Energy Proceedings
Vol 47, 2024
ISSN 2004-2965

A High-Precision Method for Photovoltaic Panel Segmentation Combining Large-Scale Model Prior Knowledge and Multimodal Information

Lingchengjia Zhou[1], Kechuan Dong[2], Hongjun Tan[3,4], Jiaze Li[5], Qing Yu[6], Zhiling Guo[3,4*], Yuhao Yao[7], Haoran Zhang[7], Jinyue Yan[3,4]

1 The Hong Kong University of Science and Technology

2 Institute of Industrial Science, The University of Tokyo

3 Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University

4 International Centre of Urban Energy Nexus, The Hong Kong Polytechnic University

5 Beijing University of Technology, No.100, Pingleyuan, Chaoyang District, Beijing, China

6 School of Urban Planning and Design, Peking University Shenzhen Graduate School

7 LocationMind Inc., 701, 3-5-2, Iwamoto-cho, Chiyoda-ku, Tokyo 101-0032, Japan

(Corresponding Author: zhiling.guo@polyu.edu.hk)

ABSTRACT

The widespread adoption of distributed photovoltaic (PV) systems highlights the need for sophisticated segmentation technologies that can accurately identify PV panels, essential for calculating potential capacity and informing development strategies. Although artificial intelligence has significantly advanced the accuracy and reliability of PV panel segmentation, real-world complexities such as diverse panel types, installation methods, and varied backgrounds pose challenges to model adaptability and generalization. This research introduces a method that enhances PV panel segmentation by employing the enhanced Segment Anything Model, which has been extensively pre-trained using a comprehensive real-world dataset to incorporate multimodal semantic information, thus improving generalization. Additionally, a fine-tuning process has been integrated to better absorb critical features from the training data, increasing the model's sensitivity to the unique characteristics of specific PV installations. Field tests in Heilbronn, Germany, confirm the method's superior performance and flexibility, underscoring its potential to support strategic planning for large-scale PV deployment.

Keywords: renewable energy, photovoltaic panel, computer vision, remote sensing, semantic segmentation, deep learning

NONMENCLATURE

Abbreviations	
PV	Photovoltaic
SAM	Segment Anything Model
CNN	Convolutional Neural Networks
RPN	Region Proposal Network
BCE	Binary Cross Entropy

1. INTRODUCTION

1.1 Background

The rapid development of solar energy systems has highlighted the urgent need for advanced image segmentation techniques. These systems are critical to the development of sustainable energy and require meticulous modeling to accurately assess potential energy capacity and assist in strategic planning. Despite significant progress in artificial intelligence models, the practical implementation of accurately segmenting photovoltaic panels still faces difficulties in adjusting and generalizing models. Given the wide range of types and sizes of solar panels, different installation methods, and complex backgrounds, the use of deep learning models has become an important choice for effectively managing these complex features.

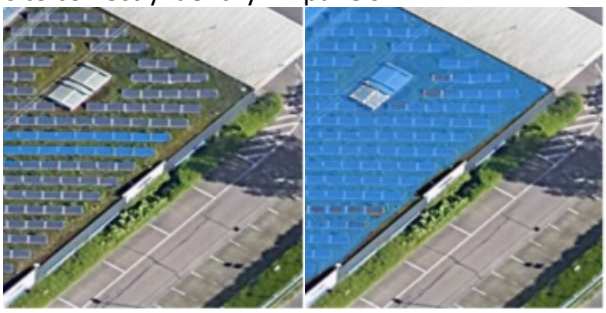
In the realm of PV panel analysis, existing studies have endeavored to enhance segmentation outcomes by modifying traditional deep-learning architectures. For instance, models such as U-Net [1], SegNet [2], and DeepLab v3+ [3] have been adapted to incorporate specific spectral and textural features. While these modifications yield commendable performance under

This is a paper for the 10th Applied Energy Symposium: Low Carbon Cities & Urban Energy Systems (CUE2024), May. 11-13, 2024, Shenzhen, China.

certain conditions, they generally lack the adaptability required to handle the unique geometric shapes and complex installation environments of PV panels. Moreover, these models typically rely on extensive labeled data and struggle to cope with the challenges posed by the diversity in PV panels.

SAM (Segment Anything Model) [4] is an image segmentation model. It contains a powerful image encoder, a versatile hint encoder, and an efficient mask decoder. The model can accept a variety of input hints, including foreground/background points, approximate boxes or masks, free-form text, etc. It then generates accurate segmentation masks based on these hints. The model's approach can be trained without the need for specific project labels, enabling seamless transfer to new segmentation tasks using hint engineering without additional training.

Given the notable transferability and dataset independence of the SAM, we considered employing the SAM framework for our study. However, when directly applied to PV panel image segmentation, the SAM model faces limitations due to a lack of relevant pre-trained data and semantic labels, which means insufficient adjustments and optimizations for specific and complex image contents. As shown in Figure 1, even with the use of bounding points(a) and boxes(b) prompts, the model fails to correctly identify PV panels.



a. Points prompt *b. Boxes prompt*

Fig. 1 Using SAM to segment PV panels

1.2 Literature Review

Currently, academics have observed the potential of SAM in the domain of remote sensing picture segmentation and researched it. This section will examine studies that are closely relevant to this study. These studies demonstrate how to utilize and enhance the SAM model to enhance the accuracy of segmenting specific cases.

In 2023, Chen et al. introduced RSPrompter [5], a method leveraging prompt learning within the SAM framework to produce semantically distinct segmentations in complex remote sensing images, even under challenging conditions with unclear boundaries. Following this, Sultan et al. developed GeoSAM [6], which fine-tunes SAM using both sparse and dense visual prompts to optimize segmentation for automated mobility infrastructure in urban environments, demonstrating the model's adaptability to varied urban scenes. Additionally, in 2024, Zhang et al. proposed RSAM-Seg [7], an adaptation of SAM that incorporates prior knowledge and internal structural modifications, such as adapter modules in the encoder, to improve recognition of specific features in remote sensing data. These enhancements not only demonstrate SAM's broad applicational potential but also highlight the importance of tailored adjustments and optimizations to meet the specific demands of diverse segmentation tasks, underscoring the need for deeper model understanding and improvement to achieve superior segmentation performance.

1.3 Motivations and Contributions

This study is motivated by the need to address adaptability and accuracy issues encountered by the SAM in photovoltaic PV panel segmentation tasks. By freezing the encoder to stabilize feature extraction, and enhancing the structure of the decoder and prompter, this research aims to improve the model's ability to adapt to and accurately segment the unique attributes of PV panels. Additionally, the study explores fine-tuning the model using high-quality annotated data to enhance its performance and generalization capabilities in specific applications. The development of this approach not only aims to increase the precision of PV panel segmentation but also to optimize the underlying structure of the large model, quickly adapt to new downstream tasks through zero-shot or few-shot learning, and improve generalization capabilities.

2. MATERIAL AND METHODS

2.1 Datasets

In this study, we used the Heilbronn Rooftop PV System Dataset (H-RPVS Dataset) to train and evaluate our PV panel segmentation models

H-RPVS Dataset is a public dataset for small-scale rooftop PV system segmentation. This dataset includes 5866 pairs of PV panel sample images from Heilbronn, Germany, each consisting of a 256×256-pixel image and its corresponding label. The samples are collected via Google Earth with a spatial resolution of 0.15 meters. We gratefully acknowledge the contributions of Wang et al.

[8] for developing this valuable dataset and related experimental results.

2.2 Model Structure

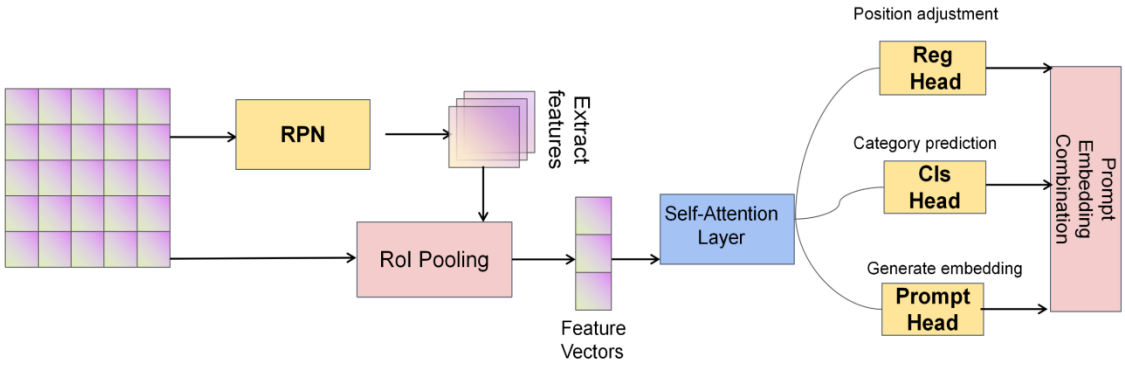


Fig. 3 Prompter Structure

This study presents an improved architecture for the SAM to enhance segmentation accuracy and model generalization capabilities. Figure 2 illustrates the overall framework of the proposed model.

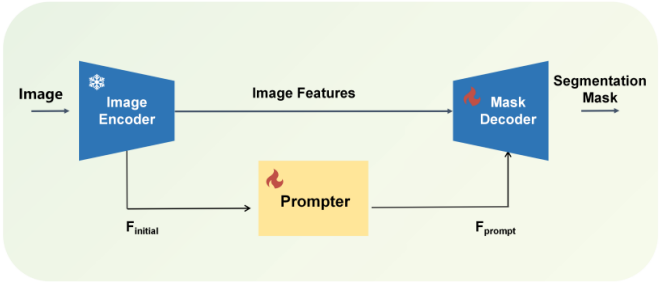


Fig. 2 Model Structure

Initially, the input image is processed by the Image Encoder to extract multi-scale features and provide rich contextual information. This encoder is based on the pre-trained SAM, utilizing deep CNN [9]. During training, the parameters of the Image Encoder are kept frozen to ensure stable feature extraction and prevent noise from fine-tuning.

Based on the contributions of Chen et al. [5], we designed an advanced Prompter adapted to the task of PV panel segmentation. The extracted features are then fed into this advanced Prompter, which generates task-specific prompt information, including category and location hints. The Prompter uses a Region Proposal Network and RoI Pooling [10] for feature extraction, and a self-attention mechanism to emphasize important features. Classification and regression operations are subsequently performed to determine the object's category and position. The enhanced features are then used to generate prompt embeddings for the Mask Decoder through the Prompt Head.

These prompt embeddings, combined with the image features, are input into the Mask Decoder. The Mask Decoder, structured around a CNN, decodes the prompt information and combines it with the image features to generate the final segmentation mask. This process captures fine details within the image to improve segmentation accuracy. The entire training process involves joint optimization of the Prompter and Mask Decoder to minimize segmentation errors and enhance segmentation quality, thus improving segmentation accuracy and model generalization across various image scenes.

By freezing the Image Encoder and enhancing the Prompter and Mask Decoder, the proposed model adapts better to complex backgrounds, improving the precision of PV panel segmentation. This method provides reliable data support for monitoring and assessing photovoltaic systems, achieving high-precision segmentation through the coordinated operation of each module as shown in Figure 2.

2.3 Advanced Prompter Design

As shown in Fig. 3, we have implemented several key design changes to the advanced prompter to enhance its performance for PV panel segmentation: an RPN [10] module after feature extraction to generate candidate bounding boxes, allowing the model to focus on key regions and improving segmentation accuracy; RoI Pooling [10] to process these bounding boxes into fixed-size feature maps, reducing computational complexity and ensuring consistent processing; a self-attention mechanism post-feature extraction to capture global information, highlight important features, and suppress irrelevant ones, thus improving feature representation quality; classification and regression heads for category prediction and fine-tuning bounding box positions, increasing prompt accuracy; and a prompt head that

generates prompt embeddings combining spatial and semantic information, effectively guiding the mask decoder to produce high-quality segmentation masks.

The mathematical formulation of our advanced prompt generator is as follows:

$$\begin{aligned} F &= \text{Encoder}\ (I) \\ B &= \text{RPN}\ (F) \\ F_{\text{pool}} &= \text{RoI}\ \ \text{Pooling}\ (F, B) \\ F_{\text{attn}} &= \text{Self}\ -\text{Attention}\ \left(F_{\text{pool}}\right) \\ \text{Class}_{\text{pred}} &= \text{Classification Head}\ (F_{\text{attn}}) \\ \text{Reg}_{\text{pred}} &= \text{Regression Head}\ \ (F_{\text{attn}}) \\ \text{Prompt}_{\text{embed}} &= \text{Prompt Head}\ (F_{\text{attn}}) \end{aligned} \quad (1)$$

This formulation describes the process where the input image I is first processed by an encoder to extract features F. These features are then processed by an RPN to generate candidate bounding boxes B. The bounding boxes are pooled using RoI Pooling to obtain fixed-size feature maps F_{pool}. A self-attention mechanism is then applied to these pooled features, resulting in F_{attn}. The enhanced features are then passed through a classification head and a regression head to predict categories and adjust positions, respectively. Finally, the prompt head generates a prompt segmentation task.

By integrating these improvements, our proposed model significantly enhances segmentation accuracy and robustness in complex and diverse scenes compared to the original SAM.

2.4 Loss Function

The design of the loss function is critical in the new structure. We constructed a complete loss function that incorporates classification loss, regression loss, region proposal network (RPN) loss, and mask loss to ensure the high accuracy of the model in the photovoltaic panel segmentation task.

2.4.1 Prompter Loss

The combined loss for the RPN and the prompt generator is defined as follows:

$$\mathcal{L}_{\text{Prompter}} = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_{\text{rpn}}^{i} + \frac{1}{N}\sum_{j=1}^{N} \left(\mathcal{L}_{\text{cls}}^{j} + \mathcal{L}_{\text{reg}}^{j}\right) \quad (2)$$

where M represents the number of candidate bounding boxes, N represents the number of classification and regression targets, $\mathcal{L}_{\text{rpn}}^{i}$ is the RPN loss for the i-th candidate box, $\mathcal{L}_{\text{cls}}^{j}$ is the classification loss, and $\mathcal{L}_{\text{reg}}^{j}$ is the regression loss.

2.4.2 Decoder Loss

To better solve the problem of data imbalance and complex background in the photovoltaic panel segmentation task, we use the Dice loss function to calculate the mask error. The Dice loss function is chosen because photovoltaic panels usually occupy only a small part of the image, while the background occupies most of the pixels. The traditional BCE loss is affected by this imbalance, causing the model to focus more on learning background pixels and ignore photovoltaic panel pixels. Dice loss effectively alleviates this problem by directly optimizing the overlap between the predicted and ground truth masks.

$$\mathcal{L}_{\text{Mask}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \quad (3)$$

2.4.3 Overall Loss Function

To integrate the promoter and mask losses, we introduce weighting parameters α and β to to balance the contributions of each loss component. The overall loss function is defined as follows:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{Prompter}} + \beta\mathcal{L}_{\text{Mask}} \quad (4)$$

3. RESULT

In this study, we performed an exhaustive evaluation of the proposed SAM optimization using the H-RPVS Dataset and compared it with several widely used deep learning models [8]. Through these evaluations, we were able to validate the performance of the optimized SAM on the PV panel segmentation task.

3.1 Evaluation Metrics

To evaluate the performance of the model, we used the following indicators and gave the corresponding formulas:

3.1.1 Precision

Precision indicates the proportion of positive samples predicted by the model that are actually positive samples. The formula is:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Where, TP is a true positive sample, FP is a false positive sample.

3.1.2 Recall rate

Recall indicates the proportion of positive samples correctly predicted by the model to be positive samples. The formula is:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

Where, FN is a false negative sample.

3.1.3 F1-Score

F1-Score is the harmonic mean of precision and recall, which is used to evaluate the performance of the model comprehensively. The formula is:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

3.1.4 IoU (Intersection over Union)

IoU is used to measure the overlap between the predicted segmentation mask and the true mask. The formula is:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (8)$$

The specific calculation is:

$$\text{IoU} = \frac{TP}{TP+FP+FN} \quad (9)$$

3.2 Model Performance Comparison

We compared the proposed method with some widely used deep learning models [8], and the results are shown in the table:

Model	Precision	Recall	F1-Score	IoU
U-Net	0.8966	0.9028	0.8997	0.8176
Deeplab v3+	0.9031	0.9386	0.9205	0.8527
U-Net++	0.9566	0.9561	0.9564	0.9164
Our method	0.9642	0.9592	0.9617	0.9262

Table 1. Comparison with some widely used Deep Learning Model

As can be seen from the table, our method outperforms other models in terms of precision, recall, F1 score, and IoU indicators, especially in the IoU indicator, our method reaches 0.9262, showing superior performance in the photovoltaic panel segmentation task.

4. DISCUSSION

To enhance the performance and versatility of the enhanced SAM structure, we want to implement the following enhancements and optimizations: Initially, we will execute ablation experiments to validate the impact of each enhanced module on the overall performance of the model. This will allow us to identify the most crucial components for enhancing performance and optimize them more efficiently. Furthermore, our objective is to incorporate multi-modal and multi-scale information input. By combining data from several modalities, such as optical images and radar data, we can enhance the model's performance in diverse settings. Furthermore, the utilization of various scales in processing information will enhance the ability to accurately delineate the distinctive features of solar panels across different levels of detail. Furthermore, we will investigate methods to streamline the model architecture or employ less resource-intensive models to minimize computational burden and enhance the model's applicability in real-world situations. Ultimately, our intention is to conduct extensive testing and implement the enhanced SAM framework over a wider range of geographical areas and varied settings. Through the examination of photovoltaic systems in various geographic locations, we can confirm the worldwide suitability of the model and make necessary modifications based on the distinct attributes of each area. By implementing these enhancements and optimizations, we are confident that the enhanced SAM structure will deliver outstanding performance in the task of segmenting photovoltaic panels. This will greatly contribute to research and application in related industries

5. CONCLUSIONS

This study proposes a high-precision PV panel segmentation method that combines large-scale model prior knowledge and multimodal information, achieving accurate identification and segmentation of photovoltaic panels through the optimization of the SAM. By pre-training on a comprehensive real-world dataset containing multimodal semantic information, the model's generalization capability is enhanced. Additionally, a fine-tuning process is integrated to better absorb critical features from the training data, increasing the model's sensitivity to specific photovoltaic installation characteristics. Field tests in Heilbronn, Germany, demonstrate the method's superior performance and flexibility. Compared to other deep learning models, our method excels in precision, recall, F1 score, and IoU metrics, achieving an IoU score of 0.9262, significantly enhancing the accuracy and generalization of photovoltaic panel segmentation tasks.

Our study results indicate that leveraging and optimizing the SAM model shows excellent performance and significant potential in the field of PV panel segmentation.

ACKNOWLEDGEMENT

We gratefully acknowledge the support from the JICA and the MOST of China for the Japan-China Cooperative Project on "Research on Human-Source-Load-Carbon Synergy Optimization Technology for Carbon Neutral City Energy System Driven by Population Trajectory Big Data."

REFERENCE

[1] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Badrinarayanan, Vijay, et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, 1 Dec. 2017, pp. 2481–2495.

[3] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Openaccess.thecvf.com.https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html

[4] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W.-Y., Dollár, P., & Girshick, R. (n.d.). Segment Anything. https://arxiv.org/pdf/2304.02643

[5] Chen, K., Liu, C., Chen, H., Zhang, H., Li, W., Zou, Z., & Shi, Z. (2023, November 29). RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation based on Visual Foundation Model. ArXiv.org. https://doi.org/10.48550/arXiv.2306.16269

[6] Sultan, R., Li, C., Zhu, H., Khanduri, P., Brocanelli, M., & Zhu, D. (2024). GeoSAM: Fine-tuning SAM with Sparse and Dense Visual Prompting for Automated Segmentation of Mobility Infrastructure. https://arxiv.org/pdf/2311.11319

[7] Zhang, J., Yang, X., Jiang, R., Shao, W., & Zhang, L. (2024, February 29). RSAM-Seg: A SAM-based Approach with Prior Knowledge Integration for Remote Sensing Image Semantic Segmentation. ArXiv.org. https://doi.org/10.48550/arXiv.2402.19004

[8] Wang, J., Chen, X., Shi, W., Jiang, W., Zhang, X., Hua, L., Liu, J., & Sui, H. (2023). Rooftop PV Segmenter: A Size-Aware Network for Segmenting Rooftop Photovoltaic Systems from High-Resolution Imagery. Remote Sensing, 15(21), 5232. https://doi.org/10.3390/rs15215232

[9] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 1–21. https://doi.org/10.1109/tnnls.2021.3084827

[10] Ren, S., He, K., Girshick, R., & Sun, J. (2017b). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031