Towards Sustainable Urban Energy: A Robust Deep Learning Framework for Solar Potential Estimation

Weiyan Lin[1], Jiasong Zhu[2], Yuansheng Hua[2], Qingyu Li[3], Lichao Mou[3], Xiao Xiang Zhu[3]

[1] School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China - linweiyan2022@email.szu.edu.cn
[2] School of Civil and Traffic Engineering, Shenzhen University, Shenzhen, China - (zjsong, yuansheng.hua)@szu.edu.cn
[3] Data Science in Earth Observation, TUM, Munich, Germany – (qingyu.li, lichao.mou, xiaoxiang.zhu)@tum.de

Abstract

Rooftop photovoltaic is considered as a cost-effective and environmentally friendly solution to energy challenges in urban areas. To ensure photovoltaic efficiency, it is essential to accurately estimate rooftop solar potential and deploy solar panels wisely. During the past few years, deep learning-based estimation methods have emerged and mainly rely on inferring rooftop orientations from aerial imagery. However, we note that rooftops often appear diversely when images are taken at different solar azimuths, and this can lead to orientation misclassification. To address this, we propose a robust solar potential estimation framework, mainly composed of a rooftop orientation prediction network and a bilateral solar potential estimation module. Specifically, we first classify rooftops into five orientations, i.e., east, west, south, north towards, and flat with a semantic segmentation network. Afterward, opposing orientations are merged to alleviate misclassification caused by variant data acquisition time. Eventually, we compute solar potentials based on PVGIS and a weighting scheme. Experimental results on the RID dataset demonstrate the effectiveness of our approach in improving the accuracy of solar energy estimation.

1. Introduction

The integration of renewable energy technologies is essential for cities to move towards a low-carbon future(Liu and Lv, 2019). In urban areas, rooftop photovoltaic systems are a viable solution(Gassar and Cha, 2021) to address the growing energy demands and environmental concerns. Moreover, rooftop photovoltaic systems can efficiently use the previously unused space in urban areas, which helps combat land scarcity issues. Additionally, decentralized rooftop photovoltaic systems can potentially reduce costs associated with long-distance power transmission and electricity consumption. Identifying the spatial distribution of rooftop solar potential is crucial to optimizing the strategic placement of photovoltaic systems within urban settings. This information can also aid the development of policies about renewable energy.

Determining the geographic potential of each rooftop is crucial for evaluating the feasibility of rooftop photovoltaic (PV) systems, involving accurately calculating the total solar radiation that each rooftop can receive. However, this task poses a unique challenge with distributed rooftop PV systems due to their scattered deployment. The most significant source of uncertainty in this assessment is the precise depiction and calculation of the building rooftops(Zhang et al., 2023). Hence, the rooftop area is a critical parameter in the evaluation process. Four methods are used for these assessments, including sampling statistical-based, geographic information-based, 3D model-based, and satellite imagery-based methods, depending on different data types.

The sampling statistics approach usually calculate one or more relevant variables related to the sample area and then using appropriate strategies to determine the overall available roof area for the entire region. Wiginton et al. use census subdivisions (CSDs) as research units to explore the correlation between population size and rooftop area. A subset of CSDs
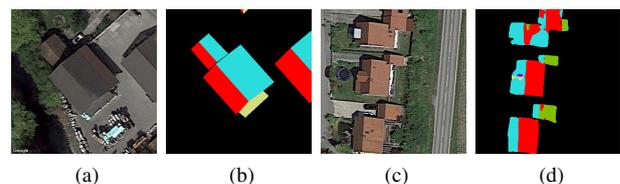
Figure 1. (a) and (c) Google satellite images. (b) Ground truth for the roof. (d) Misclassification examples, where a west-facing roof is incorrectly labeled as east-facing, and vice versa.

is sampled to calculate the per capita roof area, which is then used to estimate the total roof area based on population estimates(Wiginton et al., 2010). Similarly, Byrne et al. estimate the total number of floors by calculating the average area of each floor using statistics from the Korea Statistical Information Service (KOSIS). Utilizing this data, they determine the total rooftop area of Seoul, crucial for assessing solar potential(Byrne et al., 2015). Wang et al. utilize urban developed land and residential land area data from the *China Urban Statistical Yearbook 2019* to determine available rooftop area for installing distributed photovoltaic systems(Wang et al., 2021).

Although employing readily accessible national statistical data for indirectly determining roof area can be efficacious, the accuracy of this approach is constrained by data quality limitations. National statistical data are typically aggregated at the provincial level or higher, leading to potentially significant errors in estimating results at the city level. Therefore, more detailed geographic spatial data are imperative for estimating rooftop photovoltaic potential. Relying on professional software, geographic information-based methods can directly use existing vectorized cadastral data in various cities to calculate roof area. For example, government-provided building data(Walch et al., 2020, Wong et al., 2016) or vectorized roof

Figure 2. The variation of roof color when photographed at different solar azimuth angles. The first row is the Google satellite images, and the second row depicts the corresponding label for roof segments. (a) morning, (b) noon and (c) afternoon.

area datasets created from imagery(Zhang et al., 2022). Platforms like OpenStreetMap offer global building vectors(Buffat et al., 2018, Ni et al., 2024) and provide information on building types(Pan et al., 2022), making them valuable datasets for researchers. However, these datasets may become outdated or incomplete due to unpredictable update cycles. Therefore, researchers should integrate multiple data sources to ensure comprehensive coverage of building data across the entire study area(Buffat et al., 2018).

In recent years, researchers have explored the use of Unmanned Aerial Vehicles (UAVs) equipped with airborne Light Detection and Ranging (LiDAR) technology to create detailed 3D models(Lukač et al., 2014). These methods allow for the extraction of the 3D structure of rooftops and facilitate the analysis of environmental impacts on solar potential. For example, SPAN(Özdemir et al., 2023) is an open-source plugin designed for estimating photovoltaic potential. Users can upload 3D building data in standard formats and access detailed information on rooftop photovoltaic estimation, including surface areas, azimuth, tilt angles, daily global irradiation, and total photovoltaic output. Increasing the density of input point cloud data typically enhances the accuracy of the final results. To reduce data acquisition costs, some studies utilize open-source 3D models(Buffat et al., 2018, Zhu et al., 2020, Lan et al., 2022) for analysis of solar potential, leveraging resources like the 3D Photo-realistic Model available for Hong Kong(Ren et al., 2022, Ren et al., 2023). Similarly, Wong et al. utilize the DSM with a spatial resolution of 0.5 across Hong Kong, identifying rooftop pixels by excluding ground, obstacles, shadows, and steep slope pixels(Wong et al., 2016). LOD2-level Open 3D CityGML models have also been employed to assess the photovoltaic potential in Ludwigsburg County in southwest Germany(Rodríguez et al., 2017).

Due to legal constraints or cost concerns, many cities lack publicly available or comprehensive 3D building models, hindering the practical implementation of this technology. Acquiring DSM data for an entire city via UAVs is cost-prohibitive due to their limited range. As an alternative, satellite imagery offers a more economically feasible solution owing to its broader coverage and consistent update cycle. With the increasing spatial resolution of imagery, these images are widely employed in urban-scale rooftop availability identification, representing a more economical option for estimating photovoltaic potential on a large scale. Pan et al. utilized the vectorized building outlines of Guangzhou city from the Tianditu street map and measured the available rooftop space for different types of buildings using Google Maps(Pan et al., 2022). Mainzer et al. employ traditional image recognition techniques to detect partial roof areas, such as Canny Edge detection and Hough Transformation. They enhance publicly available aerial images of Freiburg, Germany, using histogram equalization, followed by the extraction of ridge lines. Finally, they calculate the azimuth of each roof as part of their analysis(Mainzer et al., 2017).

While traditional or manual image recognition methods are often cumbersome, there has been a growing adoption of deep convolutional neural networks (CNNs) in various complex image-processing tasks, including medical image segmentation and object detection. Recent studies indicate a rising trend in utilizing deep learning methods to extract building outlines from high-resolution imagery. To illustrate, the UNet architecture employs symmetric up-sampling and down-sampling pathways, along with skip connections to connect features from different hierarchical levels. This design enables U-Net to adeptly adjust to feature extraction across different scales, enhancing the model's capacity to recognize objects of varied sizes and shapes, such as buildings with diverse dimensions(Huang et al., 2019). Similarly, DeepLabV3 enhances its capability for detecting and segmenting objects at different scales by employing atrous convolutional structures and spatial pyramid pooling modules. The atrous convolutional structures utilize varying dilation rates to extract feature information at multiple scales, allowing for an expanded receptive field without adding parameters that could increase computational overhead. Zhong et al. optimize spatial sampling strategies using prior knowledge of land use to select training samples for training the DeepLabV3 model, allowing them to recognize buildings of different styles in Nanjing City(Zhong et al., 2021). Additionally, DeepLabV3 offers the flexibility to utilize different pre-trained backbones to adapt to various application scenarios and resource constraints, enhancing its versatility. In Yan et al.'s work, the DeepLabv3+ model is pre-trained on the Visual Object Classes Challenge and Cityscape Dataset to acquire preliminary knowledge in geographic object segmentation.Subsequent training on aerial imagery annotated with roof labels further enhances both the prediction accuracy and training efficiency of the model(Yan et al., 2023).

However, these studies assume that building roofs are flat. Neglecting roof structures, such as roof orientation, can result in an overestimation of solar potential, especially in mid to high-latitude regions, where the south rooftops receive significantly more solar radiation than the north ones. Although 3D model data offers richer roof structure features, we are inevitably confronted with the challenge of partial regions lacking LiDAR point cloud data. To address this, Lee et al. create a dataset annotated with roof orientations and propose a widely used end-to-end framework for predicting roof 3D structures. They directly infer the geometric shape and orientation of roofs from satellite imagery, achieving an average directional error of less than 10° in their predictions. When comparing the median available solar installation area estimated by the two methods, they find that this framework differs by less than ±11% compared to LiDAR-based methods(Lee et al., 2019). Li et al. point out that existing open-source datasets contain too many categories for roof orientation. This results in uneven sample distributions, potentially impacting the classification accuracy
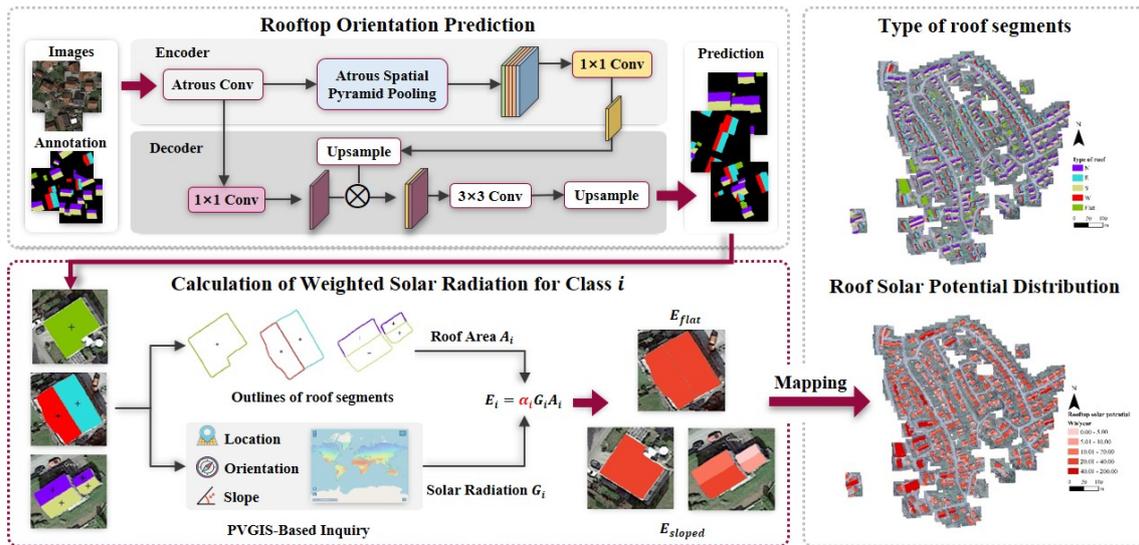
Figure 3. The flowchart of the proposed framework

of networks. Therefore, they merge 16 roof orientation categories into four and propose a multi-task learning network called SolarNet(Li et al., 2023). The results show a significant improvement in the accuracy of solar potential estimation.

We observe that rooftops often exhibit diverse appearances when images are captured at different solar azimuths, potentially leading to misclassification of orientation (as depicted in Figure 1). Specifically, the eastern rooftops exhibit enhanced brightness with the sun in the east (Figure 2a), while the western rooftops become more illuminated as the sun resides in the west (Figure 2c). Variations in rooftop color due to sun position can impact network classification accuracy. This study proposes a robust framework to estimate solar potential. The framework mainly comprises two modules: a rooftop orientation prediction network and a bilateral solar potential estimation module. To balance the data distribution, the research categorizes rooftops into five classes, including a flat roof class and four azimuth classes (east, south, west, and north). Initially, rooftop geometric boundaries are extracted from satellite imagery and classified using a semantic segmentation network. Subsequently, the two directional angle classes with a 180-degree difference are merged to reduce misclassification resulting from differences in data acquisition times. Finally, solar potential values are calculated based on the open-source solar energy database PVGIS and a weighted strategy.

2. Methodology

In this section, we present the pipeline of our proposed framework, as shown in Figure 3. A detailed description of the semantic segmentation network structure for rooftop extraction and classification is provided first, followed by the introduction of the weighted strategy for estimating solar radiation.

2.1 Rooftop Orientation Prediction Network

Considering the multiple scales of rooftop and the accuracy of boundary prediction directly influences the estimation of roof area, our framework comprises three key components: atrous convolution, atrous spatial pyramid pooling, and an encoder-decoder module, as depicted in Figure 4.

2.1.1 Atrous Convolution: In the task of rooftop segmentation, achieving a larger receptive field is crucial for improving performance, particularly due to the relatively large size of rooftop targets. Atrous convolution expands the conventional convolutional operation by introducing a dilation rate parameter, which governs the spacing between kernel elements. Unlike standard convolution, where kernel elements are positioned adjacent to each other, atrous convolution introduces gaps between kernel elements, allowing for an enlarged receptive field without increasing the number of parameters or the computational burden. Given an input feature map X and a kernel K, the atrous convolution operation is expressed as:

$$(X * K)_i = \int_{k,l} X_{i+rk,j+rl} K_{k,l} \qquad (1)$$

where i and j represent the spatial coordinates of the output feature map. The dilation rate r influences the sampling grid applied to the input feature map, effectively expanding the field of view of each layer in the network. A larger dilation rate enables the model to capture information from a broader region, facilitating efficient processing of multi-scale features.

2.1.2 Atrous Spatial Pyramid Pooling: As a pivotal component in semantic segmentation models, Atrous Spatial Pyramid Pooling (ASPP) is designed to enhance the model's ability to capture multi-scale contextual information. ASPP consists of multiple parallel convolutional branches, each utilizing atrous convolutions with different dilation rates. These dilation rates determine the sampling rates applied to the input feature map, effectively expanding the receptive field of each convolutional branch. By aggregating features from multiple scales in parallel, ASPP enables the model to capture contextual information across a range of spatial resolutions. The output Y of the ASPP layer is obtained by concatenating the feature maps produced by each convolutional branch, denoted as equ2:

$$Y = concat(X * W_n) \qquad (2)$$

Here, X represents the input feature map, while W_n refers to the convolutional kernels associated with each branch, each configured with a distinct dilation rate. In this study, we utilize

Figure 4. The structure of network

Figure 5. The classification scheme for azimuth classes

ASPP with Atrous Separable Convolution, which combines the efficiency of depthwise separable convolution with the ability of atrous convolution to capture contextual information across large spatial ranges. This approach significantly reduces the computational complexity of the model while maintaining or even improving performance. The operation consists of two main stages: depthwise atrous convolution, where filters are applied independently to each input channel with varying dilation rates to expand the receptive field, and pointwise convolution, which merges spatial features across channels.

2.1.3 Encoder-decoder Module: The encoder-decoder architecture is a widely adopted structural design in computer vision tasks, comprising two main components: the encoder and the decoder. The encoder is responsible for extracting high-level features from input images, effectively compressing the information into a lower-dimensional representation. These features capture important semantic information about the input images, such as object shapes, textures, and patterns. The decoder receives the encoded features and map the compressed features back to the original image space while preserving as much detail as possible. In our approach, the high-dimensional features obtained from the encoder undergo a fourfold up-sampling through bilinear interpolation before being concatenated with the low-level features from the backbone. Typically, low-level features contain a large number of channels, which can pose challenges during training. Therefore, a 1×1 Conv layer is employed to reduce the channel count of the low-level features. After concatenation, the features are refined using several 3×3 Conv layers, followed by another fourfold up-sampling. Through this process, the decoder performs pixel-level classification, assigning labels to individual pixels based on the information extracted by the encoder. By leveraging the encoder's ability to capture high-level semantic information and the decoder's capability to restore fine-grained details, this design proves beneficial in preserving both global context and local information in the reconstructed images.

2.2 Weighted Solar Radiation for Opposing Roof Orientations

In the preceding section, we utilized deep neural networks to determine the geometric shapes and types of roofs. In this section, we estimate the solar potential of each roof to calculate the total solar potential of the entire area. Photovoltaic Geographical Information System (PVGIS) is an open-source solar energy database that offers solar radiation data for any location worldwide except polar regions(Huld et al., 2012). By inputting relevant parameters, one can obtain the corresponding annual average solar radiation per unit area, denoted as G_i (Wh/m^2/year). As shown in equ3, s signifies the slope, θ represents the azimuth angle, and L_{lat} and L_{lon} stand for latitude

and longitude, respectively.

$$G_i = (s_i, \theta_i, L_{lat}, L_{lon}) \qquad (3)$$

For pitched roofs, the network tends to misclassify roofs of a specific orientation as belonging to another category that is 180° opposite. For instance, for the same building, a west-facing roof might be incorrectly labeled as east-facing, and vice versa. Pitched roofs are typically symmetrical, meaning the ratio of the roof area facing a particular orientation to the roof area facing the opposite orientation should be close to 1:1. Furthermore, in a study investigating the distribution of solar potential in rural areas of Northern China, the authors(Sun et al., 2022) classified rural buildings into three categories based on their architectural characteristics: the E-W pitched roof, N-S pitched roof, and flat roof. Inspired by their research and our assumptions, we assign a weighting factor for each category of orientation, denoted as α. The weighting factor α_i for the i-th category of roof is related to the category with an orientation differing by 180°. For flat roofs, the opposing category remains itself. The final formula for calculating the total solar potential is as follows:

$$E_{total} = \alpha_{flat}G_{flat}A_{flat} + \sum_{j=1}^{m} \alpha_j G_j A_j \qquad (4)$$

3. Experimental Result

3.1 Experimental Environment and Dataset

The RID dataset(Krapf et al., 2022) is a collection of data used for semantic segmentation, focusing on roof identification. The imagery used in this dataset is sourced from Google satellite images, which is known for providing high-resolution, cloud-free images with a precision of up to 0.15 meters in certain regions. Due to its wide availability and global coverage, Google satellite images has become a popular choice for roof identification and segmentation. The satellite images in this dataset are collected from WartenBerg, a city in Germany, comprising a total of 1880 images annotated with roof orientation. The dataset defines 18 categories, including 16 azimuth classes, a flat roof class with a slope defined as 0°, and a background class. The 16 azimuth classes cover a range of 22.5° for each class. In our study, we include a total of six categories, wherein the flat roof class and the background class remain unchanged. Moreover, the azimuth classes are reclassified according to rooftop orientation. To achieve a more balanced data distribution, the azimuth classes are divided into four categories (North, East, South, and

| N | E | S | W | Flat | Background |

Figure 6. Relative error comparison

Network	Scheme-1		Scheme-2		Scheme-3	
	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc
DeepLabV3+	81.44	89.15	72.72	83.68	54.33	63.18
HRNet	80.69	88.75	73.41	84.29	55.8	65.94
PSPNet	80.39	88.57	72.53	82.98	52.43	60.93
UNet	79.64	87.79	72.27	83.21	51.44	60.66

Table 1. Network performance of three classification schemes on RID dataset.

West), each spanning a range of 90 degrees. The classification scheme for azimuth classes is illustrated in Figure 5.

We employ the mmsegmentation framework to train four networks commonly used for image segmentation tasks to extract rooftop geometric shapes and predict their categories, including DeepLabV3+(Chen et al., 2018), PSPNet(Zhao et al., 2017), HRNet(Sun et al., 2019) and UNet(Ronneberger et al., 2015). Among these, the first two networks utilize ResNet18 as their backbone architecture. The dataset is partitioned into a training set, a testing set, and a validation set with a ratio of 7:2:1.

3.2 Rooftop Prediction

3.2.1 Metrics of Roof Classification Accuracy:
To assess the accuracy of predicted roof segments and orientations, Intersection over Onion (IoU) and accuracy (Acc) are computed for each category with the following equ5 and equ6, respectively:

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$Acc = \frac{TP}{TP + FP} \tag{6}$$

where TP, FP, and FN indicate the numbers of true positives, false positives, and false negatives, respectively. Afterward, the mean IoU (mIoU) and the mean accuracy (mAcc) are computed by averaging all classes.

Network	N	E	S	W	Flat
DeepLabV3+	86.65	82.98	86.23	80.19	55.96
HRNet	86.34	82.52	85.32	79.32	54.11
PSPNet	86.41	81.37	86.03	78.48	53.5
UNet	85.83	80.91	85.08	78.57	50.98

Table 2. Class-wise IoU Metrics for Scheme 1

Network	N	NE	E	SE	S	SW	W	NW	Flat
DeepLabV3+	77.98	71.53	69.44	76.04	74.99	69.04	67.1	75.8	48.78
HRNet	78.06	75.07	71.52	76.15	74.73	67.71	66.32	76.99	51.03
PSPNet	77.67	70.59	69.48	75.31	73.93	70.42	67.33	75.39	48.89
UNet	77.97	70.15	69.28	74.41	73.58	70.16	67.61	75.41	47.76

Table 3. Class-wise IoU Metrics for Scheme 2

3.2.2 Result of Roof Classification:
We evaluate the performance of different network models on the test set using three classification schemes, as depicted in Table 1. Scheme 1 corresponds to the classification scheme employed in this study, while Scheme 3 mirrors the classification scheme of the RID dataset. In Scheme 2, the 16 azimuth classes are consolidated into 8 classes, each spanning 45 degrees. Reducing the number of classes significantly improves classification accuracy. After reducing the number of classes, the performance of all networks increases by at least 7%, as revealed by comparing the mIoU of Scheme 1 and Scheme 2. DeepLabV3+ shows the largest improvement, reaching 8.72%. Furthermore, there is a significant disparity (approximately 20%) in mIoU and mAcc between Scheme 3 and the other two classification schemes. This indicates that overly detailed azimuth classes are unnecessary.

Table 2 and Table 3 respectively present the IOU of classes other than the background class under Scheme 1 and Scheme 2. DeepLabV3+ consistently outperforms other networks in Table 2. Except for flat roofs, the IoU of four azimuth classes exceeds 80%. Class N has the best classification accuracy (86.65%), with Class S coming in second with a marginal difference of only 0.42%. It is worth noting the substantial difference in classification accuracy between Classes E and W compared to Classes N and S across all models. This trend is consistent with observations made under Classification Scheme 2,

Figure 7. The zoomed-in results for the experimental area. (a) the type of roof segments, (b) the solar potential distribution map.

as depicted in Table 3. Using DeepLabV3+ as an example, the IoU of Class W is 10.88% lower than that of Class N, which may be attributed to the limited number of samples available for Class W.

While the class SE has the second-highest number of samples after the flat class, its IoU is still lower than that of the class N. This indicates that, beyond the impact of data sample quantity, the classification accuracy of roof imagery networks is notably affected by variations in roof color. Located in the mid-latitude region of the Northern Hemisphere, the experimental area consistently experiences shading on north-facing roofs. The distinct color changes observed on the east and west sides, resulting from fluctuations in solar azimuth angles, pose challenges for networks in learning stable features associated with them.

Figure 6 presents examples of roof geometry and roof category predictions generated by four networks on the RID dataset. Compared to DeepLabV3+, HRNet, and UNet, the predictions of roof geometry produced by HRNet and UNet exhibit less precise and clear borders. PSPNet shows glaring misclassifications, incorrectly categorizing certain roofs as Class N instead of Class S. The prediction of the flat roof by HRNet contains noticeable voids. It is evident that in the presence of pronounced architectural shadows, networks struggle to recognize complete flat roofs. This contributes to the relatively lower IoU of the flat class compared to the azimuth angle classes.

3.3 Solar Potential Estimation

To demonstrate the effectiveness of the proposed framework in solar potential estimation, relative error would be computed to assess the prediction accuracy of solar potential, it is defined as:

$$\varepsilon = \frac{\mid E_{pre} - E_{gt} \mid}{E_{gt}} \tag{7}$$

in which E_{pre} represents the predicted total solar energy potential, and E_{gt} represents the ground truth value.

Based on the assumptions outlined in Section2.2 and a comprehensive analysis of the study area, in this study, parameters for the flat category are set with both s and θ values at 0, and α is set to 1. For azimuth categories, s is set to 35 degrees, and θ_j is defined as the central azimuth angle value for each category.
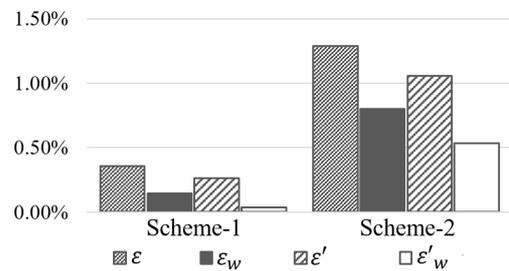
Figure 8. Comparison of relative errors

When lacking additional known geographical priors, the value of α_j is determined as follows:

$$\alpha_j = 0.5 + \frac{G_{jop}}{2G_j} \tag{8}$$

where G_{jop} represents the roof class that differs by 180° from the central azimuth of the j-th roof category.

The relative error between the total solar radiation values and the ground truth for the test area is computed, as depicted in Figure 8. The weighted relative error ε_w for Scheme 1 is 0.1450%, which represents a reduction of nearly 60% compared to the unweighted error ε. To mitigate the influence of errors in the flat category, we compute the relative error ε' and ε'_w for azimuth categories, resulting in a reduction to one-seventh of the unweighted results. Similarly, in the experiments of Scheme 2, the relative error of the orientation categories decreases by approximately 50%. These findings underscore the efficacy of our methodology. The total solar potential value of the experimental area's rooftops amounts to 12.49 GWh/year. The predicted rooftop categories and solar potential distribution are mapped, as illustrated in Figure 7.

4. Discussion

In this study, we propose a solar potential prediction framework that considers roof orientation. Since roofs of different orientations receive varying amounts of solar radiation, finer categorization of orientations is more advantageous for solar potential estimation. However, this must be established on the basis of sufficiently reliable network classification accuracy. Our experiments indicate that an excessive number of categories can lead

(a)

(b)
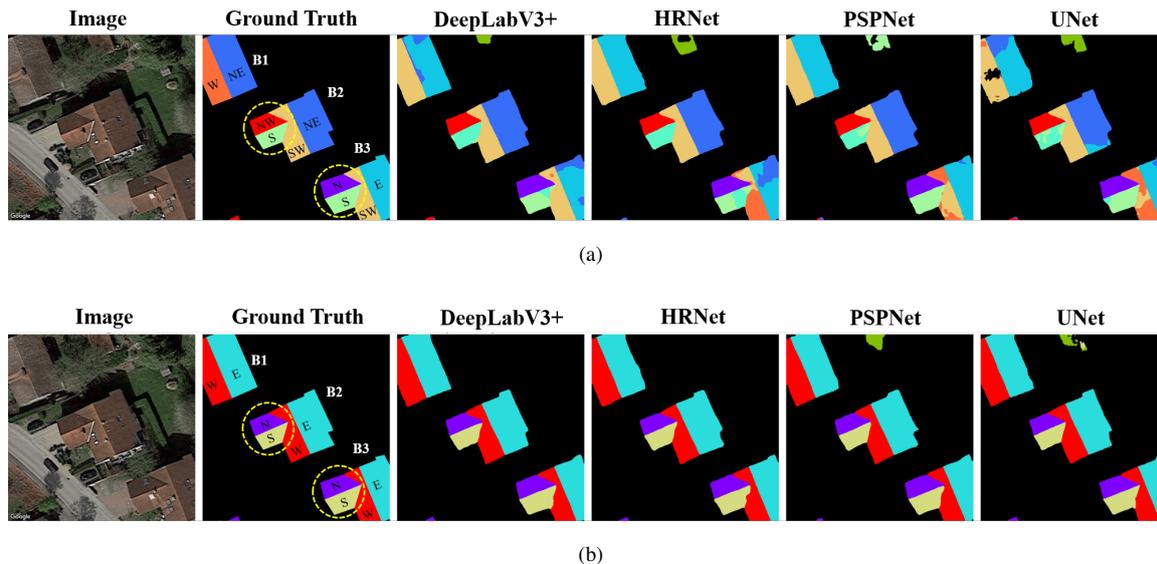
Figure 9. Examples of network prediction results for Scheme 2 and Scheme 1. (a) In Scheme 2, the azimuth angles of the two buildings circled in yellow appear similar, but their categories differ, as do those of B1, B2, and B3. (b) Under Scheme 1, all rooftop segments are correctly classified by four networks.

to lower classification accuracy. When the classification accuracy is relatively reliable (as observed in the results of Scheme 2), we observe that some roof predictions mix two categories, and these orientations are adjacent.

In Figure 9a, the UNet model mixes categories W and NE for the SW roof and the E roof of B3, respectively. It is observed that the orientations of these three buildings are nearly identical. As the azimuth angles of these roofs fall close to the threshold value, they are classified into different categories. However, this may prompt questions as to why they are not grouped under the same orientation. Simplifying the categorization of roof orientations into N, E, S, W not only streamlines the process but also reduces annotation and validation complexity compared to using 8 orientations. Therefore, we advocate for categorizing orientations into 4 groups. Unfortunately, the detection of some flat roofs is hindered by shadows cast by buildings. This issue is receiving increased attention in our ongoing research efforts.

We propose a weighted strategy based on the assumption of symmetry in pitched roofs. This strategy does not impose restrictions on the number of roof orientation categories. Instead, it only requires adherence to the condition of opposing orientations. We primarily focus on conventional pitched roofs and overlook irregular buildings or mixed-use zones. For broader applicability, we will apply our framework to other datasets in the future. For example, the DeepRoof dataset, which features more complex roof configurations, and urban datasets with clustered building heights. Variations in weather conditions significantly impact solar irradiance, but we directly utilize annual average solar radiation data obtained from PVGIS. In future work, we plan to incorporate dynamic factors to account for daily weather variations or seasonal effects.

5. Conclusion

In this study, we observe that rooftops exhibit diverse appearances when captured from different solar azimuth angles, potentially leading to misclassification of their orientation. To address this challenge, we propose a novel solar potential estima-

tion framework that considers various roof orientations. Leveraging our rooftop orientation prediction network, we achieve remarkable accuracy in determining the orientation of rooftops. By applying a weighting scheme, we effectively mitigate relative errors in the calculation of solar potential values. Experimental validation corroborates the effectiveness of our approach, demonstrating a significant enhancement in the accuracy of solar energy estimation.

Looking ahead, our future work will focus on expanding the scope of our analysis to encompass irregular and composite buildings. Additionally, we plan to integrate dynamic factors to better account for the impact of weather changes on solar potential, thereby further refining the precision and robustness of our methodology.

References

Buffat, R., Grassi, S., Raubal, M., 2018. A scalable method for estimating rooftop solar irradiation potential over large regions. *Applied energy*, 216, 389–401.

Byrne, J., Taminiau, J., Kurdgelashvili, L., Kim, K. N., 2015. A review of the solar city concept and methods to assess rooftop solar electric potential, with an illustrative application to the city of Seoul. *Renewable and sustainable energy reviews*, 41, 830–844.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *ECCV*.

Gassar, A. A. A., Cha, S. H., 2021. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. *Applied Energy*, 291, 116817.

Huang, Z., Mendis, T., Xu, S., 2019. Urban solar utilization potential mapping via deep learning technology: A case study of Wuhan, China. *Applied Energy*, 250, 283–291.

Huld, T., Müller, R., Gambardella, A., 2012. A new solar radiation database for estimating PV performance in Europe and Africa. *Solar energy*, 86(6), 1803–1815.

Krapf, S., Bogenrieder, L., Netzler, F., Balke, G., Lienkamp, M., 2022. Rid—roof information dataset for computer vision-based photovoltaic potential assessment. *Remote Sensing*, 14(10), 2299.

Lan, H., Gou, Z., Hou, C., 2022. Understanding the relationship between urban morphology and solar potential in mixed-use neighborhoods using machine learning algorithms. *Sustainable Cities and Society*, 87, 104225.

Lee, S., Iyengar, S., Feng, M., Shenoy, P., Maji, S., 2019. Deeproof: A data-driven approach for solar potential estimation using rooftop imagery. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2105–2113.

Li, Q., Krapf, S., Shi, Y., Zhu, X. X., 2023. SolarNet: A convolutional neural network-based framework for rooftop solar potential estimation from aerial imagery. *International Journal of Applied Earth Observation and Geoinformation*, 116, 103098.

Liu, F., Lv, T., 2019. Assessment of geographical distribution of photovoltaic generation in China for a low carbon electricity transition. *Journal of cleaner production*, 212, 655–665.

Lukač, N., Seme, S., Žlaus, D., Štumberger, G., Žalik, B., 2014. Buildings roofs photovoltaic potential assessment based on LiDAR (Light Detection And Ranging) data. *Energy*, 66, 598–609.

Mainzer, K., Killinger, S., McKenna, R., Fichtner, W., 2017. Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques. *Solar Energy*, 155, 561–573.

Ni, H., Wang, D., Zhao, W., Jiang, W., Mingze, E., Huang, C., Yao, J., 2024. Enhancing rooftop solar energy potential evaluation in high-density cities: A Deep Learning and GIS based approach. *Energy and Buildings*, 309, 113743.

Özdemir, S., Yavuzdoğan, A., Bilgilioğlu, B. B., Akbulut, Z., 2023. SPAN: An open-source plugin for photovoltaic potential estimation of individual roof segments using point cloud data. *Renewable Energy*, 216, 119022.

Pan, D., Bai, Y., Chang, M., Wang, X., Wang, W., 2022. The technical and economic potential of urban rooftop photovoltaic systems for power generation in Guangzhou, China. *Energy and Buildings*, 277, 112591.

Ren, H., Ma, Z., Chan, A. B., Sun, Y., 2023. Optimal planning of municipal-scale distributed rooftop photovoltaic systems with maximized solar energy generation under constraints in high-density cities. *Energy*, 263, 125686.

Ren, H., Xu, C., Ma, Z., Sun, Y., 2022. A novel 3D-geographic information system and deep learning integrated approach for high-accuracy building rooftop solar energy potential characterization of high-density cities. *Applied Energy*, 306, 117985.

Rodríguez, L. R., Duminil, E., Ramos, J. S., Eicker, U., 2017. Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach. *Solar Energy*, 146, 264–275.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. *CVPR*.

Sun, T., Shan, M., Rong, X., Yang, X., 2022. 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. *Applied Energy*, 315, 119025.

Walch, A., Castello, R., Mohajeri, N., Scartezzini, J.-L., 2020. Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. *Applied Energy*, 262, 114404.

Wang, Y., He, J., Chen, W., 2021. Distributed solar photovoltaic development potential and a roadmap at the city level in China. *Renewable and Sustainable Energy Reviews*, 141, 110772.

Wiginton, L., Nguyen, H. T., Pearce, J. M., 2010. Quantifying rooftop solar photovoltaic potential for regional renewable energy policy. *Computers, Environment and Urban Systems*, 34(4), 345–357.

Wong, M. S., Zhu, R., Liu, Z., Lu, L., Peng, J., Tang, Z., Lo, C. H., Chan, W. K., 2016. Estimation of Hong Kong's solar energy potential using GIS and remote sensing technologies. *Renewable Energy*, 99, 325–335.

Yan, L., Zhu, R., Kwan, M.-P., Luo, W., Wang, D., Zhang, S., Wong, M. S., You, L., Yang, B., Chen, B. et al., 2023. Estimation of urban-scale photovoltaic potential: A deep learning-based approach for constructing three-dimensional building models from optical remote sensing imagery. *Sustainable Cities and Society*, 93, 104515.

Zhang, Z., Chen, M., Zhong, T., Zhu, R., Qian, Z., Zhang, F., Yang, Y., Zhang, K., Santi, P., Wang, K. et al., 2023. Carbon mitigation potential afforded by rooftop photovoltaic in China. *Nature Communications*, 14(1), 2347.

Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., Zhu, R., Zhang, F., Zhang, H., Zhou, F. et al., 2022. Vectorized rooftop area data for 90 cities in China. *Scientific Data*, 9(1), 66.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. *CVPR*.

Zhong, T., Zhang, Z., Chen, M., Zhang, K., Zhou, Z., Zhu, R., Wang, Y., Lü, G., Yan, J., 2021. A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy*, 298, 117132.

Zhu, R., Wong, M. S., You, L., Santi, P., Nichol, J., Ho, H. C., Lu, L., Ratti, C., 2020. The effect of urban morphology on the solar capacity of three-dimensional cities. *Renewable Energy*, 153, 1111–1126.