



HAL
open science

Towards unsupervised assessment with open-source data of the accuracy of deep learning-based distributed PV mapping

Gabriel Kasmi, Laurent Dubus, Philippe Blanc, Yves-Marie Saint-Drenan

► To cite this version:

Gabriel Kasmi, Laurent Dubus, Philippe Blanc, Yves-Marie Saint-Drenan. Towards unsupervised assessment with open-source data of the accuracy of deep learning-based distributed PV mapping. Workshop on Machine Learning for Earth Observation (MACLEAN), in Conjunction with the ECML/PKDD 2022, Sep 2022, Grenoble, France. hal-03778289

HAL Id: hal-03778289

<https://hal.science/hal-03778289>

Submitted on 15 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards unsupervised assessment with open-source data of the accuracy of deep learning-based distributed PV mapping*

Gabriel Kasmi^{1,2}[0000-0002-7774-4302], Laurent Dubus²[0000-0002-3987-646X],
Philippe Blanc¹[0000-0002-6345-0004], and Yves-Marie
Saint-Drenan¹[0000-0003-1471-1092]

¹ MINES Paris, Université PSL, Centre Observation Impacts Energie (O.I.E.), 06904
Sophia Antipolis, France

`{name.surname}@mines-paris.psl.eu`

² RTE - Réseau de transport d'électricité, 92073 Paris La Défense, France
`{name.surname}@rte-france.com`

Abstract. Photovoltaic (PV) energy is rapidly growing and key to mitigating the energy crisis. However, distributed PV generation, which amounts to half of the PV installed capacity, is typically unavailable to transmission system operators (TSOs), making it increasingly difficult to balance the load and supply and avoid grid congestions. To assess distributed PV generation, TSOs need precise knowledge regarding the metadata of distributed PV installations. Many remote sensing-based approaches have been proposed to map these installations in recent years. However, to use these methods in industrial processes, assessing their accuracy over the mapping area, i.e., the area covered by the model during deployment, is necessary. We define the downstream task accuracy (DTA) as the accuracy over the mapping area, automatically computed using publicly available data sources and the model's outputs and expressed in an interpretable way for operators. We benchmark existing models for distributed PV mapping and show how they perform in terms of DTA. We show that the accuracy computed on the test set overestimates by about 30 percentage points the accuracy on the mapping area. Our approach paves the way for safer integration of deep-learning-based pipelines for remote PV mapping. Code is available at <https://github.com/gabrielkasmi/dsfrance>.

Keywords: Remote sensing · Solar panel mapping · Convolutional neural networks · Accuracy assessment · Robustness.

1 Introduction

Integrating growing amounts of PV energy into the electric grid is challenging for transmission system operators (TSOs). Like other renewable energy sources (RES), PV energy is weather-dependent and highly decentralized. Therefore, to

* Supported by RTE France.

ensure its optimal integration on the grid, accurate measurements of the production and accurate production forecasts are necessary. Otherwise, increased PV penetration could result in increased congestion, imbalances, and supplementary reserve requests [14].

TSOs are responsible for the balance between the load and supply of electricity. They rely on production measurements and forecasts for all energy sources to ensure this balance. In the case of PV energy, only the production stemming from power plants is accessible to the TSO through real-time measurements (when their installed capacity is larger than 1 MW_p) or estimations that use regional PV models [9]. The remaining installations, often referred to as distributed PV, are invisible to the system operator [16]. As distributed PV amounts to half the PV installed capacity, TSOs must derive methods to estimate the production of these invisible distributed PV installations.

A straightforward approach would be to expand regional PV models to distributed PV. These models estimate the PV power generation over a given area, taking a limited set of PV characteristics (installed capacity, tilt, and azimuth angles) and weather data [15, 9]. As there is no centralized record of the characteristics of distributed PV installations [7], a popular solution is to map the characteristics of distributed PV installations using earth observation data and deep learning models. Characteristics usually include the localization, tilt, azimuth, and installed capacity of the distributed PV installations [7]. Works such as [19, 6] or [11] leveraged convolutional neural networks (CNNs) and high-resolution overhead imagery to map solar installations over large areas; respectively the continental United States, the Netherlands and the German state of North-Rhine Westphalia.

However, these mapping algorithms lack accountability since it is impossible to assess their accuracy while in deployment[2]. This lack of accountability limits the practical use of these registries, as deep neural networks are sensitive to domain shift [17], which can lead to unpredictable performance drop under conditions that differ from those encountered in the training dataset. Currently, there exists no way to automatically and systematically assess the accuracy of PV mapping systems, although their sensitivity to domain shift is documented [18]. So far, the only attempts to assess the accuracy outside the training dataset have been to rely on manual verification [6, 10].

Our main contribution is to propose an unsupervised and scalable way to monitor the accuracy of deep learning-based mapping pipelines over their whole *mapping area*. The *mapping area* covers the domain beyond the test set on which the model is deployed. We call the accuracy computed over the mapping area the *downstream task accuracy (DTA)*. We express DTA in an interpretable way for operators. To derive this measurement, we use the most detailed available information on distributed PV installations and compare it to the outputs of our model. We provide an unsupervised and interpretable way of assessing the accuracy of PV mapping algorithms.

Additionally, we contribute to the ongoing effort to map distributed PV installations' tilt, localization, and installed capacity over more than 50,000 square kilometers in France, currently the largest PV registry with this level of detail.

2 Related work

Scholars proposed numerous approaches over the last few years to map solar arrays on overhead imagery. [5] provides a complete overview of the works in this field. Early works relied on hand-crafted features and classification algorithms to identify PV panels on aerial images. The advent of deep-learning [8] enabled large-scale mapping of PV panels using semantic segmentation.

The DeepSolar project [19] was a significant milestone as a deep learning-based pipeline was used to detect PV installations and estimate their surface area for the first time. The method relies on a two-step pipeline: images are classified, and if an image contains an array, it is passed to a segmentation model to identify the polygons corresponding to the PV installation. With this method, the authors achieved a precision of 93.1% (recall: 88.5%) in residential and a precision of 93.7% (recall: 90.5%) in nonresidential areas.

DeepSolar triggered efforts to construct CNN-based pipelines to map installations over large areas [11, 6]. However, as pointed out by [2], pipelines developed over one territory cannot be straightforwardly applied to another. Moreover, the data extracted with automatic pipelines is not accountable enough to be used in official statistics. As a first step towards a better assessment of the accuracy of deep learning-based PV pipelines over their mapping area, [6] leveraged manual annotators to compute the precision and recall in places that were not in the training dataset. However, this method is labor-intensive and cannot be scaled up.

[2] highlight two main research directions in PV detection. First, improving the reported data's accountability over the mapping area is necessary. To improve accountability, one needs to assess the accuracy of the pipeline's output in the mapping area. Second, one needs to mitigate domain shift as it degrades the accuracy of deep-learning mapping algorithms. In this work, we address the first question by introducing a scalable and unsupervised approach to assess the accuracy of a remote PV mapping algorithm over its mapping area.

3 Data

3.1 Training data

We train our classification and segmentation models on a new dataset called BDAPPV (*Base de données d'apprentissage profond photovoltaïque*). This dataset contains labeled thumbnails of PV panels. These PV panels come from a PV database maintained by the non-profit association of small owners of PV panels "Asso BDPV" (*Base de données photovoltaïque*). Table 1 summarizes the characteristics of our training dataset. Our training dataset is nearly balanced, and

we provide information on the PV systems’ metadata, such as the tilt, azimuth, installed capacity, and other technical characteristics. A crowdsourcing campaign enabled us to gather this training data, which we will release in a forthcoming publication.

Table 1. Training dataset characteristics. Column "Total number of samples" indicates the number of samples in the dataset (prior to augmentation), and column "positive samples" indicates the number of samples depicting a PV panel.

Dataset	Total number of samples	Positive samples (share in %)
Train	12127	5445 (44.90)
Validation	1732	755 (43.59)
Test	3466	1485 (42.84)
Total	17325	7685 (44.36)

3.2 Geographical data for PV systems mapping

We feed our classification and segmentation models into the PV mapping pipeline described in section 4.1. We then deployed this pipeline over a large deployment area of 9 French *départements*, representing an area of more than 50 000 square kilometers. Our PV mapping pipeline requires the following data inputs:

Orthoimagery We do image classification and segmentation on RGB orthoimagery. These images are provided by the *Institut Géographique National* (IGN) and are freely accessible online. This dataset is called BD ORTHO. The ground sampling distance of these images is 20 cm/pixel. These images cover all French *départements*. For our study, we downloaded the image bundles of 9 French *départements*, located in the North, West, South, and East of France, covering approximately 10% of the French metropolitan territory. Images of a given *départements* are updated every three years. In the worst case, we map the installations as they were three years ago. Since we know the overall installed capacity for a *département*, and assuming that the installations’ characteristics are comparable over three years, the relatively low revisit rate is not problematic: we construct a representative sample of the current installed capacity.

Topographic data We use topographic data, also provided by the IGN, under an open license. This dataset is called BD TOPO. This data contains the geographic footprints of all buildings registered in France. This dataset’s main aim is to filter and merge detections made on the same rooftop.

Distributed PV characteristics At the characteristics extraction stage of our PV registry pipeline (see figure 1), we use the PV characteristics gathered

in the BDPV database to calibrate the characteristics extraction module of our pipeline. These characteristics include the tilt, azimuth, surface, and installed capacity of PV installations. This database contains information on about 17000 installations in France, less than 5% of the total estimated count of distributed PV installations. Moreover, as self-reported, the installations’ geographical localization is not representative of the actual geographical localization of the distributed PV installations in France.

3.3 Monitoring data

National installations registry Our data source for accuracy assessment is the *Registre national d’installations* (RNI) [12]. The French administration compiles the RNI every year. The RNI contains the total number of distributed PV installations and the aggregated installed capacity for each city. It does not contain the individual PV systems’ characteristics.

4 Methods

4.1 Automated registry pipeline

We build on [11] to construct an automated PV registry pipeline. This pipeline takes as input orthoimages and topological data and outputs PV panels’ characteristics (surface, installed capacity, tilt, and azimuth angles) and localization. Figure 1 summarizes our approach.

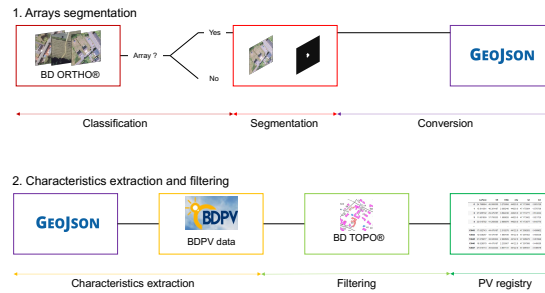


Fig. 1. Automated PV characteristics extraction pipeline

Classification and segmentation We fine-tuned our classification (Inception-v3) and segmentation (DeepLab-v3) models on our training dataset BDAPPV (see section 3.1). During inference, we classify all images and segment only images on which a PV panel is detected. Segmentation masks are then converted into polygons. We use the weights of [11] as initialization for our fine-tuned models. Training details are provided in section 4.2.

Characteristics extraction The output of the segmentation stage is a file with all PV installations' polygons detected in a given *département*. Based on this polygon, we extract the following characteristics: location, projected surface area, tilt angle, surface area, and installed capacity. The computation of the location and projected surface area are straightforward. Like previous works [11], we derive the installed capacity from the surface area and the tilt angle using a linear fit. The regression coefficient captures the efficiency of the PV panels.

We computed a look-up table based on the BDPV database to estimate the tilt angle. The BDPV database spreads across France. This LUT enables us to impute a tilt angle based on the PV polygon's projected surface and localization. We defined four projected surface clusters based on their statistical occurrence in the BDPV database. Then, for each surface cluster (see fig. 2), we divide France into geographical squares, and for each square, we compute the average tilt angle. Figure 2 depicts the look-up-table. We chose this approach as it is intuitive, computationally very efficient, and does not require surface models as in [11]. Our main aim is to capture the geographical variability of the tilt angle. As shown in figure 2, tilts are steeper in the north and for small installations. Besides, we also capture the fact that tilt angles are steeper for smaller installations [15].

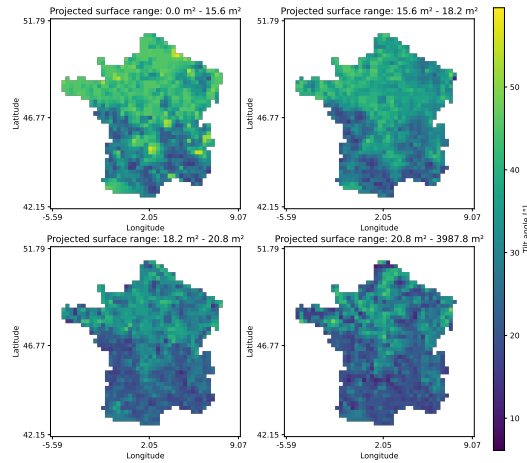


Fig. 2. Look-up table (LUT) to infer the tilt angle of the PV installations based on its localization and projected surface area. The LUT gives the average tilt over each square, computed based on the BDPV dataset.

Post-processing We use the BD TOPO to filter all polygons that are not on a building. We also filter installations that are either too small (we set the threshold at 1.7 square meters, the typical area of a single PV module) or too large. The upper threshold is 36 kWp as the RNI focuses on installations with an installed capacity lower than 36 kWp.

4.2 Training procedure

Augmentations During training, images are flipped vertically and rotated 90 degrees clockwise and counter-clockwise. We avoid rotations that lead to having arrays pointing north (i.e., upwards). Input images have a resolution of 400×400 , so we randomly crop the images to generate input images with the correct resolution of 299×299 pixels. Finally, images are also normalized with the standard ImageNet values (i.e., a mean of (0.485, 0.456, 0.406) and a standard deviation of (0.229, 0.224, 0.225)). No augmentation, but the normalization is applied to the validation and testing sets.

Classification branch We retrain all the layers of the model. We train the model for 25 epochs. We pick the model that achieved the lowest validation accuracy after the end of the training. We evaluate the final performance after threshold fine-tuning on the testing dataset. We use the binary cross entropy loss (BCE, without weighting) and Adam optimizer with a learning rate of 0.0001. We used a batch size of 128.

Segmentation branch The model is trained for 25 epochs. As for classification, we pick the model that achieved the lowest validation accuracy after the end of the training. We also use BCE loss with a learning rate of 0.0001 and Adam optimizer. The batch size is 64, spread across multiple GPUs.

4.3 Downstream task accuracy metrics

The RNI provides city-wise the total number of installations and the aggregated installed capacity. To use it as a reference, we aggregate our detections city-wise. To the best of our knowledge, this is the first work to introduce a systematic assessment of the accuracy in terms of installed capacity over the whole deployment area. [19] assessed the accuracy of the estimation of the surface of the PV panels on the test set only. [11] measured the accuracy of the installed capacity estimation per installation but only in cities for which reference data is available.

Definition The *downstream task accuracy (DTA)* is a set of metrics that measures the accuracy of a model over the whole mapping area, where no labeled data is available. We express de DTA in interpretable terms for operators. In our case, we compute the DTA in terms of installed capacity using the models' outputs and the aggregated information of the RNI. The DTA requires no additional human labor as we automatically compute it.

DTA metrics Using the RNI dataset and the aggregations from our pipeline, we define three metrics to ensure that the model estimates the installed capacity correctly. We denote \hat{C}_i our estimation of the installed capacity based on the aggregation of our detections in city i , C_i the reference installed capacity from the RNI for city i . Similarly, \hat{k}_i denotes our estimation of the number of installations in city i , and k_i is the reference number of installations recorded in the RNI.

- The **average percentage error (APE)** $\frac{|C_i - \hat{C}_i|}{C_i}$ and the **mean APE (MAPE)** $\frac{1}{n} \sum_{i=1}^n \frac{|C_i - \hat{C}_i|}{C_i}$ computed over the whole *département*. The APE and MAPE are computed for the installed capacity C_i in city i ,
- The **detection ratio** $\Delta := \frac{\hat{k}_i}{k_i}$, based on [11] computed at the city level and averaged over the *départements*. We compute this ratio for the number of installations k_i in city i ,
- The **average installation percentage error (AIPE)** $-\frac{C_i/k_i - \hat{C}_i/\hat{k}_i}{C_i/k_i}$ which is the APE computed for the *average* installation. By construction, a negative AIPE indicates that we underestimate the installations’ size, and a positive AIPE indicates that we overestimate them.

The MAPE ensures we do not overestimate or underestimate the overall installed capacity. The detection ratio ensures that we detect the correct number of installations, and the AIPE, which is a function of the installed capacity and the number of installations, ensures that, on average, we correctly estimate the size of the installations.

5 Results

5.1 Accuracy on the test set

Classification and segmentation accuracies Our fine-tuned model achieves competitive results compared to state-of-the-art models (see table 2). For the classification branch, we reach an F1 score of 0.84. For the segmentation branch, we reach an Intersection-over-Union (IoU) of 0.86. Our aim is not to establish a new state-of-the-art (SOTA) for classification or segmentation but to see how current performance translates in terms of accuracy over the mapping area.

Comparison to existing accuracy metrics As mentioned in section 4.3, there is no counterpart to our DTA metrics in the existing literature. However, [19] computed the area estimation’s mean relative error (MRE) and reported an error of 3.0% in residential and 2.1% in nonresidential areas. Our model achieves 0.44% MRE on the test set. See [19] for more details on the computation of the MRE.

Table 2. Classification and segmentation accuracy. The ground sampling distance (GSD) indicates how detailed the image is. The lower the GSD, the more detailed the image. PV panels cannot be detected on images that have a GSD greater than 30 cm/pixel ([2]).

	Classification	Segmentation	
Work	F1-score	IoU	GSD (cm/pixel)
[11]	0.87	0.74	10
[10]	-	0.67	30
[20]	0.82	-	10
[13]	0.97	0.86	10
Ours	0.84	0.86	20

[11] computed the installation-wise accuracy over the cities of Borcken, Unna, and Dortmund, Germany. They reported a mean average error (MAE) between 1.92 kWp and 24.84 kWp and a mean average percentage error (MAPE) between 27% and 37%. Our model achieves an MAE of 0.38 kWp and a MAPE of 11.59% when comparing the estimated installed capacity with the reference recorded in the BDPV dataset.

5.2 Downstream task accuracy on the mapping area

We deploy our model on nine French *départements*: Nord (north), Loire-Atlantique (west), Hérault (south), and 6 out of 8 *départements* of the Rhône-Alpes region (southeast). This way, we ensure sufficient geographical diversity in landscapes, population densities, and architectural styles. The total area covered is 51858 square kilometers. It is currently the largest area mapped with this level of detail, as [11] covered an area of roughly 34000 square kilometers.

We compute our DTA metrics over this mapping area using the RNI. To compute a baseline, we consider the test set as one city and compute the associated mean and median APE, detection ratio, and AIPE. We report the results in table 3.

Main results As we can see in table 3, our model performs worse on the mapping area than on the test set. The MAPE is about 30 percentage points higher on the mapping area than on the test set. This result is consistent with existing literature, and our accuracy tracking metrics enable us to quantify this decrease in accuracy. The MAPE is 47%, and the mean ratio is slightly above 1. In detail, we can see that we tend to slightly overestimate (by 16%) the size of the installations compared to the RNI. Performance for all metrics is relatively constant across all *départements* but the Nord.

Effect of filtering on accuracy As an example use case for the DTA, we discuss the effect of filtering using the topological data. Filtering by buildings decreases the number of detections and especially of small detections. Its effects

on accuracy are ambiguous. On the one hand, it slightly biases the size of the installations (the ratio decreases and the AIPE increases). On the other hand, it can dramatically improve the detection accuracy in places where performance is low: in the *départements* Nord and Hérault, we can see that the post-processing improves all evaluation metrics by a large margin.

Table 3. Downstream task accuracy across the mapping area. Values in parentheses correspond to the results without filtering by buildings. The line "Test" considers the training dataset as one city. k_i and C_i denote the count of installations and the installed capacity, respectively. The hat indicates the estimation by our pipeline.

<i>Dép.</i>	MAPE [%]	med. APE [%]	mean Δ [-]	mean AIPE [%]	k_i [-]	\hat{k}_i [-]	C_i [kWp]	\hat{C}_i [kWp]
Test	17.61	-	0.92	-0.10	1485	1362	6473.8	5334.02
44	39.09 (33.20)	38.05 (26.69)	0.67 (0.91)	22.83 (18.54)	12683	6838 (9325)	51955.17	34197.71 (45206.58)
69	31.99 (39.45)	28.91 (23.29)	0.83 (1.26)	12.18 (10.33)	8944	6508 (9808)	36500.6	31361.59 (45433.94)
59	130.06 (224.22)	88.13 (168.29)	2.23 (3.22)	61.16 (41.56)	6453	9524 (15393)	22083.8	52790.18 (73697.38)
34	26.80 (45.57)	17.78 (30.05)	1.01 (1.33)	6.82 (4.29)	9199	8408 (11445)	35398.41	39897.16 (52841.79)
01	35.90 (38.77)	35.35 (27.07)	0.77 (1.13)	6.18 (7.34)	4940	3654 (5259)	18433.19	14659.39 (21371.97)
38	33.41 (31.29)	31.15 (21.80)	0.81 (1.07)	9.18 (6.20)	10672	7835 (10680)	39691.43	32391.11 (42617.43)
42	29.12 (46.30)	23.81 (28.66)	1.00 (1.41)	15.42 (11.06)	6892	5831 (8384)	28594.08	27916.48 (38222.44)
26	30.46 (55.35)	23.51 (25.80)	0.92 (1.43)	3.74 (6.15)	5808	4933 (7121)	28261.94	25833.96 (37645.72)
74	44.08 (41.45)	41.18 (28.53)	0.67 (0.93)	-4.61 (-6.10)	7004	5287 (6600)	32202.1	21760.4 (25930.6)
Overall	47.45 (66.20)	32.81 (30.66)	1.03 (1.46)	16.33 (12.03)	72595	58818 (84015)	293120.70	280807.91 (382967.83)

5.3 Generalization to other areas

The need for the French TSO for detailed information on distributed PV motivates this work. However, this need is ubiquitous in many European countries and beyond [7, 16]. It is possible to apply our accuracy assessment in these countries: for instance, the RNI can be substituted by the *Marktstammdatenregister (MaStR)* for Germany [1], the *Stammdateregister* for Denmark [3].

Besides, high-resolution orthoimagery is increasingly available across European countries due to the INSPIRE directive, which enforces open access to

geographical information services (GIS) data [4]. [11] used orthoimagery available over North Rhine Westphalia and [6] images available for the Netherlands. Although this case study focuses on France, the data required for replication is available throughout Europe.

6 Conclusion and future work

We built on existing literature to construct an automated pipeline for large-scale distributed photovoltaic (PV) mapping and characterization. We mapped about 50000 square kilometers over France, whose distributed PV installations are not well mapped, resulting in the largest with this level of detail.

Crucially, we address the need for a more transparent and systematic assessment of the accuracy beyond the test set through downstream task accuracy (DTA). DTA measures the accuracy in terms of installed capacity over the whole mapping area, i.e., the area over which the model is deployed. It is unsupervised and does not require additional human labor. DTA allows quantifying the decrease in accuracy when we shift from the test set to the mapping area. Such quantification paves the way for safer integration of deep learning-based pipelines for remote PV mapping.

We benchmarked state-of-the-art detection mapping pipelines with our newly introduced downstream task accuracy metrics. We showed that test set performance overestimates the accuracy of such pipelines in estimating the installed capacity of distributed PV installations by about 30 percentage points.

In future work, we will derive downstream task accuracy metrics for tilt and azimuth estimation. We also intend to pursue the model’s deployment until eventually mapping the whole of metropolitan France. Finally, we shall discuss the benefit of more advanced classification and segmentation models on downstream task accuracy.

References

1. Bundesnetzagentur: Marktstammdatenregister. <https://www.marktstammdatenregister.de/> (2022), [Online; accessed 4-July-2022]
2. De Jong, T., Bromuri, S., Chang, X., Debusschere, M., Rosenski, N., Schartner, C., Strauch, K., Boehmer, M., Curier, L.: Monitoring Spatial Sustainable Development: semi-automated analysis of Satellite and Aerial Images for Energy Transition and Sustainability Indicators. arXiv preprint arXiv:2009.05738 (2020)
3. Energistyrelsen: Stamdataregister. <https://ens.dk/service/statistik-data-noegletal-og-kort/data-oversigt-over-energisektoren> (2022), [Online; accessed 4-July-2022]
4. European Union: Directive 2007/2/EC establishing an infrastructure for spatial information in the European Community (Inspire) (2007), <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=legisum%3A128195>
5. de Hoog, J., Maetschke, S., Ilfrich, P., Kolluri, R.R.: Using satellite and aerial imagery for identification of solar pv: State of the art and research opportunities. In: Proceedings of the Eleventh ACM International Conference on Future Energy Systems. pp. 308–313 (2020)

6. Kausika, B.B., Nijmeijer, D., Reimerink, I., Brouwer, P., Liem, V.: GeoAI for detection of Solar Photovoltaic Installations in the Netherlands. *Energy and AI* p. 100111 (2021)
7. Killinger, S., Lingfors, D., Saint-Drenan, Y.M., Moraitis, P., van Sark, W., Taylor, J., Engerer, N.A., Bright, J.M.: On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading. *Solar Energy* **173**, 1087–1106 (2018)
8. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *nature* **521**(7553), 436–444 (2015)
9. Lorenz, E., Ruiz-Arias, J.A., Martin, L., Wilbert, S., Köhler, C., Fritz, R., Betti, A., Lauret, P., David, M., Huang, J., Perez, R., Kazantzidis, A., Wang, P., Saint-Drenan, Y.M.: Forecasting solar radiation and photovoltaic power. In: Sengupta, M., Habte, A., Wilbert, S., Gueymard, C., Remund, J. (eds.) *Best practices handbook for the collection and use of solar resource data for solar energy applications*, chap. 8, pp. 206–272. National Renewable Energy Lab.(NREL), Golden, CO (United States) (2021)
10. Malof, J.M., Li, B., Huang, B., Bradbury, K., Stretslov, A.: Mapping solar array location, size, and capacity using deep learning and overhead imagery. arXiv preprint arXiv:1902.10895 (2019)
11. Mayer, K., Rausch, B., Arlt, M.L., Gust, G., Wang, Z., Neumann, D., Rajagopal, R.: 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Applied Energy* **310**, 118469 (2022)
12. Opendata réseaux-énergies: Registre national des installations de production et de stockage d'électricité (au 30/04/2022). <https://opendata.reseaux-energies.fr/> (2022), [Online; accessed 23-August-2022]
13. Parhar, P., Sawasaki, R., Todeschini, A., Reed, C., Vahabi, H., Nusaputra, N., Vergara, F.: HyperionSolarNet: Solar Panel Detection from Aerial Images. In: *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning* (2021), <https://www.climatechange.ai/papers/neurips2021/41>
14. Pierro, M., Gentili, D., Liolli, F.R., Cornaro, C., Moser, D., Betti, A., Moschella, M., Collino, E., Ronzio, D., van der Meer, D.: Progress in regional pv power forecasting: A sensitivity analysis on the italian case study. *Renewable Energy* **189**, 983–996 (2022)
15. Saint-Drenan, Y.M., Good, G.H., Braun, M.: A probabilistic approach to the estimation of regional photovoltaic power production. *Solar Energy* **147**, 257–276 (2017)
16. Shaker, H., Zareipour, H., Wood, D.: A data-driven approach for estimating the power generation of invisible solar sites. *IEEE Transactions on Smart Grid* **7**(5), 2466–2476 (2015)
17. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: *CVPR 2011*. pp. 1521–1528. IEEE (2011)
18. Wang, R., Camilo, J., Collins, L.M., Bradbury, K., Malof, J.M.: The poor generalization of deep convolutional networks to aerial imagery from new geographic locations: an empirical study with solar array detection. In: *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. pp. 1–8. IEEE (2017)
19. Yu, J., Wang, Z., Majumdar, A., Rajagopal, R.: DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States. *Joule* **2**(12), 2605–2617 (2018)
20. Zech, M., Ranalli, J.: Predicting PV Areas in Aerial Images with Deep Learning. In: *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. pp. 0767–0774. IEEE (2020)