Data Science Lab 1474
WS 2023/24

PV-Panel Detection from Aerial Pictures

German Gnetov
Alexander Hillisch
Sebastian Ostovits

Project Parnter: SIEMENS AG

Institution: Wirtschaftsuniversität Wien

Data Coaches: Danilo Valerio, Erwin Flitz, Stefan Bischof

Date of Submission: 31.01.2024

Department of Information Systems & Operations Management, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

Contents

List of Figures

1 Abstract

This project, conducted by Group Siemens 1 in collaboration with Siemens AG Österreich and advised by Data Coaches from Wirtschaftsuniversität Wien, focuses on developing a machine learning model to detect solar panels from Aerial pictures from Vienna. The primary aim is to assist electricity network operators, such as Wiener Netze, in identifying the location of photovoltaic (PV) panels, which is crucial for planning network extensions. The approach involves employing Convolutional Neural Networks (CNNs), with a particular emphasis on using the U-Net model for its proficiency in image segmentation tasks. The project utilizes two open available datasets, which in total contain 2395 pairs of features and labels. These images are preprocessed using techniques like data augmentation and efficient image and mask management. Preliminary results have demonstrated the model's effectiveness, achieving a precision on validation images of 92 percent. Furthermore, after conducting additional tests using screenshots from Google Maps of Vienna, our photovoltaic (PV) detection algorithm has demonstrated promising results. However, challenges like limited GPU capacity have been encountered, suggesting the need for more robust computational resources. Future steps also could include the creation of a labeled training dataset of Google Maps images from Vienna, to further enhance the models performance. Additional, steps that were out of scope for this project include the creation of a user friendly interface for a better user experience.

2 Introduction

In the realm of urban energy management, the transition towards renewable energy sources, especially solar energy, is not only an environmental imperative but also a complex logistical challenge. This critical aspect of this transition is addressed by energy companies like Siemens. In order to get a better understanding of the energy network, they are in need of efficient and accurate detection of solar panels in urban settings. Therefore, the integration of advanced machine learning techniques in traditional sectors like energy management is becoming increasingly vital. The accurate mapping of these PV installations is crucial for effective grid management and planning, especially in the context of increasing reliance on decentralized renewable energy sources. This project aims to provide a solution that not only enhances the operational efficiency of these operators but also contributes to the broader goal of sustainable urban energy management.

Central to the project's methodology is the use of Convolutional Neural Networks (CNNs), with a particular focus on the U-Net model. The choice of U-Net is underpinned by its proven effectiveness in detailed and precise image segmentation tasks, a feature crucial for accurately identifying solar panels from aerial images. Developed initially for biomedical image segmentation, U-Net's architecture is particularly adept at handling the intricacies of complex image data, making it an ideal choice for the project's objectives [1]. Moreover, the U-Net model has been extensively validated and refined across various studies, further establishing its robustness and suitability for such high-precision tasks [2].

The project's approach to data processing and model training involved two comprehensive datasets, comprising 2395 pairs of features and labels. [3, 4]This substantial dataset provided the necessary foundation for training and refining the machine learning model. Key to this process was the implementation of advanced preprocessing techniques, including data augmentation. This step was critical in enhancing the modelâ€™s capability to generalize from the training data to real-world scenarios, thereby improving its overall accuracy and reliability.

Preliminary results from the project have been highly promising, indicating a high degree of accuracy in PV panel detection. The model achieved a precision rate of 92 percent on validation images, a testament to the effectiveness of the chosen methodologies and the potential of machine learning in enhancing urban energy infrastructures. However, the project also faced challenges, such as limitations in computational resources, which highlight areas for future improvement. The next steps for the project include the exploration of creating a specialized training dataset using Google Maps images of Vienna and the development of a user-friendly interface to enhance the technology's accessibility and applicability.

3 Methodology

3.1 Overview of Image Segmentation Models

Image segmentation, a fundamental aspect of computer vision, involves partitioning a digital image into multiple segments to simplify and change the representation of an image into something more meaningful and easier to analyze. This process is typically used to locate objects and boundaries (lines, curves, etc.) in images. The effectiveness of image segmentation directly impacts the performance of an imaging system, influencing tasks such as object recognition, tracking, and even image compression.

3.1.1 Threshold Based Segmentation

Threshold based segmentation is one of the simplest methods of image segmentation. It involves partitioning an image into foreground and background by selecting an intensity threshold. Pixels above the threshold are classified as one object (usually the foreground), while those below are classified as another (usually the background). This technique is particularly effective in scenarios with high contrast between the object and the background. Despite its simplicity, thresholding has various forms like global thresholding, adaptive thresholding, and Otsu's method, each catering to different types of images and applications. Its wide use in various fields, from medical imaging to traffic control systems, highlights its fundamental role in image processing [5].

3.1.2 Graph Based Segmentation

Graph-based segmentation methods consider the image as a graph, where pixels are represented as nodes, and edges define the relationships between them. In these methods, segmentation is achieved by partitioning this graph into subgraphs, often using techniques like min-cut/max-flow algorithms. These approaches are particularly adept at capturing and utilizing the complex spatial relationships between pixels, making them suitable for detailed and nuanced segmentation tasks, such as in satellite imagery and biological imaging. Graph-based methods are known for their precision and ability to handle intricate structures within images, but they can be computationally intensive, especially for large-scale images [6].

3.1.3 Clustering Based Segmentation

Clustering based segmentation techniques, like K-means, Fuzzy C-means, and DBSCAN, are based on the concept of organizing pixels into clusters based on their properties, such as color or texture. K-means clustering, one of the most popular methods, partitions an image into K distinct non-overlapping clusters, where each pixel is assigned to the cluster with the nearest mean. Fuzzy C-means, on the other hand, allows pixels to belong to multiple clusters with varying degrees of membership, providing a more flexible segmentation approach. These methods are particularly well-suited for natural scene analysis and satellite imagery, where the segmentation of different landforms or objects based on their textural or color properties is essential [7].

3.1.4 Bayesian Based Segmentation

Bayesian based segmentation utilizes probabilistic models and is particularly robust in uncertain and ambiguous environments. This approach is grounded in Bayesian inference, where the probability of a hypothesis (segmentation) is updated as more evidence or information becomes available. Bayesian segmentation methods are adept at handling noise and other uncertainties in images, making them particularly useful in applications such as medical imaging, where the distinction between different tissues or structures might not be very clear. These methods, however, require a good understanding of the statistical properties of the image data and can be computationally demanding [8].

3.1.5 Neural-Network Based Segmentation

The introduction of neural networks, particularly Convolutional Neural Networks (CNNs), has revolutionized the field of image segmentation. These models, through their layered architecture, learn to capture hierarchical patterns in the data, making them extremely effective for high-dimensional data analysis found in complex image segmentation tasks. CNNs, with their ability to automatically and adaptively learn spatial hierarchies of features from the input images, have been at the forefront of this revolution. They have been widely applied in diverse fields ranging from autonomous vehicles to medical diagnostics, offering unparalleled accuracy in object detection and segmentation [9].

3.2 Data Handling and Model Training

Effective image segmentation heavily relies on the methodologies employed in data handling and model training. These components are crucial in determining the performance of the segmentation models, especially when using Convolutional Neural Networks (CNNs).

3.2.1 Data Collection for Object Detection Using CNNs

The data collection process for object detection using CNNs is multifaceted and critical for the success of the model. Quality and volume of the dataset play a pivotal role in the accuracy of CNN models. High-quality datasets ensure that the model learns relevant features, while the volume of the data helps in generalizing these features to new, unseen data. The process involves collecting a diverse range of images that represent various scenarios where the object of interest appears. This diversity is key in training a robust model capable of detecting objects under different conditions, such as varying lighting, angles, and backgrounds. Additionally, the importance of temporal and spatial diversity in datasets cannot be overstated, as it equips the model to handle real-world variability effectively [9].

3.2.2 Challenges and Techniques in Data Collection

Data collection for CNN-based object detection is laden with challenges. One of the primary challenges is the identification and labeling of target objects in the collected images, which is often labor-intensive and requires meticulous attention to detail. Automated image scraping techniques can expedite the process of gathering a large number of images, but they often need to be supplemented with manual methods to ensure the relevance and quality of the collected data. Manual annotation, although time-consuming, provides the high-quality, accurately labeled data necessary for effective training of CNNs. Techniques like crowd-sourcing annotations can also be employed to gather a large volume of data within a relatively short time frame [7].

3.2.3 Organization and Management of Data

Organizing and managing the collected data efficiently is vital for its effective utilization during training. This involves structuring the data in a way that is easily accessible and processable by the training algorithms. Effective data organization not only facilitates smoother training but also aids in quick iteration over different models during the experimental phase. Additionally, proper data management involves creating a balance between the different classes of data, ensuring that the model does not become biased toward one class due to uneven data distribution [8].

3.2.4 Efficiency in Data Handling

Handling large datasets efficiently is a significant challenge in CNN-based image segmentation. Techniques to manage memory usage and processing efficiency include data compression without losing critical information, efficient data loading techniques like data generators, and the use of databases to store and retrieve data efficiently. The use of cloud-based storage and computing resources can also play a crucial role in handling large datasets effectively, providing scalability and flexibility in data processing [5].

3.2.5 Data Augmentation and Normalization

Data augmentation and normalization are critical preprocessing steps in preparing the dataset for CNN training. Data augmentation involves artificially expanding the dataset by applying various transformations like rotation, scaling, cropping, and flipping to the images. This not only increases the volume of data but also introduces variability, making the model more robust to changes in input. Normalization, on the other hand, involves scaling the pixel values of the images so that they have a specific mean and standard deviation. This standardization of data helps in accelerating the training process and improving the model's performance by ensuring consistent data input [6].

3.3 Model Development: U-Net

The development of the U-Net model represents a significant advancement in the field of image segmentation, especially in applications requiring precise and detailed segmentation such as biomedical imaging.

3.3.1 Advantages and Capabilities of U-Net

U-Net's architecture, characterized by its symmetric contracting and expansive paths, is specifically designed for efficient segmentation of images. This model is particularly known for its precision in segmenting intricate structures, a capability crucial in biomedical image segmentation. Its efficiency stems from its ability to use context information (from the contracting path) and precise localization (from the expansive path) to perform segmentation with a high degree of accuracy. The U-Net model is also capable of working with fewer training samples compared to other segmentation models, making it an ideal choice in scenarios where acquiring large datasets is challenging [1].

3.3.2 Adaptation and Application of U-Net

The adaptability of the U-Net model to various imaging conditions makes it a versatile tool for image segmentation. This flexibility is evident in its application across a wide range of fields, from medical imaging to aerial imagery analysis. For the project at hand, the U-Net model was chosen for its ability to accurately segment solar panels from aerial images, a task that requires high precision due to the complexity and variability of the images. The model was adapted to handle the specific characteristics of the aerial images, such as varying lighting conditions and angles, ensuring accurate and reliable segmentation [2].

4 Data Handling and Model Training

4.1 Data Handling and Model Training

Efficient data handling and model training are critical in the development of machine learning models, especially when dealing with multiple datasets that may have varied naming conventions and structures.

4.1.1 Automatic File Name Handling

For this project, automatic file name handling was a crucial step due to the use of two different datasets with distinct naming conventions. To harmonize these datasets for preprocessing, a systematic approach was employed where filenames were programmatically standardized. This process involved mapping the naming conventions of one dataset to another, ensuring consistency in data structure. Such standardization was not only necessary for effective data management but also crucial for seamless integration and processing during the model training phase.

4.1.2 Preprocessing and Data Augmentation

Preprocessing played a pivotal role in preparing the data for model training. Data augmentation was a key technique used in this phase, which involved modifying image qualities and zoom levels to add variety to the training data. This augmentation included transformations like rotation, scaling, and flipping of images, simulating various real-world conditions under which the solar panels might be captured. By artificially expanding the dataset and introducing variability, the model was trained to be robust against changes in input, enhancing its ability to accurately segment solar panels in diverse conditions.

4.1.3 Importance of Data Variety

The inclusion of varied data is essential in training a model that is both accurate and generalizable. Through data augmentation, the range of conditions under which the model could accurately detect solar panels was significantly expanded. This step ensured that the model was not only trained on a wide range of scenarios but also capable of handling new, unseen data effectively in practical applications.

4.1.4 Zoom Level Adjustments

Adjusting zoom levels during data augmentation was particularly important in this project. Given the aerial nature of the imagery, varying zoom levels simulated different altitudes and angles of image capture, which are common in aerial photography. This variability in zoom levels contributed to training a model that was adaptable to different scales of imagery, further enhancing its effectiveness in real-world scenarios.

5 Results

As of our current stage in the project, we have successfully trained and saved our U-Net model for the task of segmenting solar panels from satellite imagery.

After testing the model on the validation images of our datasets, it achieved a precision, recall and F1 score between 92 and 93 percent. The resulting confusion matrix (see Figure 1 below) states how many pixels where correctly or incorrectly predicted. True negatives (TN) are pixels that the model correctly identified as not a solar panel. False positives (FP) on the other hand are pixels that the model predicted to be a solar panel, even though they are not. False negative (FN) are pixels that the model predicted to not be a solar panel, even though they are. True positives (TP) are pixels, which the model correctly identified as solar panels. The sum of TN, FP, FN and TP is the number of all the pixels in the image validation data set.

Confusion matrix

```
    TN           FP
[[39333870  1072083]
 [  922600 12476503]]
    FN           TP
Precision: 0.9208712259714777
Recall: 0.9311446445332945
F1 Score: 0.9259794411312969
```

Figure 1: Test performance visualization.

For the time being such an evaluation of our model on actual images from Vienna are not possible, since we do not have a labeled dataset to compare with the prediction of the model. Therefore, we selected random representative images from Vienna and put them into the model.

The example were the model made the best prediction, is an image, in which the solar panels are on the ground. (see Figure 2) This is due to the fact that the two datasets, which were used to train the model, mostly consisted of such images.

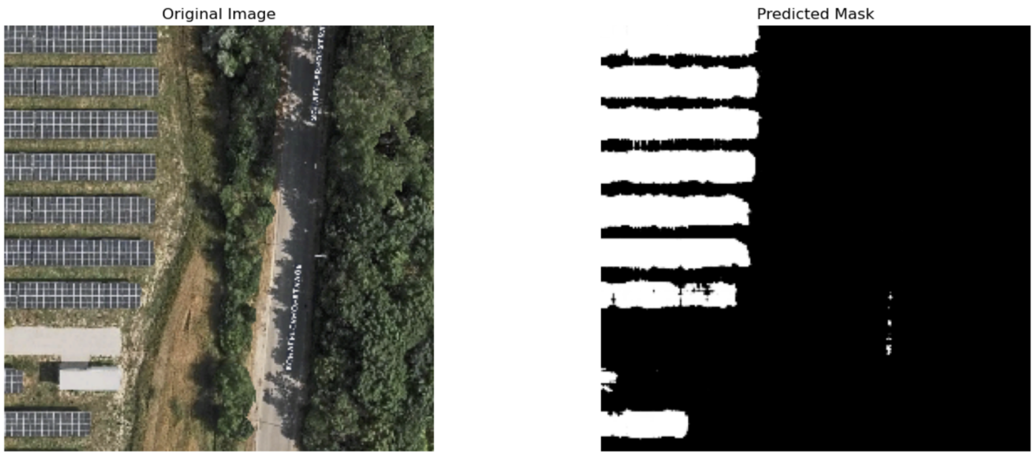

Figure 2: Image prediction: Example Vienna

However, the model also achieves a robust performance on correctly predicting solar panels on roofs, which is pretty common for urban cities like Vienna. (see Figure 3)

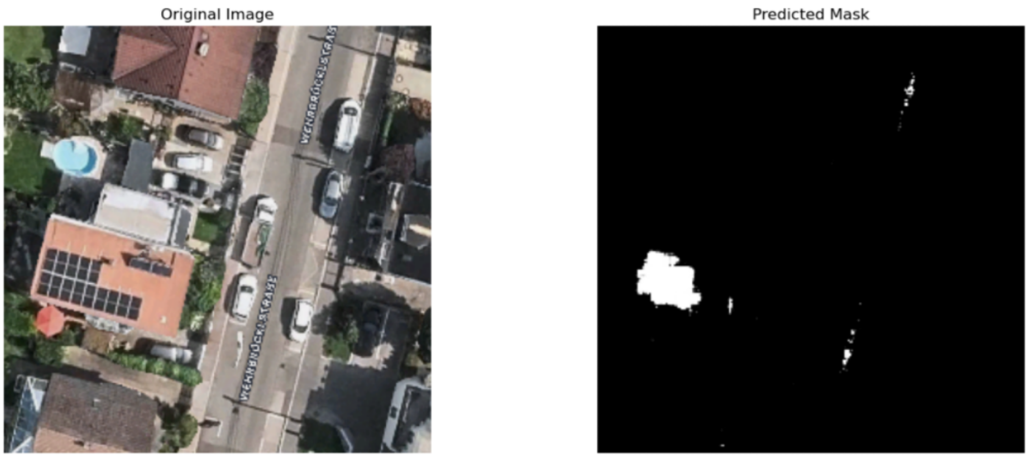

Figure 3: Image prediction: Example Vienna

The correct identification of solar panels on roofs is highly dependent on the quality and zoom level of the image. If these factors are decreasing the model is unable to make correct predictions. Moving forward this can be improved by adding such images with their corresponding labels to the training dataset. (see Figure 4)

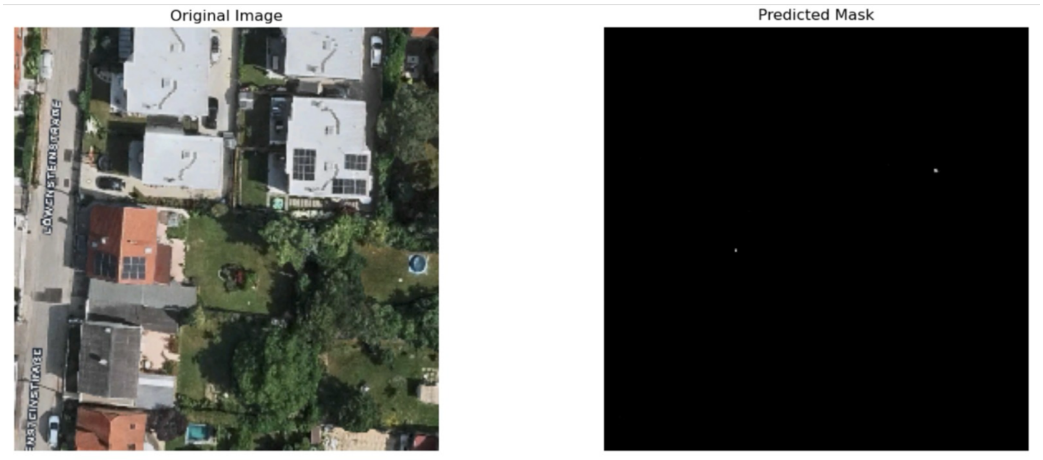

Figure 4: Image prediction: Example Vienna

6 Challenges Encountered

In this project, we encountered a range of challenges, each of which provided valuable insights and learning opportunities. One of the primary challenges was dealing with limited computational resources, particularly GPU capacity. This limitation significantly impacted our ability to process large datasets and train complex models like the U-Net with optimal efficiency. The process of synchronizing and standardizing data from multiple sources with different naming conventions and formats also presented considerable difficulties. This necessitated the development of an automated file name handling system, which, while ultimately successful, required significant time and effort to implement effectively.

Another challenge was the inherent complexity of image segmentation tasks, especially when dealing with high-resolution aerial images where solar panels need to be detected with high precision. Variability in image quality, lighting conditions, and angles further compounded the difficulty of accurately segmenting the images. Data augmentation helped address some of these issues, but the development of robust models capable of generalizing across such varied conditions remained a complex task.

Furthermore, balancing the need for a large, diverse dataset with the practical constraints of manual data labeling was a constant challenge throughout the project. Due to the complexity of the project and the given deadlines, we did not have any resources to perform manual labeling of data for training purposes given how time-consuming and labor-intensive this task is. Therefore, we had to work with publicly available datasets, which are not an ideal solution, since they lacked the variety of images, which contain solar panels on the roof of buildings.

7 Next Steps of Improvement

Looking ahead, several steps could be implemented to further improve the performance and applicability of our solar panel detection model. A primary focus could be on enhancing computational resources. Access to more powerful GPUs and increased computational capacity allows for more efficient data processing and the ability to train more complex models on higher resolution images. This improvement is expected to directly impact the accuracy and efficiency of the model, enabling more nuanced image segmentation and faster processing times.

Another critical area for improvement is the expansion and diversification of the training dataset. Incorporating a wider range of images, particularly those representing different types of urban environments and varying lighting conditions, are crucial. This expansion may improve the model's robustness and accuracy. In this area it might be useful to implement a semi-automated or fully automated data labeling process to significantly reduce the time and effort required for manual labeling, thereby accelerating the dataset expansion process.

Some further steps would be to wrap the code and the whole application up with a focus on enhancing user experience. Navigating through the application and getting a result is at this point really time consuming and not very clear. A improvement would to make it smoother and more intuitive, with simplified menus and clear prompts guiding you. Another part of the GUI could be a feedback option to give instant feedback about occuring problems or bugs.

8 Summary

This project, aimed to develop a sophisticated machine learning model for detecting solar panels from aerial images of Vienna. Employing Convolutional Neural Networks with a focus on the U-Net model, we tackled the significant challenge of accurately identifying solar panels in urban settings. Despite facing challenges such as limited GPU capacity and the complexities of data synchronization and standardization, the project achieved promising results.

The precision of the model in identifying solar panels from aerial imagery was noteworthy, demonstrating the potential of machine learning models in urban infrastructure and energy management. The experience gained from overcoming the various challenges encountered has provided valuable insights, which will be instrumental in guiding the future direction of this project. The next steps include enhancing computational resources, expanding and diversifying the training dataset, and refining data processing techniques. These improvements are anticipated to further elevate the performance and applicability of the model.

To summarise, the entire process of the project, from the first meeting with our data coaches to the final presentation, was very pleasant and smooth. The communication with them was very uncomplicated and they responded fast to our problems and helped us out where needed. Overall the workload of the project was in a realizable scope and the partitioning of it between us was no problem at all.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[2] "Medical image segmentation review: The success of u-net," 2022.

[3] C. Weber, "Aerial imagery object identification dataset for building and road detection, and building height estimation," https://zenodo.org/record/5171712, Aug 2021.

[4] T. Zheng, Z. Chen, and C. Kwan, "Connecticut solar pv semantic segmentation dataset," https://figshare.com/articles/dataset/Connecticut_Solar_PV_Semantic_Segmentation_Dataset/18982199, Apr 2022.

[5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Prentice Hall, 2002.

[6] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[7] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: a review.* ACM computing surveys (CSUR), 1999.

[8] C. M. Bishop, *Pattern recognition and machine learning.* springer, 2006.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT press, 2016.

A Appendix

Under this link the whole dataset and coding environment can be found: https://digitaleconomybasepics.s3.eu-north-1.amazonaws.com/Data+Science+Lab/Data+Science+Lab.zip