Applied Energy 298 (2021) 117273

Contents lists available at ScienceDirect

Applied Energy

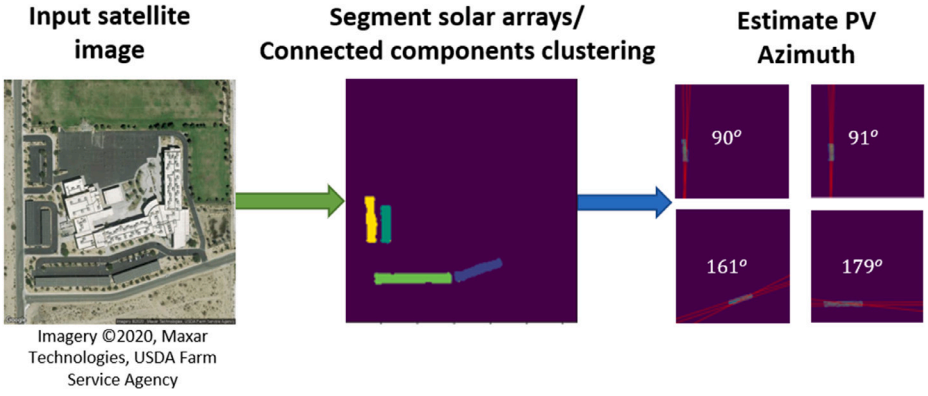
journal homepage: www.elsevier.com/locate/apenergy

Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform ☆

Ayobami S. Edun [a,*], Kirsten Perry [b], Joel B. Harley [a], Chris Deline [b,*]

[a] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, USA
[b] National Renewable Energy Laboratory, CO, USA

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Photovoltaics
Semantic segmentation
Convolutional neural network
Azimuth estimation
Computer vision

ABSTRACT

This paper explains the use of a convolutional neural network (CNN) to segment solar panels in a satellite image containing solar arrays, and extract associated metadata from the arrays. A novel unsupervised technique is introduced to estimate the azimuth of each individual solar panel from the predicted mask of the convolutional neural network. This pipeline was developed with the aim of extracting necessary metadata for a solar installation, using only a set of latitude–longitude coordinates. Azimuth prediction results for 669 individual solar installations associated with 387 sites located across the United States are provided. A mean average error and median average error of 21.65 degrees and 1.0 degrees were obtained, respectively, when predicting the azimuth of the solar fleet data set, with about 80% of the results within an error of zero degrees of the ground truth azimuth value and about 85% within an error of 25 degrees. The predicted azimuth was then used to estimate the energy conversion of the solar arrays. Results show a 90.9 and 90.6 R-squared value for estimating alternating current (AC) and direct current (DC) energy, respectively, and a mean absolute percentage error (MAPE) of 1.70% in estimating the alternating current (AC) energy using the fully automated algorithm.

☆ This document is the results of the research project funded by the United States Department of Energy.
* Corresponding authors.
E-mail addresses: aedun@ufl.edu (A.S. Edun), kirsten.perry@nrel.gov (K. Perry), joel.harley@ufl.edu (J.B. Harley), chris.deline@nrel.gov (C. Deline).
URL: http://ayobamiedun.mystrikingly.com/ (A.S. Edun).

https://doi.org/10.1016/j.apenergy.2021.117273
Received 2 March 2021; Received in revised form 2 June 2021; Accepted 9 June 2021
Available online 23 June 2021

1. Introduction

There has been a rapid growth in the deployment of photovoltaics (PV) in the world. In the United States, over 13 GW of PV was installed in 2019, representing 40% of all sources of new electricity generation (fossil fuels included) installed in the United States that year. Over the last decade, PV experienced an average annual growth rate of about 49%, with now more than 81 gigawatts (GW) of solar capacity installed nationwide, enough to power 16 million homes [1]. The efficiency of these installed PV systems is dependent on parameters such as panel size, azimuth, tilt, and mounting configuration of the solar array. In particular, the tilt and azimuth angles determine the amount of solar energy that PV modules absorb [2–4] and subsequently the amount of power they can deliver. Hence, they must be properly monitored. In this paper, a novel unsupervised technique is presented for estimating the azimuth of PV arrays in satellite imagery.

Azimuth estimation is important in monitoring the performance of a solar fleet, as correct azimuth is paramount for calculating system degradation and other factors. Most institutions rely on fleet owners sending this data manually. Gaps in system owner data motivate the need to automatically extract this metadata correctly from a satellite image, rather than relying on data sources which may be incorrect. Due to the importance of azimuth in the monitoring of grid-connected PV systems for quality assessment, reliability, inspection, compliance, data gathering, and efficiency, several works [5,6] have focused on predicting or estimating the optimal azimuth for maximizing the energy delivered by a PV system rather than manual inspection. A nonlinear least-squares approach was used in [7] to estimate the azimuth of a PV system. This approach relies on the ambient temperature of the location. Other works have compared simulations with PV output to determine the azimuth [8,9]. However, these methods are sensitive to measurement errors and can be error-prone because some electrical characteristics of PV systems are hard to simulate [10].

Deep learning methods have also been used to estimate the azimuth of a solar array. Authors in [11] used a deep neural network to estimate the size, tilt, and azimuth of solar panels using simulated time series data. The minimum net load observed at 12 p.m. throughout the year divided by that observed at 9 a.m. and 3 p.m. was used as input to the network. However this work is completely supervised, and dependent on summer and winter net loads. Additionally, simulated irradiance data at the associated locations was used to train and validate the method. Authors in [12] attempted to estimate azimuth using satellite imagery but did not provide extensive details. Our work aims to improve current methods of extracting PV metadata, by estimating solar panel azimuth from satellite images in an unsupervised manner. The advantage of our unsupervised method is that it requires no additional site data when estimating azimuth, other than the site's satellite image.

The use of satellite data and satellite imagery has received a growing effort in the PV field. For example, authors in [13] developed a fully automated performance check based on solar irradiance information derived from satellite data to replace on-site measurements. This research focuses on detecting system malfunctions and changing operating conditions as early as possible to prevent energy and financial losses. Authors in [14–16] proposed a solar power and irradiance prediction model based on various historical satellite images. In [17–19], methods were developed using satellite imagery to estimate the potential of installing PV arrays on rooftops of cities.

Deep learning methods have also been explored in combination with satellite imagery in PV systems. Before metadata can be extracted from the satellite image of a solar installation, the installation must be segmented. Authors in [20] used a deep learning-based approach to carry out image segmentation of building rooftops on low-resolution rooftop satellite images taken in Bangalore, India. The result of the segmentation was then used to estimate the solar energy that could be generated from rooftop areas. Similarly, authors in [21] developed a method to estimate building energy consumption from aerial and satellite imagery. In [22], a computer vision algorithm was developed to detect rooftop PV installations in satellite imagery and high-resolution aerial images. Similarly, in [23], a convolutional neural network was used to detect solar photovoltaic panels in low-quality satellite photos. DeepSolar was developed in [24,25] to create, publish, and, in the future, update and maintain a comprehensive open database on the location and size of solar PV installations; however, the DeepSolar project only estimated the horizontal projection areas of solar panels from satellite imagery and none of these works have automated the extraction of solar panel azimuth.

To address this gap, our research has two primary objectives. First, a convolutional neural network was developed, with the intent of segmenting the solar panels in the image, using transfer learning. Second, after performing image segmentation, the azimuth of the solar array in the satellite image was predicted using edge detection and a Hough transform. To address the case where there are multiple solar arrays, a connected components algorithm was implemented, which can be used after the panels have been segmented, to cluster the panels at the same or similar orientations in a solar installation. Then, for each cluster, the azimuth of the solar array was estimated. To the best of our knowledge, this is the first work that combines computer vision and transfer learning to segment solar panels from a satellite image and predict the azimuth of the solar panels.

Our results show an average dice coefficient of 81.74% was obtained as the accuracy of the panel segmentation. The results for azimuth prediction show a mean average error and median average error of 21.65 degrees and 1.0 degrees respectively. About 80% of the results are within an error of zero degrees of the ground truth azimuth value and about 85% are within an error of 25 degrees. This implies that, with further improvements, the approach proposed in our work [26] may become an excellent method for solar array metadata extraction via satellite images.

2. Data sets

Satellite images at a series of latitude–longitude coordinates were automatically generated using the Google Maps API. Latitude–longitude coordinates were provided via the National Renewable Energy Laboratory's PV Fleet Performance Data Initiative database [27]. The PV Fleet Initiative is a US Department of Energy-funded project, where operational PV plant data is aggregated into a centralized cloud repository, for the purpose of large-scale degradation analysis across the US. In addition to containing PV time series data, the Fleet database contains site metadata, including site latitude–longitude coordinates, azimuth, and tilt information. Solar installations in the initiative range from commercial to utility-scale sites and can include rooftop systems, carports, and ground-mounted systems. Satellite images were generated via the Google Maps API with a zoom level of 18, 0.596 m/pixel resolution, and 640 × 640 pixel dimensions. Each image was taken using the longitude and latitude coordinates of known NREL partner sites. While satellite images can be captured from different orientations, we aligned the images to face due north as that will be the reference point when detecting the azimuth.

For the segmentation algorithm, the solar arrays in each satellite image were manually annotated using the Python labelme package [28]. This package allowed for the specific pixels in each image belonging to solar arrays to be masked. Using labelme, a secondary, matching semantic segmentation mask image was generated for each satellite image. A subset of the satellite images and their associated masks is available via the National Renewable Energy Laboratory's DuraMAT Data Hub, for public Ref. [29]. Images were split into training, test, and validation sets, with 343 locations represented in the training set, 98 locations represented in the validation set, and 49 locations represented in the test set.

A subset of the satellite images used when training and validating the segmentation algorithm was used again to validate the clustering and azimuth estimation algorithms. In total, 387 satellite images, each representing a different geographic location, were used when comparing predicted azimuth to its associated ground truth. Azimuth data for each solar array in the images was provided by site owners in the initiative. Images could contain multiple solar array installations, with varying azimuth values. In total, 669 solar installations are present across the 387 images. The image set contains 185 rooftop installations, 225 ground mount installations, and 259 carport installations. 37 of the installations were utility installations, and 632 of the installations were commercial. Additionally, 653 of the installations were fixed-tilt configuration, and 16 of the installations were single-axis tracking.

3. Methodology

A CNN was used to segment and determine the boundaries of the solar array in the satellite image. CNN's are very good feature extractors, and can learn relevant features from an image for the purpose of segmentation [30]. To train the CNN architecture for segmentation task, training images with their corresponding masks, described in Section 2, were used.

3.1. Segmentation model

For the segmentation task, an encoder–decoder architecture was implemented. This architecture takes an input satellite image of dimension 640-by-640 and outputs a mask of the same size. The encoder part of the network is VGG16. Since VGG16 was originally designed for classification, the network architecture was modified briefly. Specifically, the input size was modified to 640-by-640-by-3 to match our input dimension, and the dense layers were removed; however, the weights of the architecture were fixed to speed up our training process. This technique is called transfer learning. Several works have shown that using a pre-trained model can help avoid overfitting [31].

Numerous works [32,33] have demonstrated the use of CNNs for segmentation of solar panels. For this reason, the main focus of this work is on the azimuth prediction algorithm. Details about the segmentation model used in this paper can be found in the supplementary materials section.

3.2. Azimuth detection

The second objective of our work is to detect the azimuth of the solar arrays after segmenting the solar panels. Azimuth detection was done through a combination of steps. This consists of Canny Edge Detection and the Hough transform. For images with solar arrays with different orientations, a clustering algorithm was used to cluster the solar arrays in the segmentation output. The connected components algorithm [34] was used to automatically determine the optimal number of clusters. The algorithm does not make strong assumptions on the statistics of the clusters, unlike K-Means clustering, which assumes that data points assigned to a cluster are spherical about the cluster center. Hence, the connected components algorithm helps create more accurate clusters.

3.2.1. Solar array clustering

In the case where the output of the segmentation showed multiple solar arrays in a single satellite image with dissimilar orientations, the connected components algorithm was implemented, which clusters individual solar arrays before estimating each solar array azimuth.

The connected components algorithm has gained increased popularity because of its performance in graph-based clustering problems. For example, it was used in [35,36] to cluster orthophoto images. The algorithm aims to find the number of connected components (often called sub graphs) in an undirected graph. Two nodes in the graph are connected if there is a path between them. In terms of images, connected components refers to the set of pixels that can reach each other without a change in intensity. The goal is to assign unique labels to each connected component, making it a cluster. Each cluster can then be extracted as an independent image with the same size as the input image.

Do note that our implementation does automatically decide the best number of clusters. This requires no human input and adds to the unsupervised nature of our work in an attempt to make the whole process fully automated. Fig. 1 shows an overview of the tasks considered in our work.

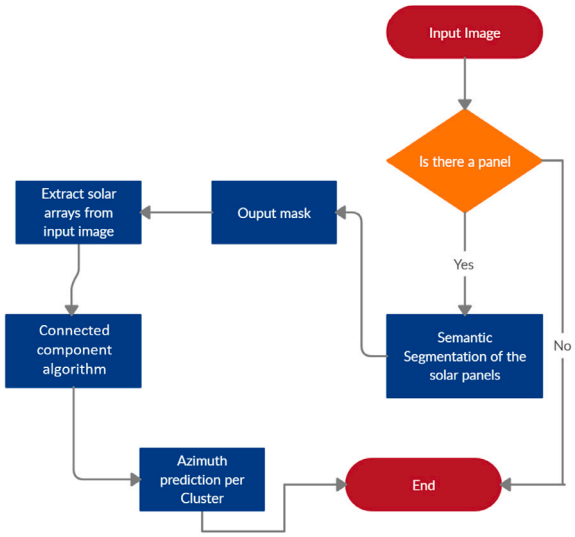


Fig. 1. A flow chart showing the tasks covered in our methodology.

3.2.2. Edge detection

Canny Edge Detection was used to identify edges in the segmented image. The Canny Edge Detector, developed by John F. Canny in 1986, is an edge detection variant that detects a wide range of edges in images using a sequential algorithm [37]. To reduce noise, the detector first smooths the input image using a Gaussian filter with a given standard deviation. Then, the local gradient and edge direction are computed at each pixel using edge detection operators such as the Sobel, Prewitt, Roberts, or the Laplacian of a Gaussian. This process is used to determine edge points. This is then followed by a non-maximum suppression, a double threshold step to identify strong, weak, and non-relevant pixels. Finally, the algorithm performs edge linking by transforming weak pixels into strong ones.

In our case, Canny Edge Detection was not used directly on the segmentation output from the CNN architecture. The panels were first clustered if solar arrays with multiple orientations were present in the image. The predicted mask in each cluster was then used to isolate the regions with solar panels from the input satellite image. Subsequently, Canny Edge Detection was applied on each segmented solar array cluster before passing through the Hough Transform, which is discussed in the following section.

3.2.3. Hough transform

The Hough Transform is a feature extraction technique used in image processing and computer vision tasks [38]. The algorithm was patented by Paul V. C. Hough in 1962. It was originally invented to detect complex lines in images [38]. The algorithm has been modified and enhanced to recognize other shapes such as circles and curves if their parametric equation is known. The Hough Transform detects lines

and circles via a voting procedure, which is performed in a parameter space called the Hough Space. In this space, lines are represented in a parametric form using radius (r) and angle (θ).

The algorithm takes an edge image and a range of radius values (r) and angles (θ) with their respective step sizes as input. Generally, the angle ranges from 0 to 180 or −90 to 90 degrees while the radius ranges from $-r$ to r, where r is the length of the edge image's diagonal. A 2D array, called the accumulator, represents the Hough Space, and is initialized with zeros of dimension (number of radii, number of angles).

The algorithm then checks for edge pixels on the edge image. For each edge pixel, it loops through all possible values of θ and calculates the corresponding radius r. It then checks the accumulator for the index that matches the pair (θ and r) and increments the accumulator. To decide the most probable lines, the (θ and r) pair that exceeds a certain threshold is chosen. These points are returned and can then be converted to the equation of the line for plotting purposes.

In calculating the radius and angle, the algorithm takes the top left of the image as the origin. The angle ranges from 0 to −90 degrees (in the anti-clockwise direction) and 0 to 90 degrees (in the clockwise direction). The algorithm returns the top three most probable lines it finds in the image, where the value three was experimentally derived. The orientation of these lines is then used to detect the azimuth of the solar array using two conditions, described below.

To estimate the azimuth, the frequency of each unique orientation that is returned was observed and used to determine the orientation with the maximum frequency. If the maximum frequency orientation is estimated more than once, it is chosen as the global orientation of the solar arrays; however, if all unique orientations have a frequency of 1, the orientation of the longest line was chosen because it goes through the largest section of a panel. To get the azimuth of the solar array, 90 degrees was added to the orientation. This was done because of how the azimuth is defined. When the panel faces east, it has an azimuth of 90 degrees; when it faces south, it has an azimuth of 180 degrees. Since the Hough Transform output takes the top left as the origin, this means a vertical line (facing east) will have an orientation of 0 degrees and a horizontal line will have an angle of 90 or −90 degrees. Based on this knowledge, the orientation of the output of the Hough Transform is 90 degrees off the azimuth value; consequently, 90 degrees was added to the output values to obtain a final azimuth estimate.

3.3. Solar energy conversion model

For each solar array where azimuth was estimated in the validation set, the difference in calculated total estimated energy between the system with its ground truth azimuth and the system with its predicted azimuth was assessed. To do this, NREL's PVWatts model [39] was run for each system, using default system parameters. System-specific latitude and longitude coordinates were used to select the nearest climate data file. The array's specific mounting type (fixed tilt or tracking) was selected based on known metadata. The PVWatts simulation was run twice, once with known azimuth and again with the estimated azimuth. The AC kWh energy was summed over a year-long period, and a comparison was made between the ground truth azimuth simulation and the predicted azimuth simulation. In using the percentage difference between the two estimates, all system input parameters canceled out, and estimate differences could be entirely attributed to azimuth variation.

4. Results

4.1. Segmentation results

During the first stage of our algorithm, specific panels are segmented from the image using the CNN. The network architecture was trained using the training data set described in Section 2, which contains panels and their corresponding masks. The training accuracy started as high as 85%, while the dice coefficient started at about 7%. This shows how the euclidean accuracy metric is deceptive and overly optimistic. Over the 300 epochs, the training dice coefficient increased steadily and the training loss decreased. Similarly, validation statistics show the validation dice coefficient increased steadily until it peaked at about 83%. During training, the validation loss was constantly monitored, and the convolutional neural network state was saved each time there was a reduction in the validation loss.

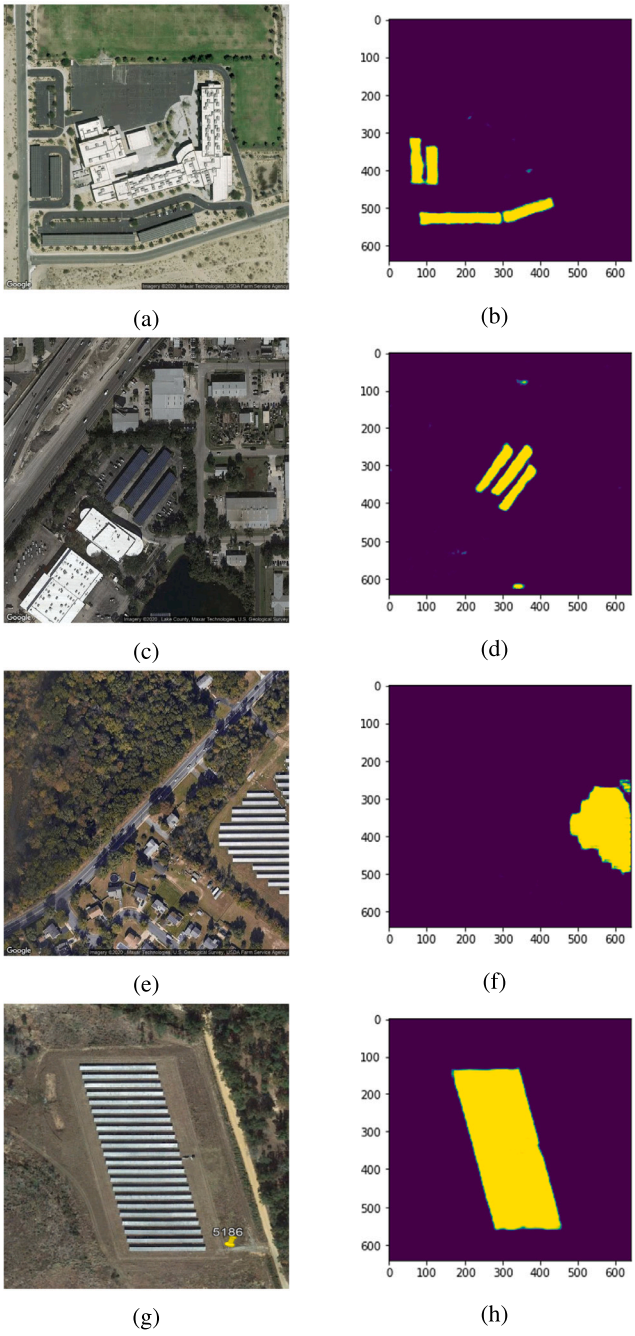


Fig. 2. Semantic segmentation results ((b), (d), (f), (h)) for four test images ((a), (c), (e), (g)) using the network architecture described in 3.1.

Our test data set, described previously, contains 49 images. Fig. 2 shows segmentation results for four test input images. The 49-image test set was passed through the segmentation architecture; results in Table 1 show an accuracy of 98.66% and an average dice coefficient of

Table 1
Classification and segmentation metrics using the network architecture described in 3.1.

Metric	Value
Panel classification statistics	
Average training accuracy	99.36%
Average validation accuracy	97.55%
Average training loss	1.78%
Average validation loss	10.51%
Test accuracy	99.03%
Test loss	1.16%
Panel segmentation statistics	
Average training accuracy	98.54%
Average validation accuracy	98.62%
Average training loss	4.42%
Average validation loss	5.06%
Average training dice coefficient	80.11%
Maximum training dice coefficient	91.69%
Average validation dice coefficient	80.74%
Average test accuracy	98.66%
Average test dice coefficient	81.74%

81.74% for the test images. However, the model achieved a maximum dice coefficient of 98.38% for a single image.

In comparison to the work in [40], where similar computer vision algorithms were used, our work has several differences. Our dataset considers rooftops, ground mounts, and carports across multiple parts of the United States while previous work has focused on rooftop data in California. Our approach operates with lower resolution images (0.596 m/pixel) relative to the prior work ($\leq$ 0.3 m/pixel). Our work also uses only a single convolutional neural network (CNN) and does not require several pre-processing and post-processing steps as in previous efforts. The work in [23,41,42] uses CNNs to detect panels in image patches or identify the general location of panels rather than directly segmenting the images.

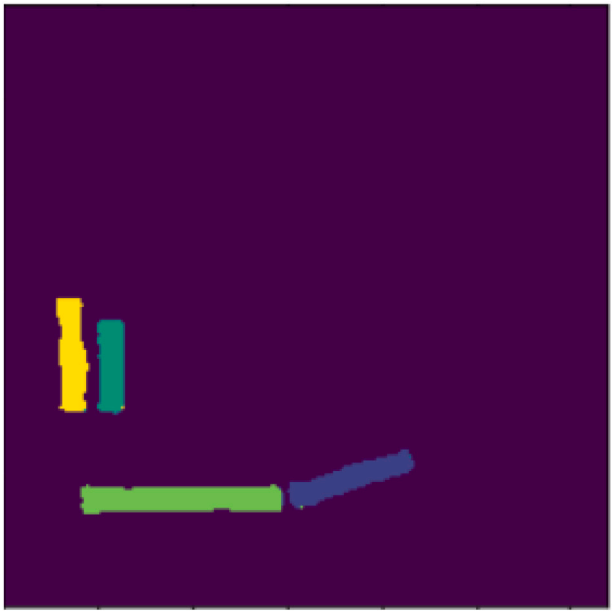
4.2. Connected components algorithm results

The connected components algorithm's ability to cluster individual solar arrays was validated using the data set described above. A trained analyst counted the number of individual solar arrays present in each satellite image, and this ground-truth value was compared to the clustering algorithm results.

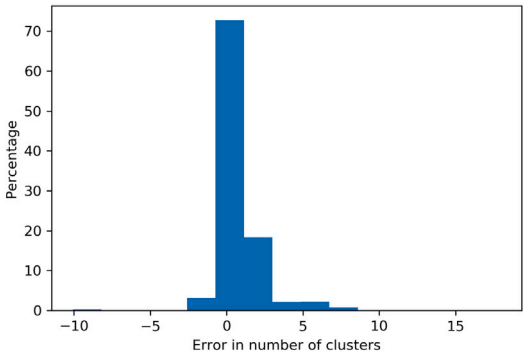
Fig. 3(a) shows the result of implementing the connected components algorithm on Test image 1. The algorithm was able to correctly cluster each solar array without the need to manually input the number of optimal clusters. The azimuth of each cluster was then predicted, as explained in Section 3.2. Fig. 3(b) shows an histogram of the error in estimating the optimal number of clusters. The predicted number of clusters was subtracted from the ground truth optimal number of clusters to obtain the errors. The histogram shows that about 90% of the error lies with +/− 1 clusters. This is a great result given that the algorithm is completely unsupervised.

4.3. Azimuth results

The second task our architecture performs is the estimation of solar array azimuths. As explained in Section 3.2, after passing the satellite image through the segmentation network, the output is a mask of the region with solar panels. The mask of the segmentation output was clustered and the mask of each cluster was used to extract the solar panels from the satellite image. The extracted solar panels were then passed through a Canny Edge Detector and Hough Line Transform to detect the dominant lines present in the image. Fig. 4 shows the results of this process without clustering. Using the lines obtained via the Hough Line transform, the azimuth of the solar panels can be predicted, as shown in the same figure. Observe that the algorithm predicts the



(a)



(b)

Fig. 3. (a) Connected components algorithm on test image 1 as described in Section 3.2.2 (b) The histogram of error between predicted number of clusters and actual number of clusters.

global azimuth and works well for Figs. 4(b–d). However, in the case of Fig. 4a, a global azimuth prediction is not enough. This emphasizes the need for clustering. Note that without clustering, ten Hough lines were used, while three Hough lines were used for each cluster when clustering is applied.

The azimuth prediction algorithm was tested on satellite images collected from 387 distinct locations across the United States, as described in Section 2. Note that for our results, the algorithm automatically clusters the segmentation output before predicting the azimuth in each cluster. Fig. 5 shows the comparison of our predictions with ground truth azimuth values. The scatter plot in Fig. 5(a) shows an obvious linear trend in our model, which illustrates strong accuracy between prediction and ground truth.

Fig. 5(b) shows a cumulative distribution function (CDF) of the residuals. About 80 percent of the residuals are within a difference of 0 degrees. Incorrectly predicted azimuths arise for a variety of reasons. First, some predictions were approximately 180 degrees off of ground truth values. This occurs because the algorithm cannot differentiate between east and west orientations. That is, when the generated Hough

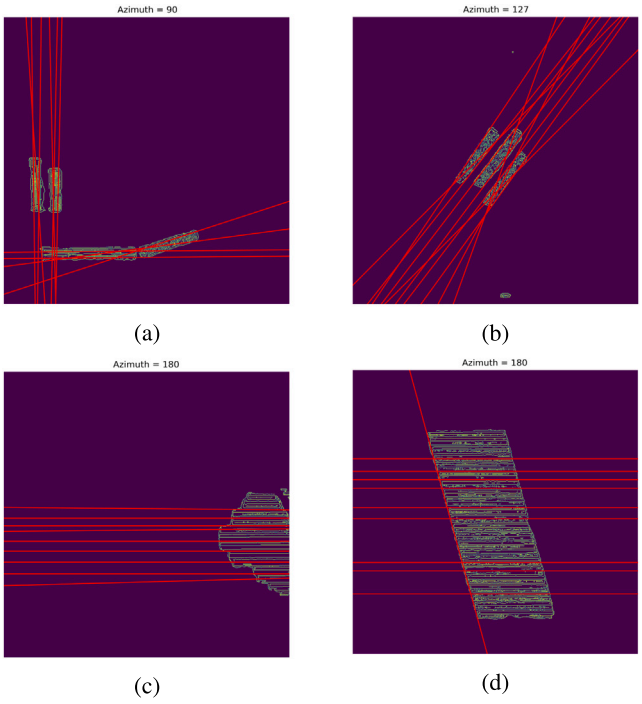


Fig. 4. Edge detection, line detection, azimuth prediction of (a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4 using the procedure described in 3.2.

Table 2
Azimuth results based on the methodology described in 3.2.

Azimuth estimation		
Mount type	MAE (in degrees)	MEAE (in degrees)
Rooftop	30.53	3.73
Carport	17.87	1.00
Ground	15.62	1.00

Line for an estimate is vertical, ambiguity arises because the array could either be facing east or west. Secondly, some estimations were off by approximately 90 degrees, as the algorithm's selected Hough Line went in the direction perpendicular to solar array length, as opposed to parallel. Additionally, for single-axis tracking installations, solar azimuth is calculated along the solar array length, as opposed to perpendicular to length, so these estimations were off by approximately 90 degrees.

Overall, the results show a mean absolute error(MAE) of 21.65 degrees and median absolute error of 1 degree even though our predictions are completely unsupervised. Table 2 shows the mean absolute error (MAE) and median absolute error (MEAE) broken down by different mounting types. The predictions for the rooftop installations have the largest error, compared to ground and carport mounting types. This is because rooftop installations are more likely to be unusual in shape and can span multiple azimuth configurations. The work in [11] used a deep neural network to estimate azimuth and they achieved a mean absolute error of 4.71 degrees. However, note that the simulation data used in the work was derived from irradiance measurement data, and required meter data from the building; in contrast, we only require a satellite image. The advantage of our method is that it does not rely on simulated data, meter readings, weather data, and it is completely unsupervised.

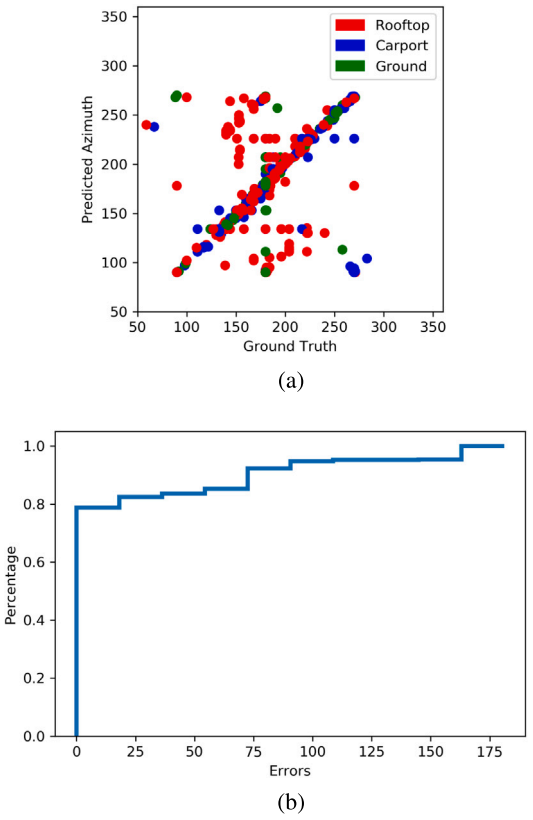


Fig. 5. Comparison of azimuth results with ground data. (a) Shows a scatter plot of predicted azimuth vs. ground truth based on mount type. (b) Shows the cumulative distribution function of residuals.

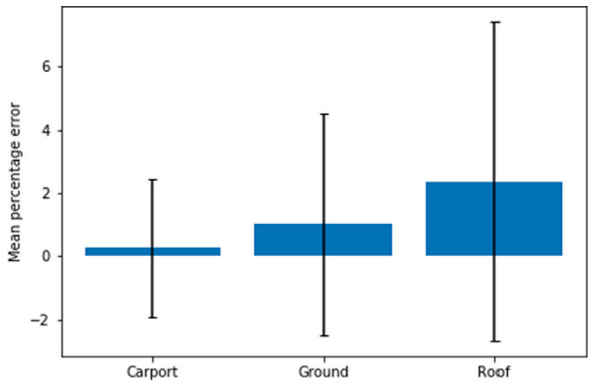


Fig. 6. Percentage difference of predicted and actual AC energy.

4.4. AC energy results

Fig. 6 shows the percentage difference in the actual and predicted energy based on an installation's mounting type. The figure shows a 0.27%, 1.13%, and 2.45% mean percentage absolute error in AC energy output for carport, ground-mounted, and rooftop sites, respectively. This is a consequence of the better azimuth prediction by the algorithm for both carport and ground-mounted sites. Overall, when tested on 387 sites, our algorithm, which is fully automated and requires no supervision, results in a mean absolute prediction error (MAPE) of 1.70% in AC energy output.

4.5. Proposed technique benefits

This work offers a great reduction in error, as compared to self-reporting and voluntary surveys, which is the normal data collection method for solar owners and operators. Self-reporting requires the presence of technicians in the field. Additionally, when operators acquire new solar installations, metadata such as azimuth may not be provided, and the operator must manually survey the newly acquired site. These on-site surveys are both expensive and time-consuming. Using the algorithm proposed here rapidly returns accurate results, and saves the operator on survey costs.

Although there remains a 1.70% MAPE in energy estimation, this can be improved in a future work by advancing the algorithm.

5. Conclusions and future work

This research demonstrates that computer vision techniques can be successfully employed to extract metadata from the satellite images containing solar arrays. Using a convolutional neural network framework, this work demonstrates that using transfer learning, individual solar arrays in satellite images can be successfully segmented. Promising results are obtained using the connected components algorithm to cluster individual solar arrays in satellite imagery, and predict the azimuth for each individual cluster using Canny Edge Detection and Hough Line Transform techniques.

In future research, authors plan to further develop our computer vision pipeline to obtain other metadata information, including the following:

- Automatic classification of installation type (roof-mounted, ground-mounted, and so on) for solar installations in a satellite image.
- Automatic classification of mounting type (fixed tilt, tracking) for solar installations in satellite imagery.
- Calculation of the panel-to-site ratio for solar array installations, via various computer vision techniques.
- Automatic detection of stuck trackers in single-axis tracking systems, using satellite imagery. Inspection of photovoltaic cells for hotspots via satellite imagery.

The pipeline developed in this research aids in completing the future research above, as the current work can now successfully detect and isolate individual solar arrays in satellite imagery.

CRediT authorship contribution statement

Ayobami S. Edun: Methodology, Software, Formal analysis, Investigation, Writing - original draft, Visualization. **Kirsten Perry:** Conceptualization, Resources, Software, Data curation, Validation. **Joel B. Harley:** Methodology, Resources, Writing - review & editing. **Chris Deline:** Conceptualization, Writing - review & editing, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory (NREL). Funding provided by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 34348. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.apenergy.2021.117273.

References

[1] Solar Energy Industries Association. Solar industry research data. 2020, https://www.seia.org/solar-industry-research-data. (Accessed 20 December 2020).
[2] Dhimish M, Silvestre S. Estimating the impact of azimuth-angle variations on photovoltaic annual energy production. Clean Energy 2019;3(1):47–58.
[3] Oladiran MT. Mean global radiation captured by inclined collectors at various surface azimuth angles in Nigeria. Appl Energy 1995;52(4):317–30.
[4] Mohammadi K, Khorasanizadeh H. A review of solar radiation on vertically mounted solar surfaces and proper azimuth angles in six iranian major cities. Renew Sustain Energy Rev 2015;47:504–18.
[5] Le Roux W. Optimum tilt and azimuth angles for fixed solar collectors in South Africa using measured data. Renew Energy 2016;96:603–12.
[6] Rowlands IH, Kemery BP, Beausoleil-Morrison I. Optimal solar-PV tilt angle and azimuth: An Ontario (Canada) case-study. Energy Policy 2011;39(3):1397–409.
[7] Killinger S, Engerer N, Muller B. QCPV: A quality control algorithm for distributed photovoltaic array power output. Sol Energy 2017;143:120–31.
[8] Saint-Drenan YM, Bofinger S, Fritz R, Vogt S, Good GH, Dobschinski J. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. Sol Energy 2015;120:479–93.
[9] Haghdadi N, Copper J, Bruce A, MacGill I. A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data. Renew Energy 2017;108:390–400.
[10] Meng B, Loonen RCGM, Hensen JLM. Data-driven inference of unknown tilt and azimuth of distributed PV systems. Sol Energy 2020;211:418–32.
[11] Mason K, Reno MJ, Blakely L, Vejdan S, Grijalva S. A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation. Sol Energy 2020;196:260–9.
[12] Li Q, Feng Y, Leng Y, Chen D. SolarFinder: Automatic detection of solar photovoltaic arrays. In: 2020 19th ACM/IEEE international conference on information processing in sensor networks. 2020, p. 193–204. http://dx.doi.org/10.1109/IPSN48710.2020.00024.
[13] Drews A, de Keizer A, Beyer HG, Lorenz E, Betcke J, van Sark W, et al. Monitoring and remote failure detection of grid-connected PV systems based on satellite observations. Sol Energy 2007;81(4):548–64.
[14] Jang HS, Bae KY, Park H, Sung DK. Solar power prediction based on satellite images and support vector machine. IEEE Trans Sustain Energy 2016;7(3):1255–63.
[15] Lago J, De Brabandere K, De Ridder F, De Schutter B. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. Sol Energy 2018;173:566–77.
[16] Kallio-Myers V, Riihelä A, Lahtinen P, Lindfors A. Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. Sol Energy 2020;198:68–80.
[17] Singh R, Banerjee R. Estimation of roof-top photovoltaic potential using satellite imagery and GIS. In: 2013 IEEE 39th photovoltaic specialists conference. 2013, p. 2343–7. http://dx.doi.org/10.1109/PVSC.2013.6744945.
[18] Zhang C, Li Z, Jiang H, Luo Y, Xu S. Deep learning method for evaluating photovoltaic potential of urban land-use: A case study of Wuhan, China. Appl Energy 2021;283:116329.
[19] Buffat R, Grassi S, Raubal M. A scalable method for estimating rooftop solar irradiation potential over large regions. Appl Energy 2018;216:389–401.

[20] Sampath A, Bijapur P, Karanam A, Umadevi V, Parathodiyil M. Estimation of rooftop solar energy generation using satellite image segmentation. In: 2019 IEEE 9th international conference on advanced computing. 2019, p. 38–44. http://dx.doi.org/10.1109/IACC48062.2019.8971578.
[21] Streltsov A, Malof JM, Huang B, Bradbury K. Estimating residential building energy consumption using overhead imagery. Appl Energy 2020;280:116018.
[22] Malof JM, Hou Rui, Collins LM, Bradbury K, Newell R. Automatic solar photovoltaic panel detection in satellite imagery. In: 2015 international conference on renewable energy research and applications. 2015, p. 1428–31. http://dx.doi.org/10.1109/ICRERA.2015.7418643.
[23] Golovko V, Bezobrazov S, Kroshchanka A, Sachenko A, Komar M, Karachka A. Convolutional neural network based solar photovoltaic panel detection in satellite photos. In: 2017 9th IEEE international conference on intelligent data acquisition and advanced computing systems: Technology and applications, vol. 1. 2017, p. 14–9. http://dx.doi.org/10.1109/IDAACS.2017.8094501.
[24] Sutherland BR. Locating photovoltaic installations with deep learning. Joule 2018;2(12):2512–3.
[25] 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 2018;2(12):2605–17.
[26] Edun A, Perry K, Deline C, US DOE Office of Energy Efficiency and Renewable Energy. Panel-segmentation. 2020, http://dx.doi.org/10.11578/dc.20201130.12, URL: https://www.osti.gov/biblio/1726007.
[27] Deline C, Muller M, Deceglie M, Jordan D, Anderson K, Simpson L, et al. PV fleet performance data initiative: March 2020 methodology report. 2020, http://dx.doi.org/10.2172/1659786.
[28] Wada K. labelme: Image Polygonal Annotation with Python. 2016, https://github.com/wkentaro/labelme.
[29] Perry K. Satellite images training and validation set. 2020.
[30] Maggiori E, Tarabalka Y, Charpiat G, Alliez P. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Trans Geosci Remote Sens 2017;55(2):645–57.
[31] Peng C, Li Y, Jiao L, Shang R. Efficient convolutional neural architecture search for remote sensing image scene classification. IEEE Trans Geosci Remote Sens 2020;1–14.
[32] Zhuang L, Zhang Z, Wang L. The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. Appl Soft Comput 2020;92:106283.
[33] de Hoog J, Maetschke S, Ilfrich P, Kolluri RR. 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. New York, NY, USA: Association for Computing Machinery; 2020, p. 308–13. http://dx.doi.org/10.1145/3396851.3397681.
[34] He L, Ren X, Gao Q, Zhao X, Yao B, Chao Y. The connected-component labeling problem: A review of state-of-the-art algorithms. Pattern Recognit 2017;70:25–43.
[35] Dornaika F, Moujahid A, Merabet Y El, Ruichek Y. Building detection from orthophotos using a machine learning approach: an empirical study on image segmentation and descriptors. Expert Syst Appl 2016;58:130–42.
[36] Abdollahi A, Pradhan B. Integrated technique of segmentation and classification methods with connected components analysis for road extraction from orthophoto images. Expert Syst Appl 2021;176:114908.
[37] Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;PAMI-8(6):679–98.
[38] Hough PVC. Method and means for recognizing complex patterns. 1962.
[39] Dobos A. PVWatts version 5 manual. Technical Report NREL/TP-6A20-62641, National Renewable Energy Laboratory; 2014, http://dx.doi.org/10.2172/1158421, URL: https://www.nrel.gov/docs/fy14osti/62641.pdf.
[40] Malof JM, Bradbury K, Collins LM, Newell RG. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. Appl Energy 2016;183:229–40.
[41] Yuan J, Yang HL, Omitaomu OA, Bhaduri BL. Large-scale solar panel mapping from aerial images using deep convolutional networks. In: 2016 IEEE international conference on big data. 2016, p. 2703–8. http://dx.doi.org/10.1109/BigData.2016.7840915.
[42] Malof JM, Collins LM, Bradbury K. A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery. In: 2017 IEEE international geoscience and remote sensing symposium. 2017, p. 874–7. http://dx.doi.org/10.1109/IGARSS.2017.8127092.

Ayobami S. Edun received his B.Eng. degree in Electrical and Electronics engineering from Federal University of Technology, Akure, Nigeria in 2014. He received the M.S. degree in Electrical and Computer engineering from University of Florida, Gainesville, FL, USA in 2019. He is currently working towards his PhD degree in Electrical and Computer Engineering at the University of Florida, Gainesville, FL, USA. He currently works as a Research Assistant at the SmartDATA Lab, University of Florida, Gainesville, FL where he focuses on developing algorithms to detect, localize, and characterize faults within solar arrays. His research interests include data science, machine learning, renewable energy, and smart grids. Ayobami's awards include 2020 College of Engineering outstanding achievement award and 2020 Scarborough-Maud Fraser award for commitment and dedication to academics and service within the University of Florida community.

Kirsten Perry received the B.S. and B.A. degrees in mechanical engineering and mathematics, respectively, from the University of Oklahoma, Norman, OK in 2017. She currently works full-time at the National Renewable Energy Laboratory as a data scientist on the PV Fleet Performance Data Initiative, while pursuing an M.S. in computer science from the Georgia Institute of Technology.

Kirsten's research interests include applying machine learning and deep learning techniques to assess photovoltaic field performance data.

Joel B. Harley (S'05-M'14) received his B.S. degree in Electrical Engineering from Tufts University in Medford, MA, USA. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Carnegie Mellon University in Pittsburgh, PA, USA in 2011 and 2014, respectively. In 2018, he joined the University of Florida, where he is currently an assistant professor in the Department of Electrical and Computer Engineering. Previously, he was an assistant professor in the Department of Electrical and Computer Engineering at the University of Utah. His research interests include integrating novel signal processing, machine learning, and data science methods for the analysis of waves and time-series data. Dr. Harley's awards and honors include 2020 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society Star Ambassador Award, a 2020 and 2018 Air Force Summer Faculty Fellowship, a 2017 Air Force Young Investigator Award, a 2014 Carnegie Mellon A. G. Jordan Award (for academic excellence and exceptional service to the community). He has published more than 90 technical journal and conference papers, including four best student papers. He is a student representative advisor for the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society, a member of the IEEE Signal Processing Society, and a member of the Acoustical Society of America.

Chris Deline (S'06-M'08-SM'19) received the B.S., M.S., and Ph.D. degrees from the University of Michigan, Ann Arbor, in 2003, 2005, and 2008, respectively, all in electrical engineering. Since 2008 he has been a research engineer at the National Renewable Energy Laboratory in Golden, CO and currently leads the photovoltaic system performance group. He manages the US Department of Energy Regional Test Center program at NREL for field assessment of novel PV technologies, and is principal investigator for multiple PV field performance projects including degradation rate assessment and bifacial system design and production modeling. Dr. Deline also manages the PV Fleet Performance Data Initiative which works with commercial fleet owners to aggregate and analyze PV field performance data using novel data quality and performance analysis tools. Dr. Deline is an expert in the field performance of PV modules, particularly partially shaded or mismatched PV with distributed power electronics. His research also includes characterization and energy simulation of advanced PV technologies including thin-film module metastability and bifacial system energy production.