Investigating the viability of detecting rooftop solar photovoltaics via convolutional neural networks using remote sensing images

Prakash P S
Irish Centre for High End Computing
University of Galway
Galway, Ireland
psprakashgis@gmail.com

Bharath Raj
Computer Science
Jain University
Bangalore, India
bharathrajpalivela@gmail.com

***Abstract*: One of the finest solutions to producing electricity from non-renewable sources is rooftop solar photovoltaics. It is crucial to diversify energy sources from renewable sources as much as possible from a consumption point of view in order to reach global net zero rapidly. In order to reduce loss and use the sun's energy, we must install rooftop solar photovoltaics for captive consumption to a significant level. Understanding current scale installations of solar photovoltaics on building roofs under various built-up conditions is crucial for achieving this. This study uses high resolution remote sensing images to train a lightweight convolutional neural network architecture referred to Faster R-CNN to identify solar panels installed on building roofs. Pre-trained models are trained using publicly accessible datasets to achieve generalization. The model is also evaluated using manually prepared labels in several cities around the world. The limits of such models in generalizing to different geographies are demonstrated through a visual examination and statistical model evaluation.**

Keywords: Faster R-CNN, Rooftop solar PV, object detection

I. INTRODUCTION

The International Energy Agency (IEA) estimates that solar energy currently accounts for 3.6% of all energy produced. According to a report from PV magazine, this production might expand 65-fold by 2050. According to reports, solar PV capacity has increased 20 times since 2010, from 40 GW to 800 GW. The price of modules reduced by nearly 90% over the same time span. As a result, renewable energy is now the cheapest source of electricity in many parts of the world. One of the finest solutions to producing electricity from non-renewable sources is rooftop solar photovoltaics. It is crucial to diversify energy sources from renewable sources as much as possible from a consumption point of view to reach global net zero quickly. For instance, household energy consumption makes up about 25% of total energy consumption in India [1]. The present 15% average rate of electricity losses during transmission and distribution has a big influence on sustainability and overall energy efficiency. In this situation, households have a prospect to actively contribute to lowering these losses and advancing towards a future of net-zero energy by utilizing rooftop solar photovoltaics for captive consumption.

Considering this, it would be useful to investigate on the number of building roofs that already have solar photovoltaics installed to monitor and plan for potential future actions. Using high-resolution remote sensing images from satellites or drones, it is feasible to effectively identify solar panels that have been installed on roofs. Process automation has been attempted in a number of different ways in the past. When developing models to identify solar panels from high-resolution images, convolutional neural networks are used. This method requires a large number of training datasets to start the process of training a model and assessing how well it works. Bounding boxes will be formed around the identified objects of interest, specifically solar panels, as a consequence of the object detection challenge under consideration.

II. LITERATURE REVIEW

The detection of rooftop solar panels has been studied extensively in the literature using methods like random forest [2] and segmentation based on convolutional neural networks [3]. The framework must, according to the authors, be expanded to include a variety of built environments. An extensive strategy to evaluate building integrated solar PV installations is laid out in a case study from the Netherlands [4]. The authors of the study created a modified U-Net architecture using VGG16 encoding. A software tool named SolarFinder was created to identify solar panels using a hybrid approach made up of K-means, support vector machines, and CNNs [5]. The study was conducted utilizing a significant amount of imagery from 13 different regions in northern America. Object detection approaches are also evaluated in the paper, in addition to image segmentation. To determine the most important socioeconomic characteristics that correlate with solar PV density, the DeepSolar machine learning framework was developed at the census track level [6]. In another case study, image segmentation models are trained using datasets from the Google Maps Static API [7] to detect solar PV panels.

Extensive training datasets are required to create convolutional neural network-based models for solar PV module detection. In the published literature, we discovered that some open repositories are helpful for initial model training. Many articles, however, lacked complete datasets that included images and labels [2]. The dataset from a study used several aerial photographs of varying spatial resolution to train a segmentation model to identify solar panels was made available to the scientific community so that it could be used to advance the development of models [8]. A research project [9] has created a crowd-sourced dataset that includes aerial images and annotations for solar PV installations, focusing specifically on the detection of solar panels on building rooftops.

III. THE IMAGERY DATASET

The dataset utilized in this study was obtained from the Zenedo open repository, as specified in the published research [9]. The dataset includes accurate installation masks for a total of 13,303 images sourced from Google Earth. Every picture has dimensions of 400×400 pixels and is in the 8-bit PNG format. The collection of pictures presented encompasses a diverse array of geographical areas, spanning around 11 thousand distinct cities. The dataset was chosen for the study partly because of its diverse range of places from which the samples were obtained. This characteristic renders it well-suited for training a model that can effectively generalize across numerous urban settings and surroundings. Figure 1 displays a representative example image along with the associated mask and bounding boxes created for the current research using the mask overlaid on top of the image.

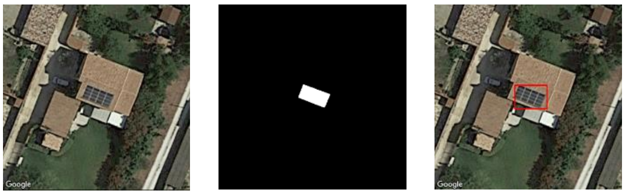

Fig. 1. Example of PV panels and their annotations

IV. Method

The four stages of the approach used in this study are shown in Figure 2. The dataset must first go through the necessary pre-processing stages, including organising the images and masks into folders, in order to be ready for the model's training. During the pre-processing stage, the extent of each masked PV panel is turned into bounding boxes because the training data is primarily created to do image segmentation. This takes place before the actual processing begins. The second stage involves training a FasterRCNN model and training it to specifically target rooftop solar PV panels using images and their associated bounding boxes. The model is assessed in the third step using certain samples from the dataset that were not utilized in the initial assessment metrics' training. The model's performance is finally assessed in the fourth stage utilizing manually compiled datasets covering various cities around the globe. Visual interpretation and statistical measurements like intersection over union are used in the evaluation to determine the accuracy of the model's predictions.

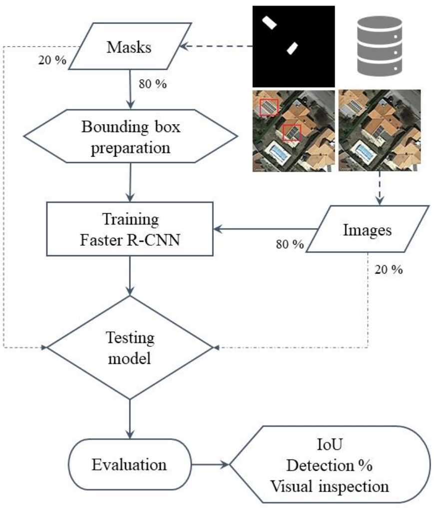

Fig. 2. The method implemented in the study

A. Data pre-processing

The selected dataset includes samples that are sets of pictures and masks divided into two distinct subfolders according to the origin of the images. By extracting the contours from each closed shape and creating a bounding box for each one in each mask, the masks were transformed into bounding boxes. These bounding boxes were used to construct a dataset in Pascal Visual Object Classes (VOC) format. In this annotation format, bounding boxes for object instances are represented as rectangles using coordinates xmin, ymin, xmax, and ymax. The dataset was then divided into train and test datasets with an 8:2 ratio

B. Model training

The method of fine-tuning a pre-trained CNN was chosen for the training process since it has been demonstrated to give significantly superior results than training a model from scratch. In this scenario, the Faster R-CNN architecture was chosen over comparable models like U-Net or Google DeepLabV3+ due to its quicker period for training and more effective use of computational resources. The selected architecture comprises a Faster R-CNN model with a ResNet-50-FPN backbone from its introductory paper [10]. The optimizer stochastic gradient descent (SGD) was chosen due to its computational efficiency and its ability to help the model generalize better by introducing stochasticity and noise tolerance during optimization. A learning rate scheduler that decreases the learning rate by 10 times after every 3 epochs was constructed to ensure stability and better convergence, preventing overshooting and improving generalization of the model. The model was then fine-tuned for 4 epochs each with 108 iterations with a total training time of about two hours on a NVIDIA Tesla T4 GPU at which point the loss reached convergence.

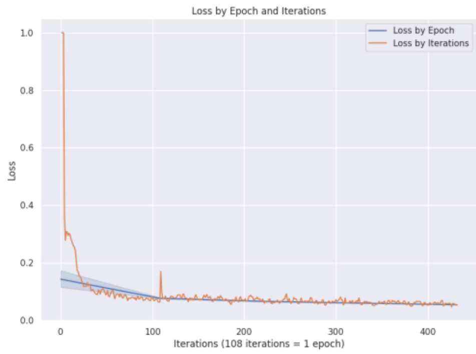

Fig. 3. Training loss by epoch and iterations

TABLE I. MODEL TRAINING RESULTS

Average precision	Average recall	Loss
0.805	0.850	0.053

C. Model evaluation

A dataset that was manually constructed was used to test the model. This dataset contains 85 images and the bounding boxes that correspond to those images from 9 distinct cities: Bangalore, Beijing, Kyoto, London, Los Angeles, Mauritius, New Delhi, New York City and Taipei City. These cities were chosen to make sure that a diverse representation of geographical places is achieved. These cities were chosen to represent a wide range of geographic areas, including various continents and regions. The dataset offers a thorough evaluation of the model's performance across various urban and sub-urban environments by including cities from various areas of the world. The Intersection Over Union (IoU), precision, and recall parameters were used to assess the model's correctness. The IoU is the overlap between predicted and ground truth bounding boxes serving as a metric to measure the accuracy of object localization. Precision helps assess the accuracy of positive predictions and recall provides insight into the model's ability to avoid missing instances

V. RESULTS AND DISCUSSIONS

Testing was carried out using a different dataset than that used for training because the metrics collected as given in table 1 alone do not provide a comprehensive evaluation of the model's performance. To obtain a more accurate evaluation of the model's effectiveness, the model was tested with the manually created dataset. According to Table II, this evaluation's findings indicate that the model performs on average with an IoU of 61%, a precision of 83%, and a recall of 67%. The high precision shows that the model correctly predicts boxes, but the moderate recall shows that the algorithm may be producing incorrect bounding boxes or failing to recognize some panels. This claim is further supported by the moderate IoU, which shows an average overlap between the anticipated and actual areas of only 61%. The visual inspection of the predictions suggests the same with predictions containing many inaccurate bounding boxes and an inconsistent detection of panels present in the images. The model may not be directly deployed for practical applications, as evidenced by the moderate IoU and recall. The initial metrics showed that the training loss was at 5%, indicating that overfitting would have resulted from further adjusting the model's parameters or increasing model complexity. It appears that a fine-tuned CNN is not a robust enough architecture for real-world applications. Figure 4 depicts three randomly selected predictions alongside their actual values for Bangalore, Los Angeles, and Taipe.

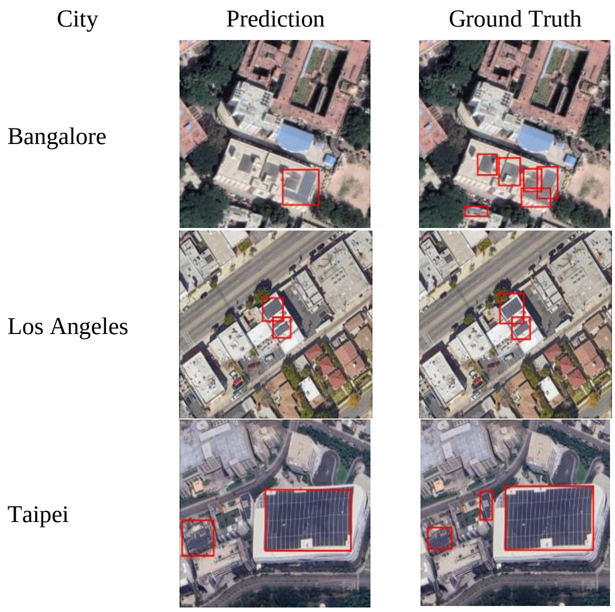

Fig. 4. An illustration of detecting PV panels from building roofs

TABLE II. EVALUATION METRICS

City	IoU	Precision	Recall
Bangalore	0.57	0.82	0.63
Beijing	0.56	0.68	0.58
Kyoto	0.73	0.84	0.86
London	0.70	0.90	0.77
Los Angeles	0.59	0.92	0.65
Mauritius	0.62	0.75	0.66
New Delhi	0.45	0.80	0.49
New York	0.75	0.95	0.78
Taipei City	0.55	0.85	0.62
Overall	**0.62**	**0.84**	**0.67**

VI. CONCLUSIONS

In brief, this study looked into whether or not it would be possible to use CNN's to detect rooftop solar panels from high resolution satellite images. By reusing a large dataset of remote sensing images from different places, fine-tuning a CNN architecture with the dataset, and testing the model with a manually-created dataset from different places, the study found that the architecture works okay but isn't good enough for real-world applications. This work does in depth measurement of the model's performance by evaluating it against a manually generated dataset covering a variety of locations. These metrics suggest that the fine-tuned CNN architecture might not be suitable for use in real-world scenarios. For the use of rooftop solar panel detection, since they are tiny compared to other spatial feat experimenting on more sophisticated architectures like Vision Transformers and fine-tuning them with larger datasets may lead to more practically useful results. In related computer vision applications, such as Meta's Segment Anything model created and constructed for a more generic semantic segmentation application, the Vision Transformers architecture has demonstrated considerable potential. Similar work with a comprehensive focus on the task of solar panel detection or segmentation has a lot of potential.

REFERENCES

[1] L. Saini, C. S. Meena, B. P. Raj, N. Agarwal, and A. Kumar, "Net Zero Energy Consumption building in India: An overview and initiative toward sustainable future," International Journal of Green Energy, pp. 1–18, Jul. 2021.

[2] J. M. Malof, K. Bradbury, L. M. Collins, and R. G. Newell, "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery," Applied energy, vol. 183, pp. 229-240, 2019.

[3] R. Castello, S. Roquette, M. Esguerra, A. Guerra, and J. L. Scartezzini, "Deep learning in the built environment: Automatic detection of rooftop solar panels using Convolutional Neural Networks," Journal of Physics: Conference Series, IOP Publishing, vol. 1343, no. 1, pp. 012034, October 2019.

[4] B. B. Kausika, D. Nijmeijer, I. Reimerink, P. Brouwer, and V. Liem, "GeoAI for detection of solar photovoltaic installations in the Netherlands," Energy and AI, vol. 6, pp.100111, 2021.

[5] Q. Li, Y. Feng, Y. Leng, and D. Chen, "SolarFinder: Automatic Detection of Solar Photovoltaic Arrays," 19th ACM/IEEE International Conference on Information Processing in Sensor Networks," pp. 193-204, Apr. 2020.

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

[7] P. Parhar, R. Sawasaki, N. Nusaputra, F. Vergara, A. Todeschini, and H. Vahabi, "HyperionSolarNet Solar Panel Detection from Aerial Images." doi: https://arxiv.org/pdf/2201.02107.pdf

[8] H. Jiang et al., "Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery," Earth System Science Data, vol. 13, no. 11, pp. 5389–5401, Nov. 2021.

[9] G. Kasmi et al., "A crowdsourced dataset of aerial images with annotated solar photovoltaic arrays and installation metadata," Scientific Data, vol. 10, no. 1, Jan. 2023.

Ren, Shaoqing, et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" arXiv, 6 Jan. 2016.