Implementierung einer KI-Infrastruktur zur automatisierten Erkennung von landesweiten Gebäudeveränderungen aus Luftbildern

Implementation of an AI-Infrastructure for the Automated Detection of State-Wide Building Changes in Aerial Images

Robert Roschlaub | Clemens Glock | Karin Möst | Frank Hümmer | Qingyu Li | Stefan Auer | Anna Kruspe | Xiao Xiang Zhu

Zusammenfassung
Aufbauend auf einer vorangegangenen Studie über den Einsatz von Künstlicher Intelligenz (KI) zur Detektion von Gebäudeveränderungen im amtlichen Liegenschaftskataster, wird in diesem Beitrag die Transformation von der Projektphase in einen Produktivbetrieb vorgestellt, es werden verschiedene KI-Architekturen angesprochen, miteinander verglichen und eine Auswahlentscheidung gegeben. Die erzielten Ergebnisse einer landesweiten Gebäudedetektion werden gemeinsam mit Untersuchungen zu den Trainingsläufen, zur Performance und den von einem Vermessungsamt validierten KI-Ergebnissen präsentiert.

Schlüsselwörter: Gebäudeerkennung, Neuronale Netze, KI-Infrastruktur, Luftaufnahmen, Oberflächenmodelle

Summary
Based on a previous study on the use of Artificial Intelligence (AI) for the detection of building changes in the official real estate cadaster, this paper presents the transformation from the project phase to the phase of production, addresses different AI architectures, compares them and explains the selection decision. The obtained results of state wide building detection are presented together with investigations on the training runs, the performance and the AI results validated by a surveying office.

Keywords: *building detection, neural networks, AI infrastructure, aerial photography, surface models*

1 Einführung

In vielen Bundesländern besteht eine Pflicht zur Einmessung von Gebäuden oder deren Veränderungen. Informationen über die erfolgten Gebäudeveränderungen sind unabhängig hiervon wertvoll, um eine möglichst vollständige Fortführung und hohe Aktualität bezüglich des Gebäudebestandes im amtlichen Liegenschaftskataster zu gewährleisten. Am Beispiel von 14 bayerischen Landkreisen wurde bereits gezeigt, wie aus Luftbildern mittels KI-basierter Methoden (Deep Learning) Gebäudeveränderungen gegenüber dem amtlichen Liegenschaftskataster automatisiert detektiert werden können, die katastertechnisch noch nicht eingemessen worden sind. Die detektierten Gebäudeveränderungen lassen sich in Gebäudeneubauten bzw. -anbauten und Altbauten klassifizieren (Roschlaub et al. 2020, Li et al. 2020). Im letzteren Fall (dem sog. Altbaufall) unterblieb die katastertechnische Einmessung und Dokumentation im amtlichen Liegenschaftskataster bereits mehrere Jahre.

Die Segmentierung von Gebäuden aus amtlichen Geodaten[1] mittels Deep Learning ist durch die landesweite Prozessierung der Luftbilder aus den Bayernbefliegungen der Jahre 2019/2020 für Nord- und Südbayern erstmalig von der Evaluierungs- und Testphase in einen produktiven Prozess überführt worden. Es erforderte erheblichen Programmieraufwand, um die verschiedenen Datenbestände in automatisierte Prozesse zu integrieren.

Dies gilt sowohl für die Bereitstellung der Ausgangsdaten, deren Überführung in einheitliche Datenformate, die für das Trainieren des KI-Netzwerks benötigt werden, als auch für die anschließende automatische Erkennung (Inferenz) von Gebäudeveränderungen aus den maßstabsgetreuen TrueOrthophotos (TrueDOP)[2].

1 Als Datengrundlage dienen die Digitale Flurkarte (DFK) mit den bestehenden amtlichen Gebäudegrundrissen, das TrueOrthophoto (TrueDOP), das gegenüber dem klassischen Digitalen Orthophoto (DOP) maßstabsgetreue Grundrisse enthält. Als weitere Daten dienen das aus dem Airborne Laserscanning (LiDAR) abgeleitete Digitale Geländemodell (DGM) und das aus dem Verfahren des Dense Image Matching abgeleitete regelmäßige Gitter eines bildbasierten Oberflächenmodells (bDOM). Aus zwei unterschiedlichen Luftbildbefliegungen lässt sich durch eine Differenzbildung der zugehörigen bDOM-Epochen ein zeitliches Differenzmodell (tDOM) berechnen und aus der Differenz eines bDOM zum DGM ein normalisiertes Digitales Oberflächenmodell (nDOM).

2 Das TrueDOP soll nach einem Beschluss der AdV von den Ländern bis Anfang des Jahres 2023 flächendeckend bereitgestellt werden (AdV 2017). Die Darstellung von Gebäuden und deren Dachflächen ist im TrueDOP lagerichtig zu den Katastergrundrissen, wenngleich die Dachflächen je nach Dachüberstand über die Gebäudegrundrisse im Kataster hinausragen können. Damit eignet sich das TrueDOP in idealer Weise für eine semantische Segmentierung von Gebäudedächern bzw. von Gebäudeveränderungen.

DOI 10.12902/zfv-0389-2022

Der vorliegende Beitrag baut auf den gewonnenen Erkenntnissen auf und widmet sich den Herausforderungen, die sich auf dem Weg von einem Ansatz für die Erkennung undokumentierter Gebäude zu einem operationellen System am bayerischen Landesamt für Digitalisierung, Breitband und Vermessung (LDBV) stellten. Aspekte der Wissenschaft, Softwareentwicklung und Systemtestung fließen dabei zusammen. Im Folgenden werden die wesentlichen Arbeitsschritte zum Aufbau einer KI-Infrastruktur beschrieben, die notwendig waren, um das Verfahren von der Projektphase in einen regulären Produktionsprozess zu überführen. Dabei werden:

- Architekturmodelle vorgestellt,
- die Anzahl notwendiger Rechendurchläufe (Epochen) beim Trainieren des KI-Systems diskutiert und getestet,
- Aspekte der Leistungsfähigkeit eines KI-Systems hinsichtlich der eingesetzten Grafikkarten (Graphics Processing Unit, GPU) betrachtet,
- die Nutzung von Python und Anaconda (in KI oft eingesetztes Paketverwaltungssystem für Python) angesprochen,
- Vorteile einer Datenbankanbindung erläutert
- sowie auf die erforderliche Umstellung des Betriebssystems von Ubuntu auf Red Hat Enterprise Linux eingegangen.

Des Weiteren wird das entwickelte semantische Segmentierungsverfahren zur Detektion von Gebäudeneubauten und Altbauten um die Detektion von Gebäudeaufstockungen und Gebäudeabrissen erweitert und die erzielten Ergebnisse aus der landesweiten semantischen Klassifizierung von Gebäudeveränderungen aus Luftbildern werden vorgestellt.

Der Beitrag ist in fünf Abschnitte untergliedert. Abschnitt 2 beschreibt den aktuellen Stand der Technik und stellt das gewählte Segmentierungsverfahren zur Gebäudeerkennung vor. Dabei werden zudem verschiedene Neuronale Netze als Optionen für den Einsatz von künstlicher Intelligenz verglichen. Abschnitt 3 befasst sich mit den Herausforderungen und Entscheidungen auf dem Weg zur notwendigen KI-Infrastruktur am LDBV. Der Ansatz zur Erkennung von in der DFK nicht dokumentierten Gebäuden, der im Kern auf den Gebäudedetektor aus Abschnitt 2 zurückgreift, wird in Abschnitt 4 vorgestellt. Außerdem werden die erzielten Ergebnisse und deren Auswirkungen in der praktischen Verwendung am Vermessungsamt präsentiert und diskutiert. Abschnitt 5 fasst als Abschluss die gewonnenen Erkenntnisse zusammen und wirft einen Blick auf denkbare Erweiterungen in der Zukunft.

2 Architekturmodelle

Das Verfahren zur Gebäudeerkennung, welches neben den amtlichen Geobasisdaten das Kernelement für die Lösung der Aufgabe darstellt, basiert auf einem Neuronalen Netz. Dieses beschreibt ein Modell, das es erlaubt, Merkmale in Bildausschnitten durch Filterung und nichtlineare Funktionen in einen alternativen Merkmalsraum zu projizieren (Convolutional Neural Network – CNN). Mit Hilfe der Projektion lassen sich die beiden Klassen »Gebäude« und »kein Gebäude« effektiver voneinander trennen als mit traditionellen Verfahren, wie z. B. Random Forest Klassifikatoren oder Support Vector Machines, für welche die entsprechenden Bildmerkmale händisch definiert werden müssten. Die Parameter für die Filterung und Projektion werden mit Hilfe von Referenzdaten gelernt, d. h. dem Neuronalen Netz werden Hunderttausende von Beispielen für das Erscheinungsbild von Gebäuden in Luftbildern und die dazugehörige zu treffende Entscheidung gegeben. Das Neuronale Netz nutzt diese Trainingsdaten, um charakteristische Gebäudemerkmale deutlicher hervorzuheben und für jedes Bildpixel die Grundlage für eine automatische Entscheidung »Gebäude/kein Gebäude« zu liefern.

Die Aufgabe einer Gebäudeerkennung mit CNN ist ein Beispiel für die semantische Segmentierung in der Computer Vision, die darauf abzielt, für jedes Pixel eines Bildes eine Klassifizierung festzulegen. Für die semantische Segmentierung werden häufig verschiedene CNN-Architekturen verwendet, z. B. Fully Convolutional Nets (FCN) (Long et al. 2015) und Encoder-Decoder-basierte Architekturen, z. B. U-Net (Ronneberger et al. 2015), Residual Network (ResNet; He et al. 2015), Densely Convolutional Network (DenseNet; Huang et al. 2016), sowie solche, die beide Grundprinzipien vereinen, wie z. B. FC-DenseNet (engl. Fully Connected Convolutional Dense Network) (Jégou et al. 2017) und DeepLabV3+ (Chen et al. 2018). Solche Architekturen übertreffen traditionelle Ansätze in der Erkennungsleistung deutlich. Ein FCN ist ein traditionelles Deep-Learning-Verfahren (DL-Verfahren) für Aufgaben der Klassifikation und basiert auf einer klassischen Abfolge von Faltungsoperationen und nichtlinearen Projektionen. Neben der FCN-Architektur ist auch die Erkennungsleistung anderer Varianten, wie z. B. Encoder-Decoder-basierter Architekturen, bemerkenswert. Bilddetails werden dabei im Encoder schrittweise reduziert und im Decoder wiederhergestellt, während gleichzeitig die Berücksichtigung von räumlichem Kontext im Bild zu- und abnimmt.

Die Aufgabe einer landesweiten Erkennung von Gebäuden in Bayern ist für die Verwendung von DL-Verfahren prädestiniert. Als Gründe hierfür sind aufzuführen:

- Die zyklische Digitale Luftbildbefliegung mit der Abdeckung Bayerns im Zweijahresturnus (ca. 50.000 Luftbilder pro Jahr mit einem Speichervolumen von 150 TB), ihrer hohen räumlichen Auflösung, ihrer Georeferenzierung und ihrer nachweislichen Qualitätssicherheit bietet einen optimalen Startpunkt für die Aufgabe.
- Die Digitale Flurkarte stellt mit ihrer hohen Aktualität, Genauigkeit und Objektstrukturierung eine herausragende Grundlage für das Training und anschließende Validieren der Erkennungsleistung dar.

- Der hohen Rechenintensität des maschinellen Lernens lässt sich mehr und mehr mit der Leistung von potenten Grafikkarten und verbesserten DL-Umsetzungen begegnen. DL-Verfahren lassen sich als Modul im Prozess austauschen, um mit den neuen Entwicklungen Schritt halten zu können.

2.1 Verwendetes Architekturmodell

Für die semantische Segmentierung der Gebäude (Gebäudedetektion) wird der FC-DenseNet-Ansatz nach Jégou et al. (2017) verwendet, ein mit dichten Blöcken (Dense Blocks) und Sprungverbindungen (Skip Connections) arbeitendes Neuronales Netz. Die Netzwerkarchitektur FC-DenseNet ist im Vergleich zu alternativen Netzwerkarchitekturen am besten für Detektionsaufgaben städtischer Szenen geeignet. Sie wird im vorliegenden Fall gemeinsam mit Oberflächenmodellen zur Erkennung von Gebäuden in hochaufgelösten Luftbildern (Bildkanäle: Rot, Grün, Blau) eingesetzt.

Die in Roschlaub et al. (2020) bereits beschriebene Netzwerkarchitektur FC-DenseNet weist die charakteristischen Merkmale eines sogenannten U-Netzes auf (vgl. Abb. 1). Im Encoder- und Decoder-Bereich des Netzes werden insgesamt neun Dense-Blöcke aufgebaut (Huang et al. 2016). Die Bildinformation wird in diesen mit Filterungen und Projektionen mit Aktivierungsfunktion weiterverarbeitet, fließt an den Dense Blocks aber über Abkürzungen auch direkt in tiefere Ebenen ein. Das Wiederzusammenführen der Information zu Bildstapeln wird über eine Tensor-Verknüpfung realisiert. Dadurch kann die aktuelle Ebene innerhalb eines Dense Blocks auf Rohdaten vorheriger Ebenen zugreifen. Die nichtlineare Projektion von Bildinformation durch das vom Netzwerk definierte Modell wird »Forward Propagation« genannt. Die Anzahl von Bildausschnitten, die für den Informationsfluss verwendet wird, heißt »Batch Size«.

Zur Verringerung des Detailierungsgrads wird zwischen den Dense Blocks die räumliche Auflösung über eine Pooling-Operation auf die Hälfte reduziert. Die Größe des Merkmalraums wird gleichzeitig auf ein Viertel reduziert. Zudem wird die eingehende Information der Bildausschnitte radiometrisch normiert, um die Generalisierung und Robustheit des Netzwerks zu verbessern (Batch Normalization mittels Mittelwert und Varianz der für den aktuellen Trainingsschritt verwendeten Bildausschnitte). Die Bildinformation wird im Anschluss durch eine nichtlineare Aktivierungsfunktion (ReLU) projiziert und mit Filteroperationen verändert, um repräsentative Merkmale herauszuarbeiten.

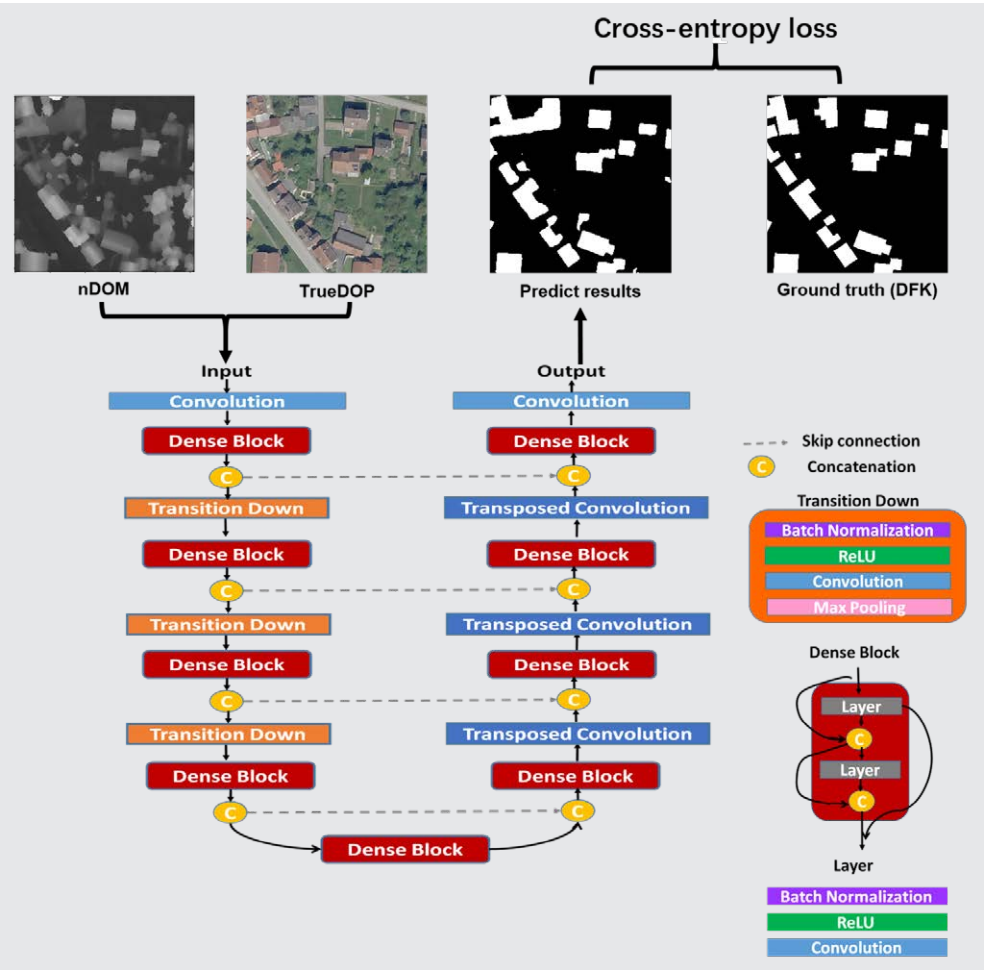

Abb. 1: Netzwerkarchitektur (FC-DenseNet). Die Integration von Dense Blocks dient dem verbesserten Informationsfluss im Netzwerk.

Nachdem die Bildausschnitte auf der untersten Ebene im Encoder-Bereich die minimale Größe von 16 × 16 Pixel erreicht haben (diese betonen Merkmale in minimaler Auflösung und mit maximalem Kontext), wird im Decoder-Bereich des U-förmigen Netzes die originale räumliche Auflösung wiederhergestellt. Die Erhöhung der Bildabtastung erfolgt im Rahmen von Blöcken zur Dekonvolution (Umkehrung der sog. Faltungsoperation) und eingebautem Unpooling (zur Wiederherstellung der Auflösung), während bei den Dense Blocks weiter Projektionen und Filtermatrizen eingesetzt werden. Mit jedem Unpooling wird die Größe der Bildmatrizen und die Abtastung der darin enthaltenen Objekte schrittweise um den Faktor 2 erhöht. Am Ende des Prozesses gibt das

FC-DenseNet für jedes Bildpixel im Ausgangsbild einen Vektor mit zwei Elementen aus, in denen die Wertungen für »Gebäude« und »kein Gebäude« enthalten sind. Der größere Wert gibt die Entscheidung vor.

Das Neuronale Netz ist in einen Entscheidungsbaum integriert (siehe Abschnitt 4.2), der für die Suche nach undokumentierten Gebäuden auf die DFK zurückgreift und mit Hilfe eines vorhandenen tDOM und nDOM nachfolgend eine verfeinerte Einordnung zur Unterscheidung zwischen Neubaufall oder Altbaufall vornimmt (Roschlaub et al. 2020).

2.2 Untersuchung verschiedener Netzwerke

Im Zusammenhang der Verfahrensentwicklung zur Gebäudeerkennung wurde die Erkennungsleistung verschiedener CNN-Varianten im Vergleich zu FC-DenseNet untersucht, nämlich zu Fully Convolutional Nets (FCN), U-Net und DeepLabV3+. Gegenüber FCN erzielt FC-DenseNet einen Zuwachs von 3,9 % im F1-Score (Li et al. 2020), einem Maß, das auf der Richtigkeit und Vollständigkeit der Detektionsergebnisse beruht. Dank seiner Architektur, die den feinen geometrischen Bilddetails mehr Bedeutung zuordnet, ist FC-DenseNet in der Lage, schärfere Gebäudegrenzen zu erhalten als FCN. Im Vergleich zu U-Net, einem Netzwerk mit U-Form aus Encoder- und Decoder-Struktur (nachfolgend U-Netz genannt), erreicht FC-DenseNet Verbesserungen von 3,2 % im F1-Score (Li et al. 2020).

In einem weiteren Vergleich wurde festgestellt, dass FC-DenseNet in der Gegenüberstellung zu DeepLabV3+ eine Verbesserung von 9,0 % im F1-Score erzielt. Dies könnte darauf zurückzuführen sein, dass die Down- und Upsampling-Module in DeepLabV3+ die Segmentierungsleistung auf Luft- bzw. Satellitenbildern erheblich beeinträchtigen, insbesondere wenn die Bilder Objekte enthalten, die nur einen kleinen Bereich abdecken (Tasar et al. 2020), d. h. im vorliegenden Fall kleine Gebäude. Außerdem gibt es in DeepLabV3+ keine Sprungverbindungen (Skip Connections) zwischen den beiden Seiten des U-Netzes für die Erhaltung von räumlichen Detaillierungen im Netzwerk. Diese Verbindungen sind für die Gebäudeerkennung von hoher Wichtigkeit, da sie in der Lage sind, Merkmale in den oberen und unteren Schichten des Netzwerks zu reproduzieren. Dadurch wird ein effizienterer Pfad für einen Informationsfluss geschaffen, durch den Details der Gebäude erhalten bleiben.

In der Gruppe der getesteten CNN-Varianten ist das FC-DenseNet das beste Netzwerk in Bezug auf die numerische Genauigkeit und die visuellen Ergebnisse. Einerseits werden in eng vermaschten Blöcken (DenseNet-Blöcken) Merkmalsarten verschiedener Auflösung miteinander verknüpft, wodurch die Information für nachfolgende Schichten des Netzwerks variabler wird. Andererseits können hochfrequente Informationen durch direkte Verbindungen zwischen dem Encoder und dem Decoder übertragen werden, wodurch räumliche Details besser erhalten bleiben (Li et al. 2020). Die Vorteile von FC-DenseNet bei der Erkennung von Gebäuden wurden auch in anderen Forschungsarbeiten nachgewiesen (Shi et al. 2020, Li et al. 2018).

2.3 Äquivalenz zweier Netzwerke

Aus Sicht von Chen et al. (2017) besteht eine Äquivalenz von Residual Network (ResNet) und Densely Convolutional Network (DenseNet), auf dem das FC-DenseNet basiert. Grundsätzlich beruhen beide auf dem gleichen Prinzip: der Propagation von Informationen in spätere Schichten, die ihnen nicht direkt nachgeschaltet sind.

Während der DenseNet-Ansatz durch Verkettung weitere Merkmale aufbaut, die auch für das Erkennen der räumlichen Verteilung der Objekte wichtig sind, kann ein ResNet nach He et al. (2015) Filterergebnisse seiner in der ersten Ebene durchgeführten Merkmalsextraktion durch Abbilden auf die folgenden Ebenen, die auch eine Einheitsabbildung (Identity-Mapping) verwendet, optimal wiederverwenden und vorhandene Nicht-Linearitäten über die Tiefe des Netzes immer weiter modellieren. Durch die Einheitsabbildung wird bei ResNet das Training stabilisiert, die Kapazität der Merkmalsrepräsentation bleibt im Vergleich zum DenseNet-Ansatz jedoch beschränkt (Zhang et al. 2020).

Ein wichtiger Aspekt bei der Gebäudeerkennung über semantische Segmentierung sind möglichst scharfe Objektgrenzen. Dazu ist aus unserer Sicht eine Encoder-Decoder-Architektur mit Sprungverbindungen notwendig, um feine Strukturen innerhalb des Netzwerks zu erhalten. Innerhalb des Encoder-Bereichs werden ResNet- oder DenseNet-Ansätze eingesetzt, die das Ziel haben, auf der untersten Ebene des Encoders Gebäude- und Nicht-Gebäudemerkmale in sehr grober Auflösung darzustellen. Der Decoder-Bereich, der die originale räumliche Auflösung für die erkannten Objekte wiederherstellen kann, baut die Ebenen nacheinander mit Hilfe von Umkehrfunktionen in größerer Auflösung wieder auf. Dazu kommen Umkehrungen der Filter-Operationen, die Dekonvolution-Filterungen und Unpooling-Ebenen zum Einsatz. Die in Drozdzal et al. (2016) als lange Sprungverbindungen (englisch: long skip connection) bezeichneten Verknüpfungen zwischen dem Encoder- und Decoder-Teil des U-Netzes dienen der Ergänzung von geometrischen Details von Gebäuden bis in die Entscheidungsebene des Netzwerks hinein.

Merkmale von innerhalb des Encoders eingesetzten ResNet-Ansätzen werden über eine lange Sprungverbindung durch Summation mit den Merkmalen im Decoder-Bereich verbunden, die dort durch Upsampling-Operationen aus Merkmalen einer tieferen Ebene gebildet werden. Dagegen werden bei DenseNet-Ansätzen im Encoder-Bereich die Merkmale der entsprechenden Ebenen des Encoder- und Decoder-Bereichs wieder über eine Tensor-Verknüpfung (englisch: tensor concatenation) in Beziehung gebracht (Drozdzal et al. 2016).

2.4 Notwendige Trainingsläufe (Epochen)

Ein Durchlauf für das Training eines Neuronalen Netzes wird als Epoche bezeichnet. Alle Trainingsdaten werden hierbei einmal für die Optimierung der Parameter des Neuronalen Netzes verwendet. Es stellt sich die Frage, wie viele Epochen notwendig sind, um das Netzwerk für seine Aufgabe optimal zu trainieren. Ziel ist es, den Unterschied zwischen der Entscheidung des Netzwerks und der bekannten Referenzinformation, der mit einer Verlustfunktion gemessen wird, so weit wie möglich zu reduzieren, d. h. die durch das Netzwerk prädizierte Information an Gebäuden muss in den Trainingsgebieten an den amtlich dokumentierten Datenbestand der DFK (GroundTruth) »herankommen«. Die Differenz wird als Kreuzentropie (Cross-Entropy Loss) bezeichnet und beschreibt die Genauigkeit der prädizierten Gebäudeerkennung im Vergleich zur Referenz (vgl. Abb. 2). Die Kreuzentropie wird mit Hilfe einer Verlustfunktion Pixel für Pixel berechnet (Validation Loss) und über alle verwendeten Bildkanäle aufsummiert.[3]

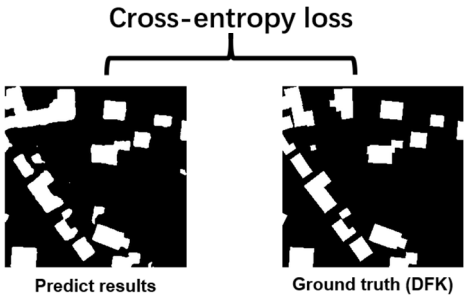

Abb. 2: Beziehung zwischen den durch das FC-DenseNet prädizierten Gebäude (links) und dem amtlichen Datenbestand der DFK (rechts)

Verlustfunktion (Kreuzentropie):

$$L\left(\overline{y}^{ij}, y^{ij}\right) = -\sum_{l=1}^{k} y^{ij,l} \log\left(\overline{y}^{ij,l}\right) \quad (1)$$

Übertragungsfunktion (Gesamtverlust):

$$\sum_{i=1}^{m}\sum_{j=1}^{n} L\left(\overline{y}^{ij}, y^{ij}\right) = c \equiv F\left(W, b\right) \quad (2)$$

Der Wert c der Übertragungsfunktion F ist die Summe an Verlustfunktionen L über alle Pixel aus den prädizierten Werten $\overline{y}^{ij}$ und den aktuellen Werten y^{ij} zum ij-ten Pixel über das gesamte Bild. Der Wert des Gesamtverlustes der Kreuzentropie c in Gleichung (2) gibt darüber Aufschluss, in welchem Umfang die Gewichte in den Filteroperatoren des Neuronalen Netzes für ein verbessertes Ergebnis angepasst werden müssen. Im vorliegenden Fall werden vier Kanäle (RGB, nDOM) als Eingangsinformation verwendet.

Die Optimierung der Gewichte im Netzwerk erfolgt durch Minimierung der Übertragungsfunktion F zur Aktualisierung der trainierbaren Parameterwerte der Gewichte W und der Translation b. Dazu wird die Ableitung der Verlustfunktion in Abhängigkeit der Gewichte des Netzwerks bestimmt, um die Gradienten zu erhalten. Die Optimierung folgt im Anschluss mit Hilfe eines Gradientenverfahrens (Stochastic-Gradient-Descent-Verfahren) auf der Suche nach einem Minimum der Verlustfunktion. Die Gewichte im Netzwerk werden dazu ausgehend von der tiefsten Ebene bis hin zur obersten Ebene auf einen neuen Wert aktualisiert (Back Propagation).

Nach der Verwendung aller Trainingsdaten für die Optimierung ist eine Epoche durchlaufen. Über eine Abfolge vieler Epochen erreicht die Optimierung ein lokales Minimum der Verlustfunktion.

Am Beispiel der Befliegungen von Nordbayern aus dem Jahr 2019 wurden die Ergebnisse für das Trainieren des FC-DenseNet für 80 und 130 Epochen untersucht. Dabei wurden 20 % der Referenzdaten als unabhängige Validierungsdaten genutzt, also als Beispiele, die das Netzwerk

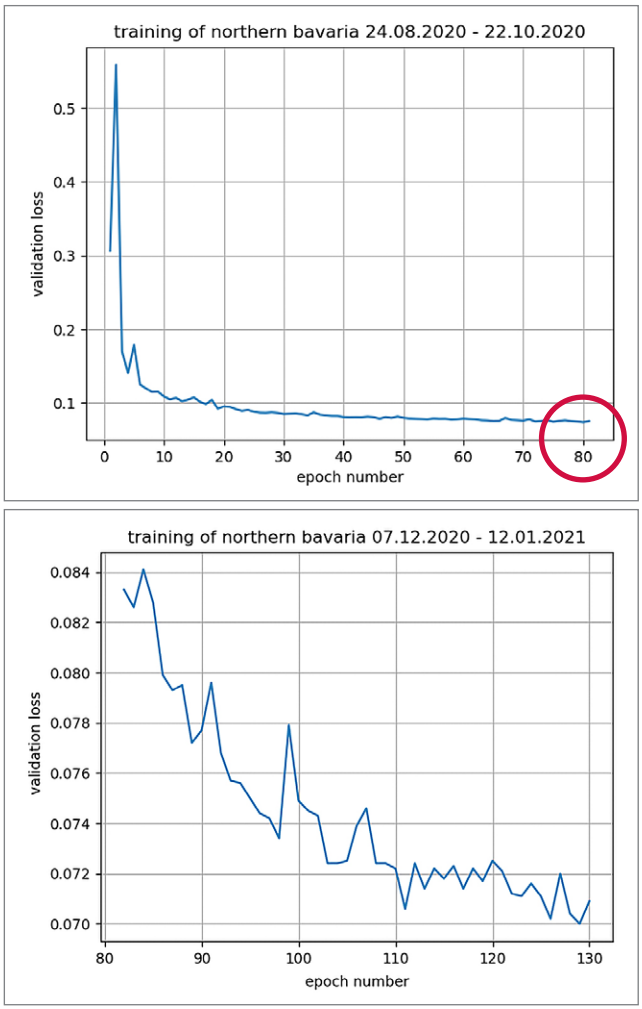

Abb. 3: Verlauf des Gesamtverlusts (Betrag c in Gleichung 2) bis Epoche 80 (oben) und zwischen den Epochen 80 und 130 (unten)

3 Im vorliegenden Fall wurden die TrueDOP-Kacheln mit den drei Farbkanälen RGB verwendet. Zukünftig könnte auch der Infrarotkanal genutzt werden.

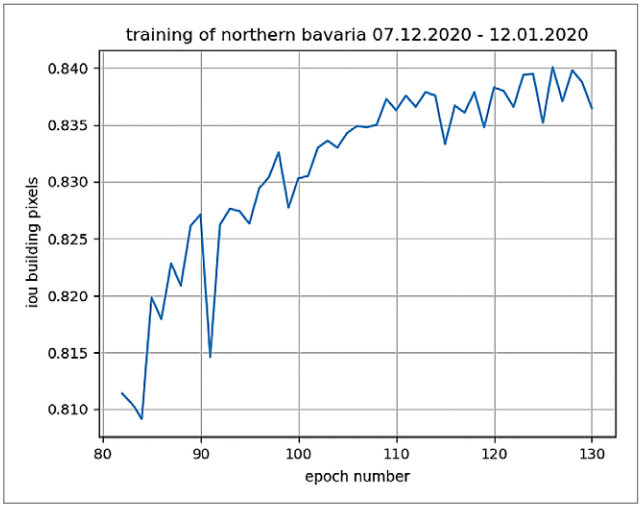

Abb. 4: Genauigkeitsmaß »Intersection over Union« (IoU) zwischen den Trainingsepochen 80 und 130, welches das Ausmaß der räumlichen Überlagerung von Detektion und Referenz beschreibt.

während des Trainings nicht sieht und auf denen dann Gütemaße berechnet werden, um die Übertragbarkeit der gelernten Klassifikation auf neue Daten zu bestimmen. Für die Berechnung auf einem Rechner mit 2 GPU wurde für die 130 Epochen insgesamt eine Rechenzeit von vier Monaten benötigt. In Abb. 3 zeigt sich, dass bereits ab 30 Epochen das System nur sehr langsam konvergiert, d. h. der Gesamtverlust langsam abnimmt, und dass ab 110 Epochen das System bereits alternierende Amplituden des Gesamtverlusts zeigt und kaum noch konvergiert. Mit Blick auf das Qualitätsmaß der Intersection over Union (IoU) für die räumliche Überlagerung zwischen dem prädizierten Detektionsergebnis von Gebäuden und dem Datenbestand der DFK als amtliche Referenz, siehe auch Roschlaub et al. (2020), ergibt sich ebenfalls ab 110 Epochen das Konvergenzverhalten mit alternierenden Amplituden (vgl. Abb. 4). Die genaue Anzahl der Epochen ist abhängig von Anfangsinitialisierung, Lernrate (siehe auch Abschnitt 3.2) und Datenkonfiguration; dennoch zeigt sich hier ein deutliches Konvergenzverhalten mit einer Erkennungsrate von 83,5 % für Nordbayern als Ergebnis.

3 Aufbau einer KI-Infrastruktur

3.1 Software-Plattform, Entwicklungen für Prä- und Postprozessierung

Als Plattform für das Trainieren der Gebäudedetektion sowie für die Vor- und Nachprozessierung (Prä- und Postprozessierung) der notwendigen Raster- und Vektordaten wurde Python gewählt. Neben seiner Funktion als Sprache für die Steuerung des KI-Frameworks PyTorch (https://pytorch.org/) und der Definition der Convolutional Neural Networks (CNN) hat sich Python als vielseitige und universale Programmiersprache für die Entwicklung von

Abb. 5: Ergebnis der Gebäudeerkennung für 80 und 130 Epochen

Verfahrensabläufen in den verschiedenen Prozessierungsschritten herausgestellt.

Das Ziel der Präprozessierung liegt in der Erstellung von Trainings- und Testkacheln für die definierten Gebiete, wie es bereits in Roschlaub et al. (2020) dargestellt wurde. Diese Daten beliefern das CNN über einen Datenlader (englisch: Data Loader). Bei der Präprozessierung werden zunächst unterschiedliche Raster- und Vektordaten miteinander verschnitten. Neben der Vektordatenverarbeitung sowie der Erstellung von Prozess- und Metadaten in offenen Datenbanken – bei der Baufallerkundung die Datenbanken SpatiaLite (www.gaia-gis.it/fossil/libspatialite/index) und PostGis (https://postgis.net/) – zeigt Python hauptsächlich bei der Rasterverarbeitung seine Stärken. Die in der Programmbibliothek NumPy (https://numpy.org/) verwendeten Werkzeuge zur Verarbeitung von numerischen Daten, z. B. von Matrizen und Rasterbildern, ermöglichen eine einfache Kombination und Verschneidung von Rasterbildern unterschiedlicher Datensätze.

Die Erfahrung zeigt, dass die Einarbeitung in das Geoprozessing mit Python erleichtert wird, weil hinter den Python-Routinen meist bekannte Open-Source-Bibliotheken arbeiten, wie z. B. die im OpenGIS-Bereich häufig eingesetzte Bibliothek Gdal (https://gdal.org/), deren Funktionalität in der Regel schon aus Vorprojekten vertraut ist. Die Funktionalität wird in Python in einfacher Syntax zur Verfügung gestellt. Neben den bereits angesprochenen Paketen zur elementaren Verarbeitung von Geodaten (Geoprozessierung) werden auch allgemeine Pakete zur Datenanalyse wie Pandas (https://pandas.pydata.org/) oder

Geopandas (https://geopandas.org/) eingesetzt. Aufgrund der hohen Dynamik der Python-Pakete ist eine Administration der Pakete über eine Weboberfläche nicht geeignet. Gerade in den Bereichen KI und Big Data ist man aufgrund der derzeit großen Fortentwicklungen auf aktuelle Pakete angewiesen. Anaconda (www.anaconda.com) ist bei der Verwaltung der Python-Pakete benutzerfreundlich und leistungsstark, weil es vor allem die Abhängigkeiten der Pakete untereinander berücksichtigt. Der Einsatz derartig verteilter Python-Pakete und deren Zugriff auf offen im Netz liegende Paketquellen ist nicht ganz gefahrlos. In Bayern ist dazu eine Prüfung der Zugriffe innerhalb des Behördennetzes durch das Landesamt für Sicherheit (LSI) Bayern erforderlich.

Als Ziel für eine erste Ausbaustufe einer KI-Infrastruktur für die Baufallerkundung am LDBV wurde eine weitgehende Automatisierung der notwendigen Prozesse für die landesweite Gebäudedetektion angestrebt. Sie erfolgte aufgrund des Zyklus der Bayernbefliegung getrennt nach Nordbayern (2020) und Südbayern (2021). Diese Automatisierung betrifft die Präprozessierung der Ausgangsdaten und die Postprozessierung der über KI detektierten Gebäude. Eine zukunftsgerichtete Konzeption mit dem ausschließlichen Einsatz einer Datenbank für die kombinierte Datenhaltung und Prozesssteuerung zur automatisierten Verarbeitung der Vektor- und Rasterdaten konnte in den Jahren 2020/2021 noch nicht abgeschlossen werden. Zum Einsatz kam eine dateibasierte Datenhaltung der Rasterdaten auf Basis von Landkreisen zum Aufbau der Trainingsdaten, die über eine Datenbank angesprochen werden, welche neben zugehörigen Prozessdaten auch einen sekundären ALKIS-Gebäudebestand enthält. Auf dieser Grundlage können automatisierte Prozesse für Nord- oder Südbayern ohne Benutzer-Interaktion ablaufen.

Für das Postprozessing wurde analog ein Dateisystem aufgebaut, das Bilder und Folgeprodukte für den Zweck der späteren Abgabe in Gebietseinheiten von Vermessungsämtern speichert. Dort werden u. a. die vom KI-System erzeugten Binärbilder abgelegt, die die detektierten Gebäude in Weiß auf einem schwarzen Hintergrund darstellen.

3.2 Hardware, Betriebssystem und erzielte Trainings-Performance

Im Frühjahr 2020 wurde ein erster KI-Server mit zwei Quadro RTX 6000 GPUs der Fa. Nvidia mit dem Betriebssystem Ubuntu LTS 20.04 in Betrieb genommen. Auf Basis von Ubuntu gelang das Einrichten des Systems, auf dem innerhalb weniger Tage erste Test-Trainingsläufe erfolgreich durchgeführt werden konnten. Hilfreich zeigte sich Ubuntu ebenfalls beim Einrichten von Test-Plattformen, die im Umfeld von KI auszuprobieren waren. Mit Aufbau eines zweiten KI-Servers, ein Jahr später, mit 8 GPUs des Typs A100 der Fa. Nvidia wurde das System zur Baufallerkundung für den produktiven Einsatz im Rechenzentrum LDBV IT-DLZ konsolidiert und relativ problemlos auf das Betriebssystem Red Hat RHEL 8.4 (Oopta) umgestellt. Aufgrund der guten Kapselung des KI-Systems konnte mit Hilfe der Python-Plattform Anaconda das gesamte System einschließlich Prä- und Postprozessierung erfolgreich über eine Installation der Python-Umgebung nach Red Hat portiert werden.

Sowohl Nordbayern im Jahr 2020 als auch Südbayern im folgenden Jahr wurden auf dem ersten KI-Rechner mit seinen 2 GPUs mit jeweils 24 GB Speicher prozessiert. Das KI-System wurde für Nord- und Südbayern getrennt voneinander trainiert und validiert. Für Nordbayern konnte das Training auf Basis von 1,2 Millionen Trainings- und 300.000 Testkacheln im August 2020 begonnen werden. Für Südbayern wurden für den Trainingslauf 2021 1,74 Millionen Trainingskacheln und 435.000 Testkacheln bereitgestellt. Die Referenzdaten Nordbayerns und Südbayerns wurden bei den KI-Anwendungen nicht gemischt.

Bei der Projektion einer einzelnen Trainingskachel in die Entscheidungsebene des Netzwerks wurden beim FC-DenseNet-Ansatz bei der Vorwärtspropagierung 27,7 GFlops Rechenoperationen ausgeführt. Da eine GPU des ersten KI-Servers dafür ca. 4 Sekunden Rechenzeit benötigt, betrug die Berechnungsdauer für Südbayern auf dem ersten KI-Server bei 102 Epochen insgesamt 106,5 Tage, also mehr als vier Monate. Der hohe zeitliche Aufwand und die mit 127 PetaFlops (= $1{,}27 * 10^{17}$ Flops) enorme Anzahl an GPU-Rechenoperationen zum Training des Netzes von Südbayern wurden auch bei der im Projekt anfänglich zur Verfügung stehenden schwachen Hardware aufgewendet, um bei den DenseNet-Blöcken, die die Bausteine des DenseNet-Ansatzes darstellen, eine hohe Parameter-Effizienz und damit eine gute Erkennungsrate zu erreichen.

Mit dem Einsatz von A100-GPUs für die bei der Baufallerkundung Südbayerns zu leistenden 127 PetaFlops wird beim LDBV eine Rechentechnologie für die KI auf- und ausgebaut, die sich durch Performance wegen konsequenter Verlagerung vieler Rechenoperationen von CPUs auf leistungsstarke GPUs enorm fortschrittlich gegenüber dem im Jahr 2018 ebenfalls hohe Rechenkapazität erfordernden System zeigt, das für die Ausgleichungen bei der UTM-Umstellung des Liegenschaftskatasters (Glock et al. 2019) konzipiert wurde. Ein allein auf die Rechenkapazität abzielender Blick auf die vorliegende Hardware beider Projekte, die für die Rechenläufe Ende 2018 und Mitte 2021 im Abstand von gut 2½ Jahren am LDBV zur Verfügung stand, erlaubt die Aussage, dass mit der neuen, modernen Rechnerarchitektur eine Steigerung der Rechenkapazität um den Faktor 30 von 4,1 GFlops (für die Ausgleichungen zur UTM-Umstellung des Liegenschaftskatasters) auf 137 GFlops (für die Baufallerkundung auf dem neuen KI-Server 2) erreicht worden ist.

Für die Lernrate des Netzwerks wurde auf dem ersten KI-Server die Lernrate α auf den Wert $\alpha = 0{,}000001$ festgelegt und für die Anzahl der Bilder pro Optimierungsschritt, die sogenannte Batch Size, die Anzahl 5 gewählt. Mit diesen Parametern zeigte sich für den ersten KI-Server

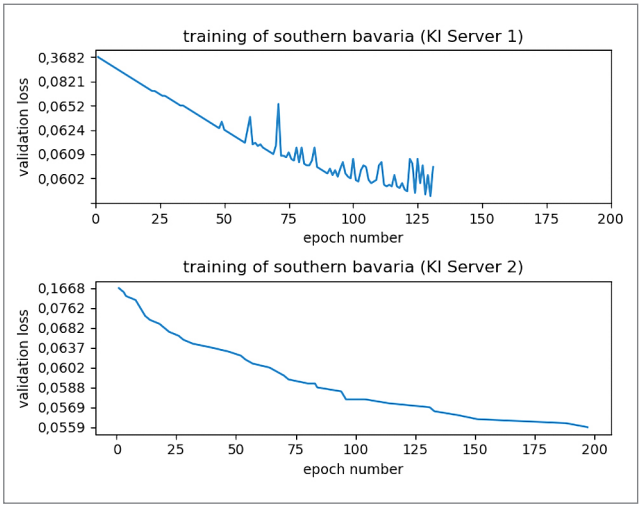

Abb. 6: Verlauf des Gesamtverlusts bei dem Validierungsdatensatz (Betrag *c* in Gleichung 2) für Südbayern bis Epoche 125 bzw. 200 berechnet auf dem KI-Server 1 mit einer Batch Size von 5 und einer Lernrate $\alpha = 0{,}000001$ und auf dem KI-Server 2 mit höherer Batch Size 16 und höheren Lernrate $\alpha = 0{,}00001$.

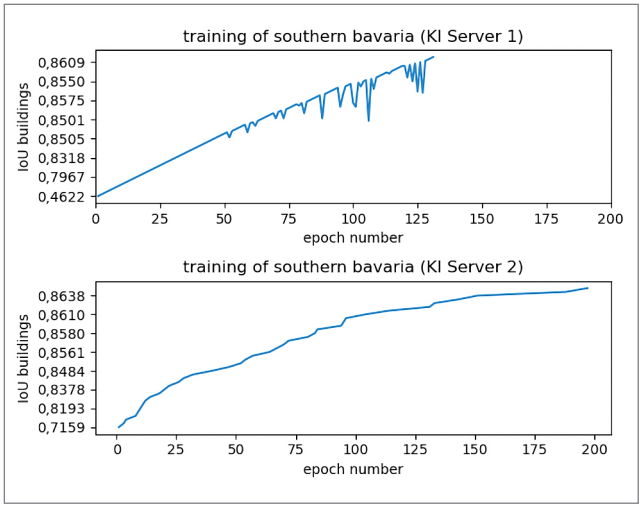

Abb. 7: Verlauf des IoU-Genauigkeitsmaßes für Südbayern

eine Auslastung der beiden GPUs im Mittel[4] von lediglich ca. 22 %.

Untersuchungen im Jahr 2021 für die Testregion München ergaben auf dem zweiten KI-Rechner mit 8 A100-GPUs sowie mit der neu gewählten Batch Size von 16 und einer höheren Lernrate von $\alpha = 0{,}00001$, dass beim 40 GByte großen Speicher für jede der A100-GPUs eine Auslastung von 73 % erreicht wurde. Durch die deutlich höhere Performance der GPUs und ihre verbesserte Speicherauslastung verkürzte sich die Datenprozessierung der Testregion München von 5,4 Tagen auf 13 Stunden. Damit ist der zweite KI-Rechner um den Faktor 10 schneller als der erste Server.

Eine erneute Prozessierung der südbayerischen Daten auf dem zweiten KI-Rechner bestätigte die drastische Verkürzung der Rechenzeit um 89,86 %: von drei Monaten (106,6 Tage) auf 10,8 Tage.

Neben einer hardwarebezogenen Performancesteigerung sind bei der Prozessierung der Daten für Südbayern auf den zwei parallel eingesetzten KI-Servern die Einflüsse der Parameteränderungen für die Größe der Batch Size und den Wert der Lernrate in Abb. 6 deutlich sichtbar. Der Kurvenverlauf auf dem KI-Server 1 mit der Batch Size 5 und der Lernrate $\alpha = 0{,}000001$ zeigt ab der 50. Epoche deutliche Ausschläge. Dieses Verhalten zeigt sich nicht beim KI-Server 2, da dort eine höhere Batch Size von 16 und eine um eine 10-er Potenz höhere Lernrate von $\alpha = 0{,}00001$ gewählt wurde. Dennoch zeigen beide Kurvenverläufe fallende Tendenz mit einem ähnlich guten Konvergenzverhalten.

Im Vergleich zu dem für Nordbayern erzielten Gesamtverlust von 0,071 nach 125 Epochen (vgl. Abb. 3 unten) ist die Konvergenz für Südbayern in Abb. 6 deutlich besser und der stetig degressive Kurvenverlauf bleibt selbst bis zur 200. Epoche erhalten.

Analog stellt sich im Genauigkeitsmaß IoU durch Wahl einer größeren Batch Size und höheren Lernrate das bessere Konvergenzverhalten mit einem glatteren Kurvenverlauf dar. Für Südbayern ergibt sich nach 130 Epochen ein IoU von 0,863 (vgl. Abb. 7) gegenüber einem niedrigeren IoU von 0,835 für Nordbayern (vgl. Abb. 4).

3.3 PostGIS-Anbindung

Mittelfristig kann das bisher dateibasierte Verfahren abgelöst und der dateibasierte Datenlader durch einen Mechanismus ersetzt werden, der direkt auf eine PostGIS-Datenbank zugreift. Über Python ist es möglich, die Patches aus einer Datenbank zu laden und der KI-Software für das Training bereitzustellen.

Für die Ausgangsdaten wie TrueDOP, DFK, DGM und bDOM wurden mittels Python rasterbasierte Importroutinen über GDAL nach PostGIS entwickelt, während die Differenzprodukte nDOM und tDOM mit den LAStools[5] effizient prozessiert wurden. Zusätzlich war es notwendig, das im ASCII-Grid vorliegende DGM in ein GeoTIFF umzuwandeln, unter Beachtung eines notwendigen Versatzes für die Höhe auf die Pixelmitte (derzeit bezieht sich die Höhe des DGM bundeseinheitlich auf die Position der linken unteren Pixelecke anstelle der notwendigen Pixelmitte). Zur Berechnung des nDOM haben damit alle Datenbestände den gleichen Bezugspunkt. Der testweise für Nordbayern bestimmte Zeitaufwand für den Datenimport betrug auf einer provisorisch aufgebauten Datenbank-Plattform etwas mehr als eine Woche. Für das Jahr 2022 ist der Aufbau eines produktiven Datenbankservers geplant.

4 Bei der Batch Size von 5 belegt der FC-DenseNet-Ansatz (v. a. der Forward- und Backward-Tracking-Graph mit u. a. über Filterungen definierten Tensoren, die automatisch differenziert werden) auf GPU 1 6058 MB von 24.200 MB (25,03 %) und auf GPU 2 4696 MB von 24.200 MB (19,40 %).

5 LAStools der Fa. rapidlasso GmbH

4 Ergebnisse der semantischen Segmentierung

4.1 Farb- und Intensitätsunterschiede zwischen den Befliegungslosen

Für eine automatisierte Klassifikation von Objekten in Rasterbildern sind stets Trainingsgebiete mit möglichst repräsentativen Eigenschaften der zu klassifizierenden Objekte erforderlich. Die notwendigen Trainingsgebiete für die Klassifizierung von Gebäudedächern in einem TrueDOP können in einfacher Weise über das Verschneiden der Gebäudegrundrisse aus der DFK mit dem TrueDOP gewonnen werden. Nachdem das Training des KI-Systems für eine semantische Detektion von Gebäuden aus Luftbildern der Bayernbefliegung mit ca. 100 Epochen je nach zur Verfügung stehender Hardware sehr zeitaufwendig sein kann, stellen sich folgende Fragen:

- Sollte das für Nordbayern trainierte System auf die Prozessierung der Luftbilder für Südbayern übertragen werden, ohne das KI-System neu zu trainieren?
- Ist das für Nordbayern trainierte System um Trainingsdaten aus Südbayern zu erweitern?
- Oder sollten die Luftbilder der Bayernbefliegung für Nord- und Südbayern getrennt voneinander trainiert und prozessiert werden?
- Kann eine Steigerung der Erkennungsrate erzielt werden, wenn die trainierten Systeme nach jeder Neubefliegung weiter trainiert werden?

Frühere Untersuchungen haben gezeigt, dass die höchste Erkennungsrate erzielt wird, wenn das KI-System auf diejenigen Bereiche beschränkt wird, für die entsprechende Trainingsdaten vorliegen. Offen bleibt jedoch, ob lediglich die aktuellsten Daten der Bayernbefliegung genutzt werden sollten. Dafür sprechen die nachfolgenden Überlegungen.

Für die Segmentierung oder Klassifikation von Objekten aus Luftbildern ist zunächst ein möglichst unverändertes Bildmaterial notwendig, denn jeder Eingriff durch Bildverarbeitungsmethoden führt zu einer Veränderung der Ausgangsqualität und vielfach zum Verlust von Informationen. Auch in dem aus der Bayernbefliegung in einer Bodenauflösung von 20 cm abgeleiteten und aus Gründen der Prozessorleistung auf 40 cm neu abgetasteten TrueDOP werden die zugrundeliegenden Bilder des Bildflugs nur sehr behutsam radiometrisch bearbeitet, sodass an den Losgrenzen der Bildflüge sich radiometrische Unterschiede unterschiedlich stark abzeichnen und Fluglose im Mosaik der TrueDOP zu erkennen sind (Roschlaub, Krey, Möst 2020). Unveränderte Gebäudedächer erscheinen in den jeweiligen Befliegungsepochen sehr unterschiedlich. Zur Segmentierung von ein und demselben Gebäude tragen ältere Befliegungsepochen keinen Informationsgewinn bei, womit das Training des KI-Systems stets auf die aktuellsten Befliegungen des TrueDOP-Mosaiks beschränkt werden sollte. Ebenso sollten nur die für den Untersuchungsbereich repräsentativen Dächer von Gebäuden verwendet werden. Aus den vorgenannten Gründen sollten Nord- und Südbayern getrennt voneinander prozessiert werden.

4.2 Inferenz für Nord- und Südbayern

Im Rahmen einer erweiterten KI-basierten Detektion von Gebäuden sollen neben den im amtlichen Liegenschaftskataster fehlenden Neu- und Altbauten (siehe Roschlaub et al. 2020) zusätzlich Gebäudeaufstockungen identifiziert werden. Charakteristisch für eine Gebäudeaufstockung ist, dass die Gebäudegrundrisse bereits in der DFK enthalten sind, aber die Höhen sich im tDOM gegenüber früheren Befliegungen geändert haben. Zur semantischen Segmen-

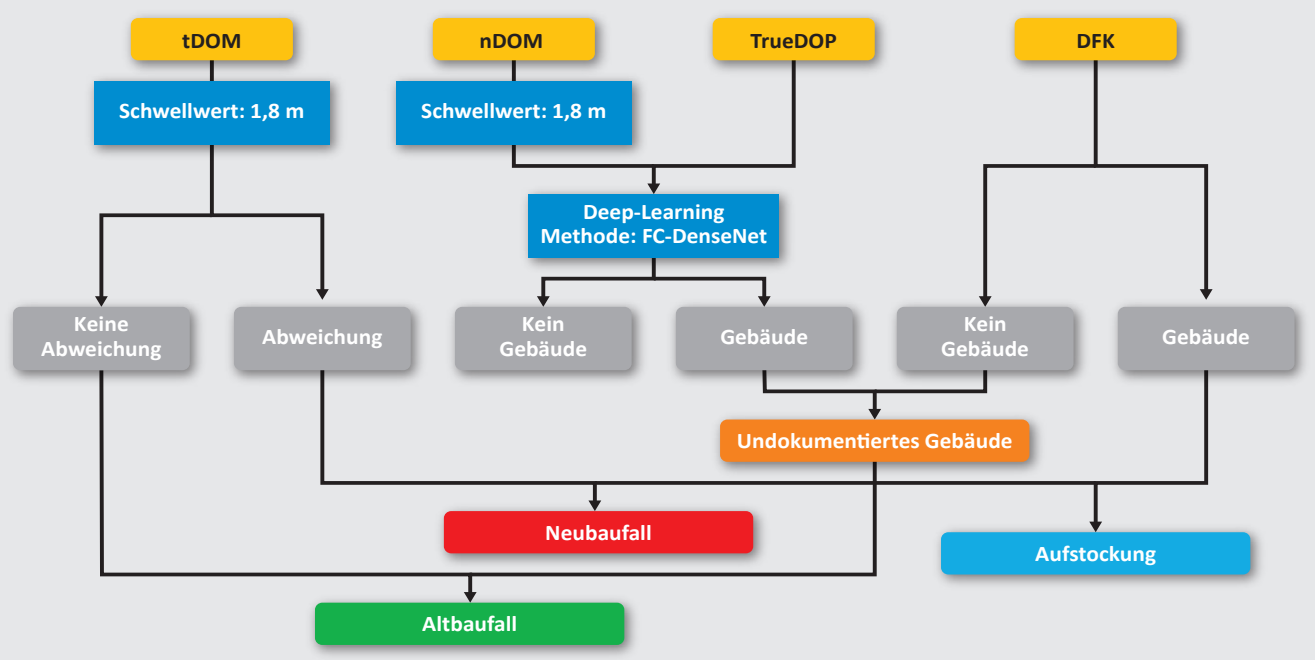

Abb. 8: KI-Entscheidungsbaum zur Segmentierung von Gebäudeneubauten, Altbauten und Aufstockungen

Abb. 9: Ergebnisse der Klassifikation (Inferenz) nach dem Training: Rot = Neubau, Grün = Altbau, Türkis = Aufstockungen

tierung ist eine Adaption des bisher verwendeten Entscheidungsbaums, wie in Abb. 8 dargestellt, erforderlich.

Mit dem bereits trainierten KI-System können im Rahmen der Inferenz sämtliche Gebäude im TrueDOP identifiziert und entsprechend des in Abb. 8 dargestellten Entscheidungsbaums klassifiziert werden (vgl. Abb. 9).

4.3 Auswertung der Daten am ADBV

Die Prozesskette zur Erkennung von undokumentierten Gebäuden wurde in 2020 am LDBV etabliert. Erste Ergebnisse für die Nordhälfte Bayerns wurden Ende 2020 produziert und testweise an Vermessungsämter übergeben.

Das Vermessungs- und Katastergesetz (VermKatG) in Bayern schreibt, wie auch in anderen Bundesländern, vor, die Liegenschaften des Staatsgebiets im Liegenschaftskataster zu beschreiben und darzustellen. Zu den Liegenschaften gehören die Grundstücke und Gebäude. Diese sind durch Fortführung auf dem Laufenden zu halten. Der gesetzliche Auftrag umfasst u. a., Veränderungen im Bestand der Gebäude durch Katastervermessungen zu erfassen. Diese werden entweder von Amtswegen durch die Ämter für Digitalisierung, Breitband und Vermessung (ÄDBV) oder auf Antrag durch private Büros durchgeführt. Zu den Veränderungen im Bestand der Gebäude gehören Neubauten, Veränderungen am Umfang des Grundrisses bestehender Gebäude, Abbrüche und die Zerstörung von Gebäuden. Aus Kostengründen sind jedoch nicht alle Gebäudeveränderungen in Bayern einmessungs- und gebührenpflichtig.

Die Mindestgrundfläche einmessungspflichtiger Gebäude beträgt grundsätzlich 13 m^2; bei Gebäudeanbauten 5 m^2. Für sonstige Gebäude mit seitlich offener Bauweise beträgt die Mindestgröße hingegen 35 m^2. Unter letztere fallen überwiegend Überdachungen und Carports. Daraus folgt jedoch, dass ein von der KI erkannter Baufall nicht automatisch zu einer Einmessungspflicht im Liegenschaftskataster führen muss. Dies impliziert eine differenzierte Betrachtung des von der KI erkannten Hinweises einer Gebäudeveränderung.

Fünf Jahre nach Abschluss der Baumaßnahme entfällt zwar die Gebührenpflicht für den Eigentümer, nicht aber die Einmessungspflicht für die ÄDBV. Die mit KI detektierten Altbaufälle beziehen sich auf ein tDOM, das im vorliegenden Fall für Nordbayern eine Zeitspanne von drei Jahren und für Südbayern von zwei Jahren berücksichtigt. Demgegenüber wird im Liegenschaftskataster ein Gebäude erst als Altbau eingestuft, wenn die Gebührenpflicht entfällt. Die ÄDBV sind daher gehalten, den Gebäudebestand fortlaufend aktuell zu halten und im Einzelfall die Einmessungs- und Gebührenpflicht zu überprüfen. Neben den Mitteilungen der Baugenehmigungsbehörden nutzen die ÄDBV bisher die DOPs aus der Bayernbefliegung. Der visuelle Vergleich zwischen Luftbild und Gebäudebestand ist jedoch aufwendig und mit Unsicherheiten behaftet, da die manuelle Sichtkontrolle vom Bearbeiter ausschließlich im 2D-Geoinformationssystem vorgenommen wird. Im Einzelfall werden Gebäudeveränderungen durch Erkundung vor Ort verifiziert.

Die erste Überprüfung der mittels KI erzielten Gebäudedetektion und der für Nordbayern in Neubauten und Altbauten klassifizierten Gebäudeveränderungen erfolgte durch das Amt für Digitalisierung, Breitband und Vermessung Nürnberg. Der Amtsbezirk umfasst die Großstädte Nürnberg und Fürth sowie acht Gemeinden des Landkreises Nürnberger Land mit insgesamt 96 Gemarkungen. Zur Baufallerkundung in den Wintermonaten 2020/21 wurden die KI-Ergebnisse in 51 Gemarkungen genutzt.

Dem Bearbeiter stehen im Geoinformationssystem neben dem aktuellen Gebäudebestand des Liegenschaftskatasters auch Informationen zu Erkundungsergebnissen aus vorausgegangenen Erkundungen mittels DOP und vor Ort zur Verfügung. Auf diese Weise kann ein bereits bekannter, nicht einmessungspflichtiger Baufall, der von der KI als Gebäudeveränderung klassifiziert wurde, leicht als Nichtbaufall eingestuft werden. Beispielhaft seien hierzu Carports, Überdachungen, Dachüberstände oder Gartenhäuser genannt, für die aufgrund der o. g. Kriterien keine Einmessungspflicht besteht. Ein weiterer entscheidender Vorteil ist die Ortskenntnis der Bearbeiter.

Um die Nutzbarkeit der KI-Ergebnisse besser beurteilen zu können, wurden im Rahmen der Überprüfung nicht nur Baufälle aus den Daten extrahiert, sondern jedes von der KI erkannte Objekt durch den Bearbeiter klassifiziert. Es sei bemerkt, dass die KI teilweise mehr als ein Objekt pro Gebäude ausweist. Die Anzahl ist deshalb nicht mit der Anzahl der Gebäudeveränderungen gleichzusetzen. Für die in den 51 Gemarkungen insgesamt erkannten ca. 17.250 Objekte mutmaßlicher Gebäudeveränderungen wurden die in Tab. 1 aufgeführten Klassen jeweils mit Anzahl und Anteil an der Gesamtsumme vergeben.

In 150 Fällen lagen im Liegenschaftskataster bereits aktuellere Informationen einer vor kurzem, erst nach der Bayernbefliegung erfolgten Einmessung des Gebäudes vor. Die der KI zugrundeliegenden Oberflächendaten hatten hier einen älteren Stand als das Liegenschaftskataster. In 3500 Fällen handelt es sich um offensichtliche Erkennungsfehler der KI. Hier sind Böschungen, Bauwerke wie beispielsweise Tiefgarageneinfahrten, aber auch Brücken

Tab. 1: Quantifizierung der KI-Ergebnisse

Klasse	Anzahl	Anteil
Baufall	3.400	19 %
Nichtbaufall	10.200	60 %
Erkennungsfehler	3.500	20 %
Bereits eingemessen	150	1 %

und versiegelte Flächen unterschiedlicher Farbwerte, vor allem auf Parkplätzen zu verzeichnen, die fälschlicherweise als Gebäude eingestuft wurden. Dieser Anteil sollte in zukünftigen Prozessierungen mit verbesserter Kalibrierung deutlich geringer ausfallen.

Der Anteil der Nichtbaufälle beträgt 60 %, was jedoch plausibel ist, da hier vor allem Gartenhäuser, Überdachungen und Dachüberstände, wie oben erwähnt, richtigerweise von der KI erkannt wurden. Diese Objekte konnten unter Hinzunahme vorhandener Informationen und der Erfahrung der Bearbeiter leicht als Nichtbaufall klassifiziert werden. Da davon ausgegangen werden muss, dass diese Nichtbaufälle in zukünftigen Berechnungen erneut erkannt werden, wurden die Schwerpunkte der Flächen berechnet. Zukünftige KI-Daten können dann mit den Schwerpunkten verschnitten werden, sodass diese Objekte automatisch klassifizierbar sind. Bei 19 % aller Objekte wurde eine Gebäudeveränderung identifiziert, die eine weitere Überprüfung und ggf. erforderliche Gebäudeeinmessung nach sich zieht. In diesem Anteil sind die tatsächlichen Mehrwerte für das Liegenschaftskataster zu finden. Diese werden in der nachfolgenden Tabelle näher differenziert und erläutert.

Da ein Auftrag im Baufallerkundungssystem aus mehreren Gebäuden bestehen kann, die KI jedoch teilweise mehrere Objekte pro Gebäude identifiziert, wurden insgesamt 6365 Objekte aus den KI-Daten extrahiert und in das Baufallerkundungssystem übernommen (Tab. 2).

In 336 Fällen wurden Gebäudeveränderungen identifiziert, die unmittelbar einen Auftrag zur gebührenpflichtigen Einmessung auslösen (Baufall). Bei 1205 Fällen liegt die Fertigstellung bereits über der Fünfjahresfrist, was zu einer gebührenfreien Einmessung durch das Amt führt (Altbaufall). Diese Fälle sind oftmals in unzugänglichen Innenhöfen von komplexen Stadtquartieren oder Gewerbebetrieben zu finden. In der Klasse »vorläufig erkundet« sind sowohl bei den Baufällen als auch bei den Altbaufällen 301 bzw. 765 Gebäudeveränderungen erfasst worden, die eine Entscheidung zur Einmessungspflicht vor Ort erfordern. Gründe hierfür sind Grundrissgrößen im Grenzbereich der Mindestgrößen von 35, 13 oder 5 m^2 zur Übernahme in das Liegenschaftskataster. Bei 40 Fällen konnte ein Gebäudeabbruch erkannt werden, der eine Fortführung des Liegenschaftskatasters ohne Außendienst auslöst. Weiterhin wurden 3718 Nichtbaufälle in das Baufallerkundungssystem übernommen. Ihre hohe Anzahl liegt in der Erfassungsmethodik dieser Objekte begründet. Bisher

Tab. 2: Differenzierung der potenziellen Baufälle

Klasse	Anzahl	Anteil
Baufall	336	5,3 %
Altbaufall	1.205	18,9 %
Abbruch	40	0,6 %
Baufall vorläufig erkundet	301	4,7 %
Altbaufall vorläufig erkundet	765	12,0 %
Nichtbaufall	3.718	58,4 %

wurden »offensichtliche« Nichtbaufälle nicht erfasst. Um jedoch zukünftige Klassifizierungen weiter zu automatisieren, wurden diese ebenfalls vollständig zur Ergänzung des Bestandes überführt.

Insgesamt übertrafen die KI-Ergebnisse dieser ersten frühen Berechnung bereits die Erwartungen. Für das ADBV Nürnberg ist der Mehrwert als zusammenfassendes Fazit deutlich gegeben:

- Die mittels KI erkannten Gebäudeveränderungen müssen auch weiterhin einer Überprüfung und Beurteilung durch das ADBV unterzogen werden. Dies schließt die teilweise Vor-Ort-Prüfung mit ein. Der große Mehrwert besteht jedoch in der gezielten Markierung einzelner Objekte, die direkt geprüft werden können, ohne das gesamte Gebiet durchforsten zu müssen.
- Die Einzelsichtung der ausschließlich von der KI erkannten Gebäudeveränderungen anstelle der bisherigen flächendeckenden Sichtkontrolle des DOP minimiert einerseits den Aufwand und beseitigt andererseits Unsicherheiten, Gebäudeveränderungen bei der Bearbeitung zu übersehen.
- Bei Gebäudekomplexen und stark verdichteter Bauweise findet die KI Gebäudeveränderungen zuverlässiger als der Bearbeiter.
- Werden die KI-Ergebnisse aus vorausgegangenen Berechnungen klassifiziert, können diese bei zukünftigen Auswertungen berücksichtigt werden. Ferner sollten zur Differenzierung zwischen Baufall und Nichtbaufall vorhandene Informationen aus vorausgegangenen Vor-Ort-Erkundungen einfließen. Werden sämtliche Nichtbaufälle systematisch erfasst, minimiert sich der Sichtungsaufwand zukünftig nur noch auf den tatsächlichen Bestand neu hinzugekommener Gebäudeveränderungen.
- Die KI liefert auch solche Gebäudeveränderungen, die von den Mitteilungen der Baugenehmigungsbehörden aufgrund der Genehmigungsfreiheit nicht übermittelt werden.
- Die Fehlerquote bei der Erkennung sollte bei zukünftigen Berechnungen aufgrund verbesserter Kalibrierung deutlich niedriger ausfallen.

5 Schluss

In einem Benchmark der KI-Modelle zur semantischen Segmentierung in urbanen Gebieten (https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes) schneidet das im Jahr 2020 entwickelte HRNet-OCR (Hierarchical Multi-Scale Attention) im Ranking mit einer IoU von 85,1 % am besten ab, während ein klassisches Fully Convolutional Network (FCN) aus dem Jahr 2016 eine Erkennungsrate von lediglich 65,3 % erzielte und auf Platz 84 rangiert. Demgegenüber wurde mit dem von der Bayerischen Vermessungsverwaltung favorisierten FC-DenseNet, wie in Abschnitt 2.4 gezeigt, für Nordbayern nach 100 Epochen eine Erkennungsrate von 83 % erreicht. Dieser Wert liegt sehr nahe an dem Bestwert des HRNet-OCR und ist deutlich höher als die ausgewiesenen 65,3 % für das FCN. Das FC-DenseNet läge in den Benchmark eingereiht auf den Plätzen 12 bis 15. Neuere Entwicklungen gehen bereits in die Richtung der Optimierung der GPU-Auslastung und der Laufzeit des Netzwerkes, beispielsweise durch ein automatisiertes Engineering der Netzwerkarchitektur mittels neuronaler Architektursuche (Neural Architecture Search, NAS). Nachdem die Qualität der Trainingsdaten im vorliegenden Fall der Baufallerkundung mit 9,4 Mio. Gebäuden in Bayern kaum zu steigern ist, bleibt schlussendlich bei den aktuell rasant steigenden Entwicklungen die Aufgabe für die Verwaltung, von Zeit zu Zeit den Einsatz neuerer Netzwerkarchitekturen zu prüfen, um die Erkennungsraten bei der semantischen Segmentierung beständig zu steigern. Ebenso darf die Rolle der Datenqualität nicht unterschätzt werden; die geringen Unterschiede in den Spitzenwerten der oben erwähnten Netze deuten darauf hin, dass hier schnell ein »Glass Ceiling« erreicht wird, bei dem die Architektur nur noch eine untergeordnete Rolle spielt.

In den letzten Jahren zeigt sich in der KI-Forschung oft, dass immer größere und leistungsfähigere Ansätze quasi automatisch bessere Ergebnisse liefern. Demgegenüber steht deren Bedarf an Hardware und Trainingsdaten, der für die meisten Nutzer nicht abbildbar ist. Es zeigt sich auf der anderen Seite, dass auch intelligente neue Architekturmodelle und Trainingsstrategien zu verbesserten Ergebnissen mit begrenzten Ressourcen führen können. Dementsprechend steht zukünftig nicht allein der stetig wachsende Ausbau immer leistungsstärkerer KI-Hardware mit immer schnelleren GPUs im Fokus, sondern eine ausgewogene Nutzung zwischen effizienten Netzwerkarchitekturen mit einem entsprechend linearen Ausbau leistungsstarker Hardware.

Für die ResNet-Architektur sind beispielsweise für den Bereich der Modelldefinition in Verbindung mit schnellerem Trainieren aktuelle Weiterentwicklungen zu beobachten, wie mit den aktuellen EfficientNet-Ansätzen (Tan und Le 2021). Bereits auf die Optimierung des EfficientNet-Ansatzes zielende ResNet-Fortentwicklungen (Bello et al. 2021) setzen Squeeze-and-Excitation-Layer (Hu et al. 2017) zur Erhöhung der Erkennungsrate ein. Bei der ResNet-Architektur ist von weiteren großen Fortschritten auszugehen, es kann sogar aktuell von der Renaissance der ResNet-Ansätze gesprochen werden.

Insgesamt bleibt festzustellen, dass der Aufbau einer KI-Infrastruktur bekannte Themen wie Rasterdatenverarbeitung und Web-Dienste einschließt. Die Detektion von Gebäudeveränderungen, welche früher mit hohem Personaleinsatz verbunden war, kann heute weitestgehend automatisiert werden.

Der Einsatz von KI stellt neuartige und veränderte Anforderungen an das Personal. Insbesondere sind vertiefte Python- und Linux-Kenntnisse erforderlich. Zusätzlich sind hohe Investitionsmittel für die Beschaffung einer leistungsstarken KI-Infrastruktur bereitzustellen. Erst bei einer Rechenzeit von wenigen Tagen für die Prozessierung landesweiter Daten ist der Einsatz von KI attraktiv. Eine Übertragung des vorgestellten Verfahrens auf Satellitendaten ist vorstellbar.

Literatur

AdV (2017): AK GT Beschluss 30/2 zur Überführung des ATKIS-DOP20 in die Qualitätsstufe TrueDOP. Apr 2017 in Saarlouis, Deutschland.

Bello, I., Fedus, W., Du, X., Cubuk, E., Srinivas, A., Lin, T.-Y., Shlens, J., Zoph, B. (2021): Revisiting ResNets: Improved Training and Scaling Strategies. arXiv: 2103.07579, 1–11.

Casanova, A., Cucurull, G., Drozdzal, Ml., Romero, A., Bengio, Y. (2018): On the iterative refinement of densely connected representation levels for semantic segmentation. arXiv: 1804.11332, 1–12.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018): Encoderdecoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), 801–818.

Chen, Y., Li, J., Jin, H. X., Yan, S., Feng, J. (2017): Dual Path Networks. arXiv: 1701.01629, 1–11.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C. (2016): The Importance of Skip Connections in Biomedical Image Segmentation. arXiv: 1608.05117, 1–9.

Glock, C., Bauer, R., Wunderlich, T., Pail, R., Bletzinger, K.-U. (2019): Das Ortra-Verfahren für die Überführung des Liegenschaftskatasters nach ETRS89/UTM in Bayern. zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement, Heft 1/2019, 144 Jg., S. 25–40. DOI: 10.12902/zfv-0237-2018.

He, K., Zhang, X., Ren, S., Sun, J. (2015): Deep residual learning for image recognition. arXiv: 1512.03385, 1–12.

Hu, J., Shen, J., Albanie, S., Sun, G., Wu, E. (2017): Squeeze-and-Excitation Networks. arXiv:1709.01507, 1–13.

Huang, G., Liu, Z., Maaten, L. v. d., Weinberg, K. Q. (2016): Densely Connected Convolutional Networks. arXiv: 1608.06993, 1–9.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio Y. (2017): The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: 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): Detection of Undocumented Building Constructions from Official Geodata Using a Convolutional Neural Network. Remote Sensing, 2020, 12, 3537. DOI: 10.3390/rs12213537.

Li, Q., Shi, Y., Huang, X., Zhu, X. X. (2020): Building footprint generation by integrating convolution neural network with feature pairwise conditional random field (FPCRF). IEEE Transactions on Geoscience and Remote Sensing, 58(11), 2020, 7502–7519.

Li, X., Yao, X., Fang, Y. (2018): Building-A-Nets: Robust Building Extraction from High-Resolution Remote Sensing Images with Adversarial Networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 3680–3687.

Long, J., Shelhamer, E., Darrell, T. (2015): Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 3431–3440.

Ronneberger, O., Fischer, P., Brox, T. (2015): U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer. 234–241.

Roschlaub, R., Krey, T., Möst, K. (2020): Automated Classification of Building Roofs for the Updating of 3D Building Models Using Heuristic Methods. PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science, Vol. 88, Number 1 (2020), 85–97. DOI: 10.1007/s41064-020-00099-9.

Roschlaub, R., Li, Q., Auer, S., Shi, Y., Möst, K., Glock, C., Schmitt, M., Shi, Y., Zhu, X. (2020): KI-basierte Segmentierung von Gebäuden mittels Deep Learning und amtlichen Geodaten zur Baufallerkundung. zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement, Heft 3/2020, 145 Jg., S. 180–189. DOI: 10.12902/zfv-0299-2020.

Shi, Y., Li, Q., Zhu, X. X. (2020): Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. ISPRS Journal of Photogrammetry and Remote Sensing, 159, 2020, 184–197.

Tan, M., Le, Q. V. (2021): EfficientNet V2: Smaller Models and Faster Training. arXiv:2104.00298, 1–11.

Tasar, O., Happy, S. L., Tarabalka, Y., Alliez, P. (2020): ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. IEEE Transactions on Geoscience and Remote Sensing, 58(10), 7178–7193.

Zhang, C., Benz, P., Argaw, D. M., Lee, S., Kim, J., Rameau, F., Bazin, J.-C., Kweon, I. S. (2020): ResNet or DenseNet? Introducing Dense Shortcuts to ResNet., 1–10.

Kontakt

Dr.-Ing. Robert Roschlaub | Dipl.-Ing. (FH) Karin Möst | Dr.-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

Dipl.-Ing. Frank Hümmer
Amt für Digitalisierung, Breitband und Vermessung, Nürnberg
Flaschenhofstraße 59, 90402 Nürnberg
frank.huemmer@adbv-n.bayern.de

Dr.-Ing. Anna Kruspe | M. Sc. Qingyu Li
Professur für Data Science in Earth Observation, TUM School of Engineering and Design, Technische Universität München
Arcisstraße 21, 80333 München
anna.kruspe@tum.de | qingyu.li@tum.de

Dr.-Ing. Stefan Auer
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung, Photogrammetrie und Bildanalyse
Oberpfaffenhofen
stefan.auer@dlr.de

Prof. Dr.-Ing. habil. Xiao Xiang Zhu
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung, EO Data Science, Oberpfaffenhofen
xiaoxiang.zhu@dlr.de
und
Professur für Data Science in Earth Observation, TUM School of Engineering and Design, Technische Universität München
Arcisstraße 21, 80333 München
xiaoxiang.zhu@tum.de

Dieser Beitrag ist auch digital verfügbar unter www.geodaesie.info.