

# Ein effizienter KI-Workflow zur Detektion und Segmentierung von Freiflächen-Photovoltaikanlagen in ganz Bayern

## An Efficient AI Workflow for the Detection and Segmentation of Ground-mounted Photovoltaic Systems throughout Bavaria

Jonathan Kraus | Kilian Sperber | Ulf Kreuziger | Henry Meißner

### Zusammenfassung

Der Artikel beschreibt einen effizienten KI-Workflow zur Erkennung und Segmentierung von Freiflächen-Photovoltaikanlagen in Bayern, der sich auf beliebige Regionen übertragen lässt. Es wurde ein zweistufiger Ansatz entwickelt, der auf Digitalen Orthophotos mit 20 cm (DOP20) bzw. 100 cm (DOP100) Auflösung basiert. Zwei verschiedenartige KI-Modelle, U-Net und YOLO, werden genutzt, um PV-Anlagen zu identifizieren sowie zu segmentieren. Der Kombinationsansatz, der zunächst niedrigauflösende DOP100-Bilder analysiert und dann gezielt hochauflösende DOP20-Kacheln einsetzt, steigert die Effizienz, indem nur relevante Bildbereiche mit hoher Auflösung verarbeitet werden. Erste Tests in drei bayerischen Landkreisen zeigen eine hohe Genauigkeit beider KI-Modelle. Der Ansatz soll für die flächendeckende Segmentierung in Bayern eingesetzt werden und so zur Änderungsdetektion und Qualitätssicherung des Digitalen Basis-Landschaftsmodells (ATKIS® Basis-DLM) beitragen.

**Schlüsselwörter:** Bildsegmentierung, Freiflächen-Photovoltaikanlage, KI, U-Net, YOLO, zweistufiger Workflow

### Summary

*The article presents an efficient AI workflow for detecting and segmenting ground-mounted photovoltaic systems across Bavaria. A two-step approach was developed based on high-resolution digital orthophotos with 20 cm (DOP20) and 100 cm (DOP100) resolution. Two AI models, U-Net and YOLO, are used to identify and segment PV installations. The approach, which first processes DOP100 images and then selectively applies DOP20 tiles, significantly increases efficiency by only processing relevant areas in high resolution. Initial tests in three Bavarian counties indicate high accuracy for both AI models. The approach is intended for statewide segmentation in Bavaria, contributing to change detection and quality assurance of the Digital Basic Landscape Model (ATKIS® Basis-DLM).*

**Keywords:** AI, ground-mounted photovoltaic system, image segmentation, two-step workflow, U-Net, YOLO

### 1 Einführung

Die Aktualisierung und Qualitätssicherung des Digitalen Basis-Landschaftsmodells (ATKIS® Basis-DLM) der Bayerischen Vermessungsverwaltung beruht im Wesentlichen auf Informationen der Gebietstopographen, denen hierfür regelmäßig alle Geobasisdaten des Landesamtes für Digitalisierung, Breitband und Vermessung (LDBV) zur Verfügung stehen. Da jedoch der Umfang und die Detailtiefe der zu erfassenden Informationen immer weiter zunehmen, sollen die Gebietstopographen zukünftig durch Methoden der Künstlichen Intelligenz (KI) am Gegenstand der Geobasisdaten des LDBV unterstützt werden. Die mit KI erkannten Veränderungen können somit vom Gebietstopographen zielgerichtet vor Ort angefahren werden, sodass lediglich eine Verifizierung notwendig ist oder eine Anfahrt im Außendienst gegebenenfalls obsolet wird.

Aus diesem Grund wurde die Forschungsk Kooperation »GeoDataObjektDetektor« (GDOD) zwischen der Ostbayerischen Technischen Hochschule Amberg-Weiden und dem LDBV ins Leben gerufen. Ziel der Forschungsk Kooperation ist die automatisierte Detektion von Veränderungen verschiedener Topographieobjekte des ATKIS® Basis-DLM anhand der Geobasisdaten des LDBV und mittels Satellitendaten unter dem Einsatz von KI. Die Verwendung von KI ist beim LDBV keine Neuheit, da bereits vor einigen Jahren die automatisierte Detektion von Gebäuden zur Baufallerkundung mit KI-Methoden realisiert wurde (Roschlaub et al. 2020).

Auch in anderen deutschen Bundesländern wird bereits KI in der Geoinformatik eingesetzt. Ein prominentes Beispiel ist die Ableitung der Landbedeckung mittels des Verfahrens Cop4ALL, welches aus NRW stammt (Sandmann et al. 2022).

Die Evaluierung der Tatsächlichen Nutzung im ATKIS® Basis-DLM mittels Deep Learning wurde bereits von Reiter et al. (2023) untersucht. In dieser Arbeit wurde ein Convolutional Neural Network (CNN) anhand von Digitalen Orthophotos der Auflösung 20 cm (DOP20) trainiert, um die Tatsächliche Nutzung zu klassifizieren.

Anders als in der Arbeit von Reiter et al. (2023) wird in diesem Artikel die Detektion und Segmentierung von einzelnen Objektarten am Beispiel der Freiflächen-Photovoltaikanlagen vorgestellt. Dabei wird auch ein zweistufiger

ger KI-Workflow aufgezeigt, bei dem zunächst DOP100-Kacheln mit Freiflächen-Photovoltaikanlagen identifiziert werden, sodass anschließend nur ein Bruchteil eines Gebietes mit DOP20-Kacheln segmentiert wird. Dieser KI-Workflow wird anhand von zwei unterschiedlichen KI-Architekturen (U-Net und YOLO) verifiziert und jeweils mit der Segmentierung aller DOP20-Kacheln verglichen.

## 2 Datengrundlage

### 2.1 Digitales Orthophoto

Das LDBV stellt eine Vielzahl amtlicher Geobasisdaten zur Verfügung. Für das Training und die Evaluierung verschiedener KI-Modelle dient das DOP als Datengrundlage, welches in einem zweijährigen Befliegungszyklus bayernweit erfasst wird. Seit 2024 erfolgt die Befliegung mit einer Längs- bzw. Querüberdeckung von mindestens 80 bzw. 55 %. Die Aufnahmen werden sowohl im Farbraum Rot, Grün, Blau und Infrarot (RGBI) als auch in Graustufen (PAN) erstellt. (LDBV 2024)

Für die Detektion und Segmentierung von Freiflächen-Photovoltaikanlagen wird zunächst das DOP-RGB mit einer Bodenauflösung von 20 cm bzw. 100 cm verwendet, da dieses einen höheren Informationsgrad als das DOP-PAN besitzt.

### 2.2 Trainingsdaten

Für das Training unterschiedlicher KI-Modelle sind qualitativ hochwertige sowie heterogene Trainingsdaten von entscheidender Bedeutung. Bei der Auswahl der Trainingsdaten wird deshalb unter anderem auf eine gleichmäßige Verteilung folgender Aspekte geachtet: Aufnahmedatum (Jahreszeit), Anlagentyp (Struktur, Farbe), Ausrichtung, Belichtung, Vegetation und Region in Bayern. Die Trainingsdatensätze setzen sich aus Kacheln des DOP-RGB mit einer Gesamtfläche von ca. 250 km<sup>2</sup> zusammen, die gleichmäßig über ganz Bayern verteilt sind und in einer Auflösung von 20 bzw. 100 cm vorliegen.

Die Modelle werden mit verschiedenen Auflösungen trainiert, da bereits in Jiang et al. (2021) nachgewiesen wurde, dass die Segmentierung von Photovoltaikanlagen in hochauflösenden Bildern durch ein Modell, das auf niedrig auflösenden Bildern trainiert wurde, ebenso wie die Segmentierung in niedrig auflösenden Bildern durch ein Modell, das auf hochauflösenden Bildern trainiert wurde (Kreuzanwendung), keine zufriedenstellenden Ergebnisse liefert.

Für das Training des Modells, das auf dem DOP20-Datensatz basiert, werden die Photovoltaik-Module sorgfältig manuell digitalisiert, wobei der Untergrund von Freiflächen-Photovoltaikanlagen nicht mit segmentiert wird.

Ziel ist es, ausschließlich die Photovoltaik-Module zu extrahieren, um eine möglichst präzise Trainingsgrundlage zu gewährleisten. Der Datensatz umfasst insgesamt etwa 20.000 Geoobjekte, welche im Verhältnis 80/20 in Trainings- bzw. Validierungsdaten gesplittet werden. Um eine potenzielle Verfälschung der Trainingsdaten zu vermeiden, werden Gebäude in dem DOP anhand ihrer Hausumringe geschwärzt, da Photovoltaik-Module auf Gebäudedächern nicht digitalisiert werden. Dies ist erforderlich, da sich die visuellen Merkmale von Photovoltaik-Modulen auf Freiflächenanlagen und Dachanlagen stark ähneln. Ohne diese Maßnahmen könnte es zu Widersprüchen in den Trainingsdaten kommen, was die Modellgenauigkeit beeinträchtigen könnte. Diese Methode wird nur beim U-Net angewendet, da sich dadurch bei YOLO keine Verbesserung der Ergebnisse zeigte.

Für das Training des Modells auf Basis des DOP100-Datensatzes wird ein anderer Ansatz verfolgt. Aufgrund der geringeren räumlichen Auflösung ist eine exakte Unterscheidung einzelner Photovoltaik-Module nicht mehr möglich. Daher werden die Module aggregiert, um eine angemessene Repräsentation zu gewährleisten. Im Gegensatz zum DOP20-Modell wird bei diesem Ansatz auf die Schwärzung von Gebäuden verzichtet, da Photovoltaik-Module auf Hausdächern in der geringeren Auflösung der DOP100-Kacheln nur schwer bzw. nicht mehr zu erkennen sind.

Die Detektion und Segmentierung von Freiflächen-Photovoltaikanlagen/-modulen stellt grundsätzlich eine binäre Segmentierungsaufgabe mit den Klassen (PV/keine PV) dar, weshalb aus den (aggregierten) Geoobjekten und den DOP-Kacheln Schwarz-Weiß-Masken mit der Feature Manipulation Engine (FME) erstellt werden. Die RGB-Bilder und Schwarz-Weiß-Masken werden anschließend auf die vom jeweiligen Modell geforderte Größe gekachelt (vgl. Abb. 1).

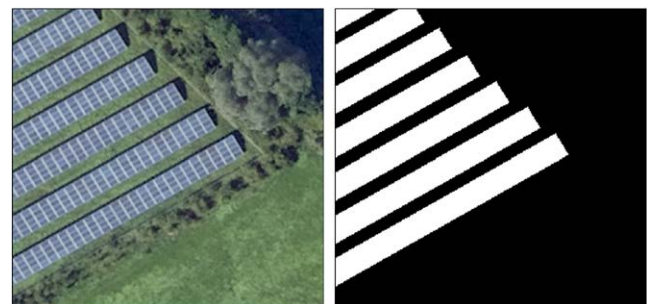


Abb. 1: DOP-RGB20-Bild (links) und korrespondierende Maske (rechts)

Um die Robustheit des Modells zu erhöhen und die Menge an Trainingsdaten zu erweitern, werden zusätzliche Bildvariationen durch verschiedene Methoden der Image Augmentation generiert (Xu et al. 2023). Als Image Augmentation wird neben Intensitätsänderungen auch eine geometrische Veränderung implementiert, die auf der Rotation der Bilder und Masken um 90 Grad basiert. Diese Maßnahme dient dazu, die unterschiedliche Ausrichtung

von Photovoltaik-Modulen zu simulieren, was durch die Auswahl der Trainingsdaten allein nur eingeschränkt möglich ist, da Photovoltaik-Module in der Realität überwiegend nach Süden ausgerichtet sind.

Die Wahl der 90°-Rotation ist zudem technisch motiviert, da andere Rotationswinkel mit einem Schärfeverlust im Bild einhergehen, der sich negativ auf die Qualität des Trainingsdatensatzes auswirken würde. Der Schärfeverlust tritt bei Rotationen auf, die nicht durch 90°-Intervalle teilbar sind, da diese eine Neuberechnung der Pixelpositionen im Bild erfordern, was Unschärfen und Verzerrungen verursacht. (Parker et al. 1983)

### 2.3 Testgebiete

Zur Bestimmung der Generalisierungsfähigkeit der Modelle werden drei Testgebiete in ganz Bayern ausgewählt, in denen keine Trainingsdaten enthalten sind. Um eine ausreichend große Stichprobe zu verwenden, fällt die Wahl dabei auf ganze Landkreise. Diese werden basierend auf mehreren Kriterien ausgewählt, die sie als geeignet auszeichnen.

Ein hoher Anteil an Photovoltaikanlagen in diesen Gebieten gegenüber anderen Landkreisen ist essenziell, um die Modelle auf viele verschiedene Arten von Freiflächen-Photovoltaikanlagen zu testen und Schwachstellen identifizieren zu können. Auf Basis des Datenbestands aus dem ATKIS® Basis-DLM werden deshalb Landkreise mit einem großen Bestand an Freiflächen-Photovoltaikanlagen gefiltert.

Zudem ist die Größe der Landkreise ein entscheidendes Kriterium. Durch die Wahl großflächiger Testgebiete kann eine größere Vielfalt an Landschafts- und Bebauungstypen abgedeckt werden. Diese Vielfalt ist wichtig, um die Flexibilität des Modells zu testen, sodass es auch auf größere Gebiete mit unterschiedlichen Topografien und Landnutzungen anwendbar ist.

Durch gängige GIS-Operationen wird mit Hilfe des Trainingsdatensatzes, des ATKIS® Basis-DLM und der Verwaltungseinheiten in Bayern eine Liste möglicher Kandidaten für Testgebiete erstellt. Die Wahl fällt dabei schließlich auf den Landkreis Bad Kissingen (Unterfranken), den Landkreis Regensburg mit der kreisfreien Stadt Regensburg (Oberpfalz) sowie den Landkreis Ostallgäu mit der kreisfreien Stadt Kaufbeuren (Schwaben), vgl. Abb. 2.

Die drei ausgewählten Testgebiete bieten eine repräsentative Verteilung über Bayern, sowohl in geographischer als auch landschaftlicher Hinsicht. Diese Regionen unterscheiden sich deutlich in ihrer Vegetation, Topographie und Siedlungsstruktur, was es ermöglicht, die Leistungsfähigkeit der Modelle in verschiedenen Umweltkontexten zu prüfen. Zudem stammen die Luftbilder aus Bad Kissingen und Regensburg aus dem Jahr 2023, während die Daten aus dem Landkreis Ostallgäu aus 2022 stammen.

Durch die Berücksichtigung all dieser Kriterien – ein hoher PV-Anteil, große Testgebiete und geografische sowie

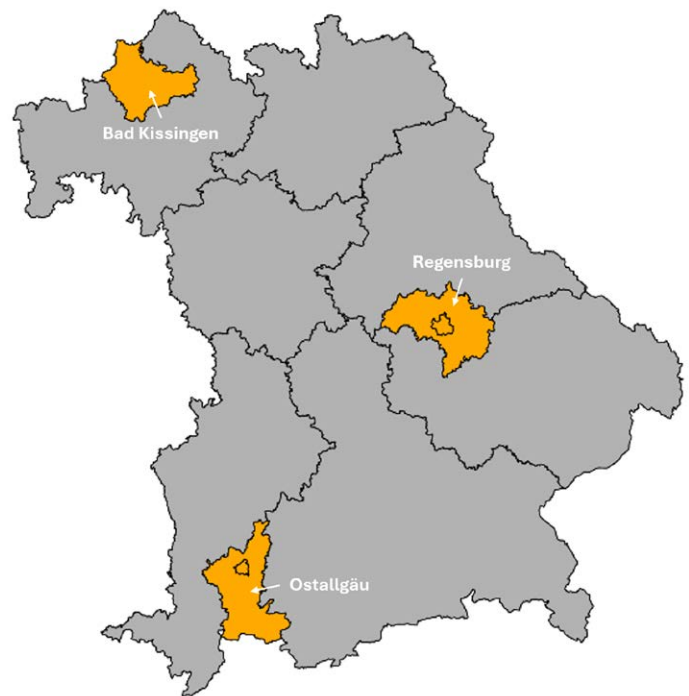


Abb. 2: Testgebiete

landschaftliche Vielfalt – kann die Generalisierungsfähigkeit der KI-Systeme zur Detektion von Photovoltaikanlagen in ganz Bayern fundiert validiert werden.

## 3 Methodik

### 3.1 U-Net

Für die Segmentierung der Freiflächen-Photovoltaikanlagen wird ein CNN-basiertes U-Net ausgewählt, vgl. Abb. 3. Diese Netzarchitektur wurde ursprünglich von Ronneberger et al. (2015) für die Segmentierung neuronaler Strukturen und Zelleinheiten in Mikroskop-Bildern eingesetzt. Allerdings konnte diese Architektur auch erfolgreich auf Segmentierungs- bzw. Klassifizierungsaufgaben der Fernerkundung adaptiert werden, so wurde beispielsweise in Knopp et al. (2022) ein U-Net zur Klassifikation von Brandflächen in Sentinel-2-Aufnahmen entwickelt.

Die U-Net-Architektur folgt einem symmetrischen Encoder-Decoder-Design. In dieser Struktur werden zunächst Eingabebilder der Dimension  $256 \times 256$  Pixel mit je drei Kanälen verarbeitet. Der Encoder besteht aus fünf Faltungsblöcken, wobei jeder Block zwei Schichten umfasst: eine Schicht besteht dabei aus einem Convolution-Layer mit einem  $3 \times 3$ -Kernel gefolgt von einem Dropout-Layer. Durch den Einsatz von Mirror-Padding bleibt die Bildgröße unverändert. Am Ende eines jeden Faltungsblocks wird die Bildauflösung mittels eines Pooling-Layers um den Faktor 4 reduziert, während gleichzeitig die Anzahl der Merkmalskarten pro Block verdoppelt wird.

Im Decoder-Teil wird eine symmetrische Struktur zum Encoder verwendet, wobei die Auflösung der Bilddaten

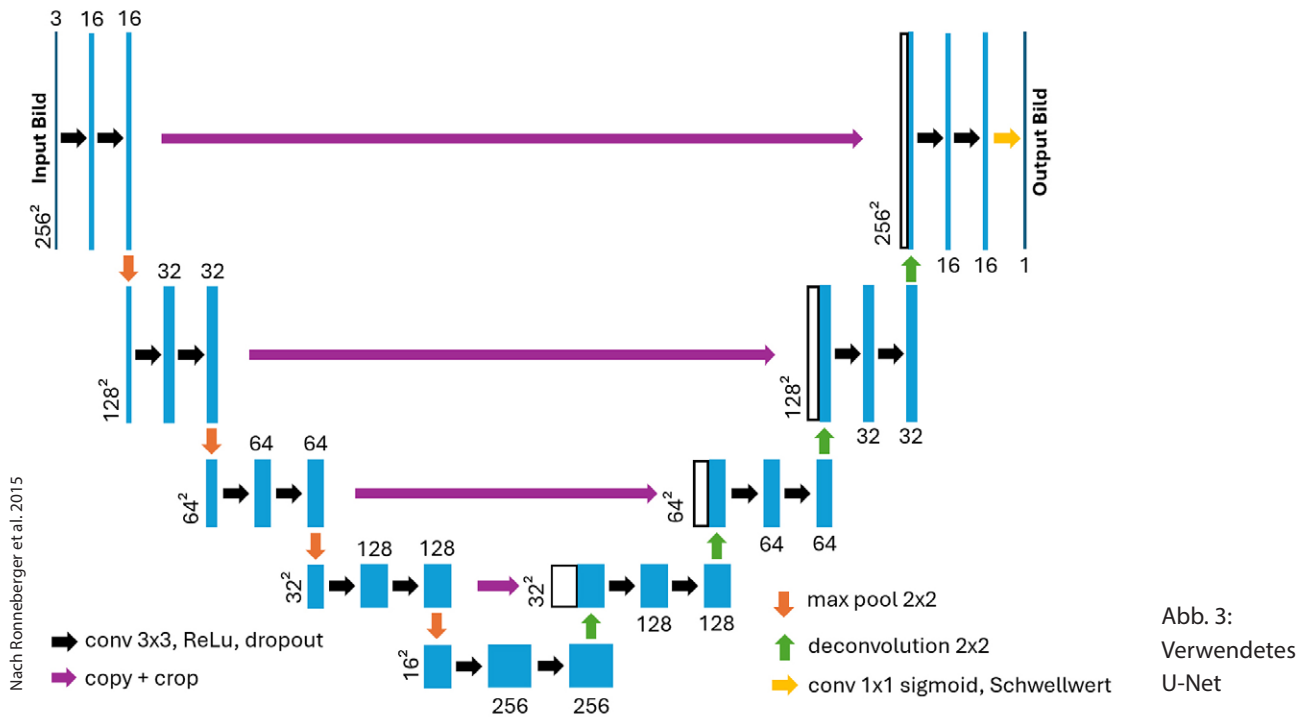


Abb. 3:  
Verwendetes  
U-Net

durch Deconvolution-Schichten pro Block um den Faktor 4 erhöht wird. Die Anzahl der Merkmalskarten wird dabei halbiert. Um den Verlust von Detailinformationen zu kompensieren, werden durch Copy-and-Crop-Operationen (skip connections) die Merkmale des korrespondierenden Encoder-Blocks auf den entsprechenden Decoder-Block übertragen. Dies ermöglicht eine präzisere Rekonstruktion der Bildinformationen.

Abschließend wird eine  $1 \times 1$ -Faltungsschicht mit einer Sigmoid-Aktivierungsfunktion eingesetzt, um eine Wahrscheinlichkeitskarte zu erzeugen, die jedem Pixel einen Wert zwischen 0 und 1 zuweist. Durch eine anschließende Schwellwertbildung erhält man die Klasse PV bzw. keine PV.

### 3.2 YOLO

YOLO (You Only Look Once) ist ein bekanntes Modell zur Objekterkennung, das erstmals 2016 eingeführt wurde und seitdem in verschiedenen Versionen weiterentwickelt wurde. Es basiert auf einem einheitlichen CNN, das den Objekterkennungsprozess als ein einzelnes Regressionsproblem formuliert. Im Gegensatz zu älteren Methoden wie R-CNN, die zuerst Regionen vorschlagen und dann klassifizieren, führt YOLO sowohl die Erkennung als auch die Lokalisierung von Objekten in einem Durchgang durch, was zu einer schnelleren Ergebnisfindung führt. (Redmon et al. 2016)

In diesem Beitrag wird YOLOv8x-seg verwendet, eine erweiterte Version des YOLOv8-Modells, die speziell auf Segmentierungsaufgaben angepasst ist. Es wurde 2023 von Glenn Jocher vorgestellt. Dieses Modell kombiniert die bewährte Architektur von YOLO mit einer speziellen Erweiterung für das maskenbasierte Segmentieren von Objekten.

Die Variante »x« ist für größere bzw. kompliziertere Modelle mit mehr Parametern und damit höherer Genauigkeit ausgelegt. YOLOv8x-seg ermöglicht es, Objekte nicht nur zu erkennen, sondern diese auch pixelgenau zu segmentieren. (Ultralytics 2024)

Obwohl bereits neuere YOLO-Versionen vorgestellt wurden, wird in dieser Arbeit zunächst weiterhin YOLOv8 eingesetzt, da diese eine stabile, gut dokumentierte und weit verbreitete Version ist (RUiNtheExtinct 2024).

### 3.3 Nachverarbeitungsprozess

Die Ergebnisse der beiden KI-Architekturen durchlaufen nach der Segmentierung einen Nachverarbeitungsprozess. Beim U-Net werden die Schwarz-Weiß-Masken zunächst vektorisiert, während die Ergebnisse bei YOLO bereits als Polygone vorliegen. Anschließend werden die Geoobjekte mit dem jeweiligen Testlandkreis verschnitten, um die Ergebnisse von landkreisübergreifenden DOP-Kacheln zu bereinigen. Dabei werden zudem Löcher in den Geometrien geschlossen. Durch eine anschließende Pufferung der Features um den Wert 0,2 m sowie anschließendes Aggregieren und Zurückpuffern um den gleichen Wert können die Übergänge zwischen zwei DOP-Kacheln bereinigt werden. Die Übergänge zwischen den einzelnen Schnipseln werden bereits durch vertikale und horizontale Überlappung berücksichtigt. Eine weitere Verschneidung eliminiert fälschlicherweise segmentierte Objekte, die sich auf Häusern (Photovoltaik), Verkehrsflächen, Eisenbahnflächen oder Gewässern befinden. Auf diesen Flächen ist die Existenz von Freiflächen-Photovoltaikanlagen ausgeschlossen. Anschließend durchlaufen die bearbeiteten Geoobjekte einen Kleinstflächenfilter, sodass Features mit einer Fläche

kleiner 8 m<sup>2</sup> gefiltert werden. Die irrtümliche Filterung von sehr kleinen, vereinzelt vorkommenden PV-Modulen wird dabei toleriert. Als Ergebnis des Nachverarbeitungsprozesses liegen georeferenzierte Features vor.

### 3.4 Klassischer Ansatz

Der klassische Ansatz zur Segmentierung der Freiflächen-Photovoltaikanlagen ist die vollständige Prozessierung aller DOP20-Kacheln eines Testgebietes. Diese werden zunächst durch das KI-Modell gegeben. Die segmentierten Kacheln durchlaufen dann den Nachverarbeitungsprozess, sodass man als Ergebnis die georeferenzierten Features erhält. Die vollständige Prozesskette ist in Abb. 4 dargestellt.

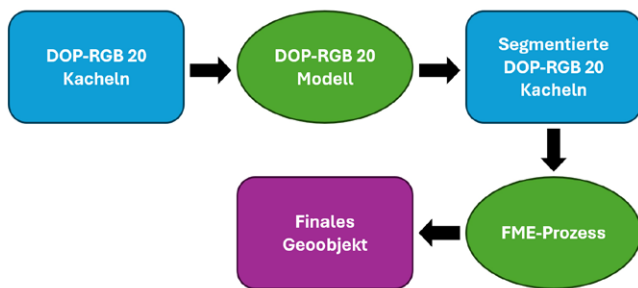


Abb. 4: Prozesskette »Klassischer Ansatz«

### 3.5 Kombinationsansatz

Der Kombinationsansatz unterscheidet sich hingegen vom klassischen Ansatz dadurch, dass das Testgebiet zunächst

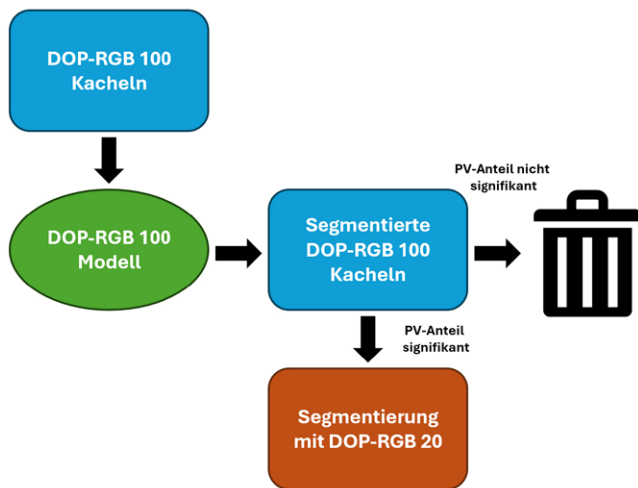


Abb. 5: Prozesskette »Kombinationsansatz«

in Meter-Auflösung dem DOP100-Modell zugeführt wird (Vorschaltung des DOP100-Modells). Der Vorteil liegt in der schnelleren Verarbeitung der DOP100-Kacheln, da bei diesen im Vergleich zu den DOP20-Kacheln nur 1/25 der Datenmenge verarbeitet werden muss. Durch eine Signifikanzprüfung der segmentierten DOP100-Kacheln können die Kacheln ermittelt werden, die anschließend noch mit dem DOP20-Modell verarbeitet werden. Dadurch muss nur ein Bruchteil des Testgebietes mit den rechenaufwändigeren DOP20-Kacheln segmentiert werden. Diese durchlaufen anschließend analog den klassischen Ansatz. Der KI-Workflow kann der Abb. 5 entnommen werden. Dieser Ansatz verspricht demzufolge eine deutliche Effizienzsteigerung.

### 3.6 Evaluierung in Testgebieten und Kennzahlen

Das sogenannte Testing ist ein essenzieller Bestandteil bei der Entwicklung und Validierung von KI-Systemen, insbesondere bei der Anwendung auf großflächige und komplexe Datensätze wie bayernweite Geodaten. Es dient dazu, die Leistungsfähigkeit und Zuverlässigkeit eines Modells in realen Anwendungsszenarien zu überprüfen. Ein richtig durchgeführtes Testing stellt sicher, dass das KI-Modell nicht nur in den Trainingsdaten, sondern auch in bisher unbekanntem Regionen und unter unterschiedlichen Umweltbedingungen verlässliche Ergebnisse liefert. Das ist bei der Anwendung auf die bayernweiten Luftbilder besonders wichtig, da das System eine Vielzahl von Photovoltaikanlagentypen in verschiedenen Regionen, die unter verschiedenen Bedingungen aufgenommen wurden, erkennen muss.

Die maßgeblichen Kenngrößen, die durch Testing bestimmt werden, beschreiben die Genauigkeit, die ein Endnutzer bei der Anwendung auf unbekannte Datensätze erhält.

Nach dem Segmentieren der drei Testgebiete liegen Ergebnisse vor, die den Nachverarbeitungsprozess durchlaufen und als georeferenzierte Shapedatei abgespeichert werden.

Durch eine Verschneidung der Ergebnisse mit dem entsprechenden manuell digitalisierten Ground Truth (»Soll«) können flächenmäßig die gängigen Kennzahlen »Precision« und »Recall« ermittelt werden. Bei der Verschneidung wird ein Toleranzwert von 0,3 m (1,5 Pixel) implementiert, um den Einfluss von Digitalisierungsungenauigkeiten in den Trainings- und Testdaten zu minimieren. In Abb. 6 ist die Verschneidung zwischen den Ergebnissen und dem Ground Truth exemplarisch dargestellt.

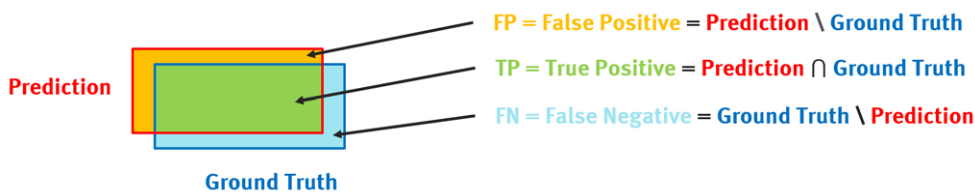


Abb. 6: Verschneidung Prediction und Ground Truth

Für die Verschneidung gelten folgende mathematische Zusammenhänge:

$$FP = Prediction \setminus Ground Truth$$

$$TP = Prediction \cap Ground Truth$$

$$FN = Ground Truth \setminus Prediction$$

Daraus können die Kennzahlen Precision («Präzision») und Recall («Vollständigkeit») sowie das harmonische Mittel aus beiden (F1-Score) berechnet werden (Murphy 2022):

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1-Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Ziel ist die vollständige Erfassung aller Freiflächen-Photovoltaikanlagen in dem jeweiligen Testgebiet, um daraus Rückschlüsse auf die Veränderung zu ziehen. Allerdings geben diese drei Kennzahlen keinen Aufschluss darüber, ob in einem Testgebiet tatsächlich alle PV-Parks erfasst wurden. Ein flächenmäßig kleiner PV-Park könnte selbst bei einem Recall von weit über 90 % nicht erkannt worden sein. Deshalb wird an dieser Stelle die Kennzahl »Erkannte Parks« eingeführt. Diese wird ermittelt, indem die einzelnen PV-Module in der Prediction («Vorhersage») und im Ground Truth durch Aggregation zu einzelnen PV-Parks zusammengefasst werden. Über die Intersection over Union (IoU) der aggregierten Parks kann somit die Anzahl an erkannten Parks bzw. die Erkennungsrate («ER») bestimmt werden. Ein Park wird dabei als erkannt eingestuft, sofern seine IoU mindestens 20 % beträgt:

$$IoU = \frac{Prediction \cap Ground Truth}{Prediction \cup Ground Truth}$$

$$IoU \geq 0.2 \rightarrow \text{Park erkannt}$$

$$IoU < 0.2 \rightarrow \text{Park nicht erkannt}$$

$$ER = \frac{\text{Park erkannt}}{\text{Park erkannt} + \text{Park nicht erkannt}}$$

Da die beiden zuvor beschriebenen Ansätze hinsichtlich ihrer Effizienz (Einsparquote) verglichen werden, ist noch eine weitere Kennzahl erforderlich, welche den Anteil an segmentierten Kacheln  $K$  beschreibt, auf denen keine Freiflächen-Photovoltaikanlage liegt.

$$Effizienz = 1 - \frac{K_{negativ \cap segmentiert}}{K_{negativ}}$$

Man nehme an, ein Testgebiet bestehe aus 1200 Kacheln, von denen auf 200 Kacheln der Ground Truth liegt, so ergeben sich 1000 Kacheln ohne Ground Truth ( $K_{negativ}$ ). Werden von diesen Kacheln 200 irrtümlicherweise segmentiert ( $K_{negativ \cap segmentiert}$ ), so ergibt sich eine Effizienz von 80 %. Eine Effizienz von 100 % bedeutet zwar, dass keine Kachel ohne Photovoltaikanlage segmentiert wurde, allerdings lässt sich daraus nicht ableiten, dass alle Kacheln mit Freiflächen-Photovoltaikanlagen segmentiert wurden. Es sind daher stets alle Kennzahlen gesamtheitlich zu betrachten.

## 4 Ergebnisse

In diesem Abschnitt werden die Leistungen der beiden verwendeten Modelle YOLO und U-Net zur Segmentierung von PV-Anlagen gegenübergestellt. Der Vergleich umfasst sowohl quantitative Metriken als auch eine qualitative Bewertung der Segmentierungsergebnisse. Auf die Darstellung der Trainings- bzw. Validierungskennzahlen wird an dieser Stelle verzichtet, da diese die Generalisierungsfähigkeit des Modells nur unzureichend beschreiben. Vielmehr werden die zuvor beschriebenen Nutzergenauigkeiten anhand der drei Testgebiete Bad Kissingen (KG), Regensburg (R) und Ostallgäu (OAL) ermittelt. Als Grafikkarte wird bei beiden Netzarchitekturen eine NVIDIA Quadro RTX 5000 verwendet.

### 4.1 U-Net

Die beiden trainierten U-Net-Modelle (DOP20-Modell und DOP100-Modell) werden in den Workflows (Abb. 4 und 5) verwendet. Die Segmentierung eines Landkreises in DOP100-Auflösung dauert ca. 3 Minuten, wobei hier die Größe eines Landkreises zeitlich betrachtet nur eine untergeordnete Rolle spielt. Die Segmentierung einer DOP20-Kachel benötigt ca. 2,2 Sekunden, sodass ein Landkreis mit 1500 km<sup>2</sup> dahingegen etwa 55 Minuten zur Prozessierung in Anspruch nimmt. In den Ergebnissen (Tab. 1) zeigt sich, dass beide Ansätze in den Kennzahlen Precision, Recall, F1-Score und Erkennungsrate jeweils stets über 90 % liegen, ausgenommen der Precision und demzufolge des F1-Scores im Landkreis Ostallgäu beim klassischen Ansatz. Die Erkennungsrate liegt bei beiden Ansätzen bei 100 %. Die Effizienz ist beim klassischen Ansatz hingegen immer 0 %, da bei diesem Ansatz alle rechenintensiveren DOP20-Kacheln segmentiert werden. Die durchschnittliche IoU liegt im Mittel bei jeweils über 96 %. Die Ergebnisse der drei Testgebiete werden für jeden Ansatz anschließend noch gemittelt. In Abb. 7 ist die beispielhafte Segmentierung einer Freiflächen-Photovoltaikanlage dargestellt.

Tab. 1: Ergebnisse U-Net

Landkreis	Precision	Recall	F1-Score	ER	Effizienz	Ø IoU
<b>Bad Kissingen (KG)</b>						
Klassisch	97,38 %	96,11 %	96,74 %	100,00 %	0,00 %	97,51 %
Kombi	99,36 %	96,00 %	97,65 %	100,00 %	99,27 %	97,22 %
<b>Regensburg (R)</b>						
Klassisch	93,62 %	97,61 %	95,57 %	100,00 %	0,00 %	98,28 %
Kombi	97,13 %	97,59 %	97,36 %	100,00 %	98,05 %	98,28 %
<b>Ostallgäu (OAL)</b>						
Klassisch	82,48 %	98,94 %	89,96 %	100,00 %	0,00 %	94,66 %
Kombi	95,54 %	98,48 %	96,99 %	100,00 %	96,81 %	92,99 %
<b>Mittel</b>						
Klassisch	91,16 %	97,55 %	94,09 %	100,00 %	0,00 %	96,82 %
Kombi	97,34 %	97,36 %	97,33 %	100,00 %	98,04 %	96,16 %



Abb. 7: Segmentierungsergebnis einer Freiflächen-Photovoltaikanlage (DOP20 links, U-Net Ergebnis Mitte, YOLO Ergebnis rechts)

### 4.2 YOLO

Bei der YOLO-Architektur werden ebenfalls die beiden trainierten Modelle in den Workflows (Abb. 4 und 5) verwendet und anhand der drei Testgebiete evaluiert. Die Dauer einer Segmentierung eines Landkreises in DOP100-Auflösung beträgt etwa 5 bis 6 Minuten. Für die Verarbeitung einer DOP20-Kachel können rund 3,5 Sekunden veranschlagt werden, sodass ein Landkreis mit 1500 km<sup>2</sup> in rund 88 Minuten prozessiert werden kann. Die Ergebnisse (Tab. 2) weisen dabei ähnliche Größenordnungen zu den U-Net-Ergebnissen auf. Hierbei liegen alle Kennzahlen bei über 90 % (ausgenommen Effizienz beim klassischen Ansatz). Die Erkennungsrate liegt ebenfalls bei beiden Ansätzen bei 100 %. Die durchschnittliche IoU liegt im Mittel bei über 97 %. Auch bei dieser Netzarchitektur werden die gemittelten Kennzahlen für beide Ansätze berechnet. Das Segmentierungsergebnis einer Freiflächen-Photovoltaikanlage ist in Abb. 7 erkennbar.

### 4.3 Vergleich Klassischer Ansatz und Kombinationsansatz

Um die beiden Ansätze zu vergleichen, werden die gemittelten Kennzahlen beider Ansätze gegenübergestellt (Abb. 8). Es zeigt sich, dass die Precision beim Kombinationsansatz etwas höher ist als beim klassischen Ansatz, während der Recall sowie die durchschnittliche IoU nur minimal schlechter sind. Demzufolge ist auch der F1-Score beim Kombinationsansatz leicht besser. Die Erkennungsrate ist bei beiden Ansätzen hingegen identisch. Die größte Differenz in den Kennzahlen weist die Effizienz auf. Beim klassischen Ansatz beträgt diese durchgehend 0 %, während diese beim Kombinationsansatz stets bei über 96 % liegt.

Tab. 2: Ergebnisse YOLO

Landkreis	Precision	Recall	F1-Score	ER	Effizienz	Ø IoU
<b>Bad Kissingen (KG)</b>						
Klassisch	94,51 %	98,45 %	96,44 %	100,00 %	0,00 %	98,11 %
Kombi	94,95 %	98,22 %	96,56 %	100,00 %	99,84 %	97,30 %
<b>Regensburg (R)</b>						
Klassisch	92,50 %	98,18 %	95,26 %	100,00 %	0,00 %	98,37 %
Kombi	94,16 %	97,56 %	95,83 %	100,00 %	98,05 %	98,22 %
<b>Ostallgäu (OAL)</b>						
Klassisch	91,56 %	98,28 %	94,80 %	100,00 %	0,00 %	96,99 %
Kombi	93,14 %	97,78 %	95,41 %	100,00 %	98,81 %	97,03 %
<b>Mittel</b>						
Klassisch	92,86 %	98,30 %	95,50 %	100,00 %	0,00 %	97,82 %
Kombi	94,08 %	97,85 %	95,93 %	100,00 %	98,90 %	97,52 %

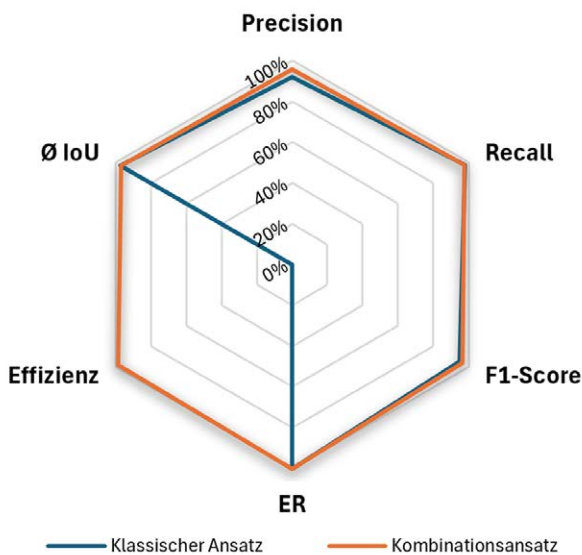


Abb. 8: Vergleich der beiden Ansätze

## 5 Diskussion

In diesem Kapitel wird die Effizienz und Leistungsfähigkeit der U-Net- und YOLO-Architekturen zur Segmentierung von Freiflächen-Photovoltaikanlagen analysiert. Die vergleichende Betrachtung zeigt, dass die Ergebnisse beider Architekturen Kennzahlen in der gleichen Größenordnung liefern. Dies ist vor allem auf die hohe Qualität der verwendeten Datensätze zurückzuführen.

### 5.1 Datengrundlage

Die hohe Genauigkeit der Ergebnisse zeigt deutlich die Qualität und Konsistenz der verwendeten Trainings- und

Testdaten. Je präziser und konsistenter die Datensätze, desto höher die Performance des Modells. Dank gezielter Maßnahmen wie dem Schwärzen von Hausdächern beim U-Net konnte die Klassifikation weiter optimiert werden, da das Modell dadurch weniger widersprüchliche Daten verarbeitet.

Um die Modelle zukünftig noch flexibler zu gestalten, bieten sich historische DOP als ergänzende Trainingsdatenquelle an. Diese ermöglichen es, das Modell auf mehrere Befliegungszeiträume zu trainieren, da bislang nur Daten aus den Jahren 2022 und 2023 verwendet wurden.

### 5.2 Bewertung des Kombinationsansatzes

Als besonders geeigneter Ansatz zur Effizienzsteigerung hat sich die Vorselektion von DOP-Kacheln mit einem DOP100-Modell herausgestellt. Durch diese Technik konnte die Analysegeschwindigkeit bei annähernd gleichbleibender Qualität erheblich gesteigert werden. Die Effizienzkennzahl liegt beim Kombinationsansatz stets bei über 96 %.

Außerdem konnte eine deutliche Reduzierung von Fehlklassifikationen (False Positives) durch die Vorselektion erreicht werden, was vor allem anhand der Precision zu erkennen ist.

Durch diesen Ansatz ist es mit der U-Net-Architektur möglich, PV-Anlagen in ganz Bayern über eine Dauer von ca. 5 Stunden vollständig zu segmentieren.

Der einzige Nachteil, den der Kombinationsansatz mit sich bringt, ist, dass sehr kleine Teile von PV-Anlagen, die sich am Rand mancher Kacheln befinden, vom DOP100-Modell nicht gefunden werden und so vereinzelt Kacheln nicht in DOP20-Qualität segmentiert werden (Abb. 9). Größere Strukturen werden jedoch zuverlässig erkannt,



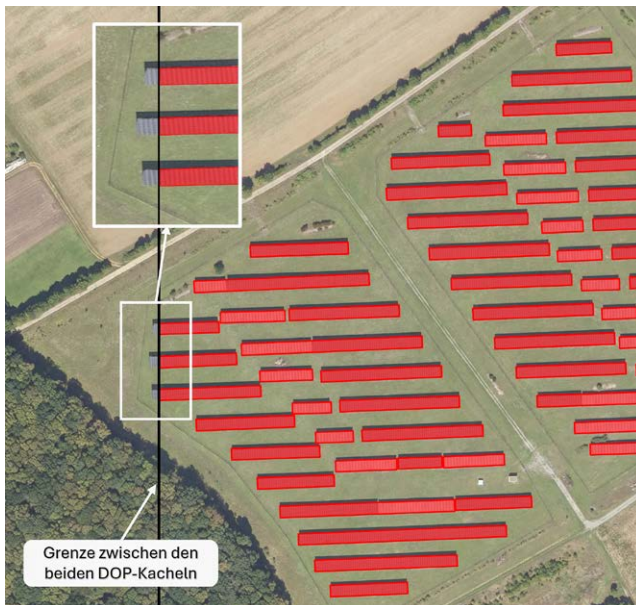


Abb. 9: Unvollständig segmentierter PV-Park

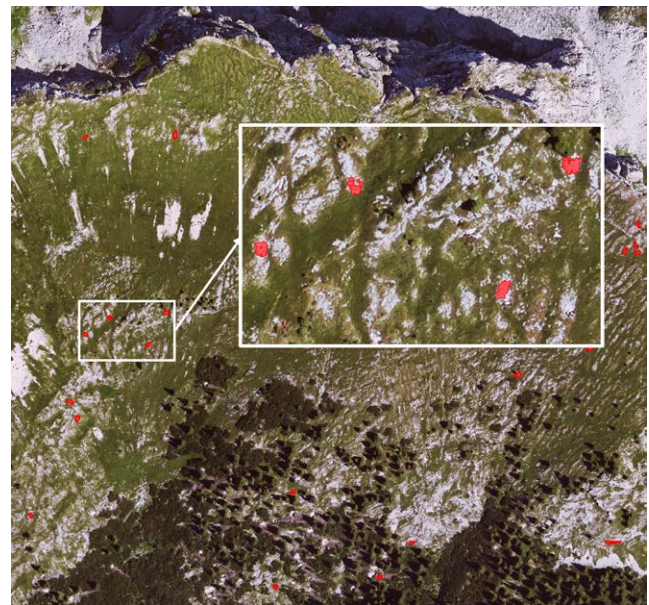


Abb. 10: Falsch segmentierte Gebirgsstrukturen

sodass dies keinen Einfluss auf die insgesamt Erkennung des jeweiligen PV-Parks hat.

Um diese Auswirkung zahlenmäßig zu betrachten, kann man die Recall-Werte von Kombinationsansatz und klassischem Ansatz vergleichen. Der Recall ist beim klassischen Ansatz um höchstens 0,6 % höher. Der Kombinationsansatz ist in diesem Gesichtspunkt minimal im Nachteil, was sich auch in einer leicht niedrigeren durchschnittlichen IoU auswirkt.

Jedoch überwiegt der große Vorteil: vollständige Erkennung aller PV-Anlagen bei mindestens 96 % höherer Effizienz, was mit einer erheblichen Zeitersparnis einhergeht.

### 5.3 Netzarchitekturen

Ein weiterer Schwerpunkt der Untersuchung lag auf der Analyse der Netzarchitekturen. Die Unterschiede zwischen U-Net und YOLO sind zwar klein, jedoch lassen sich einige Erkenntnisse aus den Ergebnissen ableiten.

YOLO erzielt konstantere Ergebnisse als das U-Net, da es in allen Testgebieten nahezu identische Kennzahlen liefert. Das deutet auf eine bessere Generalisierungsfähigkeit durch die größere Modellarchitektur von YOLO hin. Das U-Net unterliegt hingegen leichten Schwankungen, was sich insbesondere bei der Precision im Landkreis Ostallgäu bemerkbar macht, da sich diese beim klassischen Ansatz mit 82,48 % deutlich von den anderen Vergleichswerten unterscheidet. Im Landkreis Ostallgäu werden insbesondere sehr viele Kleinstflächen falsch segmentiert. Grund hierfür sind unter anderem falsch segmentierte Gebirgsstrukturen (Abb. 10). Setzt man den Flächenfilter auf 15 m<sup>2</sup> so erhält man bereits eine Steigerung der Precision um ca. 4,2 %, einhergehend mit einem geringfügigen Verlust beim Recall. Deshalb wird der Flächenfilter mit einem Schwellenwert von 8 m<sup>2</sup> beibehalten, um den Recall durch die Elimination von kleinen PV-Modulen nicht nachhaltig

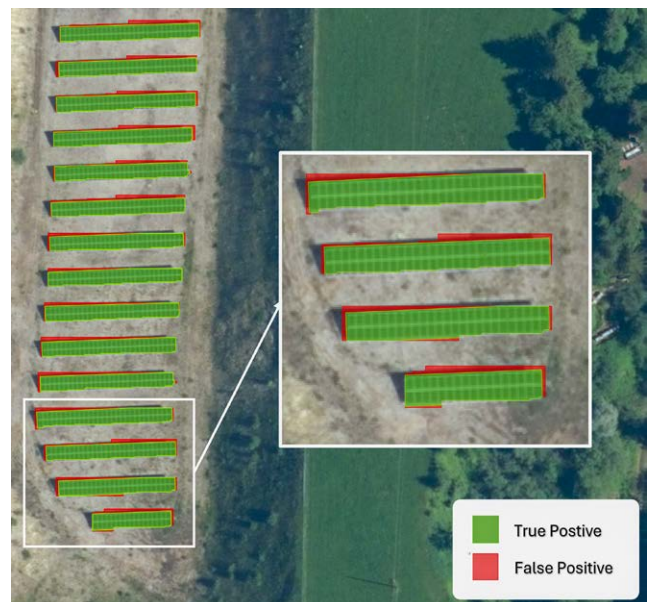


Abb. 11: YOLO-Prediction aufgeteilt in True/False Positives

zu beeinflussen. Vielmehr liefert der Kombinationsansatz eine zuverlässige Steigerung der Precision ohne einen signifikanten Rückgang des Recall-Wertes.

Die Precision von U-Net ist beim Kombinationsansatz durchschnittlich um über 3 % höher als die des YOLO-Modells. Beim klassischen Ansatz zeigt sich ein ähnliches Verhalten, wobei die Precision von U-Net im Mittel leicht niedriger ausfällt. Dies ist vor allem auf die Ergebnisse im Landkreis Ostallgäu zurückzuführen. Generell neigt YOLO dazu, die einzelnen Masken größer zu generieren, als es den tatsächlichen Objekten entspricht. Durch die Wahl des niedrigen Toleranzwertes von 0,3 m bei der Verschneidung der Ergebnisse werden somit Teile der Prediction als falsch positiv gewertet, was negative Auswirkungen auf die Precision hat (Abb. 11).

Neben der besseren Precision-Ergebnisse bietet die U-Net-Lösung aufgrund der verhältnismäßig einfachen

Architektur eine deutlich schnellere Datenverarbeitung. Während es einen Landkreis durchschnittlich in 6 Minuten verarbeitet, benötigt die YOLO-Lösung 13 Minuten. U-Net ist somit um fast 55 % schneller als YOLO.

Für einen Einsatz in der Bayerischen Vermessungsverwaltung eignen sich beide Modelle. In allen Testgebieten liegen die Erkennungsraten bei stets 100 %. Somit werden alle PV-Anlagen erkannt.

## 6 Fazit und Ausblick

Der vorgestellte Workflow zeigt anhand zweier unterschiedlicher KI-Netzarchitekturen, dass sich Freiflächen-Photovoltaikanlagen in Bayern durch die Kombination von DOP100- und DOP20-Daten effizient segmentieren lassen. Die dabei abgeleiteten Ergebnisse können vom LDBV zur Qualitätssicherung und Aktualisierung des ATKIS® Basis-DLM genutzt werden. In den nächsten Schritten ist geplant, die gesamte Segmentierung Bayerns in der DOP20-Qualität umzusetzen.

Darüber hinaus bietet es sich an, die trainierten Modelle und den entwickelten Ansatz auf Daten anderer Bundesländer anzuwenden. Dies würde die Generalisierungsfähigkeit der Modelle weiter evaluieren und könnte die Grundlage für eine breitere Anwendung schaffen.

Die Methode zur Kombination von Daten in unterschiedlichen räumlichen Auflösungen bietet zudem Potenzial für die Übertragung auf andere Objektarten des ATKIS® Basis-DLM. Dadurch könnten auch weitere Landnutzungsarten effizient und hochauflösend erfasst werden.

Insgesamt stellt der entwickelte Ansatz eine robuste Grundlage für zukünftige Anwendungen im Bereich der Geoinformatik und Geodatenverarbeitung dar und eröffnet neue Möglichkeiten für die präzise und gleichzeitig effiziente Analyse von Geodaten.

### Danksagung

Unser Dank gilt dem LDBV und dem Bayerischen Staatsministerium der Finanzen und für Heimat für die Bereitstellung der Geodaten sowie die Forschungskooperation.

### Literatur

- Jiang, H., Yao, L., Lu, N., Qin, J., Liu, T., Liu, Y., Zhou, C. (2021): Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery. In: *Earth System Science Data*, Vol. 13 (11), 5389–5401.
- Knopp, L., Wieland, M., Rättich, M., Martinis, S. (2022): Ein Deep-Learning-Ansatz zur Klassifikation von Brandflächen in Sentinel-2-Aufnahmen. In: Grunau, W. (Hrsg.): *Künstliche Intelligenz in Geodäsie und Geoinformatik. Potenziale und Best-Practice-Beispiele*. Wichmann, Berlin/Offenbach, 95–110.

- LDBV – Landesamt für Digitalisierung, Breitband und Vermessung (2024): Bayernbefliegung. Seit 1987 aktuelle Luftbilder von ganz Bayern. <https://www.ldbv.bayern.de/produkte/luftbild/bayernbefliegung.html>, letzter Zugriff 11/2024.
- Murphy, K. P. (2022): *Probabilistic Machine Learning. An Introduction*, The MIT Press, Cambridge/London.
- Parker, J. A., Kenyon, R. V., Troxel, D. E. (1983): Comparison of Interpolating Methods for Image Resampling. In: *IEEE Transactions on Medical Imaging*, Vol. 2 (1), 31–39. DOI: 10.1109/TMI.1983.4307610.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016): You Only Look Once: Unified, Real-Time Object Detection. In: *CVPR*. 779–788.
- Reiter, D., Meinel, G., Forkel, M. (2023): Evaluierung der Tatsächlichen Nutzung im ATKIS Basis-DLM mittels Digitaler Orthophotos und Deep Learning. In: *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, Heft 3/2023, 148. Jg., 126–136. DOI: 10.12902/zfv-0422-2023.
- Ronneberger, O., Fischer, P., Brox, T. (2015): U-Net Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W. M., Frangi, A. F. (Eds.): *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer, Cham, 234–241.
- Roschlaub, R., Li, Q., Auer, S., Möst, K., Glock, C., Schmitt, M., Shi, Y., Zhu, Z. (2020): KI-basierte Detektion von Gebäuden mittels Deep Learning und amtlichen Geodaten zur Baufallerkundung. In: *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, Heft 3/2020, 145. Jg., 180–189. DOI: 10.12902/zfv-0299-2020.
- RUiNtheExtinct (2024): YOLO Models – An objective comparison. <https://ruintheextinct.medium.com/yolo-models-an-objective-comparison-1566e6b6dcb7>, letzter Zugriff 11/2024.
- Sandmann, S., Hochgürtel, G., Piroška, R., Steffens, C. (2022). Cop4ALL NRW – Ableitung der Landbedeckung in Nordrhein-Westfalen mit Fernerkundung und künstlicher Intelligenz. In: *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, Heft 5/2022, 147. Jg., 299–310. DOI: 10.12902/zfv-0407-2022.
- Ultralytics (2024): Ultralytics YOLOv8/. <https://docs.ultralytics.com/de/models/yolov8/>, letzter Zugriff 11/2024.
- Xu, M., Yoon, S., Fuentes, A., Park, D. (2023): A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. In: *Pattern Recognition*, 137, 109347. DOI: <https://doi.org/10.1016/j.patcog.2023.109347>.

### Kontakt

Jonathan Kraus | Kilian Sperber |  
 Prof. Dr. Ulf Kreuziger, Professur für Geomatik und Geodäsie |  
 Prof. Dr. Henry Meißner, Professur für Geodatenmanagement und Geovisualisierung  
 Ostbayerische Technische Hochschule Amberg-Weiden  
 Kaiser-Wilhelm-Ring 23, 92224 Amberg  
 j.kraus3@oth-aw.de | k.sperber@oth-aw.de  
 u.kreuziger@oth-aw.de | h.meissner@oth-aw.de

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