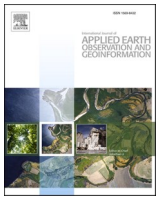
International Journal of Applied Earth Observations and Geoinformation 116 (2023) 103134

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Deep solar PV refiner: A detail-oriented deep learning network for refined segmentation of photovoltaic areas from satellite imagery

Rui Zhu [a,b,c], Dongxue Guo [a], Man Sing Wong [a,b,*], Zhen Qian [d,e,f], Min Chen [d,e,f], Bisheng Yang [g], Biyu Chen [g], Haoran Zhang [h], Linlin You [i], Joon Heo [j], Jinyue Yan [k]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[b] Research Institute for Land and Space, The Hong Kong Polytechnic University, Hong Kong, China
[c] Systems Science Department, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[d] Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China
[e] State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing 210023, China
[f] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[g] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[h] School of Urban Planning and Design, Peking University, Shenzhen 100871, China
[i] School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen 518107, China
[j] Department of Civil and Environmental Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea
[k] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

ARTICLE INFO

Keywords:
Remote Sensing
Deep Learning
Semantic Segmentation
Photovoltaic Area Segmentation
Solar Energy

ABSTRACT

To estimate electricity generation and evaluate the socio-economic effects of solar photovoltaic (PV) systems, it is critical to calculate the installed PV areas and quantify the installed capacity over a large region. Although general deep learning networks can be used to extract PV areas from satellite imagery, the capability of segmenting small and distributed ones with accurate and refined boundaries is still lacking. This is because significantly small foreground objects (i.e., PV areas) severely impeded by large and highly diverse background contexts, background objects having similar characteristics to PV modules are easily misclassified, and PV modules under various daylighting conditions present varying textures and colours. To overcome the challenges, this study proposes Deep Solar PV Refiner, a detail-oriented deep learning network, to enhance PV segmentation from satellite imagery. The proposed network advances the backbone by incorporating Split-Attention Network, combines Dual-Attention Network with Atrous Spatial Pyramid Pooling using four different structures, and integrates PointRend Network that refines PV boundary prediction. With transfer learning, a synthetic strategy, hybrid loss functions, and ablation experiments, the optimal network is obtained that outperforms the benchmark by 5%, 2%, 3%, 3%, and 2% for IoU, Accuracy, F1-score, Precision, and Recall, respectively. The network is also competitive with the state-of-the-art semantic segmentation networks and has a favourable generalization capability, with the mean IoU increasing by 0.63–11.18%. The new network effectively improves the capability of segmenting hard and small PV samples, which is deliverable to different areas and is significant for estimating the installed capacity of PV systems.

1. Introduction

Fossil fuel consumption has caused severe environmental problems, such as air pollution (Wong et al., 2021), global warming (McGlade & Ekins, 2015), and the urban heat island effect (Zhu et al., 2017), which urges people to promote energy transition with the penetration of renewable energy. For solar power generation, photovoltaic (PV) panels are increasingly being used for solar farming (Inderberg et al., 2018) and a substantial number of PV power production systems have been installed in many countries (Zhang et al., 2021; Sabadini & Madlener 2021).

To facilitate PV installation, the academic community has estimated land surface solar irradiation (Liao et al., 2022), modelled PV potential on three-dimensional urban surfaces (Wong et al., 2016; Zhong et al.,

* Corresponding author at: Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China.
E-mail address: Ls.charles@polyu.edu.hk (M.S. Wong).

https://doi.org/10.1016/j.jag.2022.103134
Received 26 August 2022; Received in revised form 25 November 2022; Accepted 27 November 2022
Available online 6 December 2022

2021; Zhu et al., 2019; Zhu et al., 2020; Zhu et al., 2022a), and optimized the location and installed capacity of PV modules (Zhu et al., 2022b; Zhu et al., 2022c). In general, these studies effectively supported PV planning considering electricity generation and life-cycle techno-economic assessment. Since massive PV modules have been installed in the past few years, increasing attention has been paid to extracting installed PV areas from satellite imagery over large geographical areas (Wang et al., 2018; Li et al., 2021), which is vital for estimating the total installed capacity and PV potential (Vries et al., 2020), analyzing the economic and environmental impacts (Lukač et al., 2016), and guiding PV policy-making (Chapman et al., 2016).

Recently, the methods for PV area extraction have been developed using remote sensing image segmentation (Wang et al., 2018), machine learning (Joshi et al., 2021; Chen et al., 2022a, b), and advanced deep learning (Costa et al., 2021; Kruitwagen et al., 2021; Li et al., 2021; Jiang et al., 2021). They suggested that machine learning models can achieve high accuracy on PV area segmentation, and deep learning models, such as Convolutional Neural Networks (CNN), are capable of extracting PV areas that are affected by various background environments and multi-source satellite images with different resolutions.

As deep learning networks have shown competitiveness in image segmentation (Zhang et al., 2022a; Liu et al., 2022; Li et al., 2022), PV areas have been extracted from satellite imagery based on deep learning, such as DeepSolar (Yu et al., 2018) and ConvNet (Yuan 2018). However, general deep learning models still meet difficulties in PV area segmentation. Previous studies indicate that (i) certain PV modules, such as small and distributed on rooftops, sometimes are hard to be detected and thus cannot be segmented from the images (Li et al., 2021); (ii) PV modules under various daylighting conditions present varying textures and colours that are confusable to other geo-objects and thus can be easily misclassified (Costa et al., 2021); and (iii) boundaries of the predicted PV areas are irregularly saw-toothed, implying inaccurate classification of pixels along the PV edges (Kruitwagen et al., 2021). These inconspicuous errors may be accumulated and propagated to be apparent mistakes when estimating the total installed capacity of PV modules, causing uncertain and unreliable analysis.

Furthermore, sophisticated CNN architectures were used to separate PV areas from satellite images. One study developed a cross-learning driven U-Net to extract PV areas, which achieved the maximum IoU of 74.92 % (Bradbury et al., 2016; Zhuang et al., 2020). Most recently, Deeplabv3+, which is a type of CNN employing an Encoder-Decoder structure and utilizing Atrous Spatial Pyramid Pooling (ASPP) to expand the receptive field, has been used to extract PV areas (Jiang et al., 2021; Costa et al., 2021), rooftop areas (Zhang et al., 2022b; Lin et al., 2022), and rooftop structural lines (Qian et al., 2022a). These studies suggested that Deeplabv3 + was competitive compared to models such as U-Net, Fully Convolutional Network, and Feature Pyramid Network.

Further study examined the PV segmentation performance of the U-Net, RefineNet, and DeepLabv3 + on various satellite image resolutions in Jiangsu, China (Jiang et al., 2021), which found that DeepLabv3 + outperformed the other two models in terms of Precision, F1-score, and IoU. It was also noticed that DeepLabv3 + tends to guarantee the extracted PV area dependable by sacrificing some parts of PV, whereas U-Net and RefineNet attempt to identify as many PV modules as possible at the expense of precision and yield more misclassified regions than DeepLabv3+ (Jiang et al., 2021). Therefore, DeepLabv3 + is preferable to U-Net and RefineNet for extracting PV areas.

Although PV segmentation results obtained from Deeplabv3 + were generally superior to some CNN approaches and may provide reasonable PV area estimation, it still has limitations since some PV boundaries were over-smoothed or saw-toothed with a drop of IoU, some geo-objects were misclassified as PV modules when their spatial textures were similar with each other, and some PV modules could not be detected due to the reflection of sunlight, resulting in significantly different spectral information than usual. As DeepLabv3 + has demonstrated the best performance and dependability among the studied deep learning approaches, this study will propose a detail-oriented deep learning network based on DeepLabv3 + to address the above limitations. Specifically, this study will develop a synthetic strategy to optimize the backbone to enhance feature map attention and multipath representation (Qian et al., 2022a), incorporate the Dual-Attention mechanism to alleviate the effects of confusable background context and the complex object forms (Lu et al., 2020), add the PointRend module to improve the edge detection accuracy (Qiu et al., 2021), and optimize loss functions to solve the class imbalance problem (Das et al., 2021).

This study aims to propose a detail-oriented deep learning network to enhance the PV area segmentation from satellite images, which has three major contributions. First, a detail-oriented deep learning network is developed for PV segmentation, which is competitive with the basic Deeplabv3 + and other state-of-the-art networks. Second, this study provides a better understanding of deep learning networks for extracting the features of positive and negative samples for effective semantic segmentation. Third, the well-trained network can be utilized for transfer learning to perform fine-scale PV segmentation in different areas, and the network can be generalized and deliverable for the other imaged segmentation applications.

2. Materials and methodology

2.1. Study area and the datasets

This study selects Heilbronn as the study area, which is a city in northern Baden-Württemberg, Germany, with an abundance of distributed PV installations (Fig. 1). A set of patch images were collected from Google Earth Satellite with three optical bands (i.e., Red, Green, and Blue) that covered a 115.6 km^2 area at a resolution of 0.15 m. The dataset has been proved being effective for PV area segmentation (Li et al., 2021), and it has been demonstrated that images at this resolution can fulfil the PV segmentation task (Jiang et al., 2021). Meanwhile, 7197 manually labelled PV areas covering a total area of 1.0 km^2 were obtained from the related literature (Li et al., 2021). The ratio of PV to background is about 1:20.1, indicating the challenge of PV segmentation task because of the significant imbalance of the positive and negative samples. Also, rooftop polygons were downloaded from open-sourced OpenStreetMap.

2.2. Research framework

Fig. 2 describes the research framework of this study. To begin with, satellite images and rooftop PV labels were processed to fulfill the requirements of deep learning. Then, detail-oriented deep learning networks with various structures and the implementation of loss functions were proposed to improve PV segmentation capability, and the optimal network was determined based on a series of ablation experiments. After that, the new network was compared with other networks to illustrate the effectiveness of the new network on rooftop PV segmentation. Specifically, this study (i) utilizes transfer learning to fine-tune initial parameters, (ii) advances the backbone using the Split-Attention module to enhance the channel-wise attention on different network branches, which is expected to improve the capability in capturing cross-feature interactions and learning diverse representations between PV and non-PV classes, (iii) integrates Dual-Attention module with ASPP to alleviate the effects of confusable background context and the complex object forms on PV area segmentation, (iv) incorporates PointRend module to improve the PV edge detection accuracy, and (v) optimizes loss functions to solve the class imbalance problem.

2.3. Refinement of the backbone

As shown in Fig. 3, the backbone serves as the input of the ASPP

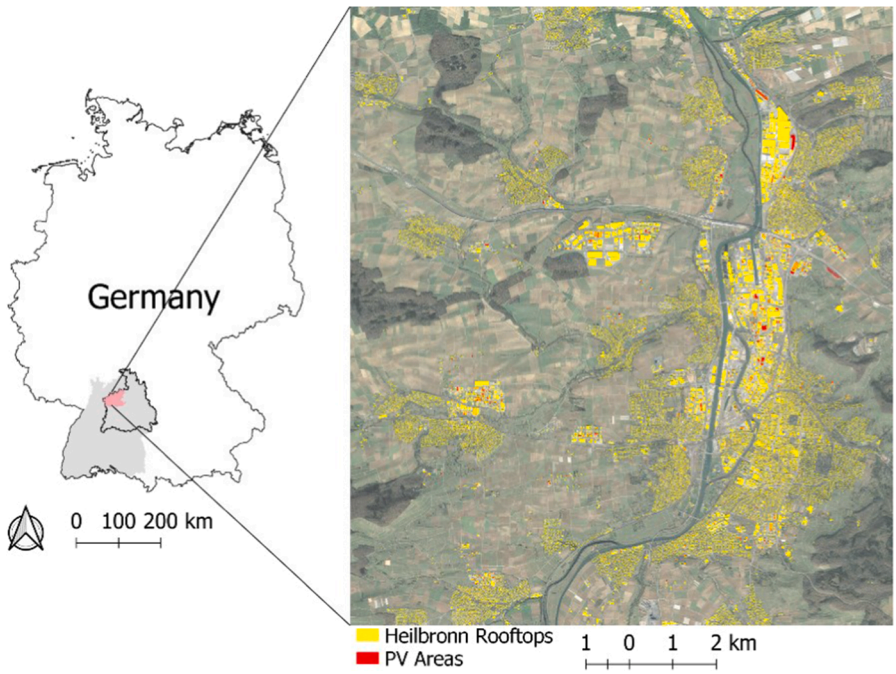


Fig. 1. Study area in Heibronn, Germany.

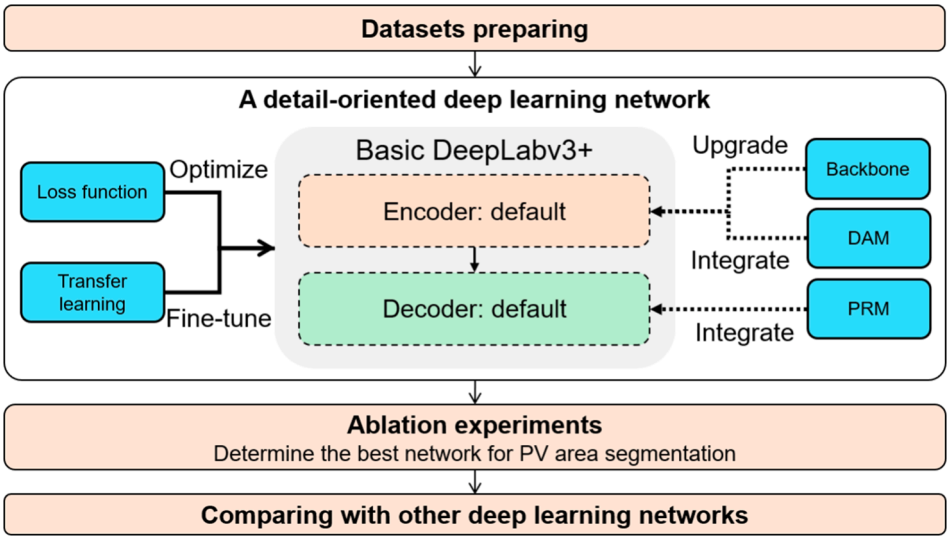


Fig. 2. Research framework.

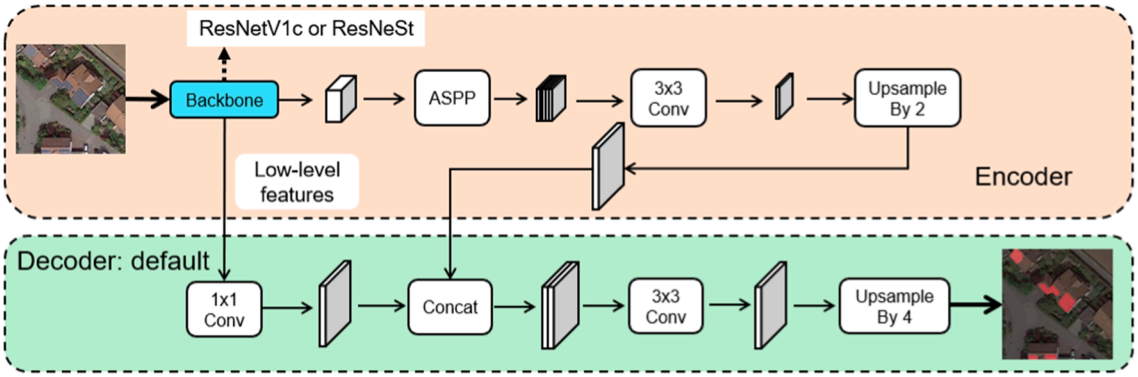


Fig. 3. The network architecture using ResNetV1c or ResNeSt as backbone.

module in Deeplabv3+, which was scaled and concatenated with the low-level feature map created by the backbone. This combined feature map was then up-sampled to provide the final forecast. Compared to the conventional ResNet (Qian et al., 2022b), ResNetV1c substitutes the 7 × 7 convolution layers in the input stem with three 3 × 3 convolutional layers to save computation and sets stride on the 3 × 3 convolutional layer in the down-sampling part to reduce feature loss (He et al., 2019; Chen et al., 2022a, b).

Furthermore, ResNeSt incorporated the Split-Attention Network (SAN) into ResNetV1c to improve model performance (Zhang et al., 2022c), and it has been demonstrated to be competitive with the plain ResNet when used as backbones in segmentation methods like DeepLabv3. In comparison to ResNetV1c, the 3 × 3 convolutional layer in the down-sampling and residual modules was replaced by SAN, and the average pooling layer was used to reshape the image instead of the convolutional layer with a stride in the down-sampling module. The SAN will not alter the input's size or the number of channels in this study. Therefore, ResNetV1c and ResNeSt were employed independently as the backbone in DeepLabv3 + to compare the performance of the two modules. It was noticed that increasing the number of layers could improve the backbone's capacity to extract features, while the inference latency was also increased. This study used 101 layers as a trade-off since 50 and 101 layers were frequently employed.

2.4. Integration of the Dual-Attention module

Considering PV areas are significantly small samples and PV segmentation can be easily affected by the complex environment in a satellite image, there is a demand for strengthening the extraction of PV characteristics. Thus, the Dual-Attention Module (DAM) was applied in

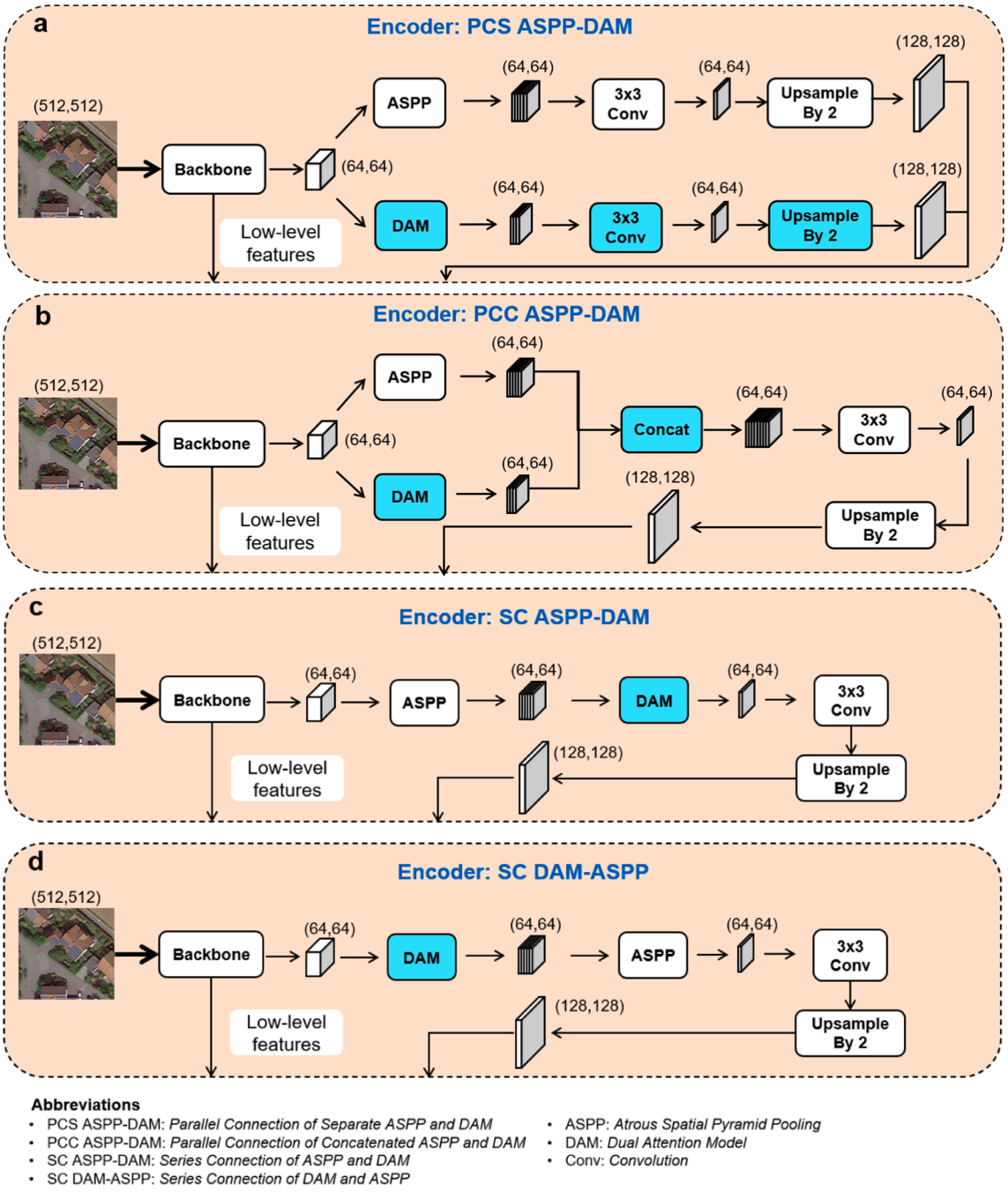


Fig. 4. Integration of DAM into Deeplabv3 +. (a) PCS ASPP-DAM; (b) PCC ASPP-DAM; (c) SC ASPP-DAM; (d) SC DAM-ASPP.

this study to adaptively combine local characteristics with their global dependencies (Lu et al., 2020). DAM comprises two types of attention modules, which respectively describe semantic interdependencies in spatial and channel dimensions. By adding DAM into the network, detailed objects can be more lucid, certain misclassified categories caused by obstructions can be accurately classified, and semantic consistency can be significantly enhanced. Therefore, it is possible to mitigate the issue caused by varying object sizes, occlusion, and lighting changes, and thus enhance segmentation results when dealing with complex and diverse scenes. This study proposed four structures to integrate DAM with ASPP, as shown in Fig. 4.

The first structure was the *Parallel Connection of Separate ASPP and DAM* (PCS ASPP-DAM), which was depicted in Fig. 4(a). In this structure, the output of Backbone was separately entered into the ASPP module and the DAM, and then the outputs of these two modules were concatenated with Backbone's low-level features. After convolution and up-sampling, this combined feature map could produce predictions.

The second structure was the *Parallel Connection of Concatenated ASPP and DAM* (PCC ASPP-DAM), as shown in Fig. 4(b). Separately, the output of Backbone was fed into the ASPP module and the DAM module before the outputs of these two modules were concatenated. The concatenated feature map was then processed by a convolution module before being concatenated with Backbone's low-level features. After convolution and up-sampling, this newly combined feature map produced predictions. There were considerable differences between the modified model with PCS ASPP-DAM and PCC ASPP-DAM, such as the number of network parameters.

The third structure was the *Series Connection of ASPP and DAM* (SC ASPP-DAM), which was presented in Fig. 4(c). In this case, the output of the Backbone was firstly entered into the ASPP module. Then, the output of the ASPP module was processed by the DAM module, which was concatenated with Backbone's low-level features. After convolution and up-sampling, the combined feature map performed predictions.

The fourth structure was the *Series Connection of DAM and ASPP* (SC DAM-ASPP), as shown in Fig. 4(d). Different from the previous structure, the output of Backbone was first entered into the DAM module, and then processed by the ASPP module. Furthermore, the output of the ASPP module was coupled with the low-level features of Backbone. This new concatenated feature map yielded predictions following convolution and up-sampling. The sole variation between the SC ASPP-DAM and SC DAM-ASPP models is the order of the ASPP and DAM components.

2.5. Integration of the PointRend module

During the up-sampling procedure, the border of the projected PV areas may be fuzzy; thus, the PointRend Module (PRM) was utilized in this work to create clear boundaries of PV areas. The PRM identified contour points of foreground objects in both the course prediction mask and the fine-grained feature map, and then refined these uncertain regions using Multi-Layer Perceptron (MLP). It has been demonstrated that DeepLabv3 with PRM outperformed DeepLabv3 for semantic segmentation on the Cityscapes datasets (Cordts et al., 2016). As shown in Fig. 5, PRM was integrated into the Decoder after Concatenation and 3 × 3 Convolution to refine the PV segmentation results.

2.6. Development of the hybrid loss functions

2.6.1. Dice loss function

The Dice loss function offers a solution to the problem that network predictions are heavily skewed toward the background since the foreground is considerably smaller than the backdrop (Qian et al., 2022a). Eq. (1) describes the Dice loss function used in this study, in which H denotes the height of an image, W denotes the weight of the image, $p_{(r,c)}$ denotes the predicted class membership probability of pixel (r,c), $g_{(r,c)}$ denotes the true classification of pixel (r,c), and ε denotes the smoothing factor that prevents the denominator from being zero.

$$L_D = 1 - \frac{2\sum_{r=1}^{H}\sum_{c=1}^{W} p_{(r,c)} g_{(r,c)} + \varepsilon}{\sum_{r=1}^{H}\sum_{c=1}^{W} [{p_{(r,c)}}^2 + {g_{(r,c)}}^2] + \varepsilon} \tag{1}$$

2.6.2. Cross-entropy loss function

The Cross-entropy loss function is commonly employed in classification and semantic segmentation tasks because the negative log function is suited for tackling inaccurate predictions (Bahri et al., 2020). Eq. (2) provides Cross-entropy loss when there are two classes to be identified, in which H, W, $p_{(r,c)}$, and $g_{(r,c)}$ denote the same meaning as used in the Dice loss function, and ω denotes the weight of foreground.

$$L_{wC} = -\sum_{(r=1,c=1)}^{(H,W)} [\omega g_{(r,c)} \log\left(p_{(r,c)}\right) + (1 - g_{(r,c)}) \log(1 - p_{(r,c)})] \tag{2}$$

2.6.3. Focal loss function

The Focal loss function is a solution to the problem that certain classes are easy to classify while others are difficult, which is also helpful in the segmentation task. Eq. (3) provides the Focal loss when the class number is equal to 2 (Lin et al., 2017). In the equation, $p_{(r,c)}$ denotes the predicted class membership probability of pixel (r,c), α is the parameter that controls the class weights, and γ is used to decrease the weight of easy-to-classify pixels. $\alpha = 0.5$ and $\gamma = 2$ in this study. When $\alpha = 1$ and $\gamma = 0$ the Focal loss is recast as Cross-entropy loss.

$$L_{focal} = \alpha(1 - p_{(r,c)})^{\gamma} \bullet L_{BCE} \tag{3}$$

2.6.4. IoU loss function

The IoU loss provides a solution to the imbalance between foreground and background in binary segmentation tasks, which directly enhances IoU. Eq. (4) defines the IoU loss function, which utilizes the same parameters as the above functions.

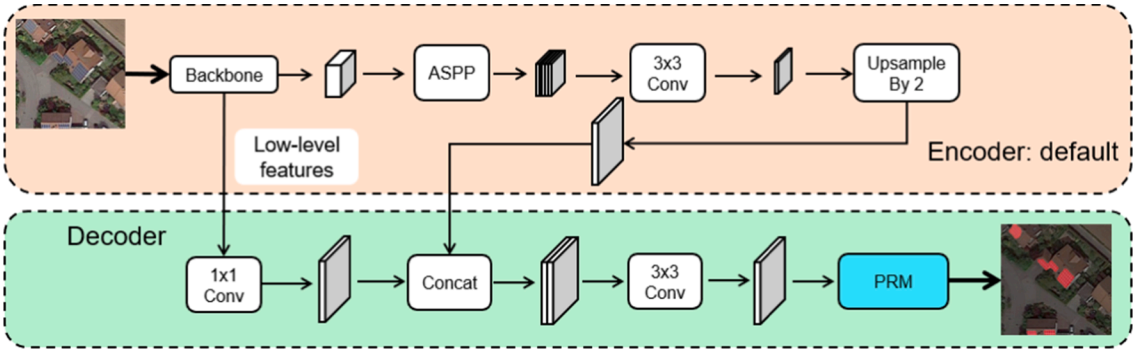


Fig. 5. Integrate PRM into Deeplabv3 +.

$$L_{IoU} = 1 - \frac{\sum_{r=1}^{H}\sum_{c=1}^{W} p_{(r,c)} g_{(r,c)}}{\sum_{r=1}^{H}\sum_{c=1}^{W} [p_{(r,c)} + g_{(r,c)} - p_{(r,c)} g_{(r,c)}]} \quad (4)$$

2.6.5. Hybrid loss functions

The Hybrid loss function (HLF) is a mixture of two or more types of loss functions and could be more efficient for satellite image object recognition and semantic segmentation tasks. The nonlinear combination of Cross-entropy loss and Dice loss has been utilized to successfully extract land cover and roads from satellite images (Wan et al., 2021; Zhang et al., 2020; Qian et al., 2022a). This study employed HLFs by incorporating Cross-entropy loss (L_C), Dice loss (L_D), IoU loss (L_{IoU}), and Focal loss (L_F). In the following equations, λ_C, λ_D, and λ_I refer to the coefficients of Cross-entropy loss, Dice loss, and IoU loss, respectively. By assigning coefficients and adjusting Cross-entropy loss, this study designs four HLFs.

$$CDI = L_C + 2L_D + L_{IoU} \quad (5)$$

$$CDhI = L_C + 2L_D + 3L_{IoU} \quad (6)$$

$$wCDI = L_{wC} + 2L_D + L_{IoU} \quad (7)$$

$$FDI = L_F + 2L_D + L_{IoU} \quad (8)$$

In CDI, $\lambda_C = 1$, $\lambda_D = 2$, $\lambda_I = 1$ are allocated, with the coefficient of Dice loss increased from 1 to 2 due to the extreme imbalance of the dataset (Eq. (5)). In CDhI, the factor of IoU loss is increased from 1 to 3 to increase the weight of foreground influence during training (Eq. (6)). In wCDI, the weighted Cross-entropy loss replaces the standard Cross-entropy loss in CDI to increase the weight of the PV class during training (Eq. (7)). Focal loss may be superior to Cross-entropy loss for addressing the imbalance issue, and it replaces Cross-entropy loss in CDI to generate FDI (Eq. (8)).

2.7. Evaluation metrics

Four indices were used to evaluate the performance of trained models, i.e., Intersection-over-Union (IoU), F1-score, Precision, and Recall. The greater value of the indices, the better the model performances. Specifically, IoU is extensively used to evaluate the performance of semantic segmentation models since it accounts for the class imbalance issue. As demonstrated by Eq. (9), IoU provides the similarity between the predicted area and the ground-truth region for an object. mIoU denotes the average IoU value of all classes, which refers to PV and background in this study. In the equation, intersection refers to the point where the ground truth and prediction area intersect, and the union is the union of the ground truth with the forecast region. Here, TP, FP, and FN denote true-positive pixels, false-positive pixels, and false-negative pixels, respectively. Meanwhile, Accuracy, Precision, Recall, and F1-score are calculated based on Eqs. (10)–(13).

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

$$Accuracy = \frac{TN + TP}{FN + TP + FP + TN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 - score = 2\frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

3. Empirical evaluation

3.1. Data preprocessing

Fig. 6 demonstrates the dataset preprocessing steps. To start with, rooftop polygons and labels of the PV areas are combined to create masks for identifying environmental backgrounds from satellite images. All satellite images and binarized labels are cropped sequentially into 512 × 512 patch images. Because of the unbalanced distribution of PV panels, many patch images do not contain PV labels, which are thus filtered out. The remaining pairs of images and labels are separated into three parts for training (3345 patches), validation (372 patches), and testing (295 patches), with a 11.34 to 1.26 and to 1.

3.2. Training strategies

Transferring weights trained by other datasets at the start of training can enhance the performance of a deep neural network compared to randomly initializing the network's parameters, which has been frequently utilized as a training approach to improve training outcomes (Zhuang et al., 2020). Thus, transfer learning is employed in this study. The weights of backbones pre-trained using Cityscapes datasets are imported into the models at the beginning of training (Cordts et al., 2016).

3.3. Experiment environments

Experiments were conducted using MMSegmentation (MMS 2022), an open-source semantic segmentation toolkit based on PyTorch. Details of the experiment and the hyperparameter settings are given in Table 1.

4. Results

4.1. Selection of the backbone

DeepLabv3 + using ResNetV1c and ResNeSt as the backbone was trained and validated independently for 100 epochs, and the experiments were repeated for three times. Based on the trained models, the prediction accuracy metrics based on the test dataset are summarized in Table 2. It shows that ResNeSt obtained slightly higher accuracy than ResNetV1c regarding the means of IoU, Accuracy, F1-score, and Recall for PV segmentation. Thus, ResNeSt is employed in the following experiments as getting a higher PV segmentation accuracy is the main objective of this study.

4.2. Integration of the DAM or PRM

Then, the Dual Attention Module (DAM) was integrated into DeepLabv3 + using ResNeSt, which resulted in the production of four different models with various connections between ASPP and DAM. By repeating the same training and validation experiments for three times, this study also obtained robustly trained models. The results of the four models are presented in Table 3. All the models containing DAM obtain increased PV mean IoU, F1-score, and Precision compared to the benchmark using ResNetV1c without DAM (Table 2). In detail, the modified model structured with SC DAM-ASPP shows better performance than the other three ones, with PV mean IoU equaling 80.68 % for PV segmentation. Fig. 7 compares PV segmentation results obtained from different models, which demonstrates that integrating DAM into Deeplabv3 + enhances feature extraction of PV modules even PV area is significantly small and there are strong obstacles from surrounding environments.

Table 3 and Fig. 7 also present the PV segmentation metrics by using ResNeSt integrated with PRM. It shows that the mean and max IoU are 81.59 % and 82.28 %, respectively, which obtains a 1.91 % and 2.11 % increase compared to the benchmark using ResNetV1c. This means that

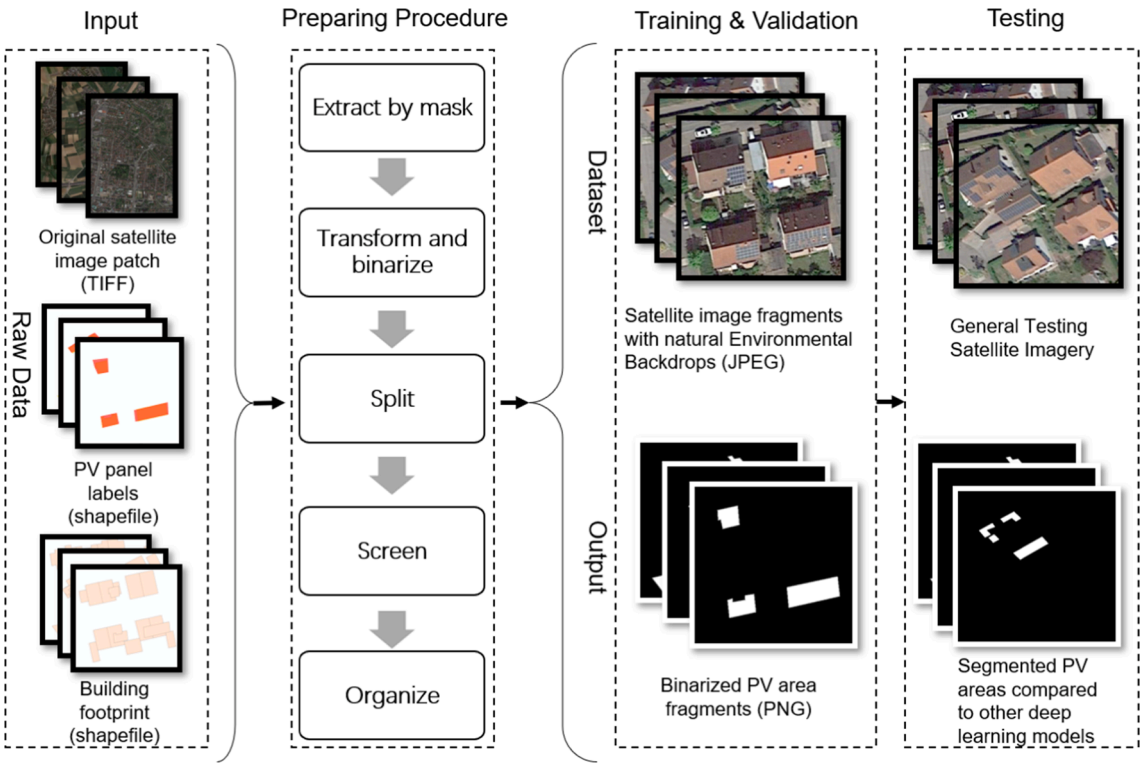


Fig. 6. Procedure for creating datasets.

Table 1
Details of the experimental environment setting.

Category	Item	Configuration
Hardware	Cloud Platform	AutoDL
	GPU	RTX 3090 × 2
	CPU	Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz
Environment	PyTorch	1.9.0
	Python	3.8
	Cuda	11.1
	MMCV	1.3.9
	MMSegmentation	0.15.0+
Hyperparameters	Batch size	4
	Optimizer	AdamW
	Learning rate	0.01
	Weight decay	0.001
	Learning rate scheduler	Polynomial (power = 0.9, minimum lr = 0.0001)
	Loss function	Dice loss (when hybrid loss is undefined)

Table 2
Evaluation of segmented PV areas using different backbones (%).

Backbone	Category	IoU	Accuracy	F1-score	Precision	Recall
ResNetV1c	Mean	79.68	87.91	88.69	89.49	87.91
	Max	80.17	88.65	88.99	89.95	88.65
ResNeSt	Mean	79.72	88.40	88.71	89.03	88.40
	Max	80.17	88.91	88.99	90.16	88.91

there is a notable improvement in PV segmentation when PV areas are significantly small and imbalanced to negative ones. For instance, small PV areas that are easily influenced by rooftops with similar textures are successfully identified and extracted from images by the model having PRM (Fig. 7). Moreover, Table 3 reveals that the model integrating PRM obtains better performance than that containing DAM in terms of PV the mean IoU, F1-score, and Precision. Considering that DAM is used to

Table 3
Evaluation of the modified models respectively with DAM and PRM (%).

Structure	Category	IoU	Accuracy	F1-score	Precision	Recall
PCS ASPP-DAM	Mean	79.71	87.82	88.71	89.63	87.82
	Max	79.83	88.13	88.79	90.14	88.13
PCC ASPP-DAM	Mean	80.33	87.70	89.09	90.53	87.80
	Max	**81.52**	88.38	89.82	91.31	88.38
SC ASPP-DAM	Mean	80.57	88.61	89.24	89.87	88.61
	Max	81.41	89.01	89.75	90.51	89.01
SC DAM-ASPP	Mean	**80.68**	88.89	89.31	89.72	88.89
	Max	81.42	89.21	89.76	90.32	89.21
PRM	Mean	81.59	87.76	89.86	92.07	87.76
	Max	82.28	88.68	90.28	92.82	88.68

exploit the context while PRM is used to acquire crisp edges of foregrounds, the results imply that improving the prediction of PV boundaries might play a more important role in enhancing the performance of PV segmentation.

4.3. Integration of DAM and PRM

The above experiments illustrate that respectively integrating DAM or PRM into Deeplabv3 + with the backbone of ResNeSt can obtain increased accuracy on PV segmentation compared to the benchmark. In this section, DAM and PRM will be integrated into DeepLabv3 + simultaneously for further investigation. By repeating the experiments for three times, it shows that all four networks containing DAM and PRM can compete with the benchmark (Table 2) and the networks only integrating DAM (Table 4) based on the mean and max IoU for segmented PV areas.

Specifically, the network utilizing ResNeSt and incorporating PCC ASPP-DAM & PRM performs the best with the mean and max IoU equaling 82.87 % and 83.48 %, which also outperformed the network with PRM alone (Table 4). Also, both the mean and max F1-score are larger than 90 % for PCC ASPP-DAM & PRM, which suggests a

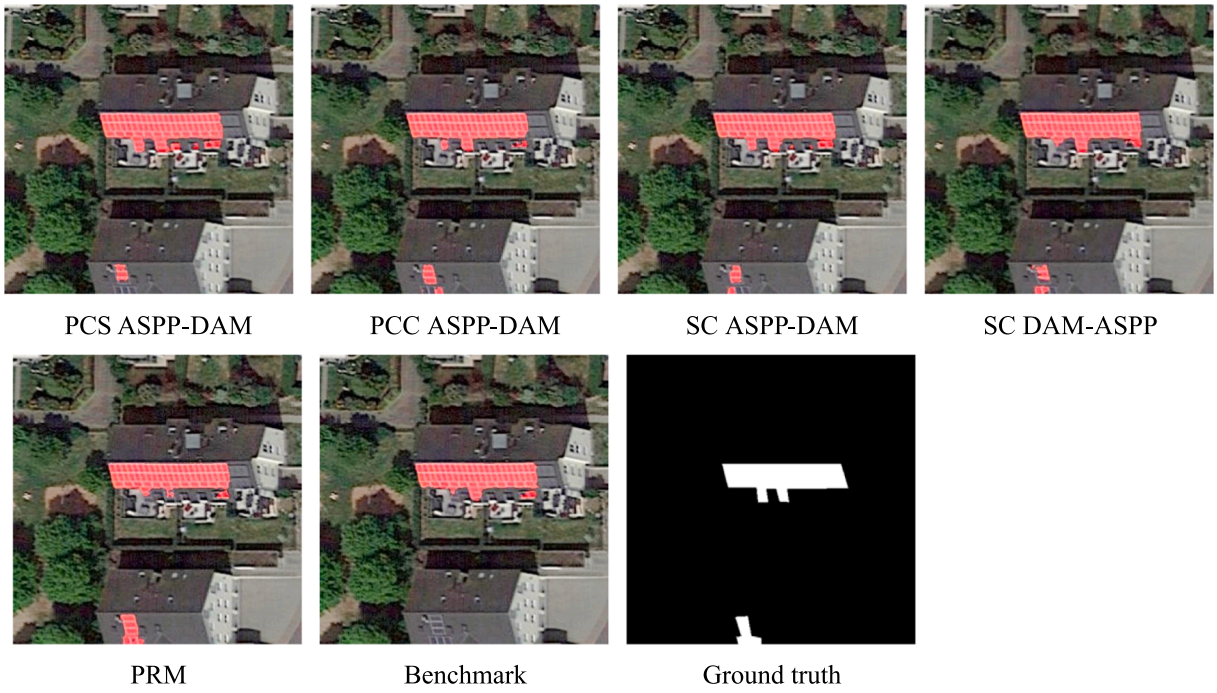


Fig. 7. Segmented PV areas in the red colour obtained from different models.

Table 4
Evaluation of the models with the integration of both DAM and PRM (%).

Structure	Category	IoU	Accuracy	F1-score	Precision	Recall
PCS ASPP-DAM & PRM	Mean	81.58	88.74	89.85	91.01	88.74
	Max	82.52	90.36	90.42	92.05	90.36
PCC ASPP-DAM & PRM	Mean	**82.87**	89.40	90.63	91.90	89.40
	Max	**83.48**	89.65	90.99	92.44	89.65
SC ASPP-DAM & PRM	Mean	82.49	89.30	90.40	91.56	89.30
	Max	82.94	89.85	90.68	93.14	89.85
SC DAM-ASPP & PRM	Mean	81.44	88.60	89.77	90.97	88.60
	Max	82.11	89.02	90.18	91.62	89.02

satisfactory balance between Precision and Recall and indicates highly accurate PV segmentation when PV area and background objects have an uneven class distribution. This means that the most effective network structure, as shown in Fig. 8, is derived at this stage based on the default Dice loss function. It is also noticeable that SC ASPP-DAM & PRM, a different type of connection between ASPP and DAM, also get the mean and max IoU larger than 82 % and F1-score larger than 90 %, suggesting an improved PV segmentation.

4.4. Integration of hybrid loss functions

The previous section demonstrates that PCC ASPP-DAM & PRM aligns optimally with the Dice loss function. Therefore, this structure is used to explore whether PV segmentation can be further enhanced by applying hybrid loss functions. By integrating a hybrid loss function into the deep learning network composed by the PCC ASPP-DAM & PRM structure, this study repeated the training and validation experiments over 100 epochs for three times. Results with the proposed hybrid loss functions are presented in Table 5. It shows that the means of IoU are 82.36 %, 82.81 %, 82.12 %, and 82.61 % for CDl, FDl, wCDl, and CDhl, respectively, all of which are slightly lower than that integrates the Dice loss function having the mean IoU equaling 82.87 %. This means that the Dice loss function is effective to solve the prediction challenges when positive samples are significantly smaller than the background samples.

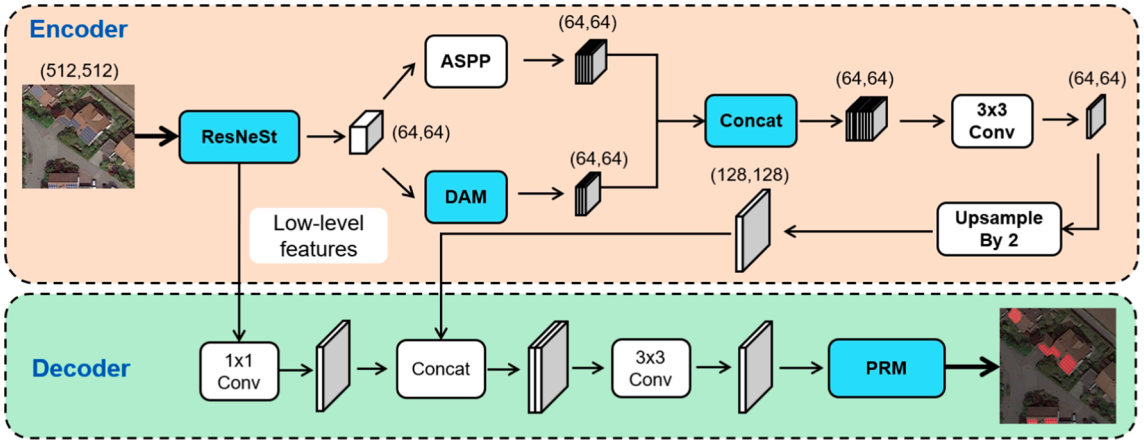


Fig. 8. The architecture of the optimal network.

Table 5
Evaluation of PCC ASPP-DAM & PRM integrating four HLFs (%).

HLFs	Category	IoU	Accuracy	F1-score	Precision	Recall
CDl	Mean	82.36	88.75	90.32	91.96	88.75
	Max	82.92	89.15	90.66	93.23	89.15
CDhl	Mean	82.61	88.32	90.47	92.75	88.32
	Max	**83.62**	89.83	91.08	93.27	89.83
wCDl	Mean	82.12	88.43	90.18	92.00	88.43
	Max	83.34	89.36	90.91	93.63	89.36
FDl	Mean	**82.81**	89.15	90.59	92.09	89.15
	Max	83.29	90.08	90.88	92.80	90.08

It is also found that PCC ASPP-DAM & PRM with CDhl obtained the max IoU equaling 83.62 % out of the three experiments (Table 5), which got the best PV segmentation performance compared to all other networks either with- or without hybrid loss functions. Moreover, the network has Accuracy, F1-score, Precision, and Recall equaling 89.83 %, 91.08 %, 93.27 %, and 89.83 %, respectively (Table 5), all of which are higher than the same deep learning structure using the Dice loss function (Table 4). The results suggest that CDhl can help get the largest IoU and has a high capability to further enhance the PV segmentation accuracy, although the mean IoU is slightly lower. This is because CDhl assigning increased weights on Cross-entropy loss, Dice loss, and IoU loss functions cannot only pay the most important attention on getting a large IoU but also tackle inaccurate predictions along the PV boundary areas, using a negative log function that identifies significantly imbalanced classes with improved PV area segmentation.

4.5. Comparison with the state-of-the-art deep learning networks

The latest deep learning methods have been proposed for the semantic segmentation task, such as DNLNet (Yin et al., 2020), UPerNet (Huang et al., 2020), U-Net (Bradbury et al., 2016), DMNet (Li et al., 2020), and PSPNet (Zhao et al., 2017). The results for testing the proposed detail-oriented network and comparing with these networks using a different dataset, as introduced in Sections 2.1 and 3.1, are presented in Table 6. By repeating the experiments for three times, these networks achieve the PV mean IoU at 72.16 %, 78.28 %, 80.20 %, 82.65 %, and 82.71 %, respectively. In comparison, our network (ResNeSt & PCC ASPP-DAM & PRM & Dice) has the PV mean IoU equaling 83.34 %, getting IoU increased by 11.18 %, 5.06 %, 3.14 %, 0.69 %, and 0.63 % and F1-score increased by 7.08 %, 3.09 %, 1.90 %, 0.41 %, and 0.37 %, respectively. Compared to the basic DeepLabv3+ (ResNetV1c & ASPP & Dice), our network increases PV IoU by 4.55 %, Accuracy by 2.30 %, F1-score by 2.71 %, Precision by 3.19 %, and Recall by 2.30 %.

Fig. 9 visualizes the segmented PV areas obtained from the proposed detail-oriented network and the labelled PV areas used as the ground truth. Compared to previously developed models, it shows that our model can (i) sensitively identify bright PV areas caused by solar reflection (in the first and second columns); (ii) successfully exclude negative and confusable pixels and segment the PV area with a regular boundary, most similar to the labelled PV area (in the third column); (iii) exclude fibreglass roof that is pretty similar to but essentially not PV areas (in the fourth column); (iv) successfully detect all small and distributed PV modules (in the fifth column); (v) segment all PV arrays with clear boundaries (in the sixth column).

Table 6
Segmented PV area metrics using different networks (%).

Models	IoU	Accuracy	F1-score	Precision	Recall
DNLNet	72.16	82.15	83.83	85.58	82.15
UPerNet	78.28	82.76	87.82	93.54	82.76
U-Net	80.20	85.76	89.01	92.53	85.76
DMNet	82.65	88.12	90.50	92.95	88.18
PSPNet	82.71	89.53	90.54	91.57	89.53
Deeplabv3+	78.79	85.85	88.20	90.67	85.85
Detail-oriented network	**83.34**	88.15	90.91	93.86	88.15

Lastly, to investigate the generalization capability of the proposed network, this study utilized another independent PV dataset for testing. The dataset has been published by Jiang et al. (2021), which contains a set of satellite images at the 0.1-m resolution, pretty like ours at 0.15-m, and the PV areas were well-labelled in Jiangsu province, China. The results show that our network achieves an overall IoU equaling 85.30 % (Table 7), higher than others by 0.18–6.99 % except for U-Net. This suggests that our network has a favourable generalization ability. It is also noted that the IoU is even higher than that obtained from our dataset, implying a high quality of the dataset and few confusable geo-objects that leads to an easy segmentation.

5. Discussion and conclusion

This study proposes a detail-oriented semantic segmentation network to accurately segment PV areas from satellite images. By coupling with transfer learning, a synthetic strategy that systematically refines a deep learning network structure and optimizes the loss function, and a series of well-designed ablation experiments, the optimal network is obtained that exceeds DeepLabv3 + in terms of PV mean IoU, Accuracy, F1-score, Precision, and Recall by 5 %, 2 %, 3 %, 3 %, and 2 %, respectively. The developed network remains competitive compared to the most advanced semantic segmentation models developed in recent years. These improvements mainly come from the successful identification of hard and positive PV samples, which demonstrates an important contribution to accurately segmenting small and distributed PV areas.

Three important findings are revealed from this study. First, it is useful to include a Split-Attention Network to leverage the success in capturing cross-feature interactions and learning diverse representations and integrate a Dual-Attention Network to enhance the spatial and channel attention. This is especially important for successfully extracting small PV areas that are influenced by surrounding geo-objects with similar feature characteristics. Second, integrating the PointRend Model into Decoder is useful to improve the prediction accuracy by accurately identifying positive and negative samples along the PV boundaries with the help of Multi-Layer Perceptron. Third, PCC ASPP-DAM is outperformed when using the Dice loss function, and it can also obtain the highest IoU when the hybrid loss function CDhl is used. It is because PCC ASPP-DAM maintains the features of ASPP and DAM well when they are processed by Concatenated convolution and up-sampling, which allows effective training with the combination of Cross-entropy loss, Dice loss, and IoU loss functions.

This study is significant in three aspects. First, this study is innovative in accurate segmenting hard and small PV areas from images. The significant improvement in IoU suggests that the new network has a greater number of correctly classified positive samples than the original. Second, this study is vital for evaluating the economic and environmental impacts made by solar farming. With the increasing penetration of distributed PV systems, it is important to estimate the total amount of electricity generation and the consequent carbon reduction made by PV modules at a large geographical scale. Our network can help to achieve this by providing accurate statistics of the total PV installed capacity. Third, our trained model can be used for transfer learning, and the proposed model is effective for PV segmentation with an adaption of other satellite images with various spatial resolutions. Meanwhile, it is noted that the extracted rooftop PV size may deviate from the actual one because the tilted PV panels are projected onto the orthographical satellite images. Nevertheless, the accuracy of the model will not be affected by this constraint as PV labels are manually created based on the same orthographical images.

To conclude, this study proposes Deep Roof Refiner, an end-to-end and detail-oriented deep learning network, for accurate segmentation of PV areas from satellite imagery. The new network contributes to studies related to remote sensing image segmentation, deep learning

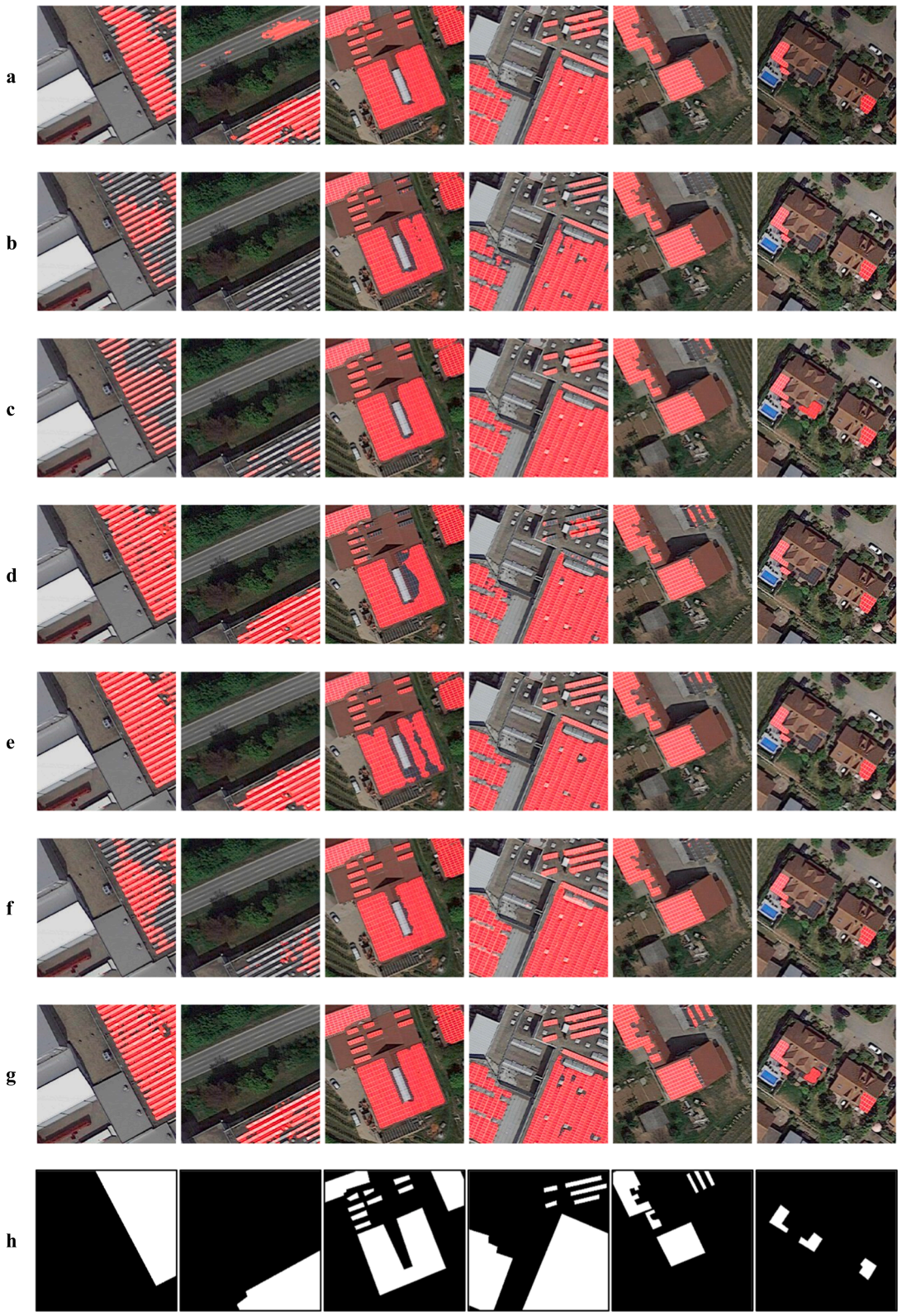


Fig. 9. Comparison of PV segmentation. (a) DNLNet; (b) UPerNet; (c) U-Net; (d) DMNet; (e) PSPNet; (f) DeepLabv3+; (g) Detail-oriented network; (h) Ground truth.

refinement, and PV installed capacity estimation, which is also deliverable for other geo-object segmentation. Future work may predict the inclination and orientation of PV modules based on deep learning to estimate PV electricity generation.

CRediT authorship contribution statement

Rui Zhu: Conceptualization, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. **Dongxue Guo:** Formal analysis, Software, Validation, Visualization, Writing – original draft. **Man Sing Wong:** Supervision,

Table 7
Segmented PV area metrics using different networks and another dataset (%).

Models	IoU	Accuracy	F1-score	Precision	Recall
DNLNet	83.37	96.36	90.93	86.08	96.36
UPerNet	85.12	93.09	91.96	90.86	93.09
U-Net	87.60	95.84	93.39	91.06	95.84
DMNet	79.66	88.16	88.68	89.2	88.16
PSPNet	78.31	88.43	87.83	87.25	88.43
Deeplabv3+	81.53	96.82	89.82	83.77	96.82
Detail-oriented network	**85.30**	94.77	92.06	89.51	94.77

Writing – review & editing, Funding acquisition. **Zhen Qian:** Methodology, Supervision, Writing – review & editing. **Min Chen:** Supervision, Writing – review & editing. **Bisheng Yang:** Supervision, Writing – review & editing. **Biyu Chen:** Supervision, Writing – review & editing. **Haoran Zhang:** Supervision, Writing – review & editing. **Linlin You:** . **Joon Heo:** Supervision, Writing – review & editing. **Jinyue Yan:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Rui Zhu thanks the funding support from the Strategic Hiring Scheme (Grant No. P0036221) and the Projects of RILS (Grant No. P0039240) at the Hong Kong Polytechnic University. Man Sing Wong thanks the funding support from the General Research Fund (Grant No. 15602619 and 15603920), and the Collaborative Research Fund (Grant No. C5062-21GF) from the Research Grants Council, Hong Kong, China.

References

Bahri, A., Majelan, S.G., Mohammadi, S., Noori, M., Mohammadi, K., 2020. Generalized cross entropy loss for training deep neural networks with noisy labels. IEEE Geosci. Remote Sens. Lett. 17 (6), 1087–1091.

Bradbury, K., Saboo, R., Johnson, T.L., Malof, J.M., Devarajan, A., Zhang, W., Collins, L. M., Newell, R.G., 2016. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. Sci. Data 3, 160106.

Chapman, A.J., McLellan, B., Tezuka, T., 2016. Residential solar PV policy: An analysis of impacts, successes and failures in the Australian case. Renew. Energy 86, 1265–1279.

Chen, Z., Kang, Y., Sun, Z., Wu, F., Zhang, Q., 2022a. Extraction of Photovoltaic Plants Using Machine Learning Methods: A Case Study of the Pilot Energy City of Golmud, China. Remote Sens. 14, 2697.

Chen, J., Shao, S., Zhu, Y., Wang, Y., Rao, F., Dai, X., Lai, D., 2022b. Enhanced Automatic Identification of Urban Community Green Space Based on Semantic Segmentation. Land 11, 905.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, 3213-3223.

Costa, M.V.C.V.d., Carvalho, O.L.F.d., Orlandi, A.G., Hirata, I., Albuquerque, A.O.d, Silva, F.V.e., Guimarães, R.F., Gomes, R.A.T., Júnior, O.A.d.C., 2021. Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. Energies 14, 2960.

Das, S., Fime, A.A., Siddique, N., Hashem, M.M.A., 2021. Estimation of Road Boundary for Intelligent Vehicles Based on DeepLabV3+ Architecture. IEEE Access 9, 121060–121075.

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M., 2019. Bag of Tricks for Image Classification with Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16-20 June 2019, 558-567.

Huang, J., Liu, G., Wang, B., Li, Y., Yang, L., 2020. Semantic Segmentation under a Complex Background for Machine Vision Detection Based on Modified UPerNet with Component Analysis Modules. Math. Probl. Eng. 2020, 6903130.

Inderberg, T.H.J., Tews, K., Turner, B., 2018. Is there a prosumer pathway? Exploring household solar energy development in Germany, Norway, and the United Kingdom. Energy Res. Soc. Sci. 42, 258–269.

Jiang, H., Yao, L., Lu, N., Qin, J., Liu, T., Liu, Y., Zhou, C., 2021. Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery. Earth Syst. Sci. Data 13, 5389–5401.

Joshi, S., Mittal, S., Holloway, P., Shukla, P.R., Gallachóir, B., Glynn, J., 2021. High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation. Nat. Commun. 12, 5738.

Kruitwagen, L., Story, K.T., Friedrich, J., Byers, L., Skillman, S., Hepburn, C., 2021. A global inventory of photovoltaic solar energy generating units. Nature 598, 604–610.

Li, Q., Mou, L., Hua, Y., Shi, Y., Zhu, X.X., 2022. CrossGeoNet: A Framework for Building Footprint Generation of Label-Scarce Geographical Regions. Int. J. Appl. Earth Obs. Geoinf. 111, 102824.

Li, W., Zhang, X., Peng, Y., Dong, M., 2020. DMNet: A Network Architecture Using Dilated Convolution and Multiscale Mechanisms for Spatiotemporal Fusion of Remote Sensing Images. IEEE Sens. J. 20 (20), 12190–12202.

Li, P., Zhang, H., Guo, Z., Lyu, S., Chen, J., Li, W., Song, X., Shibasaki, R., Yan, J., 2021. Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. Adv. Appl. Energy 4, 100057.

Liao, X., Zhu, R., Wong, M.S., 2022. Simplified estimation modeling of land surface solar irradiation: A comparative study in Australia and China. Sustain. Energy Technol. Assess. 52, 102323.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. Proc. International Conference on Computer Vision (ICCV), 2999-3007.

Lin, S., Zhang, C., Ding, L., Zhang, J., Liu, X., Chen, G., Wang, S., Chai, J., 2022. Accurate Recognition of Building Rooftops and Assessment of Long-Term Carbon Emission Reduction from Rooftop Solar Photovoltaic Systems Fusing GF-2 and Multi-Source Data. Remote Sens. (Basel) 14, 3144.

Liu, T., Yao, L., Qin, J., Lu, N., Jiang, H., Zhang, F., Zhou, C., 2022. Multi-scale attention integrated hierarchical networks for high-resolution building footprint extraction. Int. J. Appl. Earth Obs. Geoinf. 109, 102768.

Lu, R., Wang, N., Zhang, Y., Lin, Y., Wu, W., Shi, Z., 2020. Extraction of Agricultural Fields via DASFNet with Dual Attention Mechanism and Multi-scale Feature Fusion in South Xinjiang, China. Remote Sens. 12, 582.

Lukač, N., Seme, S., Dežan, K., Žalik, B., Štumberger, G., 2016. Economic and environmental assessment of rooftops regarding suitability for photovoltaic systems installation based on remote sensing data. Energy 107, 854–865.

McGlade, C., Ekins, P., 2015. The geographical distribution of fossil fuels unused whenlimiting global warming to 26 °C. Nature 517 (7533), 187–190.

MMS. Welcome to MMSegmentation's documentation. https://mmsegmentation.readthedocs.io/en/latest/ (assessed August 10, 2022).

Qian, Z., Chen, M., Zhong, T., Zhang, F., Zhu, R., Zhang, Z., Zhang, K., Sun, Z., Lü, G., 2022a. Deep Roof Refiner: A detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. Int. J. Appl. Earth Obs. Geoinf. 107, 102680.

Qian, Z., Chen, M., Yang, Y., Zhong, T., Zhang, F., Zhu, R., Zhang, K., Zhang, Z., Sun, Z., Ma, P., Lü, G., Ye, Y., Yan, J., 2022b. Vectorized dataset of roadside noise barriers in China. Earth Syst. Sci. Data 14, 4057–4076.

Qiu, T., Liang, X., Du, Q., Ren, F., Lu, P., Wu, C., 2021. Techniques for the Automatic Detection and Hiding of Sensitive Targets in Emergency Mapping Based on Remote Sensing Data. ISPRS Int. J. Geo Inf. 10 (2), 68.

Sabadini, F., Madlener, R., 2021. The economic potential of grid defection of energy prosumer households in Germany Author links open overlay panel. Adv. Appl. Energy 4, 100075.

Vries, T.N.C., Bronkhorst, J., Vermeer, M., Donker, J.C.B., Brielsc, S.A., Ziara, H., Zemana, M., Isabella, O., 2020. A quick-scan method to assess photovoltaic rooftop potential based on aerial imagery and LiDAR. Sol. Energy 209, 96–107.

Wan, J., Xie, Z., Xu, Y., Chen, S., Qiu, Q., 2021. DA-RoadNet: A dual-attention network for road extraction from high resolution satellite imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 6302–6315.

Wang, M., Cui, Q., Sun, Y., Wang, Q., 2018. Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching. ISPRS J. Photogramm. Remote Sens. 141, 100–111.

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. Renew. Energy 99, 325–335.

Wong, M.S., Zhu, R., Kwok, Y.T., Kwan, M.P., Santi, P., Lee, K.H., Heo, J., Li, H., Ratti, C., 2021. Association between NO_2 concentrations and spatial configuration: A study of the impacts of COVID-19 lockdowns in 54 US cities. Environ. Res. Lett. 16, 054064.

Yin, M., Yao, Z., Cao, Y., Li, X., Zhang, Z., Lin, S., Hu, H., 2020. Disentangled Non-local Neural Networks. Computer Vision – ECCV 2020 12360, 191-207.

Yu, J., Wang, Z., Majumdar, A., Rajagopal, R., 2018. DeepSolar: a machine learning framework to efficiently construct a solar deployment database in the United States. Joule 2 (12), 2605–2617.

Yuan, J., 2018. Learning Building Extraction in Aerial Scenes with Convolutional Networks. IEEE Trans. Pattern Anal. Mach. Intell. 40 (11), 2793–2798.

Zhang, C., Chen, X., Ji, S., 2022a. Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks. Int. J. Appl. Earth Obs. Geoinf. 112, 102885.

Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., Zhu, R., Zhang, F., Zhang, H., Zhou, F., Yu, J., Zhang, B., Lü, G., Yan, J., 2022b. Vectorized rooftop area data for 90 cities in China. Sci. Data 9, 66.

Zhang, H., Yan, J., Yu, Q., Obersteiner, M., Li, W., Chen, J., Zhang, Q., Jiang, M., Fredrik Wallin, F., Song, X., Wu, J., Wang, X., Shibasaki, R., 2021. 1.6 Million transactions replicate distributed PV market slowdown by COVID-19 lockdown. Appl. Energy 283, 116341.

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., Smola, A., 2022. ResNeSt: Split-Attention Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2736-2746.

Zhang, Y., Zhang, S., Li, Y., Zhang, Y., 2020. Coarse-to-fine satellite images change detection framework via boundary-aware attentive network. Sensors 20 (23), 6735.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2881–2890.
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. Appl. Energy 298, 117132.
Zhu, R., Wong, M.S., Guilbert, E., Chan, P.W., 2017. Understanding heat patterns produced by vehicular flows in urban areas. Sci. Rep. 7, 16309.
Zhu, R., You, L., Santi, P., Wong, M.S., Ratti, C., 2019. Solar accessibility in developing cities: A case study in Kowloon East. Hong Kong. Sustainable Cities and Society 51, 101738.
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. Renew. Energy 153, 1111–1126.
Zhu, R., Kondor, D., Cheng, C., Zhang, X., Santi, P., Wong, M.S., Ratti, C., 2022a. Solar photovoltaic generation for charging shared electric scooters. Appl. Energy 313, 118728.
Zhu, R., Wong, M.S., Kwan, M.P., Chen, M., Santi, P., Ratti, C., 2022b. An economically feasible optimization of photovoltaic provision using real electricity demand: A case study in New York city. Sustain. Cities Soc. 78, 103614.
Zhu, R., Cheng, C., Santi, P., Chen, M., Zhang, X., Mazzarello, M., Ratti, C., 2022c. Optimization of photovoltaic provision in a three-dimensional city using real-time electricity demand. Appl. Energy 316, 119042.
Zhuang, L., Zhang, Z., Wang, L., 2020. The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. Appl. Soft Comput. 92, 106283.