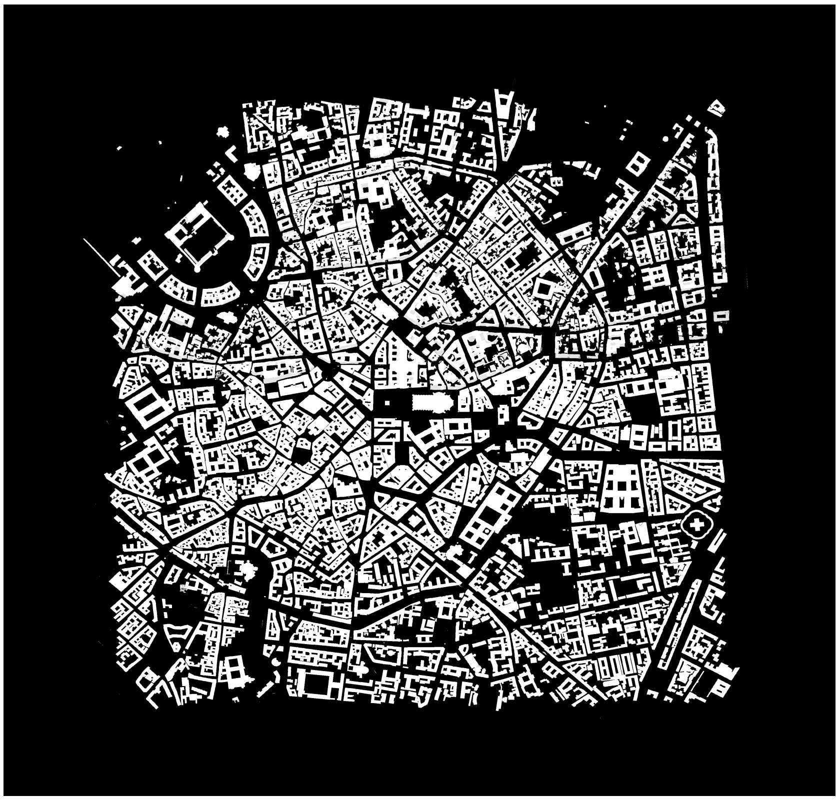
Roof Segmentation for Solar Energy Potential Estimation Using Convolutional Neural Networks With A U-Net Architecture

Brian Bisetti, Andrea Duico, Filip Juren

with BSML | Bocconi Students for Machine Learning

brian.bisetti@studbocconi.it, andrea.duico@studbocconi.it, filip.juren@studbocconi.it

February 23, 2025



1 Introduction

The transition to renewable energy is crucial to address climate change. Solar photovoltaic (PV) panels are an accessible solution, particularly for residential and rural areas. Estimating the potential for solar PV power generation on rooftops is key to optimizing solar energy usage and guiding investments. This report proposes using convolutional neural networks (CNNs) for roof segmentation from satellite images to assess solar energy potential.

Accurate estimates of solar energy generation potential are essential for informed investment decisions. Rooftop PV power generation depends on factors such as roof area, slope, and azimuth. While data like solar irradiation can be obtained from resources such as PVGIS, extracting roof geometry and suitable installation areas remains challenging. Automating rooftop segmentation from satellite imagery is necessary to efficiently identify rooftops suitable for PV installation, promoting informed investments and the adoption of renewable energy.

This work develops a deep-learning model for automatic rooftop segmentation using CNNs. Unlike traditional methods relying on complex feature extraction and machine learning approaches like Support Vector Machines (SVMs), the proposed CNN offers an end-to-end solution for feature learning. This approach builds on previous research to determine the optimal CNN architecture for accurate segmentation, particularly for rural rooftops.

Several studies have explored machine learning for rooftop segmentation and PV potential estimation. Tao Sun et al. (2022) [3] used a CNN for estimating solar PV power distribution on rural rooftops, though details on the model were limited. Eslam Muhammed et al. (2023) [1] applied SVMs with complex feature extraction for rooftop extraction. The most relevant study by Peiran Li et al. (2021) [2] focused on rooftop PV panel segmentation using deep learning, providing insights into optimal architectures and challenges. This work serves as a key reference for data handling and model architecture.

2 Problem Analysis

The data acquisition and processing began by gathering high-resolution satellite imagery for Ljubljana and Milan using the Mapbox API. We defined a bounding box for each area of interest, visualized it using Folium, and used Selenium to take high-resolution screenshots. This process resulted in the satellite images used for subsequent analysis.

Building structures were identified using OpenStreetMap data, and processed with the JOSM application. The GeoJSON file containing building geometries was filtered to retain only relevant rooftops suitable for solar panel installation. However, due to the variability in OpenStreetMap data quality—being an open-source platform—misalignment between the satellite image and building masks was a recurring issue. Manual adjustments were made using Photoshop to correct this, improving the accuracy of alignment, which was crucial for the reliable training of machine learning models.

Selective labeling was also an essential part of this process. We excluded structures unsuitable for solar panels, such as roofs with very steep slopes, like the Duomo in Milan. We also excluded rooftops heavily shadowed by surrounding buildings or dense vegetation, as they were not practical for solar installations. These considerations aimed to ensure that the dataset contained only rooftops with real potential for solar power generation.

The processed imagery resulted in two final maps: a 3000x3000-pixel map for Ljubljana and a 3000x1500-pixel map for Milan. We used a sliding window approach with a step size of 25 pixels, creating approximately 15 GB of 200x200-pixel patches from the satellite images and corresponding masks. This provided the training data for a convolutional neural network (CNN) intended to segment rooftops for photovoltaic potential.

The use of OpenStreetMap was suggested in an article by J. Fitzgerald Weaver (2019), [4] which highlights how combining machine learning with accessible satellite data can effectively map rooftop solar potential. Despite the manual alignment required, these steps ensured the quality of our dataset, enabling more accurate segmentation and better solar potential estimates for urban rooftops.

3 Method

This section presents the methodology used to develop the convolutional neural network (CNN) for rooftop segmentation. Specifically, a U-Net architecture was chosen due to its effectiveness in semantic segmentation tasks, particularly for applications requiring precise localization, like rooftop identification.

The U-Net model utilized in this project follows an encoder-decoder architecture with skip connections that help retain spatial information lost during downsampling. The input to the model is an image of size $200 \times 200 \times 3$.

The encoder consists of multiple convolutional blocks, each involving two convolutional layers followed by batch normalization and ReLU activations. The encoder gradually downsamples the input image using max-pooling layers, extracting increasingly abstract feature representations. The bottleneck layer captures the most abstract features, after which the decoder reconstructs the segmentation mask through transposed convolutions (upsampling), combining feature maps from the encoder via skip connections to ensure precise localization.

The output layer consists of a 1×1 convolution followed by a sigmoid activation function, producing a binary segmentation mask:

$$\hat{y} = \sigma(\text{Conv}_{1\times 1}(f_{\text{decoder}})), \tag{1}$$

where σ is the sigmoid activation function, and f_{decoder} is the output from the last decoder block.

The model was trained using the Dice loss, which is particularly well-suited for binary segmentation tasks that involve class imbalance. The Dice loss is defined as:

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2 \cdot \sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}, \tag{2}$$

where y_i and $\hat{y}_i$ are the ground truth and predicted pixel values, respectively, and ϵ is a small smoothing term to prevent division by zero.

Dice loss was chosen because it directly optimizes for overlap between the predicted and true masks, making it highly effective in handling class imbalance, which is common in rooftop segmentation since the foreground (roof) area is often significantly smaller compared to the background. By maximizing the overlap between predicted and actual segments, Dice loss improves the accuracy of the segmentation mask, especially for small or narrow rooftop regions.

In addition to Dice loss, the Intersection over Union (IoU) metric was used for evaluating model performance:

$$\text{IoU} = \frac{\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i - \sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}, \qquad (3)$$

where ϵ is used to prevent division by zero. IoU provides a useful measure of the accuracy of segmentation by quantifying the overlap between the predicted mask and the true label.

The training, validation, and test sets were split geographically. The entire city of Ljubljana was used as the training set, while Milan was used for both validation and testing. By using different cities for training and evaluation, this approach prevents the model from overfitting to specific local features, ensuring better generalization to new areas with diverse characteristics.

The U-Net model was compiled using the Adam optimizer, with an initial learning rate of 1×10^{-3}. The combination of Dice loss and IoU as metrics helped ensure that the model learned to effectively segment rooftops while focusing on maximizing the overlap between predicted and actual regions. Everything was trained on local CPU.

4 Results

The results obtained from our roof segmentation CNN demonstrate significant discrepancies between the performance on the training set and the validation/test sets, highlighting critical areas for improvement. While the training metrics were notably high, with accuracy reaching around 95% and IoU approximately 85%, the performance on the validation set was considerably worse. Validation accuracy remained below 80%, and IoU hovered around 50%, indicating a lack of generalization. A similar trend was observed in the test set, further underscoring the model's limitations in handling unseen data.

This performance gap suggests that the model overfits the training data, capturing patterns specific to plain roofs, which constituted the majority of the training set. However, it failed to generalize to more complex roof structures, such as sloped roofs, which are common in real-world scenarios. This limitation greatly impacts the practical utility of the model, as its focus on plain roofs renders it ineffective for broader applications. Addressing this overfitting and improving the model's capability to segment diverse roof types is critical for future iterations.

One of the most significant challenges observed in the results is the model's bias toward detecting plain roofs while failing to adequately segment sloped roofs. This limitation severely impacts the model's utility, as sloped roofs are prevalent in many urban and rural settings. Literature, such as the work by Peiran Li et al. (2021), discusses similar challenges, emphasizing the influence of roof orientation on segmentation performance. Different roof inclinations and orientations, particularly those influenced by latitude and satellite angle, require specialized handling that our current model does not address adequately.

Moreover, our dataset's limitations contributed to this issue. Training exclusively on images from a specific region likely resulted in the model's inability to generalize to diverse roof types and orientations. Expanding the dataset to include masks from multiple European cities could significantly enhance the model's robustness and adaptability.

As noted in prior research, segmentation models for roof detection often struggle with generalization due to regional variations in architectural styles and roof orientations. Developing separate models for different latitudes, or incorporating region-specific augmentations, could mitigate this challenge. However, such an approach requires substantial computational resources and diverse data, which were unavailable during this project.

Another critical issue was the computational environment. The training was conducted on a CPU due to technical constraints related to data management on Kaggle, which severely limited the training process. Training a deep learning model like a

CNN on a CPU is notably inefficient, resulting in extended training times and potentially suboptimal convergence. Due to time limitations, the model was trained for fewer epochs than planned, which might have hindered its ability to achieve better performance.

5 Experiment

While training the model, we developed a module that can be used with the trained models to facilitate practical applications. By providing latitude and longitude coordinates, the module can generate a segmentation map of rooftops and estimate the potential placement of solar panels. Using tools such as PVGIS, users can calculate the solar power output based on the panel placement identified by the segmentation model. This functionality allows for an estimation of the potential power generation for a given architectural structure, providing valuable insights for renewable energy planning.

6 Conclusions

This study focused on developing a CNN-based model for rooftop segmentation with potential applications in solar panel placement and energy estimation. While the model demonstrated promising performance on the training data, the results highlighted critical challenges, particularly in generalizing to unseen data and accurately segmenting diverse roof types.

A key area for improvement is the masking process. For more accurate segmentation and a better understanding of roof orientations, we propose enhancing the dataset with masks that include not only rooftop boundaries but also information about azimuthal orientation. One potential method is to assign different colors to roofs based on their orientation. This approach could help distinguish between roof slopes and improve the accuracy of solar panel placement estimation.

Moreover, to ensure the model performs well in diverse environments, it is essential to expand the dataset to include rooftops from various architectural styles and geographic locations. Incorporating urban and rural areas will make the model more robust and widely applicable. Employing data augmentation techniques, such as rotating images to simulate different azimuths, can further enhance the model's capability to handle diverse scenarios.

7 Appendix

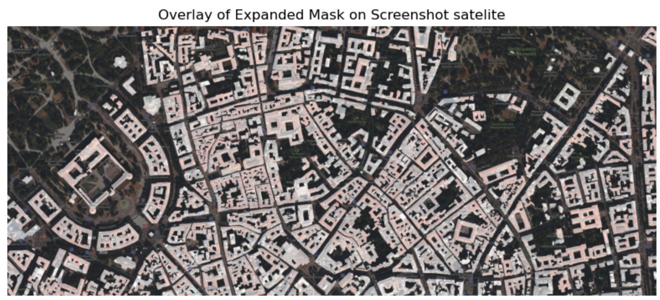


Figure 1: This is the labeled city of Milan

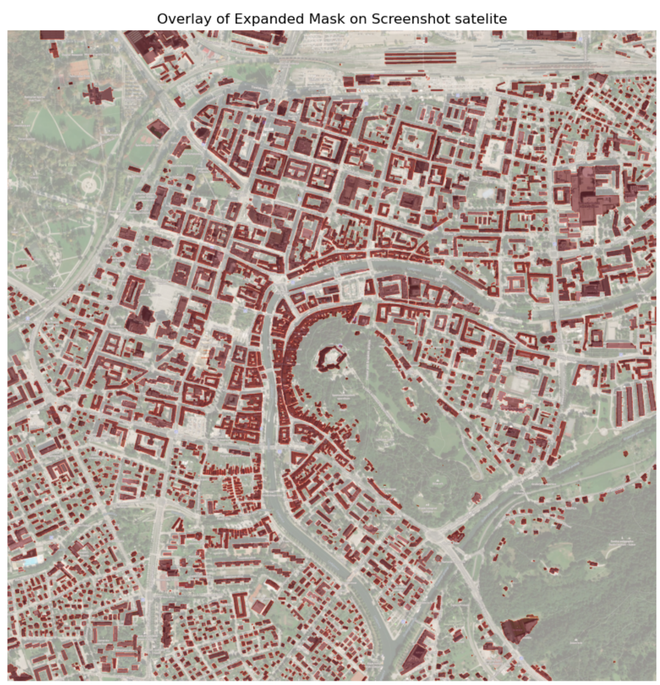


Figure 2: This is the labeled city of Ljubljana

Figure 3: 200x200 pixel mask example for the taining data

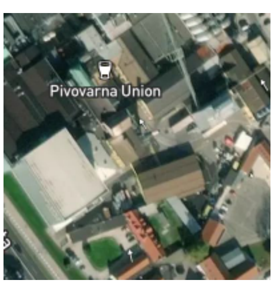


Figure 4: 200x200 pixel satelite example for the taining data

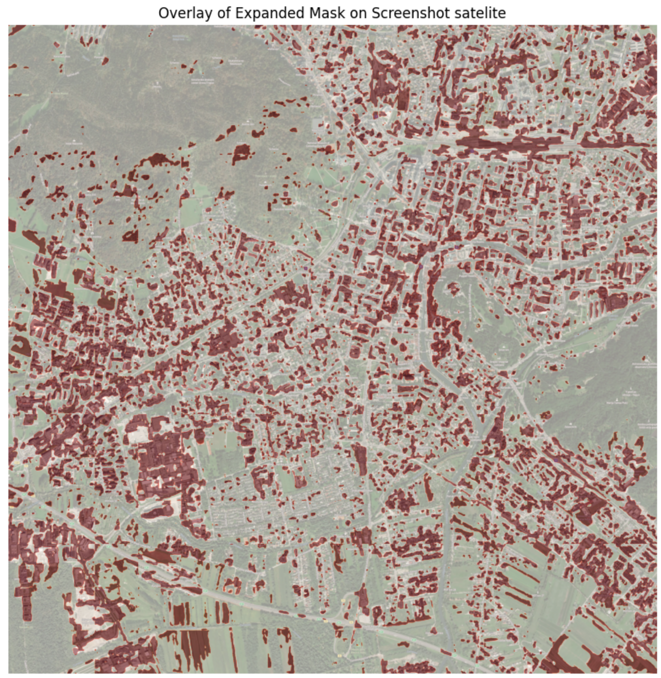


Figure 5: Example of output from our model

References

[1] S. M. Eslam Muhammed, Adel El-Shazly. Building rooftop extraction using machine learning algorithms for solar photovoltaic potential estimation. *mdpi*, 2023.

[2] Z. G. S. L. J. C. W. L. X. S. R. S. J. Y. Peiran Li, Haoran Zhang. Understanding rooftop pv panel semantic segmentation of satellite and aerial images for better using machine learning. *sciencedirect*, 2021.

[3] X. R. X. Y. Tao Sun, Ming Shan. Estimating the spatial distribution of solar photovoltaic power generation potential on different types of rural rooftops using a deep learning network applied to satellite images. *sciencedirect*, 2022.

[4] J. F. Weaver. Using machine learning and cheap satellite data to design rooftop solar power. *pv magazine*, 2019.