

A robust and transferable AI workflow for segmenting ground-mounted Photovoltaic Systems

Kilian Sperber, Jonathan Kraus, Henry Meißner, Ulf Kreuziger

Ostbayerische Technische Hochschule Amberg-Weiden, 92224 Amberg, Germany

Keywords: Image segmentation, Photo Voltaic System, AI, U-Net, YOLO.

Abstract

The given contribution describes an efficient artificial intelligence (AI) workflow for the detection and segmentation of ground-mounted photovoltaic (PV) systems in Bavaria (Germany), which can be transferred to any region. A two-stage approach was developed based on digital orthophotos (DOP) with a resolution of 20 cm (DOP20) or 100 cm (DOP100). Two different AI models, U-Net and YOLO, are used to identify and segment PV systems. The combined approach, which first analyses low-resolution DOP100 images and then uses targeted high-resolution DOP20 tiles, increases efficiency, by processing only relevant image areas with high resolution. Initial tests in three Bavarian districts show a high level of accuracy for both AI models. The approach is designed to be used for area-wide segmentation in Bavaria and thus contribute to change detection and quality assurance of the Digital Basic Landscape Model (ATKIS®Base-DLM). Furthermore, the generalisation capability of the workflow was validated using an independent high-resolution dataset from the Piedmont region in Italy, where the models achieved promising recognition rates even without applying the post-processing pipeline.

1. Introduction

Updating and quality assurance of the Digital Basic Landscape Model (ATKIS®Base-DLM) of the Bavarian Surveying Administration, Germany, is essentially based on information from area topographers, who regularly have all the geodata of the State Office for Digitization, Broadband and Surveying (LDBV) at their disposal. However, as the scope and depth of detail of information to be recorded continues to increase, the area topographers need to be supported in future by methods of artificial intelligence (AI) to fulfill their task maintaining LDBV's geo-database. Changes detected with the presented AI-workflow can be targeted on site by the area topograph, so that only a verification is necessary or a field trip may become obsolete.

For this reason, the research cooperation *GeoDataObjektDektor* was established between OTH Amberg-Weiden and the LDBV. The aim of the research cooperation is automated detection of changes in various topography objects in the ATKIS®Base-DLM using the LDBV's basic geo-database and satellite data with the aid of AI.

Using AI workflows is not new in the LDBV, as automated building detection for construction investigation and monitoring using AI methods was already realized a few years ago (Roschlaub et al., 2020). AI is also already being used in geoinformatics in other German federal states. A prominent example is the derivation of land cover using the Cop4ALL process, which originated in North Rhine Westphalia (Sandmann et al., 2022). Evaluation of actual land use in the ATKIS®Base-DLM using deep learning has already been investigated by (Reiter et al., 2023). In that article, a Convolutional Neural Network (CNN) was trained using digital orthophotos with a resolution of 20 cm (DOP20) in order to classify the actual land use.

In contrast to the work by (Reiter et al., 2023), this paper presents the detection and segmentation of individual object types using the example of ground-mounted photo-voltaic

installations. A two-stage AI workflow will be presented, in which DOP100-tiles showing ground-mounted photovoltaic systems are first identified so that only a fraction of an area is subsequently segmented with DOP20-tiles. This AI workflow is verified using two different AI architectures, U-Net (Ronneberger et al., 2015) and YOLO (Redmon et al., 2016) and compared in each case with ground truth segmentation of all DOP20-tiles.

2. Data Basis

2.1 Digital Orthophoto

The LDBV provides a wide range of official geodata. The DOP serves as the data basis for the training and evaluation of various AI models, which is recorded throughout Bavaria in a two-year aerial survey cycle. Since 2024, the aerial survey has been carried out with a longitudinal and transverse coverage of at least 80 and 55 % respectively. The images are created in both the red, green, blue and infrared colour space (RGBI) and panchromatic (PAN) (LDBV, 2024). For the detection and segmentation of ground-mounted photovoltaic systems, the DOP-RGB with a ground resolution of 20 cm or 100 cm is initially used, as this has a higher degree of information than the DOP-PAN.

2.2 Training Data

High-quality and heterogeneous training data is crucial for the training of different AI models. When selecting the training data, attention is therefore paid to an even distribution of the following aspects: Recording date (season), installation type (structure, colour), orientation, exposure, vegetation and region in Bavaria. The training data sets are composed of DOP-RGB tiles with a total area of approx. 250 km², which are evenly distributed throughout Bavaria and are available at a resolution of 20 or 100 cm.

The models are trained with different resolutions, it has already been shown that segmentation of photovoltaic systems in high-resolution images by a model trained on low-resolution images,

as well as segmentation in low-resolution images by a model trained on high-resolution images (cross-application), does not provide satisfactory results (Jiang et al., 2021). To train the model based on the DOP20 dataset, the photovoltaic modules are carefully digitised manually, whereby the subsurface of ground-mounted photovoltaic systems is not segmented. PV systems on buildings are explicitly not taken into account. The aim is to extract only the photovoltaic modules in order to ensure the most precise training basis possible. The data set comprises a total of around 20,000 geo-objects, which are split into training and validation data at a ratio of 80/20. In order to avoid potential falsification of the training data, buildings in the DOP are blacked out based on their building perimeters, as photovoltaic modules on building roofs are not digitised. This is necessary because the visual characteristics of photovoltaic modules on ground-mounted and roof-mounted systems are very similar. Without these measures, there could be inconsistencies in the training data, which could affect the accuracy of the model. This method is only used for U-Net, as it did not improve the results for YOLO.

A different approach is used to train the model based on the DOP100 dataset. Due to the lower spatial resolution, an exact differentiation of individual photovoltaic modules is no longer possible. Therefore, the modules are aggregated to ensure an appropriate representation. In contrast to the DOP20 model, buildings are not blacked out in this approach, as photovoltaic modules on house roofs are difficult or impossible to recognise in the lower resolution of DOP100 tiles.

Detection and segmentation of ground-mounted photovoltaic systems/modules is basically a binary segmentation task with the classes (PV/no PV), which is why black and white masks are created from the (aggregated) geo-objects and the DOP tiles using the Feature Manipulation Engine (FME). The RGB images and black and white masks are then tiled to the size required by the respective model (see Fig. 1).

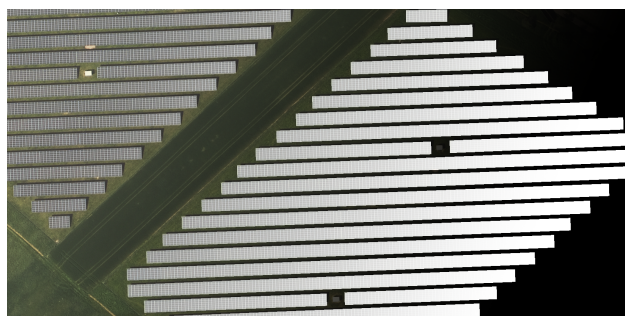


Figure 1. DOP RGB20 image fading to corresponding mask.

To increase robustness of the model and expand the amount of training data, additional image variations are generated using various image augmentation methods (Xu et al., 2023). In addition to intensity changes, a geometric change based on rotation of the images and masks by 90 degrees is also implemented as image augmentation. This measure is used to simulate the different orientations of photovoltaic modules, which is only possible to a limited extent through the selection of training data alone, as photovoltaic modules are predominantly orientated south in northern hemisphere (especially when installed north of tropic of Cancer). The choice of 90° rotation is also technically motivated, as other rotation angles are associated with a loss of sharpness in the image, which would have a negative effect on quality of the training data set. The loss of sharpness

occurs with rotations that are not divisible by 90° intervals, as these require a re-sampling of the pixel positions in the image, which causes blurring and distortion (Parker et al., 2007).

2.3 Test Areas

To determine the generalisation capability of the models, three test areas throughout Bavaria are selected that do not contain any training data. In order to use a sufficiently large sample, entire administrative districts are selected. Selection is based on several criteria that characterise them as suitable.

A high proportion of photovoltaic systems in these areas compared to other districts is essential in order to test the models for many different types of ground-mounted photovoltaic systems and to be able to identify weak points. Counties with a large number of ground-mounted photovoltaic systems are therefore filtered on the basis of ATKIS@Base-DLM dataset.

Size of the districts is also a decisive criterion. By selecting large-scale test areas, a greater variety of landscape and development types can be covered. This diversity is important in order to test flexibility of the model so that it can also be applied to larger areas with different topographies and land uses.

Using standard geoinformation system (GIS) operations, a list of possible candidates for test areas is compiled with the help of the training data set, the ATKIS@Base-DLM and the administrative units in Bavaria. The district of Bad Kissingen (Lower Franconia, area: 1136 km²), the district of Regensburg with the independent city of Regensburg (Upper Palatinate, area: 1472 km²) and the district of Ostallgäu with the independent city of Kaufbeuren (Swabia, area: 1434 km²), have been found as best fit (see Fig. 2).

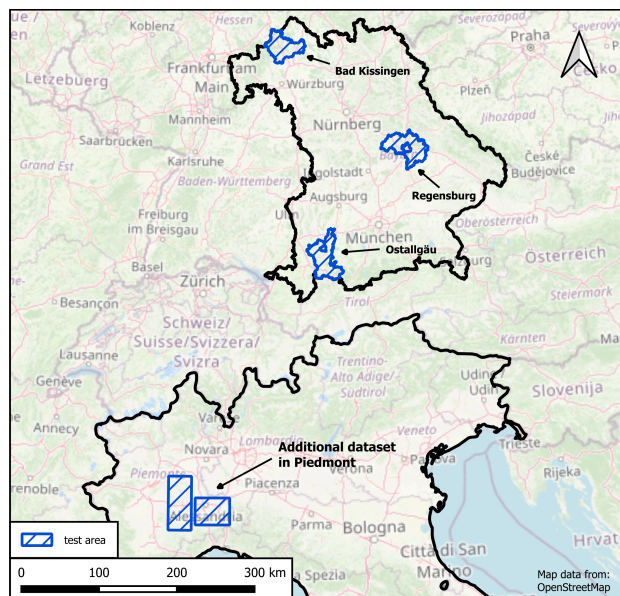


Figure 2. Test Areas.

The three selected test areas offer a representative distribution across Bavaria, both geographically and in terms of landscape. These regions differ significantly in their vegetation, topography and settlement structure, which makes it possible to test the performance of the models in different environmental contexts. In addition, the aerial images from Bad Kissingen and Regensburg are from 2023, while the data from the district of Ostallgäu is from 2022.

3. Methodology

3.1 U-Net

A CNN-based U-Net is used for the segmentation of the ground-mounted photovoltaic systems. This network architecture was originally used by (Ronneberger et al., 2015) for the segmentation of neuronal structures and cell units in microscope images. However, this architecture has also been successfully adapted to segmentation and classification tasks in remote sensing, for example, (Knopp et al., 2022) developed a U-Net for the classification of burnt areas in Sentinel-2 images.

The U-Net architecture follows a symmetric encoder-decoder design. In this structure, input images of the dimension 256×256 pixels are initially processed with three channels each. The encoder consists of five convolutional blocks, with each block comprising two layers: one layer consists of a convolution layer with a 3×3 kernel followed by a dropout layer. The image size remains unchanged through the use of mirror padding. At the end of each convolution block, the image resolution is reduced by a factor of 4 by means of a pooling layer, while the number of feature maps per block is doubled at the same time.

In the decoder section, a symmetrical structure to the encoder is used, whereby the resolution of the image data is increased by a factor of 4 per block using deconvolution layers. The number of feature maps is halved in the process. To compensate for the loss of detailed information, copy-and-crop operations (skip connections) are used to transfer the features of the encoder block to the corresponding decoder block. This enables a more precise reconstruction of the image information.

Finally, a 1×1 convolutional layer with a sigmoid activation function is used to generate a probability map that assigns a value between 0 and 1 to each pixel. Subsequent thresholding yields the distinction between PV or no PV class.

3.2 YOLO

YOLO (You Only Look Once) is a well-known model for object recognition that was first introduced in 2016 and has since been further developed in various versions. It is based on a unified CNN that formulates the object recognition process as a single regression problem. In contrast to older methods such as R-CNN, which first propose regions and then classify them, YOLO performs both the detection and localisation of objects in one pass, which leads to faster result finding (Redmon et al., 2016). This paper uses YOLOv8x-seg, an extended version of the YOLOv8 model that is specifically adapted to segmentation tasks. It was introduced in 2023. This model combines the proven architecture of YOLO with a special extension for mask-based segmentation of objects. The 'x' variant is designed for larger or complicated models with more parameters and therefore higher accuracy. YOLOv8x-seg makes it possible not only to recognise objects, but also to segment them with pixel accuracy (Ultralytics, 2024). Although newer YOLO versions have already been introduced, YOLOv8 will continue to be used in this work for the time being, as it is a stable, well-documented and widely used version (RUiNtheExtinct, 2024).

3.3 Post Processing

Results of the two AI architectures undergo post-processing after segmentation. This step is fully automated. In case of U-Net, the black and white masks are first vectorised, while

results for YOLO are already available as polygons. The geo-objects are then blended with the respective test county in order to clean up the results of cross-county DOP-tiles. Holes in the geometries are also closed in the process. By subsequently buffering the features by the value 0.2 m followed by aggregating and buffering back by the same value, the transitions between two DOP-tiles can be cleaned up. The transitions between individual snippets are already taken into account by vertical and horizontal overlapping. A further intersection eliminates erroneously segmented photovoltaic-objects that are located on houses, traffic areas, railway areas or bodies of water. The existence of ground-mounted photovoltaic systems on these areas is excluded. The processed geo-objects then pass through a small area filter so that features with an area of less than 8m^2 are filtered out. The erroneous filtering of very small, isolated PV modules is tolerated. Result of the post-processing stage is geo-referenced features.

3.4 Classic approach

The classic approach to segment ground-mounted photovoltaic systems is the complete processing of all DOP20-tiles of a test area. These are first given by the AI model. Segmented tiles then go through the post-processing process, resulting in the georeferenced features. The complete process chain is shown in Fig. 3.

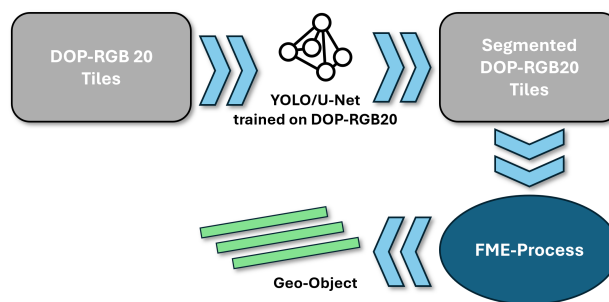


Figure 3. Classic approach process chain.

3.5 Combined approach

The combination approach differs from the classic approach as the test area is first fed to the DOP100 model in metre resolution. The advantage lies in the faster processing of the DOP100-tiles, as only 1/25 of the data volume has to be processed compared to the DOP20-tiles. A significance test of the segmented DOP100-tiles can be used to determine the tiles that are subsequently processed with the DOP20 model. This means that only a fraction of the test area needs to be segmented with the more computationally intensive DOP20-tiles. These then run through the classic approach in the same way. The AI workflow can be seen in Fig. 4. This approach therefore promises a significant increase in efficiency.

3.6 Evaluation of Test Areas and Key Parameters

Testing is an essential part of the development and validation of AI systems, especially when applied to large-scale and complex data sets such as Bavaria-wide geodata and other European datasets. It is used to check the performance and reliability of a model in real application scenarios. Proper testing ensures that the AI model delivers reliable results not only in the training data, but also in previously unknown regions and under different environmental conditions. This is particularly important

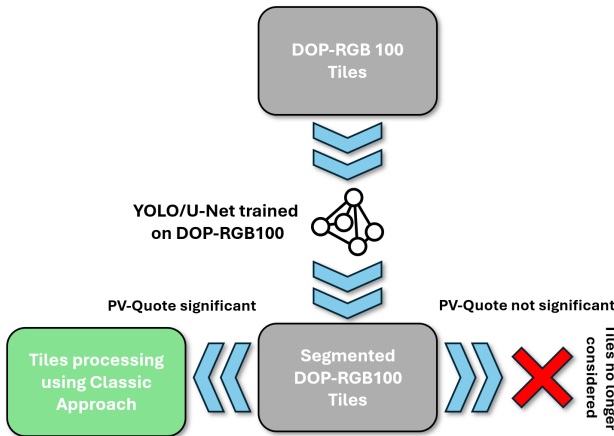


Figure 4. Combined approach process chain.

when applied to the Bavaria-wide aerial images, as the system has to recognise a large number of photovoltaic system types in different regions that were recorded under different conditions.

The key parameters that are determined by testing describe the accuracy that an end user obtains when applying to unknown data sets. After segmenting the three test areas, results run through the post-processing process and saved as a geo-referenced shapefile. By intersecting the results with the corresponding manually digitised ground truth ('target'), the common key parameter "precision" and "recall" can be determined in terms of area. A tolerance value of 0.3 m (1.5 pixels) is implemented for the intersection in order to minimise the influence of digitisation inaccuracies in the training and test data.

The following mathematical relationships apply to the intersection:

$$FP = \text{Prediction} \setminus \text{Ground Truth} \quad (1)$$

$$TP = \text{Prediction} \cap \text{Ground Truth} \quad (2)$$

$$FN = \text{Ground Truth} \setminus \text{Prediction} \quad (3)$$

Where FP stands for False Positive, TP for True Positive and FN for False Negative. This can be used to calculate the precision and recall ('completeness') key parameters as well as the harmonic mean of both (F1 Score) (Murphy, 2022):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

The aim is to fully record all ground-mounted photovoltaic installations in the respective test area in order to draw conclusions about the change. However, these three key parameters do not provide any information as to whether all PV parks in a test area have actually been recorded. A PV park with a small area may not have been identified even with a recall of well over 90%. For this reason, the key parameter 'Identified parks' is introduced at this point. This is determined by aggregating the

individual PV modules in the prediction and in the ground truth to form individual PV parks. The intersection over union (IoU) of the aggregated parks can thus be used to determine the number of identified parks or the recognition rate ('RR'). A park is classified as identified if its IoU is at least 20%:

$$\text{IoU} = \frac{\text{Prediction} \cap \text{Ground Truth}}{\text{Prediction} \cup \text{Ground Truth}} \quad (7)$$

$$\text{IoU} \geq 0.2 \rightarrow \text{PV Park identified} \quad (8)$$

$$\text{IoU} \leq 0.2 \rightarrow \text{PV Park not identified} \quad (9)$$

$$\text{RR} = \frac{\text{PV Park identified}}{\text{PV Park identified} + \text{PV Park not identified}} \quad (10)$$

Since the two approaches described above are compared in terms of their efficiency (savings rate), a further key figure is required which describes the proportion of segmented tiles T on which there is no ground-mounted photovoltaic system.

$$\text{Efficiency} = 1 - \frac{T_{\text{negative} \cap \text{segmented}}}{T_{\text{negative}}} \quad (11)$$

Assuming that a test area consists of 1200 tiles, of which 200 tiles representing the ground truth, then consequently the remaining 1000 tiles do not contain ground truth (T_{negative}). If 150 of these tiles are erroneously segmented ($T_{\text{negative} \cap \text{segmented}}$), this results in an efficiency of 85%. Although an efficiency of 100% means that no tile without a photovoltaic system was segmented, it cannot be deduced from this that all tiles with ground-mounted photovoltaic systems were segmented. All key parameters must therefore always be considered as a whole.

4. Results

This section compares performance of the two models used, YOLO and U-Net, for segmenting PV systems. The comparison includes both quantitative metrics and a qualitative evaluation of the segmentation results. Training and validation metrics are not presented here, as they do not adequately describe the generalisation capability of the model. Instead, the user accuracies described above are determined using the three test areas Bad Kissingen (KG), Regensburg (R) and Ostallgäu (OAL). An NVIDIA Quadro RTX 5000 is used as the graphics card for both network architectures.

4.1 U-Net

The two trained U-Net models (DOP20 model and DOP100 model) are used for both workflows (Fig. 3 and 4). The segmentation of a district in DOP100 resolution takes approx. 3 minutes, whereby the size of a district only plays a subordinate role in terms of time. The segmentation of a DOP20-tile takes approx. 2.2 seconds, so that a county of 1500 km² takes about 55 minutes to process. The results (see Table 1) show that both approaches are always above 90% in terms of precision, recall, F1 score and recognition rate, with the exception of precision and therefore the F1 score in the Ostallgäu district for the classic approach. The detection rate is 100% for both approaches. In contrast, the efficiency is always 0% for the classic approach, as this approach segments all the more computationally intensive DOP20-tiles. The average IoU is more than 96% in each case. The results of the three test areas are then averaged for each approach. An exemplary segmentation result of a ground-mounted photovoltaic system is given in Fig. 5.

Admin. District	Precision	Recall	F1-Score	RR	Efficiency	Ø IoU
Bad Kissingen (KG)						
Classic	97,38 %	96,11 %	96,74 %	100,00 %	0,00 %	97,51 %
Combined	99,36 %	96,00 %	97,65 %	100,00 %	99,27 %	97,22 %
Regensburg (R)						
Classic	93,62 %	97,61 %	95,57 %	100,00 %	0,00 %	98,28 %
Combined	97,13 %	97,59 %	97,36 %	100,00 %	98,05 %	98,28 %
Ostallgäu (OAL)						
Classic	82,48 %	98,94 %	89,96 %	100,00 %	0,00 %	94,66 %
Combined	95,54 %	98,48 %	96,99 %	100,00 %	96,81 %	92,99 %
Mean						
Classic	91,16 %	97,55 %	94,09 %	100,00 %	0,00 %	96,82 %
Combined	97,34 %	97,36 %	97,33 %	100,00 %	98,04 %	96,16 %

Table 1. Results U-Net.



Figure 5. Segmentation result of a ground-mounted photovoltaic system (left: DOP20, centre: U-Net result, right: YOLO result).

4.2 YOLO

For the YOLO architecture, the two trained models are also used for workflows (Fig. 3 and 4) and evaluated using the three test areas in Bavaria. It takes around 5 to 6 minutes to segment a county in DOP100 resolution. The processing of a DOP20-tile can be estimated at around 3.5 seconds, meaning that a county of 1500 km² can be processed in around 88 minutes. The results (see Table 2) are similar to the U-Net results. All key parameters are above 90% (with the exception of efficiency in the classic approach). The detection rate is also 100% for both approaches. The average IoU is over 97%. The averaged key parameters for both approaches are also calculated for this network architecture. An exemplary segmentation result of a ground-mounted photovoltaic system is given in Fig. 5.

4.3 Comparison of Classic and Combined Approach

In order to compare the two workflows, the averaged key parameters of both approaches are compared. It can be seen that the precision is slightly higher with the combined approach than with the classic approach, while the recall and the average IoU are only slightly worse. As a result, the F1-score is also slightly better with the combined approach. However, the recognition rate is identical for both approaches. The biggest difference in the key parameters is the efficiency.

5. Additional Experiments on the Piedmont Dataset

5.1 Generalisation Experiment

To further investigate the robustness and generalisation capability of the developed workflow, additional experiments were conducted on an external dataset covering photovoltaic installations in the Piedmont region of Italy (Compagnia Generale Riprese et al., 2023). This dataset differs substantially from the Bavarian DOP20 and DOP100 imagery used in the primary study. In particular, the Piedmont aerial data exhibits a coarser ground sampling distance of 30 cm, and its annotations consist of more than 9.000 manually verified PV polygons, including additional metadata such as installation category.

The experiments on this dataset serve a dual purpose. First, they allow an evaluation of cross-regional transferability under substantially different imaging conditions. Second, they reflect realistic use cases for large-scale solar infrastructure monitoring within the European Union (EU), where the contribution of solar power to the energy mix has reached approximately 12% in 2022 due to increased deployment of PV systems (Compagnia Generale Riprese et al., 2023). Automated, scalable mapping solutions are therefore essential for agencies and stakeholders aiming to track the expansion of renewable energy infrastructure and assess its spatial distribution.

Admin. District	Precision	Recall	F1-Score	RR	Efficiency	Ø IoU
Bad Kissingen (KG)						
Classic	94,51 %	98,45 %	96,44 %	100,00 %	0,00 %	98,11 %
Combined	94,95 %	98,22 %	96,56 %	100,00 %	99,84 %	97,30 %
Regensburg (R)						
Classic	92,50 %	98,18 %	95,26 %	100,00 %	0,00 %	98,37 %
Combined	94,16 %	97,56 %	95,83 %	100,00 %	98,05 %	98,22 %
Ostallgäu (OAL)						
Classic	91,56 %	98,28 %	94,80 %	100,00 %	0,00 %	96,99 %
Combined	93,14 %	97,78 %	95,41 %	100,00 %	98,81 %	97,03 %
Mean						
Classic	92,86 %	98,30 %	95,50 %	100,00 %	0,00 %	97,82 %
Combined	94,08 %	97,85 %	95,93 %	100,00 %	98,90 %	97,52 %

Table 2. Results Yolo.

Notably, the tests were performed *without* re-training on the new resolution, *without* domain adaptation, and *without* applying the post-processing pipeline used in the main experiments (tile merging, geometric filtering, small-area removal). Thus, the results presented here reflect raw model output, based on the classic approach, under more challenging and unconstrained conditions.

5.2 YOLO11-m Training Configuration

For the Piedmont dataset evaluation, the newer YOLO11-m segmentation model was trained using the following configuration. Only the most relevant arguments are shown for clarity, while preserving reproducibility.

```
task: segment
model: YOLO11-seg_3_channel_m.yaml
epochs: 120
batch: 24
imgsz: 512
pretrained: false
translate: 0.1
fliplr: 0.5
erasing: 0.4
auto_augment: randaugment
weight_decay: 0.0005
```

The training procedure was run for 120 epochs with a batch size of 24 and an input resolution of 512 px. The default YOLO11 loss formulation (box, segmentation mask, class confidence, and the distribution-focused loss) was used. No pretrained weights were employed, ensuring that the model learned the task from scratch based solely on the imagery of Bavaria described in section 2.2.

5.3 YOLO11-m Training Metrics

Across 120 epochs, the YOLO11-m model exhibited consistent improvement in both localisation and segmentation performance. The following trends summarise the training dynamics:

- **Mask segmentation loss** decreased from 4.15 at epoch 1 to 0.34 by epoch 120.

- **Box regression loss** improved from 3.99 to 0.31 over the same period.
- **Mean Average Precision (mAP)** for masks (M) reached:
 - $mAP_{50} \approx 0.96$
 - $mAP_{50-95} \approx 0.82$
- **Precision and recall (mask mode)** converged to:
 - Precision ≈ 0.93
 - Recall ≈ 0.93
- **Learning rate** decayed from $1.0 \cdot 10^{-2}$ to $1.8 \cdot 10^{-4}$

5.4 Results on the Piedmont Dataset

After training, the YOLO11-m model was applied to the full Piedmont test imagery using the classic approach. Importantly, no post-processing was performed on the predictions. Under these conditions, the model identified:

- **6.037 of 9.462 annotated PV polygons**
- Recognition rate: **63.80%**
- Area-level coverage: **69.02%**

For comparison, the previously trained YOLOv8 model achieved:

- **6.226 of 9.462 annotated polygons**
- Recognition rate: **65.74%**
- Area-level coverage: **65.80%**

These results are particularly noteworthy given the substantial differences in spatial resolution (30 cm vs. 20 cm), the absence of post-processing, and the lack of adaptation to regional characteristics. Achieving raw recognition rates above 60% and area coverage nearing 70% demonstrates strong cross-regional

generalisation and suggests that the learned visual representations of PV modules extend well beyond the Bavarian domain.

Overall, this experiment highlights the robustness of YOLO-based segmentation architectures and supports the applicability of the proposed workflow to other European regions without extensive calibration or re-annotation efforts.

6. Discussion

6.1 Data Basis

High accuracy of the results clearly shows the quality and consistency of the training and test data used. The more precise and consistent the data-sets, the better performance of the model. Due to deliberately measures such as the blackening of house roofs in U-Net, the classification could be further optimised, as the model processes less contradictory data as a result. To make the models even more flexible in the future, historical DOPs can be used as a supplementary source of training data. These make it possible to train the model for several flight periods, as only data from the years 2022 and 2023 has been used so far.

6.2 Evaluation of the Combined Approach

The pre-selection of DOP tiles using a DOP100 model has proven to be a particularly suitable approach for increasing efficiency. This technique has significantly increased the speed of analysis while maintaining almost the same quality. The key parameter 'Efficiency' for the combination approach is always over 96%.

In addition, a significant reduction in misclassification (false positives) could be achieved through pre-selection, which can be seen above all in precision. Thanks to this approach, the U-Net architecture makes it possible to fully segment PV systems throughout Bavaria over a period of approx. 5 hours. The only disadvantage of the combination approach is that very small parts of PV systems, which are located at the edge of some tiles, are not found by the DOP100 model and therefore isolated tiles are not segmented in DOP20 quality. However, larger structures are reliably recognised so that this has no influence on the overall recognition of the respective PV park.

To analyse this effect numerically, the recall values of the combination approach and the classic approach can be compared. The recall is at most 0.6% higher with the classic approach. The combination approach is at a slight disadvantage in this respect, which is also reflected in a slightly lower average IoU.

However, the major advantage is the following: complete detection of all PV systems with at least 96% higher efficiency, which is accompanied by considerable time savings.

6.3 Network architectures

Another focus of the study was the analysis of network architectures. While the differences between U-Net and YOLO are small, some insights can be derived from the results. YOLO achieves more consistent results than U-Net, delivering almost identical metrics across all test areas. This indicates better generalization capabilities due to YOLO's larger model architecture. U-Net, on the other hand, is subject to slight fluctuations, which is particularly noticeable in the precision in the Ostallgäu district, where, at 82.48%, this value differs significantly compared to other values for the classic approach. In the Ostallgäu

district, in particular, a large number of very small areas are incorrectly segmented. This is due, among other things, to incorrectly segmented mountain structures. Rather, the combination approach delivers a reliable increase in precision without a significant decrease in recall. The precision of U-Net is on average



Figure 6. Example YOLO prediction and ground truth for the Piedmont dataset.

over 3% higher than that of the YOLO model using the combination approach. The classic approach shows similar behavior, although the precision of U-Net is slightly lower on average. This is primarily due to the results in the Ostallgäu district. In general, YOLO tends to generate individual masks larger than the actual objects (see Fig. 6). By choosing the low tolerance value of 0.3m when blending the results, parts of the prediction are considered false positives, which negatively impacts precision. In addition to the improved precision results, the U-Net solution offers significantly faster data processing due to its relatively simple architecture. While it processes a district in an average of 6 minutes, the YOLO solution takes 13 minutes. U-Net is thus almost 55% faster than YOLO. Both models are suitable for use in the Bavarian Surveying Administration. The detection rates are consistently 100% in all Bavarian test areas. This means that all PV systems are detected.

6.4 Piedmont Dataset

Although the numerical performance indicators appear comparatively low at first glance (recognition rate: 63.8%–65.7%; area-level coverage: 65.8%–69.0%), a visual inspection reveals that a substantial portion of the missing detections corresponds to rooftop PV installations (see Fig. 7, bottom). This effect is fully expected, as rooftop systems were neither included in the training data nor explicitly learned by the models (see Section 2.2).

It should also be emphasized that no post-processing was applied in this experiment (see Section 3.3). In the Bavarian workflow, incorrectly segmented objects located on buildings, transportation infrastructure, railway areas, or water bodies are removed through an additional intersection step. Consequently,



Figure 7. Missing ground truth annotations but found with YOLO11-m (top), Roof mounted PV panels and corresponding ground truth (bottom).

raw predictions — such as those evaluated for Piedmont — naturally contain more false positives and unfiltered artefacts.

A further observation concerns the incompleteness of the Piedmont ground truth. The YOLO11-m model detected several PV parks that were not annotated in the reference dataset (see Fig. 7, top). This highlights a limitation of the benchmark and suggests that the true recognition rate may be higher than reported.

Considering these factors — particularly the substantially different ground resolution of 30 cm instead of 20 cm — the results are highly promising. They demonstrate that the trained models not only transfer effectively to other geographic regions (including different federal states or countries), but can also be applied to imagery with different spatial resolutions.

7. Conclusion and Outlook

The presented workflow demonstrates, that ground-mounted photovoltaic systems in Bavaria can be efficiently segmented over large-scale areas by combining DOP100 and DOP20 data using two different AI network architectures. The results derived can be used by the LDBV for quality assurance and updating the ATKIS®Base-DLM but also in other countries, as data from Piedmont shows. Next steps plan to implement the entire segmentation of Bavaria in DOP20 quality. Furthermore, it would be appropriate to apply the trained models and the developed approach to data from other federal states and countries. This would further evaluate the generalization capabilities of the models and could lay the foundation for broader application. The method for combining data at different spatial resolutions also offers potential for transfer to other object types of the ATKIS®Base-DLM. This would allow additional land-use types to be recorded efficiently and with high resolution.

Acknowledgements

We would like to thank the LDBV and the Bavarian State Ministry of Finance and Home Affairs for providing the geodata

and for their fruitful and constructive research cooperation.

References

- Compagnia Generale Riprese, Pasquali, P., Arnaudo, E., Blanco, G., 2023. Piedmont photovoltaic panels dataset, IEEE Dataport.
- 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. *Earth System Science Data Discussions*, 2021, 1–17.
- Knopp, L., Wieland, M., Rättich, M., Martinis, S., 2022. Ein Deep Learning Ansatz zur Klassifikation von Brandflächen in Sentinel-2 Aufnahmen. *Künstliche Intelligenz in der Geodäsie*, 95–110.
- LDBV, 2024. Landesamt für Digitalisierung, Breitband und Vermessung. Bayernbefliegung. Seit 1987 aktuelle Luftbilder von ganz Bayern. <https://www.ldbv.bayern.de/produkte/luftbilder/orthophotos/bayernbefliegung.html>. Last access July 2025.
- Murphy, K. P., 2022. *Probabilistic machine learning: an introduction*. MIT press.
- Parker, J. A., Kenyon, R. V., Troxel, D. E., 2007. Comparison of interpolating methods for image resampling. *IEEE Transactions on medical imaging*, 2(1), 31–39.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.
- Reiter, D., Meinel, G., Forkel, M. et al., 2023. Evaluierung der Tatsächlichen Nutzung im ATKIS Basis-DLM mittels Digitaler Orthophotos und Deep Learning. *ZfV-Zeitschrift für Geodäsie, Geoinformation und Landmanagement*.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18, Springer, 234–241.
- Roschlaub, R., Li, Q., Auer, S., Möst, K., Glock, C., Schmitt, M., Shi, Y., Zhu, X. X., 2020. KI-basierte Detektion von Gebäuden mittels Deep Learning und amtlichen Geodaten zur Baufallerkundung. *ZfV-Zeitschrift für Geodäsie, Geoinformation und Landmanagement*.
- RUintheExtinct, 2024. YOLO Models — An objective comparison. Summary. <https://ruintheextinct.medium.com/yolo-models-an-objective-comparison-1566e6b6dcb7> (3 July 2024).
- Sandmann, S., Hochgürtel, G., Piroška, R., Steffens, C., 2022. Cop4ALL NRW—Ableitung der Landbedeckung in Nordrhein-Westfalen mit Fernerkundung und künstlicher Intelligenz. *ZfV-Zeitschrift für Geodäsie, Geoinformation und Landmanagement*.
- Ultralytics, 2024. Ultralytics YOLOv8. Documentation. <https://docs.ultralytics.com/models/yolov8/> (3 July 2024).
- Xu, M., Yoon, S., Fuentes, A., Park, D. S., 2023. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137, 109347.