spektralbändern der jeweiligen Rasterdatensätzen. Die Bildobjekte werden

dann über eine überwachte Klassifizierungsstrategie den Zielklassen zugeordnet (Belgiu und Csillik 2018).

4.6 Segmentierung und Klassifikation

Abschnitt 4.6.1 thematisiert die Bestimmung der optimalen Segmentierungsparameter des MRS für die DOP- und SAT-Daten.

Der anschließende Abschnitt 4.6.2 befasst sich mit der Klassifizierung der erstellten segmentierten Bildobjekte basierend auf einer Kombination aus Nachbarschaftsbeziehungen, räumlicher Zuordnung, den Wahrscheinlichkeitsergebnissen des CNN-Modells und einer nachgeschalteten regelbasierten Anpassung. Dies soll zu einer Verbesserung der Ergebnisse im Vergleich zu einer unüberwachten Klassifikation führen (Ghorbanzadeh et al. 2021).

4.6.1 Abschätzung der Segmentierungsparameter für DOP und SAT

Nachdem der lernbasierte Klassifikationsalgorithmus eine Wahrscheinlichkeit über das Vorhandensein etwaiger Klassen im Gebiet ermittelt hat, folgt die Segmentierung des Gebietes mit dem MRS. Dazu bedarf es der richtigen Wahl des SP, der die Größe und Homogenität der resultierenden Objekte bestimmt (Belgiu und Csillik 2018). Der am häufigsten verwendete Ansatz, um die optimale Parametereinstellung herauszufinden, ist die Anwendung von heuristischen Methoden wie *trial and error* (El-naggar 2018). Die Parameter werden gemäß bisherigen Erfahrungen festgelegt (Ma et al. 2017) und die Bewertung über die Genauigkeit basiert anschließend auf visuellen Faktoren (Whiteside et al. 2011).

Um diese sehr zeitintensive Arbeitsweise, die auf subjektiven Bewertungskriterien beruht, zu verbessern, entwickelten Drăguț et al. (2010) das Werkzeug „*Estimation of Scale Parameter*“ (ESP), mit dem die optimalen Parametereinstellungen einer Segmentierung automatisiert bestimmt werden können. Mit dem Tool ESP II folgte 2014 der Nachfolger der Entwicklungen von Drăguț et al. (2014), dessen Konzept auf der lokalen Varianz (LV) (Woodcock und Strahler 1987) über verschiedene Skalen hinweg beruht, um automatisch drei geeignete SP für MRS zu identifizieren (Belgiu und Csillik 2018). ESP/ESP II implementiert dabei einen automatisierten Prozess der Bildsegmentierung für drei standardisierte Segmentierungsgrößen (1, 10 und 100), wobei der Nutzer entscheiden kann, ob ein hierarchischer oder ein nicht-hierarchischer Ansatz verfolgt werden soll (Drăguț et al. 2010). Drăguț et al. (2010) führen weiter aus, dass bei der Wahl für eine hierarchische Herangehensweise jede Objektebene basierend auf der vorherigen Segmentierung abgeleitet wird und den Segmentierungsprozess somit stark beeinflusst. Die Bewertung dieses datengestützten, unüberwachten Segmentierungsansatzes basiert ausschließlich auf Bildstatistiken (Belgiu und Csillik 2018) und

die Ermittlung eines optimalen SP funktioniert daher unabhängig von Referenzdaten (Li et al. 2015).

Die Daten werden dabei automatisch von ESP mit festen Skalierungsparameter segmentiert und die lokale Varianz als durchschnittliche Standardabweichung der Objekte für jede durch die Segmentierung erhaltene Objektebene berechnet (Drăguț et al. 2010). Eine zweite wichtige Kennzahl ist die Änderungsrate (*Rate of Change* (ROC)), diese misst die Dynamik der LV von einer Objektebene zu einer anderen (Drăguț et al. 2010). In ihrer Publikation gehen Drăguț et al. (2010) davon aus, dass der Höhepunkt des ROC-LV-Diagramms die Objektebene anzeigt, in denen die Szenenelemente am besten segmentiert wurden. Somit visualisiert die Kombination der Kurven *ROC* und *LV* einen Schwellenwert als ein Abbrechen der Kurve nach einem kontinuierlichen Abfallen (Drăguț et al. (2010)).

In dieser Arbeit wurde der hierarchische Ansatz angewendet. Dabei wurde der Parameter *Shape* auf 0.1 und der Parameter *Compact* auf 0,5 gesetzt und 50 *Loops* (DOP) bzw. 100 *Loops* (SAT) durchgeführt. Die Standardparameter der drei Objektebenen wurden bei 1, 10 und 100 belassen. Zusätzlich wurde eine txt-Datei nach dem Prozess ausgegeben, die eine Visualisierung des Graphen ermöglicht (vgl. Abbildung 40 und Abbildung 41).

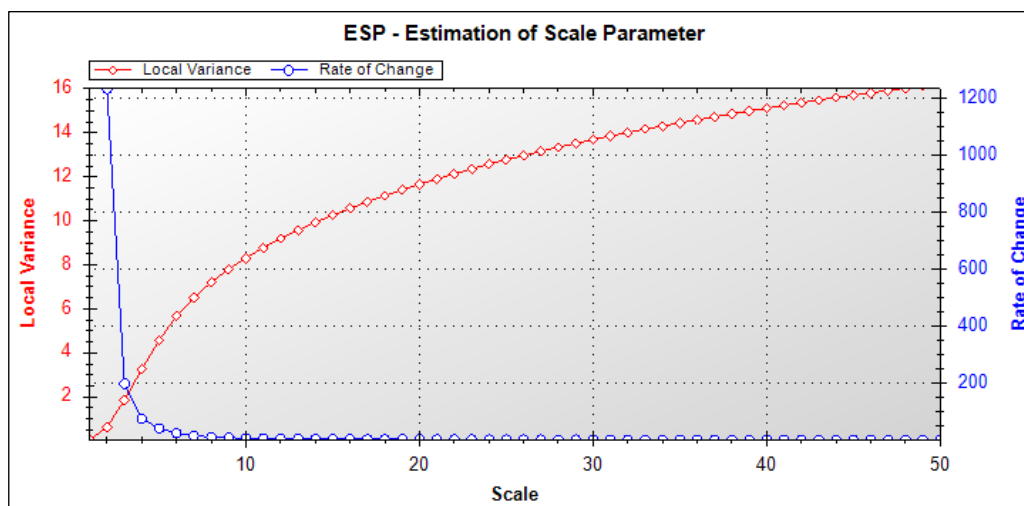


Abbildung 40 - Darstellung der Abschätzung der Segmentierungsparametern für DOP mit der Varianz (rot) und der Veränderungsrate (blau)

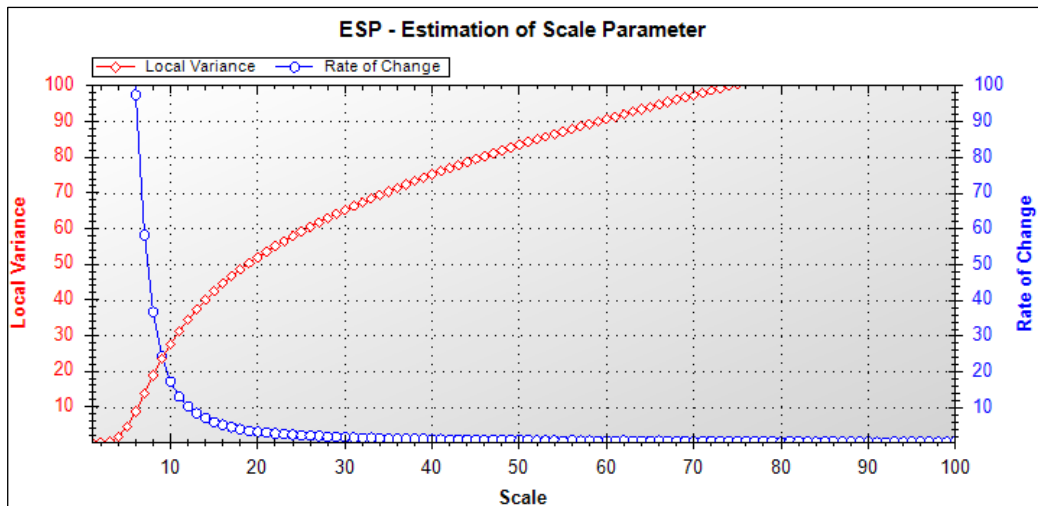
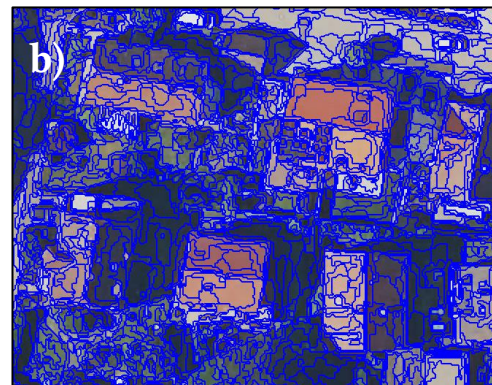


Abbildung 41 - Darstellung der Abschätzung der Segmentierungsparametern für *PlanetScope*-Daten mit der Varianz (rot) und der Veränderungsrate (blau)

ESP2 zeigt ein starkes Abfallen der ROC-Kurve, sowohl für die DOP- (vgl. Abbildung 40), als auch für die *PlanetScope*-Daten (vgl. Abbildung 41). ROC schneidet die LV-Kurve bei 4 (DOP) bzw. bei 7 (SAT), zeigt aber in beiden Graphen keinen Peak nach dem starken Abfallen, sondern läuft kontinuierlich gegen 0.

Die Auswertungen mit dem ESP2-Tool sind schwer interpretierbar und zeigen keine eindeutige Empfehlung für einen optimalen Parameter. Eine Tendenz für den optimalen *Scale Parameter* kann aber aus den Ergebnissen abgeleitet werden. Für die Umsetzung der Arbeit wird somit der SP für DOP-Daten auf 10 festgelegt, da der Parameter 4 sehr feingliedrige Objekte generiert, die einer pixelweisen Herangehensweise ähnelt. Der SP für die *PlanetScope*-Daten wird auf 7 gesetzt.



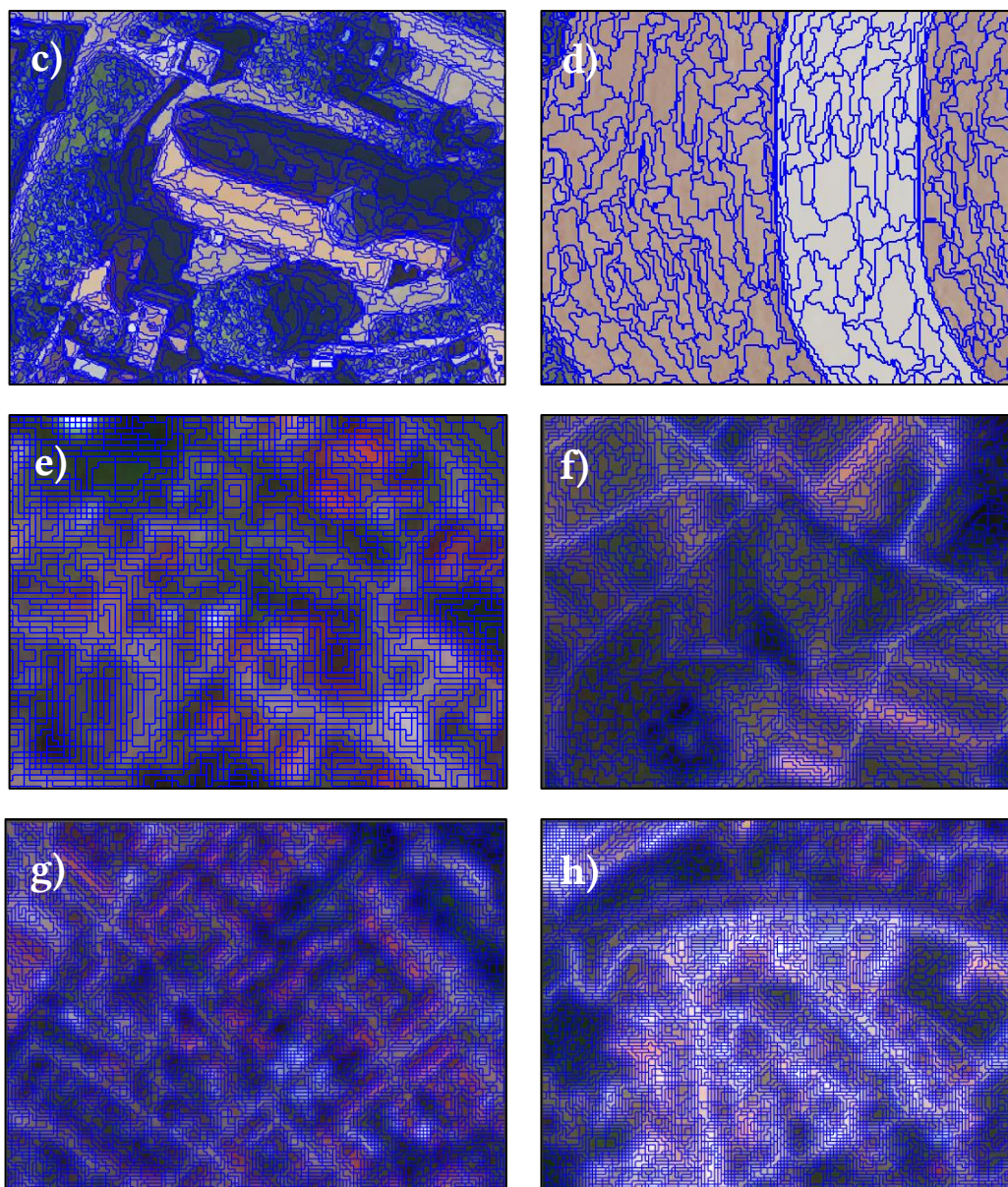


Abbildung 42- Beispiel der Segmentierung mit MRS und einem SP von 10 für DOP (a-d) und 7 für *PlanetScope* (e-h)

4.6.2 Wissensbasierte OBIA-Verfeinerung

OBIA ist eine iterative Methode, welche mit der Segmentierung von Bildaufnahmen in homogene und zusammenhängende Bildsegmente (auch Bildobjekte genannt) beginnt (Blaschke 2010). Die Bildung von homogenen Bildobjekten wird als großer Vorteil von OBIA angesehen (Blaschke et al. 2014b; Liu und Xia 2010). Durch die Analyse der Bildobjekte (OBIA) lassen sich räumliche, textuelle und kontextbezogene Objektmerkmale extrahieren. Die Nutzung der Informationen zu Form, Textur und Größe der Objekte, sowie zu den Nachbarschaftsbeziehungen zueinander (auch *Features* genannt) können zu erheblichen Verbesserungen der Klassifizierungsleistung führen (Blaschke et al. 2014a; Zhang und Xie 2012; Liu et al. 2019). Auch Höhendaten, beispielsweise aus

LiDAR-Messungen oder Digitalen Oberflächenmodellen, können das Erkennungsverfahren erheblich verbessern, werden jedoch noch nicht als kosteneffiziente Lösung für die großflächigen Anwendungen der Veränderungsdetektion oder der Kartierung angesehen (Vakalopoulou et al. 2015).

Bei der Herstellung der TrueDOP-Daten im LGL BW wird beim Prozess des DIM zusätzlich ein bDOM generiert, welches mithilfe des landesweit verfügbaren DGM in ein Normalisiertes Digitales Oberflächenmodell (*Normalized Digital Surface Model* (nDSM)) überführt werden kann. Dieses ist flächendeckend für jede TrueDOP-Aufnahme in einer Auflösung von 0,25 m verfügbar und kann für die Anwendung der wissensbasierten Verfeinerung der Detektionsergebnisse genutzt werden. In dieser Thesis wird die Verwendung der nDSM-Daten für die Verfeinerung mittels Höhendaten exemplarisch dargestellt.

Neben dem (1) nDSM werden für die DOP-Daten außerdem weitere bekannte Indizes wie der (2) *Normalized Difference Vegetation Index* (NDVI), der (3) *Normalized Difference Water Index* (NDWI) und der (4) *Normalized Difference Soil Index* (NDSI) verwendet. Zusätzlich dazu, werden für die *PlanetScope*-Daten weitere Indizes, wie beispielsweise der (5) *Soil Adjusted Vegetation Index* (SAVI), der (6) *Enhanced Vegetation Index* (EVI) und der (7) *Green-Red-Vegetation Index* (GRVI) genutzt.

Zur besonderen Charakteristik der *PlanetScope*-Daten gehören die zwei Grünkanäle Green_I und Green_II (vgl. 2.2.3) und der Red-Edge-Kanal zwischen dem Rot- und NIR-Kanal. Letzterer ermöglicht die Berechnung eines (8) *Normalized Difference RedEdge Red Index* (NDRR) (Shafri und Hamedianfar 2015). Ergänzend folgt ein (9) Schattenindex (SI), bei dem alle Kanäle aufsummiert und durch deren Anzahl geteilt werden. Dieser Kanal kann dienlich sein, um verschattete Orte ausfindig zu machen. Da weder die DOP- noch die *PlanetScope*-Daten einen kurzwelligen Infrarotkanal besitzen, kann der *Normalized Difference Built-Up Index* (NDBI), der in eCognition standardmäßig implementiert ist, nicht genutzt werden. Durch die beiden Kanäle *Coastal-Blue* und *RedEdge* bei *PlanetScope* kann allerdings der auf *WorldView*-basierende (10) Built-Up Index (WV-BI) berechnet werden. Somit stehen beispielhaft 10 verschiedene mögliche Indizes zur Verfügung, mit denen eine wissensbasierte OBIA-Klassifizierung durchgeführt werden kann.

1 $nDSM = DSM - DGM$

2 $NDVI = (NIR - Red) / (NIR + Red)$

3a $NDWI_G1 = (Green - NIR) / (Green + NIR)$

3b $NDWI_G2 = (Green_II - NIR) / (Green_II + NIR)$

4 $NDSI = (Red - Blue) / (Red + Blue)$

- 5 $SAVI = ((NIR-Red) / (NIR+Red+L))*(1+L)$
- 6 $EVI = 2,5*((NIR-Red) / (NIR+C1*Red - C2*Blue +L))$
- 7a $GRVI_G1 = Green_1-Red-Vegetation\ Index\ (Green_1-Red)/(Green_1+Red)$
- 7b $GRVI_G2 = Green_2-Red-Vegetation\ Index\ (Green_2-Red)/(Green_2+Red)$
- 8 $NDRR = (RedEdge-Red) / (RedEdge+Red)$
- 9 $SI = (Red+Green+Blue+NIR) / 4$
- 10 $WV-BI = (Coastal - RedEdge) / (Coastal + RedEdge)$

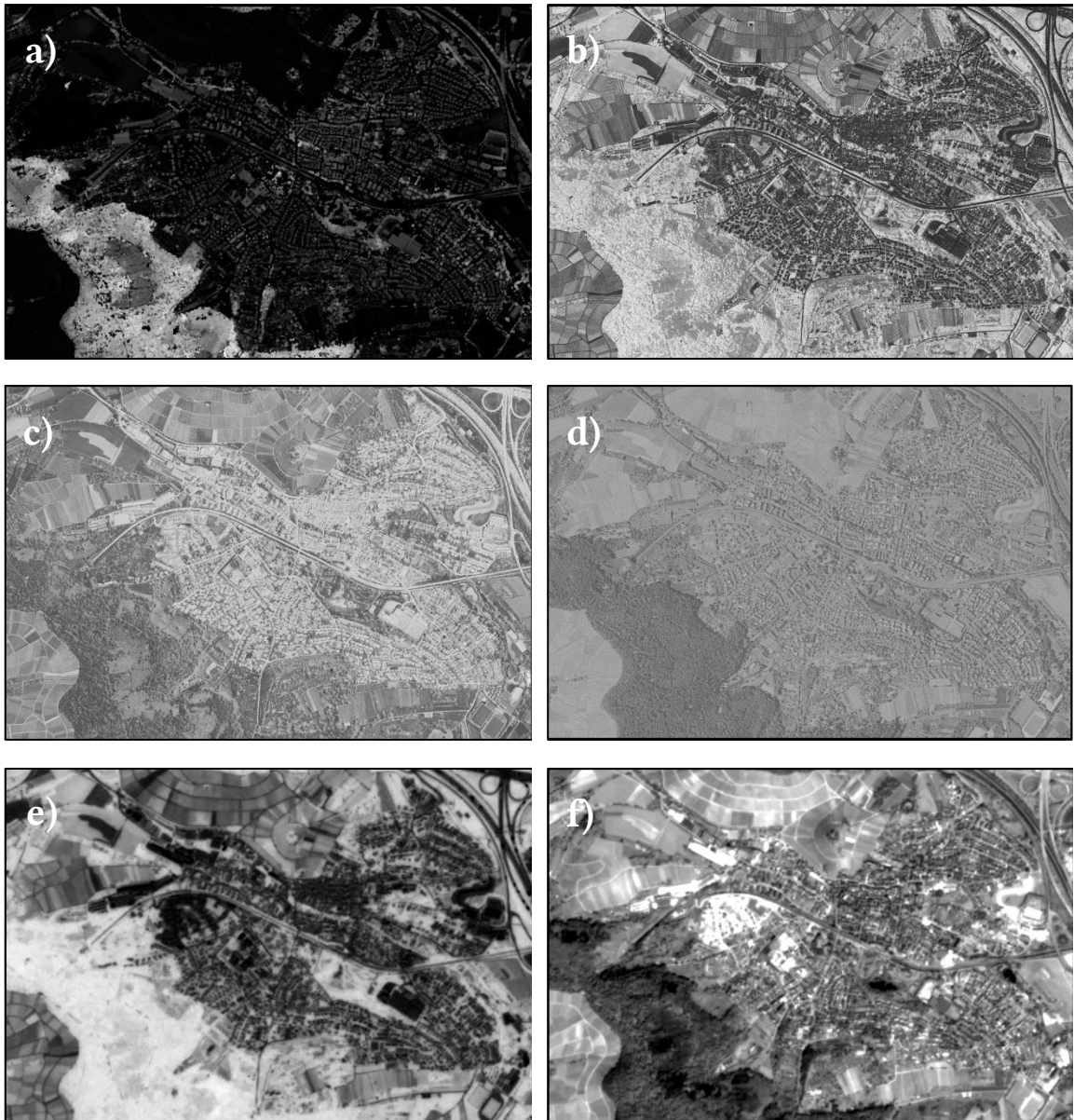




Abbildung 43 - Darstellung der DOP Index-Berechnung mit a) nDSM b) NDVI c) NDWI und d) NDSI und SAT-Index-Berechnung e) NDRR f) SI und g) WV-BI

5 Ergebnisse

Dieses Kapitel beschreibt die abschließenden Ergebnisse der Arbeit, wobei jeweils die Resultate für die flugzeug- und satellitengestützten Rasterdaten getrennt betrachtet werden. Die Workflows wurden unter zwei verschiedenen Gesichtspunkten entwickelt. Die erste Variante zeichnet sich durch eine laufzeitoptimierte Einstellung der Parameter aus, deren Trainingszeit bei ungefähr 6-10 h liegt, was einem realistischen Anwendungsszenario in der Verwaltung entspricht. Die zweite Variante wird als qualitätsoptimierte Parametereinstellung bezeichnet. Hier liegt der Fokus auf einer langen Trainingszeit, bei dem eine hohe Trainingsdauer in einem Maximum der Erkennungsqualität resultieren soll. Die Trainingslaufzeit erreicht dabei eine Dauer von circa 30 - 35 Stunden. Beide Varianten sind in den folgenden Abschnitten aufgeführt. Anzumerken sei noch, dass längere Trainingszeiten als die angegebenen 30 - 35 Stunden zu Programmabstürzen geführt haben, die im Ausschöpfen des Speicherplatzes begründet sein kann. Die qualitätsoptimierte Version stellt also die bestmögliche Konfiguration dar, die mit den Rahmenbedingungen der IT-Komponenten, die für diese Thesis zur Verfügung standen, erfüllt werden konnte.

5.1 Evaluierung der Klassifikationsgüte für DOP-Daten

Die Evaluierung der Klassifikationsgüte beginnt mit der Betrachtung der Klassifizierung des CNN-Prozesses im 3-Klassenmodell. Da die Bestimmung von drei Klassen bei der vorhandenen Datenlage zu Unregelmäßigkeiten in der Detektion einer der gesuchten Zielklasse geführt hat, wurde anschließend zusätzlich ein zweiter Ansatz, welcher sich auf zwei Klassen beschränkt, untersucht (vgl. 5.1.2).

Es wurde eine Vielzahl an Parametereinstellungen für die Umsetzung der Thesis getestet. Diese sind im Anhang aufgeführt.

5.1.1 Klassifikation mit 3-Klassen

Laufzeitoptimierte Trainingsparameter

Nachdem das trainierte CNN auf das Kontrollgebiet angewendet wurde, generiert der Prozess *Heatmaps* (vgl. Abschnitt 4.5). Anschließend wird, wie im Abschnitt 4.6.1 beschrieben, das Gebiet basierend auf den ermittelten Segmentierungsparametern des ESP2-Tools segmentiert. Dabei werden allerdings die automatisiert ermittelten Parameter für die Segmentierung der DOP-Daten leicht angepasst, da der empfohlene SP zu fragmentierten und kleingliedrigen Objekten führt, die sich nachteilig auf die Prozessierungslaufzeiten auswirken. Die Anpassung des Wertes führt weiterhin zu einer erwarteten Übersegmentierung, reduziert aber im Anschluss die Berechnungszeiten der nachgeschalteten Prozesse.

Als Trainingsparameter wurde die *batch size* auf 100 gesetzt und 9.600 Trainings-schritte durchgeführt. Aufgrund der Anzahl an *Training Samples* (160.000) sind somit 1.600 Trainingsschritte notwendig, um einen vollständigen Trainingsdurchlauf aller *Samples* zu erreichen. Im Summe werden 6 Epochen berechnet, was einer Verdopp- lung der Mindestanforderung seitens Trimble entspricht (Trimble 2020). Die Dauer des Trainings liegt mit 5 Stunden im mittleren Bereich.

Über die drei definierten Kontrollklassen „Gebäudeobjekte“, „Kleine Objekte“ und „Nicht-Gebäude“ wird das Ergebnis aus dem CNN mittels einer Genauigkeitsbewer- tung (*Accuracy Assessment*) evaluiert und in diesem Abschnitt präsentiert.

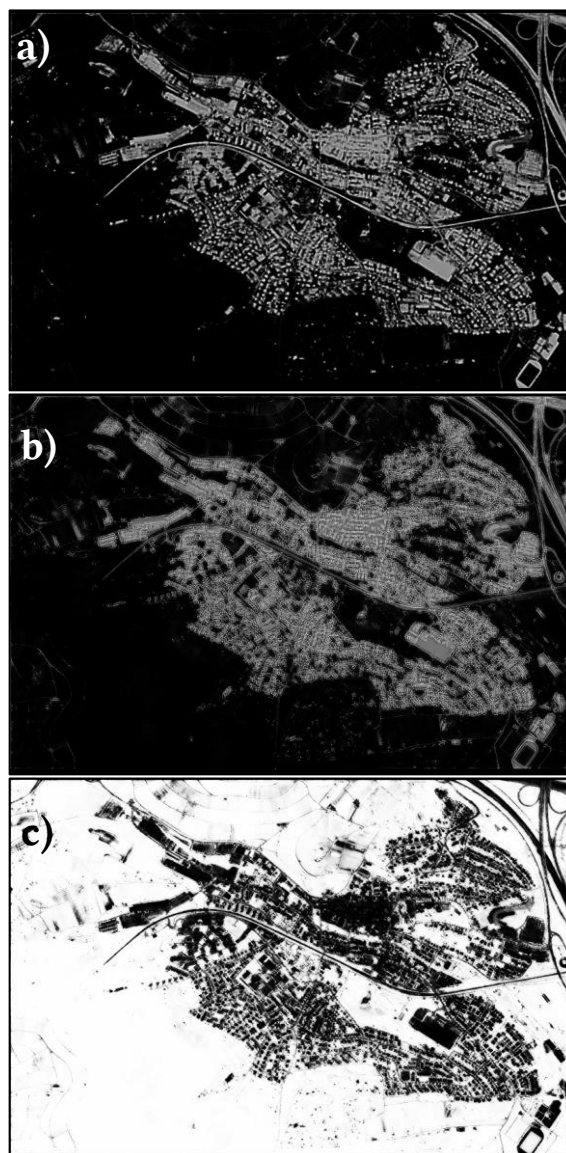


Abbildung 44 – Überblick der Darstellung der *Heatmaps* nach der Berechnung des trainierten CNN für DOP für die 3 Klassen a) Gebäudeobjekte b) Kleine Objekte c) Nicht-Gebäude. Helle Pixel entsprechen dabei jeweils einer hohen Zugehörigkeitswahrscheinlichkeit.

Die Darstellung der Klassenwahrscheinlichkeiten in Abbildung 44 ermöglicht bereits eine erste subjektive Einschätzung der Erkennungsqualität ohne statistische Kennzahlen zu verwenden. Hellere Farben stehen für eine höhere Erkennungswahrscheinlichkeit. Dunkle Farben für eine geringere Wahrscheinlichkeit, wobei die Farbe schwarz für 0 % Wahrscheinlichkeit steht. Die Darstellungen der Klassen „Gebäudeobjekte“ und „Kleine Objekte“ beinhalten wenig helle Pixel innerhalb der *Heatmap*. Auch ist auf den ersten Blick erkennbar, dass sich Infrastrukturelemente, wie Straßen in Bild a) befinden. Hier liegen Fehlklassifikationen vor. Besonders deutlich ist dieser Fehler im nordöstlichen Bereich zu erkennen und in der Bildmitte in Ost-West-Richtung.

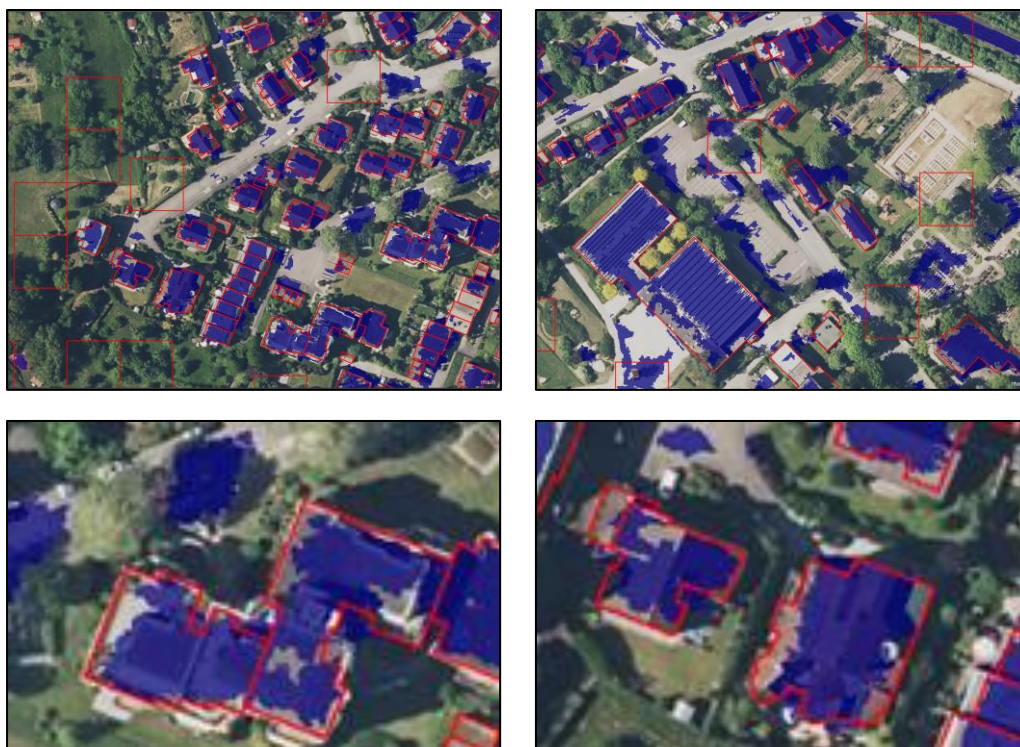


Abbildung 45 - Darstellung CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen) im Vergleich zur *Ground-Truth*-Klasse Gebäudeobjekte (rote Gebäudeumringe)

Abbildung 45 zeigt Beispielbilder der Klassifikation der Zielklasse „Gebäudeobjekte“ im Kontrollgebiet. Ein erster subjektiver Eindruck vermittelt Fehlklassifikationen auf versiegelten Flächen wie Straßen, wobei der Großteil der Gebäudeobjekte richtig klassifiziert werden konnte. In Abbildung 45 ist weiter zu erkennen, dass die Detektion innerhalb der Umringe nicht flächendeckend erfolgte.

Abbildung 46 zeigt die Probleme der CNN-Klassifikation bei der Erkennung der Klasse „Kleine Objekte“. Die Bereiche, mit den höchsten Zugehörigkeitswahrscheinlichkeiten befinden sich größtenteils in Bereichen von Infrastruktur, wie Straßen oder Parkplätzen. Somit ist zu erkennen, dass die Klasse „Kleine Objekte“ eine große Zahl an falsch-positiven Erkennungen generiert und mit den gewählten Einstellungen der Trainingsparameter nicht eindeutig klassifizierbar ist.

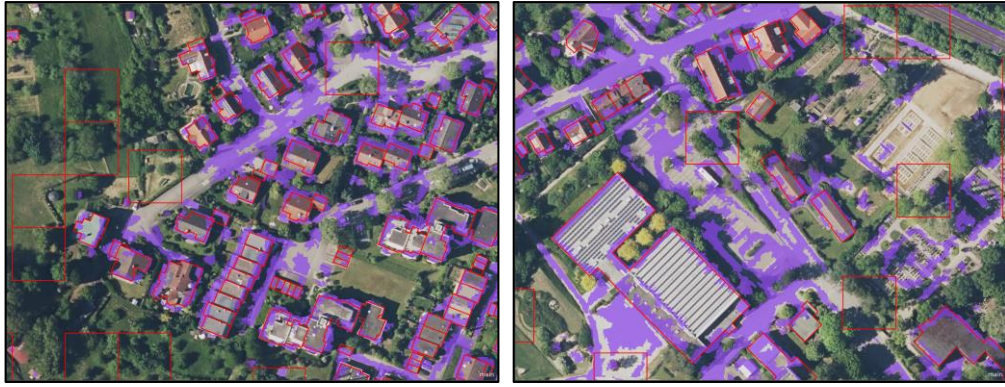


Abbildung 46 - Darstellung der CNN-Klassifizierung der Zielklasse "Kleine Objekte" (lila Flächen) im Vergleich zur Ground-Truth-Klasse (rote Gebäudeumringe)

Abbildung 47 zeigt die Klassifikation der dritten Klasse „Nicht-Gebäude“. Die weiß transparenten Flächen zeigen dabei die Ableitung von Pixeln, die laut dem CNN nicht der Gebäudeklasse angehören. Es lässt sich erkennen, dass die Kontrollkacheln (rote quadratische Kacheln) in Bild c) eine hohe Genauigkeit aufweisen. Die Kontrollfläche in Bild d) konnte allerdings nicht zu 100 % der richtigen Klasse zugeordnet werden.



Abbildung 47 - Darstellung der CNN-Klassifikation der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Flächen) im Vergleich zu den *GroundTruth*-Klassen (rote Umringe)

Visueller Eindruck Neubaugebiet

In Abbildung 48 ist das Neubaugebiet in Weinsberg zu sehen. Die Ableitung der Klassifikation der 3-Klassen verdeutlicht den bereits vorherrschenden Eindruck. Die Zielklasse „Gebäudeobjekte“ klassifiziert zwar die Gebäude in einem hohen Maße, allerdings sind innerhalb der GT-Daten Flächen erkennbar, die nicht klassifiziert wurden. Bei der Klassifikation „Kleine Objekte“ sind viele Artefakte und eine unvollständige Abdeckung zu erkennen. Die höchste Qualität bei der Klassifikation erreicht die Nicht-Zielklasse. Hier sind die relevanten GT-Daten, welche im Neubaugebiet verteilt sind, vollständiger erkannt als die beiden Zielklassen.

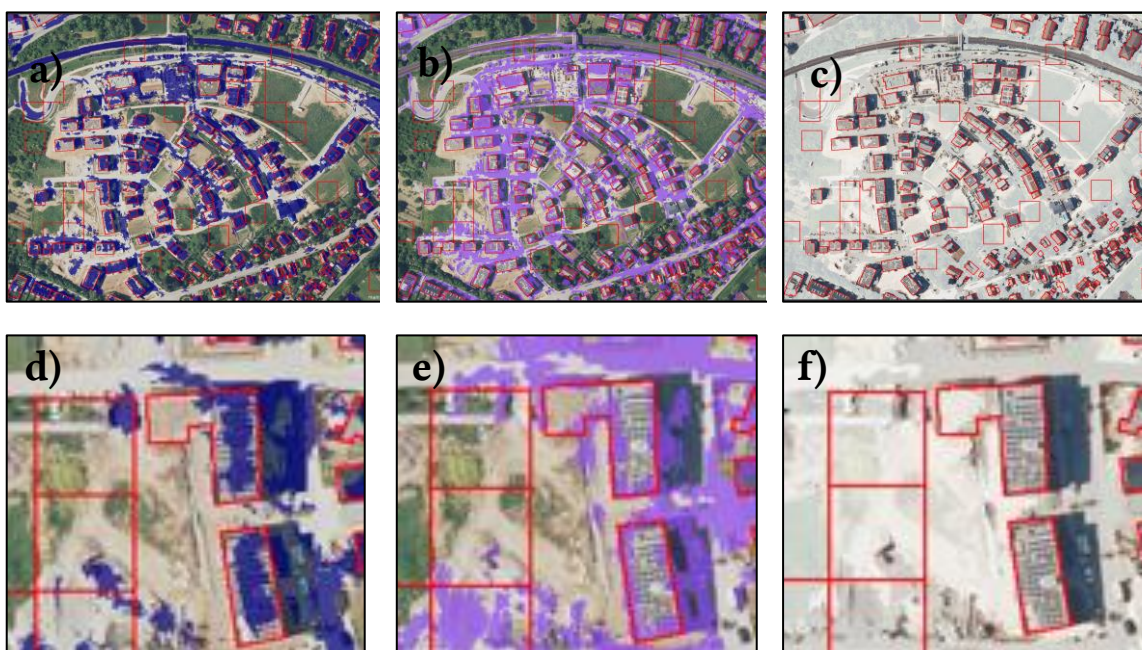


Abbildung 48 - Darstellung der Erkennungsqualität für das Neubaugebiet (a-c) und als Vergrößerung der drei Klassen d) Gebäudeobjekt e) kleine Objekte f) Nicht-Gebäude

Auch die Konfusionsmatrix bestätigt die Vermutung der schlechten Erkennungsqualitäten, welche darauf hindeuten, dass sich das 3-Klassenmodell eher nachteilig im Hinblick auf die Fragestellung dieser Thesis auswirkt. Tabelle 16 beinhaltet die Konfusionsmatrix und verdeutlicht diese Problematik. Die Nicht-Zielklasse hat eine sehr hohe Nutzergenauigkeit von fast 95 %, somit gehören die klassifizierten Objekte mit einer großen Sicherheit dieser Klasse an. Etwas geringer ist die Herstellergenauigkeit, die bekanntlich angibt, ob die Referenzobjekte tatsächlich richtig klassifiziert sind. Kaum schlechter ergibt sich die Nutzergenauigkeit bei der Klasse „Gebäudeobjekte“ mit 89,8 %, welche allerdings eine deutlich schlechtere Herstellergenauigkeit zeigt. Knapp die Hälfte der Pixel, die eigentlich der Klasse „Gebäudeobjekte“ zugehörig sind, werden einer anderen Klasse zugeordnet. Das schlechteste Ergebnis ergibt sich bei der Klasse

„Kleine Objekte“, also die Gebäude mit einem Flächeninhalt unter 20 m². Hier liegt die Nutzergenauigkeit bei lediglich 0,5 %. Es kann also festgehalten werden, dass das CNN-Modell mit einem Trainingsatz von 9.600 Trainingsschritten, oder sechs Epochen, die Klasse „Kleine Objekte“ nur unzureichend erkennt.

Tabelle 16 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 3-Klassen

DOP 3-Klassen Konfusionsmatrix						
KONTROLLDATEN						
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Kleine Objekte	Gebäude-Objekte	Summe	Nutzergenauigkeit
	Nicht-Gebäude	396.123	2.307	19.355	417.785	94,8 %
	Kleine Objekte	16.042	6.264	99.323	121.629	0,5 %
	Gebäudeobjekte	12.609	5.198	156.730	174.537	8,98 %
	Summe	424.774	13.769	275.408	713.951	
	Herstellergenauigkeit	93,3 %	45,5%	56,7 %		Gesamt: 78,3 %

Anpassung der Trainingsparameter

Zur Erhöhung der Erkennungsqualität im Besonderen für die Klasse „Kleine Objekte“, wurden die Trainingsparameter (Gewichtung und Anzahl der Trainingsschritte) angepasst. Die grundsätzliche CNN-Architektur (Anzahl der verborgenen Schichten und *Kernel*-Größe), wurde dabei nicht verändert. Mit diesem Versuch soll ausgeschlossen werden, dass eine zu geringe Trainingsdauer für die nicht ausreichende Erkennung der Klasse „Kleine Objekte“ verantwortlich ist.

Die *batch size* wird nun auf 1.000 gesetzt und insgesamt 10.000 Trainingsschritte vorgenommen. Die Anzahl der *Samples* wurde pro Klasse bei 160.000 Einheiten belassen. Somit werden 160 Trainingsschritte für einen vollständigen Trainingsdurchlauf benötigt, was folglich einer Gesamtlaufzeit von 62 Epochen entspricht. Die Berechnungszeit für die qualitätsoptimierte Anpassung liegt nun im Bereich von 30 - 35 Stunden.

Abbildung 49 zeigt als Ergebnis der Berechnung die drei *Heatmaps* der Wahrscheinlichkeiten des trainierten CNN. Dabei fällt unmittelbar auf, dass bei Bild a) für die Klasse „Gebäudeobjekte“ deutlich höhere Wahrscheinlichkeiten dargestellt werden. Besonders auffällig ist dabei die *Heatmap* der Wahrscheinlichkeiten für die Klasse „Kleine Gebäude“ (Bild b). Es ist zu erkennen, dass das CNN fast keine Ergebnisse für diese Klasse ausgegeben hat, lediglich für vereinzelte Bereiche. Die Klasse ist nahezu

nicht vertreten. Obwohl das Training des Neuronalen Netzes mehr als 30 Stunden benötigte, konnte das angewendete CNN-Modell bis auf einige vereinzelte Stellen keine „*Kleinen Objekte*“ im Kontrollgebiet ableiten.

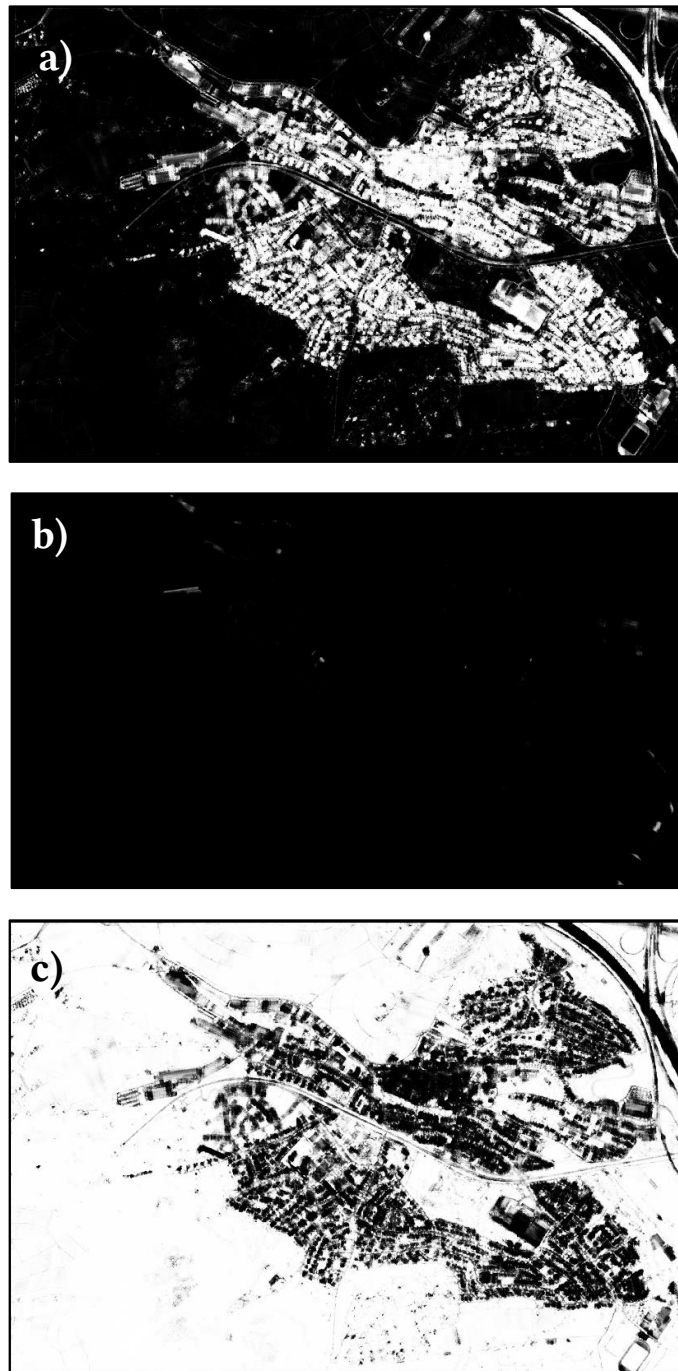


Abbildung 49 - Überblick der Darstellung der *Heatmaps* nach Anpassung der Trainingsparameter für DOP (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit für die 3 Klassen a) Gebäudeobjekte b) Kleine Objekte c) Nicht-Gebäude

Die Konfusionsmatrix aus Tabelle 17 zeichnet ein eindeutiges Bild bezüglich der Klasse „*Kleine Objekte*“. Mit 0 % Herstellergenauigkeit erscheint der Versuch einer gezielten Erkennung für Objekte kleiner 20 m² als ergebnislos. Das trainierte CNN detektiert

diese Objekte zwar als Gebäudeobjekte, scheitert aber daran, diese der richtigen Klasse zuzuordnen. Die Herstellergenauigkeit der beiden anderen Klassen „Nicht-Gebäude“ und „Gebäudeobjekte“ liegt bei 96,5 %, respektive 89,5 %, was einem sehr guten Ergebnis entspricht.

Auch die Nutzergenauigkeiten der Klassen „Nicht-Gebäude“ und „Gebäudeobjekte“ liegen bei über 90 %. Das zeigt, dass das allgemeine Verfahren zur Erkennung der Gebäudeobjekte grundsätzlich erfolgreich ist. Diese Aussage wird durch die Tatsache gestützt, dass die Gesamtgenauigkeit dieses Versuchs bei 91,9 % liegt. Zu bedenken ist allerdings, dass die Klasse „Kleine Objekte“ keine Werte bei der Klassifikation aufweist und somit nicht in die Gesamtbewertung mit einfließt. Selbst weitere Anpassungen der Trainingsparameter waren nicht zielführend und führten infolgedessen zu keiner nennenswerten Verbesserung dieser Klasse hinsichtlich der Nutzer- und Herstellergenauigkeit.

Tabelle 17 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 3-Klassen (angepasste Trainingsparameter)

DOP 3-Klassen Konfusionsmatrix (angepasste Parameter)						
KONTROLLDATEN						
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Kleine Objekte	Gebäudeobjekte	Summe	Nutzergenauigkeit
	Nicht-Gebäude	410.124	1.907	28.831	440.862	93,0 %
	Kleine Objekte	0	0	0	0	nan
	Gebäudeobjekte	14.650	11.862	246.577	273.089	90,3 %
	Summe	424.774	13.769	275.408	713.951	
	Herstellergenauigkeit	96,5 %	0 %	89,5 %		Gesamt: 91,9 %

Das Problem der 3-Klassen-Herangehensweise liegt möglicherweise in der Schwierigkeit der Extraktion herausstehender Merkmale von kleinen Objekten begründet. Der Klassentyp weist zwar Objekte auf, die eine eindeutige und bestimmbare Charakteristik besitzen, wie Garagenreihen. Gleichzeitig können solche Objekte aber auch ein Teil eines größeren Objektes sein, wie bei Anbauten und Überdachungen, und bilden dann optisch eine Einheit mit diesem Objekt. Das erschwert offenbar dem CNN die Merkmalsextraktion dieser Klasse.

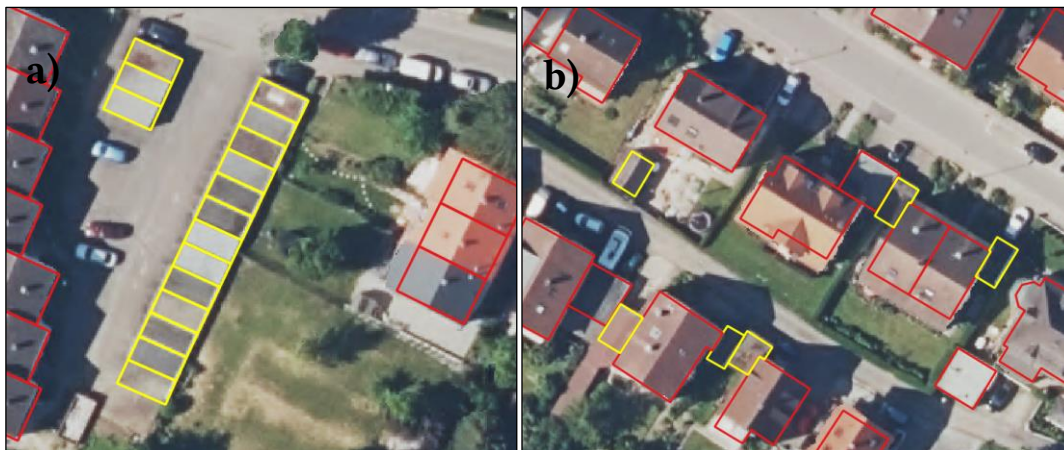


Abbildung 50 – Schwierigkeiten der Differenzierbarkeit der Klasse "Kleine Objekte". a) eindeutige Charakteristik am Beispiel Garagenreihe b) nicht eindeutige Charakteristik am Beispiel Anbauten und Carports

Eine weitere Möglichkeit für die fehlerhafte Detektion dieser Klasse kann an der geringen Anzahl an *Samples* liegen, die für das Training zur Verfügung standen. Lediglich 5.000 unikale Referenzobjekte konnten in den Trainingsdatensatz einfließen, der anschließend mithilfe der *Data Augmentation* quantitativ erhöht wurde. Es ist vorstellbar, dass die Anzahl nicht ausreichend für die Ableitung eindeutiger Objektmerkmale war.

Weiterhin kann auch die generelle *Samples Patch*-Größe ein Grund für die unzureichende Abbildung von kleinen Objekten sein. Da sich „Kleine Objekte“ hauptsächlich in unmittelbarer Nähe von größeren Objekten wie Wohngebäuden befinden, ist es denkbar, dass das CNN andere Merkmalseigenschaften der *Samples* priorisierte. Die *Samples Patch*-Größe, die für alle Klassen gleich gewählt wurde, deckt im Falle von kleinen Objekten womöglich zu viele Hintergrundpixel ab.

Wegen der unzureichenden Bestimmbarkeit wird die Klasse „Kleine Objekte“ aus dem Klassifizierungsdatensatz entfernt und vergleichbar mit den Satellitendaten auf eine binäre Klassifikation gewechselt. Die Ableitung erfolgt somit für die beiden Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“.

5.1.2 Binäre Klassifikation

Standardparametereinstellung (laufzeitoptimiert)

Das bereits trainierte CNN wird hinsichtlich der Klassenreduktion auf nunmehr zwei Klassen angepasst und ausschließlich für die Erkennung der „Gebäudeobjekte“ und „Nicht-Gebäude“ angewendet.

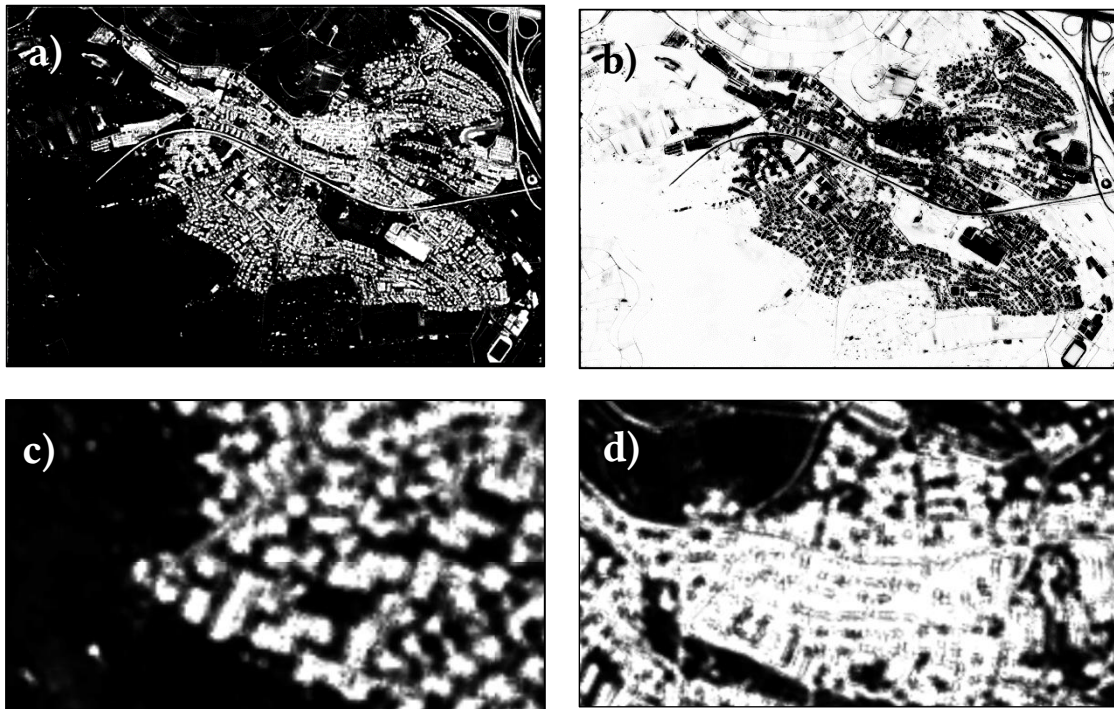


Abbildung 51 - Überblick der Darstellung der *Heatmaps* nach Anwendung des trainierten CNN für DOP für 2 Klassen a) *Gebäudeobjekte* b) *Nicht-Gebäude* c) - d) Vergrößerung der Zugehörigkeitswahrscheinlichkeit zu „*Gebäudeobjekte*“

Ersten subjektiven Einschätzungen nach bewirkt der Übergang auf ein 2-Klassenmodell eine Erkennung höherer Güte. Durch die Fokussierung auf eine einzige Gebäudeklasse wird optisch eine vollständigere Abdeckung der Dachflächen erreicht (vgl. Abbildung 52).



Abbildung 52 – Beispiel 1: CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche) im Vergleich zur *Ground-Truth*-Klasse „Gebäudeobjekte“ (rote Gebäudeumringe) und „Nicht-Gebäude“ (rote regelmäßige Quadrate)

Die Ergebnisse zeigen aber weiterhin Fehlklassifikationen. So ordnet das CNN Infrastrukturobjekte wie Straßen abermals der Gebäudeklasse zu. Ein möglicher Grund hierfür könnte die mangelhafte Repräsentation von Straßenelementen im Trainingsdatensatz der Nicht-Zielklasse sein.

Eine positive Änderung im Vergleich zum 3-Klassen-Ansatz ist die vollständige Abdeckung der Dachflächen. Es muss allerdings festgehalten werden, dass die Klassifikationsergebnisse optisch zu einer Übererkennung der Gebäude neigen.

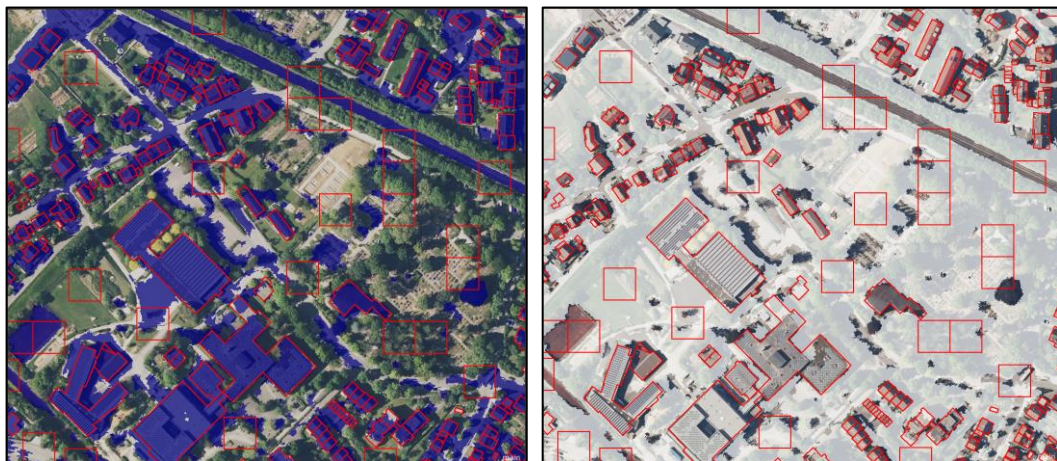


Abbildung 53 – Beispiel 2: Darstellung CNN-Klassifizierung der Zielklasse „Gebäudeobjekte“ (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche) im Vergleich zur *Ground-Truth*-Klasse „Gebäudeobjekte“ (rote Gebäudeumringe) und „Nicht-Gebäude“ (rote regelmäßige Quadrate)

Abbildung 53 zeigt diese Überrepräsentation der Gebäudeklasse. Es ist zu erkennen, dass Straßen, aber auch die Flächen zwischen den Wohnhäusern, sowie einige Grünflächen fälschlicherweise der Klasse „Gebäudeobjekte“ zugeordnet werden. Auch Schattenbereiche werden weiterhin der Gebäudeklasse zugewiesen. Hier zeigt sich weiteres Verbesserungspotenzial.

Positiv hervorzuheben bei diesem Ergebnis ist jedoch, dass sowohl die Dachflächen der Wohngebäude als auch große Industrieobjekte annähernd flächendeckend klassifiziert wurden.

Tabelle 18 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen und den laufzeitoptimierten Parametern

DOP 2-Klassen Konfusionsmatrix (laufzeitoptimierte Parameter)					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzergenauigkeit
	Nicht-Gebäude	403.344	38.519	441.863	91,3 %
	Gebäudeobjekte	21.430	250.658	272.088	92,1 %
	Summe	424.774	289.177	713.951	
	Herstellergenauigkeit	95,0 %	86,7 %		Gesamt: 91,6 %

Die Konfusionsmatrix des 2-Klassen-Ansatzes ist in Tabelle 18 ersichtlich. Die Nicht-Zielklasse hat eine weiterhin hohe Nutzergenauigkeit von 91,3 %. Für die Klasse „Gebäudeobjekte“ liegt die Nutzergenauigkeit etwas höher und erreicht einen Wert von 92,1 %. Somit gehören die klassifizierte Objekte mit einer sehr hohen Wahrscheinlichkeit tatsächlich der Gebäudeklasse an.

Etwas geringer ist die Herstellergenauigkeit der Gebäudeobjekte, erreicht mit 86,7 % aber weiterhin einen guten Wert. Die Herstellergenauigkeit der Klasse „Nicht-Gebäude“ liegt bei 95 %, sodass davon ausgegangen werden kann, dass vergleichsweise wenige Gebäudeobjekte als „Nicht-Gebäude“ klassifiziert wurden. Die Entscheidung die Klassifizierbarkeit der Gebäudeobjekte nicht von deren Flächengröße abhängig zu machen, erweist sich als grundsätzlich richtig.

Durch den Ausschluss der ursprünglich sehr schlecht klassifizierbaren Klasse „Kleine Objekte“ kann eine deutliche Verbesserung erreicht werden. Die Klasse „Gebäudeobjekte“, die nun als alleinstehende Klasse auch die Objekte kleiner als 20 m² beinhaltet, ermöglicht nun offenbar auch die Klassifikation von Gebäudeobjekten, unabhängig von ihrer Größe. Die Unterteilung der zu klassifizierbaren Objekte nach ihrer Größe hat sich für eine gesicherte Klassifikation als hinderlich erwiesen.

Anpassung der Trainingsparameter (qualitätsoptimiert)

Für die Klassifikation mit den qualitätsoptimierten Parametereinstellungen wird dieselbe CNN-Architektur verwendet. Nachdem die Anpassung auf zwei Klassen (Zielklasse und Nicht-Zielklasse) bereits gute Ergebnisse im Abschnitt 5.1.1 geliefert hat, werden die angepassten Parametereinstellungen für die Qualitätsoptimierung im Modell infolgedessen nun auch für die binäre Klassifikation verwendet. Durch die bisherigen Erkenntnisse ist naheliegend, dass die Erhöhung der Anzahl an Trainingsschritten zu einer Qualitätssteigerung der Klassifikation führt.

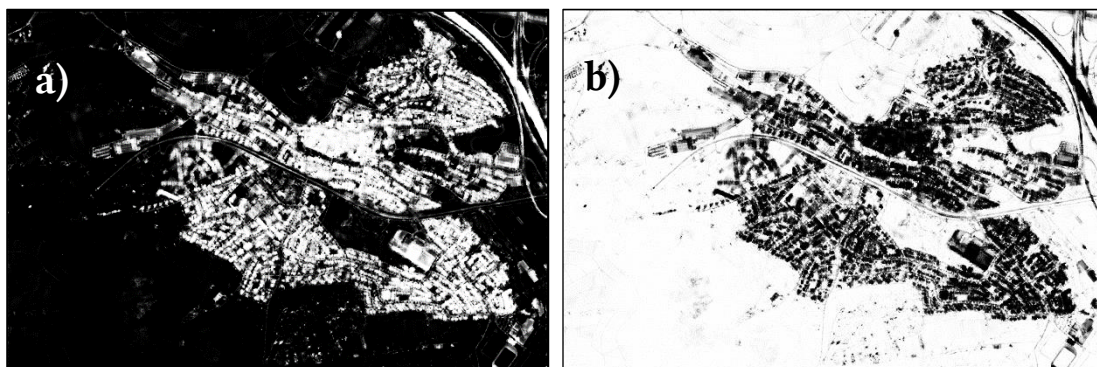


Abbildung 54 - Überblick der Darstellung der *Heatmaps* nach der Berechnung des trainierten CNN für DOP mit angepassten Parametern für die 2 Klassen a) „Gebäudeobjekte“ b) „Nicht-Gebäude“

Das ursprüngliche Training aus Abschnitt 5.1.1 wurde bekanntlich mit jeweils 160.000 *Samples* für die beiden Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“ durchgeführt. Die *Samples* werden vor dem Training gemischt. Es wird eine *Batch Size* von 1.000 gewählt und 10.000 Trainingsschritte ausgeführt (in Summe somit 62 Epochen). Die Berechnungszeit liegt mit insgesamt 30 Stunden in einem hohen Bereich.

Die hohe Trainingszeit offenbart auf dem ersten Blick keine sonderliche Verbesserung gegenüber den Ergebnissen mit der laufzeitoptimierten Parameterwahl. Abbildung 55 visualisiert die Klassen „Gebäudeobjekte“ als blaue Flächen und die Klasse „Nicht-Gebäude“ in einer grauen Farbe.

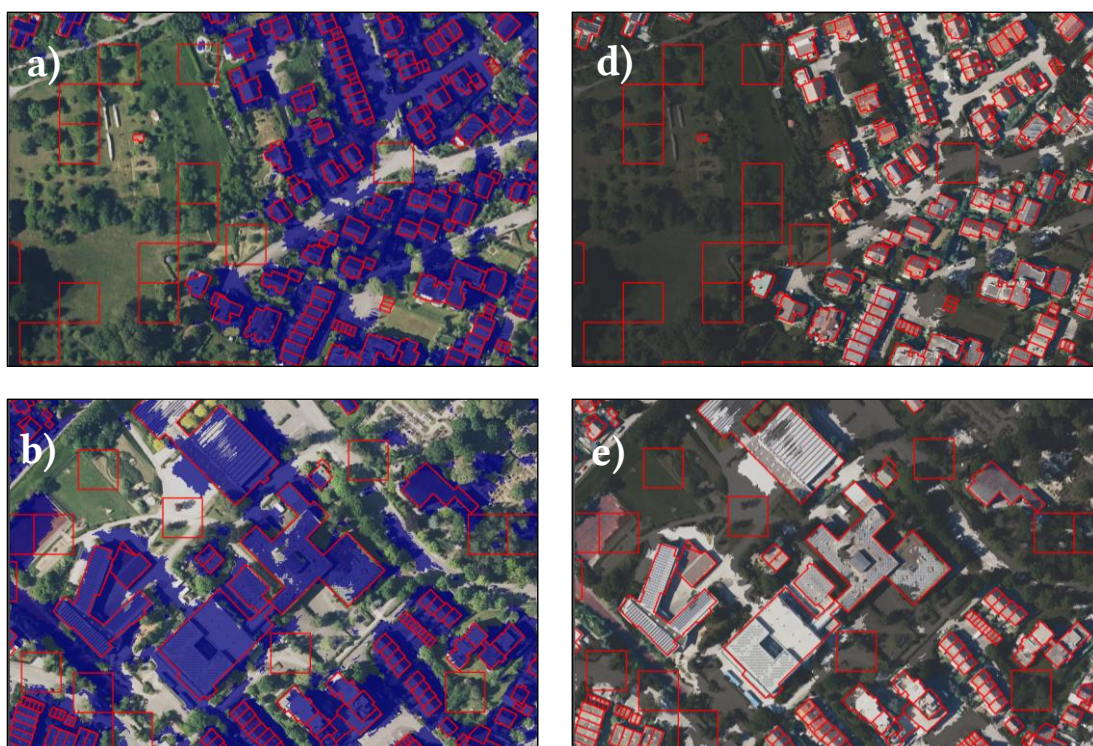




Abbildung 55 - Beispiel der CNN-Klassifizierung der Zielklasse "Gebäudeobjekte" (blaue Flächen a-c) und der Nicht-Zielklasse „Nicht-Gebäude“ (graue Fläche d-f) im Vergleich zur *Ground-Truth*-Klasse (rote Umringe)

In den bebauten Gebieten wird eine deutliche Überrepräsentation der Zielklasse sichtbar. Die Objekterkennung zeichnet sich zwar fast durchgehend mit einer hohen Vollständigkeit der Gebäude und Dachflächen aus, allerdings neigt der Prozess dazu mehr zu klassifizieren als tatsächlich vorliegt. Besonders auffällig ist dabei, der im Bild c) in der Abbildung 55 dargestellte Innenstadtkern. In diesem Beispiel wird deutlich, dass der Algorithmus Gebäude und entsprechend angrenzende, versiegelte Flächen (z.B. Straßen oder Parkflächen) nicht voneinander unterscheiden kann. Diese Problematik hat sich auch in den anderen Durchläufen gezeigt: Infrastrukturobjekte werden häufig der Klasse „Gebäudeobjekte“ zugeordnet.

Visueller Eindruck Neubaugebiet Weinsberg

Dieser erste visuelle Eindruck wird auch für das Neubaugebiet bestätigt. Die Bilder b-d) in Abbildung 56 zeigen die beschriebene Problematik für die Klasse „Gebäudeobjekte“. Einige Objekte konnten nicht vollständig erkannt werden (Bild b). Auch Pixel, die in Schattenbereichen und auf Straßen liegen, werden als Gebäude detektiert. Zusätzlich werden auch Gebäudeobjekte erkannt, die sich noch im Bau befinden.

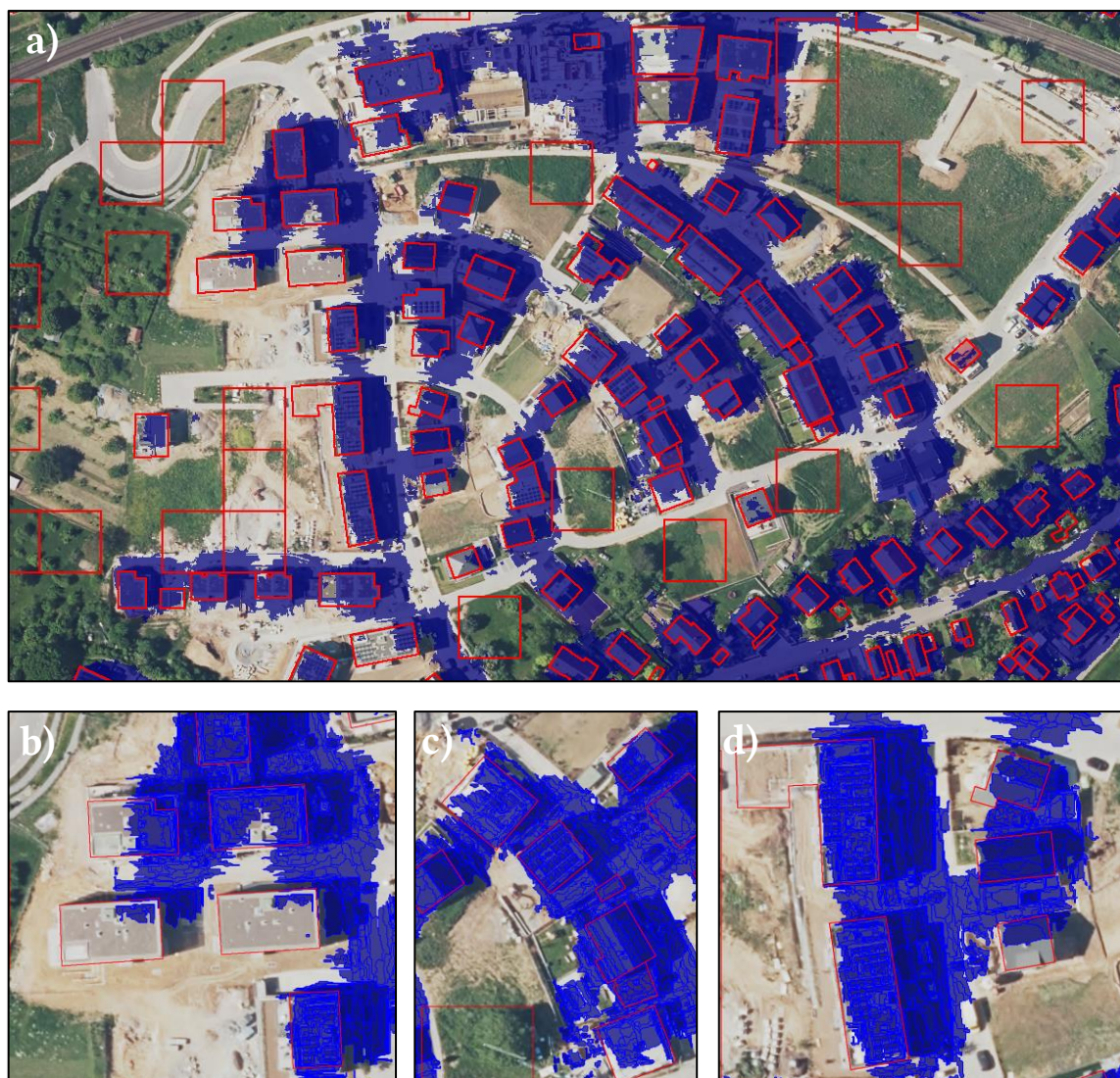


Abbildung 56 – Klassifikation der Gebäudeobjekte (blau) mit den Kontrolldaten der Zielklasse (rote Gebäudeumrandung) und Nicht-Gebäude (rote quadratische Flächen) für das Neubaugebiet in Weinsberg (a) und der Vergrößerung der Darstellung mit dargestellten Segmentierungsobjekten (b – d)

Trotz der visuellen Überrepräsentation der Gebäudeklasse spricht die Gesamtgenauigkeit für das durchgeführte Verfahren. Wie in Tabelle 19 dargestellt, liegt diese bei 93,6 %. Die Herstellergenauigkeit der Gebäudeobjekte erreicht einen Wert von 89,4 %. Die Herstellergenauigkeit bei den „Nicht-Gebäude“ liegt bei 96,5 %. Wird die Nutzergenauigkeit betrachtet, befinden sich beide Werte in einen sehr hohen Bereich, 93,0 % bei „Nicht-Gebäude“ und 94,6 % bei „Gebäudeobjekten“.

Grundsätzlich anzumerken ist allerdings, dass sich nur eine untergeordnete Zahl an GT-Daten der Nicht-Zielklasse in unmittelbarer Nähe der Gebäudeobjekte befindet. Nur eine geringe Anzahl an GT-Daten für die Klasse „Nicht-Gebäude“ fällt in den Raum zwischen einzelnen Gebäudeobjekten in dichtbesiedelten Gebieten. Dadurch ist die Überrepräsentation zwischen den Einzelobjekten zwar visuell erkennbar, fällt aber

statistisch kaum als Fehlklassifikation ins Gewicht. Diese Problematik wurde erkannt und wird in Folge gesondert betrachtet.

Tabelle 19 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen (qualitätsoptimiert)

DOP 2-Klassen Konfusionsmatrix (qualitätsoptimierte Parameter)					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzer-genauigkeit
	Nicht-Gebäude	410.124	30.738	440.862	93,0 %
	Gebäudeobjekte	14.650	258.438	273.089	94,6 %
	Summe	424.774	289.177	713.951	
	Herstellergenauigkeit	96,5 %	89,4 %		Gesamt: 93,6 %

In der folgenden Tabelle 20 werden die Ergebnisse der vier verschiedenen Ansätze inklusive der Berechnungsdauer und der Gesamtgenauigkeit für den DOP-Datensatz zusammengefasst. Klassen- und Parameteranpassungen konnten eine Steigerung der Erkennungsqualität erreichen, allerdings auf Kosten der Laufzeit.

Tabelle 20 - Zusammenfassung der Ausweltergebnisse für die Ableitung von Gebäudeobjekten aus DOP

Anzahl Klassen	Parameter	Dauer in h	Gesamtgenauigkeit in %
3 Klassen	laufzeitoptimiert	5	78,5
	qualitätsoptimiert	35	91,9
2 Klassen	laufzeitoptimiert	5	91,6
	qualitätsoptimiert	30	93,6

5.2 Evaluierung der Klassifikationsgüte für *PlanetScope*-Daten

Wie in Abschnitt 4.3 beschrieben, wird für die Ableitung der Gebäudeobjekte aus Satellitendaten ein binärer Klassifikator angewendet. Dabei ist das Ziel, die Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“ in einer hohen Qualität zu erkennen.

Auch im Fall der Satellitendaten wird das trainierte CNN auf das Kontrollgebiet angewendet und als Ergebnis *Heatmaps* mit Zugehörigkeitswahrscheinlichkeiten für die beiden Klassen ausgegeben (vgl. Abschnitt 4.5). Anschließend erfolgt die Segmentierung anhand der im Abschnitt 4.6.1 ermittelten Parametereinstellungen auf deren Grundlage die Daten anhand der Wahrscheinlichkeiten klassifiziert werden.

Übereinstimmend mit der Evaluierung der DOP-Daten werden auch für die SAT-Daten laufzeit- und qualitätsoptimierte Parameterkonfigurationen untersucht.

Standardparametereinstellung (laufzeitoptimiert)

Die Pixelgröße der Satellitendaten hat einen Einfluss auf den Aufbau der zugrundeliegenden Architektur des CNN. Als Trainingsparameter wurde die *batch size* auf 200 gesetzt und 6.000 Trainingsschritte durchgeführt, somit sind 2.000 Trainingsschritte notwendig, um einen vollständigen Trainingsdurchlauf aller *Samples* zu erreichen. In der Summe werden drei Epochen berechnet, was der Mindestanforderung der Trainingsdurchläufe seitens Trimble entspricht (Trimble 2020). Diese geringe Anzahl an Epochen wird zur Reduktion der Trainingszeit gewählt und liegt bei 9 Stunden.

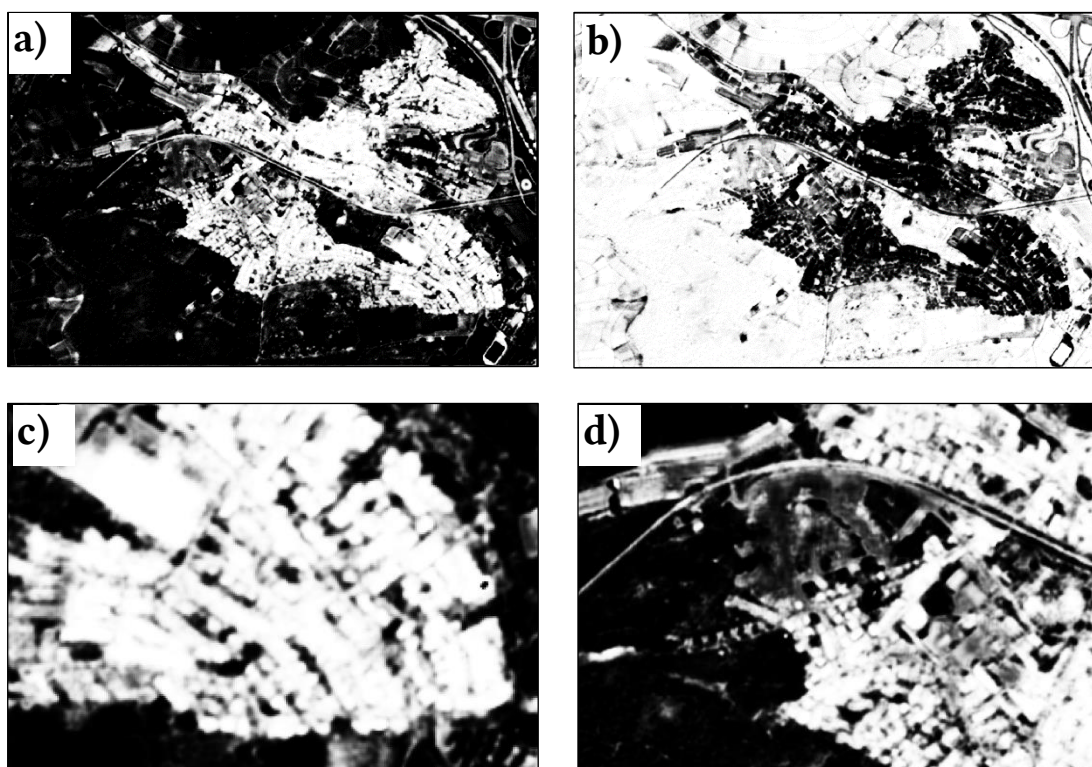
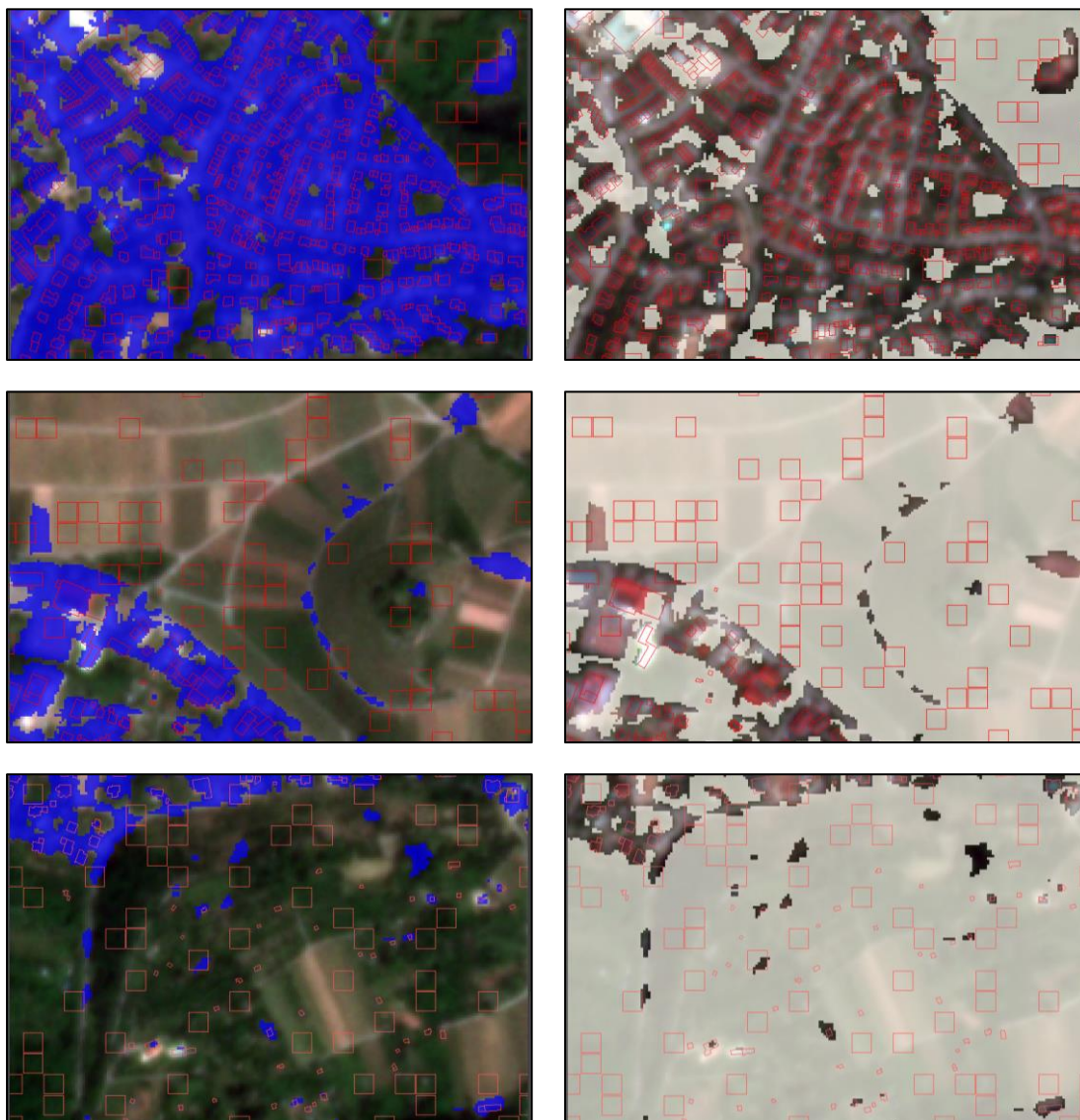


Abbildung 57 – Satellitendaten: Überblick der Darstellung der *Heatmaps* nach Anwenden des trainierten CNN mit Standardparametern (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit) für die 2 Klassen a) *Gebäudeobjekte* b) *Nicht-Gebäude* c) und d) Vergrößerung der Zugehörigkeitswahrscheinlichkeit zur Klasse „Gebäudeobjekte“

Der erste subjektive Eindruck des Ergebnisses mit den standardmäßigen Einstellungen für den Trainingsprozess vermittelt ein positives Bild (vgl. Abbildung 57). Die dargestellten Zugehörigkeitswahrscheinlichkeiten deuten darauf hin, dass eine prinzipielle Erkennung von Gebäudeobjekten aus *PlanetScope* grundsätzlich möglich ist. Die Abbildung 58 zeigt die Klassifikation der Gebäudeobjekte in blau. Es wird deutlich, dass zwar eine Erkennung stattgefunden hat, sie allerdings zu einer deutlichen Überschätzung der Gebäudeklasse neigt.



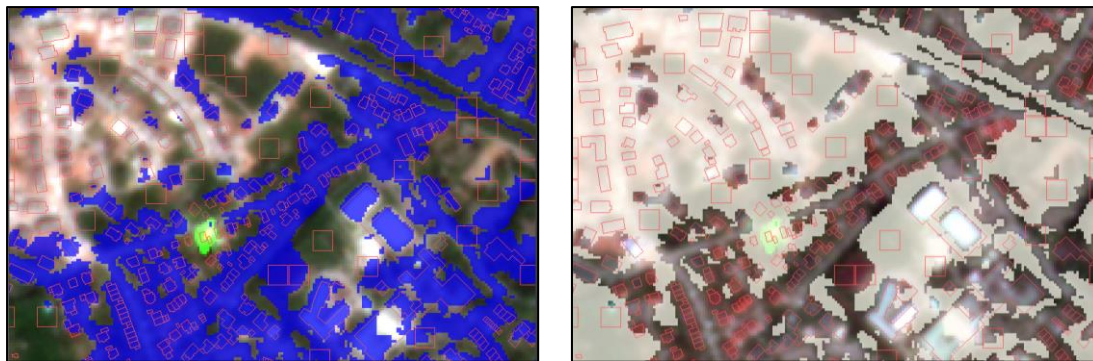


Abbildung 58 - Beispiele aus der CNN-Klassifizierung der Satellitendaten für die Zielklasse "Gebäudeobjekte" (blaue Flächen – linke Spalte), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Fläche – rechte Spalte) im Vergleich zur *Ground-Truth*-Klasse Gebäudeobjekte (rote Gebäudeumringe) und Nicht-Gebäude (rote regelmäßige Quadrate)

Die Beispiele in Abbildung 58 zeigen die Erkennungsqualität der jeweiligen Klassen. Es wird deutlich, dass eine Klassifikation von Gebäudeobjekten grundsätzlich auf dem *PlanetScope*-Datensatz durchführbar ist. Es treten aber auch offensichtliche Fehlklassifikationen auf. Die Gebäude werden im Groben erkannt, eine Separation von einzelnen Objekten ist allerdings mit der in der Thesis verwendeten Parametereinstellungen nicht möglich. Das unterste Beispiel in Abbildung 58 zeigt deutlich, dass das Neubaugebiet falsch klassifiziert wurde. Der Prozess detektiert hauptsächlich „Nicht-Gebäude“ an dieser Stelle. Offenbar fehlt es hier an vergleichbaren Trainingsdaten, alternativ ist ein verlängerter Trainingsdurchlauf eine gangbare Lösungsmöglichkeit.

Tabelle 21 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten mit 2-Klassen und den laufzeitoptimierten Parametern

SAT 2-Klassen Konfusionsmatrix (laufzeitoptimierte Parameter)					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzer-genauigkeit
	Nicht-Gebäude	18.539	3.081	21.620	85,8 %
	Gebäudeobjekte	2.665	25.287	27.952	90,5 %
	Summe	21.204	28.368	49.725	
	Herstellergenauigkeit	87,4 %	90,5 %		Gesamt: 88,4 %

Der visuelle Eindruck spiegelt sich nicht in den statistischen Genauigkeiten aus der Konfusionsmatrix wider, die vergleichsweise gute Ergebnisse suggerieren. Hier zeigt sich die Schwäche der statistischen Auswertung; durch die Verwendung von einzelnen Kontrolldaten kann das Ergebnis aus dem CNN nur punktuell und nicht flächenhaft validiert werden. Die Beispiele in Abbildung 58 zeigen deutlich, dass die erwartete

Klassifizierung auf Einzelobjekt-Ebene nicht erreicht wurde, obwohl die zugehörigen Metriken als ausreichend gut eingestuft werden können.

Angepasste Parametereinstellung (qualitätsoptimiert)

Nach der laufzeitoptimierten Ableitung von zwei Klassen (Zielklasse und Nicht-Zielklasse) aus den Satellitendaten, folgt anschließend der Schritt der Parameteranpassung. Zur Steigerung der Klassifikationsgüte wird zunächst die Zahl der Trainingsschritte erhöht.

Die CNN-Architektur entspricht dabei dem im Abschnitt 4.5 beschriebenen Vorgehens. Die Parametereinstellungen ähneln den unter Abschnitt 5.1.1 verwendeten qualitätsoptimierten Parametern. Das Training wurde mit jeweils 400.000 *Samples* für die beiden Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“ durchgeführt. Die *Samples* werden vor dem Training abermals gemischt. Es wird eine *Batch Size* von 500 gewählt mit 10.000 Trainingsschritten, in Summe somit 13 Epochen. Die Gesamtlaufzeit des Trainings liegt mit 25 Stunden in einem sehr hohen Bereich.

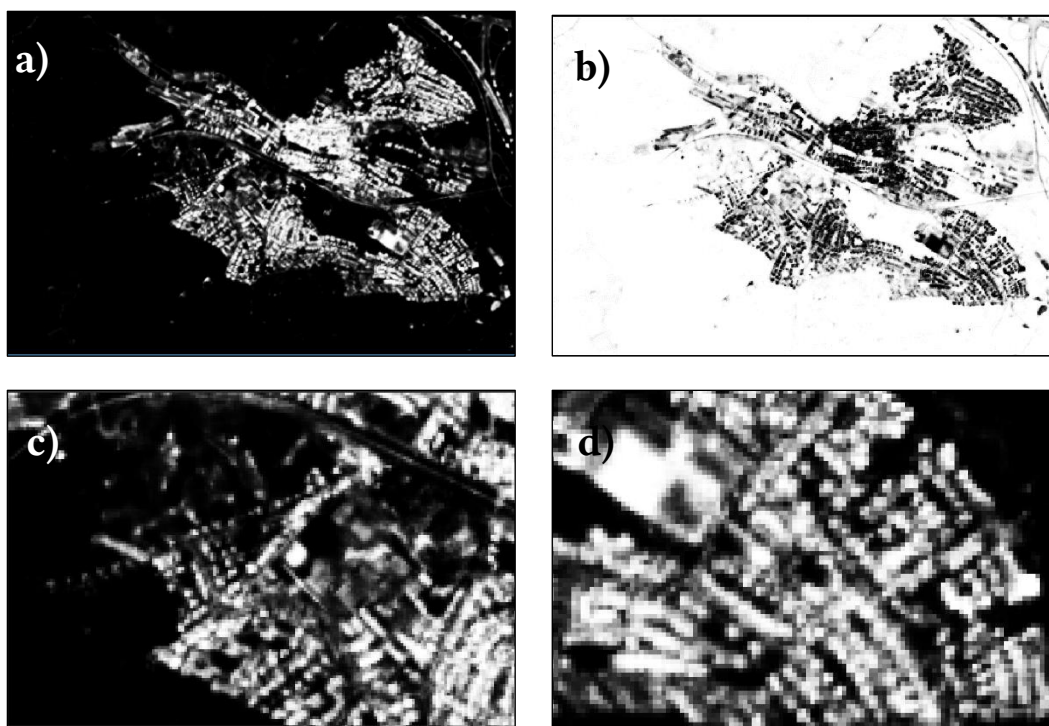
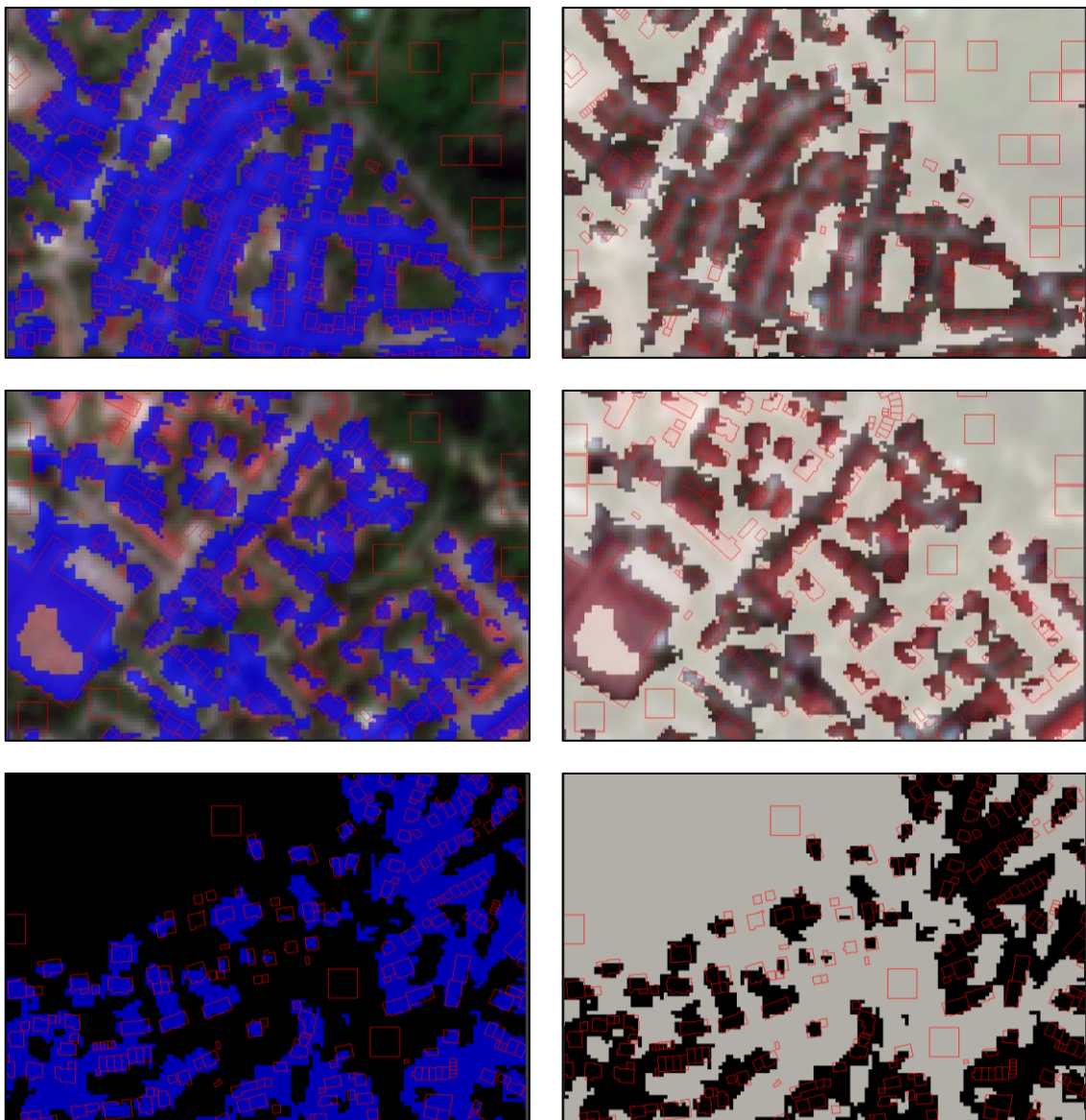


Abbildung 59 – Satellitendaten: Überblick der Darstellung der *Heatmaps* nach der Berechnung des trainierten CNN mit angepassten Parametern (helle Werte stehen für eine hohe, dunkle Wert für eine geringe Wahrscheinlichkeit der 2 Klassen) a) Gebäudeobjekte b) Nicht-Gebäude c) und d) Vergrößerung der Wahrscheinlichkeit „Gebäudeobjekte“

Abbildung 59 zeigt die Wahrscheinlichkeiten des CNN für die beiden Klassen abgeleitet aus Satellitendaten. Der erste Eindruck vermittelt eine differenziertere Ableitung der Gebäudeobjekte. Im Gegensatz zur vorherigen Variante sind nun einzelne Gebäude in der *Heatmap* erkennbar. Die Anpassung hat offenbar zu einem veränderten Erkennungsniveau mit einer saubereren Klassenseparation geführt. Auch erscheint der Effekt der Überrepräsentation abgeschwächt und die Abgrenzungen der Objekte untereinander wirken geschärfter. Artefakte und Fehlklassifikationen treten reduzierter auf.



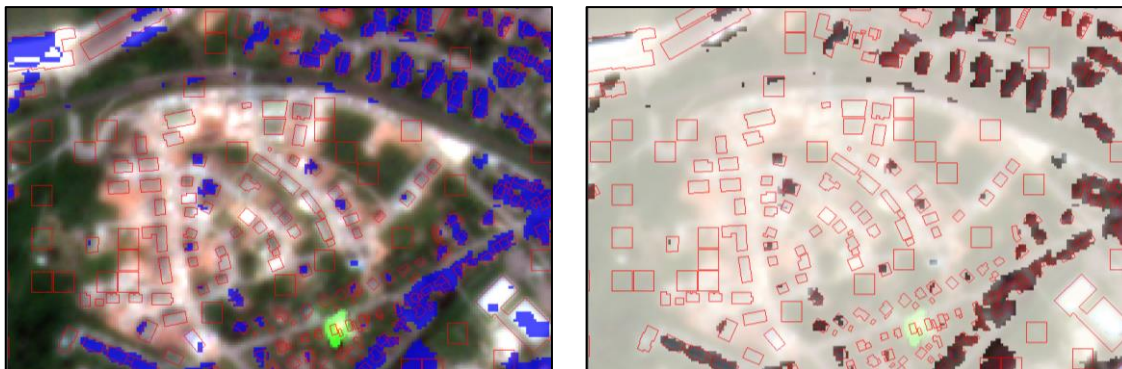


Abbildung 60 - Beispiele aus der CNN-Klassifizierung der Satellitendaten für die Zielklasse "Gebäudeobjekte" (blaue Flächen), sowie der Nicht-Zielklasse „Nicht-Gebäude“ (weiße Flächen) im Vergleich zur *Ground-Truth*-Klasse Gebäudeobjekte (rote Gebäudeumringe) und Nicht-Gebäude (rote regelmäßige Quadrate)

Eine genauere Prüfung der Ergebnisse zeigt jedoch, dass grundsätzlich detektierbare Gebäudeobjekte nicht klassifiziert werden konnten. Besonders deutlich wird das am Beispiel des Neubaugebietes in Abbildung 60. Das CNN hat die Neubauten nicht als Gebäudeobjekte erkannt. Die Erkennung zeigt zwar eine schärfere Abgrenzung der einzelnen Objekte und auch die Überrepräsentation ist deutlich reduzierter als zuvor, insgesamt werden aber nun weniger Gebäudeobjekte selektiert.

Tabelle 22 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten der 2-Klassen mit qualitätsoptimierten Parametern

SAT 2-Klassen Konfusionsmatrix (ergebnisoptimierte Parameter)					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzergenauigkeit
	Nicht-Gebäude	20.860	10.709	31.569	66,1 %
	Gebäudeobjekte	344	17.659	18.003	98,1 %
	Summe	21.204	28.368	49.725	
	Herstellergenauigkeit	98,4 %	62,3 %		Gesamt: 77,7 %

Die Konfusionsmatrix bestätigt den visuellen Eindruck. Die Herstellergenauigkeit der Gebäudeobjekte ist auf 62,3 % gefallen und somit nur noch im Mittelfeld angesiedelt. Die Herstellergenauigkeit bei den Nicht-Gebäuden ist mit 98,4 % weiterhin sehr hoch, eine deutliche Mehrheit der Kontrolldaten werden also richtig klassifiziert. Im Gegenzug liegt die entsprechende Nutzergenauigkeit allerdings lediglich bei 66,1 %, somit gehören nur 2/3 der klassifizierten Objekte tatsächlich der Klasse „Nicht-Gebäude“ an. Die Nutzergenauigkeit der Gebäudeobjekte liegt bei 98,1 %. Schlussfolgernd kann fest-

gehalten werden, dass die Gebäude durch die veränderten Parametereinstellungen zwar selektiver ermittelt werden, insgesamt die Gesamtgenauigkeit im Vergleich zur vorherigen Variante allerdings abgenommen hat.

Zusammenfassend lässt sich sagen, dass die hier vorgestellte Erstableitung mit lauffeiertoptimierten Parametern zwar zu einer hohen Gesamtgenauigkeit geführt hat, das visuelle Ergebnis aufgrund der Überpräsentation im Bereich dicht bebauter Siedlungen allerdings nicht zufriedenstellend war. Diese lässt sich vermutlich auf die im Vergleich zu den DOP-Daten gröbere räumliche Auflösung der *PlanetScope*-Daten erklären. Um ihr entgegenzuwirken, wurden die Trainingsparameter entsprechend angepasst, um dem CNN das Erlernen weiterer relevanter Merkmale aus den Satellitendaten zu ermöglichen. Durch diese Anpassung konnte tatsächlich eine wahrnehmbare Qualitätsverbesserung erzielt werden. Das Ergebnis zeigt nun eine deutlich bessere Trennung und Kantenbildung zwischen den Objekten. Allerdings führte es zu einer sichtbaren Unterklassifikation der Gebäudeobjekte. Die Gesamtgenauigkeit fällt von 88,4 % auf 77,7 %.

Es lässt sich also kein abschließendes Fazit aus den bisherigen Untersuchungen mit *PlanetScope*-Daten ziehen. Die Beispiele zeigen, dass eine gezielte Ableitung von Gebäudeobjekten aus *PlanetScope*-Daten mit akzeptablen Ergebnissen möglich ist und somit großes Potential hinsichtlich der Verwendung von den Satellitendaten im Hinblick auf die Fragestellung besteht. Zur Ableitung optimaler Parametereinstellungen sind allerdings noch weitere Untersuchungen notwendig.

5.3 OBIA-Verfeinerung der Klassifikation

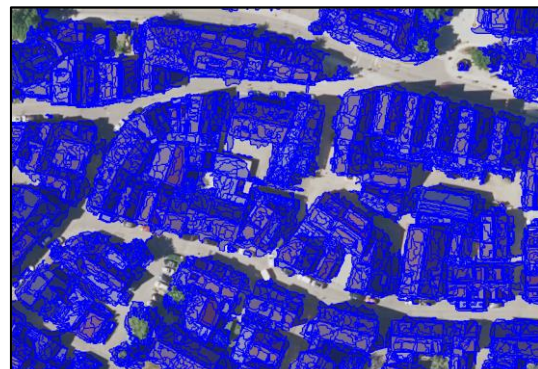
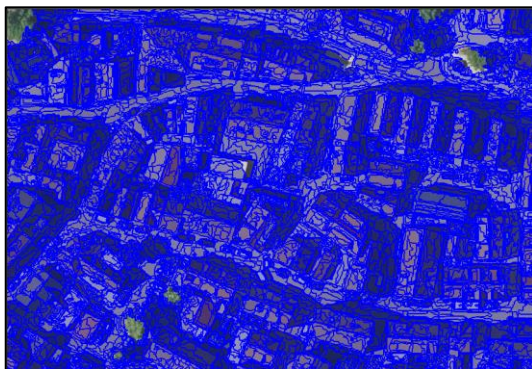
5.3.1 Wissensbasierte Bildanalyse der TrueDOP-Daten

Zur Reduktion der Effekte, die durch Fehlklassifikationen oder durch die Zusammenfassung von mehreren zusammenstehenden Objekten zu einem Zielobjekt entstehen, kann anschließend an die CNN-Prozessierung eine Verfeinerung in Form einer regelbasierten Klassifikation durchgeführt werden. Durch die IT-Rahmenbedingungen, die für die vorliegende Thesis gelten, erscheint das mögliche Verbesserungspotential der Klassifikation durch Anpassung von Trainingsdaten und den entsprechenden Trainingsparametern im CNN ausgereizt. Zur Verbesserung der Ergebnisse wird deswegen eine objektbasierte Bildanalyse, bei der etwa Nachbarschaftsbeziehungen, die Form und Größe von Objekten sowie deren spektrale Eigenschaften berücksichtigt werden, angewendet. Exemplarisch wird in diesem Abschnitt das Ergebnis der nachgelagerten OBIA-Verfeinerung für die Untersuchung mit den qualitätsoptimierten Parametern der binären Klassifikation der TrueDOP dargestellt.

Wie in Abschnitt 4.6.2 beschrieben, wird anhand der verfügbaren Bildkanäle und mit daraus abgeleiteten Bandarithmetiken eine Verfeinerung der Klassifizierungsergebnisse angestrebt. Für die TrueDOP-Aufnahmen steht nach dem Herstellungsprozess standardmäßig ein nDSM zur Verfügung. Dieses wird in Kombination mit dem (2) NDVI, dem (3) NDWI und dem (4) NDSI in Form von Bedingungsabfragen zur Qualitätsabsicherung von segmentierten Bildobjekten genutzt. Fehlerhaft segmentierte Objekte können dadurch etwa der gewünschten Zielklasse hinzugefügt werden. Für DOPs, die keine Höheninformation aufweisen, können dennoch die Indizes genutzt werden.

Der Schattenindex ermöglicht beispielsweise den Ausschluss von Objektsegmenten, deren Farbwert in einen zu dunklen Bereich fällt, wie es in abgeschatteten Bereichen einer Hausseite der Fall ist. Exemplarisch für das unter 4.6.2 beschriebene Expertenwissen werden die Indizes (1) bis (4) und der Schattenindex (9) verwendet. Dabei werden ausschließlich Bildobjekte der überrepräsentierten Klasse „Gebäudeobjekte“ in die Klasse „Nicht-Gebäude“ verschoben. Beispielphaft sind die folgenden Bedingungen:

- (a) Shadowindex < 50 and Mean nDSM < 2
- (b) Mean NDSI $< -0,4$
- (c) Mean NDWI < 0
- (d) Mean NDVI $> 0,3$ and Mean nDSM < 1



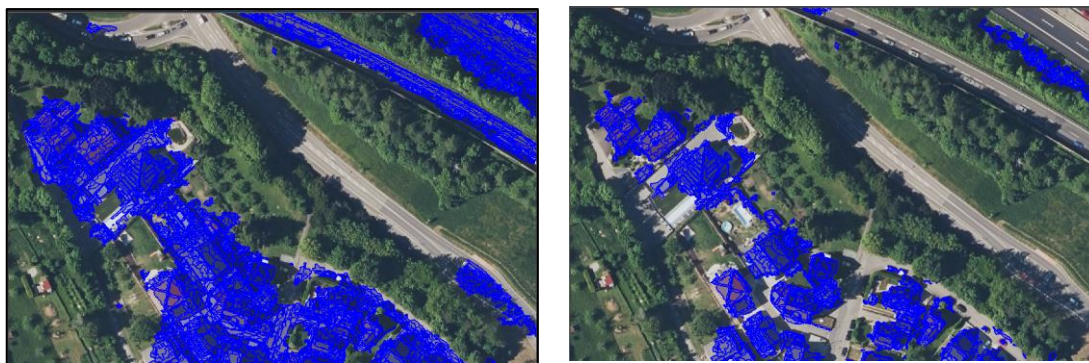


Abbildung 61 - Vergleich der Segmentierung der Gebäudeobjekte vor (linke Seite) und nach (rechte Seite) der OBIA-Verfeinerung für DOP

Tabelle 23 zeigt ein Ergebnis mit 94,6 % Gesamtgenauigkeit bei automatischer Ableitung mit CNN und anschließender wissensbasierter Bildanalyse. Bei der Klasse „Nicht-Gebäude“ werden 99,2 % der Kontrolldaten korrekt abgeleitet. Bei den Gebäudeobjekten ist der Wert der Herstellergenauigkeit allerdings im Vergleich zur vorherigen Ableitung (vgl. Tabelle 19) geringfügig gefallen und befindet sich bei 87,7 %. Es ist vorstellbar, dass die regelbasierte Anpassung zu viele segmentierte Gebäudeobjekte in die Nicht-Zielklasse übertragen hat und somit den leichten Rückgang verursacht.

Tabelle 23 - Konfusionsmatrix für die CNN-Klassifikation der DOP-Daten mit 2-Klassen nach wissensbasierter Bildanalyse

DOP 2-Klassen Konfusionsmatrix (qualitätsoptimierte Parameter)					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzer-genauigkeit
	Nicht-Gebäude	421.294	35.385	456.679	92,3 %
	Gebäudeobjekte	3.480	253.792	257.272	98,7 %
	Summe	424.774	289.177	713.951	
	Herstellergenauigkeit	99,2 %	87,7 %		Gesamt: 94,6 %

5.3.2 Wissensbasierte Bildanalyse der PlanetScope-Daten

In diesem Abschnitt wird analog zum Kapitel 5.1.2 der Versuch unternommen mithilfe von wissensbasierter Bildanalyse die Klassifikation auf Basis der PlanetScope-Daten zu verfeinern. Für diese exemplarische Darstellung werden die Ergebnisse der ersten Klassifikation aus 5.2, also diejenige mit laufzeitoptimierten Parametern, untersucht. Der Ablauf ähnelt dem zuvor bereits beschriebenen Verfahren und zielt darauf ab,

mithilfe der Index-Berechnungen und Bedingungsabfragen mit logischen Operatoren für beide Klassen Fehlklassifikationen nachträglich zu minimieren.

Die insgesamt 8 Bänder der *PlanetScope*-Daten gestatten eine größere Bandbreite an möglichen Indizes, die für die wissensbasierte Bildanalyse herangezogen werden können.

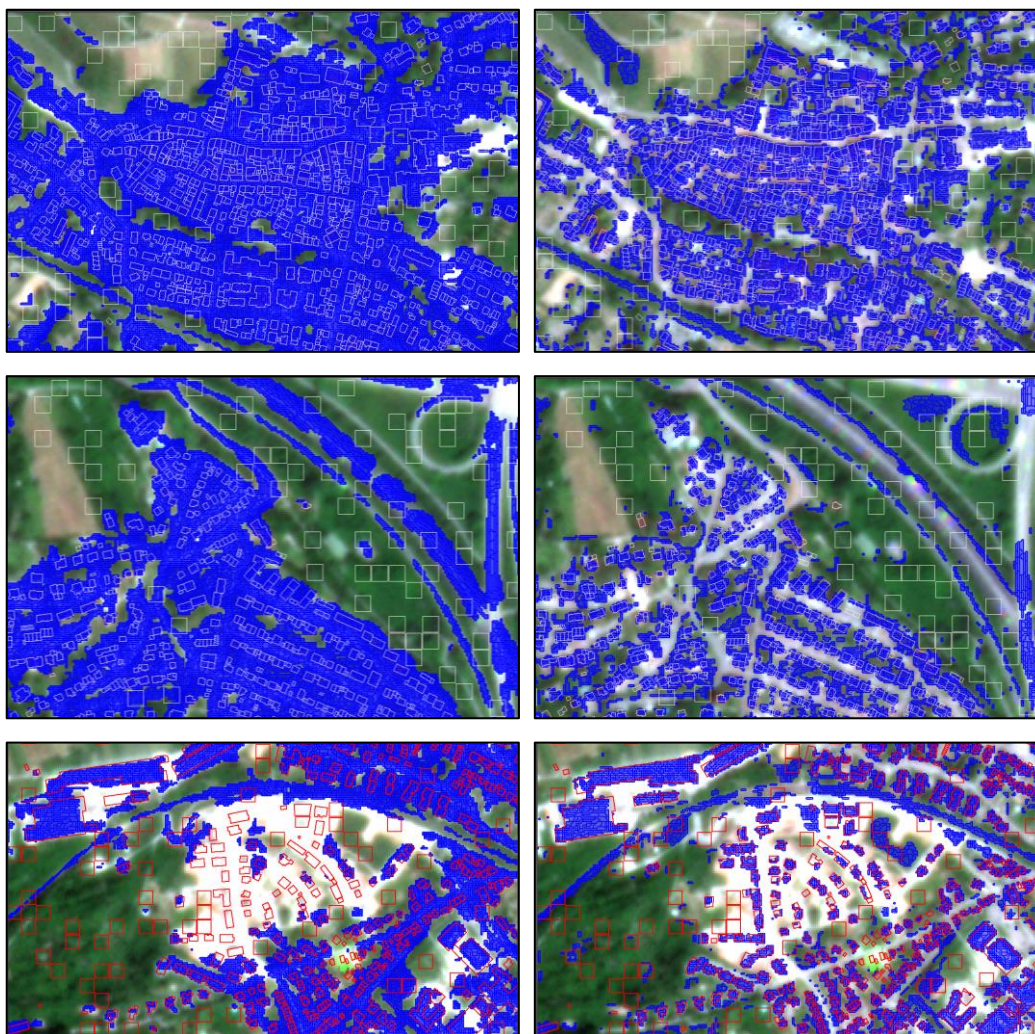


Abbildung 62 - Vergleich der Segmentierung der Gebäudeobjekte vor (linke Spalte) und nach (rechte Spalte) der OBIA-Verfeinerung für SAT

Tabelle 24 zeigt für die automatische Ableitung mit CNN und anschließender wissensbasierter Bildanalyse ein Ergebnis mit 84,4 % Gesamtgenauigkeit. Bei der Klasse „Nicht-Gebäude“ werden 90,9 % der Kontrolldaten korrekt abgeleitet. Bei der Klasse „Gebäudeobjekte“ kann ein Rückgang bei der Herstellergenauigkeit von 90,5 % auf 79,5 % beobachtet werden (vgl. Tabelle 21). Visuell erscheint die Abgrenzung der Gebäudeobjekte untereinander allerdings klarer. Einer deutlichen Überrepräsentation der Klasse „Gebäudeobjekte“ konnte mit einer regelbasierten Anpassung entgegengesteuert werden, welche allerdings in einem Rückgang der Gesamtgenauigkeit resultiert. Auch

hier ist es vorstellbar, dass zu viele segmentierte Gebäudeobjekte in die Nicht-Zielklasse übertragen wurden und den leichten Rückgang verursacht haben.

Tabelle 24 - Konfusionsmatrix für die CNN-Klassifikation der SAT-Daten mit 2-Klassen nach wissensbasierten Bildanalyse

SAT 2-Klassen Konfusionsmatrix					
KONTROLLDATEN					
KLASSIFIZIERUNG	Klassen	Nicht-Gebäude	Gebäudeobjekte	Summe	Nutzer-genauigkeit
	Nicht-Gebäude	19.295	5.817	25.082	76,8 %
	Gebäudeobjekte	1.939	22.551	24.490	92,0 %
	Summe	21.204	28.368	713.951	
	Herstellergenauigkeit	90,9 %	79,5 %		Gesamt: 84,4 %

6 Diskussion

Das primäre Ziel der vorliegenden Arbeit war zu klären, ob und wie gut es möglich ist, Objekte des amtlichen ALKIS-Gebäudedatenbestandes automatisiert aus Fernerkundungsdaten zu klassifizieren. Dazu wurden Rasterdaten zweier unterschiedlicher bildgebender Trägersysteme mithilfe von Deep Learning-Methoden und objektbasierter Bildklassifikation klassifiziert, untersucht und die Ergebnisse gegenübergestellt. Damit soll auch eine Aussage über die Vor- und Nachteile der jeweiligen Rasterdaten, bezüglich ihrer Eignung für die Bestimmung neuer Elemente des Liegenschaftskatasters, getroffen werden. Bei den beiden untersuchten Bilddatensätzen handelte es sich um die geometrisch höher aufgelösten flugzeuggestützten Digitalen Orthophotos im Qualitätsstandard TrueDOP, sowie die zeitlich höher aufgelösten satellitengestützten *PlanetScope*-Daten. Der amtliche Gebäudedatenbestand aus ALKIS diente als *Ground-Truth*-Daten zur Quantifizierung der jeweiligen Erkennungsqualitäten. Dafür wurde eine Genauigkeitsbewertung der Klassifikationsergebnisse auf Basis von Kontrolldaten für die jeweiligen Trägersysteme durchgeführt. Die angewendete Methodik ähnelt in den Grundzügen der Methodik anderer Publikationen mit vergleichbarer Fragestellung, etwa Ghorbanzadeh et al. (2021). Besonders die Kombination aus Deep Learning und der objektbasierten Herangehensweise als Standardmethode der Bildanalyse (Blaschke et al. 2014a) ermöglichen eine qualitativ hochwertige Klassifikation von besonders hochauflösenden Bilddaten. So konnte eine Gesamtgenauigkeit von 94,5 % für die Digitalen Orthophotos in der Qualitätsstufe TrueDOP und 88,4 % für die *PlanetScope*-Satellitendaten erreicht werden.

Für eine Ableitung von bestmöglichen Ergebnissen wurde eine Vielzahl an Testdurchläufen des CNN durchgeführt. Diese Testdurchläufe unterschieden sich deutlich in ihrer jeweiligen Trainingsdauer.

Die ursprüngliche Ableitung der Gebäudeobjekte aus den TrueDOP-Daten für zwei Zielklassen „Gebäudeobjekte“ und „Kleine Objekte“ (Objekte kleiner als 20 m²), sowie einer Nicht-Zielklasse zeigte im ersten Durchgang eine Gesamtgenauigkeit im mittleren Bereich von 78,3 % (vgl. 5.1.1). Auffällig war hierbei, dass besonders die Zielklasse mit Objekten kleiner als 20 m² zu einer geringen Erkennung führte. Durch die Erhöhung der Trainingsdurchläufe sollte eine Verbesserung der Erkennung dieser Klasse herbeigeführt werden. Dieser sogenannte qualitätsoptimierte Durchlauf führte allerdings zu einer Verschlechterung der Detektion der Klasse „Kleine Objekte“. Das CNN konnte keine Zugehörigkeitswahrscheinlichkeiten für das Vorhandensein der Klasse im Kontrollgebiet ableiten. Die Herstellergenauigkeit wurde mit 0 % und die Nutzergenauigkeit musste mit N/A angegeben (vgl. 5.1.1).

Als Konsequenz aus diesem Zwischenergebnis wurde die Klassenzahl auf zwei reduziert und mit den Klassen „Gebäudeobjekte“ und „Nicht-Gebäude“ auf das Kontrollge-

biet angewendet. Dieselben Parameter, die bei der 3-Klassen-Klassifikation zu einem mittleren Ergebnis von unter 80 % geführt haben, erreichten nun im ersten Durchgang eine Gesamtgenauigkeit von 91,6 % (vgl. 5.1.2). Durch Anpassung der Parameter und Erhöhung der Berechnungsdauer konnte das CNN im Fall dieser Untersuchung eine maximale Gesamtgenauigkeit von 93,6 % erreichen.

Daraus lässt sich folgern, dass die Schwierigkeiten einer semantischen Trennung der beiden Gebäude-Zielklassen durch die Reduktion von drei auf zwei Klassen umgangen werden konnte. Die ursprünglich fehlerbehaftete und unzureichende Merkmalsextraktion bei einer gleichzeitigen Ableitung der Klassen „Gebäudeobjekte“ und „Kleine Objekte“, konnte nun überwunden werden. Die Detektion der Objekte kleiner als 20 m² wurde nun über die Ableitung der Zielklasse „Gebäudeobjekte“ erreicht. Offenbar ist eine klare Separation der beiden Klassen mit einem gleichzeitigen Training nicht ohne weiteres möglich.

Als mögliche Gründe hierfür konnten drei Ursachen ausgemacht werden: (1) die Klasse „Kleine Objekte“ besitzt keine eindeutigen und unterscheidbaren Merkmale. Sie beinhaltet zwar Objekte mit einem Alleinstellungsmerkmal wie Garagenreihen und Carports, aber auf der anderen Seite sind solche Elemente oftmals Anbauten eines größeren Objektes, welche sich dann entsprechend eine Dachform und -farbe teilen, was in einer fehlenden Merkmalsausprägung resultiert. (2) Für die Klasse „Kleine Objekte“ standen zu wenig *Training Samples* zur Verfügung. Durch die geringe Anzahl an *Samples* kann das Netz möglicherweise nur unzureichend trainiert werden. (3) Die gewählte *Samples-Patch*-Größe ist zu groß. Es wurde eine feste Fenstergröße für die *Samples* aller drei Klassen gewählt. Dadurch sind bei kleinen Objekten viele Hintergrundpixel, also nicht dem gesuchten Objekt zugehörige Pixel, in den *Samples-Patches* abgebildet. Da sich die zu klassifizierenden Objekte erfahrungsgemäß oft in unmittelbarer Nähe von größeren Objekten wie Wohngebäude befinden, besteht die Möglichkeit, dass das CNN hierbei falsche Charakteristiken ableitet.

Durch die Änderung auf eine binäre Klassifikation konnte anschließend eine sehr hohe Gesamtgenauigkeit erreicht werden. Diese wurde sogar noch durch eine nachträgliche objektbasierte Verfeinerung gesteigert. Die nachgelagerte Betrachtung und Berücksichtigung von Indizes aus den Bändern beziehungsweise Hilfsdaten können helfen Fehlklassifikationen aus dem CNN zumindest teilweise auszugleichen. Durch die OBIA-Verfeinerung wird eine Gesamtgenauigkeit von 94,5 % erreicht (vgl. 5.1.2).

Das Teilziel der Ableitung der Gebäudeobjekte auf Basis von TrueDOP kann abschließend wie folgt zusammengefasst werden: Eine Klassifikation von Gebäudeobjekten aus 4-Kanal-TrueDOPs ist mit einer sehr hohen Güte möglich und kann präzise Ergebnisse liefern. Allerdings treten auch Fehlklassifikationen auf. Die Unterteilung der Trainingsklassen ist nicht zwingend von deren Objektgröße abhängig. Eine zu klassifizierende Objektart muss nicht entsprechend ihren Größenvorkommnissen unterteilt

werden, wenn das Ziel eine größenunabhängige Ableitung ist, wie im Fall dieser Thesis. Ein robust vorbereitetes CNN mit ausreichender Anzahl an *Samples* kann alle Gebäude unabhängig ihrer Größe und Form trainieren und erreicht in dieser Thesis eine Genauigkeit von über 90 %. Die Erkennung von „*Nicht-Gebäuden*“ hat sogar noch einen höheren Wert erreicht. Die vorhandenen *Ground-Truth*-Daten der Klasse „*Nicht-Gebäude*“ im Kontrollgebiet wurden vom CNN mit einer Genauigkeit von 99,2 % klassifiziert.

Als zweites großes Teilziel wurden die Ableitungsmöglichkeiten von Gebäudeobjekten aus den *PlanetScope*-Daten untersucht und bewertet. Aufgrund der im Vergleich geringeren Bodenauflösung dieses Rasterdatensatzes wurden Anpassungen bezüglich der Trainingseinstellungen durchgeführt (vgl. 4.3). Die erzielten Ergebnisse zeigen, dass auch auf Basis dieser Daten eine Ableitung von Gebäudeobjekten in einer hohen Qualität möglich ist. Die dabei angewendete binäre Klassifikation erreicht im ersten Durchlauf eine Gesamtgenauigkeit von 88,4 %. Dabei ist allerdings anzumerken, dass eine deutliche Überklassifikation der Klasse „*Gebäudeobjekte*“ auftritt. Eine fehlerfreie Klassifikation von einzelnen Bildelementen in dicht bebauten Gebieten konnte nicht erreicht werden. Die Kontrolldaten für die Klasse „*Nicht-Gebäude*“ liegen in Form von quadratischen Flächen mit Pixelabmessungen analog zu den Trainingsdaten vor. Die Problematik der zusammenhängenden Gebäudeobjekte spiegelt sich so nicht ohne weiteres in der statistischen Genauigkeitsbewertung wider. Dieser Umstand bedarf einer zusätzlichen visuellen Bewertung. So werden Straßen zwischen Wohnhäusern oftmals als Gebäudeobjekte fehlklassifiziert. Eine Anpassung der Trainingszeiten um den Faktor fünf auf knapp 30 Stunden führte zu einer schärferen Abgrenzung zwischen den einzelnen Objekten und einer korrekten Zuordnung der Zwischenräume zur Klasse „*Nicht-Gebäude*“. Allerdings wurden auch viele Gebäudeobjekte dadurch der falschen Klasse zugewiesen. Die Gesamtgenauigkeit fiel auf 77,7 % und lag damit im mittleren Bereich.

Anpassungen durch das Einbringen von a-priori-Wissen durch einen nachgelagerten regelbasierten Ansatz konnten die Gesamtgenauigkeit von 77,7 % auf 84,4 % erhöhen.

Fehlklassifikationen traten auch in Bezug der generellen Detektierbarkeit der Objekte des Neubaugebietes auf. Dem Klassifikationsprozess gelang es nicht, die Mischbebauung, die sich deutlich im Kontrollgebiet von den anderen Objekten abhebt, zu erkennen. Die dort befindlichen Objekte, die teilweise im Bau und teilweise bereits abgeschlossen sind, konnten nicht fehlerfrei klassifiziert werden.

Im Rahmen der Thesis konnten Defizite in der Hard- und Softwareausstattung ausgemacht werden. Diese äußerten sich vor allem bei den Untersuchungen auf Basis der flugzeuggestützten Bilddaten in Form von langen Lade- und Prozessierungszeiten und damit verbundenen häufigen Softwareabstürzen.

Die Probleme bei der Verwendung der *PlanetScope*-Daten beliefen sich weniger auf deren Handhabung als vielmehr auf die Qualität der Ausgangsdaten. Es erfolgte keine weitere radiometrische Anpassung der gelieferten Rohdaten und die Aufnahmen zeigten bei hellen Objekten stellenweise eine deutliche Überbelichtung und vereinzelt traten Überstrahlungsfragmente auf (vgl. 2.2.3).

Beide untersuchten Datensätze zeigten eine deutliche Fehlklassifikation von Infrastrukturelementen, wie Straßen. Die spektralen Eigenschaften von Straßen ähneln denen von Gebäuden. Mithilfe von objektbasierter Klassifikation wurde über Indizes wie dem NDSI und dem NDVI versucht, die Daten nachträglich zu korrigieren und die Erkennung dadurch zu verbessern. Auch die Verwendung von Höheninformationen kann hier Abhilfe schaffen.

7 Schlussfolgerung und Ausblick

Der direkte Vergleich der Erkennung von Gebäudeobjekten aus TrueDOP und aus den Satellitendaten von *PlanetScope* hat gezeigt, dass erwartungsgemäß die flugzeuggestützten Rasterdaten eine höhere Genauigkeit erreichen als die *PlanetScope*-Daten. Allerdings konnte festgestellt werden, dass die erreichte Genauigkeit dabei nicht so hoch ist, wie es durch die erheblichen Unterschiede in den jeweiligen Bodenauflösungen zu erwarten gewesen wäre. Die höchste erreichte Gesamtgenauigkeit der TrueDOP liegt bei 94,6 % im Vergleich zu 88,4 % bei den *PlanetScope*-Daten. Die Ergebnisse liegen somit nah beieinander, obwohl die Bodenauflösung um den Faktor 15 gröber ist.

Die zeitlich besser aufgelösten Satellitendaten besitzen sogar Vorteile gegenüber den räumlich besser aufgelösten flugzeuggestützten Bildaufnahmen. Durch die geringere Auflösung können mehr *Samples* in kürzerer Zeit generiert und erweitert werden. Der Speicherplatz dieser *Samples* ist im Vergleich geringer und die Generierung und Speicherung kommt somit der Ausstattung eines handelsüblichen PCs entgegen. Allerdings ist hier zu bedenken, dass ein Training von mehr als doppelt so vielen *Samples* der Satellitendaten im Vergleich zu den der Orthophotos (400.000 zu 160.000) trotzdem ebenfalls zu einer sehr langen Laufzeit führt. Diese ist zwar im Vergleich zu den hochauflösenden *Samples* der TrueDOP geringer, durch die vielen kleinen *Samples* werden aber auch erhebliche Rechenzeiten im zweistelligen Stundenbereich erreicht. Je nach Parameterwahl führen längere Trainingszeiten allerdings nicht zwingend zu besseren Ergebnissen (vgl. Abschnitt 5). Die Ergebnisse zeigen, dass eine Ableitung von Gebäudeobjekten aus gröber aufgelösten Satellitendaten mit einer hohen Güte möglich ist, aber Abstriche in der Erkennungsqualität in Kauf genommen werden müssen.

Die vielen unterschiedlichen Ausprägungen eines Gebäudeobjektes hinsichtlich Form, Größe und Farbe führen zu einer hohen Heterogenität der Klasse. Dieser Umstand hat große Auswirkungen auf die Klassifikationsqualität. Weitere Faktoren, wie atmosphärische Einflüsse, Schattenwurf, Jahreszeit und Sonnenstand erschweren eine hochwertige Klassifikation zusätzlich. Zudem besitzen Gebäude ähnliche spektrale Eigenschaften wie Straßen, wodurch es häufig zu Verwechslungen zwischen dem Hoch- und dem Tiefbau kommt, wie in den Ergebnissen zu beobachten war. Diesem Problem kann entgegengesteuert werden, indem Höheninformationen wie ein nDSM genutzt werden oder mithilfe von Indizes versucht wird, die spektralen Überschneidungen auszugleichen. Um der Heterogenität der Gebäudeobjekte Genüge zu tun, bietet sich die Erhöhung der Anzahl von Trainingsdaten an. Damit sollen möglichst viele Variationen von Gebäudeobjekten abgedeckt werden, auch wenn mit einer deutlichen Erhöhung der Trainingsdauer und der Berechnungslaufzeit zu rechnen ist.

Der Landesvermessung von Baden-Württemberg stehen mehrere Millionen Gebäudeobjekte aus dem ALKIS-Gebäudedatenbestand als Referenzdaten und ab 2023 das flächendeckende Bildmaterial der DOP-Daten als Trainingsdaten zur Verfügung. Leider erfüllen die technischen vorliegenden Hardwarevoraussetzungen nicht die notwendigen Kapazitäten, um eine Ableitung auf Basis der CNN-Klassifikation für große Flächen in BW zu ermöglichen. Die Lade- und Berechnungszeiten von Segmentierungen oder für die Generierung von *Samples* erhöhen sich mit zunehmender Untersuchungsfläche deutlich. Eine lokale Umsetzung für einen Teil eines Landkreises auf TrueDOP-Basis ist mit den zur Verfügung stehenden Mitteln möglich und realistisch. Damit aber flächendeckende Ableitungen erreicht werden können, müssen weitere technische Anforderungen erfüllt sein.

Anders sieht es bei den Satellitendaten aus: Durch die geringe Hardware- und Speicherlast beim Vorhalten der Daten und bei der Generierung der *Training Samples* kann eine landkreisweite Ableitung mit den aktuellen Mitteln umgesetzt werden. Wie sich allerdings ein erweiterter Trainingsdatensatz auf die Prozessierungszeiten auswirkt, muss in weiteren Untersuchungen geklärt werden.

Ein Vorteil seitens der *PlanetScope*-Daten liegt in der möglicherweise geringen Anzahl an zu trainierenden *Samples*, da die *PlanetScope*-Daten eine größere Fläche täglich erfassen können. Externe Einflüsse wie Farbunterschiede und Sonnenstand aufgrund von differierenden Befliegungszeiten wirken sich hier nicht so stark aus. So muss nicht prinzipiell für jeden Befliegungstag ein neuer Trainingsdatensatz erstellt werden, wie das bei den DOP vermutlich der Fall wäre. Allerdings gilt es zu beachten, dass die Erkennungsqualität der satellitengestützten *PlanetScope*-Daten nicht an die Genauigkeit von flugzeuggestützten Aufnahmen heranreicht.

Somit kann die ursprünglich aufgestellte Frage mithilfe der erarbeiteten Ergebnisse zu den Erkennungsqualitäten für die flugzeug- und satellitengestützten Bildaufnahmen und ihrer Erkennungsqualität wie folgt beantwortet werden:

Die Klassifikation von Gebäudeobjekten aus TrueDOP erreicht eine sehr hohe Genauigkeit, benötigt aber viel Speicherplatz, große Rechenleistung und lange Berechnungszeiten.

Die Klassifikation von Gebäudeobjekten aus *PlanetScope*-Daten erreicht eine hohe Genauigkeit, allerdings mit Einschränkungen: Anbauten, Überdachungen und andere kleine Objekte waren mit den vorhandenen Möglichkeiten nicht eindeutig klassifizierbar. Die Erkennbarkeit von Einzelobjekten ist gut, vorausgesetzt die Trainingsdaten und -parameter sind von hoher Qualität.

Als Ausblick können folgende Verbesserungsmöglichkeiten für zukünftige Untersuchungen angegeben werden:

Möglicherweise können Dynamische Neuronale Netze (Han et al. 2021) in Zukunft das Problem der unterschiedlichen Objektgrößen mit der Möglichkeit einer Generierung von Trainingsdaten mit unterschiedlicher *Samples*-Größe ausgleichen. Alternativ sollten weitere Untersuchungen die hier verwendete *Samples*-Größe kritisch betrachten. Die Verwendung von anderen *Sample Patch*-Größen, die nicht auf der Annahme basieren, dass ein Wohngebäude womöglich vollständig von einem *Sample*-Ausschnitt erfasst ist, ist naheliegend.

Allgemein gesprochen, würden sich weitere Untersuchungen zur Kernfrage, ob und wie die Erkennbarkeit von Objekten mit einer Fläche kleiner als 20 m² in TrueDOP gesteigert werden kann, anbieten.

Die Übertragbarkeit der entwickelten Workflows auf geographisch andere Gebiete wurde in dieser Thesis nicht untersucht. Eine zusätzliche Untersuchung zur Erprobung des trainierten CNN der DOP- und SAT-Daten auf andere Bereiche innerhalb des Landkreises ist deswegen sinnvoll. Das trainierte Netz kann auch auf Bereiche außerhalb des Landkreises angewendet werden, um die Möglichkeiten für Gebiete unterschiedlicher Aufnahmezeitpunkte auszuloten.

Im Falle einer Hardwareaufrüstung können für weitere Untersuchungen mehr *Samples* erstellt und darauf aufbauend der Fokus auf eine Klassifikation auf Basis der Gebäudefunktion gelegt werden. Vorstellbar wäre hierbei die Ableitung von Gebäudefunktionen wie etwa Wohnhäuser, Carports, Überdachungen oder Industrie. Dieser Aspekt wurde im Rahmen der Thesis nicht untersucht, da die vorhandenen *Samples* keine eindeutigen Charakteristiken der Gebäudefunktion abdeckten.

Möglicherweise können Anpassungen an der *Samples*-Größe auch bei den Satellitendaten zu besseren Ergebnissen führen. Diese Hypothese konnte nicht untersucht werden, da das Hauptaugenmerk der Arbeit auf der Vergleichbarkeit der beiden Rasterdatensätze lag und die verwendete Größe in beiden Fällen ähnlich gehalten wurde.

Viele Tiefbauobjekte wurden fälschlicherweise der Klasse „Gebäudeobjekte“ zugeordnet, was in einer Überrepräsentation der Gebäudeklasse resultierte. Nachfolgende Untersuchungen können dieser Problematik gegensteuern, indem gesonderte Trainingsdaten für Infrastrukturobjekte wie Straßen eingeführt werden, wodurch das CNN Unterschiede besser feststellen kann.

Abschließend bleibt die Frage unbeantwortet, ob durch die Erhöhung der *Sample*anzahl innerhalb eines begrenzten Trainingsgebietes zu einer höheren Erkennungsqualität führen würde. Auch der Einfluss der Trainingsdauer konnte nicht tiefergehend untersucht werden. Die Annahme, dass eine längere Trainingszeit automatisch zu besseren Klassifikationsergebnissen führt, konnte in der Thesis nicht bestätigt werden. Weitere Untersuchungen hierzu bieten sich an.

Literaturverzeichnis

- AdV-Projektgruppe LB/LN (2020). Erläuterungen zum Anwendungsschema Landbedeckung (LB). AAA-Anwendungsscheman 7.1. Online verfügbar unter <https://www.adv-online.de/GeoInfoDok/GeoInfoDok-NEU/Landbedeckung/binarywriterservlet?imgUid=76e2d078-9d3f-5711-1fea-f5203b36c4c2&uBasVariant=11111111-1111-1111-1111-111111111111> (abgerufen am 13.11.2021).
- Ahmad, A./Quegan, S. (2012). Analysis of Maximum Likelihood classification technique on Landsat 5 TM satellite data of tropical land covers. In: 2012 IEEE International Conference on Control System, Computing and Engineering, 2012 IEEE International Conference on Control System, Computing and Engineering, 280–285.
- Albawi, Saad/Mohammed, Tareq Abed/Al-Zawi, Saad (2017). Understanding of a convolutional neural network. In: Proceedings of 2017 International Conference on Engineering & Technology (ICET'2017). Akdeniz University, Antalya, Turkey, 21-23 August 2017, 2017 International Conference on Engineering and Technology (ICET), Antalya, 8/21/2017 - 8/23/2017. Piscataway, NJ, IEEE, 1–6.
- Alidoost, F./Arefi, H. (2016). Knowledge Based 3D Building Model Recognition Using Convolutional Neural Networks From LiDAR And Aerial Imageries. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B3, 833–840. <https://doi.org/10.5194/isprsarchives-XLI-B3-833-2016>.
- Alidoost, Fatemeh/Arefi, Hossein (2018). A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 86 (5-6), 235–248. <https://doi.org/10.1007/s41064-018-0060-5>.
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2018). Tätigkeitsbericht 2017/2018. Amtliches deutsches Vermessungswesen. Online verfügbar unter <https://www.adv-online.de/Veroeffentlichungen/Taetigkeitsberichte-der-AdV/> (abgerufen am 22.01.2022).
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2019a). Leitfaden zur Ausschreibung einer Luftbildbefliegung für die Zwecke der Landesvermessung. 32. Tagung AdV-Arbeitskreis Geotopographie, Beschluss GT 2019/02. Online verfügbar unter <https://www.adv-online.de/icc/extdeu/nav/8dd/binarywriterservlet?imgUid=2d220307-0b71-ee71-7657-80b6a757628a&uBasVariant=11111111-1111-1111-1111-111111111111> (abgerufen am 24.03.2022).
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2019b). Leitfaden zur Qualitätssicherung von True Orthophotos (TrueDOP). 32. Tagung AdV-Arbeitskreis Geotopographie, Beschluss GT 2019/07. Online verfügbar unter <https://www.adv-online.de/AdV-Produkte/Standards-und-Produktblaetter/Standards-der-Geotopographie/binarywriterservlet?imgUid=1e220307-0b71-ee71-7657-80b6a757628a&uBasVariant=11111111-1111-1111-1111-111111111111> (abgerufen am 20.12.2021).
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2019c). Produkt- und Qualitätsstandard für Digitale Orthophotos. Version 4.0. Online verfügbar unter <https://www.adv-online.de/binarywriterservlet?imgUid=87940abe-c42d-ba61-cdcd-175401fa2e0c&uBasVariant=11111111-1111-1111-1111->

- 111111111111#:~:text=Digitale%20Orthophotos%20(DOP)%20sind%20grunds%C3%A4tzlich,ei%2D%20nem%20Digitalen%20H%C3%B6henmodell%20abgeleitet. (abgerufen am 22.01.2022).
- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (2022). Liegenschaftskataster. Online verfügbar unter <https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/> (abgerufen am 22.01.2022).
- Audebert, Nicolas/Boulch, Alexandre/Randrianarivo, Hicham/Le Saux, Bertrand/Ferecatu, Marin/Lefevre, Sebastien/Marlet, Renaud (2017). Deep learning for urban remote sensing. In: 2017 Joint Urban Remote Sensing Event (JURSE). IEEE.
- Aytekin, Örsan/Erener, Arzu/Ulusoy, İlkey/Düzgün, Şebnem (2012). Unsupervised building detection in complex urban environments from multispectral satellite imagery. *International Journal of Remote Sensing* 33 (7), 2152–2177. <https://doi.org/10.1080/01431161.2011.606852>.
- Baatz, M./Schäpe, A. (2000). Multiresolution Segmentation : an optimization approach for high quality multi-scale image segmentation. Online verfügbar unter <https://www.semanticscholar.org/paper/Multiresolution-Segmentation-%3A-an-optimization-for-Baatz-Sch%C3%A4pe/364cc1ff514a2e11d21a101dc072575e5487d17e>.
- Baltrusch, S. (2016). TrueDOP - A New Quality Step For Official Orthophotos. 2194-9034 XLI-B4 (66), 619–624. <https://doi.org/10.5194/isprsarchives-XLI-B4-619-2016>.
- Belgiu, M./Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment* 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>.
- Belgiu, Mariana/Drăguț, Lucian (2014). Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, 67–75. <https://doi.org/10.1016/j.isprsjprs.2014.07.002>.
- Benz, Ursula C./Hofmann, Peter/Willhauck, Gregor/Lingenfelder, Iris/Heynen, Markus (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4), 239–258. <https://doi.org/10.1016/j.isprsjprs.2003.10.002>.
- Beveridge, J. Ross/Griffith, Joey/Kohler, Ralf R./Hanson, Allen R./Riseman, Edward M. (1989). Segmenting images using localized histograms and region merging. *International Journal of Computer Vision* 2 (3), 311–347. <https://doi.org/10.1007/BF00158168>.
- Bezdek, James C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, Springer.
- Bill, Ralf (2016). *Grundlagen der Geo-Informationssysteme*. 6. Aufl. Berlin/Offenbach, Wichmann.
- Blaschke, Thomas (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1), 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>.
- Blaschke, Thomas/Burnett, Charles/Pekkarinen, Anssi (2004). Image Segmentation Methods for Object-based Analysis and Classification. In: Steven M. de Jong/Freek D. van der Meer (Hg.). *Remote sensing image analysis. Including the spatial domain*. Dordrecht, Springer; Kluwer Acad. Publ, 211–236.
- Blaschke, Thomas/Feizizadeh, Bakhtiar/Holbling, Daniel (2014a). Object-Based Image Analysis and Digital Terrain Analysis for Locating Landslides in the Urmia Lake Basin, Iran. *IEEE Journal of*

- Selected Topics in Applied Earth Observations and Remote Sensing 7 (12), 4806–4817.
<https://doi.org/10.1109/jstars.2014.2350036>.
- Blaschke, Thomas/Hay, Geoffrey J./Kelly, Maggi/Lang, Stefan/Hofmann, Peter/Addink, Elisabeth/Queiroz Feitosa, Raul/van der Meer, Freek/van der Werff, Harald/van Coillie, Frieke/Tiede, Dirk (2014b). Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87 (100), 180–191.
<https://doi.org/10.1016/j.isprsjprs.2013.09.014>.
- Blaschke, Thomas/Kelly, Maggi/Merschdorf, Helena (2015). Object Based Image Analysis: Evolution, History, State-of-the-Art and Future Vision. In: *Remote Sensing Handbook; Remotely Data Characterization, Classification, and Accuracies*, 275–290.
- Blaschke, Thomas/Lang, Stefan/Hay, Geoffrey (2008). Object-based image analysis. Spatial concepts for knowledge-driven remote sensing applications. Berlin, Heidelberg, Springer Berlin Heidelberg.
- Blaschke, Thomas/Strobl, Josef (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS [Was ist mit den Pixeln los? Neue Entwicklungen zur Integration von Fernerkundung und GIS]. *Geo-Information-Systeme* 14 (6), 12–17. Online verfügbar unter <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035382401&partnerID=40&md5=5af9148739dd984bcffb99e2cec9dd46>.
- BMI (2017). Fernerkundung - Entscheidungs-hilfe durch Erd-beobachtung. Bundesministerium des Innern und für Heimat vom 25.09.2017. Online verfügbar unter <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/geoinformationen/fernerkundung/fernerkundung-node.html> (abgerufen am 09.04.2022).
- Boonpook, Wuttichai/Tan, Yumin/Ye, Yinghua/Torteeeka, Peerapong/Torsri, Kritanai/Dong, Shengxian (2018). A Deep Learning Approach on Building Detection from Unmanned Aerial Vehicle-Based Images in Riverbank Monitoring. *Sensors* 18 (11), 3921. <https://doi.org/10.3390/s18113921>.
- Breiman, Leo (2001). Random Forests. *Machine Learning* 45 (1), 5–32.
<https://doi.org/10.1023/A:1010933404324>.
- Carranza-García, M./García-Gutiérrez, J./Riquelme, J. C. (2019). A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing* 11 (3).
<https://doi.org/10.3390/rs11030274>.
- Chang, Chih-Chung/Lin, Chih-Jen (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology* 2 (3), 1–27. <https://doi.org/10.1145/1961189.1961199>.
- Chen, Qi/Wang, Lei/Waslander, Steven L./Liu, Xiuguo (2020). An end-to-end shape modeling framework for vectorized building outline generation from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 170, 114–126. <https://doi.org/10.1016/j.isprsjprs.2020.10.008>.
- Chen, Qi/Zhang, Yuanyi/Li, Xinyuan/Tao, Pengjie (2021). Extracting Rectified Building Footprints from Traditional Orthophotos: A New Workflow. *Sensors (Basel, Switzerland)* 22 (1).
<https://doi.org/10.3390/s22010207>.
- Chen, Yangyang/Ming, Dongping/Lv, Xianwei (2019). Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation. Springer. Online verfügbar unter <https://link.springer.com/article/10.1007/s12145-019-00383-2> (abgerufen am 30.01.2022).

- Ciresan, D./Meier, U./Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012). Providence, Rhode Island, USA, 16 - 24 [i.e. 16 - 21] June 2012, 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, 6/16/2012 - 6/21/2012. Piscataway, NJ, IEEE, 3642–3649.
- Congalton, Russell G. (2019). *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*, Third Edition. 3. Aufl. Milton, Chapman and Hall/CRC.
- Cortes, Corinna/Vapnik, Vladimir (1995). Support-vector networks. *Machine Learning* 20 (3), 273–297. <https://doi.org/10.1007/BF00994018>.
- Cui, S. Y./Yan, Q./Liu, Z. J. (2008). Building Detection and Recognition from High Resolution Remotely Sensed Imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Online verfügbar unter https://www.isprs.org/proceedings/XXXVII/congress/3b_pdf/79.pdf (abgerufen am 16.03.2022).
- Dao, Phuong/Liou, Yuei-An (2015). Object-Based Flood Mapping and Affected Rice Field Estimation with Landsat 8 OLI and MODIS Data. *Remote Sensing* 7 (5), 5077–5097. <https://doi.org/10.3390/rs70505077>.
- Deng, Jia/Dong, Wei/Socher, Richard/Li, Li-Jia/Li, Kai/Fei-Fei, Li (2009). ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009 ; Miami [Beach], Florida, USA, 20 - 25 June 2009, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), Miami, FL, 6/20/2009 - 6/25/2009. Piscataway, NJ, IEEE, 248–255.
- Deng, Li/Dong, Yu (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing* 7 (3-4), 197–387. <https://doi.org/10.1561/20000000039>.
- Drăguț, L./Csillik, O./Eisank, C./Tiede, D. (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing* 88 (100), 119–127. <https://doi.org/10.1016/j.isprsjprs.2013.11.018>.
- Drăguț, Lucian/Tiede, Dirk/Levick, Shaun R. (2010). ESP: a tool to estimate scale parameter for multi-resolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science* 24 (6), 859–871. <https://doi.org/10.1080/13658810903174803>.
- El Naqa, Issam/Murphy, Martin J. (2015). What Is Machine Learning? In: Issam El Naqa/Ruijiang Li/Martin J. Murphy (Hg.). *Machine Learning in Radiation Oncology. Theory and Applications*. Cham/s.l., Springer International Publishing, 3–11.
- El-naggar, Aly M. (2018). Determination of optimum segmentation parameter values for extracting building from remote sensing images. *Alexandria Engineering Journal* 57 (4), 3089–3097. <https://doi.org/10.1016/j.aej.2018.10.001>.
- ESA (2022). *Newcomers Earth Observation Guide | ESA Business Applications*. Online verfügbar unter <https://business.esa.int/newcomers-earth-observation-guide> (abgerufen am 08.04.2022).
- Feng, Tiantian/Zhao, Junqiao (2009). Review and Comparison: Building Extraction Methods Using High-Resolution Images. In: Wen Chen (Hg.). 2009 Second International Symposium on Information Science and Engineering. (ISISE 2009) ; Shanghai, China, 26 - 28 December 2009, 2009 International Symposium on Information Science and Engineering (ISISE), Shanghai, China, 12/26/2009 - 12/28/2009. Piscataway, NJ, IEEE, 419–422.

- Ferro, Adamo/Brunner, Dominik/Bruzzone, Lorenzo (2013). Automatic Detection and Reconstruction of Building Radar Footprints From Single VHR SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 51 (2), 935–952. <https://doi.org/10.1109/tgrs.2012.2205156>.
- Foody, Giles M./Mathur, Ajay/Sanchez-Hernandez, Carolina/Boyd, Doreen S. (2006). Training set size requirements for the classification of a specific class. *Remote Sensing of Environment* 104 (1), 1–14. <https://doi.org/10.1016/j.rse.2006.03.004>.
- Fradkin, M./Maître, H./Roux, M. (2001). Building Detection from Multiple Aerial Images in Dense Urban Areas. *Computer Vision and Image Understanding* 82 (3), 181–207. <https://doi.org/10.1006/cviu.2001.0917>.
- Fujita, Aito/Sakurada, Ken/Imaizumi, Tomoyuki/Ito, Riho/Hikosaka, Shuhei/Nakamura, Ryosuke (2017). Damage detection from aerial images via convolutional neural networks. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). IEEE.
- Fukushima, Kunihiro/Miyake, Sei (1982). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition. In: *Competition and Cooperation in Neural Nets*. Springer, Berlin, Heidelberg, 267–285.
- Gao, Yan/Mas, Jean Francois/Kerle, Norman/Navarrete Pacheco, Jose Antonio (2011). Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing* 32 (13), 3747–3763. <https://doi.org/10.1080/01431161003777189>.
- GEObasis.nrw (2020). Orthophotos in der Qualitätsstufe TrueDOP. Online verfügbar unter <https://docplayer.org/186326128-Orthophotos-in-der-qualitaetsstufe-truedop.html> (abgerufen am 09.04.2022).
- Ghanea, Mohsen/Moallem, Payman/Momeni, Mehdi (2016). Building extraction from high-resolution satellite images in urban areas: recent methods and strategies against significant challenges. *International Journal of Remote Sensing* 37 (21), 5234–5248. <https://doi.org/10.1080/01431161.2016.1230287>.
- Ghorbanzadeh, Omid/Blaschke, Thomas/Gholamnia, Khalil/Meena, Sansar/Tiede, Dirk/Aryal, Jagannath (2019). Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing* 11 (2), 196. <https://doi.org/10.3390/rs11020196>.
- Ghorbanzadeh, Omid/Tiede, Dirk/Wendt, Lorenz/Sudmanns, Martin/Lang, Stefan (2021). Transferable instance segmentation of dwellings in a refugee camp - integrating CNN and OBIA. *European Journal of Remote Sensing* 54 (sup1), 127–140. <https://doi.org/10.1080/22797254.2020.1759456>.
- Girshick, Ross/Donahue, Jeff/Darrell, Trevor/Malik, Jitendra (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. Online verfügbar unter <https://arxiv.org/pdf/1311.2524>.
- Graf, Roland J./Wegenkittl, Stefan (2012). Integration von Support Vector Machines in die objektbasierte Bildklassifizierung am Beispiel der Entwicklung eines Plug-Ins für eCognition. Online verfügbar unter https://www.researchgate.net/publication/262698139_Integration_von_Support_Vector_Machine_s_in_die_objektbasierte_Bildklassifizierung_am_Beispiel_der_Entwicklung_eines_Plug-Ins_fur_eCognition.

- Guo, Qinghua/Kelly, Maggi/Gong, Peng/Liu, Desheng (2007). An Object-Based Classification Approach in Mapping Tree Mortality Using High Spatial Resolution Imagery. *GIScience & Remote Sensing* 44 (1), 24–47. <https://doi.org/10.2747/1548-1603.44.1.24>.
- Hairie, Ilkham Sibaruddin/Helmi, Zulhaidi Mohd Shafri/Biswajeet, Pradhan/Nuzul, Azam Haron (2018). UAV-based Approach to Extract Topographic and As-built Information by Utilising the OBIA Technique. *Journal of Geosciences and Geomatics* 6 (3), 103–123. <https://doi.org/10.12691/jgg-6-3-2>.
- Han, Dongmei/Liu, Qigang/Fan, Weiguo (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications* 95, 43–56. <https://doi.org/10.1016/j.eswa.2017.11.028>.
- Han, Yizeng/Huang, Gao/Song, Shiji/Le Yang/Wang, Honghui/Wang, Yulin (2021). Dynamic Neural Networks: A Survey. Online verfügbar unter <https://arxiv.org/pdf/2102.04906>.
- Haris, K./Efstratiadis, S. N./Maglaveras, N./Katsaggelos, A. K. (1998). Hybrid image segmentation using watersheds and fast region merging. *IEEE Transactions on Image Processing* 7 (12), 1684–1699. <https://doi.org/10.1109/83.730380>.
- Hashem, Nadeem/Balakrishnan, Perumal (2015). Change analysis of land use/land cover and modelling urban growth in Greater Doha, Qatar. *Annals of GIS* 21 (3), 233–247. <https://doi.org/10.1080/19475683.2014.992369>.
- Hay, G. J./Castilla, G. (2008). Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: *Object-Based Image Analysis*. Springer, Berlin, Heidelberg, 75–89.
- Hay, G. J./Marceau, D. J./Dubé, P./Bouchard, A. (2001). A multiscale framework for landscape analysis: Object-specific analysis and upscaling. *Landscape Ecology* 16 (6), 471–490. <https://doi.org/10.1023/A:1013101931793>.
- Hay, Geoffrey J./Blaschke, Thomas/Marceau, Danielle J./Bouchard, André (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing* 57 (5-6), 327–345. [https://doi.org/10.1016/S0924-2716\(02\)00162-4](https://doi.org/10.1016/S0924-2716(02)00162-4).
- Hermosilla, Txomin/Ruiz, Luis A./Recio, Jorge A./Estornell, Javier (2011). Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. *Remote Sensing* 3 (6), 1188–1210. <https://doi.org/10.3390/rs3061188>.
- Hinton, Geoffrey E./Srivastava, Nitish/Krizhevsky, Alex/Sutskever, Ilya/Salakhutdinov, Ruslan R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. Online verfügbar unter <https://arxiv.org/pdf/1207.0580>.
- Hossain, Mohammad D./Chen, Dongmei (2019). Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 150, 115–134. <https://doi.org/10.1016/j.isprsjprs.2019.02.009>.
- Hu, Fan/Xia, Gui-Song/Hu, Jingwen/Zhang, Liangpei (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing* 7 (11), 14680–14707. <https://doi.org/10.3390/rs71114680>.
- Hu, Qintao/Zhen, Liangli/Mao, Yao/Zhou, Xi/Zhou, Guozhong (2021). Automated building extraction using satellite remote sensing imagery. *Automation in Construction* 123, 103509. <https://doi.org/10.1016/j.autcon.2020.103509>.

- Ji, Shunping/Wei, Shiqing/Lu, Meng (2019). A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *International Journal of Remote Sensing* 40 (9), 3308–3322. <https://doi.org/10.1080/01431161.2018.1528024>.
- Jin, Baoxuan/Ye, Peng/Zhang, Xueying/Song, Weiwei/Li, Shihua (2019). Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images. *Journal of the Indian Society of Remote Sensing* 47 (6), 951–965. <https://doi.org/10.1007/s12524-019-00945-3>.
- Jin, Xiaoying/Davis, Curt H. (2005). Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information. *EURASIP Journal on Advances in Signal Processing* 2005 (14), 1–11. <https://doi.org/10.1155/ASP.2005.2196>.
- Jog, Sayali/Dixit, Mrudul (2016). Supervised classification of satellite images. 2016 Conference on Advances in Signal Processing (CASP), 93–98. <https://doi.org/10.1109/CASP.2016.7746144>.
- Jozdani, Shahab Eddin/Johnson, Brian Alan/Chen, Dongmei (2019). Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. *Remote Sensing* 11 (14), 1713. <https://doi.org/10.3390/rs11141713>.
- Kamiński, Bogumił/Jakubczyk, Michał/Szufel, Przemysław (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research* 26 (1), 135–159. <https://doi.org/10.1007/s10100-017-0479-6>.
- Karpathy, Andrej/Toderici, George/Shetty, Sanketh/Leung, Thomas/Sukthankar, Rahul/Fei-Fei, Li (2014). Large-Scale Video Classification with Convolutional Neural Networks, 1725–1732. <https://doi.org/10.1109/CVPR.2014.223>.
- Kattenborn, Teja/Leitloff, Jens/Schiefer, Felix/Hinz, Stefan (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>.
- Kohavi, R./Provost, F. (1998). Glossary of terms. *Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. Machine Learning. Machine Learning* 30 (2/3), 271–274. <https://doi.org/10.1023/A:1017181826899>.
- Kriesten, Markus (2017). *Vermessungsrecht, Grenzstreitigkeiten und Recht der Öffentlich bestellten Vermessungsingenieure*. Richard Boorberg Verlag.
- Krizhevsky, Alex/Sutskever, Ilya/Hinton, Geoffrey E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60 (6), 84–90. <https://doi.org/10.1145/3065386>.
- Lang, Stefan/Schöpfer, Elisabeth/Hölbling, Daniel/Blaschke, Thomas/Moeller, Matthias/Jekel, Thomas/Kloyber, Elisabeth (2008). Quantifying and Qualifying Urban Green by Integrating Remote Sensing, GIS, and Social Science Method. In: *Use of Landscape Sciences for the Assessment of Environmental Security*. Springer, Dordrecht, 93–105.
- Lange, Norbert de (2020). *Geoinformatik in Theorie und Praxis. Grundlagen von Geoinformationssystemen, Fernerkundung und digitaler Bildverarbeitung*. 4. Aufl. Berlin/Heidelberg, Springer Spektrum.
- Längkvist, Martin/Kiselev, Andrey/Alirezaie, Marjan/Loutfi, Amy (2016). Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. *Remote Sensing* 8 (4), 329. <https://doi.org/10.3390/rs8040329>.

- LeCun, Yann/Bengio, Yoshua (1995). Convolutional Networks for Images, Speech, and Time-Series. Online verfügbar unter <http://yann.lecun.com/exdb/publis/pdf/lecun-bengio-95a.pdf> (abgerufen am 27.02.2022).
- LeCun, Yann/Bengio, Yoshua/Hinton, Geoffrey (2015). Deep learning. *Nature* 521 (7553), 436–444. <https://doi.org/10.1038/nature14539>.
- LeCun, Yann/Boser, Bernhard E./Denker, John S./Henderson, Donnie/Howard, R. E./Hubbard, Wayne E./Jackel, Lawrence D. (1989). Handwritten Digit Recognition with a Back-Propagation Network. Online verfügbar unter <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf> (abgerufen am 21.03.2022).
- LGL BW (2011). Das AFIS-ALKIS-ATKIS-Konzept. Vom AdV-Konzept zur Implementierung. Online verfügbar unter https://www.lgl-bw.de/export/sites/lgl/unsere-themen/Geoinformation/AFIS-ALKIS-ATKIS/Galerien/Dokumente/01_AAA_Konzept.pdf (abgerufen am 05.02.2022).
- LGL BW (2022). Digitale Landschaftsmodelle (DLM). Online verfügbar unter <https://www.lgl-bw.de/unsere-themen/Geoinformation/Topographie/Digitale-Landschaftsmodelle-DLM/> (abgerufen am 08.04.2022).
- LGL BW (2022b). Digitale Orthophotos. Landesamt für Geoinformation und Landentwicklung Baden-Württemberg. Online verfügbar unter <https://www.lgl-bw.de/Produkte/Geodaten/Digitale-Orthophotos/> (abgerufen am 22.01.2022).
- LGL BW (2022c). Hausumringe. Landesamt für Geoinformation und Landentwicklung Baden-Württemberg. Online verfügbar unter <https://www.lgl-bw.de/Produkte/Liegenschaftskataster/Hausumringe/> (abgerufen am 12.04.2022).
- LGL BW (2022d). Liegenschaftskataster. Landesamt für Geoinformation und Landentwicklung Baden-Württemberg. Online verfügbar unter <https://www.lgl-bw.de/unsere-themen/Geoinformation/Liegenschaftskataster/> (abgerufen am 30.01.2022).
- Li, Qingting/Wang, Cuizhen/Zhang, Bing/Lu, Linlin (2015). Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data. *Remote Sensing* 7 (12), 16091–16107. <https://doi.org/10.3390/rs71215820>.
- Lin, Chungan/Nevatia, Ramakant (1998). Building Detection and Description from a Single Intensity Image. *Computer Vision and Image Understanding* 72 (2), 101–121. <https://doi.org/10.1006/cviu.1998.0724>.
- Lin, Zeming/Lanchantin, Jack/Qi, Yanjun (2016). MUST-CNN: A Multilayer Shift-and-Stitch Deep Convolutional Architecture for Sequence-based Protein Structure Prediction. Online verfügbar unter <https://arxiv.org/pdf/1605.03004>.
- Liu, Desheng/Xia, Fan (2010). Assessing object-based classification: advantages and limitations. *Remote Sensing Letters* 1 (4), 187–194. <https://doi.org/10.1080/01431161003743173>.
- Liu, Li/Ouyang, Wanli/Wang, Xiaogang/Fieguth, Paul/Chen, Jie/Liu, Xinwang/Pietikäinen, Matti (2020). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision* 128 (2), 261–318. <https://doi.org/10.1007/s11263-019-01247-4>.
- Liu, Shengjie/Qi, Zhixin/Li, Xia/Yeh, Anthony (2019). Integration of Convolutional Neural Networks and Object-Based Post-Classification Refinement for Land Use and Land Cover Mapping with Optical and SAR Data. *Remote Sensing* 11 (6), 690. <https://doi.org/10.3390/rs11060690>.

- Liu, Tao/Abd-Elrahman, Amr (2018). Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 139, 154–170. <https://doi.org/10.1016/j.isprsjprs.2018.03.006>.
- LRA Heilbronn (2021). Einwohnerzahlen der Städte und Gemeinden - Landkreis Heilbronn. Statistisches Landesamt. Online verfügbar unter <https://www.landkreis-heilbronn.de/einwohnerzahlen-der-staedte-und-gemeinden.41.htm> (abgerufen am 13.11.2021).
- Luo, Lin/Li, Pengpeng/Yan, Xuesong (2021). Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review. *Energies* 14 (23), 7982. <https://doi.org/10.3390/en14237982>.
- M. Rothermel/K. Wenzel/D. Fritsch/N. Haala (2012). SURE: Photogrammetric Surface Reconstruction from Imagery.
- Ma, Lei/Li, Manchun/Ma, Xiaoxue/Cheng, Liang/Du, Peijun/Liu, Yongxue (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 277–293. <https://doi.org/10.1016/j.isprsjprs.2017.06.001>.
- Ma, Weixuan/Wan, Youchuan/Li, Jiayi/Zhu, Sa/Wang, Mingwei (2019). An Automatic Morphological Attribute Building Extraction Approach for Satellite High Spatial Resolution Imagery. *Remote Sensing* 11 (3), 337. <https://doi.org/10.3390/rs11030337>.
- Mardia, K. V./Hainsworth, T. J. (1988). A spatial thresholding method for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (6), 919–927. <https://doi.org/10.1109/34.9113>.
- Marpu, P. R./Neubert, M./Herold, H./Niemeyer, I. (2010). Enhanced evaluation of image segmentation results. *Journal of Spatial Science* 55 (1), 55–68. <https://doi.org/10.1080/14498596.2010.487850>.
- Meinel, Gotthard/Neubert, M. (2004). A comparison of segmentation programs for high resolution remote sensing data.
- Möller, M./Lymburner, L./Volk, M. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation* 9 (3), 311–321. <https://doi.org/10.1016/j.jag.2006.10.002>.
- Myint, Soe W./Gober, Patricia/Brazel, Anthony/Grossman-Clarke, Susanne/Weng, Qihao (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* 115 (5), 1145–1161. <https://doi.org/10.1016/j.rse.2010.12.017>.
- Nguyen, Kim-Anh/Liou, Yuei-An (2019). Mapping global eco-environment vulnerability due to human and nature disturbances. *MethodsX* 6, 862–875. <https://doi.org/10.1016/j.mex.2019.03.023>.
- Norman, M./Shahar, H. M./Mohamad, Z./Rahim, A./Mohd, F. A./Shafri, H.Z.M. (2021). Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. *IOP Conference Series: Earth and Environmental Science* 620 (1). <https://doi.org/10.1088/1755-1315/620/1/012010>.
- Oczipka, Martin Eckhard (2007). Objektbasierte Klassifizierung hochauflösender Daten in urbanen Räumen unter besonderer Berücksichtigung von Oberflächenmodellen. <https://doi.org/10.17169/REFUBIUM-5353>.

- Ohta, Yu-Ichi/Kanade, Takeo/Sakai, Toshiyuki (1980). Color information for region segmentation. *Computer Graphics and Image Processing* 13 (3), 222–241. [https://doi.org/10.1016/0146-664X\(80\)90047-7](https://doi.org/10.1016/0146-664X(80)90047-7).
- Peña-Barragán, José M./Ngugi, Moffatt K./Plant, Richard E./Six, Johan (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment* 115 (6), 1301–1316. <https://doi.org/10.1016/j.rse.2011.01.009>.
- Perez, Luis/Wang, Jason (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. Online verfügbar unter <https://arxiv.org/pdf/1712.04621>.
- Perona, P./Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (7), 629–639. <https://doi.org/10.1109/34.56205>.
- PlanetLabs (2022). Planet Imagery Product Specification. Online verfügbar unter https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (abgerufen am 24.03.2022).
- Pong, Ting-Chuen/Shapiro, Linda G./Watson, Layne T./Haralick, Robert M. (1984). Experiments in segmentation using a facet model region grower. *Computer Vision, Graphics, and Image Processing* 25 (1), 1–23. [https://doi.org/10.1016/0734-189X\(84\)90046-X](https://doi.org/10.1016/0734-189X(84)90046-X).
- Qin, Rongjun/Fang, Wei (2014). A Hierarchical Building Detection Method for Very High Resolution Remotely Sensed Images Combined with DSM Using Graph Cut Optimization. *American Society for Photogrammetry and Remote Sensing*.
- Radovic, Matija/Adarkwa, Offei/Wang, Qiaosong (2017). Object Recognition in Aerial Images Using Convolutional Neural Networks. *Journal of Imaging* 3 (2), 21.
- Safavian, S. R./Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics* 21 (3), 660–674. <https://doi.org/10.1109/21.97458>.
- Sameen, Maher Ibrahim/Pradhan, Biswajeet (2019). Landslide Detection Using Residual Networks and the Fusion of Spectral and Topographic Information. *IEEE Access* 7, 114363–114373. <https://doi.org/10.1109/ACCESS.2019.2935761>.
- Sarker, Iqbal H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science* 2 (6), 420. <https://doi.org/10.1007/s42979-021-00815-1>.
- Schiewe, J. (2002). Segmentation of high-resolution remotely sensed data-concepts, applications and problems.
- Schuegraf, Philipp/Bittner, Ksenia (2019). Automatic Building Footprint Extraction from Multi-Resolution Remote Sensing Images Using a Hybrid FCN. *ISPRS International Journal of Geo-Information* 8 (4), 191. <https://doi.org/10.3390/ijgi8040191>.
- Scott, Grant J./England, Matthew R./Starns, William A./Marcum, Richard A./Davis, Curt H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters* 14 (4), 549–553. <https://doi.org/10.1109/LGRS.2017.2657778>.
- Shafri, Helmi/Hamedianfar, Alireza (2015). Detailed intra-urban mapping through transferable OBIA rule sets using WorldView-2 very-high-resolution satellite images. Unpublished.

- Shao, Zhenfeng/Tang, Penghao/Wang, Zhongyuan/Saleem, Nayyer/Yam, Sarath/Somma, Chatpong (2020). BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images. *Remote Sensing* 12 (6), 1050. <https://doi.org/10.3390/rs12061050>.
- Shih, Frank Y./Cheng, Shouxian (2004). Adaptive mathematical morphology for edge linking. *Information Sciences* 167 (1-4), 9–21. <https://doi.org/10.1016/j.ins.2003.07.020>.
- Shorten, Connor/Khoshgoftar, Taghi M. (2019). A survey on Image Data Augmentation for Deep Learning. SpringerOpen. Online verfügbar unter <https://link.springer.com/article/10.1186/s40537-019-0197-0?code=a6ae644c-3bfc-43d9-b292-82d77d5890d5> (abgerufen am 10.03.2022).
- Shultz, Thomas R./Fahlman, Scott E./Craw, Susan/Andritsos, Periklis/Tsaparas, Panayiotis/Silva, Ricardo/Drummond, Chris/Ling, Charles X./Sheng, Victor S./Lanzi, Pier Luca/Gama, João/Wiegand, R. Paul/Sen, Prithviraj/Namata, Galileo/Bilgic, Mustafa/Getoor, Lise/He, Jun/Jain, Sanjay/Stephan, Frank/Sammut, Claude/Harries, Michael/Ting, Kai Ming/Pfahring, Bernhard/Case, John/Wagstaff, Kiri L./Nijssen, Siegfried/Wirth, Anthony/Zhang, Xinhua/Cancedda, Nicola/Renders, Jean-Michel/Michelucci, Pietro/Oblinger, Daniel/Keogh, Eamonn/Mueen, Abdullah (2010). Confusion Matrix. In: Claude Sammut/Geoffrey I. Webb (Hg.). *Encyclopedia of Machine Learning*. Boston, MA, Springer US, 209.
- Sirmacek, Beril/Unsalan, Cem (2008). Building detection from aerial images using invariant color features and shadow information. In: 2008 23rd International Symposium on Computer and Information Sciences. IEEE.
- Su, Tengfei/Zhang, Shengwei (2017). Local and global evaluation for remote sensing image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 256–276. <https://doi.org/10.1016/j.isprsjprs.2017.06.003>.
- Talukdar, Swapan/Singha, Pankaj/Mahato, Susanta/Shahfahad/Pal, Swades/Liou, Yuei-An/Rahman, Atiqur (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing* 12 (7), 1–24. <https://doi.org/10.3390/rs12071135>.
- Tamés-Noriega, A./Rodríguez-Cuenca, B./Alonso, M. C. (2015). Automatic extraction of buildings and trees using fuzzy K-means classification on high-resolution satellite imagery and LiDAR data. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 593–596.
- Tavakkoli Piralilou, Sepideh/Shahabi, Hejar/Jarihani, Ben/Ghorbanzadeh, Omid/Blaschke, Thomas/Gholamnia, Khalil/Meena, Sansar/Aryal, Jagannath (2019). Landslide Detection Using Multi-Scale Image Segmentation and Different Machine Learning Models in the Higher Himalayas. *Remote Sensing* 11 (21), 2575. <https://doi.org/10.3390/rs11212575>.
- Trimble (2020). Deep Learning (CNN) Algorithms. eCognition Suite - Document Version 10.0.1. Online verfügbar unter [https://docs.ecognition.com/v10.0.1/eCognition_documentation/Reference%20Book/23%20Deep%20Learning%20\(CNN\)%20Algorithms/Deep%20Learning%20\(CNN\)%20Algorithms.htm](https://docs.ecognition.com/v10.0.1/eCognition_documentation/Reference%20Book/23%20Deep%20Learning%20(CNN)%20Algorithms/Deep%20Learning%20(CNN)%20Algorithms.htm) (abgerufen am 21.03.2022).
- Tuia, Devis/Volpi, Michele/Copa, Loris/Kanevski, Mikhail/Munoz-Mari, Jordi (2011). A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Signal Processing* 5 (3), 606–617. <https://doi.org/10.1109/JSTSP.2011.2139193>.

- Vakalopoulou, M./Karantzas, K./Komodakis, N./Paragios, N. (2015). Building detection in very high resolution multispectral data with deep learning features. In: Ieee (Hg.). 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1873–1876.
- Verma, V./Kumar, R./Hsu, S. (2006). 3D Building Detection and Modeling from Aerial LIDAR Data. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) vom 2006.
- Vincent, L./Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (6), 583–598. <https://doi.org/10.1109/34.87344>.
- Walton, Jeffrey/Nowak, David/Greenfield, Eric (2008). Assessing Urban Forest Canopy Cover Using Airborne or Satellite Imagery. *Arboriculture & Urban Forestry* 34 (6), 334–340. <https://doi.org/10.48044/jauf.2008.046>.
- Weidner, Uwe/Förstner, W. (1995). Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50 (4), 38–49. [https://doi.org/10.1016/0924-2716\(95\)98236-S](https://doi.org/10.1016/0924-2716(95)98236-S).
- Whiteside, Timothy G./Boggs, Guy S./Maier, Stefan W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation* 13 (6), 884–893. <https://doi.org/10.1016/j.jag.2011.06.008>.
- Wild-Pfeiffer, Franziska/Wampach, Maryse/Rokus, Marius/Holuba, Karl-Heinz (2019). Kompetenzfeld Satellitendaten – Beitrag der Fernerkundung zur Digitalisierungsstrategie digital@bw des Landes Baden-Württemberg. *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement* (4/2019), 253–259. <https://doi.org/10.12902/zfv-0271-2019>.
- Woodcock, Curtis E./Strahler, Alan H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment* 21 (3), 311–332. [https://doi.org/10.1016/0034-4257\(87\)90015-0](https://doi.org/10.1016/0034-4257(87)90015-0).
- Yamashita, Rikiya/Nishio, Mizuho/Do, Richard Kinh Gian/Togashi, Kaori (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging* 9 (4), 611–629. <https://doi.org/10.1007/s13244-018-0639-9>.
- Yu, Qian/Gong, Peng/Clinton, Nick/Biging, Greg/Kelly, Maggi/Schirokauer, Dave (2006). Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering & Remote Sensing* 72 (7), 799–811. <https://doi.org/10.14358/PERS.72.7.799>.
- Yuan, Jiangye (2016). Automatic Building Extraction in Aerial Scenes Using Convolutional Networks. Online verfügbar unter <http://arxiv.org/pdf/1602.06564v1>.
- Yu-Jin Zhang (2006). An Overview of Image and Video Segmentation in the Last 40 Years. In: *Advances in Image and Video Segmentation*. IGI Global, 1–16.
- Zhang, Caiyun/Xie, Zhixiao (2012). Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery. *Remote Sensing of Environment* 124, 310–320. <https://doi.org/10.1016/j.rse.2012.05.015>.
- Zhang, Guangyun/Zhang, Rongting/Zhou, Guoqing/Jia, Xiuping (2018). Hierarchical spatial features learning with deep CNNs for very high-resolution remote sensing image classification. *International Journal of Remote Sensing* 39 (18), 5978–5996. <https://doi.org/10.1080/01431161.2018.1506593>.

- Zhang, Qinchuan/Wang, Yunhong/Liu, Qingjie/Liu, Xiangyu/Wang, Wei (2016). CNN based suburban building detection using monocular high resolution Google Earth images. In: 2016 IEEE International Geoscience & Remote Sensing Symposium. Proceedings : July 10-15, 2016, Beijing, China, IGARSS 2016 - 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 7/10/2016 - 7/15/2016. Piscataway, NJ, IEEE, 661–664.
- Zhang, Xin/Engel, Bernard/Xiong, Xiaoping/Johannsen, Chris (1995). Analysis of classification results of remotely sensed data and evaluation of classification algorithms.
- Zhang, Xin/Han, Liangxiu/Han, Lianghao/Zhu, Liang (2020). How Well Do Deep Learning-Based Methods for Land Cover Classification and Object Detection Perform on High Resolution Remote Sensing Imagery? *Remote Sensing* 12 (3), 417. <https://doi.org/10.3390/rs12030417>.
- Zhao, Wenzhi/Guo, Zhou/Yue, Jun/Zhang, Xiuyuan/Luo, Liqun (2015). On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing* 36 (13), 3368–3379. <https://doi.org/10.1080/2150704X.2015.1062157>.
- Zhou, Shuang/Mi, Liang/Chen, Hao/Geng, Yishuang (2013). Building detection in Digital surface model. In: 2013 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE.
- Zhu, Xiao Xiang/Tuia, Devis/Mou, Lichao/Xia, Gui-Song/Zhang, Liangpei/Xu, Feng/Fraundorfer, Friedrich (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine* 5 (4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>.
- Zhu, Yuting/Liang, Zili/Yan, Jingwen/Chen, Gao/Wang, Xiaoqing (2021). E-D-Net: Automatic Building Extraction From High-Resolution Aerial Images With Boundary Information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 4595–4606. <https://doi.org/10.1109/jstars.2021.3073994>.

Glossar

Interessensgebiet:

Beim Interessensgebiet der Thesis handelt es sich um den Land- und Stadtkreis Heilbronn. Dabei begründet sich der Begriff aus dem Interesse des Vermessungsamtes des Landkreises Heilbronn an den Untersuchungsergebnissen.

Untersuchungsgebiet:

Als Untersuchungsgebiet bzw. maximales Untersuchungsgebiet wird die Fläche bezeichnet, die sich aus der Überlappung der Bildaufnahmen der Digitalen Orthophotos und der Satellitendaten von *PlanetScope* ergibt. Das Untersuchungsgebiet umfasst knapp 60 % des Interessensgebietes.

Bearbeitungsgebiet:

Als Bearbeitungsgebiet wird das Gebiet bezeichnet, welches für die Untersuchung ausgewählt wurde und auf dessen Fläche das Training des Modells durchgeführt wird. Die Fläche umfasst 30 km².

Trainingsgebiet:

Umfasst die identische Fläche des Bearbeitungsgebietes. Der Begriff wird verwendet, wenn thematisch das Training des Modells behandelt wird.

Kontrollgebiet:

Das Kontrollgebiet befindet sich innerhalb des Bearbeitungsgebietes und umfasst eine Fläche von 6 km² und dient der Kontrolle der Ergebnisse aus dem Trainingsgebiet.

Anhang

Anhang 1: Metadatenauszug der <i>PlanetScope</i>-Daten	CXV
Anhang 2: Auszüge der DOP-Samples	CXVI
Für die Klasse „Gebäudeobjekte“	CXVI
Für die Klasse „Kleine Objekte“	CXVII
Für die Klasse „Nicht-Gebäude“	CXVIII
Anhang 3: Auszüge der SAT-Samples	CXIX
Klasse „Gebäudeobjekte“	CXIX
Klasse „Nicht-Gebäude“	CXIX
Anhang 4: Übersicht des Trainingsgebietes als RGB-Darstellung	CXX
Übersicht Trainingsgebiet für TrueDOP-Daten	CXX
Übersicht Trainingsgebiet für <i>PlanetScope</i> -Daten.....	CXXI
Anhang 5: Übersicht des Kontrollgebietes als RGB-Darstellung	CXXII
Übersicht Kontrollgebiet für TrueDOP-Daten.....	CXXII
Übersicht Kontrollgebiet für <i>PlanetScope</i> -Daten	CXXIII
Anhang 6: 3-Klassen DOP: <i>Heatmap</i> der Wahrscheinlichkeitswerte für Abschnitt 5.1.1	CXXIV
Klasse „Gebäudeobjekte“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern.....	CXXIV
Klasse „Kleine Objekte“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern	CXXIV
Klasse „Nicht-Gebäude“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern.....	CXXV
Klasse „Gebäudeobjekte“ für Abschnitt 5.1.1 - DOP mit (b) angepassten Trainingsparameter	CXXV
Klasse „Kleine Objekte“ für Abschnitt 5.1.1 - DOP mit (b) angepassten Trainingsparametern	CXXVI
Klasse „Nicht-Gebäude“ für Abschnitt 5.1.1 - DOP mit (b) angepassten Trainingsparametern	CXXVI
Anhang 7: 2-Klassen DOP: <i>Heatmap</i> der Wahrscheinlichkeitswerte für Abschnitt 5.1.2 –	CXXVII
Klasse „Gebäudeobjekte“ für Abschnitt 5.1.2 - DOP mit (a) Standardtrainingsparametern.....	CXXVII
Klasse „Nicht-Gebäude“ für Abschnitt 5.1.2 - DOP mit (a) Standardtrainingsparametern.....	CXXVII
Klasse „Nicht-Gebäude“ für Abschnitt 5.1.2 - DOP mit (b) angepassten Parametern.....	CXXVIII
Anhang 8: binäre Klassen SAT: <i>Heatmap</i> der Wahrscheinlichkeitswerte für Abschnitt 5.2	CXXIX
Klasse „Gebäudeobjekte“ für Abschnitt 5.2 - SAT mit (a) Standardtrainingsparametern.....	CXXIX
Klasse „Nicht-Gebäude“ für Abschnitt 5.2 - SAT mit (a) Standardtrainingsparameter	CXXIX
Klasse „Gebäudeobjekte“ für Abschnitt 5.2 - SAT mit (b) angepassten Parametern.....	CXXX
Klasse „Nicht-Gebäude“ für Abschnitt 5.2 SAT mit (b) angepassten Parametern	CXXX

Anhang 9: Getestete Parametereinstellungen für das CNN-Training	CXXXI
TrueDOP-Daten.....	CXXXI
Satellitendaten	CXXXI
Anhang 10: Darstellungen der Konfusionsmatrizen aus eCognition für die beschriebenen Ergebnisse	CXXXII
DOP 3-Klassen laufzeitoptimiert.....	CXXXII
DOP 3-Klassen qualitätsoptimiert.....	CXXXII
DOP 2-Klassen laufzeitoptimiert (ohne „Kleine Objekte“).....	CXXXII
DOP 2-Klassen qualitätsoptimiert (ohne „Kleine Objekte“)	CXXXIII
DOP 2-Klassen qualitätsoptimierte & wissensbasierte Bildanalyse	CXXXIII
SAT 2-Klassen laufzeitoptimiert.....	CXXXIII
SAT 2-Klassen qualitätsoptimiert	CXXXIV
SAT 2-Klassen laufzeitoptimierte & wissensbasierte Bildanalyse	CXXXIV
Anhang 11: Übersicht des Kontrollgebietes für die 3-Klassen für das Kontrollgebiet	CXXXV
Laufzeitoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten	CXXXV
Qualitätsoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten	CXXXVI
Anhang 12: Übersicht der binären DOP-Klassifizierung für das Kontrollgebiet CXXXVII	
Laufzeitoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten	CXXXVII
Qualitätsoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten	CXXXVIII
Qualitätsoptimiertes Ergebnis & wissensbasiertes OBIA in Kombination mit RGB für DOP-Daten	CXXXIX
Anhang 13: Übersicht der SAT-Klassifizierung für 2-Klassen für das Kontrollgebiet	CXL
Laufzeitoptimiertes Ergebnis in Kombination mit RGB für <i>PlanetScope</i> -Daten.....	CXL
Qualitätsoptimiertes Ergebnis in Kombination mit RGB für <i>PlanetScope</i> -Daten.....	CXLI
Laufzeitoptimiertes Ergebnis & wissensbasiertes OBIA in Kombination mit RGB für <i>PlanetScope</i> -Daten.....	CXLII

Anhang 1: Metadatenauszug der PlanetScope-Daten

```
{
  "id": "20210601_102105_64_2424",
  "type": "Feature",
  "geometry": {
    "coordinates": [
      [
        [
          [9.052337581491342, 49.39136541658423],
          [8.981616993057726, 49.20003601141481],
          [9.469123896344593, 49.12187248832804],
          [9.541712309091086, 49.31241494680947],
          [9.052337581491342, 49.39136541658423]
        ]
      ]
    ]
  },
  "type": "Polygon",
  "properties": {
    "acquired": "2021-06-01T10:21:05.645602Z",
    "anomalous_pixels": 0,
    "clear_confidence_percent": 100,
    "clear_percent": 100,
    "cloud_cover": 0,
    "cloud_percent": 0,
    "ground_control": true,
    "gsd": 4.1,
    "heavy_haze_percent": 0,
    "instrument": "PSB.SD",
    "item_type": "PSScene",
    "light_haze_percent": 0,
    "pixel_resolution": 3,
    "provider": "planetscope",
    "published": "2021-06-02T04:43:06Z",
    "publishing_stage": "finalized",
    "quality_category": "standard",
    "satellite_azimuth": 109.8,
    "satellite_id": "2424",
    "shadow_percent": 0,
    "snow_ice_percent": 0,
    "strip_id": "4539693",
    "sun_azimuth": 151.3,
    "sun_elevation": 60.5,
    "updated": "2021-06-03T04:53:55Z",
    "view_angle": 1,
    "visible_confidence_percent": 100,
    "visible_percent": 100
  }
}
```


Anhang 2: Auszüge der DOP-Samples

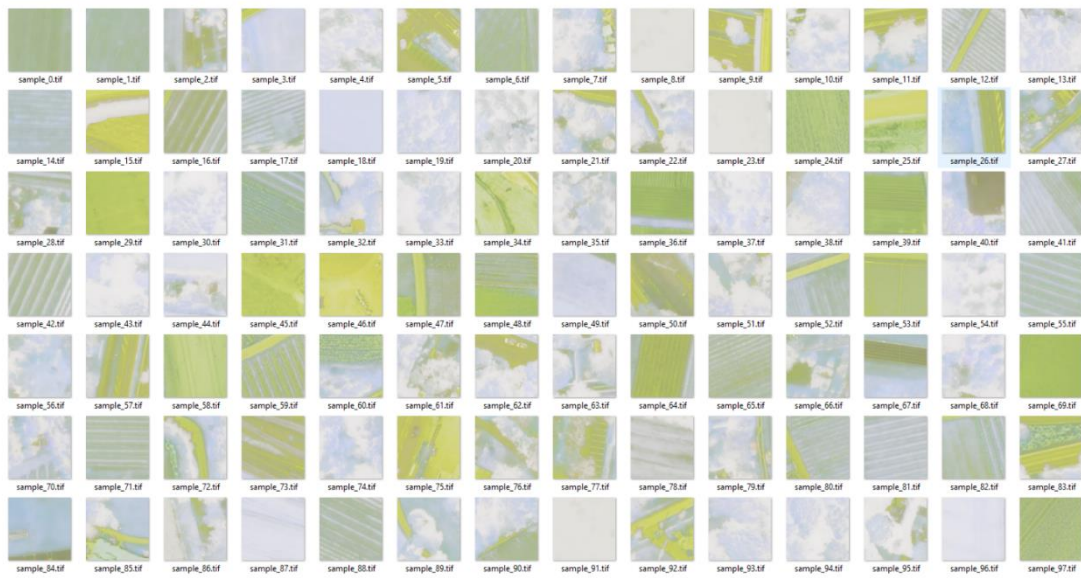
Für die Klasse „Gebäudeobjekte“



Für die Klasse „Kleine Objekte“



Für die Klasse „Nicht-Gebäude“

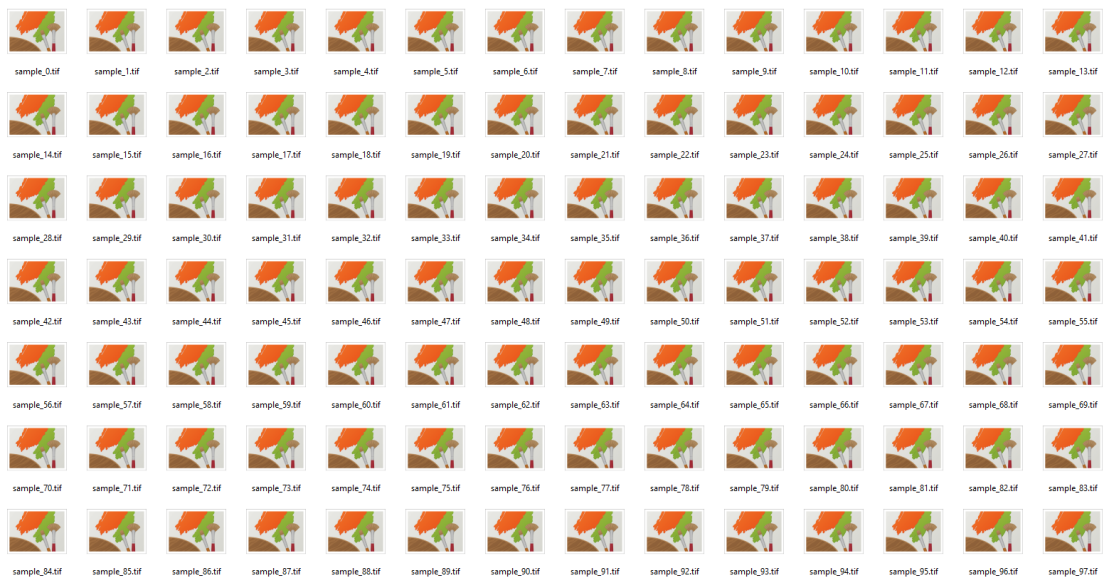


Anhang 3: Auszüge der SAT-Samples

Klasse „Gebäudeobjekte“



Klasse „Nicht-Gebäude“



(Anmerkung: Durch die 8-Kanäle der SAT-Aufnahmen kann Windows kein Vorschau bild darstellen, sondern stellt lediglich einen Platzhalter dar)

Anhang 4: Übersicht des Trainingsgebietes als RGB-Darstellung

Übersicht Trainingsgebiet für TrueDOP-Daten



Übersicht Trainingsgebiet für *PlanetScope*-Daten



Anhang 5: Übersicht des Kontrollgebietes als RGB-Darstellung

Übersicht Kontrollgebiet für TrueDOP-Daten



Übersicht Kontrollgebiet für *PlanetScope*-Daten



Anhang 6: 3-Klassen DOP: *Heatmap* der Wahrscheinlichkeitswerte für Abschnitt 5.1.1**Klasse „Gebäudeobjekte“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern****Klasse „Kleine Objekte“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern**

Klasse „Nicht-Gebäude“ für Abschnitt 5.1.1 - DOP mit (a) Standardtrainingsparametern



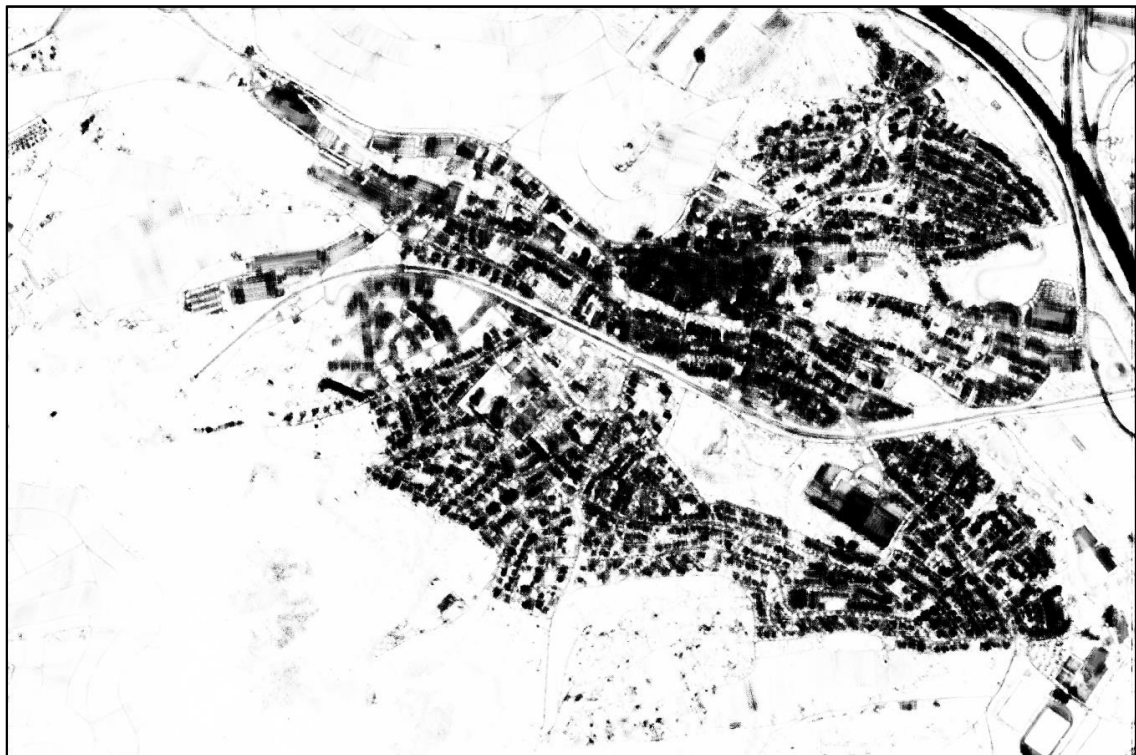
Klasse „Gebäudeobjekte“ für Abschnitt 5.1.1- DOP mit (b) angepassten Trainingsparameter



Klasse „Kleine Objekte“ für Abschnitt 5.1.1 - DOP mit (b) angepassten Trainingsparametern



Klasse „Nicht-Gebäude“ für Abschnitt 5.1.1 - DOP mit (b) angepassten Trainingsparametern



Anhang 7: 2-Klassen DOP: *Heatmap* der Wahrscheinlichkeitswerte für Abschnitt 5.1.2 –

Klasse „Gebäudeobjekte“ für Abschnitt 5.1.2 - DOP mit (a) Standardtrainingsparametern



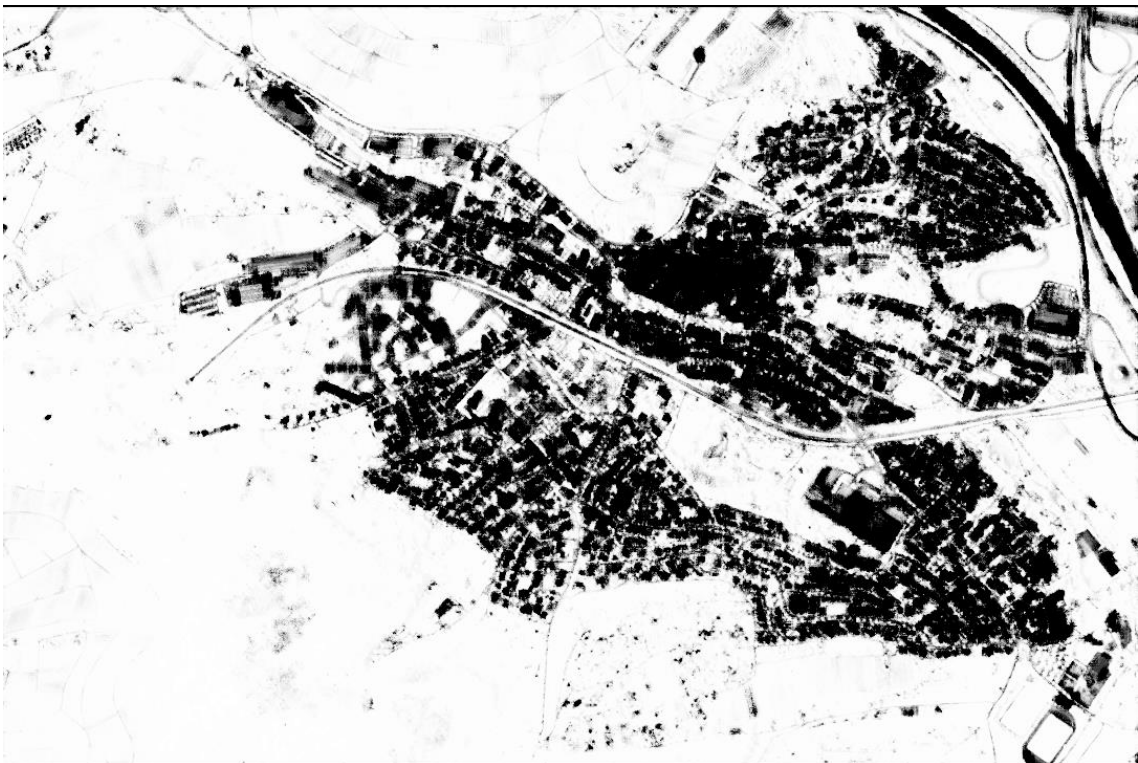
Klasse „Nicht-Gebäude“ für Abschnitt 5.1.2 - DOP mit (a) Standardtrainingsparametern

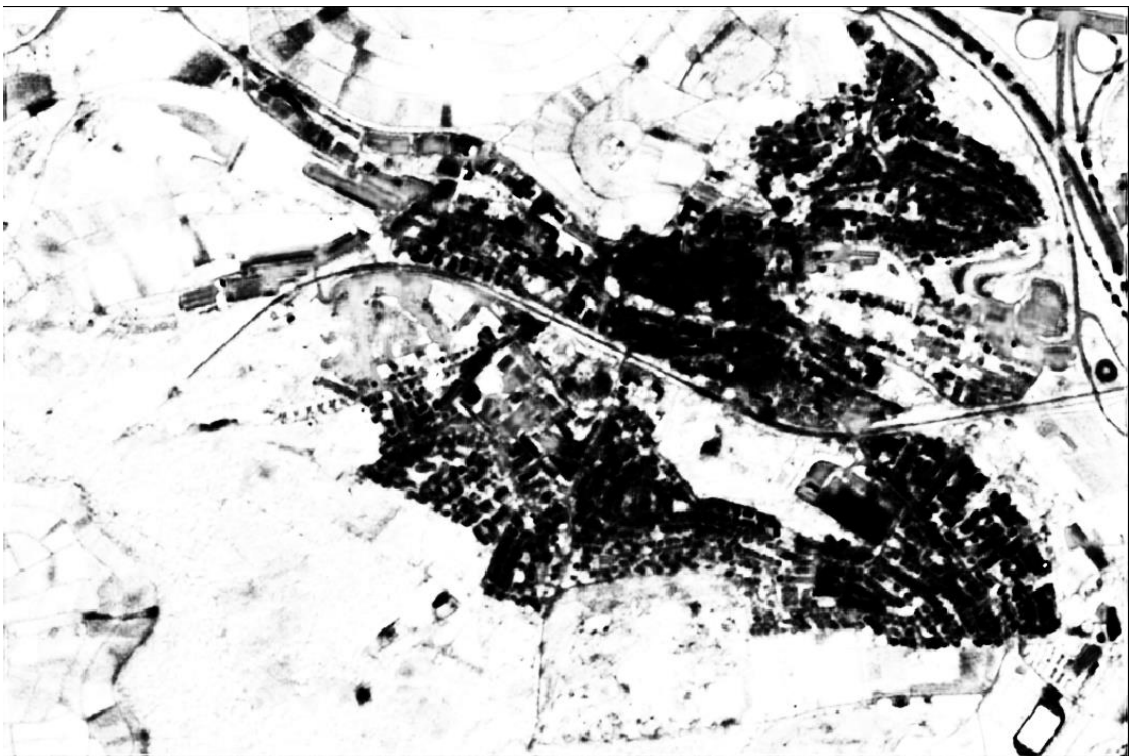


Klasse „Gebäudeobjekte“ für Abschnitt 5.1.2 - DOP mit (b) angepassten Parametern

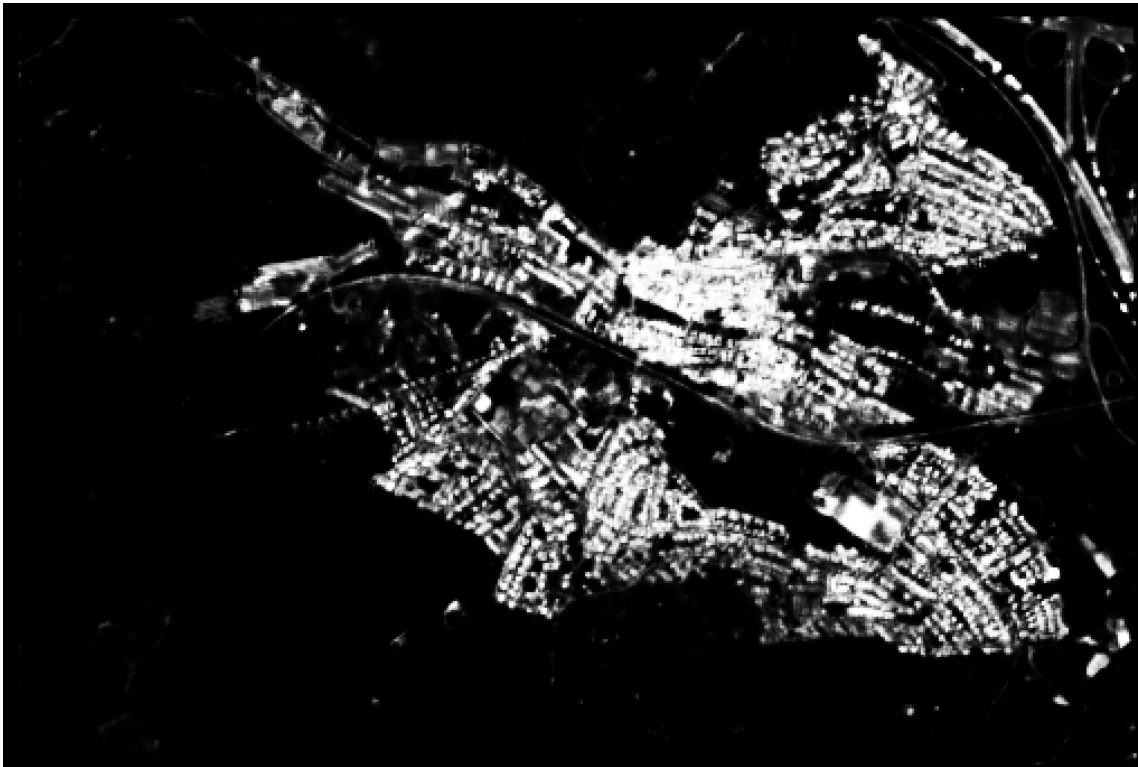


Klasse „Nicht-Gebäude“ für Abschnitt 5.1.2 - DOP mit (b) angepassten Parametern



Anhang 8: binäre Klassen SAT: *Heatmap* der Wahrscheinlichkeitswerte für Abschnitt 5.2**Klasse „Gebäudeobjekte“ für Abschnitt 5.2 - SAT mit (a) Standardtrainingsparametern****Klasse „Nicht-Gebäude“ für Abschnitt 5.2 - SAT mit (a) Standardtrainingsparameter**

Klasse „Gebäudeobjekte“ für Abschnitt 5.2 - SAT mit (b) angepassten Parametern



Klasse „Nicht-Gebäude“ für Abschnitt 5.2 SAT mit (b) angepassten Parametern



Anhang 9: Getestete Parametereinstellungen für das CNN-Training

TrueDOP-Daten

Architektur	Batch size	Trainingsschritte	Trainingsrate	Epochen	
2 Hidden Layer, 7x7 Kernel, 12 feature maps, Max-Pooling, 5x5 Kernel, 12 feature maps,	50	5.000	0,0006 (Standard)	1	
	100			3	
	200			6	
	500			15	
	1.000			31	
	50	7.500		1	
	100			4	
	200			8	
	500			23	
	750			35	
	3 Hidden Layer 7x7, 12fm, max 4x4, 12fm, max 2x2, 12fm , max	50		5.000	1
		100			3
		200			6
		500			15
		1.000		10.000	62
3 Hidden Layer 7x7, 12fm, max 5x5, 12fm, max 3x3, 12fm	50	5.000	1		
	100		3		
	200		6		
				6	

Satellitendaten

Architektur	Batch Size	Trainingsschritte	Trainingsrate	Epochen
2 Hidden Layer, 3x3, 12 fm, max 2x2, 12 fm	500	10.000	0,006	12,5
2 Hidden Layer, 1x1, 2fm, max 3x3, 2fm,	500	5000		6
		1000		12,5
2 Hidden Layer, 1x1, 2fm, max 3x3, 3fm,	320	7.000		5
	500	25.000		31
1 Hidden Layer, 2x2 Kernel, 10 feature maps	50	6000		3

Anhang 10: Darstellungen der Konfusionsmatrizen aus eCognition für die beschriebenen Ergebnisse

DOP 3-Klassen laufzeitoptimiert

User Class \ Sam...	not-building	small_object	building_object	Sum
Confusion Matrix				
not-building	396123	2307	19355	417785
small_object	16042	6264	99323	121629
building_object	12609	5198	156730	174537
Sum	424774	13769	275408	
Accuracy				
Producer	0.9326	0.4549	0.5691	
User	0.9482	0.0515	0.898	
Hellden	0.9403	0.0925	0.6967	
Short	0.8873	0.0485	0.5345	
Kappa Per Class	0.8374	0.343	0.4297	
Totals				
Overall Accuracy	0.7831			
Kappa	0.6087			

DOP 3-Klassen qualitätsoptimiert

User Class \ Sample	not-building	small_object	building_object	Sum
Confusion Matrix				
not-building	410124	1907	28831	440862
small_object	0	0	0	0
building_object	14650	11862	246577	273089
Sum	424774	13769	275408	
Accuracy				
Producer	0.9655	0	0.8953	
User	0.9303	-nan(ind)	0.9029	
Hellden	0.9476	0	0.8991	
Short	0.9004	0	0.8167	
Kappa Per Class	0.9098	0	0.8305	
Totals				
Overall Accuracy	0.9198			
Kappa	0.8347			

DOP 2-Klassen laufzeitoptimiert (ohne „Kleine Objekte“)

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	403344	38519	441863
building_object	21430	250658	272088
Sum	424774	289177	
Accuracy			
Producer	0.9495	0.8668	
User	0.9128	0.9212	
Hellden	0.9308	0.8932	
Short	0.8706	0.807	
Kappa Per Class	0.8676	0.7848	
Totals			
Overall Accuracy	0.916		
Kappa	0.8241		

DOP 2-Klassen qualitatsoptimiert (ohne „Kleine Objekte“)

Confusion Matrix - Samples

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	410124	30738	440862
building_object	14650	258439	273089
Sum	424774	289177	
Accuracy			
Producer	0.9655	0.8937	
User	0.9303	0.9464	
Hellden	0.9476	0.9193	
Short	0.9004	0.8506	
Kappa Per Class	0.9098	0.8279	
Totals			
Overall Accuracy	0.9364		
Kappa	0.8669		

reduce expand Close

DOP 2-Klassen qualitatsoptimierte & wissensbasierte Bildanalyse

Confusion Matrix - Samples

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	421294	35385	456679
building_object	3480	253792	257272
Sum	424774	289177	
Accuracy			
Producer	0.9918	0.8776	
User	0.9225	0.9865	
Hellden	0.9559	0.9289	
Short	0.9155	0.8672	
Kappa Per Class	0.9773	0.8087	
Totals			
Overall Accuracy	0.9456		
Kappa	0.885		

reduce expand Close

SAT 2-Klassen laufzeitoptimiert

Confusion Matrix - Samples

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	18539	3081	21620
building_object	2665	25287	27952
Sum	21204	28368	
Accuracy			
Producer	0.8743	0.8914	
User	0.8575	0.9047	
Hellden	0.8658	0.898	
Short	0.7634	0.8148	
Kappa Per Class	0.7771	0.751	
Totals			
Overall Accuracy	0.8841		
Kappa	0.7638		

reduce expand Close

SAT 2-Klassen qualitätsoptimiert

Confusion Matrix - Samples

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	20860	10709	31569
building_object	344	17659	18003
Sum	21204	28368	

Accuracy

Producer	0.9838	0.6225
User	0.6608	0.9809
Hellden	0.7906	0.7616
Short	0.6537	0.615
Kappa Per Class	0.9553	0.4072

Totals

Overall Accuracy	0.777
Kappa	0.571

reduce expand Close

SAT 2-Klassen laufzeitoptimierte & wissensbasierte Bildanalyse

Confusion Matrix - Samples

User Class \ Sam...	not-building	building_object	Sum
Confusion Matrix			
not-building	19265	5817	25082
building_object	1939	22551	24490
Sum	21204	28368	

Accuracy

Producer	0.9086	0.7949
User	0.7681	0.9208
Hellden	0.8324	0.8533
Short	0.713	0.7441
Kappa Per Class	0.8149	0.5947

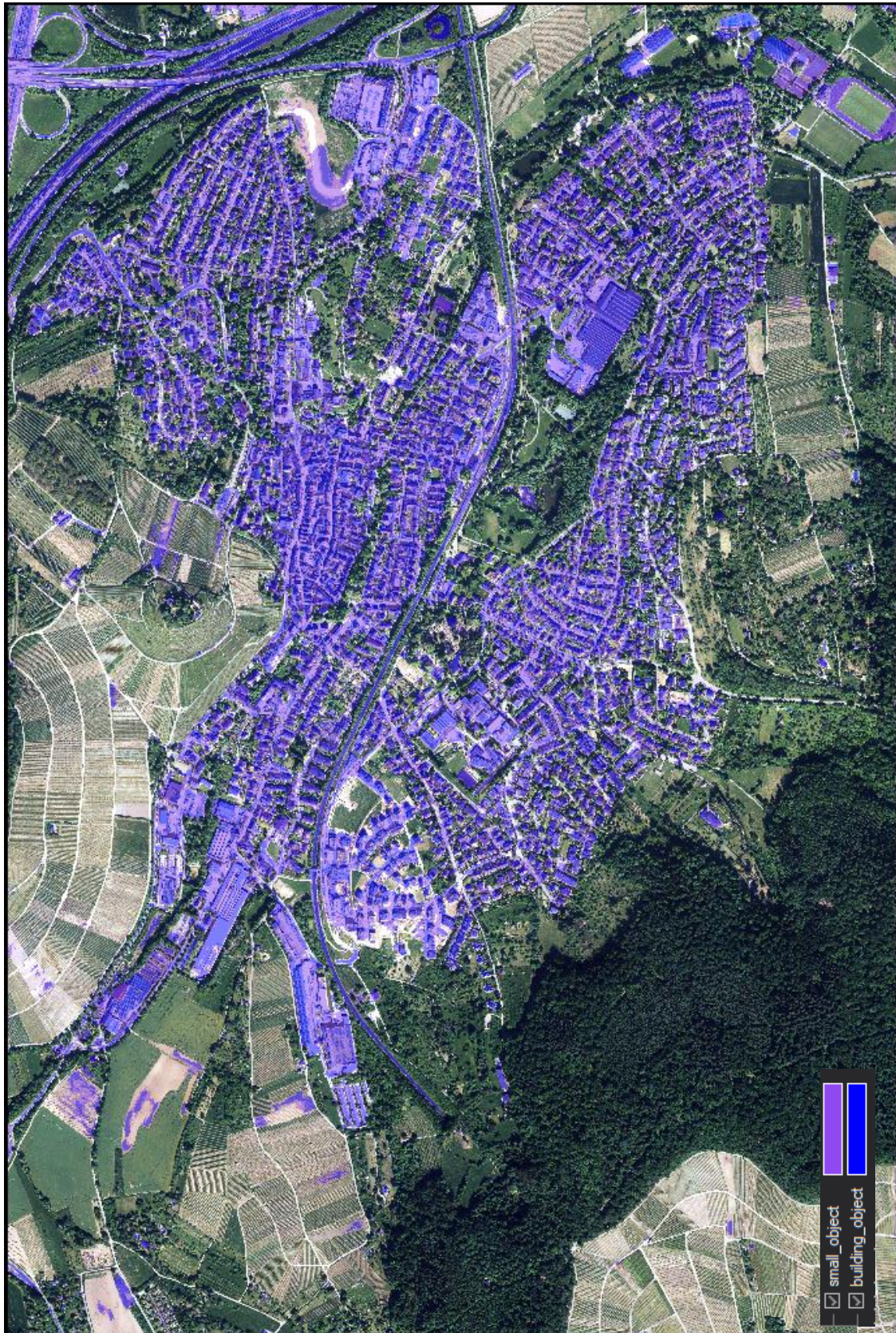
Totals

Overall Accuracy	0.8435
Kappa	0.6876

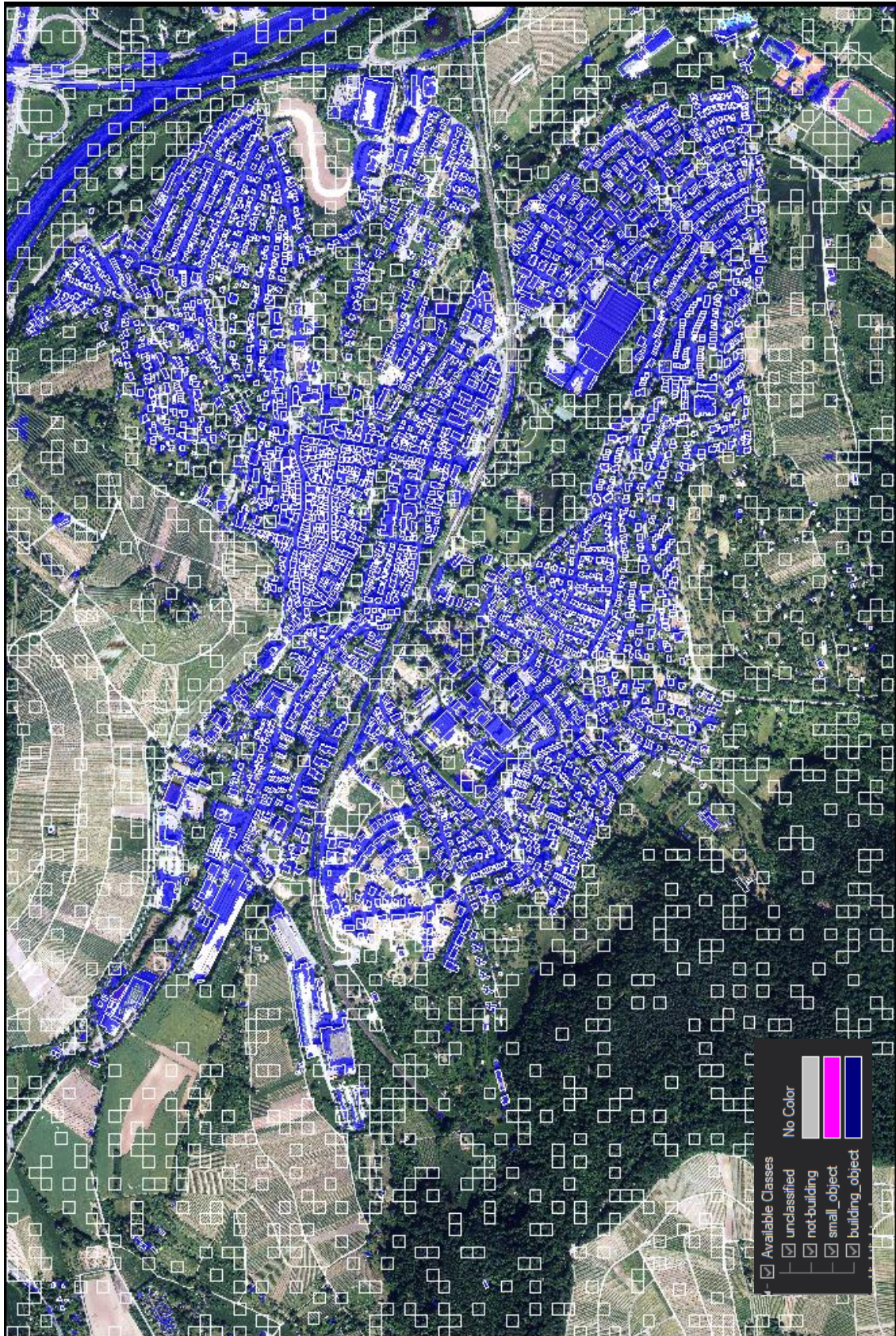
reduce expand Close

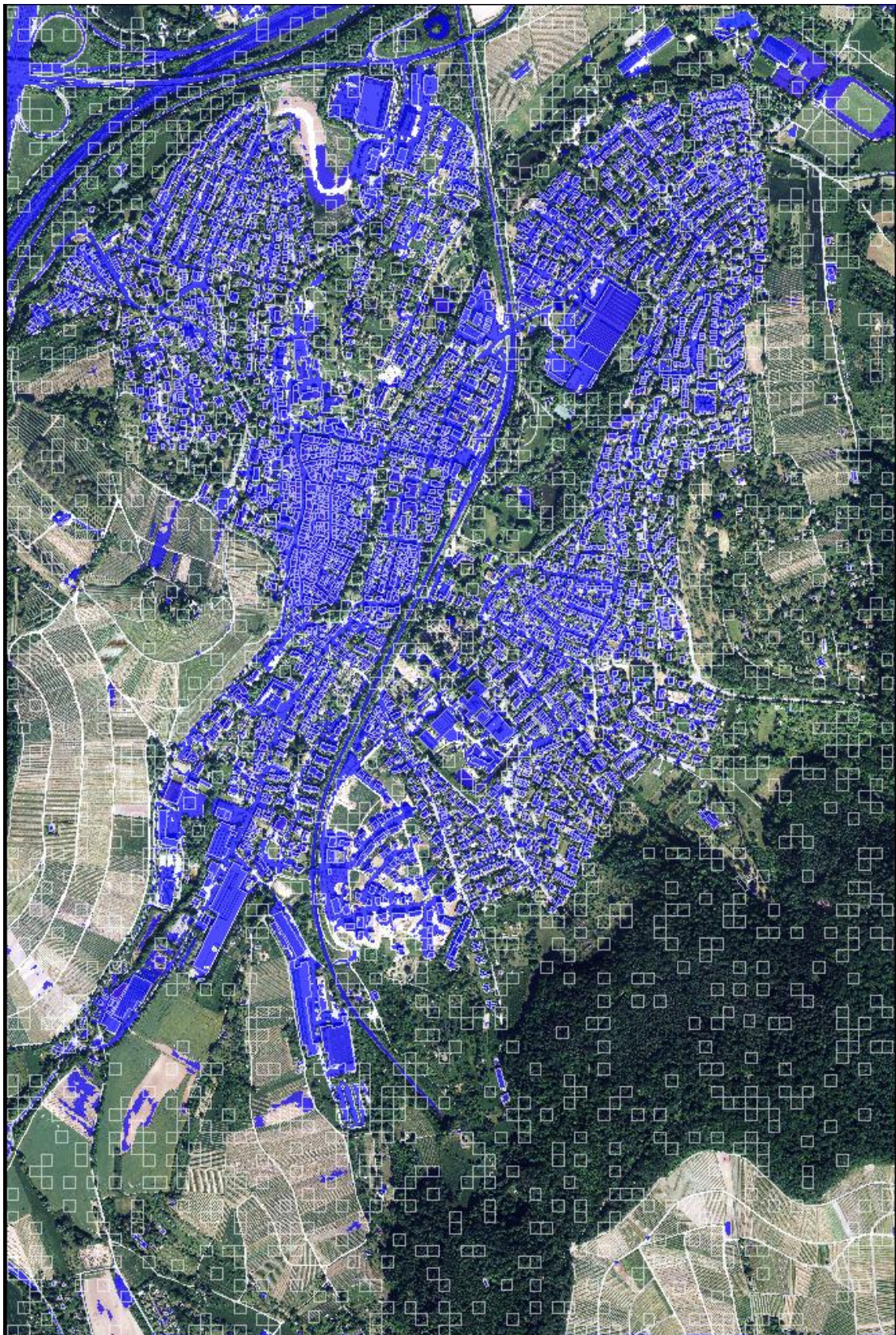
Anhang 11: Übersicht des Kontrollgebietes für die 3-Klassen für das Kontrollgebiet

Laufzeitoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten

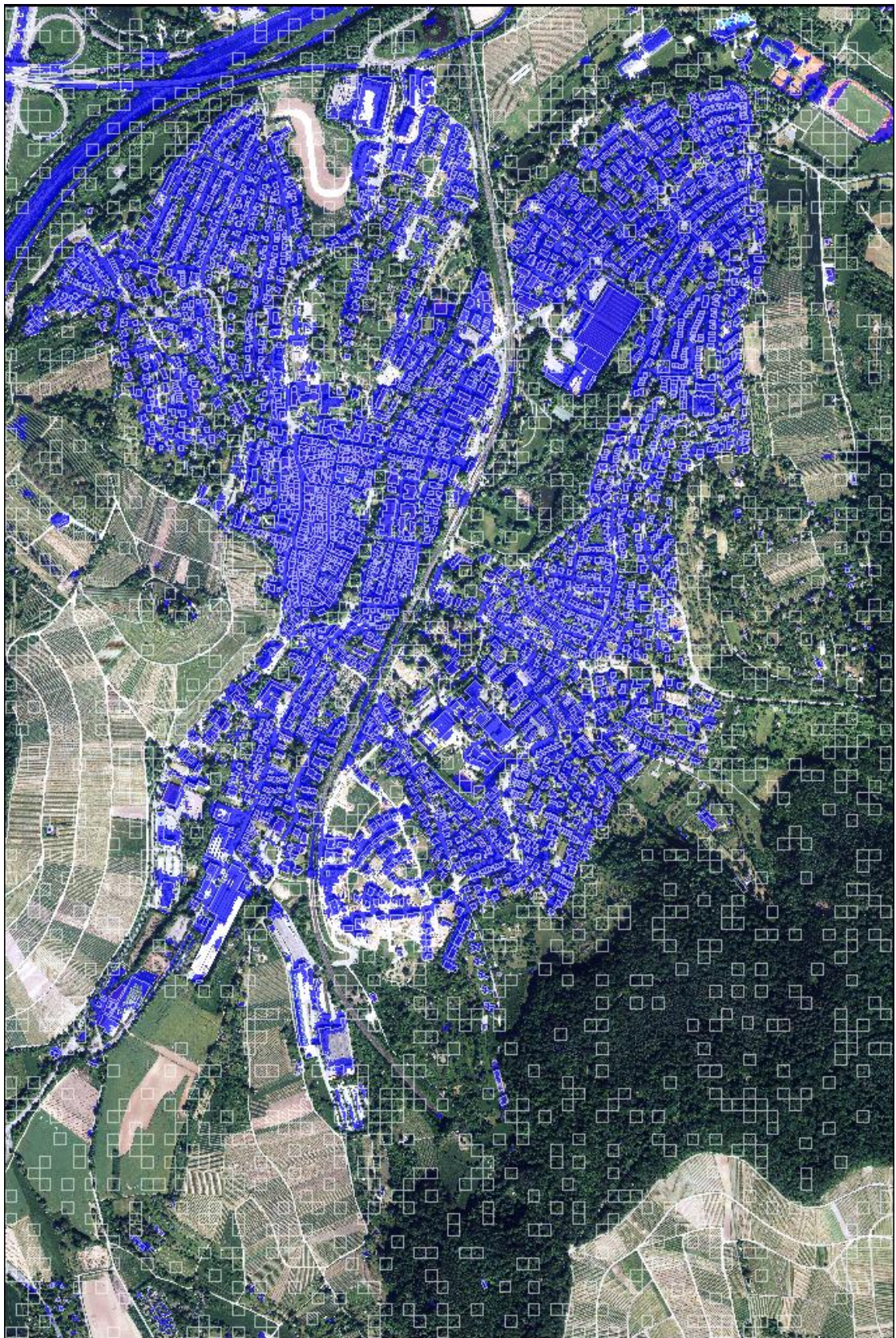


Qualitätsoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten

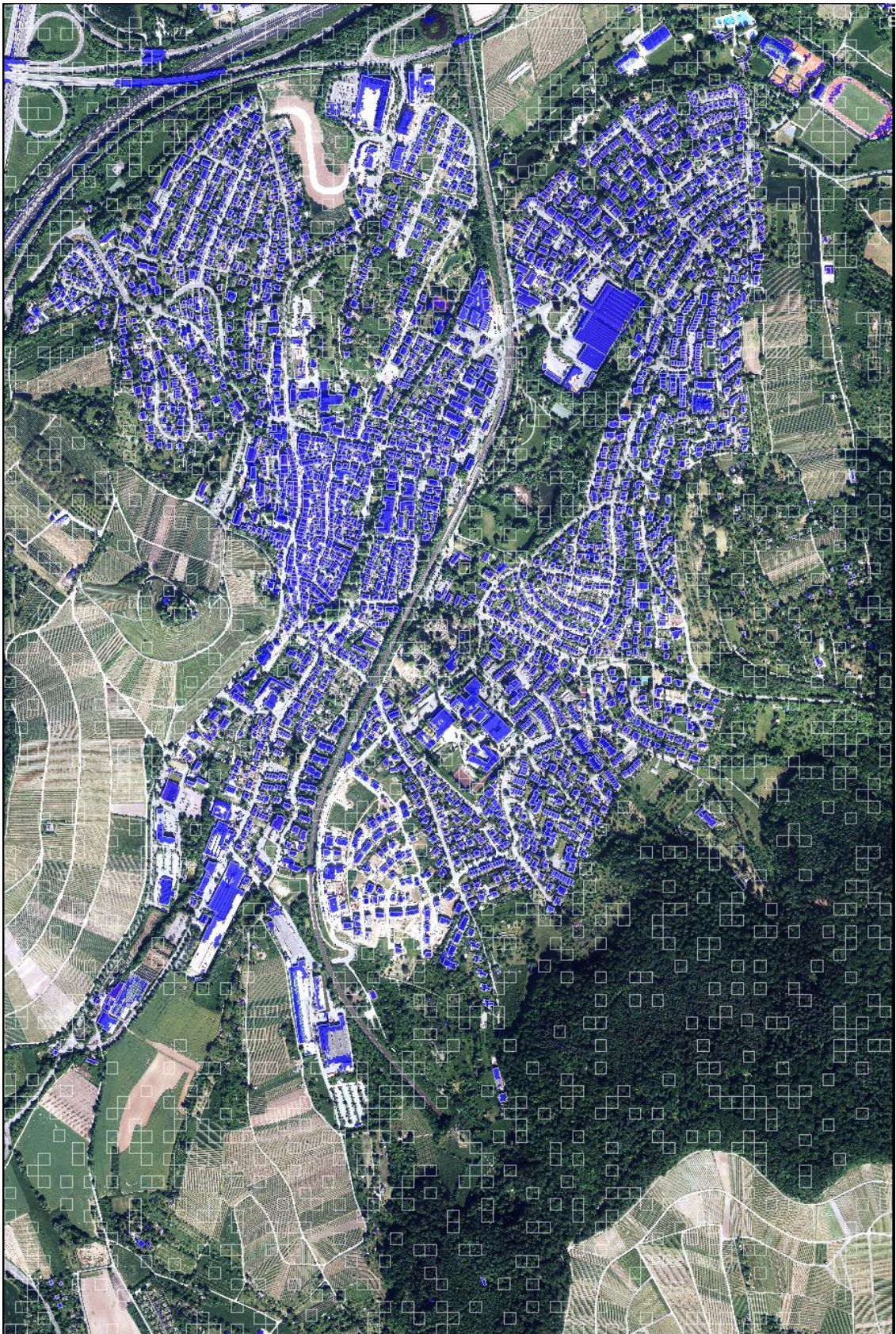


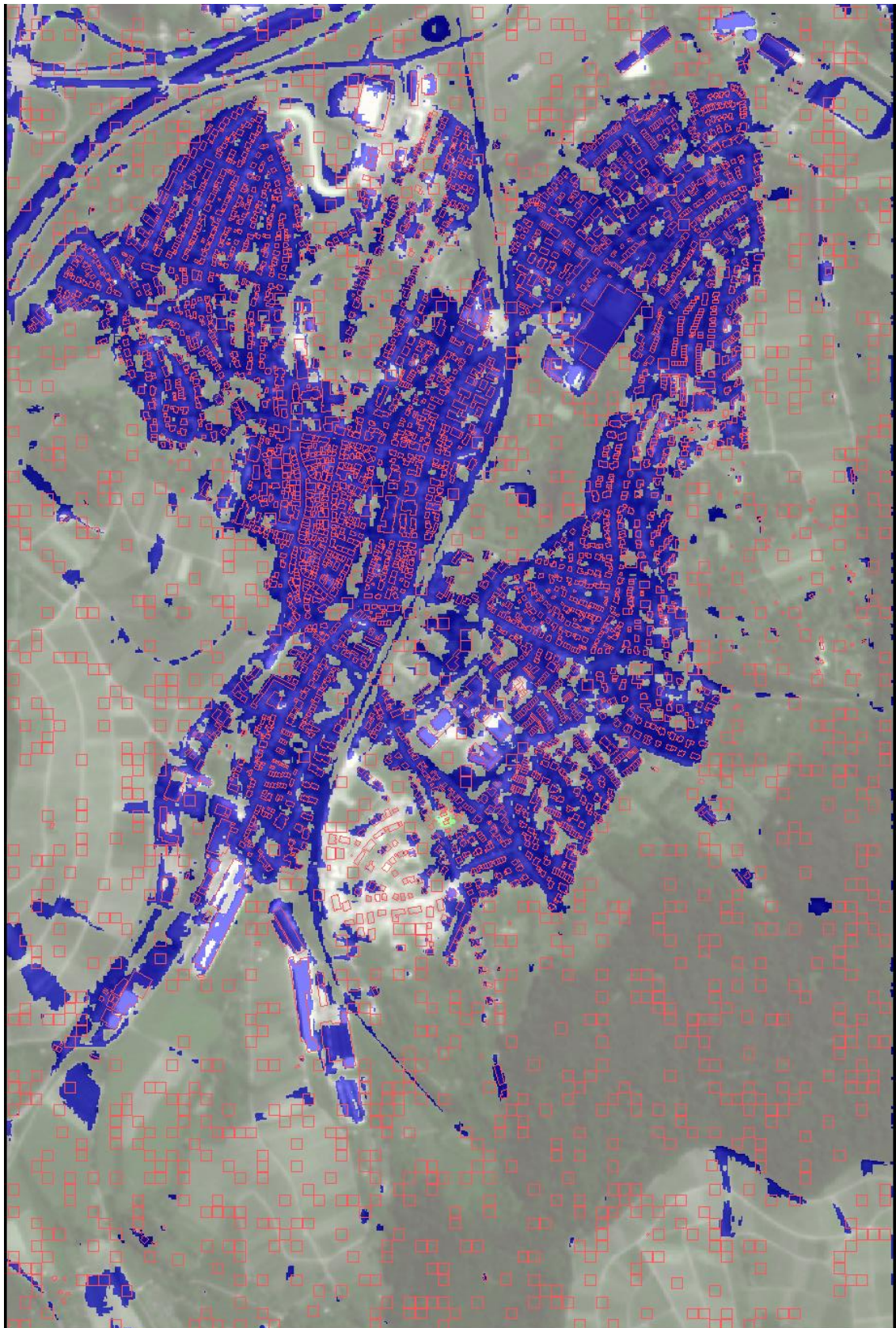
Anhang 12: Übersicht der binären DOP-Klassifizierung für das Kontrollgebiet**Laufzeitoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten**

Qualitätsoptimiertes Ergebnis in Kombination mit RGB für DOP-Daten



Qualitätsoptimiertes Ergebnis & wissensbasiertes OBIA in Kombination mit RGB für DOP-Daten



Anhang 13: Übersicht der SAT-Klassifizierung für 2-Klassen für das Kontrollgebiet**Laufzeitoptimiertes Ergebnis in Kombination mit RGB für *PlanetScope*-Daten**

Qualitätsoptimiertes Ergebnis in Kombination mit RGB für *PlanetScope*-Daten



Laufzeitoptimiertes Ergebnis & wissensbasiertes OBIA in Kombination mit RGB für PlanetScope-Daten

