Convolutional Neural Network-Based Roof Classification and Segmentation for Solar Potential Estimation

Dylan Nguyen
Valencia High School
Placentia, CA

Abstract

This study presents a convolutional neural network-based pipeline for automated roof classification and segmentation to estimate residential solar potential using satellite imagery from Google Earth, BRAILS, and Kaggle. A sequential convolutional neural network (CNN) accurately classified roof types (flat, gabled, hipped) at 90.63%, and an improved U-Net segmented rooftop areas with 89.03% accuracy. Segmentation performance declined for white-colored roofs, emphasizing the need for dataset diversification. Future work includes integrating high-resolution imagery, 3D structural analysis, and shading factors to enhance solar potential assessments.

Keywords: Convolutional Neural Network, Roof classification, Image segmentation, Solar potential, Satellite imagery, U-Net, Deep learning

1. Introduction

1.1 Background

Accurate estimation of residential rooftop solar potential is crucial for determining the viability of solar panel installations. Traditional assessment methods, such as manual inspections and large-scale urban modeling, can be expensive, time-consuming, or insufficiently detailed for individual homeowners. Recent research primarily focuses on city-wide or district-level evaluations using GIS and CNN approaches (Li et al., 2024; Ni et al., 2024), while relatively fewer studies address individual roof assessments. This research gap highlights the need for precise, accessible, and cost-effective methodologies tailored to individual consumers. Convolutional neural networks are particularly effective for rooftop analysis due to their ability to extract spatial hierarchies and local features from imagery.

1.2 Literature Review

Before the widespread adoption of CNNs, solar potential estimation relied heavily on manual assessments or rule-based image analysis methods. Traditional approaches often required expert interpretation of satellite images or on-site inspections to measure roof area, tilt, and shading, making them time-consuming and inconsistent across different

evaluators. Other non-CNN methods included thresholding techniques and classical machine learning models like support vector machines (SVMs) and decision trees, which depended on handcrafted features and lacked spatial awareness. These techniques were generally less accurate and struggled with generalizing across diverse roof geometries and lighting conditions.

Convolutional Neural Networks (CNNs) have emerged as powerful tools in estimating rooftop solar potential by automating the extraction of relevant spatial features from aerial and satellite imagery. Li et al. (2023) introduced SolarNet, a multi-task CNN framework designed to jointly estimate rooftop geometry, orientation, and usable area from high-resolution aerial images. This model outperformed traditional segmentation methods by incorporating geometry-aware learning and leveraging domain-specific datasets.

Li et al. (2024) built upon this foundation by integrating rooftop superstructures—such as chimneys and HVAC units—into its segmentation pipeline. By accounting for obstructions, the framework more accurately identified usable space for photovoltaic (PV) installations. This refinement enhanced solar suitability assessments at the individual building level.

Kurte & Kulkarni (2025) presented a complementary CNN-based approach to detect existing PV panels. Their model employed a VLAD (Vectors of Locally Aggregated Descriptors) encoding scheme to enhance the discriminative power of localized features, enabling accurate rooftop classification across varied urban settings.

These methods often incorporate preprocessing techniques like azimuth-based alignment and spatial normalization to improve generalization across diverse regions (Lin et al., 2024). Architectures typically use binary cross-entropy for segmentation tasks and categorical cross-entropy for classification, sometimes enhanced with geometry-based auxiliary losses. Multi-task learning has also proven effective in boosting shared feature representations.

Despite their strengths, CNN models still face challenges in segmenting rooftops with highly reflective surfaces, occlusions, or irregular geometries (Fu et al., 2022). These issues can lead to decreased accuracy in real-world urban scenarios and require improved generalization strategies.

Collectively, these studies showcase the growing precision and adaptability of CNN-based methods in rooftop-level solar potential assessment. This paper builds on these frameworks with a two-stage CNN pipeline that performs both roof classification and segmentation, tailored for residential applications using freely accessible satellite imagery.

1.3 Data Description

This study utilized three distinct datasets to train and evaluate the CNN models for rooftop classification and segmentation. Each dataset contributed a unique component to the two-stage pipeline: roof type classification and pixel-wise segmentation of usable rooftop areas.

The first dataset was the BRAILS Roof Dataset (SimCenter, 2023), which includes aerial images of residential buildings labeled by roof type—specifically flat, gabled, and hipped. This dataset was primarily used to train the classification model and contains diverse roof structures from various regions across the United States, improving the model's generalizability. Example images from this dataset are shown in Figure 1. For training, the classification dataset was split into 80% training and 20% validation. Images were resized to 180×180 pixels, normalized to the [0,1] range, and converted to RGB format. To improve generalization, data augmentation techniques such as random flips and rotations were applied.

Figure 1: Sample images from the BRAILS Roof Dataset. Left: flat. Right: gabled.

The second dataset was the Kaggle Rooftop Images for Semantic Segmentation dataset (Weeb, 2023). It includes paired satellite images and corresponding binary segmentation masks, where each mask distinguishes rooftop pixels from the background. These annotations enabled supervised learning for semantic segmentation of rooftop areas. A sample image-mask pair is shown in Figure 2. The segmentation masks were preprocessed by converting them to grayscale, resizing with nearest-neighbor interpolation to maintain label fidelity, and reshaping to include a single-channel format suitable for CNN input. Images were resized to 128×128 pixels, normalized to the [0,1] range, and converted to RGB. During training, the data were loaded into NumPy arrays and processed in batches of 16.

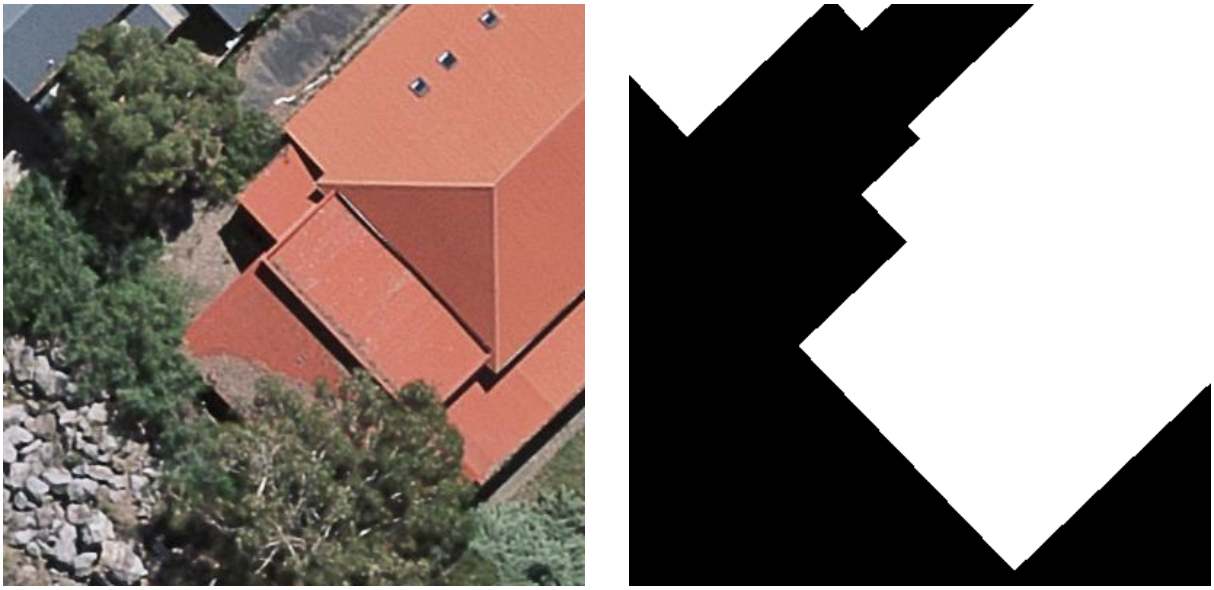

Figure 2: Sample image-mask pair from the Kaggle Rooftop Segmentation Dataset. Left: original rooftop image. Right: binary mask (roof pixel contour).

To evaluate model performance on real-world data and perform post-segmentation area calculations, a third testing dataset was constructed by manually collecting rooftops from Google Earth (Google Earth, 2023). Each image was captured at a fixed vantage point of 225 meters above ground, oriented directly downward, and constrained to a square with a 35-meter diagonal using the built-in measurement tool. This consistent capture configuration allowed derivation of a scale factor from pixels to real-world units. The area

of each pixel in square meters was calculated using the formula $A = \left(\frac{35}{\sqrt{2}\cdot 128}\right)^2$, which for a 128×128 image yields approximately $0.0745\,\mathrm{m}^2$ per pixel.

This conversion enabled accurate estimation of usable rooftop area from segmentation masks. Most of the sampled rooftops for this dataset were located in East Lake Village, a residential community in Yorba Linda, California, as shown in Figure 3.

Figure 3: Sample rooftop image from East Lake Village in Yorba Linda, California, collected via Google Earth.

To standardize training inputs, images for the classification model were resized to 180×180 pixels, while images for the segmentation model were resized to 128×128 pixels. All images were normalized to the [0,1] range and converted to RGB format.

Lastly, to support solar potential estimation, solar irradiance data were obtained from the Global Solar Atlas (Group, 2024). The metric used was the Global Tilted Irradiation at Optimum Angle ($\mathrm{GTI}_{\mathrm{opta}}$), which is most relevant for estimating photovoltaic output. The value for Yorba Linda was reported as 2243.1 $\mathrm{kWh/m^2}$ per year, which was converted to average watts per square meter ($\mathrm{W/m^2}$) by dividing by the number of hours in a year, yielding 256 $\mathrm{W/m^2}$. This irradiance value, along with the area derived from segmentation masks, was used to estimate annual solar energy output per rooftop.

1.4 Article Outline

The article begins with an overview of rooftop solar potential estimation and its significance for residential energy planning. It then introduces the CNN-based two-stage pipeline, covering both roof classification and rooftop area segmentation.

The Methods section details the theory and implementation of the CNN classifier and U-Net segmentation model. It also explains the methodology used to calculate solar potential by combining segmentation outputs with irradiance and system efficiency data.

The Results and Discussion section presents model performance metrics, visual examples of predictions, and commentary on the pipeline's strengths and limitations. The article concludes with a brief discussion of future work and potential improvements.

The full implementation of this pipeline is available on the following GitHub repository:

https://github.com/dylann4500/cnn-solar-potential

2. Methods

2.1 CNN for Roof Classification

2.1.1 Theoretical Foundation

A central operation in CNNs is the discrete 2D convolution, which systematically applies learnable filters to input feature maps. For layer l, let $\mathbf{X}^{(l-1)} \in \mathbb{R}^{H\times W\times C}$ denote the input activation map of height H, width W, and C channels. A convolutional filter $\mathbf{W}_k^{(l)} \in \mathbb{R}^{r\times r\times C}$ (with kernel size $r \times r$) and bias $b_k^{(l)}$ produce a single output channel k in the feature map $\mathbf{Z}^{(l)}$. The operation at spatial location (i, j) is:

$$Z_k^{(l)}(i,j) = \sum_{u=1}^{r}\sum_{v=1}^{r}\sum_{c=1}^{C} W_k^{(l)}(u,v,c)\, X^{(l-1)}(i+u-1, j+v-1, c) + b_k^{(l)}.$$

The network learns the filter weights $\mathbf{W}$ and bias b to capture spatial patterns in the data (e.g., edges or roof features).

After each convolution, a non-linear activation function is typically applied element-wise. In our models, the Rectified Linear Unit (ReLU) is used:

$$\mathrm{ReLU}(x) = \max(0,\, x).$$

ReLU accelerates convergence by avoiding saturation issues while maintaining sparse gradients.

Max-pooling reduces the spatial resolution, providing a degree of translational invariance and reducing the parameter count. A 2×2 max-pooling over the output $\mathbf{Z}^{(l)}$ can be expressed as

$$P^{(l)}(i,j) = \max_{0\le u,v<2} Z^{(l)}(2i+u,\; 2j+v).$$

This downsampling helps the CNN capture hierarchical features while reducing computation.

After successive convolution and pooling layers, features are flattened into a vector $\mathbf{h}$ and fed to a fully connected layer. If $\mathbf{W}^{(\mathrm{fc})}$ and $\mathbf{b}^{(\mathrm{fc})}$ denote the weights and biases of the dense layer, the output is

$$\mathbf{z}^{(\mathrm{fc})} = \mathbf{W}^{(\mathrm{fc})}\,\mathbf{h} + \mathbf{b}^{(\mathrm{fc})}.$$

For a classification task with K classes, the softmax activation in the final layer produces a probability vector $\hat{\mathbf{y}} \in \mathbb{R}^K$:

$$\hat{y}_k = \frac{e^{z_k^{(\mathrm{fc})}}}{\sum_{k'=1}^{K} e^{z_{k'}^{(\mathrm{fc})}}}, \quad k = 1, 2, \ldots, K.$$

We employ the sparse categorical cross-entropy loss. Given a ground-truth class label y (an integer $1 \le y \le K$) and the predicted probability distribution $\hat{\mathbf{y}}$, the loss for a single example is

$$\mathcal{L}_{\text{SC-CE}}(y, \hat{\mathbf{y}}) = -\ln(\hat{y}_y).$$

For a batch of N samples, the total loss is the mean of individual losses:

$$\mathcal{L}_{\text{total}} = \frac{1}{N} \sum_{n=1}^{N} -\ln(\hat{y}^{(n)}_{y^{(n)}}).$$

Both classification and segmentation models are trained via backpropagation and stochastic gradient descent. The Adam optimizer updates parameters using running estimates of gradients and second moments. Denoting the parameters by θ and gradients by g_t, Adam's updates are:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2,$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t},$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon},$$

where α is the learning rate, β_1 and β_2 are exponential decay rates for the moment estimates, and ϵ is a small constant.

2.1.2 Application

A sequential CNN was developed using TensorFlow's Keras API (Abadi et al., 2015) to classify roof types as flat, gabled, or hipped. The model included three convolutional layers with ReLU activation and max-pooling, followed by a dense layer with 128 neurons and a softmax output layer for three-class prediction.

Trained with the Adam optimizer and sparse categorical cross-entropy loss, the model processed preprocessed image tiles to assign roof type labels. This classification step enabled solar potential estimates to be tailored to roof geometry.

2.2 U-Net-Based Architecture for Roof Segmentation

2.2.1 Theoretical Foundation

The U-Net architecture is widely used for semantic segmentation due to its ability to perform precise, pixel-level classification. It consists of two main paths: a contracting encoder that captures semantic context and an expanding decoder that restores spatial resolution. Each encoder block applies convolutions followed by downsampling, while each decoder block performs upsampling and concatenation with corresponding encoder features through skip connections.

Let $\mathbf{C}_l$ represent the output of the l-th encoder block. After the bottleneck, the decoder progressively upsamples the feature maps using transposed convolutions. These are concatenated with the corresponding $\mathbf{C}_l$ from the encoder to form $\mathbf{D}'_{L-l}$, where L is the total number of encoding layers:

$$\mathbf{D}'_{L-l} = \text{Concat}\left(\text{Conv2DTranspose}(\mathbf{D}_{L-l+1}),\ \mathbf{C}_l\right).$$

This fusion of high-level and low-level features preserves fine spatial details, which is critical for accurate segmentation.

The bottleneck layer, situated at the bottom of the U-shape, typically employs deeper convolutional layers and dropout regularization to prevent overfitting. For a given activation h_i, dropout randomly zeroes out activations with probability p, and scales the

retained values:

$$h_i' = \begin{cases} 0, & \text{with probability } p, \\ \frac{h_i}{1-p}, & \text{otherwise.} \end{cases}$$

This forces the network to learn redundant, distributed representations, which improves generalization on unseen data.

The final decoder output is a single-channel map of the same spatial size as the input, with a sigmoid activation applied to produce a probability score $\hat{y}(i,j)$ for each pixel (i,j):

$$\hat{y}(i,j) = \sigma(z(i,j)), \quad \sigma(x) = \frac{1}{1+e^{-x}}.$$

The resulting output indicates the likelihood that each pixel belongs to the foreground (i.e., a roof area).

To train the model, binary cross-entropy (BCE) loss is computed for each pixel. Let $y(i,j) \in \{0,1\}$ denote the ground-truth label and $\hat{y}(i,j)$ the predicted probability. The BCE loss at a pixel is:

$$\mathcal{L}_{\text{BCE}}\left(y(i,j), \hat{y}(i,j)\right) = -\left[y(i,j)\ln\hat{y}(i,j) + (1-y(i,j))\ln(1-\hat{y}(i,j))\right].$$

The total loss is computed as the mean over all pixel locations in the training batch.

As with the classification model, optimization is performed using the Adam algorithm. Parameters are updated using estimates of the first and second moments of the gradients to achieve faster convergence and robustness to sparse gradients.

2.2.2 Application

An improved U-Net architecture was implemented to segment rooftop areas from overhead imagery. The encoder consisted of four convolutional blocks with ReLU activation, batch normalization, and max-pooling, while the decoder used transposed convolutions and skip connections to recover spatial detail. A bottleneck layer with dropout reduced overfitting, and the final sigmoid-activated output produced a binary mask of rooftop pixels.

The model was trained using binary cross-entropy loss and optimized with Adam. These predicted masks enabled precise rooftop area estimation, forming the foundation for subsequent solar potential calculations.

2.3 Solar Potential Calculation

To estimate solar potential, the usable rooftop area identified by the segmentation model was combined with regional solar irradiance and system performance parameters. The number of roof pixels exceeding a threshold of 0.5 in the predicted mask was summed and multiplied by a scale factor derived from the fixed 35-meter diagonal of each image tile, yielding area in square meters. This threshold was chosen to binarize the sigmoid output, with 0.5 representing the midpoint probability between roof and non-roof classifications.

Because not all rooftop space is usable for photovoltaic (PV) installation due to structural constraints, pitch, and obstructions, a correction coefficient was applied based on roof type. Based on ranges discussed by Jakubiec & Reinhart (2013) and Karteris et al. (2019), we assigned average usable area coefficients of 0.9 for flat roofs, 0.75 for gabled roofs, and 0.6 for hipped roofs. These values reflect reductions from features such as dormers, ridgelines, and fragmented surfaces.

Solar potential P in watts was calculated using the equation:

$$P = A \times G \times \eta_p \times \eta_s$$

where A is the adjusted usable roof area (m^2), G is the solar irradiance in W/m^2, η_p is the panel efficiency, and η_s is the system efficiency. For this study, $G = 256\,W/m^2$, based on the Global Tilted Irradiation at Optimum Angle for Yorba Linda. A panel efficiency of 20% was used, representative of modern silicon-based photovoltaic panels (Karteris et al., 2019). System efficiency, which accounts for inverter losses, temperature effects, wiring, and dust accumulation, was set at 85% (Jakubiec and Reinhart, 2013).

This calculation provided a simplified but practical estimate of peak solar power potential for each rooftop, enabling comparative assessments across diverse residential structures.

3. Results and Discussion

The classification model achieved an accuracy of **90.625%**, effectively distinguishing between flat, gabled, and hipped roofs. The segmentation model obtained an accuracy of **89.03%**, demonstrating strong performance in delineating rooftop areas from aerial images.

Figure 4: Left: original aerial image. Right: predicted rooftop mask. Estimated solar potential of 10,297 watts.

Figure 4 shows a sample output from the full pipeline. The model correctly classified the roof as gabled and segmented a rooftop area of $315.37\,m^2$. After applying the gabled roof coefficient of 0.75, the usable area was estimated at $236.53\,m^2$, resulting in a solar potential of approximately 10,297 watts.

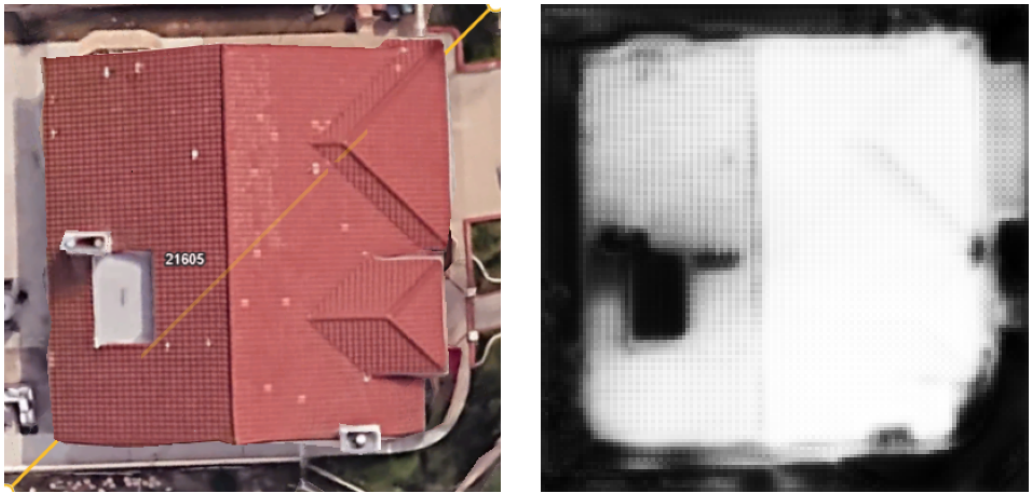

Figure 5: Left: original aerial image. Right: predicted rooftop mask. Estimated solar potential of 13,347 watts.

Figure 5 presents another gabled rooftop example. The raw segmented area was 408.79 m^2, which adjusted to 306.59 m^2 usable area after applying the coefficient. This yielded an estimated solar potential of 13,347 watts.

These values fall within the expected range for residential photovoltaic systems in Southern California and support the effectiveness of the pipeline. Although on-site photovoltaic generation data was unavailable for verification, the outputs are consistent with common system capacities.

Segmentation issues were occasionally observed on white or highly reflective roofs, which led to underestimation of usable area. However, for darker colored roofs, this indirectly allowed the model to exclude superstructures in its calculation, thus yielding more accurate solar potential calculations.

4. Potential Future Research

Future work should first address data diversity. The present models were trained primarily on conventional asphalt or tile roofs; highly reflective or white membranes, less common geometries such as mansards or shed roofs, and images captured under different illumination conditions remain under-represented. Enlarging the training corpus and exploring unsupervised or few-shot domain-adaptation techniques would help the pipeline generalise to cities that differ markedly from the original domain.

A second avenue lies in incorporating richer three-dimensional information. Height maps derived from stereo photogrammetry or LiDAR would allow direct estimation of roof tilt, azimuth, and self-shadowing. Coupling these data with physics-based or learned irradiance models would convert the present peak-power figures into seasonally resolved energy-yield forecasts and enable panel-layout optimisation that respects pitch, setback, and structural constraints.

Finally, the methodology would benefit from ground-truth validation and a broader techno-economic perspective. Comparing predicted solar potential with measured production from installed photovoltaic systems will quantify real-world accuracy, while integrating cost, pay-back, and carbon-offset calculations can turn the tool into a complete decision-support platform for homeowners, installers, and municipal planners.

Acknowledgments

I would like to sincerely thank Dr. Olga Korosteleva at California State University, Long Beach for her invaluable guidance in identifying practical applications of convolutional neural networks and for helping me build a stronger, mathematical foundation in deep learning. Her mentorship played a significant role in shaping the direction and depth of this project.

I am also grateful to my high-school mentors at Valencia High School: Ms. Wendy Umekubo, whose passion for calculus revitalised my own and deepened my appreciation of the mathematics underlying this study; and Mr. James Womack, whose long-standing support of our mathematics club created the environment and opportunities that made projects such as this possible.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org].

Fu, Y., Wang, Y., & Liu, J. (2022). A unified deep learning framework for robust rooftop extraction under complex urban scenes. *Remote Sensing*, 14(3), 547. https://doi.org/10.3390/rs14030547

Google Earth. (2023). Google Earth Pro. https://earth.google.com

Group, W. B. (2024). Global solar atlas. https://globalsolaratlas.info/

Jakubiec, J. A., & Reinhart, C. F. (2013). A method for predicting city-wide electric production from photovoltaic panels based on lidar and gis data combined with hourly daysim simulations. *Solar Energy*, 93, 127–143. https://doi.org/10.1016/j.solener.2013.03.028

Karteris, M., Theodoridou, I., & Mallinis, G. (2019). Assessing the benefits of large-scale green roofs implementation in mediterranean urban environments: The case of thessaloniki, greece. *Renewable and Sustainable Energy Reviews*, 104, 420–432. https://doi.org/10.1016/j.rser.2019.01.031

Kurte, K., & Kulkarni, K. (2025). Enhanced rooftop solar panel detection by efficiently aggregating local features. *arXiv preprint arXiv:2501.02840*. https://arxiv.org/abs/2501.02840

Li, Y., Zhang, Y., & Chen, X. (2023). Solarnet: A convolutional neural network-based framework for rooftop solar potential estimation from aerial imagery. *Expert Systems with Applications*, 186, 115764. https://doi.org/10.1016/j.eswa.2021.115764

Li, Y., Zhang, Y., & Chen, X. (2024). Solarnet+: A multi-task cnn for rooftop solar potential estimation. *Applied Energy*, 306, 117956. https://doi.org/10.1016/j.apenergy.2021.117956

Lin, Z., Wang, S., & Gao, R. (2024). Hybrid modeling of direct normal irradiance for rooftop photovoltaic applications using ground-based cloud images. *Solar Energy*, 220, 45–56. https://doi.org/10.1016/j.solener.2021.04.029

Ni, J., Wang, Q., & Liu, H. (2024). A robust deep learning framework for solar potential estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1-2024, 371–378. https://doi.org/10.5194/isprs-annals-XLVIII-1-2024-371-2024

SimCenter. (2023). Brails++: Building recognition using artificial intelligence at large-scale. https://simcenter.designsafe-ci.org/products/backend-components/brails/

Weeb, S. (2023). Rooftop images for semantic segmentation. https://www.kaggle.com/datasets/slyveinweeb/rooftop-images-for-semantic-segmentation