

Automatisierte Klassifizierung der Landnutzung: Deep Learning basierter Ansatz für die Arealstatistik der Schweiz

Das Bundesamt für Statistik (BFS) erstellt periodisch auf der Grundlage von Luftbildern eine schweizweit einheitliche Arealstatistik zur Klassifizierung der Bodennutzung und -bedeckung. Die Aufbereitung der Arealstatistik erfolgt bisher in einem zeit- und personalintensiven Prozess. In zwei am Institut Geomatik der Fachhochschule Nordwestschweiz (FHNW) durchgeführten Bachelorarbeiten wurde deshalb untersucht, inwieweit sich Möglichkeiten der künstlichen Intelligenz (KI) auf der Basis von Deep Learning (DL) eignen, eine Automatisierung der Klassifizierungsaufgabe herbeizuführen. Erste Ergebnisse dieser explorativen Untersuchung zeigen auf, dass sich durch eine Basiskonfigurierung neuronaler DL Netzwerke vielversprechende Ergebnisse erzielen lassen, die auf das hohe Anwendungspotenzial von KI für die Arealstatistik hinweisen.

L'Office fédéral de la statistique (OFS), sur la base d'images aériennes, établit périodiquement pour l'ensemble du territoire suisse une statistique de la superficie permettant la classification de l'utilisation et de l'occupation du sol. Le traitement de la statistique de la superficie a lieu jusqu'à présent par un processus exigeant du temps et du personnel. Dans le cadre de deux thèses bachelor au sein de la Haute école spécialisée du nord-ouest de la Suisse (FHNW) on a donc examiné dans quelle mesure de nouvelles possibilités de l'intelligence artificielle sur la base de deep learning (DL) se prêtent à la réalisation automatisée de la tâche de classification. Les premiers résultats de cette évaluation explorative montrent que par une configuration de base de réseaux neuronaux DL on peut s'attendre à des résultats prometteurs et que ceux-ci indiquent un haut potentiel de procédés automatisés pour la statistique de la superficie.

L'Ufficio federale di statistica (UFT) allestisce periodicamente, sulla base di riprese aeree, una statistica di superficie unitaria svizzera per classificare l'utilizzo e la copertura del suolo. Finora l'approntamento della statistica di superficie è risultato essere un processo complesso in termini di tempo e manodopera. In due tesi di Bachelor realizzate presso la Scuola universitaria superiore della Svizzera nordoccidentale (FHNW) si sono analizzate le possibilità di ricorrere all'intelligenza artificiale (IA), improntata sul Deep Learning (DL), per arrivare a un'automatizzazione dei compiti di classificazione. I primi risultati di questi test esplorativi dimostrano che sussiste un potenziale promettente nella configurazione di base delle reti neurali DL che presentano un elevato potenziale di applicazione dell'IA per la statistica di superficie.

D. Jordan, N. Lack, S. Hochuli,
A. F. Meyer, M. Schär

Einleitung

Die in einem Abstand von aktuell sechs Jahren erstellte Arealstatistik des BFS (Neuchâtel) dient zur Erfassung von Veränderungen der Landnutzung in der Schweiz und stellt die Datengrundlage für viele öffentliche Planungsentscheidungen dar. Die Erzeugung der Arealstatistik erfolgt auf der Basis flächendeckend vorliegender Luftbilder, indem Expertinnen und Experten diese in einem Stichprobenraster von 100 m Maschenweite den insgesamt 72 Bodennutzungs- und Bodenbedeckungsklassen zuordnen. Für die hohe erreichte Erfassungsgenauigkeit des aufwändigen Arbeitsprozesses werden für jede Erhebungsperiode ca. 4.1 Mio. Stichprobenpunkte, die durch je eine Referenzfläche von 50 m × 50 m umgeben sind, klassifiziert und validiert. Ziel ist es, den Aufwand für die Erstellung der Arealstatistik ohne Einbussen an die Zuverlässigkeit zu optimieren, indem automatisierte Klassifikationsmethoden auf KI Basis zukünftig eingesetzt werden.

Grundsätzlich dient KI der algorithmischen Nachbildung von Entscheidungsprozessen, mit dem Ziel, angelernte Erfahrung prädiktiv für neu gewonnene Daten anzuwenden. Für die Umsetzung bieten sich eine Vielzahl von Methoden des maschinellen Lernens an, deren Eignung spezifisch zur Aufgabenstellung zu beurteilen ist. Für die Arealstatistik soll die Aufgabe der Maschine darin bestehen, jedes ihr präsentierte Luftbild selbstständig einer der 72 Klassen zuzuordnen, indem erlernte relevante Merkmale regelbasiert für die Klassifikation genutzt werden. Dabei soll die Maschine nach erfolgtem Trainingsprozess mit bekannten Referenzdaten (Luftbilder mit Klassenzuordnungen aus abgeschlossenen Erhebungen) zukünftig erhobene Daten automatisch und möglichst fehlerfrei klassifizieren. Mit Hilfe von künstlichen neuronalen DL Netzwerken soll die typische *Big Data* Fragestellung adressiert

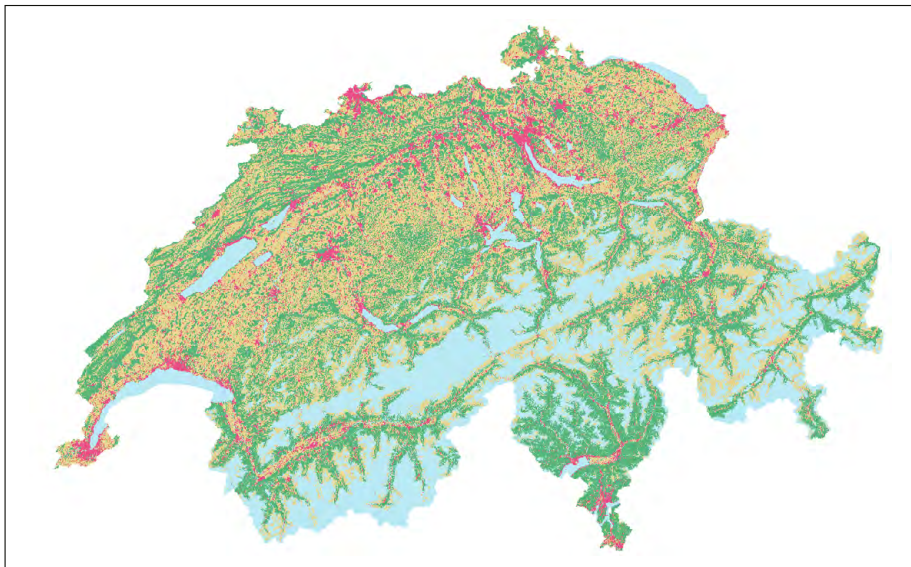


Abb. 1: Arealstatistik in der Erfassungsperiode 2004/09 mit Darstellung folgender aggregierter Klassen: Siedlungsflächen (rot, 8% Anteil der Gesamtfläche), Landwirtschaftsflächen (gelb, 36%), bestockte Flächen (grün, 31%) und unproduktive Flächen (türkis, 25%) [2].

und die Eignung von KI für die Arealstatistik untersucht werden. DL hat sich in den letzten Jahren für viele Anwendungsbereiche als State-of-the-Art Technologie des maschinellen Lernens etabliert und stellt nicht zuletzt durch die enorme Leistungssteigerung von Grafikoperationen moderner Rechnersysteme einen leistungsfähigen Ansatz in der Bilderkennung dar, der insbesondere auch für die Fernerkundung genutzt wird [1].

Im Rahmen von zwei am Institut Geomatik FHNW durchgeführten Bachelorarbeiten (Markus Schär, 2017/Stefan Hochuli, 2018) wurde ein DL basierter Workflow bereitgestellt, welcher das Anwendungspotenzial der KI für die Arealstatistik des BFS darstellen soll. Vorliegender Bericht fasst die daraus gewonnen Erkenntnisse zusammen.

Methoden

Datengrundlage der Arealstatistik

Nachfolgende Untersuchungen beruhen auf der Arealstatistik des BFS für die Erfassungsperiode von 2004/09 [2], für welche Dreikanal RGB Orthofotos *swiss-image* des Bundesamtes für Landestopografie (Wabern) mit einer Bodenauflösung von 50 cm aus Befliegungen im

Frühjahr oder Herbst verwendet wurden [3]. Die aus Luftbildern bereitgestellten 25a grossen Referenzflächen dienten dem BFS als Grundlage für die Zuordnung eines jeden zentral liegenden Stichprobenpunktes zu einer der 72 Bodennutzungs- und Bodenbedeckungsklassen. Ein wichtiges Merkmal der Arealstatistik ist, dass es sich um eine zentrumsbezogene Klassifikation handelt: Massgebend für die Klassenzugehörigkeit ist die Lage der

Stichprobenpunkte, wobei Kontextinformationen der umgebenden Referenzflächen mitberücksichtigt werden. Abbildung 1 zeigt die Arealstatistik 2004/09 bezüglich Bodennutzung in vier aggregierten Hauptbereichen (zusammengefasste Klassen) auf.

DL Netzwerke zur Automatisierung der Arealstatistik

Erste Realisierungen künstlicher neuronaler Netze entstanden in den 1960er-Jahren, deren Grundkonzepte dem neurophysiologischen Vorbild des menschlichen Gehirns nachempfunden sind: Einzelne biologische Nervenzellen (Neuronen) des Gehirns sind dafür ausgelegt, Reize tausender vorangehender Neuronen zu empfangen, diese zu integrieren und selektiv an nachgeschaltete Neuronen weiterzuleiten. Sie entscheiden für jede Eingangskonfiguration über die Weiterführung einer neuronalen Information. Erst in einem hochkomplexen Netzwerk aus ca. 100 Milliarden einzelner Neuronen ist das Gehirn in der Lage, spezifische Aufgaben wie die Prozessierung visueller Informationen bis hin zur Klassifizierung betrachteter Bilder vorzunehmen. Dabei ist das Gehirn funktionell hierarchisch und anatomisch in Schichten organisiert, indem tiefere funktionelle Einheiten grundlegende Funktionen wie

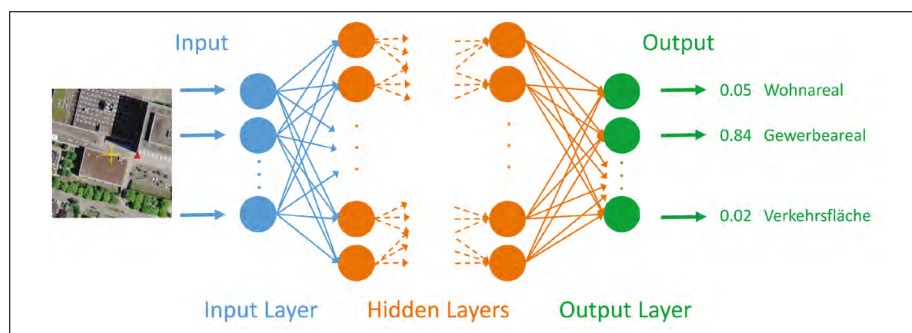


Abb. 2: Schichtweise Architektur eines DL Netzwerkes mit den Neuronen (ausgefüllte Kreise) des *Input Layers*, der *Hidden Layers* und des *Output Layers*, welche zwischen benachbarten Schichten zumindest teilweise und unidirektional verbunden sind. Angewandt auf die Arealstatistik wird für jeden zu klassifizierenden Stichprobenpunkt die zugehörige Referenzfläche durch ihre RGB Pixelwerte den Neuronen des Input Layers präsentiert und im DL Netzwerk verarbeitet. Schliesslich wird im Output Layer eine Klassifikationsentscheidung durch die Angabe einer Wahrscheinlichkeit für die Klassenzugehörigkeit angezeigt.

die Aufnahme und Vorprozessierung sensorischer Reize  bernehmen, w hrend h here Einheiten f r kognitive Funktionen zust ndig sind [4].

Die technisch-mathematische Implementierung erfolgte auf Neuronenebene erstmals durch das *Perceptron*, welches als elektrischer Schaltkreis eine Gewichtung, Integration und eine Schwellwertbasierte Weiterleitung eingehender, als elektrische Spannungswerte modellierter Informationen nachbildet. Schliesslich wurde in den 1980er Jahren intensiv an konnektionistischen Modellen technischer neuronaler Netzwerke geforscht, deren Ansatz ist, dass viele einfache Recheneinheiten ein intelligentes Verhalten erreichen, wenn sie als k nstliche Neuronen in einem f r die Aufgabenstellung spezifisch organisierten und konfigurierten Netzwerk zusammenarbeiten [5]. Abbildung 2 zeigt im Kontext der Arealstatistik die grundlegende Architektur eines DL Netzwerks. Diese zeichnen sich als spezielle neuronale Netze durch ihre hohe Anzahl Schichten aus, wobei jeder Schicht spezifische Funktionen wie Filterungen zugeordnet werden.

Es stellt sich die Frage, wie ein DL Netzwerk die Klassifikationsentscheidung herbeif hren kann. Dazu wird das Netzwerk anhand einer m glichst grossen und hinsichtlich der Fragestellung repr sentativen Stichprobe *trainiert*. F r jeden Stichprobenpunkt des *Trainingsdatensatzes* gilt es, die zuvor initialisierten Parameter (Gewichtungen, Schwellwerte) der einzelnen Neuronen so einzustellen, dass ein minimaler Klassifikationsfehler resultiert. Die Gr sse des Trainingsdatensatzes und die hohe Anzahl der Einstellm glichkeiten innerhalb des Netzwerkes f hren jedoch dazu, dass diese Aufgabe nicht ohne Weiteres beherrschbar ist. In so genannten *Backpropagation-Algorithmen* erfolgt die Konfigurierung des Netzwerkes mittels iterativem Prozess, indem die Neuronen ausgehend von hinteren zu vorderen Schichten und anhand kleiner Stichproben aus den Trainingsdaten schrittweise hinsichtlich Klassifikationsfehler optimiert werden. Dabei kann ein *Overfitting* an die Trainingsda-

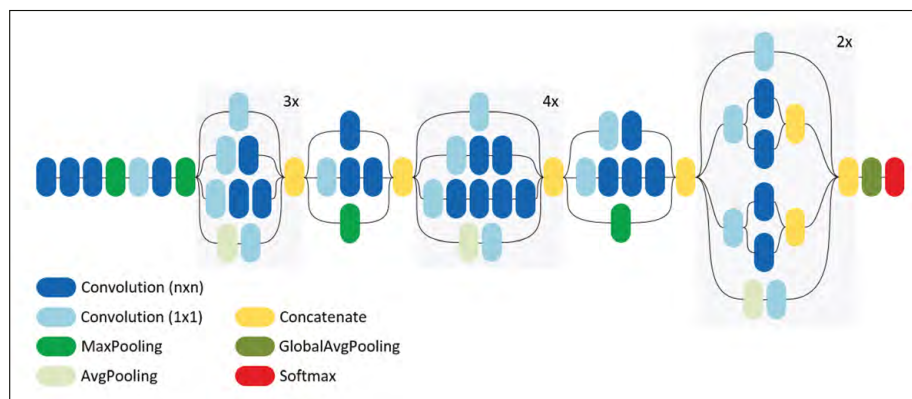


Abb. 3: Prinzipdarstellung der Architektur von Inception-v3 mit den zugeh rigen Schichten. Einzelne Bl cke werden mehrfach sequentiell wiederholt [7].

ten eintreffen: Das optimierte DL Netzwerk erkennt zuverl ssig alle Stichprobenpunkte der Trainingsdaten, jedoch ist die Klassifikation nicht f r neue, der Maschine unbekanntes Stichprobenpunkte generalisierbar bzw. pr diktiv einsetzbar. Aus diesem Grund beinhaltet das Training eines DL Netzwerkes immer eine parallel durchgef hrte *Validierung* anhand eines unabh ngigen *Validierungsdatensatzes*. Erst eine auf den Trainings- wie Validierungsdaten vergleichbare Klassifikationsleistung weist auf einen erfolgreichen und anwendbaren Lernprozess hin [6]. Die G te der Klassifikation wird schliesslich mit statistischen Metriken anhand eines *Testdatensatzes*, den der Algorithmus im Rahmen der Trainingsprozedur nicht gesehen hat, quantitativ beurteilt.

Zur Bildklassifikation sind *Convolutional Neural Networks (CNN)* als Implementierungsvariante der DL Netzwerke besonders geeignet. In solchen werden die Neuronen einer Schicht als Matrix angeordnet, wobei die Hidden Layers Faltungen (z. B. Hochpass- und Tiefpassfilter) sowie B ndelungen der Bildinformation herbeif hren. CNN erzielen mit grossen Trainingsdatens tzen und einer hohen Anzahl an Iterationen sehr gute Klassifikationsergebnisse, jedoch erfordern 20–60 Millionen konfigurierbare Parameter eine entsprechend hohe Rechenleistung und Speicherbedarf. Vortrainierte CNN steigern insofern die Effizienz, indem die Gewichtungen in den einzelnen Neuronen im Initialzustand nicht zuf llig, son-

dern bereits zur Bildklassifizierung optimiert vorliegen. Das CNN braucht anhand eigener Trainingsdaten und Klassenzuteilungen nur noch nachjustiert zu werden [6]. Abbildung 3 zeigt das weit verbreitete *Inception-v3* (Tensorflow, Google, Mountain View, Kalifornien, USA) als Umsetzung der CNN Architektur [7].

Die vorliegende Untersuchung verwendet Inception-v3 als DL Netzwerk zur Bereitstellung der Arealstatistik durch maschinelles Lernen. Der Softwarebasierte Workflow ist in drei Arbeitsprozessen gegliedert, der Datenaufbereitung, dem Training und dem Test, wobei die Implementierung in der Sprache *Python* (Version 3.6, Python Software Foundation, Wilmington, Delaware, USA) und mit Hilfe des High-Level *Application Programming Interface (API) Keras* [6] sowie dem von Google entwickelten DL Framework *TensorFlow* erfolgt:

- **Datenaufbereitung:** Das Stichprobenraster mit 100 m Maschenweite definiert die Lokalisierung der quadratisch aus den swissimage Orthofotos zugeschnittenen Referenzfl chen (50 m × 50 m) mit den zentralen Stichprobenpunkten. Die Referenzfl chen werden in einen Trainings- (70% der Daten), Validierungs- (10%) und Testdatensatz (20%) aufgeteilt, wobei die Aufteilung randomisiert  ber die gesamte Landesfl che unter Einhaltung vergleichbarer Klassenanteile und eines balancierten Umfangs pro Klasse vorgenommen wird. Um Letzteres zu erreichen, wurde der gesamte Datensatz verkleinert.

- **Training:** Die Referenzflächen der Trainings- und Validierungsdaten dienen als Input für den DL Algorithmus und werden im Rahmen des Trainingsprozesses zusammen mit der vom BFS erstellten Klassenzuordnung dem Netzwerk präsentiert [7]. Es erfolgt die Optimierung des vortrainierten Inception-v3 bezüglich des Klassifikationsparadigmas der Arealstatistik in unterschiedlichen Aggregationsstufen (4, 17 und 72 Klassen).
- **Test:** Dem trainierten DL Netzwerk wird der unabhängige Testdatensatz zur Prädiktion übergeben. Die Parametrisierung ist in diesem Schritt «eingefroren», so dass sich die erzielten Klassifikationsergebnisse mit den Zuordnungen des BFS vergleichen und als Konfusionsmatrizen bzw. daraus abgeleiteten summarischen Metriken darstellen lassen. In vorliegender Arbeit wird hierfür die *Genauigkeit* als Gütekriterium angegeben, welche durch das Verhältnis richtig klassifizierter Stichprobenpunkte zum gesamten Umfang der Testdaten definiert ist.

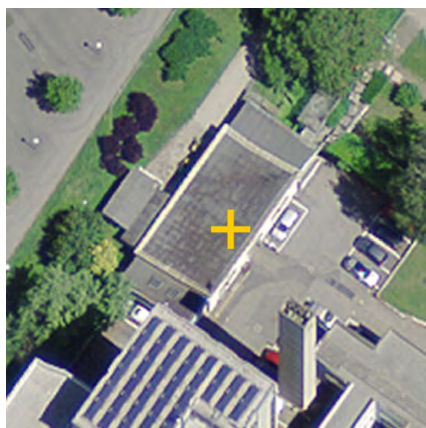


Abb. 4: Exemplarische Referenzflächen mit den Stichprobenpunkten für die Klassen *Gebäudeareal* (links) und *Industrie- und Gewerbeareal* (rechts).

ren des maschinellen Lernens ist prinzipiell die Lernaufgabe umso ambitionierter, je mehr Klassen zu unterscheiden sind. Umgekehrt stellt ein dichotomes Szenario mit der Unterscheidung zweier Klassen meist die einfachste Klassifikationsaufgabe dar. Die Beschaffenheit der Bilder, das heisst wie verschränkt oder disjunkt klassenrelevante Merkmale ausfallen, ist weiterhin mitentscheidend dafür, wie gut sich einzelne Klassen separieren lassen. Vor allem für die Klassifikationen von 72 und 17 Klassen lassen sich erhebliche Genauigkeitsdifferenzen feststellen, für die folgender Faktor primär verantwortlich ist: Stichprobenpunkte unterschiedlicher Klassen weisen teilweise vergleichbare RGB Signaturen auf und lassen sich in solchen Fällen kaum hinsichtlich ihrer Klassenzugehörigkeit unterscheiden, wie in Abbildung 4 exemplarisch illustriert ist. Beim BFS erfolgt die Zuordnung mit Hilfe

von Zusatzinformationen wie amtlichen Daten.

In vorliegender Untersuchung wurde durch die Generierung balancierter Klassengrössen eine Benachteiligung unterrepräsentierter Klassen, das sind solche mit einer vergleichsweise niedrigen Anzahl Stichprobenpunkte wie etwa *Baustellen*, vermieden. Dennoch fallen die erzielten Klassengenauigkeiten teilweise niedrig aus. Als Ursachen sind nebst der oben erwähnten überlappenden RGB Bildsignatur die zentrumsbasierte Klassifikation zu nennen: In gewissen Stichprobenpunkten reicht der Mittelpunkt des Bildes aus, um eine korrekte Klassifikation herbeizuführen, während andere Stichprobenpunkte die Hinzunahme einer Kontextinformation erfordern, beispielsweise zur Unterscheidung einer Strasse von einer Start- und Landebahn eines Flughafens. Für einen Algorithmus des maschinellen Lernens stellen

Ergebnisse und Diskussion

In Tabelle 1 sind die Ergebnisse des Inception-v3 Modells der Arealstatistik 2004/09 auf Grundlage der Testdaten in drei Aggregationsstufen (72 Klassen, 17 bzw. 4 aggregierte Klassen) zusammengefasst. Grundsätzlich wird eine Steigerung der Klassifikationsgüte mit zunehmender Aggregation beobachtet: Für ein Verfah-

Inception-v3	72 Klassen		17 aggregierte Klassen		4 aggregierte Klassen	
Gesamtgenauigkeit [%]	44.4		65.1		82.5	
höchste Klassengenauigkeiten [%]	stehende Gewässer	84.6	stehende Gewässer	93.6	Bestockte Flächen	88.8
	Rebbauflächen	83.9	Gletscher, Firn	86.6	Landwirtschaftsflächen	82.3
	Obstanlagen	80.0	Obst-, Reb- und Gartenbauflächen	76.4	Siedlungsflächen	78.0
	Autobahnen	77.1	Ackerland	75.7	unproduktive Flächen	76.3
niedrigste Klassengenauigkeiten [%]	Holzschläge	21.5	besondere Siedlungsflächen	46.8		
	Baustellen	17.1	Naturwiesen, Heimweiden	46.8		
	übrige Versorgungs-, Entsorgungsanlagen	15.2	unproduktive Vegetation	40.6		

Tab. 1: Klassifikationsgenauigkeiten des Modells Inception-v3 mit 72 Klassen sowie 17 und 4 aggregierten Klassen.

derartige Inkonsistenzen in der merkmals- und regelbasierten Entscheidungsfindung grosse Herausforderungen dar. Dennoch scheint das verwendete DL Netzwerk für eine grosse Anzahl von Stichprobenpunkten selbstständig Merkmale zu erlernen, die auf das Zentrum des Bildes hinweisen und die sich zielführend für die Klassifikation nutzen lassen.

Fazit und Ausblick

Der vorgestellte DL basierte Workflow ist hinsichtlich einer angestrebten automatisierten oder teilautomatisierten Durchführung der Arealstatistik vielversprechend: Einige Klassen werden mit einer relativ hohen Genauigkeit von über 80% erkannt. Dieses Ergebnis lässt sich beispielsweise nutzen, indem Stichprobenpunkte, die mit einer hohen prognostizierten Wahrscheinlichkeit klassifiziert werden, aus der manuellen Prüfung entfernt werden. Zudem kann das vorgestellte Klassifikationsproblem mit 72 zentrumsbasierten Klassen als dichotome Fragestellung interpretiert werden: Die Frage nach der Änderung der Klassenzugehörigkeit zwischen zwei Erhebungen der Arealstatistik lässt sich mit *Ja* oder *Nein* beantworten, wobei Änderungen nur für einen kleinen Teil der Stichprobenpunkte zu erwarten sind. Eine zuverlässige Veränderungsdetektion könnte eine Reduktion des Bearbeitungsaufwandes herbeiführen.

In weiterführenden Untersuchungen lassen sich festgestellte Schwächen adressieren und Lösungsansätze bereitstellen: RGB Bilder stellen die Grundlage der Arealstatistik dar, jedoch werden vom BFS ergänzende Informationen hinzugenommen, die zu einer sichereren Klassenzuordnung führen. Zueinander komplementäre Informationen lassen sich ebenso in ein Verfahren des maschinellen Lernens

einbinden. Zwar ist das vorgestellte Inception-v3 Netzwerk dazu konzipiert, Bildmaterial mit bis zu drei Kanälen zu analysieren, jedoch liesse sich eine angepasste DL Architektur zur Prozessierung multimodaler Daten konstruieren. Als weitere Möglichkeit könnte Inception-v3 (oder eine verwandte Architektur) in weitere Klassifikationsmethoden integriert und zu einem mehrstufigen Verfahren erweitert werden. Ein Lernverfahren, bei dem beispielsweise Luftbilder mit multispektralen Satellitendaten, amtlichen Daten und Höhenmodellen multiskalar und -temporal fusioniert werden, ist der bisherigen manuellen Erhebungsmethodik der Arealstatistik nahe und ist in der maschinellen Umsetzung erfolgversprechend.

Methoden der KI wie DL werden zurzeit stark weiterentwickelt und ermöglichen dank der verfügbaren Rechenleistungen und Datengrundlagen die Bearbeitung sehr ambitionierter Detektions- und Klassifikationsaufgaben, beispielsweise für die Fernerkundung. Die Grösse der Datensätze stellt jedoch eine Hürde für die lokale Bearbeitung dar. Cloudbasierte Analysen von Geodaten bieten einen Ansatz, diese Hürden weitgehend zu umgehen und runden das hier ausblickend diskutierte Angebot weiterer Schritte zur Anwendung der KI für die Arealstatistik ab. Aktuell werden diese Möglichkeiten in einem Forschungsauftrag des BFS am Institut Geomatik der FHNW und in Zusammenarbeit mit dem Geotechnologie Unternehmen ExoLabs GmbH (Zürich) untersucht.

Literatur

[1] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152: 166–177.

[2] Bundesamt für Statistik, 2014. Arealstatistik nach Nomenklatur 2004 – Standard.

[3] Bundesamt für Landestopografie, 2017. swissimage – Das digitale Farbornthophotomosaik der Schweiz.

[4] Jordan, D. et al., 2013. Simultaneous electroencephalographic and functional magnetic resonance imaging indicate impaired cortical top-down processing in association with anesthetic-induced unconsciousness. *Anesthesiology*, 119: 1031–1042.

[5] Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press, London, England.

[6] Chollet, F., 2018. *Deep Learning with Python*. Manning Publications Co., Shelter Island, New York, USA.

[7] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.

Bei sämtlichen aufgeführten Bildauschnitten aus dem Luftbilddatensatz swissimage liegt das Urheberrecht beim Bundesamt für Landestopografie (Vertrag JA100071).

Denis Jordan
Natalie Lack
Stefan Hochuli
Adrian F. Meyer
Markus Schär
Institut Geomatik, Hochschule für
Architektur, Bau und Geomatik
Fachhochschule Nordwestschweiz
Hofackerstrasse 30
CH-4132 Muttenz
denis.jordan@fhnw.ch
natalie.lack@fhnw.ch