DELFT UNIVERSITY OF TECHNOLOGY

THESIS PROPOSAL

Automated rooftop solar panel detection through Convolutional Neural Networks

---

Simon Pena Pereira
StudentID: *5391210*

1st Supervisor: Azarakhsh Rafiee
2nd Supervisor: Stef Lhermitte

TU Delft Delft University of Technology

Contents

1 Introduction

As one of the main drivers of the greenhouse effect, the energy sector is increasingly shifting towards more renewable and sustainable energy sources to face climate change by reaching policy-made climate targets. A popular option to contribute to this energy transition is the installation of photovoltaic (PV) systems on rooftops. That allows a renewable energy supply by solar PV systems in a decentralized manner. Depending on the country, there exist well to poorly documented registries of active PV systems, which are a hurdle for planning and achieving an efficient energy transition.

An alternative method for populating the registries with up-to-date information about installed PV panels is the use of machine learning algorithms based on satellite or aerial imagery. In the field of remote sensing, machine learning algorithms have proven their ability to classify various types of objects.

The first milestone in computer vision is the neural network architecture Necognitron introduced by Fukushima in 1980, followed by one of the most popular publications about Convolutional Neural Networks (CNNs) in 1998 by Yann LeCun which deals with the automated recognition of hand-written digits [1, 2]. Due to the increase in computing power in recent years, more and more deep neural networks were presented to solve object detection tasks in computer vision, and thereafter in the remote sensing domain. Building on that knowledge, many CNN architectures have been implemented in remote sensing analysis of the urban environment, for instance, for the purpose of detecting roads, buildings, or vehicles. With the increasing interest in PV panel installations, new applications for remote sensing techniques emerged in the solar energy domain, such as the estimation of solar energy, the automated selection of potential solar plant sites, the PV potential of building rooftops, or the localization of PV panels.

To optimize the effectiveness as well as the efficiency of CNN, research has mainly focused on the technical configurations of CNN architectures. However, the performance of CNNs is calculated based on the network's prediction in comparison to the ground truth data. Therefore, it is also prudent to understand the impact of diverse ground truth data, for instance from different geographic regions, on the performance of the network. The importance of analyzing the detection of PV panels with regard to geographical and architectural differences becomes evident when statistics about PV panels are created on a national or international scale.

2 Related work

Setting the thesis topic into a broader scientific context will be done by highlighting related work in the field of remote sensing that discusses automated object detection in images. Therefore, three different research approaches to detect PV panels will be introduced, namely (1) supervised machine learning algorithms, (2) CNNs for object detection, and (3) CNNs for semantic segmentation.

2.1 Classification and regression analysis

After numerous studies have proven the applicability of machine learning algorithms for object detection in the urban environment, researchers from Duke University carried out multiple research projects on PV panel detection. The first research on PV panel detection was published by Malof et al. in 2015 [3]. The objective was to obtain information about the locations, capacity, and energy production of installed PV panels on rooftops. The significance of the results is rather low due to a small amount of no more than 100 rooftop labels as an input. Nevertheless, having a correct detection rate of 94% proves an excellent performance for detecting PV panels. Similar to later introduced CNNs, the algorithm consists of two stages where the first one is prescreening the image to extract potential regions followed by the second stage in which features are extracted and classified. For the classification process, the support vector machine (SVM) classifier was used. For the preprocessing steps in this thesis, important insights can be taken from Bradburry et al.'s publication on creating PV array datasets for object identification. The dataset contains nearly 20.000 manually annotated PV panels from four different cities based on aerial orthoimages with a resolution of 30 cm. Besides technical insights, it also provides a broader impression of aspects to focus on when annotating the PV panels [4]. Building upon this dataset Malof et al. presented another project in 2016, in which they conducted more research on detecting PV arrays. Although this project is more comprehensive than the previous approach, the pixel-based analysis with a random forest (RF) classifier is not sufficient to predict PV array locations due to its performance drop between the training and testing datasets. The outcome of the project proved overfitting to training data which prevented an efficient classification of testing data [5].

2.2 Convolutional Neural Networks for Image Classification and Object Detection

To move on from traditional machine learning algorithms to more advanced deep learning techniques, in particular CNNs, this section is introducing the DeepSolaris project that was carried out by four national statistical offices from the Netherlands, Germany, and Belgium, together with the Open Universiteit Nederland [6]. It is especially interesting as a reference to this thesis since it includes the state of North Rhine-Westphalia (NRW) as a study area which is the area of interest (AOI) for this thesis. Therefore, similar datasets will be used in this thesis. Prior to this publication, Curier et al. published a paper describing the data, its pre-processing steps as well as the annotation process for creating label data in more detail [7]. These are crucial information for creating a suitable input dataset for this thesis. Besides NRW, the algorithms are also trained on the province of Limburg in the Netherlands. The project aims to create a new way of producing official statistics by retrieving information from aerial images with deep learning algorithms. Thus, the main objective is to map the locations of PV panels at a regional to local level and to provide a better understanding of the energy transition in the context of PV panels. For reaching that objective, two approaches were tested. Firstly, the approach of image classification, in which the algorithm predicts whether the image contains a PV panel or not. This approach demands a different method of annotating the images because it does not determine the precise PV panel location within the image but the image itself. They applied both models, InceptionResNetV2 as well as VGG16 with pre-trained weights. Secondly, the approach of object detection which is divided into two stages similar to the approach described in section 2.1. While the first stage is proposing potential pixel regions for the presence of an object, the second stage is localizing the object. The most common localization method is the use of bounding box regressions to predict the object's exact location. In this approach, they went one step further by applying the Mask R-CNN algorithm that computes pixel-based masks of the object, in addition to the bounding box. Overall, the DeepSolaris project demonstrated the ability of CNNs to detect PV panels in an almost automatic manner. It furthermore succeeded in detecting 24% of so far unknown PV panels in the cities of Bonn and Düren, which underlines the need for alternative ways of detecting PV panels for national registries. Nevertheless, improvements are required concerning the number of false-positive detections. Also, performance drops were detected when training and validating the networks for different geographic regions due to overfitting to one specific region. Additionally, the impact of using different aerial images on the same region needs to be considered. In the following, the reasons for performance drops are pinpointed as suggestions for further research. These reasons are especially relevant for the research questions of this thesis:

- differences between rural and urban areas
- differences between cities with different architectures
- a variation in time of the day/period/year in which the photos were taken
- different spatial resolutions

2.3 Convolutional Neural Network for Semantic Segmentation

In an additional project by Malof et al., previous shortcomings were tackled by shifting to CNNs which achieved major improvements in object recognition. The applied CNN architecture was inspired by the designs of the Visual Geometry Group (VGG) at Oxford [8]. Following their previous work, they closed the performance gap between training and testing datasets and achieved a recall rate of 0.8 and a precision of circa 0.95 on a testing dataset. This project proves how CNNs are superior to traditional machine learning algorithms [9]. In a similar approach for mapping the location and size of PV panels, Castello et al. proposed in 2019 a CNN with U-Net architecture for image segmentation of high-resolution aerial images. As in Malof et al.'s project in 2017, it outputs a pixel-based binary raster predicting either a PV class or no-PV class per pixel. Overall, the proposed algorithm achieved an accuracy of 0.94 [10].
One of the most recent studies on semantic segmentation (also called image segmentation) for detecting PV installations was published by Costa et al. in 2021. Unlike previous studies, the focus was on solar plants and not on small-scale PV panels on rooftops. Also, Sentinel-2 imagery is used instead of high-resolution aerial images as well as a near-infrared (NIR) band in addition to the RGB bands. The project discusses the performance differences of four CNN architectures, namely U-Net, DeepLabv3+, Pyramid Scene Parsing Network, and Feature Pyramid Network, combined with four different backbones (Efficient-net-b0, Efficient-net-b7, ResNet-50, and ResNet101). Although the U-Net-Eff-b7 combination achieved the best results, it needs to be highlighted that the performances of all model–backbone combinations were sufficient. The main outcome of their work is that results depend rather on the quality of the input labels than on the models [11].

3 Research questions

Given the name of the thesis, its generic aim is the implementation of a CNN for the automated detection of PV panels on rooftops. To narrow down the scope of the thesis, its objective lies in the understanding of technical and physical aspects that potentially have an impact on the prediction of a CNN with U-Net architecture. According to its objective, the main research question is the following:

To what extent is a CNN with U-Net architecture suitable for detecting PV panels on rooftops?

Understanding why the algorithm is performing in a certain way is crucial to improve it, in case it underperforms, but also to modify it properly for different purposes. For that reason, research is conducted regarding the model's hyperparameters. Furthermore, the impact of both NIR and building outline data as additional information to the aerial RGB image will be evaluated. A further aspect to examine is the effect of lower resolution aerial imagery. To understand the physical parameters of detecting a PV panel, it is important to consider its geographical and architectural environment. As stated in the previous section 2.2, the AOI is the state NRW due to the availability of high-resolution aerial imagery and PV panel locations. Since building a labeled training, validation, and test set for the entire state is out of the thesis' scope, its focus is set on both a small-scale urban area as well as a small-scale rural area. This comparison allows research on the suggestion to compare urban and rural areas by Jong et al. stated in section 2.2. Regarding the closer environment of the PV panels, research will be conducted on the impact of different PV panel sizes and rooftop colors. Lastly, it will be examined if the solar capacity can be derived from the pixel-based prediction mask of a PV panel. The estimation can then be compared to the actual capacity stated in the metadata of the PV panel. The research guide is listed below in the form of sub-questions:

- *What is the effect of adding near-infrared data to aerial images on the detection of PV panels?*
- *What is the impact of urban and rural configurations on the detection of PV panels?*
- *How is the roof color affecting the detection of PV panels?*
- *To what extent can the power capacity be estimated based on the predicted PV panel mask?*
- *How much does the panel size affect the detection rate?*
- *How sensitive is the model towards lower resolutions?*

4 Methodology

The methodology of this thesis is divided into three main parts, namely (1) pre-processing the data, (2) applying the CNN with U-Net architecture, and (3) the evaluation of the results.

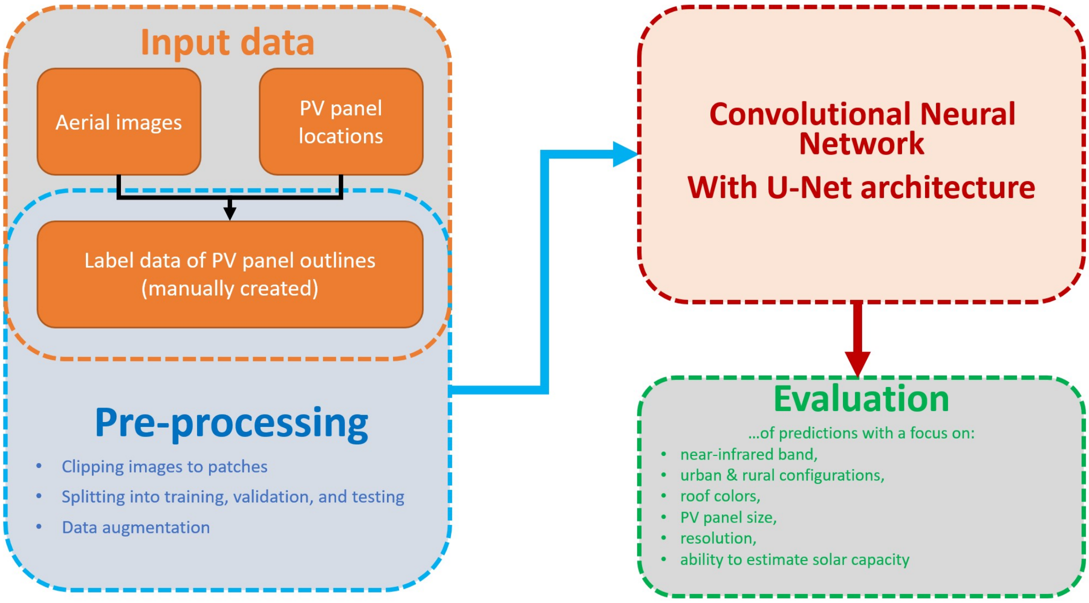

Figure 1: Methodology workflow

4.1 Pre-processing input data

There are two datasets that need to be pre-processed. Firstly, the aerial images, and secondly, the point vectors located at the addresses of commercial and public buildings with PV panels on their rooftops. PV panels on private rooftops need to be detected manually.

4.1.1 Creating ground truth label

In order to create ground truth labels of the exact PV panel outline, a grid covering the AOI will be created to facilitate a systematic labeling process. Based on the aerial image and the grid, the PV panel outline is manually drawn and stored as a polygon feature. Metadata regarding PV panel size, roof color, and solar capacity will be added to meet the research questions in the evaluation part. Then, the polygon needs to be converted to a binary raster to provide a pixel-based ground truth mask of the PV panel. The labeling process is one of the most essential aspects of a successful detection because it determines the quality of the ground truth which is the reference for validating the model's performance. Potential errors to be aware of are missing PV panels, incorrectly identified PV panels, and incorrectly drawn outlines [4]. Obstacles that need to be overcome are skylights, glass roofs of balconies or conservatories as well as thermal collectors.

4.1.2 Clipping input data

The aerial images and the label images need to be clipped to uniform patch sizes, to be forwarded as a compatible input to the CNN. A commonly used patch size is 256 x 256 pixels [12, 13]. However, it is prudent to compute results first and adapt to a different patch size if needed. Also, depending on the computational demand, the patch size might be changed to allow more efficient computations. Furthermore, it depends on the actual area size that is covered by the patch in proportion to the object size. For instance, 256 x 256 pixels at a spatial resolution of 10 cm is equivalent to 25.6 x 25.6 m, so circa 655 m^2, which could be too small to catch the entirety of larger rooftop PV panels. As a reference the DeepSolaris project is taken, in which different models used a patch size of 330 x 330 pixels on images at the same resolution as images used in this thesis [6].

4.1.3 Splitting input data

In this step, the collection of patches needs to be split into three new collections for the purpose of training, validating, and testing the CNN. Commonly, the splitting ratio is around 70% for training, 20% for validation, and 10% for testing as in the project of Costa et al. [11].

4.1.4 Data augmentation

Applying the method of data augmentation is especially important for this thesis since it enriches the size and variety of input patches for the model. The augmentation is based on horizontal and vertical flips of patches that are added to the original datasets. This method enhances the robustness of the model by reducing the effect of overfitting as proven by Jong et al. [6]. Splitting needs to be carried out before data augmentation to avoid the case that augmented patches are part of more than one dataset. Having the same patch in the training and test dataset would not allow an unbiased evaluation of the model's accuracy since the patch was already seen by the model in the training dataset.

4.2 Convolutional Neural Network with U-Net architecture

For the purpose of detecting and predicting PV panels, a CNN will be used. A CNN is a deep network with a hierarchical structure to extract high-level semantic information from images by recognizing patterns. The raw input images are fed into the model as stacks, so-called batches. The model consists of stacked operation layers, namely convolutional layers, pooling layers, and non-linear activation functions [14]. In a convolution layer, a filter matrix with random weights is sliding across the input image to compute the dot product with the pixel values of the image. The output matrix is a feature map that serves as an input for the following convolutional layer. Spatial padding operations are applied to preserve the same spatial resolutions after each convolutional filter [8, 12]. The entire process of images propagating through the network is called feed-forward operation. At the end of the network, the errors between the ground truth values and the predicted values are calculated to adjust the weights of the network in a process called back-propagation [14].
The network architecture that is going to be used in this thesis was presented by Ronneberger et al. in 2015 and is called U-Net since it is shaped like the letter U. As illustrated in figure 2, it consists of a contracting path on the left side, where the input is downsampled by max-pooling layers, and an expansive path on the right side, where upsampling operations are carried out by up-convolutional layers [12].

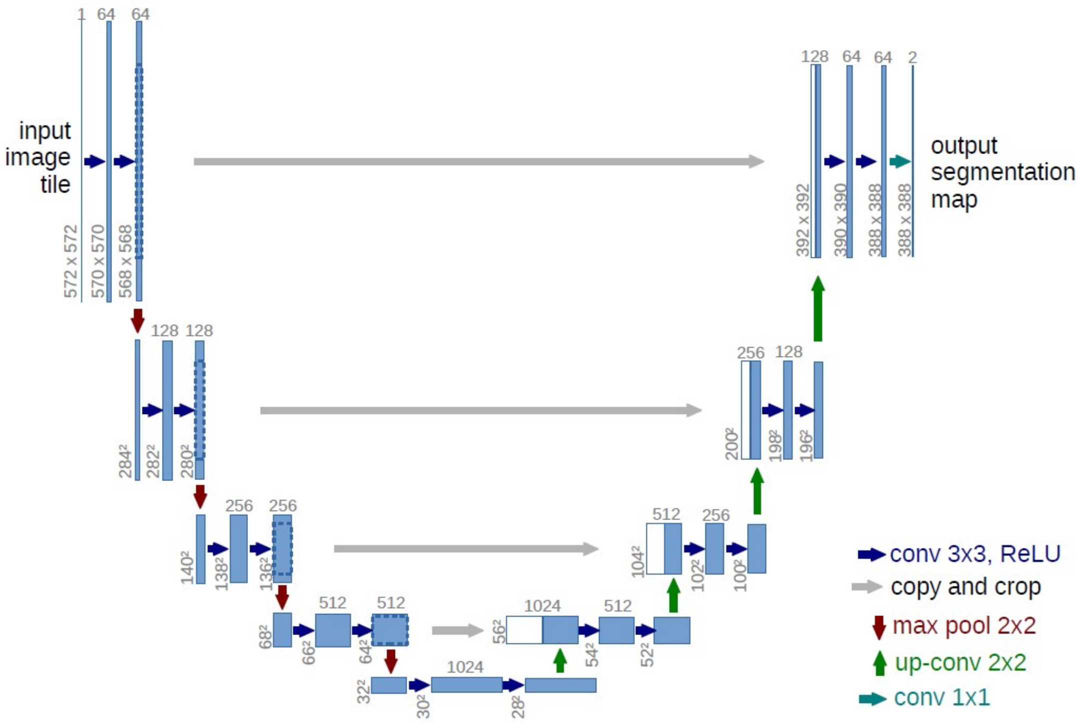

Figure 2: U-Net architecture showing a feed-forward operation with multi-channel feature maps as blue boxes [12]

4.3 Evaluation

After having the model trained and validated, the test dataset will be used to predict output segmentation maps that are evaluated based on standard metrics. The most common metric is the *accuracy* which is defined by the number of correctly predicted images divided by the total number of predictions. Correctly predicted images include true positives, where PV panels are correctly identified as well as true negatives, where the absence of PV panels is correctly identified. Another metric called *precision* calculates the proportion of true positives to the total number of actual PV panels, including true positives and not identified PV panels (false positives). To express the proportion of correctly identified PV panels to all predictions of PV panels, the *recall* metric will be applied. It is calculated by the number of true positives divided by the number of true positives and false negatives. Last but not least, there is the *F1 score* which expresses the harmonic average of precision and recall [6].
These metrics will be applied to different scenarios that are outlined by the research questions in section 3. For instance, comparing the results of urban and rural areas as well as training on urban areas while predicting on rural areas and vice versa. Furthermore, the metrics will be applied on predictions that are including NIR data, and different spatial resolutions. Lastly, the labelled test data will be compared to the predictions to examine their accuracies with regard to the roof color, the PV panel size, and the ability to estimate solar capacity.

5 Time planning

Start date	End date	Tasks
7 Feb	15 Apr	Finding a thesis topic
		P1 – Registration of topic
15 Apr	29 May	Literature and dataset review
30 May	3 Jun	Review of neural networks
		P2 – Graduation Plan
15 Jun	31 Jul	Labeling aerial images
11 Jul	19 Aug	Pre-processing labels and images
1 Sepr	30 Sep	Setting up the CNN
		P3 – Midterm progress meeting
1 Oct	14 Oct	Applying the model
10 Oct	15 Dec	Writing thesis
		P4 – Go/no go
15 Dec	9 Jan	Finalizing the thesis
9 Jan	19 Jan	Presentation preparation
		P5 – Public presentation and final assessment

6 Tools and datasets used

Two open available datasets will be used. Firstly, there are the digital orthophotos from Geobasis.NRW, the spatial data infrastructure (SDI) of NRW. The orthophotos from 2020 are provided as tiles covering areas of 1 km^2. They consist of four 8-bit bands (RGB + NIR) with a spatial resolution of 10 cm [15]. Secondly, the dataset of the PV panels is partially provided by the national German SDI as part of a dataset of renewable energy sources in NRW from 2016 [16]. The layer of rooftop PV panels is available as a shapefile and Web Feature Service (WFS) for PV panels on public and commercial buildings. The outlines of PV panels will be drawn manually in QGIS. The CNN model will be written using the Python programming language in a Google Colab-Notebook to run the model on a GPU provided by the Google cloud. The open-source machine learning library TensorFlow will be used in this environment to construct the CNN [17].

References

[1] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] J. Malof, R. Hou, L. Collins, K. Bradbury, and R. Newell, "Automatic solar photovoltaic panel detection in satellite imagery," pp. 1428–1431, 11 2015.

[4] K. Bradbury, R. Saboo, T. L. Johnson, J. M. Malof, A. Devarajan, W. Zhang, L. M. Collins, and R. G. Newell, "Distributed solar photovoltaic array location and extent dataset for remote sensing object identification," *Scientific Data*, vol. 3, no. 1, p. 160106, 2016.

[5] 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, 2016.

[6] T. de Jong, S. Bromuri, X. Chang, M. Debusschere, N. Rosenski, C. Schartner, K. Strauch, M. Boehmer, and L. Curier, "Monitoring spatial sustainable development: semi-automated analysis of satellite and aerial images for energy transition and sustainability indicators," *CoRR*, vol. abs/2009.05738, 2020.

[7] R. L. Curier, T. J. A. D. Jong, K. Strauch, K. Cramer, N. Rosenski, C. Schartner, M. Debusschere, H. Ziemons, D. Iren, and S. Bromuri, "Monitoring spatial sustainable development: Semi-automated analysis of satellite and aerial images for energy transition and sustainability indicators," *CoRR*, vol. abs/1810.04881, 2018.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2015.

[9] J. Malof, L. Collins, and K. Bradbury, "A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery," pp. 874–877, 07 2017.

[10] 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*, vol. 1343, p. 012034, 11 2019.

[11] M. V. C. V. da Costa, O. L. F. de Carvalho, A. G. Orlandi, I. Hirata, A. O. de Albuquerque, F. V. e. Silva, R. F. Guimarães, R. A. T. Gomes, and O. A. d. C. Júnior, "Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation," *Energies*, vol. 14, no. 10, 2021.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[13] O. Rukundo, "Effects of image size on deep learning," *CoRR*, vol. abs/2101.11508, 2021.

[14] M. Alam, J.-F. Wang, C. Guangpei, L. Yunrong, and Y. Chen, "Convolutional neural network for the semantic segmentation of remote sensing images," *Mobile Networks and Applications*, vol. 26, pp. 1–16, 02 2021.

[15] "Digitale orthophotos." https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/luftbildinformationen/aktuell/digitale_orthophotos/index.html. Accessed: 2022-06-05.

[16] "Kartenlayer dachflächenphotovoltaikanlagen." https://www.govdata.de/web/guest/suchen/-/details/kartenlayer-dachflachenphotovoltaikanlagen. Accessed: 2022-06-05.

[17] "Google colaboratory." https://colab.research.google.com/?utm_source=scs-index. Accessed: 2022-06-05.