

# KI-basierte Detektion von Gebäuden mittels Deep Learning und amtlichen Geodaten zur Baufallerkundung

Robert Roschlaub, Qingyu Li, Stefan Auer, Karin Möst, Clemens Glock, Michael Schmitt, Yilei Shi und Xiao Xiang Zhu

## Zusammenfassung

Es wird ein neuartiger Ansatz vorgestellt, der auf einer Erkennung von Gebäuden und Gebäudeveränderungen aus hochauflösten Luftbildern anhand von Künstlicher Intelligenz (KI) beruht. Die zum Trainieren des KI-Systems notwendigen Datenbestände sind die Gebäudeumrisse aus der amtlichen Digitalen Flurkarte (DFK) und das lagerichtige Digitale Orthophoto (TrueDOP). Die semantische Detektion der Gebäude und Gebäudeveränderungen erfolgt über optische Aufnahmen und Oberflächenmodelle, wie dem normalisierten Oberflächenmodell (nDOM) und einem aus der Differenz von zwei Zeitepochen abgeleiteten Oberflächenmodell (tDOM).

Am Beispiel der Baufallerkundung werden die Ergebnisse einer aktuellen Forschungskooperation zwischen der Bayerischen Vermessungsverwaltung (BVV) und der Technischen Universität München (TUM) aufgezeigt und bestehenden Verfahrenslösungen gegenübergestellt. Die vorgestellte KI-basierte Verfahrenslösung ist grundsätzlich auf alle Vermessungsverwaltungen der Länder bundesweit übertragbar.

## Summary

*A novel approach is introduced which is based on the detection of buildings and building changes from high-resolution aerial images using artificial intelligence (AI). The data sets necessary for training the AI system are the building outlines from the official Digital Cadastre Map (DFK) and Digital Orthophotos without building lean (TrueDOP). The semantic detection of the buildings and building changes is carried out based on optical images and digital surface models, such as the normalized digital surface model (nDOM) and a Temporal Digital Surface Model (tDOM) derived from the difference of two time epochs.*

*The results of the current cooperation between the Bavarian Agency for Digitisation, High-Speed Internet and Surveying (BVV) and the Technical University of Munich (TUM) on the detection of buildings and building changes are presented and compared with results derived from existing approaches. The presented AI-based solution is basically transferable to all surveying administrations of the federal states nationwide.*

**Schlüsselwörter:** Gebäudedetektion, Semantische Segmentation, Deep Learning, DOM, Datenfusionierung

## 1 Einleitung

Mit steigender Digitalisierung sind technische Verwaltungen zunehmend gefordert, moderne Innovationen aufzugreifen. Ein gutes Beispiel hierfür ist der Einsatz von Künstlicher Intelligenz (KI), die im Rahmen einer Forschungskooperation der Bayerischen Vermessungsverwaltung mit der TUM erstmalig auf die Baufallerkundung angewendet wurde.

Die an den Ämtern für Digitalisierung, Breitband und Vermessung (ÄDBV) durchgeführte Baufallerkundung dient zur Aktualisierung der Gebäude im amtlichen Liegenschaftskataster. Gebäude werden von den ÄDBV mit einer Katastergenauigkeit im Bereich von wenigen Zentimetern eingemessen und als Gebäudegrundrisse in der amtlichen Digitalen Flurkarte (DFK) dargestellt.

Zur Fortführung des Liegenschaftskatasters gehört in Bayern die Erfassung von Altbau- und Neubaufällen. Altbaufälle sind Gebäude, die teilweise schon vor Jahren errichtet, aber nie katastertechnisch erfasst wurden und für die in der Regel keine Gebühren für die Gebäudeeinemessung mehr erhoben werden. Neubaufälle sind Gebäude, die seit Kurzem erst errichtet, aber katastertechnisch von den ÄDBV noch nicht gebührenpflichtig eingemessen sind. Daher fehlt sowohl für die Altbau- als auch für die Neubaufälle der Nachweis der Gebäudegrundrisse in der DFK.

Zur Baufallerkundung stehen in Bayern am Landesamt für Digitalisierung, Breitband und Vermessung (LDBV) vielfältige qualitätsgesicherte Geobasisdaten zur Verfügung. Diese Daten werden auf gesetzlicher Grundlage aufgaben- und maßstabsbezogen erfasst. Hierzu zählen z.B. das aus dem Airborne Laserscanning abgeleitete und bayernweit in einer Gitterweite von 1 m verfügbare Digitale Geländemodell (DGM) sowie die als Bayernbelfliegung bekannte und seit dem Jahr 2017 in einem Zweijahreszyklus durchgeführte Digitale Luftbildbelfliegung. Sie wird in einer Bodenauflösung von 20 cm sowie einer Längsüberdeckung von 80 % und einer Querüberdeckung von 50 % beflogen. Aus den digitalen Luftbildern wird mittels Dense Image Matching (Haala 2011) ein bildbasiertes Digitales Oberflächenmodell (bDOM) in Form eines regelmäßigen Gitters mit Höhen und Rot-Grün-Blau (RGB)-Farben abgeleitet (Hirschmüller 2011), das seit dem Jahr 2016 für Bayern flächendeckend zur Verfügung steht. Die Farbe wird für jeden dreidimensionalen Gitterpunkt des bDOM mittels Kollinearitätsgleichung

aus dem Luftbild entnommen. Durch eine Parallelprojektion des bDOM in die Ebene erhält man das TrueDOP, d.h. durch das Weglassen der Höheninformation im bDOM. Das TrueDOP ist ein Digitales Orthophoto, das frei ist von Umklappeffekten bei hohen Gebäuden und Objekten.

Zur Identifizierung von Altbaufällen eignet sich in besonderer Weise das normalisierte Digitale Oberflächenmodell (nDOM), das aus der Differenz zwischen dem bDOM und dem aus Airborne Laserscanning abgeleiteten DGM berechnet wird. Höheninformationen von Fahrzeugen und Containern, Aufschüttungen und Abgrabungen sowie der Vegetation können über Mindesthöhen und Mindestdimensionen aus dem nDOM herausgefiltert werden (Geßler et al. 2019). Vereinzelte Vegetationsreste bleiben jedoch im nDOM erhalten (vgl. Abb. 1). Das nDOM enthält die Höheninformationen sämtlich bestehender Gebäude, unabhängig davon, ob sie bereits im Liegenschaftskataster dokumentiert sind – also auch sämtliche Alt- und Neubaufälle. Zudem können Neubaufälle, die von den ÄDBV noch nicht eingemessen wurden und damit nicht in der DFK dokumentiert sind, mit Hilfe eines aus zwei bDOM-Epochen der Bayernbefliegung abgeleiteten Differenzmodells (tDOM) identifiziert werden. Ebenso lassen sich die im Liegenschaftskataster undokumentierten Abrisse von Gebäuden im tDOM über eine negative Höhenänderung identifizieren.

In jüngster Zeit werden in der Fernerkundung Deep Learning Methoden (DL) bevorzugt (Zhu et al. 2017), insbesondere auch für die Aufgabe der Klassifikation von Fernerkundungsdaten (Shi et al. 2018). Dies ist auf ihre Überlegenheit in der Verallgemeinerung und Erkennungsgenauigkeit zurückzuführen, ohne dass manuell oder interaktiv eingegriffen werden muss. DL-Methoden erfordern Trainingsdaten, die in einem bestimmten Gebiet den Charakteristiken der zu klassifizierenden Objekte entsprechen. Das trainierte DL-System kann anschließend zur Detektion von Objekten auf andere Gebiete übertragen werden.

Im Folgenden wird ein Verfahren zur Baufallerkundung durch eine Datenfusionierung der verschiedenen Datenbestände und den Einsatz Künstlicher Intelligenz erläutert. In diesem Beitrag werden vier Schwerpunkte gesetzt:

1. Um für DL-Methoden sinnvolle Stichprobenstrategien für großräumige Gebäudeerkennungsaufgaben anbieten zu können, wird die Frage der Übertragbarkeit untersucht, indem amtliche Referenzdaten ausgewählter Landkreise verwendet werden.
2. In diesem Zusammenhang wird ein Verfahrensablauf zur Detektion von Alt- und Neubaufällen vorgeschlagen, der die modernsten DL-Methoden integriert und die amtlichen Geodaten wie Gebäudegrundrisse aus der DFK, das nDOM und TrueDOP nutzt. Es werden DL-Methoden implementiert, um durch einen Vergleich mit der aktuellen DFK undokumentierte Gebäude zu identifizieren.



Abb. 1: Überlagerung der DFK mit dem blauen nDOM zur Identifizierung von Altbaufällen (oben) und dem roten tDOM zur Identifizierung von Neubaufällen (unten)

3. Unter Verwendung eines einzigen optimierten DL-Netzes, das mit Daten aus 14 Vermessungsamtsbezirken trainiert wurde – dies entspricht einem Viertel des Gebiets von Bayern –, wird eine Karte mit markierten Neu- und Altbaufällen generiert. Die mittels DL-Methoden erzielten Ergebnisse zeigen eine robuste Identifizierung von Alt- und Neubaufällen im Zielmaßstab der DFK.
4. Die Validierung der erzielten Gebäudeerkennungsrate wird in einem weiteren, vom Trainingsdatensatz unabhängigen Vermessungsamtsbezirk sowohl für die DL-Methode als auch für zwei am LDBV bestehende Verfahrenslösungen anhand der in der DFK bekannten Anzahl an Gebäuden bestimmt und einem traditionellen statistischen Vergleichswert (dem sogenannten »F1 score«) gegenübergestellt.

Die zwei derzeit am LDBV angewandten Strategien zur Detektion von Neubaufällen sind das Filterverfahren und das Vergleichsklassenverfahren. Beide Verfahren basieren auf heuristischen Methoden (Geßler et al. 2019, Roschlaub et al. 2020). Beim *Filterverfahren* werden verschiedene Filter auf das tDOM angewandt, um Fehldetections von nicht vorhandenen Baufällen zu minimieren, die aufgrund des Höhenrauschens des tDOM verursacht werden können oder durch eine Höhenänderung, die nicht ursächlich aus der Bebauungssituation entstanden ist. Anschließend erfolgt eine Prozessierung des tDOM, die sich auf die Höhenänderungen, RGB-Farbinformationen, den Vegetationsindex und geometrische Filter stützt (Geßler et al. 2019). Beim *Vergleichsklassenverfahren* werden aus den Gebäudegrundrissen der DFK im TrueDOP die RGB-Farbwerte aller Pixel der entsprechenden Gebäudedächer

als Referenz gesammelt (Roschlaub et al. 2020). Dann werden die Farbwerte und deren Häufigkeit im RGB-Farbwürfel gezählt, um Gebäude von der Vegetation durch einen empirisch gewählten Schwellenwert zu trennen. Mit Hilfe des nDOM werden zusätzlich Fehlklassifikationen zwischen Gebäuden und anderen Objekten wie Straßen durch einen empirisch ermittelten Höhenschwellenwert vermieden. Da die heuristische Definition von Schwellenwerten jedoch nicht standardisiert ist, müssen diese für verschiedene Flugläufe individuell bestimmt werden, sodass größere Gebiete nicht einheitlich und standardisiert prozessiert werden können. Zudem verfälschen Bäume, die über die Dächer ragen und somit im Gebäudegrundriss liegen, die Referenz für Dächer im RGB-Farbwürfel, sodass im Rahmen der Gebäudeerkennung einige Bäume als Gebäude klassifiziert werden.

Der vorgeschlagene Ansatz über die Verwendung von Neuronalen Netzen zielt darauf ab, eine hohe Erkennungsleistung zu erreichen, bei einer gleichzeitig geringen Anzahl von falschen Hinweisen. Eine hohe Erkennungsrate führt zu einer zuverlässigen und verbesserten Identifikation von fehlenden Gebäuden in der DFK. Eine

geringe Anzahl an Fehlalarmen führt zu einer spürbar einfacheren Verwertung des Ergebnisses am LDBV und an den Vermessungsämtern.

Im nächsten Abschnitt werden zunächst die DL-Netzwerkarchitektur und der vorgeschlagene Entscheidungsbaum für die Erkennung von Altbaufällen erläutert und der Versuchsaufbau beschrieben. In Abschnitt 3 erfolgen die Validierung der KI-Ergebnisse und ein Vergleich mit bestehenden Verfahrenslösungen. Daran schließt sich am Ende die Diskussion der Ergebnisse an.

## 2 Verfahrenslösung

Die Detektion von Gebäuden stellt eine Aufgabe dar, bei der für jedes aus mehreren Bildpixeln bestehende Segment zwei Klassen zu unterscheiden sind: »Gebäude« und »kein Gebäude«. Ansätze über faltungsbasierte Neuronale Netze (Convolutional Neural Networks) stellen für Aufgaben dieser Art den Stand der Technik dar und liefern Ergebnisse mit höherer Genauigkeit und besserer Effizienz

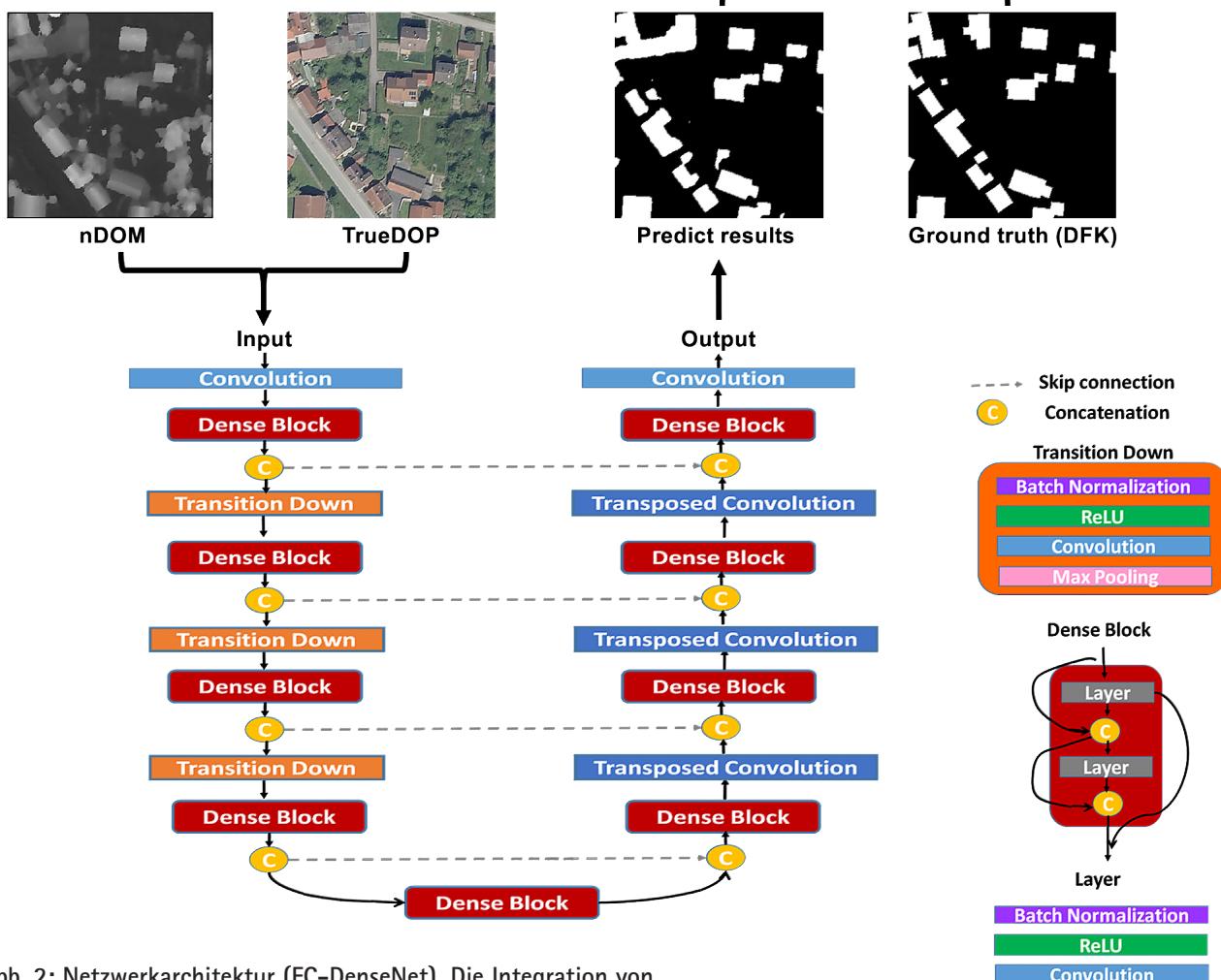


Abb. 2: Netzwerkarchitektur (FC-DenseNet). Die Integration von Dense Blocks dient dem verbesserten Informationsfluss im Netzwerk.

als bislang verwendete Verfahren wie »Random Forest« oder »Support Vector Machines«.

Die Netzwerkarchitektur FC-DenseNet (Jégou et al. 2017) hat sich im Vergleich zu alternativen Netzwerkarchitekturen als sehr geeignet für Detektionsaufgaben für städtische Szenen erwiesen. Hierbei sind vor allem eine höhere Erkennungsleistung (Li et al. 2018) und eine verbesserte Extraktion von Merkmalen (Shi et al. 2020) zu nennen.

## 2.1 Ansatz für Gebäudedetektion

Der Ansatz für die Gebäudedetektion beruht auf dem Prinzip von »Convolutional Neural Networks (CNN)«, d.h. Neuronalen Netzen mit einer Abfolge von Filteroperationen und nichtlinearen Projektionen. Eingangsdaten werden in aufeinanderfolgenden Schichten gefiltert (Linearkombination mit Filtermatrizen), wodurch repräsentative Merkmale von gesuchten Datenkomponenten verstärkt werden. Die Einträge in den Filtermatrizen werden dafür mit Trainingsdaten gelernt, um die Erkennungsleistung des Netzwerkes zu optimieren. Das Ergebnis der Filteroperationen geht darüber hinaus in nichtlineare Aktivierungsfunktionen ein. Dadurch verbessert sich die Voraussetzung für eine erfolgreiche Trennung von Datenkomponenten in einem neuen Merkmalsraum.

Im vorliegenden Fall besteht die Aufgabe des Netzwerks in der Erkennung von Gebäuden in hochaufgelösten Bilddaten und Oberflächenmodellen. Die Erken-

nungsmerkmale beruhen hierbei auf den optischen Bildkanälen (Rot, Grün, Blau) und geometrischen Oberflächenmodellen.

Das gewählte neuronale Netz FC-DenseNet (siehe Abb. 2) folgt den Grundprinzipien von »Convolutional Neural Networks« und weist dabei die charakteristischen Merkmale eines sogenannten U-Netzes auf:

- Faltung mit Filtermatrizen zur Hervorhebung von relevanten Merkmalen,
- Reduzierung der Abtastung zur Vergrößerung des Auswertebereichs auf der linken Seite (mit Max-Pooling),
- Erhöhung der Abtastung für den Rückerhalt der originalen Bildauflösung auf der rechten Seite,
- Verwendung einer Aktivierungsfunktion für die Nachbildung von Nichtlinearitäten,
- Sprungverbindungen (»Skip Connections«) für den Erhalt von hoher räumlicher Auflösung,
- Abgleich von Ergebnissen mit Referenzdaten anhand einer Verlustfunktion und
- iterative Anpassung der Einträge in den Filtermatrizen im Rahmen des Lernprozesses.

Darüber hinaus beinhaltet das Netzwerk sogenannte »Dense Blocks«, in denen Bildausschnitte an Faltungsoperationen vorbeigeschleust werden, wodurch der Informationsfluss und Lernprozess im Netzwerk verbessert wirken kann. Der Gebäudedetektor bildet das Kernstück des Entscheidungsbaums zur Unterscheidung von undokumentierten Gebäuden (Altbau- und Neubaufälle; siehe Abb. 3).

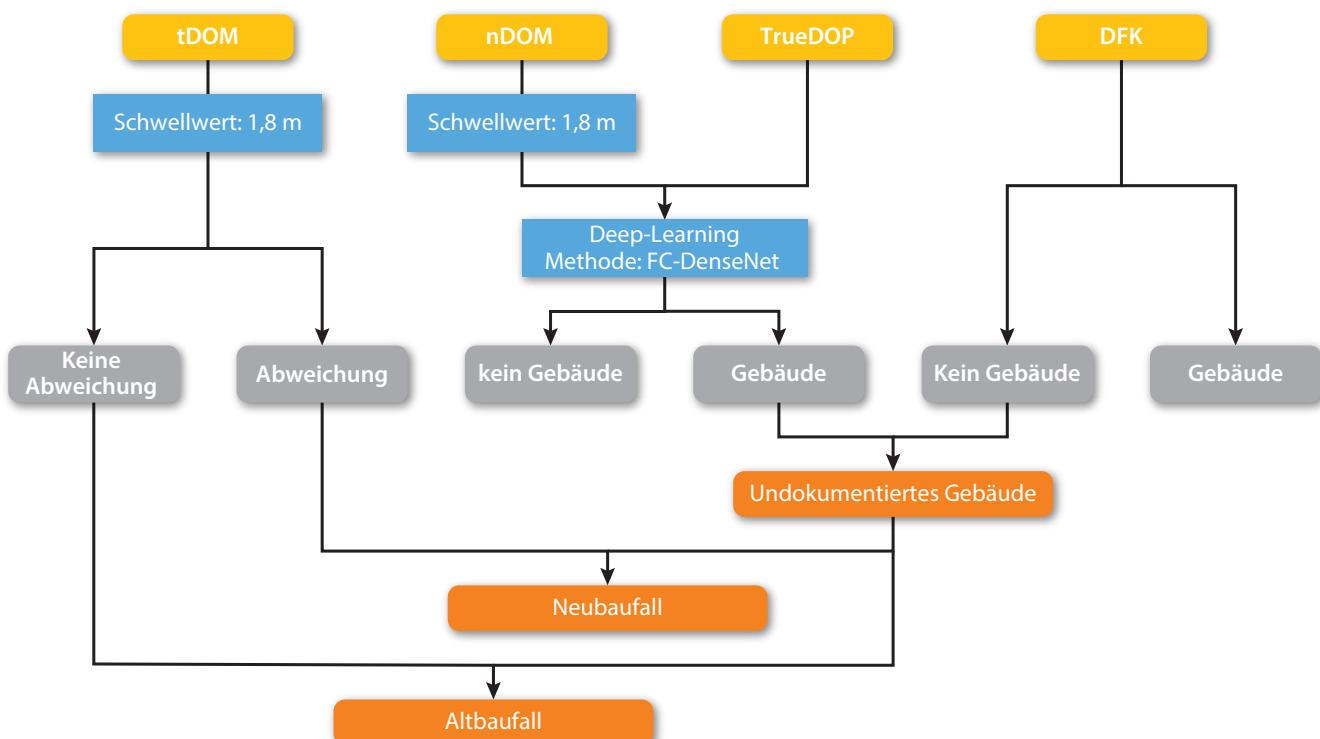


Abb. 3: KI-Entscheidungsbaum zur Identifizierung von Alt- und Neubaufällen (Anm.: Der Entscheidungsbaum enthält derzeit noch keine Identifizierung von Gebäudeaufstockungen. Hierzu müssten die im tDOM bestehenden Höhenänderungen größer des gesetzten Schwellenwertes von 1,8 m mit dem Gebäude der DFK zusätzlich verbunden werden.)

## 2.2 Entscheidungsbaum für die Erkennung undokumentierter Gebäude

Die Erkennung von Gebäuden anhand des Neuronalen Netzes erfolgt mit Hilfe von optischen Bildern und dem nDOM (siehe Abb. 2). Für die Erkennung von undokumentierten Gebäuden wird das Detektionsergebnis in einen Entscheidungsbaum (siehe Abb. 3) eingebunden, der eine weitere Kategorisierung ermöglicht. Zunächst werden die erkannten Gebäudeflächen mit der DFK-Information (Rasterformat) überlagert für die Identifikation von allen undokumentierten Gebäuden. Die Gebäudeflächen werden im zweiten Schritt mit Informationen des tDOM verschnitten, in dem neu gebaute Gebäude ein zeitliches Signal hinterlassen haben. Zur Trennung der undokumentierten Gebäude in Alt- und Neubaufälle wird ein konstanter Schwellenwert (Höhenveränderung von 1,8 m) für das tDOM in einer räumlichen Auflösung von 0,4 m verwendet. Somit spaltet sich die Menge aller undokumentierten Gebäude in alte undokumentierte Gebäude (Signal in nDOM, kein Signal in tDOM) und neue undokumentierte Gebäude (Signal in nDOM und tDOM) auf.

Die Gebäudesegmente der beiden Kategorien werden am Ende in Polygone umgewandelt und als Shape-Datei gespeichert.

## 2.3 Datenaufbereitung und experimentelle Umsetzung

Die in dieser Studie verwendeten Datensätze bestehen aus TrueDOP, nDOM, tDOM und der DFK. Zunächst werden für alle 15 Vermessungsbezirke die TrueDOP und nDOM Kacheln ausgewählt, die innerhalb der Bezirksgrenzen liegen. Dabei werden das tDOM, nDOM und das TrueDOP als Rasterdaten in eine Größe von  $2500 \times 2500$  Pixel gekachelt, mit einer räumlichen Auflösung von 0,4 m, während die DFK als Vektordatei bereitsteht. Die DFK wird innerhalb des Geobereichs der Bildkacheln beschnitten. Dann wird die Vektorinformation der DFK in ein Rasterformat umgewandelt. Das TrueDOP, das nDOM und die DFK als Bodenreferenz werden in Ausschnitte mit der Größe von  $256 \times 256$  Pixel geschnitten, mit einer Überlappung von 124 Pixel zum nächsten benachbarten Ausschnitt. Da nicht alle Gebäude in der DFK dokumentiert sind, gibt es Inkonsistenzen zwischen TrueDOP, nDOM und der DFK, wenn die DFK als Bodenreferenz für das Trainieren des Neuronalen Netzes verwendet wird. Die Auswirkungen sind jedoch unbedeutend angesichts der Größe des Trainingsdatensatzes.

Das FC-DenseNet ist in der Softwareumgebung Pytorch implementiert und arbeitet mit einem NVIDIA Tesla P100-Grafikprozessor mit 16 GB Arbeitsspeicher. Das Training des Netzwerks durch den fortlaufenden Abgleich mit Referenzdaten beruht auf einem Stochastic Gradient Descent Verfahren für die Optimierung der Lösung. Die Verlustfunktion beruht auf dem Maß der Kreuzentropie (Maß für die Passung des Modells an die Referenzinformation), die Anzahl der Referenzbildpaare pro Optimierungsschritt beträgt 5. Das FC-DenseNet arbeitet mit 12 Blöcken (Dense Blocks) und 5 zusätzlichen Schichten, in denen Bildausschnitte gefaltet werden.

Die Aufgabenstellung ist auf eine spätere operationelle Verwendung des Verfahrens ausgerichtet. In diesem Zusammenhang ist es wertvoll, die Zusammenstellung und Auswirkung der Trainingsdaten zu betrachten. Dazu werden zwei separate Netzwerke mit verschiedenen Datensätzen trainiert, um insbesondere Aspekte der Übertragbarkeit prüfen zu können. Im ersten Fall stammen die Trainingsdaten aus dem Bezirk Ansbach, im zweiten Fall wurden Trainingsdaten aus 14 Vermessungsbezirken verwendet. Die Anzahl der Trainings- und Validierungsausschnitte ist in Tab. 1 zusammengefasst, deren Verhältnis sich auf ca. 80 % zu 20 % beläuft. Die beiden trainierten Modelle wurden für die Gebäudedetektion in zwei Vermessungsbezirken (Ansbach, Bad Tölz-Wolfratshausen) eingesetzt und anschließend validiert (siehe Ergebnisse im Abschnitt 3).

Tab. 1: Anzahl der Bildausschnitte für Training und Validierung

Trainiertes Modell	Vermessungsbezirk	Anzahl von Bildausschnitten für Training	Anzahl von Bildausschnitten für Validierung
1	Ansbach	67.965	18.077
2	Ansbach	67.965	18.077
	Bad Tölz-Wolfratshausen	14.982	3.671
	Kulmbach	24.998	5.679
	Kulmbach-Kronach	19.987	5.112
	Landau	34.964	8.733
	Landau-Deggendorf	38.454	9.763
	Landshut	60.957	15.090
	München	88.364	22.213
	Regensburg	47.947	11.941
	Regensburg-Hemau	9.481	2.243
	Rosenheim	59.141	14.789
	Rosenheim-Wasserburg	14.150	3.567
	Schweinfurt	54.951	13.759
	Weilheim	76.959	19.202

### 3 Validierung der Ergebnisse

#### 3.1 Validierung auf Pixelebene

Die Bewertung der Ergebnisse der beiden trainierten Netzwerke für die Gebäudedetektion aus dem DL-Schritt erfolgt anhand der DFK für das Testgebiet Ansbach und Bad Tölz-Wolfratshausen als Referenzinformation. Diese wird in ein Rasterformat umgewandelt, mit dem Ergebnis überlagert und pixelbasiert verglichen. Im Idealfall müsste der KI-basierte Gebäudedetektor sämtliche Gebäudeflächen der DFK erkennen.

Zwei traditionellen Qualitätsmaße sind hierbei von besonderem Interesse: der »F1 score« und die Maßeinheit »Intersection-over-Union (IOU)«, welche die räumliche Überlagerung zwischen dem Detektionsergebnis und der Referenz bemisst.

Der »F1 score« berechnet sich durch

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

wobei die Maße »Precision« die Richtigkeit und »Recall« die Vollständigkeit des Erkennungsergebnisses erfassen.

In den Gleichungen bezeichnet *TP* die Anzahl der richtig erkannten Gebäudepixel, *FP* die fälschlicherweise als Gebäude detektierten Pixel und *FN* die nicht erkannten Gebäudepixel.

Das Maß *IOU* beschreibt das Verhältnis zwischen dem Überlappungsbereich (»Area of Overlap«) und Vereinigungsbereich (»Area of Union«) zwischen dem Detektionsergebnis und der Referenz. Keine Überlappung führt demzufolge zum Wert 0, eine vollständige Übereinstimmung zum Wert 1.

In den Ergebnissen (siehe Tab. 2 und Tab. 3) zeigt sich, dass der kleinere Trainingsdatensatz aus Ansbach ausreichend repräsentativ für die Gebäudeerkennung für

Tab. 2: Erkennungsgenauigkeit der trainierten Gebäude-detektoren für das Testgebiet Ansbach

Trainiertes Modell	F1 score	IOU
1	90,5 %	82,7 %
2	90,3 %	82,3 %

Tab. 3: Erkennungsgenauigkeit der trainierten Gebäude-detektoren für das Testgebiet Bad Tölz-Wolfratshausen

Trainiertes Modell	F1 score	IOU
1	73,9 %	58,7 %
2	86,5 %	76,2 %

ein Testgebiet in Ansbach ist. Für das unabhängige Testgebiet Bad Tölz-Wolfratshausen ergibt die Verwendung des größeren Trainingsdatensatzes aus den 14 Vermessungsbezirken eine Verbesserung der Gebäudedetektion (d.h. ein Unterschied von 12,6 % beim »F1 score« und von 17,5 % beim IOU-Wert, siehe Tab. 3). Die Trainingsdaten des umfassenderen Modells bestehend aus den 14 Vermessungsamtbezirken bilden die Variabilität des Erscheinungsbilds von Gebäuden im Vermessungsbezirk Bad Tölz-Wolfratshausen deutlich besser ab.

#### 3.2 Validierung für Gebäudeinstanzen

Die Gebäudedetektion mittels KI liefert unscharfe Gebäudepolygone auf Rasterebene, ohne einen kataster-technischen Bezug zu bestehenden Grundstücksgrenzen zu haben, wie es typischerweise bei Reihenhäusern der Fall ist. Im TrueDOP werden Reihenhäuser über die gesamte Dachfläche als ein Gebäude erkannt. Entsprechend können Reihenhäuser von bildbasierten Detektionsverfahren nur als ein eigenständiges Gebäude erkannt werden, unabhängig von ihrer rechtlichen Trennung durch amtliche Grundstücksgrenzen. Zur Ermittlung, ob ein Gebäudegrundriss in der DFK erkannt wurde, werden die KI-Ergebnisse des Gebäudedetektors mit jedem einzelnen Gebäudegrundriss der DFK verschnitten – bei Reihenhäusern also mit jedem einzelnen Reihenhaus der gesamten Reihenhauszeile. Zur qualitativen Bewertung der Gebäudeerkennung im aktuellen Gebäudebestand wird für jedes einzelne Gebäude der DFK der prozentuale Flächenanteil aus der Überdeckung der KI-basierten Gebäudedetektion zur amtlichen Gebäudegrundrissfläche ermittelt. Dabei sind verschiedene prozentuale Abstufungen vorgenommen worden. Ein Gebäude wird als erkannt eingestuft, wenn mehr als 20 % des Gebäudegrundrisses in der DFK von Pixeln des mittels KI ermittelten Gebäudepolygons überdeckt sind (vgl. Abb. 4). Kleinere Gebäude sind von Detektionsverfahren nur schwer zu erkennen, daher wird eine Überdeckungsrate von mehr als 20 % für ein erkanntes Gebäude als ausreichend angesehen. Je höher der Prozentsatz ist, desto vollständiger ist die Gebäudeerkennung. Altbau- und Neubaufälle sind von diesem Verfahren ausgenommen, da für diese Gebäude keine Gebäudegrundrisse in der DFK vorliegen.

##### 3.2.1 Vergleich zwischen den Landkreisen Ansbach und Bad Tölz

Mit Beschränkung auf die Trainingsdaten für  $\frac{1}{4}$  Bayerns ergeben sich im Testgebiet von Ansbach, das mit einem Flächenanteil von 20 % nicht im Trainingsdatensatz enthalten war, ein »F1 score« in Höhe von 90,3 % für die pixelbasierte Validierung und eine Erkennungsrate von 88,7 % über die Verifikation mittels der Gebäudeinstanzen in der DFK. Die Differenz der beiden Werte beträgt nur 1,6 %.

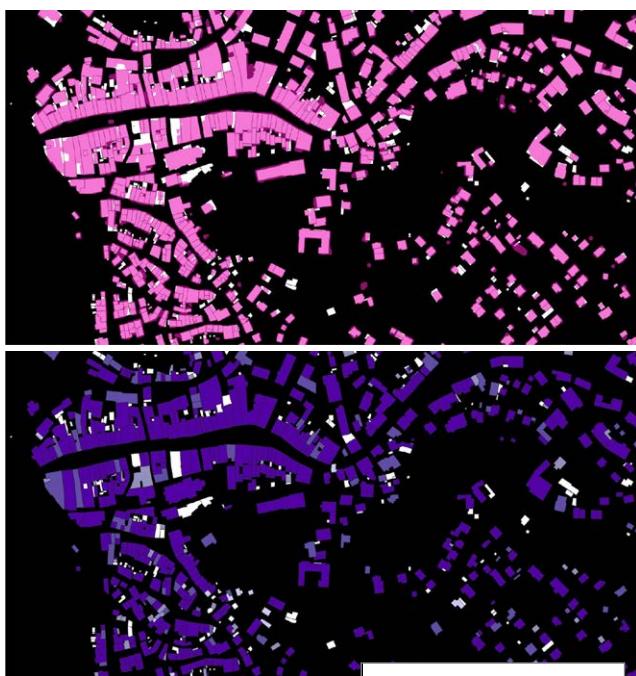


Abb. 4:  
Auf der DFK überlagertes Ergebnis der KI-basierten Gebäudedetektion (oben) und gebäudescharfe Klassifizierung der Überdeckungsrate der KI-basierten Gebäudedetektion (unten); die in der Legende in den eckigen Klammern stehenden Zahlen repräsentieren die Anzahl der zugeordneten Gebäude je Klasse.

In ähnlicher Größenordnung liegt der Unterschied in dem vom KI-Trainingsdatensatz nicht mittrainierten 20 % des Vermessungsamtsbezirks Bad Tölz-Wolfratshausen. Mit den Erkennungsraten für die KI-Lösung von 86,5 % und 83,0 % für die DFK-Verifikation ergibt sich eine Differenz von 3,5 %. Damit ist der empirische Nachweis erbracht, dass die über KI ermittelten Erkennungsraten (»F1 score«) robust und vertrauenswürdig sind.

### 3.2.2 Vergleich der Gebäudedetektoren aus unterschiedlichen Verfahren

In gleicher Weise lassen sich die Erkennungsraten für das Filter- und das Vergleichsklassenverfahren durch eine Flächenverschneidung bestimmen. Die Erkennungsraten der Gebäudedetektoren liegen in einem 40 km<sup>2</sup> großen Testgebiet von Bad Tölz-Wolfratshausen für das Filterverfahren bei 81,9 % und für das Vergleichsklassenverfahren bei 93,7 %. Das entspricht gegenüber der mittels KI erzielten 85,9 % zwar einer höheren Erkennungsrate für Gebäude, jedoch sind die Fehldetections von Vegetation als Gebäude in diesen beiden Ansätzen ohne KI deutlich höher (vgl. Abb. 5). Eine qualitative Bewertung der Gebäudedetektoren ist daher allein anhand der Anzahl der richtig erkannten DFK-Gebäude nicht ausreichend. Tab. 4 gibt diesen Sachverhalt deutlich wieder.

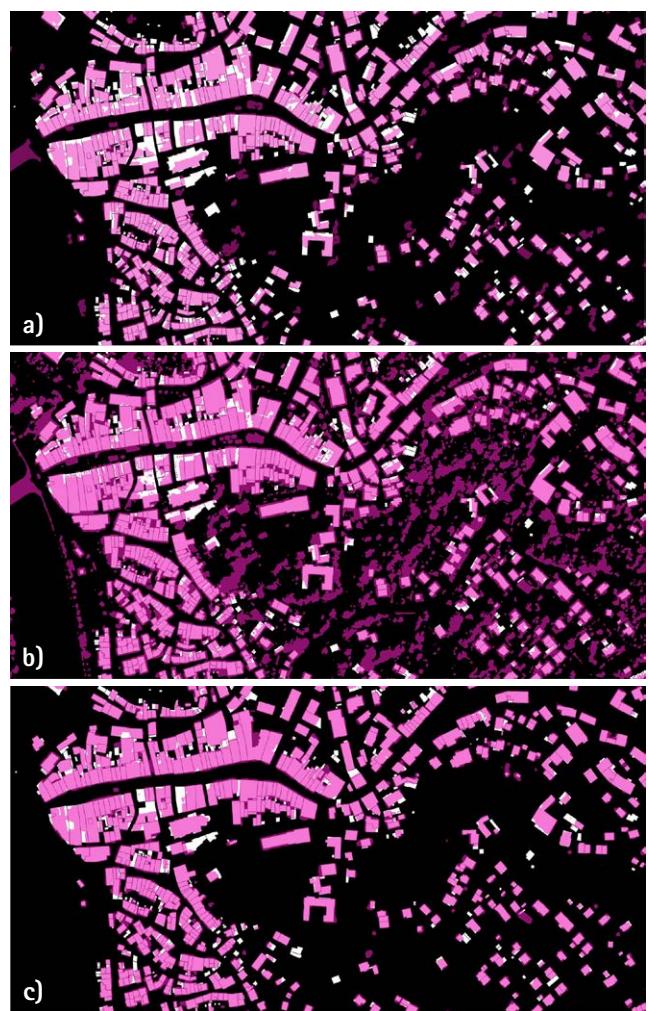


Abb. 5: Überlagerte DFK mit Gebäudedetektoren (rosa bzw. magenta Flächen) nach dem Filterverfahren (a), dem Vergleichsklassenverfahren (b) und dem KI-Verfahren (c)

Hatte das Vergleichsklassenverfahren mit 93,7 % die höchste Erkennungsrate gegenüber dem Filter- und KI-Verfahren, so kehrt sich das Ranking mit Verwendung des »F1 score« genau um und KI überflügelt deutlich die anderen Verfahren, bei dem das Vergleichsklassenverfahren schließlich am schlechtesten abschneidet. Damit ist nachgewiesen, dass im Vergleich der untersuchten Gebäudedetektoren eine sichere Gebäudeerkennung bzw.

Tab. 4: Vergleich der Gebäudedetektoren anhand des »F1 score« für ein 40 km<sup>2</sup> großes Testgebiet in Bad Tölz-Wolfratshausen

	Erkennungsrate via Gebäudeinstanzen der DFK für ca. 40 km <sup>2</sup>	F1 score flächenförmig
KI mit 1/4 Bayerns als Trainingsdaten	85,9 %	85,1 %
Filterverfahren	81,9 %	69,3 %
Vergleichsklassenverfahren	93,7 %	37,9 %

eine sichere Trennung der im nDOM verbliebenen Vegetation nur mit KI möglich ist.

Der »F1 score« ist als Vergleichsmaß hier deutlich objektiver, da er auch die Fehldetections einschließt. Der »F1 score« wurde im vorliegenden Fall anhand der Shape-Dateien der Gebäudedetektoren durch Verschneidung mit den Gebäuden in der DFK ermittelt.

### 3.3 Detektion von Neubau- und Altbaufällen

An den ÄDBV sind im operativen Betrieb möglichst fehlerfreie Hinweise über mögliche Altbau- bzw. Neubaufälle für eine wirtschaftliche Durchmusterung der TrueDOP unter Einbeziehung der aktuellen DFK erforderlich, um zuverlässig entscheiden zu können, ob im Fall von Neubaufällen Gebäude durch Messtrupps vor Ort einzumessen sind oder die abgerissenen Gebäude aus dem amtlichen Liegenschaftskataster bzw. der DFK gelöscht werden müssen. Eine sichere und möglichst vollständige Gebäudedetektion ist dafür maßgebend. Mit Blick auf die in Abb. 5 visualisierten Ergebnisse der unterschiedlichen Gebäudedetektoren scheidet das Vergleichsklassenverfahren wegen der umfangreichen Fehldetections für die Ermittlung von Neu- und Altbaufällen aus.

Die Gebäudedetektion des Filterverfahrens (via LAStools) nutzt nicht die DFK. Ebenso werden Neubaufälle bei diesem Verfahren lediglich aus dem tDOM ermittelt. Die DFK kann daher zur Überprüfung der erzielten Gebäudedetektion genutzt werden. Die mit Hilfe des Filterverfahrens erzielte Gebäudedetektion wurde mit dem in der DFK erfassten Gebäudebestand und den an den ÄDBV vorliegenden Änderungshinweisen vom ADBV visuell durchmustert. Einige der durch das Filterverfahren nicht erkannten Gebäude der DFK sind durch Kreise gekennzeichnet und die ermittelten Neubauten in Rot dargestellt (vgl. Abb. 6a). Im Vergleich zum Ausgangsbild des TrueDOP zeigt sich, dass die nicht erkannten Gebäude entweder unter einer dichten Vegetation von Bäumen liegen, sich im Schatten befinden oder sehr helle bis weiße Dächer haben. Derartige Objekte können im TrueDOP nicht erkannt werden – dies gilt sowohl für die visuelle Durchmusterung durch Mitarbeiter an den ÄDBV als auch für sämtliche Detektionsverfahren wie KI oder das Filterverfahren. Diese nicht erkennbaren Gebäude bestimmen den Bodensatz an Fehlern in der Gebäudedetektion und ließen sich nur durch eine vollständige Vor-Ort-Erkundung durch die Messtrupps an den ÄDBV vermeiden. Die Praxis hat in der Vergangenheit aber gezeigt, dass eine vollständige Erfassung durch die ÄDBV kaum gewährleistet werden kann. Daher sind alternative Methoden erforderlich geworden, die, wie Abb. 6a zeigt, ebenfalls keine hundertprozentige Erfassungssicherheit geben können.

Ein Vergleich der mittels des Filter- und des KI-Verfahrens ermittelten Neubaufälle (vgl. rote Gebäude in den Abbildungen Abb. 6c und Abb. 6d) veranschaulicht, dass

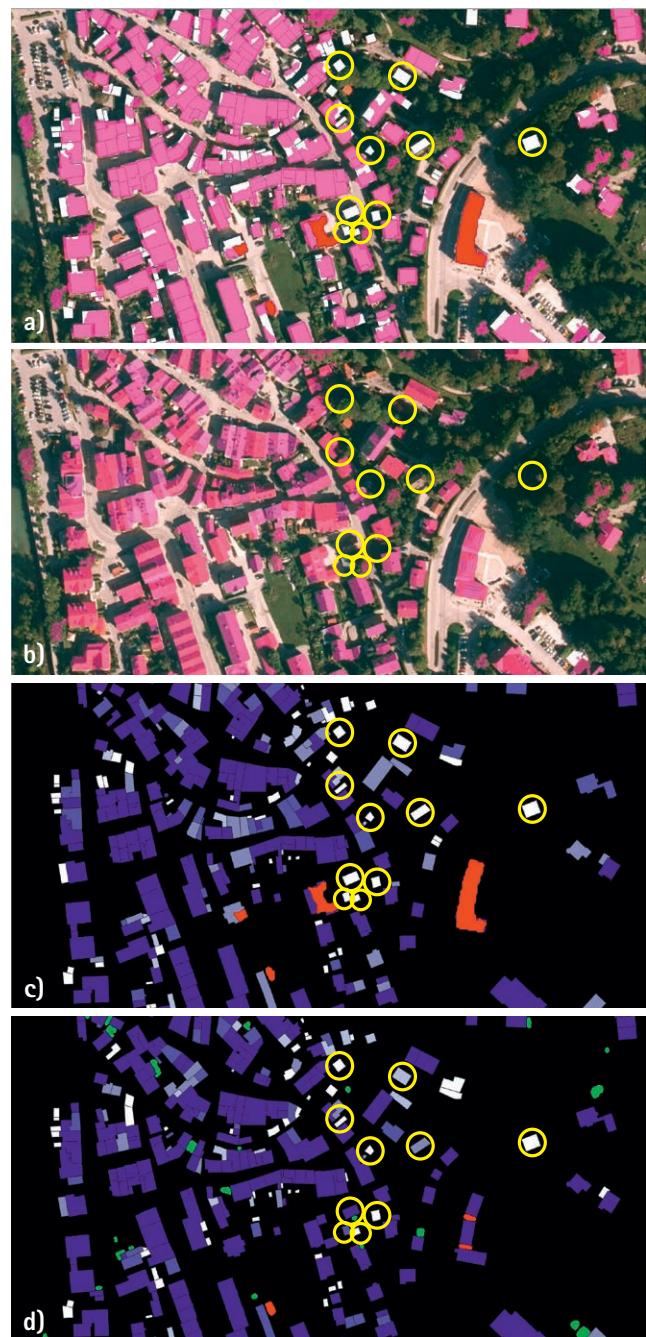


Abb. 6: a) Mit dem Filterverfahren ermittelte Gebäude-detection auf dem Hintergrund der DFK mit rot gekennzeichneten Gebäuden und Hinweiskreisen des ADBV über nicht erkannte Gebäude. b) Gebäude-detection ohne DFK. c) klassifizierte Gebäude-detection des Filterverfahrens. d) klassifizierter KI-Gebäudedetektor mit Hinweisen zu Neu- (rote) und Altbaufällen (grün).

das Filterverfahren teilweise mehr Neubauten gegenüber KI detektiert. Dies kann unter anderem darin begründet sein, dass über das Filterverfahren zusätzlich Änderungen in der Gebäudehöhe wie Gebäudeaufstockungen als Neubauten identifiziert werden, während Änderungen in der Gebäudehöhe bei KI-Verfahren unberücksichtigt bleiben. Es wäre jedoch wünschenswert, in weiteren Untersuchungen auch Höhenänderungen zusätzlich in den KI-Entscheidungsbaum zu implementieren. Derzeit

werden mit KI die Altbaufälle sicher detektiert, ohne dass offensichtliche Fehlalarme auftreten (vgl. grüne Gebäude in Abb. 6d).

Im Gegensatz zum Entscheidungsbaum des KI-basierten Verfahrens (siehe 2.2) nutzt das Filterverfahren keine DFK, sondern verwendet ausschließlich das tDOM zur Identifizierung von Neubaufällen. Daher ergeben sich Unterschiede zwischen den in Rot dargestellten Neubaufällen, wie sie über das Filterverfahren ermittelt wurden (vgl. Abb. 6c), und den über die KI-Methode ermittelten Neubaufällen (vgl. Abb. 6d). Das Filterverfahren identifiziert das rechte große rote Gebäude in Abb. 6c vollständig als Neubau, während in Abb. 6d mittels KI nur die Überdachungen der Gebäude als Neubaufälle identifiziert. Der Vergleich des in Abb. 7 dargestellten TrueDOP aus der aktuellen Epoche 2018 mit dem aus der vorhergehenden Befliegung im Jahr 2015 zeigt, dass die Schlussfolgerungen in beiden Verfahren richtig sind. Beim Filterverfahren wird aus dem tDOM, also aus der



Abb. 7: Vergleich der TrueDOP aus den Jahren 2015 (oben) und 2018 (unten)

Differenz der beiden Oberflächenmodelle der Jahre 2018 und 2015, das gesamte Gebäude als Neubau erkannt, da es im Jahr 2015 noch nicht vorhanden war. Als Ergebnis ist dieses Gebäude mit seinem gesamten Gebäudeumfang als Neubaufall in Abb. 6a und Abb. 6c richtig dargestellt.

Das KI-basierte Verfahren nutzt entsprechend des Entscheidungsbaums nur diejenigen Gebäude, die nicht in der DFK enthalten sind. Im vorliegenden Fall besteht der gesamte Gebäudekomplex aus drei eingemessenen Gebäuden, die bereits in der DFK dokumentiert sind. Deren Überdachungen sind jedoch nicht in der DFK dokumentiert, sodass mittels KI folgerichtig nur die Überdachungen als Neubaufälle detektiert werden.

Auch in den anderen in Abb. 6c und Abb. 6d in Rot dargestellten Neubaufällen ergeben sich verfahrensbedingte Unterschiede. Der Vergleich der beiden in Abb. 7 dargestellten TrueDOP-Epochen für die in Abb. 6c ausgewiesenen Neubaufälle zeigt, dass Gebäude neu errichtet oder abgerissen und an derselben Stelle neu überbaut wurden und dass diese in beiden Verfahren als Neubaufälle richtig erkannt werden.

Das KI-Verfahren weist gegenüber dem Filterverfahren durch die Verwendung der aktuellen DFK die exakteren Hinweise zu Neubaufällen aus und ist daher dem Filterverfahren vorzuziehen.

#### 4 Zusammenfassung und Ausblick

Mit dem vorgestellten neuartigen Ansatz können mit einer semantischen Detektion via KI Gebäude und Gebäudeveränderungen aus Luftbildern segmentiert werden. Voraussetzungen hierfür sind die aus der Bayernbefliegung und dem Laserscanning abgeleiteten hochgenauen und innovativen Produkte sowie die amtlichen Geobasisdaten. Die Kombination verschiedener Geobasisdaten ermöglicht in Verbindung mit neuartigen Algorithmen eine Optimierung bestehender Prozesse, wie die Baufallerkundung.

Die Untersuchungen in Abschnitt 3 haben gezeigt, dass KI im Vergleich zu den am LDBV entwickelten Filter- und Vergleichsklassenverfahren die beste Gebäudedetektion mit den geringsten Fehldetektionen aufweist. Die mittels KI abgeleiteten Neu- und Altbaufälle liefern wertvolle Hinweise und können in idealer Weise in die Baufallerkundung an den ÄDBV integriert werden.

Die Güte des eingesetzten KI-Systems hängt von der Qualität der Trainingsdaten ab. Die Gebäudedetektion stellt die einfachste Art der Unterscheidung dar, da nur zwei Klassen – Gebäude und nicht Gebäude – unterschieden werden müssen. Der in der DFK dokumentierte amtliche Bestand von über 9 Millionen Gebäuden in Bayern, der täglich durch die ÄDBV fortgeführt wird, stellt einen qualitätsgesicherten, unabhängig erfassten Trainingsdatensatz dar, der für jede Bayernbefliegungsperiode erneut zum Training des KI-Systems verwendet werden kann.

Mit den nahezu rauschfreien Ergebnissen zur Detektion der Neubaufälle liefert KI wertvolle Veränderungshinweise für eine beschleunigte Durchmusterung der Gebäudeveränderungen an den ÄDBV zur Fortführung des amtlichen Liegenschaftskatasters.

Es gilt nun, im nächsten Schritt das aus dieser Forschungscooperation erarbeitete KI-System am LDBV zu etablieren, um zukünftig im Rhythmus der Bayernbefliegung die Gebäudeveränderungen selbstständig durchführen und das Verfahren weiterentwickeln zu können – beispielsweise unter Einbeziehung der Infrarotbilder. Denkbar wäre auch eine Übertragung des KI-Systems auf weitere Fernerkundungsdaten.

**Literatur**

- Geßler, S., Krey, T., Möst, K., Roschlaub, R. (2019): Mit Datenfusionierung Mehrwerte schaffen – Ein Expertensystem zur Baufallerkundung. DVW-Mitteilungen, Heft 2, 159–187.
- Haala, N. (2011): Multiray Photogrammetry and Dense Image Matching. Photogrammetric Week 2011, Wichmann Verlag, Berlin/Offenbach, 185–195.
- Hirschmüller, H. (2011): Semi-Global Matching – Motivation, Developments and Applications. Photogrammetric Week 2011, 173–184.
- Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y. (2017): The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 11–19.
- Li, Q., Shi, Y., Auer, S., Roschlaub, R., Möst, K., Schmitt, M., Zhu, X. (2020): Segmentation of Undocumented Buildings using Deep Learning Methods and official geodata. ISPRS, Commission III, WG III/1, submitted.
- Li, X., Yao, X., Fang, Y. (2018): Building-A-Nets: Robust Building Extraction From High-Resolution Remote Sensing Images With Adversarial Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(10), 3680–3687.
- Roschlaub, R., Möst, K., Krey, T. (2020): Automated Classification of Building Roofs for the Updating of 3D Building Models using Heuristic Methods. PFG 88, 85–97. DOI: 10.1007/s41064-020-00099-9.
- Shi, Y., Li, Q., Zhu, X.X. (2018): Building Footprint Generation Using Improved Generative Adversarial Networks. IEEE Geoscience and Remote Sensing Letters, 16(4), 603–607.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F. (2017): Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8–36.

**Kontakt**

Dr.-Ing. Robert Roschlaub | Dipl.-Ing. (FH) Karin Möst |  
 Dipl.-Ing. Clemens Glock  
 Landesamt für Digitalisierung, Breitband und Vermessung  
 Alexandrastraße 4, 80538 München  
 robert.roschlaub@ldbv.bayern.de | karin.moest@ldbv.bayern.de |  
 clemens.glock@ldbv.bayern.de

M.Sc. Qingyu Li | Dr.-Ing. Yilei Shi | PD Dr.-Ing. habil. Michael Schmitt  
 Professur für Signalverarbeitung in der Erdbeobachtung, Fakultät für  
 Luftfahrt, Raumfahrt und Geodäsie, Technische Universität München  
 Arcisstraße 21, 80333 München  
 qingyu.li@tum.de | yilei.shi@tum.de | m.schmitt@tum.de

Dr.-Ing. Stefan Auer  
 Deutsches Zentrum für Luft- und Raumfahrt (DLR), Remote Sensing  
 Technology Institute, Photogrammetry and Image Analysis  
 Oberpfaffenhofen  
 stefan.auer@dlr.de

Prof. Dr.-Ing. habil. Xiao Xiang Zhu  
 Deutsches Zentrum für Luft- und Raumfahrt (DLR), Remote Sensing  
 Technology Institute, EO Data Science, Oberpfaffenhofen  
 xiaoxiang.zhu@dlr.de  
 und  
 Professur für Signalverarbeitung in der Erdbeobachtung, Fakultät für  
 Luftfahrt, Raumfahrt und Geodäsie, Technische Universität München  
 Arcisstraße 21, 80333 München  
 xiaoxiang.zhu@tum.de

Dieser Beitrag ist auch digital verfügbar unter [www.geodaeisie.info](http://www.geodaeisie.info).