SolarDetector: A Transformer-based Neural Network for the Detection and Masking of Solar Panels

Abdulrahman Salama
University of Washington
Tacoma, WA, USA
asalama@uw.edu

Abdeltawab Hendawi
University of Rhode Island
RI, USA
hendawi@uri.edu

Richard Franklin
University of Washington
Tacoma, WA, USA
rsamf@uw.edu

Eyhab Al-Masri
University of Washington
Tacoma, WA, USA
ealmasri@uw.edu

Anish Deshpande
University of Washington
Tacoma, WA, USA
amd1729@uw.edu

Mohamed Ali
University of Washington
Tacoma, WA, USA
mhali@uw.edu

ABSTRACT

As the global transition towards renewable energy sources accelerates, solar power becomes an increasingly important solution. Identifying and understanding the current distribution of solar panel installations is crucial for future planning and decision-making process. This paper introduces SolarDetector, a transformer-based neural network model, which we developed and fine-tuned for the accurate detection of solar panels. It achieves 91.0% mIoU for the task of masking solar panels on SWISSIMAGE dataset.

KEYWORDS

solar, semantic segmentation, geospatial, mask2former, swin, mask r-cnn

ACM Reference Format:
Abdulrahman Salama, Abdeltawab Hendawi, Richard Franklin, Eyhab Al-Masri, Anish Deshpande, and Mohamed Ali. 2023. SolarDetector: A Transformer-based Neural Network for the Detection and Masking of Solar Panels . In *The 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November 13–16, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589132.3625649

1 INTRODUCTION

Installation of solar panels is accelerating all over the world. For example, according to official capacity numbers reported by Swiss Federal Office of Energy, solar capacity has increased more than 200 times in Switzerland over the last 20 years from 18 megawatt (MW) in 2001 to 3,655 MW in 2021 [23]. The period between 2020 and 2021 alone has seen an increase of 23% bringing the capacity from 2,973 MW in 2020 to 3,655 MW in 2021. With the increase of solar energy as a renewable energy source, there is more need for efficient solar detection methods and visualizations at different levels. This problem is challenging for many reasons. Solar panels come with a huge variety of visual characteristics with different colors, shapes, and textures, making the process of detecting all different kinds accurately burdensome.

There are several research works in the literature focusing on using computer vision for solar panel detection. These works present various approaches on classification [13, 14, 16], object detection [15], segmentation using common machine learning techniques [20, 21, 30, 31, 34], and segmentation using CNN [17, 32].

The solar detection accuracy achieved by models in the literature is barely satisfactory and can be improved. This paper develops a model, SolarDetector, based on the latest transformer-swin architecture Mask2Former [2]. Experiments show that our developed model achieves 91.0% mIoU on SWISSIMAGE dataset.

The rest of this paper is organized as follows. Section 2 highlights related work. Sections 3 give details about the neural network component. The experiments and results are illustrated in Section 4. Finally, Section 5 concludes the paper.

2 RELATED WORK

There are several research works focusing on using computer vision for detecting solar panels. Papers [13, 14, 16] proposed approaches for classifying an image based on the presence or absence of solar panels. These approaches lack the ability to identify the precise location or surface areas of solar panels. There are other research works [20, 21, 30, 31, 34] that utilize basic techniques such as SVM and open-cv to detect solar panels based on common features such as color and texture. These approaches fail to generalize well on the different variety of solar panels' visual characteristic that exists in practice.

In [15], researchers developed a faster R-CNN neural network to detect solar panels installed on roofs. The paper utilized an object detection approach with a bounding box, which doesn't identify the exact surface area of solar panels. Paper [33] proposed a segmentation approach using CNN, but it has low precision and recall and more advanced CNN techniques have been available since then. Paper [32] introduced the DeepSolar dataset, which is a dataset having samples covering all US with negative and positive labels. Researchers used the dataset to build a CNN classifier model to indicate the presence or absence of solar panels. Furthermore, they utilized a semi-supervised segmentation approach using a greedy layer-wise training technique to estimate the boundary and size of solar panels. A small portion of the data was also labeled with ground truth masks and was used to evaluate this approach. The

This work is licensed under a Creative Commons Attribution International 4.0 License.
SIGSPATIAL '23, November 13–16, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0168-9/23/11.
https://doi.org/10.1145/3589132.3625649

approach was used to build a database for solar panels in the whole US. In [17], researchers utilized a Mask R-CNN model to perform semantic segmentation on solar panels. They also proposed the right angle algorithm to more accurately mask the sharp edges of solar panels. They used the DeepSolar dataset introduced in [32] to train their model and used it as a benchmark. Their approach outperformed the results achieved in [32] setting new state-of-the-art results on that dataset with IoU of 88.8%.

There are also many research works focusing on the area of detecting anomalies and defects in solar panels. In [1] researchers developed a real-time system to detect defects in photovoltaic (PV) modules. The images are taken through a drone with two cameras, a thermal and a Charge-Coupled Device CCD. The system then utilizes fault detection algorithms to detect various faults in real-time and send information to a ground station. In [12] researchers used thermal infrared imaging to detect anomalies in PV modules. Infrared video sequences are first collected and then sent to an image-processing algorithm to segment the solar panels from the background. Image preprocessing and pattern recognition are then used to detect common anomalies like cracks. Hot panels are also detected using DBSCAN clustering. In [22] researchers used deep learning and machine learning techniques to extract features and used feature classification to identify various types of defects. In [28] researchers proposed the use of standard thermal image processing and the Canny edge detection operator as diagnostic tools for module-related faults. There are other research works as well [25, 27] where researchers utilized machine learning techniques to detect defects in solar panels.

3 SOLAR DETECTOR

This section gives details on the neural network model, SolarDetector, which detects and masks the exact surface areas of solar panels. SolarDetector is retrained and fine-tuned on solar panel-specific training datasets. We also add augmentations during the training process such as flipping, rotation, and random brightness to diversify the training dataset and improve the model's ability to detect different variations of solar panels of different colors and orientations. We consider different neural network architectures as the potential starting backbone for the development of our SolarDetector model. The architectures we consider are Mask R-CNN [26], MaskFormer [3], and Mask2Former [2]. We evaluate and compare the performance of the different fine-tuned models based on the different architectures.

We evaluate the models' performance using the mIoU metric (mean Intersection-Over-Union), which is a standard metric for measuring the accuracy of semantic segmentation models. This metric is based on measuring the area of intersection over the area of union between the detected mask and the ground truth mask. Then this calculation is averaged across all images in the test dataset. Equation (1) and (2) represent the calculation of this metric.

$$IoU = \frac{TP}{TP + FP + FN} \tag{1}$$

where TP is the number of true positive pixels that are correctly predicted as belonging to the target class (solar panel), FP is the number of false positive pixels that are incorrectly predicted as belonging to the solar panel class, and FN is the number of false negative pixels that are incorrectly missed as being belonging to the solar panel class.

$$mIoU = \frac{\sum_{i=1}^{L} W_{\mathrm{i}} * IoU_{\mathrm{i}}}{\sum_{i=1}^{L} W_{\mathrm{i}}} \tag{2}$$

where L is the length of the test dataset, i is the index iterating over each image of the training dataset, $\mathrm{IoU_i}$ is IoU calculated for the image i, and $\mathrm{W_i}$ is the weight for the image i which is the area of the union.

We briefly explain the neural network architectures we consider in the following sections.

3.1 Mask R-CNN

These networks are based on the Mask-RCNN framework [26] and have the same network backbone based on ResNet but comes with different network heads. We consider the following variations within this architecture from detectron2 library [29] :

- **Mask R-CNN R-101-C4-3x:** Uses ResNet backbone with 101 layers. Feature extraction happens at the fourth convolution layer (C4).
- **Mask R-CNN R-101-DC5-3x:** Uses ResNet backbone with 101 layers. It uses dilation to extract features at the fifth convolution layer (DC5).
- **Mask R-CNN R-50-FPN-3x:** Uses ResNet backbone with 50 layers with an FPN head (Feature Pyramid Network). FPN extracts features at different scales.
- **Mask R-CNN R-101-FPN-3x:** Uses ResNet backbone with 101 layers with an FPN head.

All these networks are pre-trained on the ImageNet dataset which contains over 14 million images belonging to 1000 different classes.

3.2 MaskFormer Models

This architecture is introduced in paper [3]. It is a transformer-based and utilizes the SWIN architecture introduced in [19]. This model family has models available with different sizes (tiny, small, and large) and pre-trained either on the ade or coco dataset. Ade dataset is more suitable for segmentation tasks. Thus, we consider the following pre-trained variations within this architecture:

- maskformer-swin-tiny-ade [11]
- maskformer-swin-small-ade [10]
- maskformer-swin-large-ade [9]

3.3 Mask2Former Models

This architecture is an improvement on MaskFormer and is introduced in paper [2]. It is a new transformer-swin-based architecture that utilizes masked attention to extract localized features by constraining cross-attention within predicted mask regions. This model family also has models available with different sizes (tiny, small, and large) and pre-trained either on the ade or coco dataset. We consider the following pre-trained variations within this architecture:

- mask2former-swin-tiny-ade [6]
- mask2former-swin-small-ade [7]
- mask2former-swin-large-ade [8]

4 EXPERIMENTAL EVALUATION

In this section, we run experiments evaluating the performance of our model across multiple dimensions. We begin by describing the environment setup (Section 4.1). Then, we describe the different datasets used (Section 4.2). After that, we present experiments for developing and fine-tuning our model (Section 4.3). We measure and report the performance using precision, recall, and mIoU metrics.

4.1 Experiment Setup

All evaluations are conducted on Colab environment using a machine with NVIDIA T4 GPU. All components are implemented in Python. We use Pytorch as our neural network framework. To calculate mIoU, precision, and recall metrics, we use the "evaluate" module from Hugging Face [5], which provides implementations for such metrics.

4.2 Datasets

We utilize dataset SWISSIMAGE [24] as an example of real-world datasets for running many experiments. This dataset contains aerial images covering the whole of Switzerland over the last few years with specific regions being covered each year. Each tile from the SWISSIMAGE dataset is 10,000 x 10,000 pixels at 10 cm resolution, which means every tile covers an area of 1 km^2. These tiles are downloaded from the official website producing this dataset [24] as a tif image. However, such a large resolution is extremely hard to deal with and feeds into neural networks. Thus, we split every tif file into 100 tiles of 1,000 x 1,000 pixels each.

We also develop our own annotated dataset from SWISSIMAGE by selecting a random sample of images and labeling them to be used for training, validation, and testing. We adopt the online tool cvat.ai [4] to do the labeling. The tool then generates the labeled data in COCO format [18].

Table 1 summarizes the details of this annotated dataset, which we use extensively in the experiments section for retraining and validating different neural network architectures.

Table 1: SwissImage Dataset

Dataset	*# Tiles*	*# Annotated Instances*
Training Set	1000	605
Validation Set	200	343
Test Set	300	110
Total	**1500**	**1058**

4.3 Evaluations for Fine-Tuning SolarDetector

In this section, we present various experiments on retraining and fine-tuning different neural network architectures for the task of solar panel segmentation. We train all models for 200 epochs using a learning rate of 5e-5, batch size of 4, 4 workers for the data loader, and using Adam optimizer. We add augmentations of random flipping, rotation, and brightness contrast to diversify the training datasets.

Neural Network Architectures Comparison. In this experiment, we retrain different architectures on our SWISSIMAGE annotated dataset described in 4.2.

Table 2 presents these results. The inference time measurement includes data loading and is calculated using T4 machines with 4 data loader workers and a batch size of 4.

We note from Table 2 that transformer-based architectures significantly outperform the Mask R-CNN architecture by a wide margin for comparable model sizes. This is attributed to the power of transformer architectures and its ability to extract features using the attention mechanism which was found to be superior in comparison to traditional convolutions methods. We also note that Mask2Former outperforms MaskFormer significantly. Mask2Former are slightly larger than their MaskFormer counterparts and they extract localized features more efficiently using masked attention. All different sizes of Mask2Former perform very closely to each other. The tiny variation achieves 90.9% compared to 91.0% achieved by the large variation. Fig. 1 presents an example of solar panels detection by the model SolarDetector-Mask2Former-Tiny.

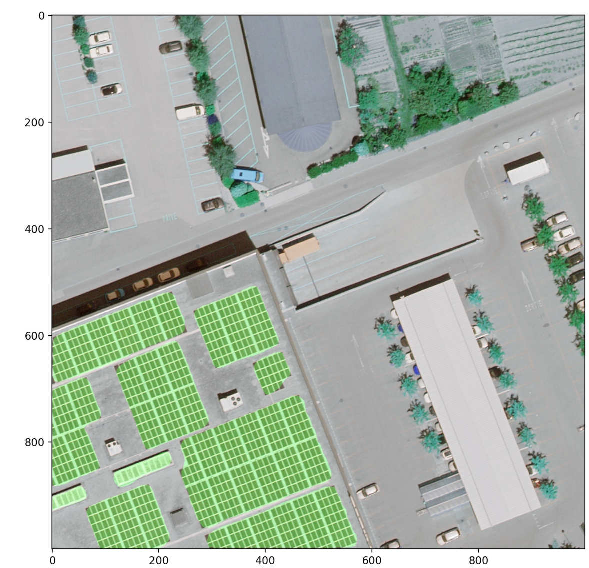

Figure 1: Example of Solar Panels Detection

In a detailed version of this work, we utilize this best performing model in a comprehensive system to process large geospatial regions to detect and quantify solar intensity. The system also utilizes a pyramid spatial index for efficient querying of solar intensity for any region represented by a spatial range query. We also include generation of heat maps showing distribution of solar intensity, and optimization algorithms to optimize the processing of geospatial regions repeatedly over time.

5 CONCLUSION

In this paper, we developed and fine-tuned a transformer-based neural network for detecting and masking solar panels. We conducted experiments comparing the performance of different neural network architectures on SWISSIMAGE dataset. Our model, SolarDetector, achieves 91.0% mIoU on SWISSIMAGE dataset. As solar

Table 2: Models Evaluation after Fine Tuning

Model	# Parameters	Time per Epoch (sec)	GPU Memory (Training)	Batch Inference Time (sec)	P	R	mIoU
SolarDetector-R-50-FPN	43.9 millions	123	2.1 GB	0.29	70.3%	51.3%	42.2%
SolarDetector-R-101-FPN	62.9 millions	140	3.5 GB	0.31	84.7%	73.6%	62.8%
SolarDetector-R-101-C4	54.0 millions	138	4.2 GB	0.71	62.7%	93.7%	60.1%
SolarDetector-R-101-DC5	190.8 millions	150	6.1 GB	1.20	55.6%	40.0%	30.3%
SolarDetector-MaskFormer-Tiny	41.7 millions	208	5.4 GB	0.48	93.8%	78.9%	75.0%
SolarDetector-MaskFormer-Small	63.0 millions	287	7.1 GB	0.82	88.3%	90.2%	80.6%
SolarDetector-MaskFormer-Large	211.5 millions	475	14.5 GB	0.86	94.3%	86.2%	81.9%
SolarDetector-Mask2Former-Tiny	47.4 millions	277	9.1 GB	0.62	94.2%	96.3%	**90.9**%
SolarDetector-Mask2Former-Small	68.7 millions	312	10.6 GB	1.09	94.3%	95.1%	**90.4**%
SolarDetector-Mask2Former-Large	215.5 millions	668	14.7 GB	1.36	94.8%	95.8%	**91.0**%

energy continues to grow in importance, accurate models like SolarDetector can empower users to analyze solar data from geospatial images helping them to make informed decisions effectively.

REFERENCES

[1] Moath Alsafasfeh, Ikhlas Abdel-Qader, Bradley Bazuin, Qais Alsafasfeh, and Wencong Su. 2018. Unsupervised Fault Detection and Analysis for Large Photovoltaic Systems Using Drones and Machine Vision. *Energies* 11, 9 (2018). https://doi.org/10.3390/en11092252

[2] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. 2022. Masked-attention Mask Transformer for Universal Image Segmentation. arXiv:2112.01527 [cs.CV]

[3] Bowen Cheng, Alexander G. Schwing, and Alexander Kirillov. 2021. Per-Pixel Classification is Not All You Need for Semantic Segmentation. arXiv:2107.06278 [cs.CV]

[4] CVAT.ai Corporation. 2022. *Computer Vision Annotation Tool (CVAT)*. https://github.com/opencv/cvat

[5] Hugging Face. 2023. *Evaluate Metric*. https://huggingface.co/evaluate-metric

[6] Hugging Face. 2023. *facebook/mask2former-swin-large-ade-semantic*. https://huggingface.co/facebook/mask2former-swin-large-ade-semantic

[7] Hugging Face. 2023. *facebook/mask2former-swin-large-ade-semantic*. https://huggingface.co/facebook/mask2former-swin-large-ade-semantic

[8] Hugging Face. 2023. *facebook/mask2former-swin-large-ade-semantic*. https://huggingface.co/facebook/mask2former-swin-large-ade-semantic

[9] Hugging Face. 2023. *facebook/maskformer-swin-large-ade*. https://huggingface.co/facebook/maskformer-swin-large-ade

[10] Hugging Face. 2023. *facebook/maskformer-swin-small-ade*. https://huggingface.co/facebook/maskformer-swin-small-ade

[11] Hugging Face. 2023. *facebook/maskformer-swin-tiny-ade*. https://huggingface.co/facebook/maskformer-swin-tiny-ade

[12] Xiang Gao, Eric Munson, Glen P Abousleman, and Jennie Si. 2015. Automatic solar panel recognition and defect detection using infrared imaging. In *Automatic Target Recognition XXV*, Vol. 9476. SPIE, 196–204.

[13] Vladimir Golovko, Sergei Bezobrazov, Alexander Kroshchanka, Anatoliy Sachenko, Myroslav Komar, and Andriy Karachka. 2017. 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 (IDAACS)*, Vol. 1. IEEE, 14–19.

[14] Vladimir Golovko, Sergei Bezobrazov, Alexander Kroshchanka, Anatoliy Sachenko, Myroslav Komar, and Andriy Karachka. 2017. 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 (IDAACS)*, Vol. 1. IEEE, 14–19.

[15] Vladimir Golovko, Alexander Kroshchanka, Sergei Bezobrazov, Anatoliy Sachenko, Myroslav Komar, and Oleksandr Novosad. 2018. Development of Solar Panels Detector. In *2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*. 761–764. https://doi.org/10.1109/INFOCOMMST.2018.8632132

[16] Dorian House, Margaret Lech, and Melissa Stolar. 2018. Using deep learning to identify potential roof spaces for solar panels. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*. IEEE, 1–6.

[17] SiMing Liang, FengYang Qi, YiFan Ding, Rui Cao, Qiang Yang, and Wenjun Yan. 2020. Mask R-CNN based segmentation method for satellite imagery of photovoltaics generation systems. In *2020 39th Chinese Control Conference (CCC)*. 5343–5348. https://doi.org/10.23919/CCC50068.2020.9189474

[18] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[19] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10012–10022.

[20] Jordan M Malof, Kyle Bradbury, Leslie M Collins, Richard G Newell, Alexander Serrano, Hetian Wu, and Sam Keene. 2016. Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier. In *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*. IEEE, 799–803.

[21] Jordan M. Malof, Rui Hou, Leslie M. Collins, Kyle Bradbury, and Richard Newell. 2015. Automatic solar photovoltaic panel detection in satellite imagery. In *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*. 1428–1431. https://doi.org/10.1109/ICRERA.2015.7418643

[22] S Naveen Venkatesh and V Sugumaran. 2022. Machine vision based fault diagnosis of photovoltaic modules using lazy learning approach. *Measurement* 191 (2022), 110786. https://doi.org/10.1016/j.measurement.2022.110786

[23] Swiss Federal Office of Energy. 2023. *Swiss Federal Office of Energy*. https://www.bfe.admin.ch/bfe/en/home/versorgung/erneuerbare-energien/solarenergie.exturl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWRtaW4uY2gvZGUvcHVibGljYX/Rpb24vZG93bmxvYWQvODc4Nw==.html

[24] Federal Office of Topography swisstopo. 2022. *SWISSIMAGE 10 cm*. https://www.swisstopo.admin.ch/en/geodata/images/ortho/swissimage10.html

[25] Xiaoliang Qian, Heqing Zhang, Huanlong Zhang, Yuanyuan Wu, Zhihua Diao, Qing-E Wu, and Cunxiang Yang. 2017. Solar cell surface defects detection based on computer vision. *International Journal of Performability Engineering* 13, 7 (2017), 1048. https://doi.org/10.23940/ijpe.17.07.p6.10481056

[26] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Eds.), Vol. 28. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[27] Du-Ming Tsai, Shih-Chieh Wu, and Wei-Yao Chiu. 2013. Defect Detection in Solar Modules Using ICA Basis Images. *IEEE Transactions on Industrial Informatics* 9, 1 (2013), 122–131. https://doi.org/10.1109/TII.2012.2209663

[28] John A Tsanakas, Dimitrios Chrysostomou, Pantelis N Botsaris, and Antonios Gasteratos. 2015. Fault diagnosis of photovoltaic modules through image processing and Canny edge detection on field thermographic measurements. *International Journal of Sustainable Energy* 34, 6 (2015), 351–372.

[29] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

[30] Zhipeng Xi, Zhuo Lou, Yan Sun, Xiaoxia Li, Qiang Yang, and Wenjun Yan. 2018. A vision-based inspection strategy for large-scale photovoltaic farms using an autonomous UAV. In *2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. IEEE, 200–203.

[31] Yi-yong Yao and Yu-tao Hu. 2017. Recognition and location of solar panels based on machine vision. In *2017 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*. 7–12. https://doi.org/10.1109/ACIRS.2017.7986055

[32] Jiafan Yu, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. 2018. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule* 2, 12 (2018), 2605–2617. https://doi.org/10.1016/j.joule.2018.11.021

[33] Jiangye Yuan, Hsiu-Han Lexie Yang, Olufemi A Omitaomu, and Budhendra L Bhaduri. 2016. Large-scale solar panel mapping from aerial images using deep convolutional networks. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2703–2708.

[34] D Zhang, F Wu, X Li, X Luo, J Wang, W Yan, Z Chen, and Q Yang. 2017. Aerial image analysis based on improved adaptive clustering for photovoltaic module inspection. In *2017 International Smart Cities Conference (ISC2)*. IEEE, 1–6.