INFRARED SOLAR MODULE DATASET FOR ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

There are currently 485 gigawatts of solar energy capacity in the world and the industry is experiencing 29% yearly growth. In addition to faults that can be caused by manufacturing, transportation, and installation, these solar assets are subject to degradation from environmental factors throughout their lifespan and need to be inspected in order to ensure that power production conforms to the expected financial model. As the solar industry scales, inspections are increasingly relying on remote sensing. Inspecting thermal imagery of solar modules typically requires a trained expert to identify anomalies. However, this data is not widely available to machine learning researchers that have the means to automate these data challenges. This paper presents a new dataset, InfraredSolarModules, that contains different types of defects, faults, and findings that can be used as a basis for automating anomaly classification. [1]

1 INTRODUCTION

Solar photovoltaic (PV) energy production is a rapidly growing industry worldwide. It is a cost effective and environmental solution for communities, businesses, and governments that aim to become carbon neutral, or carbon negative (UNG; CNC). As the use of solar PV energy continues to grow, more efficient methods are needed to build, operate, and maintain the solar industry's growth.

The solar industry has been standardizing faster inspection methods that utilize remote sensing and non-destructive testing (NDT) via infrared imaging of solar panels (IEC TS 62446-3:2017). These methods include capturing visible and infrared images from an aircraft. Once imagery is collected, anomalies are identified and classified in order to estimate losses in power production. Currently, trained experts are required to analyze this data. The authors seek to automate this process with machine learning to allow experts to focus on restoring power production, rather than manually searching through imagery.

This paper presents a labeled dataset, InfraredSolarModules, that contains real-world imagery of different anomalies found in solar farms. This dataset can be used for machine learning research to gain efficiencies in the solar industry.

2 PROJECT SUMMARY

There are over 485 gigawatts of solar energy capacity in the world, which equates to more than 1.5 billion solar modules installed (Whiteman et al.). More remote sensing data is being collected to find anomalies, but infrared imagery is not widely available to researchers that have the skills to automate image classification problems.

In order to combat the lack of publicly available data on infrared imagery of anomalies in solar PV, this project presents a novel, labeled dataset to facilitate research to solve problems well suited for machine learning that can have environmental impact. This dataset has already been collected, labeled, and organized for the purposes of this project and is ready for further analysis.

[1]Dataset is available for download at `https://github.com/RaptorMaps/InfraredSolarModules`

Table 1: Classes

Class Name	Images	Description
Cell	1,877	Hot spot occurring with square geometry in single cell.
Cell-Multi	1,288	Hot spots occurring with square geometry in multiple cells.
Cracking	941	Module anomaly caused by cracking on module surface.
Hot-Spot	251	Hot spot on a thin film module.
Hot-Spot-Multi	247	Multiple hot spots on a thin film module.
Shadowing	1056	Sunlight obstructed by vegetation, man-made structures, or adjacent rows.
Diode	1,499	Activated bypass diode, typically 1/3 of module.
Diode-Multi	175	Multiple activated bypass diodes, typically affecting 2/3 of module.
Vegetation	1,639	Panels blocked by vegetation.
Soiling	205	Dirt, dust, or other debris on surface of module.
Offline-Module	828	Entire module is heated.
No-Anomaly	10,000	Nominal solar module.

The research presented in this paper builds the development of image processing and computer vision techniques to analyze the health of PV systems. For example, the application of RGB images to analyze soiling was explored by both Mehta and Yap (Mehta et al., 2017; Yap et al., 2015). Visual spectrum imagery has been used to identify solar farms (Bradbury et al., 2016). Electroluminescence imaging and sofware has been used defect various types of defects at the cell level of a PV modules (Deitsch et al., 2018; Buerhop et al., 2018). In similar fashion to our proposed dataset, aerial thermography was used for classifying defective modules (Aghaei et al., 2016). However we seek to highlight the class imbalance that is prevalent in solar PV anomaly detection and of broad interest in the research community (Johnson & Khoshgoftaar, 2019). The InfraredSolarModules dataset presented through this paper, contains a relatively large collection of real-world samples with 12 classes.

3 Technical Component

3.1 Dataset Description

The dataset consists of 20,000 infrared images that are 24 by 40 pixels each. There are 12 defined classes of solar modules presented in this paper with 11 classes of different anomalies and the remaining class being No-Anomaly (i.e. the null case). Table 1 lists each class along with a description.

This dataset includes real, unique anomalies of solar modules. Data was aggregated by the Raptor Maps team and collected by piloted aircraft and unmanned aerial systems equipped with midwave or longwave infrared (3-13.5 μm) and visible spectrum imaging systems. The image resolution varies from 3.0 to 15.0 cm/pixel. Anomalies were cropped to the individual module and separated into classes. Corresponding visible spectrum images were used during classification to increase accuracy. A canonical example of each anomaly is shown in Figure 1.

3.2 Dataset Discussion

It is noteworthy to compare the class proportions of data presented in this paper to the real world. Ewanich's global report on aerial inspections found that 2.2% of all modules have anomalies (Ewanich et al.). We elected to include 10,000 No-Anomaly images in the dataset to broadly classify nominal conditions. This may help to identify anomalies outside of the 11 classes included in InfraredSolarModules. Figure 2 shows the class percentages compared to the global findings and highlights the class imbalance. The dataset proportions are different in magnitude because more classes existed in the real world. The authors made a best effort to provide the most verified data possible so that researchers could readjust proportions to suit their needs.

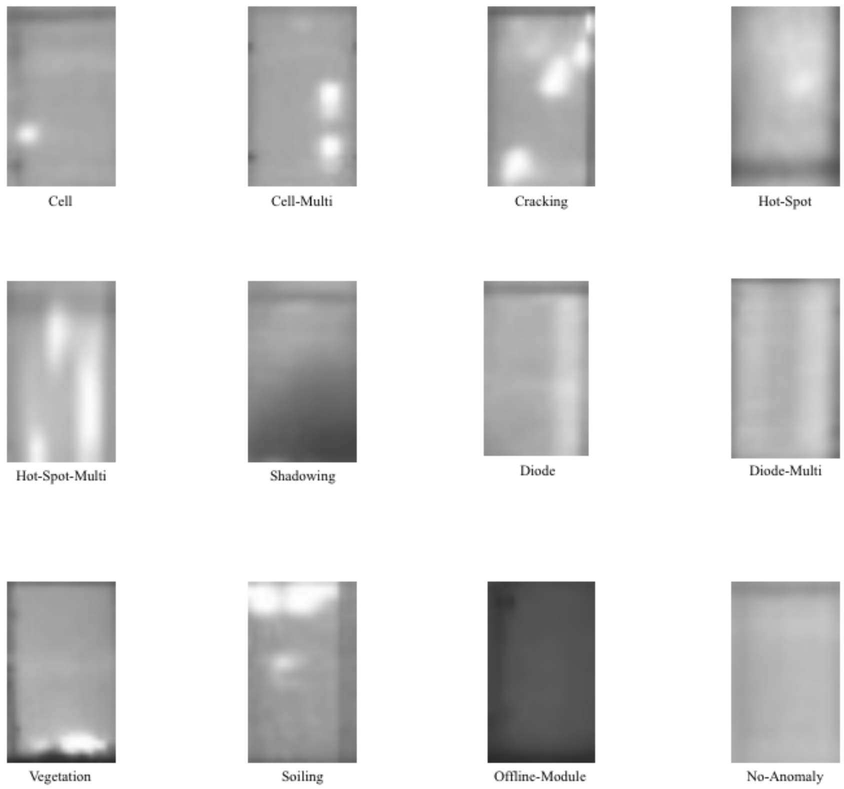

Figure 1: Canonical examples of solar module anomalies observable in infrared imagery.

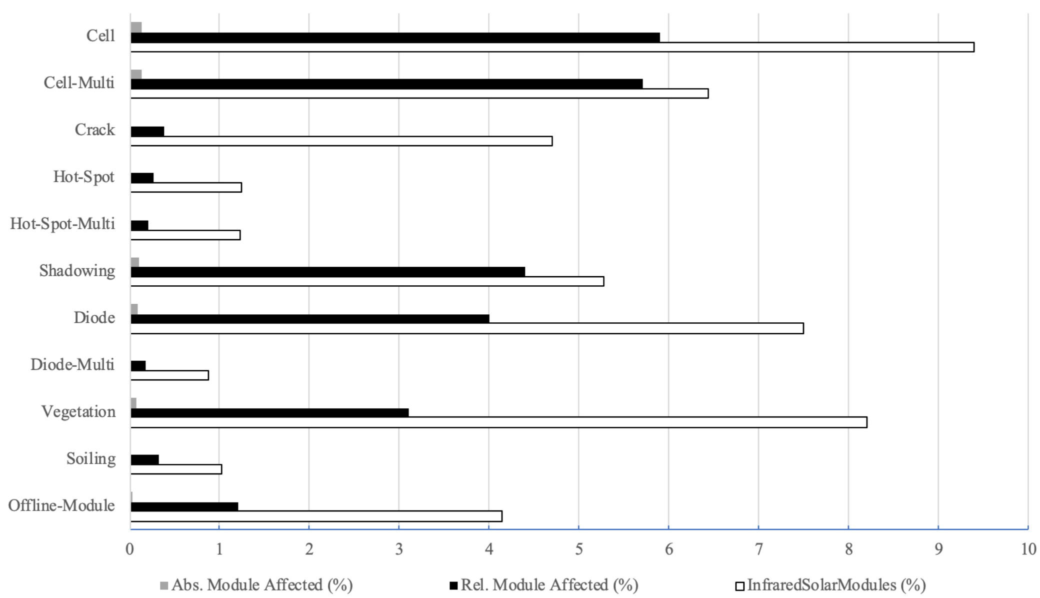

Figure 2: InfraredSolarModules versus Real World Proportions.

4 MILESTONES

The authors propose the following milestones for future work.

1. Bench marking classification results that can be shared broadly with the community.
2. Incorporation of more classes and the ability to recognize anomalies that are outside of the 12 classes of InfraredSolarModules.
3. Algorithm optimization towards real-time classification that can be used by low performance edge-computing devices.

4.1 Hosting and Processing

The authors are familiar with Azure and other cloud providers. The dataset presented in this paper would benefit from long term data storage to allow the research community to have continual and open access. The dataset is 81.9 megabytes, which is expected to grow by 30% each year as more data is collected and labeled, but may grow larger if the dataset includes higher-resolution visible spectrum data. Research projects that work with this data may need access to several GPUs to quickly train models and tune hyper-parameters.

5 Impact

This project specifically aligns with AI4Earth specifically with regards to climate change. Energy generation is one of the largest sources of carbon emissions and solar PV is one of the fastest and most economically viable forms of renewable energy in the world (Brunisholz et al.). Solar PV needs to scale to meet the demand. By efficiently identifying loss in power production, a positive feedback loop is created between environmental impact and solar PV.

In 2019, 1.6% of power production loss was identified by aerial inspections (Ewanich et al.). Restoring this production loss alone will offset 9.1 million metric tons of carbon dioxide emissions. Software and technology are lowering costs so more solar can be built without needing to train or hire as many technicians per every megawatt installed.

This software capability has a positive impact in any region that has solar or transitioning to solar. The research community can benefit from InfraredSolarModules to jump start any data collection or labeling efforts in solar PV. The dataset is provided in hopes that more datasets, techniques, and models are developed.

6 Team

Matthew Millendorf is a software engineer at Raptor Maps. A 2019 graduate of Brandeis University, Matthew holds a B.A. in Computer Science and a B.A. in Economics. With a strong interest in image processing and computer vision, Matthew actively contributes to Raptor Maps' artificial intelligence initiatives.

Edward Obropta is the co-founder, CTO of Raptor Maps. Eddie earned his B.S. and M.S. in Aeronautics and Astronautics from MIT. He has started a video game and technical clothing company and has worked on thermal dynamics and life support systems at SpaceX. Eddie has also built computer vision systems to measure material deformation for the US Navy (ONR).

Nikhil Vadhavkar is the co-founder, CEO of Raptor Maps. Nikhil earned his B.S. in Biomedical Engineering and M.S. in Engineering Management at Johns Hopkins University. He is 5 years into the Health Sciences and Technology PhD at MIT. Nikhil is a NASA Space Technology research fellow, and his research involves simulations for space suit life support systems. Nikhil has also earned a grant from the Gates Foundation to design drones for medical supply delivery in developing nations.

Acknowledgments

Funded by Raptor Maps R&D

REFERENCES

Members of the carbon neutrality coalition. https://www.carbon-neutrality.global/members/. Accessed: 2020-02-14.

Business leaders taking action. https://www.unglobalcompact.org/take-action/events/climate-action-summit-2019/business-ambition/business-leaders-taking-action. Accessed: 2020-02-09.

Mohammadreza Aghaei, Sonia Leva, and Francesco Grimaccia. Pv power plant inspection by image mosaicing techniques for ir real-time images. *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, pp. 3100–3105, 2016.

K. Bradbury, R. Saboo, T. L. Johnson, Jordan M. Malof, Arjun Devarajan, Wuming Zhang, Leslie M. Collins, and Richard G. Newell. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. *Scientific Data*, 160106, 2016. doi: 10.1038/sdata.2016.106.

Mary Brunisholz, Gaëtan Masson, and Izumi Kaizuka. 2018 snapshot of global photovoltaic markets. Technical report, Paris, France.

C. Buerhop, Sergiu Deitsch, Andreas Maier, Florian Gallwitz, Stephan Berger, Bernd Doll, Jens Hauch, Christian Camus, and Christoph Brabec. A benchmark for visual identification of defective solar cells in electroluminescence imagery. 09 2018. doi: 10.4229/35thEUPVSEC20182018-5CV.3.15.

Sergiu Deitsch, Vincent Christlein, Stephan Berger, Claudia Buerhop-Lutz, Andreas K. Maier, Florian Gallwitz, and Christian Riess. Automatic classification of defective photovoltaic module cells in electroluminescence images. *CoRR*, abs/1807.02894, 2018. URL http://arxiv.org/abs/1807.02894.

Kira Ewanich, Shane Carey, and Austin Coffin. Raptor Maps 2020 global solar aerial thermography report. Technical report, Boston, USA. URL "https://raptormaps.com/raptor-maps-releases-2020-global-aerial-thermography-report/".

IEC TS 62446-3:2017. Photovoltaic (PV) systems - Requirements for testing, documentation and maintenance - Part 3: Photovoltaic modules and plants - Outdoor infrared thermography. Standard, International Electrotechnical Commission, 2017.

Justin M. Johnson and Taghi M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27, 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0192-5. URL https://doi.org/10.1186/s40537-019-0192-5.

Sachin Mehta, Amar P. Azad, Saneem A. Chemmengath, Vikas Raykar, and Shivkumar Kalyanaraman. Deepsolareye: Power loss prediction and weakly supervised soiling localization via fully convolutional networks for solar panels. *CoRR*, abs/1710.03811, 2017. URL http://arxiv.org/abs/1710.03811.

Adrian Whiteman, Javier Esparrago, Sonia Rueda, Samah Elsayed, and Iana Arkhipova. Renewable energy capacity statistics 2019. Technical report, Abu Dhabi, UAE.

Wai Kean Yap, Roy Galet, and Kheng Cher Yeo. Quantitative analysis of dust and soiling on solar pv panels in the tropics utilizing image-processing methods. 2015.