This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

Deep Learning Framework for Multi-Class Segmentation of Photovoltaic Systems

Svea Krikau, Sina Keller

***Abstract*—Accurate mapping of photovoltaic (PV) installations supports renewable energy monitoring, capacity assessment, and policy planning. We introduce a deep learning framework for the simultaneous segmentation of rooftop (RTPV) and ground-mounted (GMPV) PV systems from $0.2\,\mathrm{m}$ aerial imagery. The approach integrates automatically generated labels from registries and crowdsourced geographic information data with a small set of manually refined annotations, enabling scalable training with high spatial fidelity. A dual-head U-Net++ with a ResNet50 encoder jointly performs pixel-wise segmentation and image-level classification, reducing false positives in heterogeneous landscapes. Exemplarily applied to the state of Hesse, Germany, the model achieves a balanced accuracy of $87.52\,\%$, precision of $94.74\,\%$, and a recall of $82.89\,\%$, with F1-scores of $75.75\,\%$ (RTPV) and $88.65\,\%$ (GMPV). Coupled with solar suitability data, the results indicate rooftop PV utilization of only $3.24\,\%$ statewide and $2.5\,\%$ in the Dietzenbach case study. This implemented deep learning framework is the first high-resolution, single-step multi-class PV segmentation applied at the federal state scale, offering a transferable tool for large-area PV monitoring.**

***Index Terms*—classification, photovoltaic, remote sensing, solar energy potential, semantic segmentation, supervised learning.**

I. Introduction

THE emission of greenhouse gases, especially carbon dioxide, for example, due to the combustion of fossil fuels, is a major driver of the globally changing climate [1]. Overcoming this challenge demands a collective commitment to move towards clean and renewable energy sources. For example, the German government aims to achieve greenhouse gas neutrality by 2045 with an intermediate milestone of reducing emissions by $65\,\%$ by 2030 [2].

Achieving these targets requires the adoption of renewable energy technologies such as solar power, hydro power, biomass, or geothermal systems [3]. Among these technologies, solar energy has been identified as "one of the most promising renewable energy sources to meet the future global energy demand", according to [4]. In particular, photovoltaic (PV) technology holds significant promise, as it can be easily installed by specialized organizations and individuals in residential households using small single-unit PV units for personal use [5]. However, there are still "some barriers to overcome, including a complex regulatory framework, policy goals, costs, lack of consumer information, and standards" [6]. The German Renewable Energy Sources Act (EEG) 2023 plans for PV expansion to 215 GWp by 2030 and 400 GWp by 2040 [7]. However, current information on actual implementation is difficult to obtain, as official records are often incomplete [8], making it difficult to monitor and plan the achievement of environmental goals.

S. Krikau (svea.krikau@kit.edu) and S. Keller (sina.keller@kit.edu) are with the Institute of Photogrammetry and Remote Sensing (IPF) at the Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany.

Accurate and up-to-date mapping of existing PV installations is crucial for assessing deployment progress, identifying untapped solar potential, and supporting data-driven energy planning. These aspects require scalable and automated methods to detect and distinguish photovoltaic systems in varying landscapes and types of installations.

In this context, deep learning (DL) methods for semantic segmentation have emerged as powerful tools to identify PV infrastructure in aerial and satellite imagery. However, as discussed in Section II, many existing approaches are limited in scope, focusing on rooftop (RTPV) or ground-mounted photovoltaic (GMPV) systems separately, or relying heavily on manually labeled data that limit scalability and generalizability.

Contributions of This Study

This study presents a DL-based framework for the joint segmentation and classification of RTPV and GMPV systems to address the limitations mentioned above (see Figure 2 in Section III). Unlike previous approaches, our DL framework leverages automatically generated labels and supports multi-modal input data, enabling accurate and scalable mapping of PV installations in diverse environments.

The main contributions are summarized as follows.

- We propose a *high-resolution semantic segmentation pipeline* that uses automatically generated training labels to detect photovoltaic systems at a spatial resolution of $0.2\,\mathrm{m}$. Our approach achieves precision values that exceed $90\,\%$ in two geographically distinct and independent datasets, which cover a wide range of landscape types such as high-density urban areas, rural settlements, agricultural fields, and forested regions.
- We integrate a *classification level within the segmentation* architecture to differentiate between RTPV and GMPV installations in a single unified process, achieving a balanced accuracy of $87.52\,\%$.
- The *scalability of the entire framework* is demonstrated through its application to large-scale regions with varying urban and rural characteristics, highlighting its potential to support the monitoring and planning of PV at the national level.
- We evaluate the *potential solar energy generation* based on the detected RTPV systems, revealing that only a fraction of the theoretically available capacity is currently being utilized.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

II. Related Work

Rapid advancement of machine learning (ML) techniques has significantly transformed the field of image analysis. In particular, DL approaches have proven highly effective for automated image interpretation, enabling precise object detection [9] and semantic segmentation [10] in diverse image domains. In this context, the identification of PV systems can be formulated as an image classification, object detection, semantic segmentation, or instance segmentation task.

Early work on PV detection relied on traditional computer vision techniques that required manual feature engineering [11]. These methods demonstrated the potential of remote sensing and aerial imagery for detecting PV systems, but environmental conditions, installation types, and variations in image quality often limited their effectiveness.

A. Multi-scale and multi-type challenges in automated PV system recognition

Identifying PV systems in aerial imagery remains a challenging task, even for human annotators, due to their visual diversity, varying spatial scales, and installation contexts. One reason is that PV arrays vary significantly, ranging from large, centralized ground-mounted systems to small, distributed rooftop installations. These systems are classified differently according to the application context, e.g., by capacity (kWp) [12], as small-scale (residential) versus large-scale solar [8], centralized or distributed [13], or alternatively, as RTPV and GMPV [14] for image classification tasks.

Although individual small-scale rooftop units generate relatively modest amounts of electricity, their cumulative contribution is substantial, accounting for approximately $40\,\%$ of the globally installed solar PV capacity [15]. In Germany, approximately $52\,\%$ of the installed capacity consists of distributed small-scale PV systems [12]. They often have subtle visual cues that are difficult to distinguish from other rooftop structures. Moreover, the background context varies significantly between urban and rural environments, adding further complexity to the segmentation task and contributing to a higher number of false positives within these areas [16]. Large-scale PV installations are predominantly located worldwide in homogeneous terrains, such as croplands, desert regions, or grasslands [17]. Variations in orientation, shading, panel configuration, weather conditions, and the presence of nearby objects introduce additional complexity. In addition, PV systems typically occupy only a fraction of the image area, compounding these issues and making them prone to false positives and missed detections in automated pipelines.

B. Reference data used in PV systems segmentation studies

Official registries are often used to obtain information on the current number and capacity of installed PV systems. However, these records can be unreliable due to missing entries, as demonstrated by [8], even though the data is generally considered credible. In contrast, crowdsourced volunteered geographic information data, such as OpenStreetMap (OSM), often provides more comprehensive coverage than official registries. However, the quality and completeness of crowdsourced data can vary significantly in terms of contributor coverage and quality, leading to substantial geographic variability [18]. The complexity of PV panel appearances, the varied installation environments, and the labor-intensive nature of manual annotation all contribute to the scarcity of well-labeled, publicly available datasets. Few researchers publish their annotated datasets, which limits model development and benchmarking. Some focus exclusively on large-scale PV plants [19], [20], while others target small-scale rooftop PV installations [21], [22] or face variable accuracy due to reliance on OSM polygons [18]. Among these, only the dataset by [14] provides sufficiently high-quality data covering both RTPV and GMPV, allowing the identification of not only entire plants but also individual panels. As illustrated in fig. 1, the automatic extraction of PV installations from crowdsourced sources, such as OSM, has notable limitations and is unsuitable as a test dataset when the objective is to obtain detailed information down to the individual panel level. In addition to the lack of high-resolution labels, high-resolution image data are not universally available. However, some countries (such as the United States of America, the United Kingdom of Great Britain, and Germany) provide orthophotos (with ground sampling distances (GSD) between $0.1\,\mathrm{m}$ to $1\,\mathrm{m}$), making them more suitable for detailed PV mapping tasks.

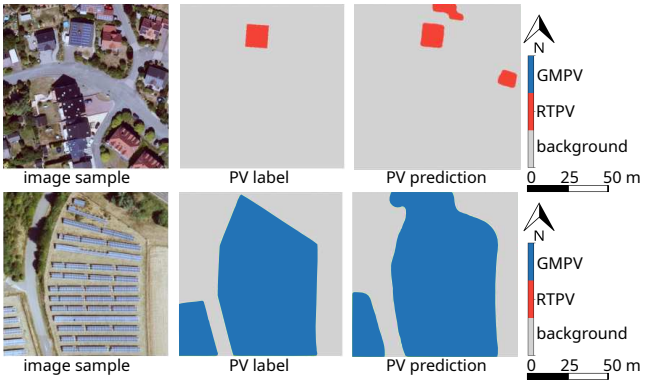

Fig. 1. Automatically generated labels often lack sufficient detail, particularly for standalone ground-mounted photovoltaic (PV) systems. These labels often include background pixels and fail to accurately capture all PV installations. The figure is adopted from [23]. The following abbreviations are used: GMPV – ground-mounted PV, RTPV – rooftop PV.

C. DL for the automatic identification of PV systems

The review by [13] presents a wide range of approaches for identifying PV systems and provides a detailed overview. In the underlying study, the focus is specifically on DL-based semantic segmentation models for detecting various types of PV systems. In general, high-resolution datasets (as shown in table I) outperform lower-resolution ones when fine structural details of PV systems are important. Although most studies focus on RTPV or GMPV systems, few address both simultaneously. For example, [14] examined a large-scale region that covered an entire province in China, but did not distinguish between PV types. Only one study, [16], has been identified that performed a multi-class segmentation of RTPV

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

TABLE I
SUMMARY OF SELECTED DATASETS AND SEMANTIC SEGMENTATION PERFORMANCE FOR HIGH-RESOLUTION PHOTOVOLTAIC SYSTEM IDENTIFICATION. DATASETS ARE SORTED BY GROUND SAMPLING DISTANCE (GSD).

Source	**GSD** in m	**Type**	**ROI**	**Input**	**Dataset*** # of Datapoints	**Classes**	**P** [%]	**R** [%]	**IoU** [%]	**F1** [%]
[14]	0.10	RTPV	Jiangsu Province, China	RGB (UAV)	645 (manual)	binary	92.8	89.4	86.8	91.1
[26]	0.10	RTPV	Shanghai, China	RGB (aerial)	1000 (manual)	binary	93.9	94.2	88.7	94.0
[27]	0.15	RTPV	Heilbronn, Germany	RGB (Google)	5866 (manual)	binary	96.4	95.9	92.6	96.2
[16]	0.26	RTPV+ GMPV	Taihuyuan Town, Hangzhou, China	RGB (various sources)	1000 (manual)	binary +**	97.0	94.2	91.58	95.6
[28]	$\sim$ 0.20	RTPV	Oldenburg, Germany	RGB (Google)	1325 (manual)	binary	86.0	76.0	68.0	81.0
[14]	0.30	RTPV+ GMPV	Jiangsu Province, China	RGB (aerial)	2308 (manual)	binary	95.9	93.1	90.8	94.5
[26]	0.30	RTPV	California, USA	RGB (aerial)	526 (manual)	binary	86.2	83.5	73.6	84.8
[14]	0.80	RTPV+ GMPV	Jiangsu Province, China	RGB (Gaofen-2, Beijing-2 Sat.)	763 (manual)	binary	87.7	85.7	79.0	86.7
[29]	2.00	GMPV	China	RGB (Gaofen-1 Sat.)	223 (manual)	binary	93.8	92.2	87.4	92.6

* Image size varies between datasets.
** Two-step process: binary classification, then rooftop-specific refinement.
P – Precision, R – Recall, IoU – Intersection over Union, F1-score – harmonic mean of P and R.

and GMPV systems. However, their two-step approach was limited to an urban area in Taihuyuan Town, China, where they first segmented rooftop and PV structures and then classified them into RTPV and GMPV categories.

Several key limitations persist. The limited availability of high-resolution benchmark datasets makes it challenging to compare developed approaches directly, since annotation quality can vary even among human annotators. Furthermore, differences in annotation objectives, such as distinguishing individual PV panels from entire PV plants, further complicate a consistent evaluation. As noted by [16], high inter-class similarity, such as between roads and rooftop elements or between PV systems and agricultural areas, can hinder the transferability of models to other regions not trained on (such as rural versus urban areas with varying land cover types). This challenge is further amplified when the available imagery differs in spatial resolution (as demonstrated by [14]).

In contrast to the studies summarized in (A) to (C), we specifically focus on a unified and scalable framework for jointly segmenting and classifying RTPV and GMPV systems. The approach combines automatically generated labels with manually annotated labels to segment PV systems from openly available high-resolution datasets. We demonstrate the applicability of the proposed DL framework in the entire region of Hesse, Germany ($21{,}115\,\mathrm{km}^2$). In addition, the labels used and generated [24], as well as the Python code [25], are freely available.

III. METHODOLOGY

The proposed methodology for simultaneously segmenting the RTPV and GMPV systems integrates heterogeneous geospatial data sources, automatic and manual annotation strategies, and a dual-head deep learning architecture. Figure 2 illustrates the proposed workflow, from the generation of multi-class PV labels and weighted patch sampling to the dual-head network architecture that produces a final multi-class PV map distinguishing RTPV, GMPV, and background for subsequent capacity and potential analyses.

A. DL model architecture

The model ('PVMultiSegNet') employs a dual-head architecture composed of a shared encoder and two task-specific branches: one for the pixel level semantic segmentation and one for image-level classification (Figure 2) of RTPV and GMPV systems. Multichannel orthophotos are the input data with a spatial resolution of 0.2 m with 512×512 pixels.

The backbone of the architecture is a `U-Net++` [30] implemented using the Segmentation Models PyTorch (SMP) library [31] in combination with PyTorch [32].

A `ResNet50` encoder [33], pre-trained on ImageNet, is used to extract deep spatial features from the input image and is shared between both heads. Empirical comparisons reported in the review by [13] indicate that increasing the depth of the ResNet architecture from 50 to 101 layers yields only marginal or negligible performance improvements. Given that this study uses a smaller training dataset compared to previous works, a simpler backbone is prioritized to mitigate the risk of overfitting. For this reason, transformer-based architectures, which typically require extensive data and computational resources, were not considered. Empirical comparisons by [14] show that although DeepLab v3+ [34] achieved slightly higher F1-scores at higher resolutions than the U-Net [35] based architecture, it tends to miss small or partially obscured PV areas, resulting in wider gaps between adjacent panels. To improve segmentation accuracy in complex and fine-grained

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

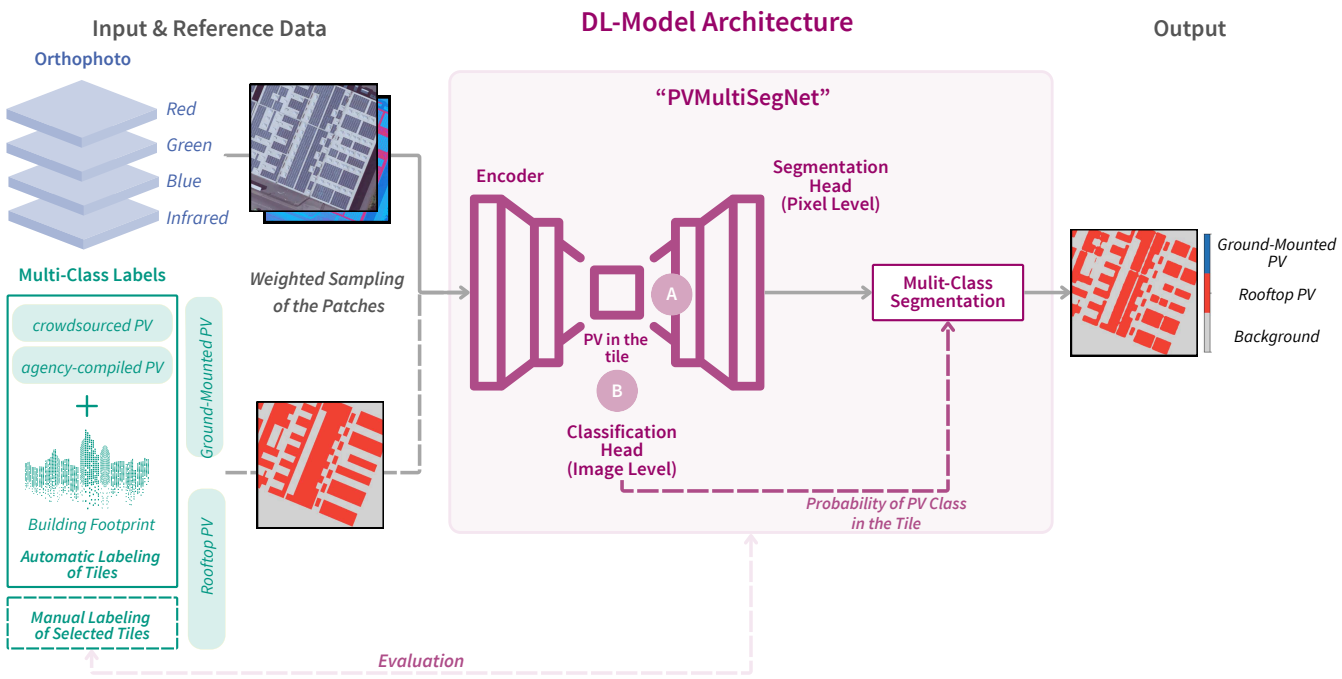

Fig. 2. Overview of the developed workflow for simultaneous segmentation of rooftop (RTPV) and ground-mounted (GMPV) photovoltaic (PV) systems from aerial imagery. The process combines multi-class PV label generation, weighted patch sampling, and a dual-head deep learning model to produce a multi-class PV map that distinguishes between RTPV, GMPV, and background for subsequent capacity and potential analyses. During training, the segmentation head (A) is trained independently of (B). During inference, (B) first checks for the presence of PV, and if positive, the pixel-wise segmentation by (A) is performed.

structures, such as PV systems, the `U-Net++` extends the base `U-Net` [35] framework by introducing dense skip connections between the encoder and decoder layers, enabling the precise delineation of object boundaries. Based on these characteristics and prior empirical findings, as well as visual inspection of different architectures during early training stages, `U-Net++` is selected as the core segmentation architecture for this study. Exploratory tests with alternative architectures were also conducted; although quantitative metrics are not reported here, the implementation is available [25] to allow reproduction of these experiments.

The segmentation head performs pixel-wise classification to identify PV structures, including RTPV and GMPV systems, while the detection head, composed of convolutional and fully connected layers, classifies entire images as PV or non-PV per class type. The fully trained model incorporates the output of the detection head to restrict segmentation to patches predicted to contain PV systems. It outputs pixel-wise class probabilities, which are converted into discrete labels by selecting the most probable class per pixel. This dual-task architecture helps filter out non-PV or ambiguous inputs before segmentation, thereby reducing computational overhead and minimizing false positives, particularly in complex environments such as agricultural regions. Depending on the application, the detection and segmentation heads can also be used independently.

B. Input samples and channel selection

For the development of the PV segmentation model, multi-spectral orthophotos (referred to as RGBI) with a spatial resolution of $0.2\,\mathrm{m}$, captured during the years 2021 and 2022 in Hesse, Germany [36] are used as input samples. We refer to this dataset as 'DOPHesse'. The OpenNRW dataset [37] from North Rhine-Westphalia (NRW) provides orthophotos at a spatial resolution of $0.1\,\mathrm{m}$ and is, therefore, down-sampled to match the resolution of the DOPHesse dataset.

Original orthophoto tiles, each with a resolution of $5{,}000 \times 5{,}000$ pixels, are divided into smaller patches of 512×512 pixels as input data of the model. Resampling to a spatial resolution of $0.2\,\mathrm{m}$ is performed, if necessary. These patches are extracted with a stride of 256 pixels.

C. Automatic generation of labels for the high resolution segmentation of PV systems

For training the DL model, sufficient reference data must be available. Therefore, we establish an automatic label generation step in combination with manual labeling.

1) Automatic labeling: For the automatic generation of labels for segmenting RTPV and GMPV systems, official registries, and crowdsourced data (here: OSM [38]) are combined. The data from the official registry are based on information from the 'Amtliches Topographisch-Kartographisches Informationssystem (ATKIS)' [39] dataset. This dataset is an official topographic and cartographic information system from the Federal Agency for Cartography and Geodesy. To ensure a consistent data format, the data are first harmonized, trimmed to the state boundary of Hesse, and then merged. During this merging process, 47 additional areas of the ATKIS dataset are added to the OSM data, as they contain more

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

detailed information about PV installations. Binary labels for PV systems are automatically generated based on the available data. To distinguish between RTPV and GMPV systems, these labels are intersected with information on building footprints. This process results in the creation of three distinct labels: 0 for the background, 1 for the RTPV systems, and 2 for the GMPV systems. In addition, information about GMPV systems is added from [40]

2) Manual labeling: Since PV labels and orthophotos originate from different acquisition times, variations in roof structures and PV systems may occur due to construction or renovations. This temporal discrepancy poses a challenge in generating consistent labels. Manual inspection of the labels revealed additional tasks to solve: some label boundaries are too large, fine structures are not adequately captured, and there are inaccuracies in aligning the labels with the underlying PV systems (see Figure 1, adapted from our previous work [23]). To address this task, 30 tiles for the region of Hesse [39] and the region of North Rhine-Westphalia (NRW) [37] are randomly selected and manually labeled. These relabeled tiles are automatically generated from the original labels to avoid overlapping information during model development. A total of 2,150 automatically labeled tiles and 56 manually labeled tiles are used to generate patches that serve as input samples to the model. The tiles containing the references are published [24], with filenames encoding information about the region, annotation type (manual or automatic), orthophoto coordinates, and corresponding year. The complete list of generated patches is provided with the code [25].

D. Dataset preparation and patch sampling strategy

Manually labeled patches are upsampled by a factor of 4 in the final dataset, and extensive data augmentation is used to improve the volume of training data. 33,423 automatically labeled (overlapping) patches and 32,317 manually relabeled patches (upsampled by a factor of 4) are provided. In total, 65,740 patches exist; 53,053 of those contain examples for PV systems.

Only manually created patches are used for validation and testing, with 30 % allocated to each. The remaining 40 % are reserved for training, in addition to the automatically created labels. The final evaluation metrics (see Table III) are calculated based on the test dataset. This dataset contains 1,013 background, 371 RTPV, 34 GMPV, and 9 mixed patches. In total, 748 patches are located in Hesse, and 679 patches are located in NRW. The pixel-level distribution in the test dataset includes 98.86 % background, 0.58 % RTPV, and 0.56 % GMPV, totaling about 374 million pixels.

Stratification is performed based on the presence of PV systems (with mixed areas labeled as GMPV within the original image tiles). For validation and testing, only non-overlapping patches are selected to ensure spatial independence. Data augmentation is applied exclusively during training. General augmentations include horizontal flipping and vertical flipping, each with a probability of 50 %, as well as random 90° rotations with the same probability. Additionally, Gaussian blur with a blur limit ranging from 3 to 7 is applied with a probability of 30 %. For RGB channels, color jittering is used to introduce variations in brightness, contrast, saturation, and hue, with respective intensity values of 0.2 for brightness, contrast, and saturation, and 0.1 for hue, all applied with a probability of 50 %. All input feature patches are standardized before being fed into the model to ensure consistency and improve training stability.

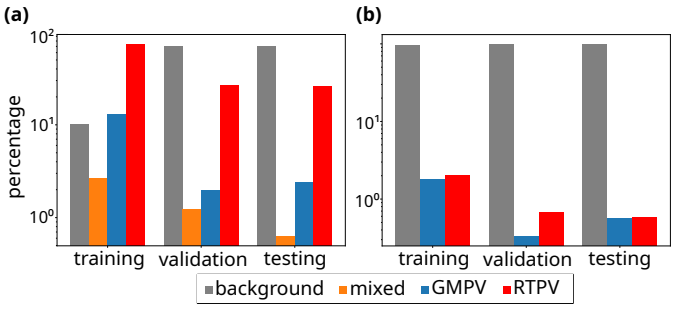

Fig. 3. Patch-wise (a) and pixel-wise (b) class distributions (background, rooftop (RTPV), and ground-mounted (GMPV) photovoltaic systems, and mixed patches) plotted with a logarithmic y-axis due to the dominant number of background patches/pixels.

Given the high prevalence of background pixels (see Figure 3) and the imbalanced class distributions at both patch and pixel levels, implementing a suitable sampling strategy is essential to ensure robust model training. To address the imbalance between pixels containing objects and those without objects, a weighted sampler is used during training to increase the frequency of samples with fewer objects; otherwise, the sampling is done randomly. Weighted sampling is applied on the basis of the pixel-wise occurrence of PV types within the images. For the segmentation task, manual labels are further amplified by a factor of 10. In the image classification task, the GMPV samples are increased by an additional factor of 4, while the RTPV samples are boosted by a factor of 2.

E. Training strategy

Model training is performed in two stages:

- The segmentation head performs pixel-wise classification to identify PV structures, including RTPV and GMPV systems, and is trained exclusively on regions containing PV examples.
- The detection head classifies entire images as PV or non-PV. This branch is trained on a broader dataset that includes both PV and background images.

In this way, the segmentation head leans towards capture fine-grained pixel-level details, while the classification head provides a broader image-level context to reduce false positives outside of the training dataset. This two-stage training approach was deemed necessary, as training the segmentation model with a high number of true negatives prevents the model from converging. However, true negative examples are necessary to prevent false positives, especially in non-urban areas, where few examples with PV systems exist.

1) Training the segmentation head: Initially, only training samples containing PV systems are used to train the decoder for the segmentation task; after two epochs, the encoder

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

is unfrozen. The hyperparameter tuning results in an initial learning rate of 5×10^{-5} with a weight decay of 0.001 using the Adam optimizer. A cosine annealing learning rate scheduler is applied with a maximum number of iterations of 50 epochs and a minimum learning rate of 1×10^{-6}. Training is carried out with a batch size of 32 for up to 25 epochs. Early stopping monitors validation loss with patience of 5 epochs and a minimum improvement threshold of 1×10^{-5}.

To address class imbalance and improve segmentation performance, particularly for underrepresented classes, a combined loss function comprising Tversky and Focal Loss is implemented. The selected parameters result from our focus on high precision (Figure S2). The Tversky Loss generalizes Dice Loss by introducing tunable weighting parameters $\alpha = 0.8$ $\beta = 0.2$, which control the penalty for false positives and false negatives. Focal loss further improves the performance on difficult or minority classes by down-weighting well-classified examples. This specific weighting is achieved here with a focus parameter $\gamma = 5.0$ and a class balancing factor $\alpha = 0.25$. The final loss is calculated as the weighted sum of both components, given by $\mathcal{L} = 0.5 \cdot \mathcal{L}_{\text{Tversky}} + 0.5 \cdot \mathcal{L}_{\text{Focal}}$.

2) Training the classification head: Subsequently, the detection head is trained on the full training dataset (including PV and non-PV samples) to classify whether a tile contains a specific PV system or not. To address the class imbalance between positive and negative samples, a weighted binary cross-entropy loss is applied, with a weight of 2 assigned to positive examples.

IV. Results

A. Qualitative results of the segmented PV systems in Hesse, Germany

The region used to develop and exemplarily apply the model is the state of Hesse, located centrally in Germany. The state is characterized by a wide range of land cover types, including urban and industrial areas, agricultural land, forests, and natural reserves, offering a diverse mix of built-up zones and open spaces, as well as a variety of PV system types. The southern part is flat with a high population density, including cities like 'Frankfurt am Main'. In contrast, the northern area is more hilly, with a lower population density and predominantly rural landscapes.

An overview of the qualitative results of the resulting PV segmentation can be found in Figure 7 and close-ups in Figure S4. From these, the following statements can be derived:

- The model tends to delineate boundaries more accurately for larger PV installations (approximately $\geq$5 m). However, smaller PV systems tend to be more challenging, with boundaries appearing less distinct or sometimes omitted.
- Objects with shapes or textures that resemble PV panels can occasionally lead to false detections, whereas high-contrast conditions generally improve the detection reliability.
- The model shows difficulties in classifying PV systems in atypical settings; for example, small ground-mounted installations in urban areas may be overlooked

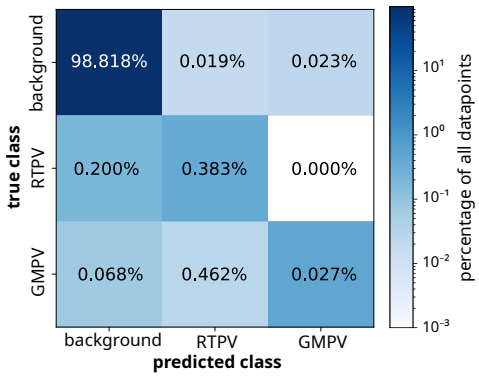

Fig. 4. Pixel-level confusion matrix for the final segmentation model, showing the percentage of the entire test set for background, rooftop (RTPV), and ground-mounted (GMPV) photovoltaic systems. The colorbar is on a logarithmic scale due to the high prevalence of background pixels.

- Classification inconsistencies: The model struggles with class assignment in atypical contexts, e.g., small ground-mounted PV in urban areas may not be recognized, while non-PV rooftops in rural areas can be misclassified as PV.
- Improved robustness in rural contexts: Despite some misclassifications, rural structures that resemble PV are generally not falsely detected.

B. Model configuration and segmentation performance

The trained segmentation head achieves an accuracy of $99.78\,\%$ with a precision and recall of $94.74\,\%$ and $82.89\,\%$ (Table II). Different channel configurations are tested as input for the segmentation model (see Table II and Figure S1). The additional infrared channel in the RGBI configuration improves model performance compared to RGB-only, although recall decreases slightly. However, adding another channel with information on the surface structure by including a normalized Digital Surface Model (nDSM) as a fifth channel does not improve model performance compared to RGBI. Due to our aim of minimizing false positives in the segmented PV systems, with an emphasis on maximizing precision, also select the hyperparameters of the loss function. Using the Focal Loss alone decreases the overall model performance, while Tversky Loss without the Focal Loss likewise under performs compared to the combined configuration (see S.I-B).

The detection head alone achieves an accuracy of $92.44\,\%$ with a precision and recall of $85.48\,\%$ and $89.00\,\%$ on the task

TABLE II
PV segmentation performances for different input data configurations and DL models: RGB only, RGBI with Infrared (RGBI), and RGBI combined with normalized Digital Surface Model (nDSM).

Metric in %	RGB	RGBI	RGBI+nDSM
Accuracy	99.74	**99.78**	99.71
Precision	87.66	**94.74**	88.80
Recall	**88.07**	82.89	83.43
IoU	79.14	**80.09**	76.43
F1-score	87.61	**88.08**	85.71

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

of detecting PV systems at the image level.

The dual-head architecture improves the precision of the overall pixel-level segmentation from 94.74 % to 95.32 % (see Table III) on the test dataset. However, this specific architecture slightly decreases the other metrics. Although the performance metrics are quite similar between the single segmentation model and the dual-head architecture, the dual-head design proves essential for deployment in regions such as agricultural zones with few or no PV installations. Figure 5 shows examples of correctly identified PV systems, while Figure 6 illustrates examples of false positives and other misclassifications.

Table III summarizes the class-wise segmentation performance, showing that while RTPV and GMPV are occasionally confused with each other, both exhibit a much higher rate of confusion with the background. The model achieves reasonably high precision for both classes, 89.30 % for RTPV and 95.20 % for GMPV. Most of the PV predictions are correct; the false positive rate is relatively low. However, recall reveals a clear limitation, particularly for RTPV, with scores dropping to 65.77 %, as also evident in the confusion matrix (Figure 4). In particular, RTPV is never misclassified as GMPV, whereas the reverse can occur; however, this seems overly optimistic. Overall, the results indicate a performance gap between PV system types, with RTPV segmentation remaining the most challenging.

C. RTPV Detection compared to official records in Kleve, NRW

Due to the slight underperformance of the RTPV another comparison to the official PV registry [37] of NRW is performed. The registry provides a point-based dataset for GMPV and RTPV installations across the federal state of NRW. For this study, the district of Kleve is selected as the reference region.

Since the official registry records installations at the address level, without indicating the exact rooftop locations, the comparison is conducted at the parcel ('Flurstück') level. Both the official point dataset and our segmented RTPV are spatially intersected with the cadastral parcel boundaries to determine the presence of PV systems.

121 parcels contain PV installations according to the official registry, whereas 1,110 parcels are identified as containing rooftop PV systems based on our segmentation approach. The intersection between the two datasets has 106 parcels, corresponding to an agreement of 87.6 %. Therefore, only

TABLE III
CLASS-WISE PERFORMANCE OF THE FINAL SEGMENTATION MODEL, INCLUDING EVALUATION METRICS FOR EACH CLASS AND THE OVERALL AVERAGE WITH A BALANCED ACCURACY OF 87.52 %.

Metric in %	Background	RTPV	GMPV	Overall (dual-head)
Accuracy	99.69	99.75	99.88	99.78 (99.77)
Precision	99.73	89.30	95.20	94.74 (95.32)
Recall	99.96	65.77	82.93	82.89 (82.29)
IoU	99.69	60.96	79.61	80.09 (79.81)
F1-score	99.84	75.75	88.65	88.08 (87.91)

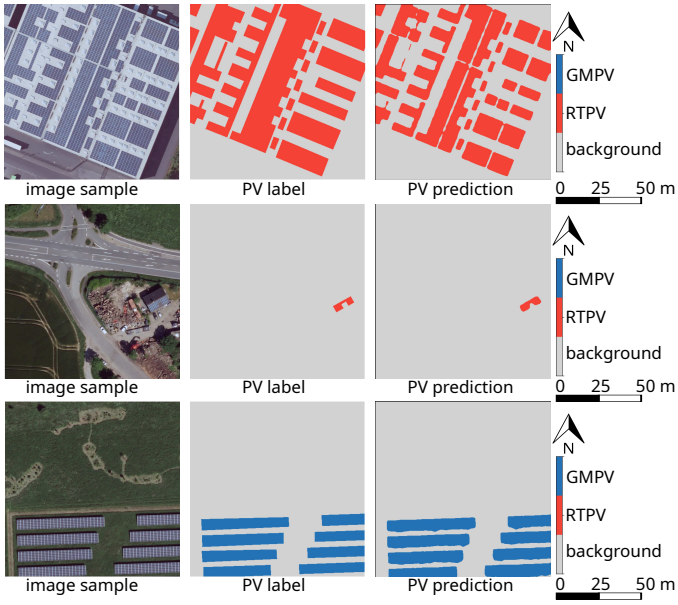

Fig. 5. Example cases of photovoltaic (PV) identification: Orthophoto (left), reference data (center), and automatically detected areas (right). Different PV installations are correctly identified by the model with fine detail. The following abbreviations are used: GMPV – ground-mounted PV, RTPV – rooftop PV.

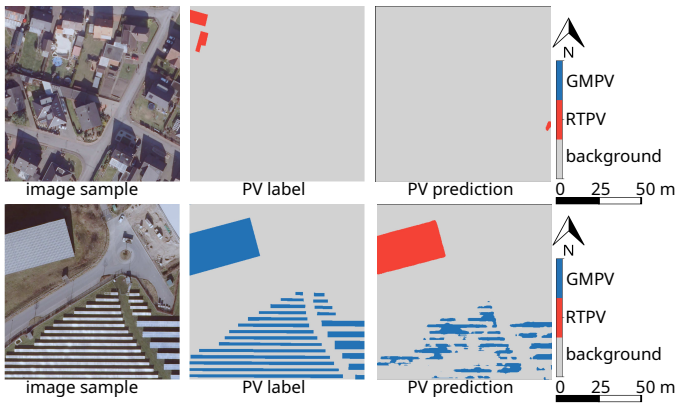

Fig. 6. Example cases of photovoltaic (PV) identification: Orthophoto (left), reference data (center), and automatically detected areas (right). Identifying narrow and small installations remains challenging. The following abbreviations are used: GMPV – ground-mounted PV, RTPV – rooftop PV.

9.55 % of the PV systems detected through our segmentation are also recorded in the official registry. Despite the low recall of the RTPV, the segmentation already provides an added value (manual inspection of the unregistered RTPV is performed to ensure their validity) by identifying additional, previously unregistered PV systems. Additional comparisons for the state of Bavaria can be found in the supplementary material S.II.

D. Additional analysis focused on installed photovoltaic capacity and energy generation in Hesse, Germany

After mapping the existing PV system, their current PV capacity and energy generation can be determined.

Therefore, we rely on the 'Solar-Kataster Hessen' data [41]. These data deliver global radiation for the entire state of Hesse at a spatial resolution of 0.5 m, derived from airborne laser

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

scanning data collected between 2015 to 2022. Building footprints from ATKIS are utilized to calculate capacity separately for building and non-building areas. Then, these are intersected with the model-segmented PV results to calculate the potential for roof and non-roof areas.

Figure 7 presents the available input data for a set of randomly selected tiles within Hesse, covering examples of urban, agricultural, and industrial areas. Note that in terms of these three types of area, most PV systems in Hesse are RTPV, accounting for $67.19\,\%$ of the total installed capacity according to available orthophotos from the years 2021 and 2022. Currently, only $3.24\,\%$ of the solar potential is utilized on the roofs of buildings, which corresponds to a possible capacity of $1{,}476.13\,\mathrm{MW}$, while GMPV has a capacity of only $720.82\,\mathrm{MW}$.

Furthermore, solar suitability data for buildings, based on the previously mentioned solar radiation data, are used for a detailed analysis of the city of Dietzenbach in Hesse, Germany. This dataset includes information for each building with the following parameters: building footprint area (in m^2), module area (in m^2), with 3D rooftop area receiving a global solar radiation of more than $850\,\mathrm{kW\,h\,m^{-2}}$ per year, with a $10\,\%$ deduction for unusable space, installed capacity (in kWp (here: p – peak) with the estimated PV system capacity based on the 3D rooftop area, assuming a module efficiency of $20\,\%$ and a performance ratio of $80\,\%$ and energy yield (in $\mathrm{kW\,h}$) with the projected electricity generation based on system capacity and average global radiation for the location.

Assessing the solar potential, exemplified by an area in Dietzenbach, reveals the contribution of installed RTPV systems while also indicating the remaining untapped potential for energy capacity (see Figure 8). The shown district, which includes both industrial and residential areas, has $40{,}467.87\,\mathrm{m}^2$ of RTPV, accounting for $2.22\,\%$ of the total 1.82 million m^2 of rooftop area. Given the lack of data on the solar suitability of GMPV installations in this dataset, the analysis focuses exclusively on building-mounted solar panels. The installed PV capacity reaches $5{,}446.52\,\mathrm{kW}$, making up $2.35\,\%$ of the total $231{,}866.39\,\mathrm{kW}$ capacity. In terms of energy generation, the RTPV systems in Dietzenbach produce approximately 5.77 million $\mathrm{kW\,h}$ annually, which corresponds to $2.38\,\%$ of the total of 242.89 million $\mathrm{kW\,h}$. Although some industrial areas in the eastern part of the city have solar panels installed, the available roof area remains underutilized, particularly on rooftops with high potential for solar energy generation.

V. Discussion

A. Integrating RTPV and GMPV segmentation in a single DL-framework: challenges and outcomes

Our developed multi-class, pixel-wise segmentation model achieves an overall balanced accuracy of $87.52\,\%$, with an overall precision of $94.74\,\%$ and recall of $82.89\,\%$. The segmentation of RTPV systems is challenging due to the variability in roof structures, materials, and orientations, resulting in an F1-score of $75.75\,\%$ for this class. The model achieves a precision of $89.3\,\%$, while recall is comparatively lower at $65.8\,\%$. It is optimized for high precision to minimize false positives. Adjusting the hyperparameter to increase recall would detect more RTPV systems, but at the cost of additional false positives. Unlike GMPV systems, which typically exhibit larger and more uniform structures with an F1-score of $88.65\,\%$, rooftop installations tend to have more complex semantics, including shading effects and variable roof inclinations. These aspects and characteristics are challenging for segmentation DL approaches. By incorporating multi-class learning targets (background, rooftop, and ground-mounted) for PV systems into our framework at the data level, the model is compelled to learn the detailed differences in features, which enhances its performance. Although the confusion matrix indicates no misclassifications between true RTPV and predicted GMPV, applying the model across the entire federal state of Hesse reveals that such confusions do occur (S4 (a.iii)). This may indicate reduced generalization capacity or suggest that data representing such cases (e.g., urban areas with GMPV) are insufficiently represented in the training and test dataset. Additionally, green roofs with PV, which cover the entire image patch, could be misclassified as GMPV due to their visual similarity to typical ground surfaces. Providing more contextual information could mitigate this issue.

Many roofs and surfaces may appear similar in visible light images, but infrared wavelengths can reveal differences in material properties. While infrared imagery helped distinguish materials and improved metrics, adding an elevation channel did not yield noticeable gains. Many rooftops and surfaces confused with PV systems share similar heights, limiting the discriminative value of elevation. Moreover, temporal discrepancies between the orthophotos and elevation data can introduce inconsistencies, while coarse automatic labels may obscure fine details, making it challenging for the model to exploit subtle elevation differences.

Additionally, creating high-quality labels for training and evaluation is a time-consuming process. Manual annotation is particularly challenging due to the difficulty in distinguishing PV installations from other rooftop structures, such as skylights, air conditioning units, or dark-colored roofing materials. This ambiguity can lead to inconsistencies in training and evaluation data, potentially introducing noise into the learning process. As visible in Figure 5, in some cases, the human annotations are too coarse, and small gaps between PV panels get wrongly attributed as false negatives, even though the model correctly identifies panel gaps. Even for human annotators, the clear delineation of PV systems is not always straightforward, which increases the risk of mislabeling and highlights the need for a high-quality benchmark dataset for consistent model comparison.

Although automatic label generation significantly reduces manual effort, it tends to produce coarser annotations that may miss fine-grained structures, particularly for small or irregularly shaped rooftop PV systems. Mixing manual and automatically generated labels introduces noise, which could cause incorrect associations and reduce generalization. Evaluation is performed only on manually labeled data, and upsampling is applied exclusively to these labels. While the loss function does not explicitly target noisy labels, the combination of Tversky and Focal Loss improves robustness: Tversky controls penalties for both false positives and negatives, and Focal Loss

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

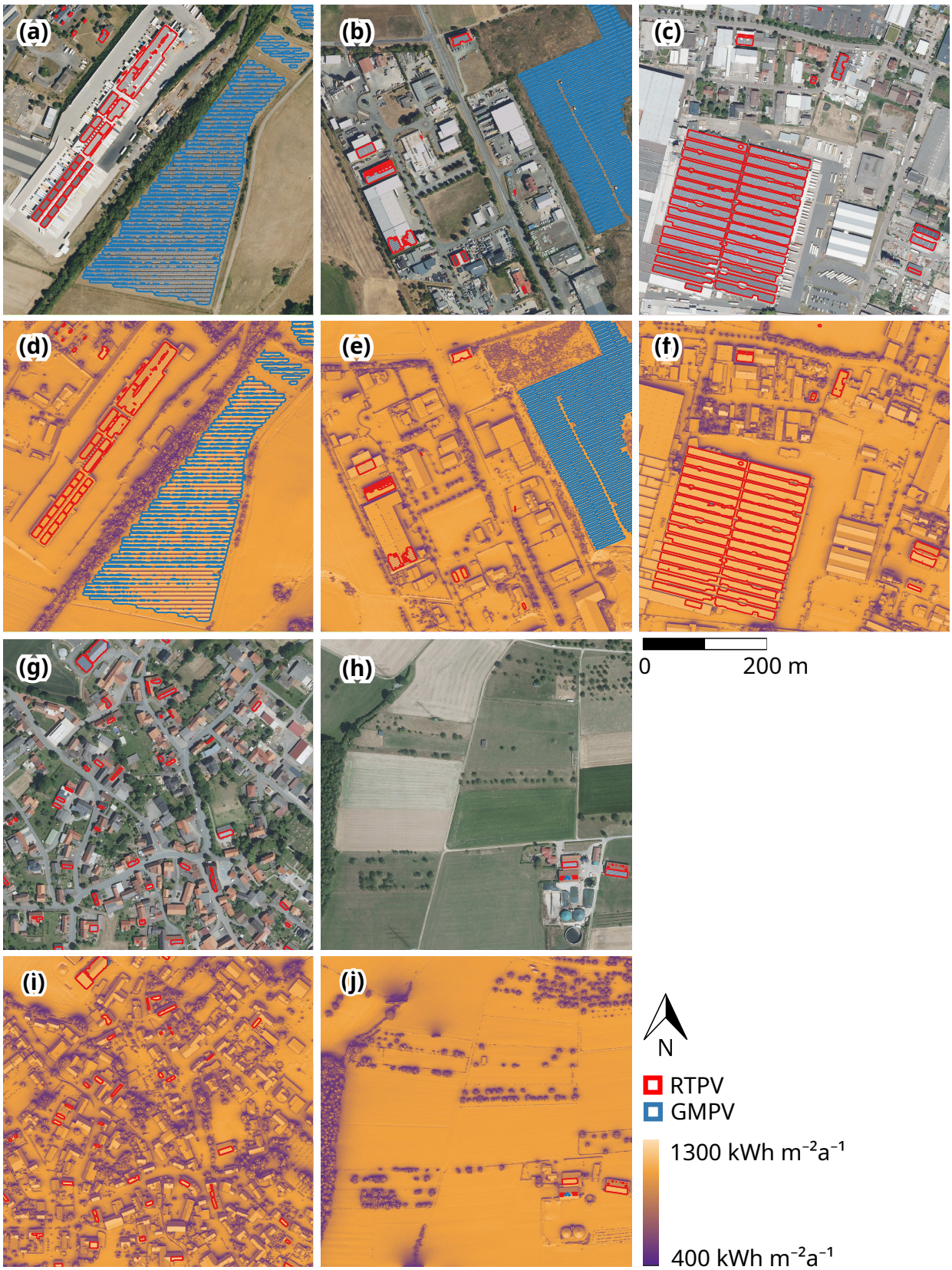

Fig. 7. The figure shows vector outlines of segmented rooftop photovoltaic (RTPV) (red) and ground-mounted photovoltaic (GMPV) (blue) systems overlaid on the corresponding orthophotos ((a) to (c) and (g) to (h)). For the same crop of each full-size image tile, corresponding values of global horizontal irradiance are shown in $\mathrm{kW\,h\,km^{-2}}$ ((d) to (f) and (i) to (j)), enabling spatial comparison between PV locations and local solar radiation potential. Digital Orthophotos © Hessian State Office for Land Management and Geoinformation, Germany (HVBG), 2021/22. Close-up views are provided in the supplementary material.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

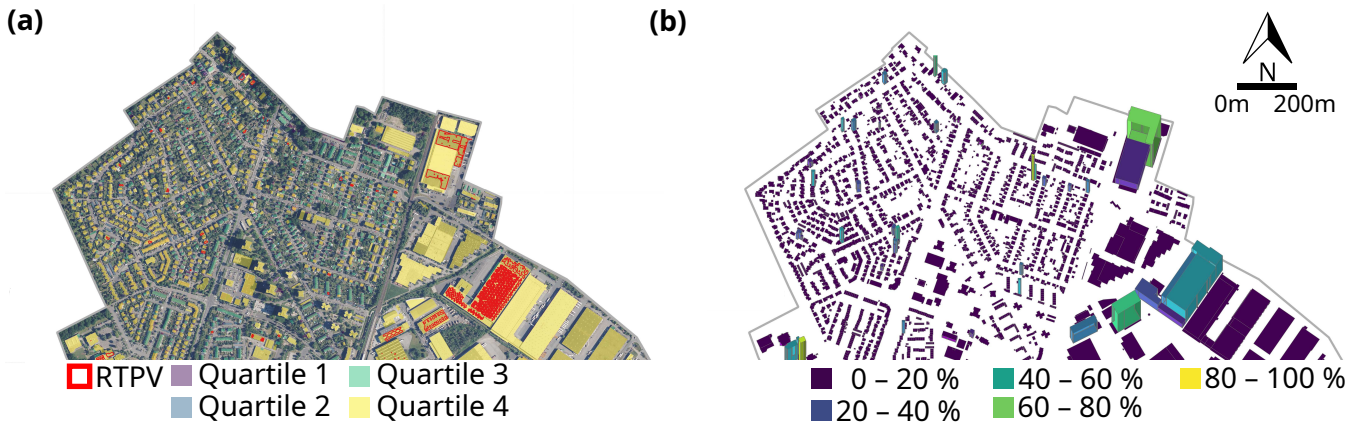

Fig. 8. Exemplary visualization of (a) the solar energy potential and rooftop photovoltaic (RTPV) systems (red outlines) in northern Dietzenbach, color-coded by quartiles. Large industrial rooftops, in particular, show high potential, falling within the fourth quartile. Buildings are color-coded and extruded based on roof area usage, with only 2.2 % currently utilized. (b) shows the notable expansion potential in the eastern industrial zones.

emphasizes hard or ambiguous pixels, which are often affected by noise. In the multi-task setup, the detection head uses automatic labels without disadvantage, providing additional noise-robust supervision. For segmentation, the probability of a PV type in the image is incorporated (currently via a simple threshold), and uncertain samples are weighted lower. While these strategies help mitigate noise, further improvements, especially confidence-based weighting, could be explored more in depth.

The results of this study indicate that manually refined annotations remain essential for the high-resolution segmentation of PV systems. Our hybrid approach strikes a balance between efficiency and segmentation accuracy. Adding more classes, such as floating PV systems, or implementing a categorization approach similar to [14], which differentiated rooftop data based on background information (e.g., flat concrete roof, steel tile roof, brick roof, cropland, and scrubland), could potentially further enhance model performance.

Although there is no benchmark dataset specifically for the multi-class PV segmentation, the OpenNRW dataset (without downsampling) has also been used by [42], who achieved precision and recall scores of 63.96 % and 86.69 % at a spatial scale of 0.1 m for PV detection. In comparison, our multi-class approach achieves on the image classification task precision and recall scores of 85.48 % and 89.00 %.

The recent study by [43] demonstrates that incorporating rooftop information enhances model performance, achieving a precision of 87.3 % and a recall of 87.5 % on the same dataset. This aligns with our findings that enhancing RGB images by including additional information, such as the infrared channel in our case, improves model performance. However, a major limitation is the availability of current orthophotos and the inconsistencies between different data sources. Direct comparisons with the metrics of other segmentation models (see Table I) are not straightforward due to differences in datasets, spatial resolutions, regional scales, and the number of classes. Our approach is currently the only one that performs segmentation of both RTPV and GMPV in a single step. By incorporating automatic labels, we were able to drastically reduce the manual annotation effort, using only 56 manually labeled image tiles compared to other approaches that required hundreds or even thousands (although differences in image scales and thereby the number of pixels exist).

B. Assessing the technical, regulatory, and practical potential of RTPV in Hesse

The energy transition targets in Hesse envision the use of photovoltaic systems covering approximately 1 % of the state's land area [44]. The current deployment still falls short of this goal, even though the utilization of available rooftop space already exceeds 3 %. The total possible capacity of 2,196.95 MW, based on the segmented PV systems from 2021 and 2022, is in a similar range to the figures reported by [44], which indicate approximately 2,777 MW for 2021. The solar suitability data for Dietzenbach represents the technical potential for solar energy. To prevent overestimation in municipal or regional assessments, a 10 % deduction is applied to exclude areas unsuitable for PV installations while considering spatial constraints such as shading and usability. However, this deduction remains an approximation and furthermore does not account for regulatory restrictions, such as building codes, heritage protection, or zoning laws, which define the regulatory potential and further refine feasibility. Although the slope orientation of the roofs is taken into account, the orientation of PV systems is currently not considered. Additionally, the economic and practical potential encompasses broader factors, including economic viability (e.g., infrastructure), social acceptance, and competing uses such as green roofs. Germany has a technical potential of approximately 504 TW h per year [45] for RTPV, with less than 10 % of rooftop potential currently utilized [7]. This aligns with our findings, where currently only about 2.5 % of the possible solar energy potential from rooftops in Dietzenbach is used.

VI. Conclusion

In summary, this study presents a pixel-wise, multi-class segmentation method for PV systems, which, unlike previous work, can differentiate between RTPV and GMPV installations

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

simultaneously in a single step. The model leverages RGBI channels from aerial imagery with a $0.2\,\mathrm{m}$ spatial resolution.

Our model achieves an overall balanced accuracy of $87.52\,\%$, with precision and recall of $94.74\,\%$ and $82.89\,\%$, respectively. Segmenting RTPV systems remains challenging due to the high variability in roof geometry, materials, and orientation, resulting in a comparatively lower F1-score of $75.75\,\%$ for this class. In contrast, GMPV installations, characterized by larger and more uniform surfaces, achieve an F1-score of $88.65\,\%$. The complex architectural and environmental factors affecting rooftop PV, such as shading and diverse roof inclinations, contribute to this difficulty. Our multi-class framework, which differentiates between background, rooftop, and ground-mounted PV, enables the model to learn fine-grained discriminative features that enhance overall segmentation performance. Benchmarking against existing work is complicated by dataset heterogeneity, differences in spatial resolution, and variations in classification schemes. Importantly, our method is unique in that it performs simultaneous segmentation of both RTPV and GMPV within a single framework. The incorporation of automatic labels enabled us to reduce manual annotation to 56 tiles, which is significantly less than other approaches that require hundreds or thousands of annotated images.

Data from two German federal states are used: the DOPHesse dataset from Hesse [39], which includes both automatically extracted and manually created labels, and the openNRW dataset from NRW [37], for which only manually annotated labels are available. Although automatic labeling greatly reduces manual effort, these labels tend to be coarser and may omit small or irregular RTPV components. Thus, manual refinement of annotations is essential to maintain high-resolution accuracy. Future work may focus on increasing class granularity (e.g., floating PV) or adopting hierarchical or contextual classification schemes.

From a broader perspective, Hesse's energy transition targets aim to cover roughly $1\,\%$ of the state's land area [44] with PV systems. Based on aerial imagery from 2021 to 2022, RTPV systems account for about $67.19\,\%$ of the total installed capacity within Hesse, with GMPV systems making up the remainder. Current utilization of rooftop solar potential is at approximately $3.24\,\%$, corresponding to a capacity of $1{,}476.13\,\mathrm{MW}$, while GMPV corresponds to a capacity of $720.82\,\mathrm{MW}$. A detailed solar suitability assessment for the city of Dietzenbach, Hesse, considers 3D rooftop areas, deducting for unusable spaces, but only for rooftop locations. This yields a technical potential capacity estimate that currently has only about $2.5\,\%$ utilization. Regulatory, economic, and social factors (such as zoning restrictions, heritage protection, and competing rooftop uses) are not accounted for here, but they are critical for refining the realistic deployment potential. Additionally, while roof slope and orientation are partially included, the precise azimuth of PV installations is not modeled, which may affect irradiance and energy yield estimates.

In total, this study demonstrates the feasibility and value of simultaneous multi-class segmentation for PV systems, providing a scalable tool to support accurate capacity assessment and inform strategic renewable energy planning. The final DL segmentation model [25] and the generated reference [24] are freely available for download.

ACKNOWLEDGMENTS

We thank Stefan Hinz, Head of the Institute of Photogrammetry and Remote Sensing at the Karlsruhe Institute of Technology, for co-funding this work. We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

REFERENCES

[1] H. L. Core Writing Team and J. R. (eds.), "IPCC, 2023: Summary for Policymakers," Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland, Climate Change 2023: Synthesis Report, Jul. 2023, edition: First.

[2] Bundesregierung, "Climate Change Act: climate neutrality by 2045." Jun. 2021, accessed: 2025-10-14. [Online]. Available: https://www.bundesregierung.de/breg-en/service/archive/climate-change-act-2021-1936846

[3] T.-Z. Ang, M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, and N. Prabaharan, "A comprehensive study of renewable energy sources: Classifications, challenges and suggestions," *Energy Strategy Reviews*, vol. 43, p. 100939, Sep. 2022.

[4] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun, and K.-H. Kim, "Solar energy: Potential and future prospects," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 894–900, Feb. 2018.

[5] Bundesregierung, "Mehr Photovoltaik mit Solarpaket," Oct. 2024, accessed: 2025-10-14. [Online]. Available: https://www.bundesregierung.de/breg-de/aktuelles/solarpaket-photovoltaik-balkonkraftwerke-2213726

[6] C. Lamnatou, C. Cristofari, and D. Chemisana, "Renewable energy sources as a catalyst for energy transition: Technological innovations and an example of the energy transition in France," *Renewable Energy*, vol. 221, p. 119600, Feb. 2024.

[7] H. Wirth and F. ISE, "Aktuelle Fakten zur Photovoltaik in Deutschland," Apr. 2025.

[8] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States," *Joule*, vol. 2, no. 12, pp. 2605–2617, Dec. 2018.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Jun. 2016, pp. 779–788, iSSN: 1063-6919.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[11] J. M. Malof, R. Hou, L. M. Collins, K. Bradbury, and R. Newell, "Automatic solar photovoltaic panel detection in satellite imagery," in *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, Nov. 2015, pp. 1428–1431.

[12] G. Lettner, H. Auer, F. Andreas, D. Schwabeneder, B. Dallinger, F. Moisl, E. Roman, D. Velte, and A. Huidobro, "Existing and future PV prosumer concepts," PVP4Grid, Tech. Rep. D2.1, 2018.

[13] H. Mao, X. Chen, Y. Luo, J. Deng, Z. Tian, J. Yu, Y. Xiao, and J. Fan, "Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images," *Renewable and Sustainable Energy Reviews*, vol. 179, p. 113276, 2023.

[14] H. Jiang, L. Yao, N. Lu, J. Qin, T. Liu, Y. Liu, and C. Zhou, "Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery," *Earth System Science Data*, vol. 13, no. 11, pp. 5389–5401, Nov. 2021, publisher: Copernicus GmbH.

[15] S. Joshi, S. Mittal, P. Holloway, P. R. Shukla, B. Ó Gallachóir, and J. Glynn, "High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation," *Nature Communications*, vol. 12, no. 1, p. 5738, 2021.

[16] Q. Qi, J. Zhao, Z. Tan, K. Tao, X. Zhang, and Y. Tian, "Development assessment of regional rooftop photovoltaics based on remote sensing and deep learning," *Applied Energy*, vol. 375, p. 124172, Dec. 2024.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3640268

[17] L. Kruitwagen, K. T. Story, J. Friedrich, L. Byers, S. Skillman, and C. Hepburn, "A global inventory of photovoltaic solar energy generating units," *Nature*, vol. 598, no. 7882, pp. 604–610, 2021.

[18] D. Stowell, J. Kelly, D. Tanner, J. Taylor, E. Jones, J. Geddes, and E. Chalstrey, "A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK," *Scientific Data*, vol. 7, no. 1, p. 394, 2020.

[19] A. M. Moradi Sizkouhi, M. Aghaei, S. M. Esmailifar, M. R. Mohammadi, and F. Grimaccia, "Automatic Boundary Extraction of Large-Scale Photovoltaic Plants Using a Fully Convolutional Network on Aerial Imagery," *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 1061–1067, Jul. 2020, conference Name: IEEE Journal of Photovoltaics.

[20] A. Ortiz, D. Negandhi, S. R. Mysorekar, S. K. Nagaraju, J. Kiesecker, C. Robinson, P. Bhatia, A. Khurana, J. Wang, F. Oviedo, and J. L. Ferres, "An Artificial Intelligence Dataset for Solar Energy Locations in India," *Scientific Data*, vol. 9, no. 1, p. 497, 2022.

[21] J. Wang, X. Chen, W. Shi, W. Jiang, X. Zhang, L. Hua, J. Liu, and H. Sui, "Rooftop PV Segmenter: A Size-Aware Network for Segmenting Rooftop Photovoltaic Systems from High-Resolution Imagery," *Remote Sensing*, vol. 15, no. 21, p. 5232, 2023.

[22] A. N. Wu and F. Biljecki, "Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability," *Landscape and Urban Planning*, vol. 214, p. 104167, 2021.

[23] S. Keller, S. Benz, M. Jehling, D. Böhnke, G. Hager, S. Wursthorn, C. Gallacher, D. Reiter, L. Hänchen, F. Kistner, S. Krikau, A. Schenk, and S. Hinz, "Entwicklung von planungshilfen für klimaschutz und klimaanpassung in der räumlichen gesamtplanung mittels fernerkundung - abschlussbericht," Hessisches Ministerium für Wirtschaft, Energie, Verkehr, Wohnen und ländlichen Raum, Tech. Rep., 2025.

[24] S. Krikau and S. Keller, "Photovoltaic systems: Reference data on rooftop and ground-mounted installations," 2025.

[25] S. Krikau, "PVMultiSegNet: Framework for Simultaneous Segmentation of Rooftop and Ground-Mounted Photovoltaics," 2025.

[26] Y. Jie, X. Ji, A. Yue, J. Chen, Y. Deng, J. Chen, and Y. Zhang, "Combined Multi-Layer Feature Fusion and Edge Detection Method for Distributed Photovoltaic Power Station Identification," *Energies*, vol. 13, no. 24, p. 6742, 2020.

[27] L. Zhou, K. Dong, H. Tan, J. Li, Q. Yu, Z. Guo, and J. Yan. A High-Precision Method for Photovoltaic Panel Segmentation Combining Large-Scale Model Prior Knowledge and Multimodal Information.

[28] M. Zech and J. Ranalli, "Predicting PV Areas in Aerial Images with Deep Learning," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020, pp. 0767–0774.

[29] Y. Jie, A. Yue, S. Liu, Q. Huang, J. Chen, Y. Meng, Y. Deng, and Z. Yu, "Photovoltaic power station identification using refined encoder–decoder network with channel attention and chained residual dilated convolutions," *Journal of Applied Remote Sensing*, vol. 14, no. 1, p. 016506, 2020.

[30] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi, Eds. Cham: Springer International Publishing, 2018, pp. 3–11.

[31] P. Iakubovskii, "Segmentation models pytorch," https://github.com/qubvel/segmentation_models.pytorch, 2019.

[32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., Dec. 2019, no. 721, pp. 8026–8037.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778.

[34] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Springer International Publishing, 2018, pp. 833–851.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[36] Hessian Administration for Land Management and Geoinformation, "Digital orthophoto (dop20)," 2023, accessed: 2025-08-11. [Online]. Available: https://www.hvbg.hessen.de/

[37] L. I. und Technik Nordrhein-Westfalen, "Opengeodaten.nrw," 2025, accessed: 2025-02-04. [Online]. Available: https://www.opengeodata.nrw.de/produkte/

[38] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2024.

[39] Hessian State Office for Land Management and Geoinformation, Germany (HVBG), "Atkis basic data and digital orthophotos," 2025, accessed: 2025-08-12. [Online]. Available: https://gds.hessen.de/

[40] D. Manske, "Geo-Locations and System Data of Renewable Energy Installations in Germany," accessed: 2025–14. [Online]. Available: https://doi.org/10.5281/zenodo.14627853

[41] LEA LandesEnergieAgentur Hessen, "Solar-kataster hessen," 2024, accessed: 2025-03-18. [Online]. Available: https://www.lea-hessen.de/buergerinnen-und-buerger/sonnenenergie-nutzen/solar-kataster-hessen/

[42] K. Mayer, Z. Wang, M.-L. Arlt, D. Neumann, and R. Rajagopal, "DeepSolar for Germany: A deep learning framework for PV system mapping from aerial imagery," in *2020 International Conference on Smart Energy Systems and Technologies (SEST)*, Sep. 2020, pp. 1–6.

[43] K. Mayer, B. Rausch, M.-L. Arlt, G. Gust, Z. Wang, D. Neumann, and R. Rajagopal, "3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D," *Applied Energy*, vol. 310, p. 118469, Mar. 2022.

[44] HMWVW, "Energiewende in Hessen · Monitoringbericht 2024," p. 182, 2024, accessed: 2025-10-14. [Online]. Available: https://wirtschaft.hessen.de/sites/wirtschaft.hessen.de/files/2024-12/monitoringbericht_2024_web.pdf

[45] J.-B. Eggers, M. Behnisch, J. Eisenlohr, H. Poglitsch, W.-F. Phung, M. Muenzinger, C. Ferrara, and T. Kuhn, "PV-Ausbauerfordernisse versus Gebäudepotenzial: Ergebnis einer gebäudescharfen Analyse für ganz Deutschland," in *35. PV-Symposium*, Sep. 2020.

Biography Section

Svea Krikau received her M.Sc. in Geodesy and Geoinformatics from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2021. She is currently a research associate and doctoral candidate at the Institute of Photogrammetry and Remote Sensing, KIT. Her research focuses on image analysis of remote sensing data using machine learning techniques for the retrieval of environmental parameters, including water constituents, urban heat islands, and renewable energy assessment, as well as on integrating these data with various geospatial datasets.

Sina Keller received the Ph.D. degree in natural sciences from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2015, and the Habilitation degree ('Privatdozentin'in 2023 with a habilitation on 'Concept and Application of Machine Learning Methods for Estimating Physical Quantities from Heterogeneous Geoinformation Data'. She is currently a Group Leader with the Institute of Photogrammetry and Remote Sensing (IPF), KIT. Her research interests include machine learning for remote sensing and geospatial data analysis, environmental monitoring, renewable energy assessment, and the integration of heterogeneous geoinformation with physical modeling.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/