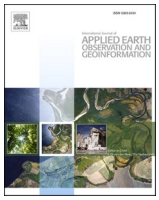
International Journal of Applied Earth Observations and Geoinformation 107 (2022) 102680

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Deep Roof Refiner: A detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery

Zhen Qian [a,b,c], Min Chen [a,b,c,d,*], Teng Zhong [a,b,c], Fan Zhang [e], Rui Zhu [f], Zhixin Zhang [g], Kai Zhang [a,b,c], Zhuo Sun [a,b,c], Guonian Lü [a,b,c]

[a] Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China
[b] State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing 210023, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[d] Jiangsu Provincial Key Laboratory for NSLSCS, School of Mathematical Science, Nanjing Normal University, Nanjing 210023, China
[e] Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[f] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[g] College of Geography & Marine, Nanjing University, Nanjing, PO Box 2100913, China

ARTICLE INFO

Keywords:
Deep learning
Roof Structure Lines
Satellite Imagery
Fine-grained Geospatial Data

ABSTRACT

Urban research is progressively moving towards fine-grained simulation and requires more granular and accurate geospatial data. In comparison to building footprints, roof structure lines (RSLs) are finer-grained elements of building roofs that provide a more sophisticated data reference. However, generating high-quality and up-to-date RSLs is arduous owing to the high expense of data sources (e.g., digital surface models and light detection and ranging data) and the low robustness of conventional image processing approaches. While the current combination of high-resolution satellite imagery and deep learning methods enables the automatic generation of RSLs, it also introduces two distinct challenges. First, the high diversity of roof sizes, forms, and spatial distribution complicates the extraction of essential RSL features from satellite imagery using general deep learning methods. Second, the significant class imbalance issue between foreground objects (i.e., RSLs) and background context in satellite imagery makes it difficult for deep learning methods to concentrate on RSL locations. To overcome these challenges and effectively delineate RSLs from satellite imagery, this study designs Deep Roof Refiner—an end-to-end and detail-oriented deep learning network and proposes a synthetic strategy to enhance the network's performance. The effectiveness of the proposed network is verified by quantitative and qualitative experiments, with the optimal dataset scale F1-score and optimal image scale F1-score of 60.89% and 63.48%, respectively. The proposed network significantly outperforms state-of-the-art deep learning methods and associated conventional research. The results indicate that the delineated RSLs can serve as a reliable data source for some urban building-based studies.

1. Introduction

With the rapid development of urbanization in recent years, many studies have focused on urban sensing and simulation, including urban traffic dynamics (Zhu et al., 2017), landscape layout evolution (Yang et al., 2019), and air quality variation (Zhang et al., 2017). The above research establishes high requirements for high-quality and up-to-date geospatial data of cities (Li et al., 2020). As a critical piece of urban geospatial data, building footprints serve as a reliable reference for a variety of urban building-based studies, such as urban layout mapping (Nouvel et al., 2017), solar energy estimation (Zhu et al., 2020), and disaster management (Cheng et al., 2021). To efficiently obtain building footprints, extensive research has been conducted over the last decade using various data sources, such as digital surface models (DSMs), light detection and ranging (LiDAR) data, and satellite imagery (Demir, 2018; Tian et al., 2017), as well as cutting-edge methods, such as machine learning and deep learning (Deng et al., 2021). Furthermore, the governments of most large cities have produced basic building footprint databases, making it easy to access high-quality building footprint data (Guo et al., 2021). However, the available building footprint data could not meet the needs of some detail-oriented urban studies (Lü et al., 2018). In-depth analyses using fine-grained spatial data of cities are

* Corresponding author at: School of Geography, Nanjing Normal University, NO.1, Wenyuan Road, Qixia District, Nanjing 210023, China.
E-mail address: chenmin0902@njnu.edu.cn (M. Chen).

https://doi.org/10.1016/j.jag.2022.102680
Received 8 December 2021; Received in revised form 2 January 2022; Accepted 8 January 2022
Available online 15 January 2022

needed to address the high-precision needs for urban sensing, modeling, and simulation, especially in the context of the complexity and dynamics of smart cities (Mainzer et al., 2017).

Roof structure lines (RSLs) are fine-grained elements of building roofs, which consist of ridge lines, valley lines, hip lines, and eave lines (Alidoost et al., 2020), as shown in Fig. 1. To obtain high-quality RSL data, some research has been conducted from both 2D and 3D viewpoints. From the 2D-viewpoint, RSLs show polyline structure with topology rules on the observed plane (Mainzer et al., 2017; Rau and Lin, 2011), as shown in Fig. 1(a). High-resolution satellite imagery is a reliable data source to delineate RSLs from the 2D plane (Alidoost et al., 2020). Based on the morphology of RSLs, building roof architecture can be efficiently classified into different types (e.g., flat, gable and hip) (Mohajeri et al., 2018). From the 3D-viewpoint, the RSLs represent lines contacting or intersecting with each other in 3D space due to changes in roof surface slope or aspect (Zhang et al., 2014), as shown in Fig. 1(b). At present, LiDAR and DSM are the primary data sources to build a 3D RSLs (Demir, 2018). Based on the spatial stereoscopic characteristics of these 3D spatial data, RSLs can be built reliably and used as a medium to reconstruct 3D models of buildings (Cao et al., 2017). However, the acquisition cost for these 3D spatial data is almost unaffordable when applying to a large area (Zhong et al., 2021).

This study will focus on delineating RSLs from 2D-viewpoint using high-resolution satellite imagery. The RSLs show the linear edge patterns because of the nonuniform illumination when projected onto images from satellite photographic surveying. The edge pattern of RSLs from satellite imagery can be detected by many conventional image processing approaches, such as the Sobel operator, Canny operator, and Laplacian filter (Mainzer et al., 2017). However, due to the limited capabilities for feature extraction, these approaches are only applicable to simple roof types, such as gables and hips, and have low robustness for generalization (Dal Poz and Fernandes, 2016). Under the trend of data sharing, an increasing number of community-based organizations or companies, such as Google Earth and Baidu Map, provide open-access and high-resolution satellite imagery data (Niu et al., 2020; Zhong et al., 2021). With the abundance of satellite imagery data, data-driven methods (e.g., deep learning) for RSL delineation may be quite feasible (Alidoost et al., 2020).

Since the success of AlexNet on the ImageNet Large Scale Visual Recognition Challenge in 2012 (Krizhevsky et al., 2012), the deep convolutional neural network (DCNN) has been of great interest and has brought a series of state-of-the-art (SOTA) achievements in the field of computer vision (Ioannidou et al., 2017; Voulodimos et al., 2018). With the high automatic feature extraction capability and spatial invariance characteristics of DCNNs (Kayhan and Gemert, 2020; Zhao et al., 2019), many studies have employed DCNNs to interpret information from satellite imagery, such as terrain classification (Qian et al., 2020), change identification (Woodcock et al., 2020), and object detection (Zhong et al., 2018). Inspired by the above works, RSL delineation can be regarded as a pixel-classification task of satellite imagery. However, two main challenges remain in applying deep learning to RSL delineation:

(1) The physical architecture and urban functions of buildings are various, resulting in highly diverse roof sizes, forms, and spatial distribution, which complicates the extraction of essential RSL features from satellite imagery using a general DCNN.
(2) The linear edge patterns of RSLs make it a very small proportion of the satellite imagery, so there is a significant class imbalance between foreground objects (i.e., RSLs) and background context, making it difficult for DCNN to concentrate on RSL locations.

To overcome the abovementioned challenges and achieve efficient RSL delineation using satellite imagery, we design the Deep Roof Refiner (DRR)—a detail-oriented deep learning network and propose a synthetic strategy. The proposed DRR is designed as an end-to-end DCNN with an encoder-decoder structure that incorporates several SOTA components to capture multi-scale RSL features and recover distinct RSL boundaries from satellite imagery. Specifically, the components include a split-attention backbone (SAB) network, an atrous spatial pyramid pooling (ASPP), a dual-attention mechanism (DAM), and a detail-refinement module (DRM). To further enhance the DRR's performance and the quality of delineated RSLs, a synthetic strategy is used which incorporates a series of techniques, including a novel hybrid loss function, a transfer learning strategy, an ensemble learning strategy, and a post-processing procedure based on morphological principles. The suggested approaches are evaluated quantitatively and qualitatively, and the results indicate that the delineated RSLs maintain both localization accuracy and edge precision. The contributions of this study are listed below:

(1) This study delineates RSLs from satellite imagery using a data-driven approach based on deep learning that yields good performance.
(2) To efficiently extract RSLs from satellite imagery using deep learning methods, this study highlights two challenges and develops a deep learning network in combination with a synthetic strategy to address them.
(3) Extensive quantitative and qualitative studies suggest that the proposed methods are robust and that the delineated RSLs can be used as a valid data source.

The rest of this paper is organized as follows. Section 2 introduces the materials and pre-analysis of this study. Section 3 presents the DRR's components, synthetic strategy, and evaluation metrics. Section 4 demonstrates the extensive experiments and analyses the effectiveness of the proposed methods. Section 5 discusses the ability of proposed methods to solve challenges and potential usage of the delineated RSLs. Finally, we conclude in Section 6.

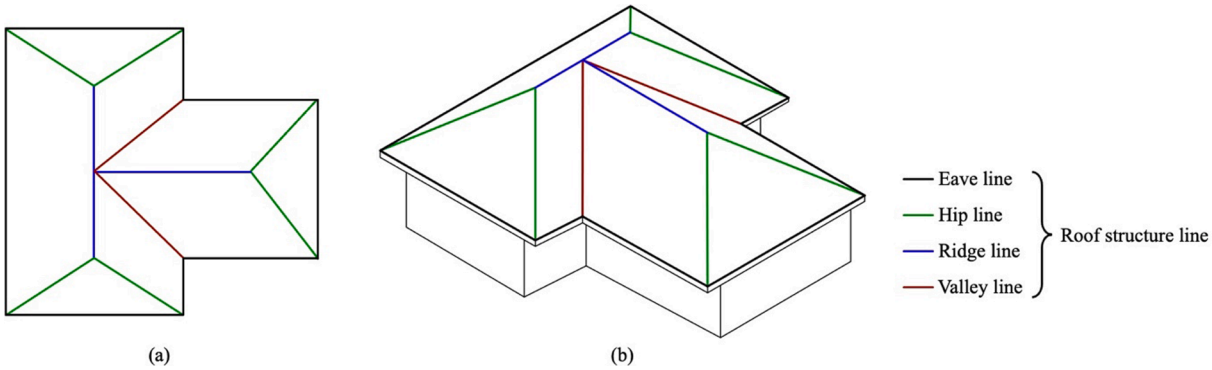


Fig. 1. Diagram of a roof structure line. (a) 2D-viewpoint illustration, (b) 3D-viewpoint illustration.

2. Materials and pre-analysis

2.1. Study area and data source

This study is conducted in the Gulou District, Nanjing, China, which has an area of 73.83 km2, as shown in Fig. 2. The Gulou District contains many functional areas, including governmental institutions, cultural and educational institutions, universities, residential areas, and commercial areas. These functional areas have a variety of building roofs, which provides sufficient and representative samples support to verify the effectiveness of the proposed methods.

This study employs level-18 Google Earth satellite (GES) imagery as the data source because of its open access and high resolution, as illustrated in Fig. 2(c). GES imagery is download based on map service application program interface provided by Google, which can be acquired from https://earth.google.com. The spatial resolution of GES images is generally 0.6 m/pixel, which clearly displays the geometries and structures of various rooftops. However, because the GES imagery is derived from numerous data sources, its resolution may vary among locations. After acquiring the GES imagery, a gamma correction algorithm (Guofu and Zhenghao, 2006) and contrast limited adaptive histogram equalization algorithm (Pizer et al., 1987) are applied to alleviate brightness and sharpness problems.

The study area is divided into three areas based on road network of Open Street Map, which can be acquired from https://www.openstreetmap.org, as shown in Fig. 2(c). These areas are the training area (52.77 km^2), the validation area (8.13 km^2), and the test area (12.93 km^2) from which image datasets are sampled using sliding window method. Based on the definition of RSLs, this study labels the RSLs on the images and generates corresponding mask annotations.

2.2. Empirical data analysis

This study explores the characteristics of the RSLs in the images by analyzing image content and comparing it with other open datasets. Two challenges are identified in delineating the RSLs in the satellite imagery.

2.2.1. Challenge I: High diversity in roof sizes, forms, and spatial distribution

This study examines roof size and discovers that, while roofs aggregate in small sizes, their sizes are not consistent and variable, as seen in Fig. 3(a). Additionally, the structure types of 21,447 roofs in the Gulou District are identified, and the statistical results are presented in Fig. 3(b). The results suggest that roof types are diverse. Specifically, the flat type accounts for around 62%. The gable, hip, and complex types account for around 15%, 5%, and 18% of the area, respectively. Fig. 3(b) also illustrates that several roof samples of different structure types exhibit a variety of forms. The spatial distribution of structures in Gulou District is illustrated in Fig. 3(c), indicating that the spatial distribution of buildings is not uniform, with various uneven dense and sparse zones. As a result, building roofs exhibit a significant degree of variability in

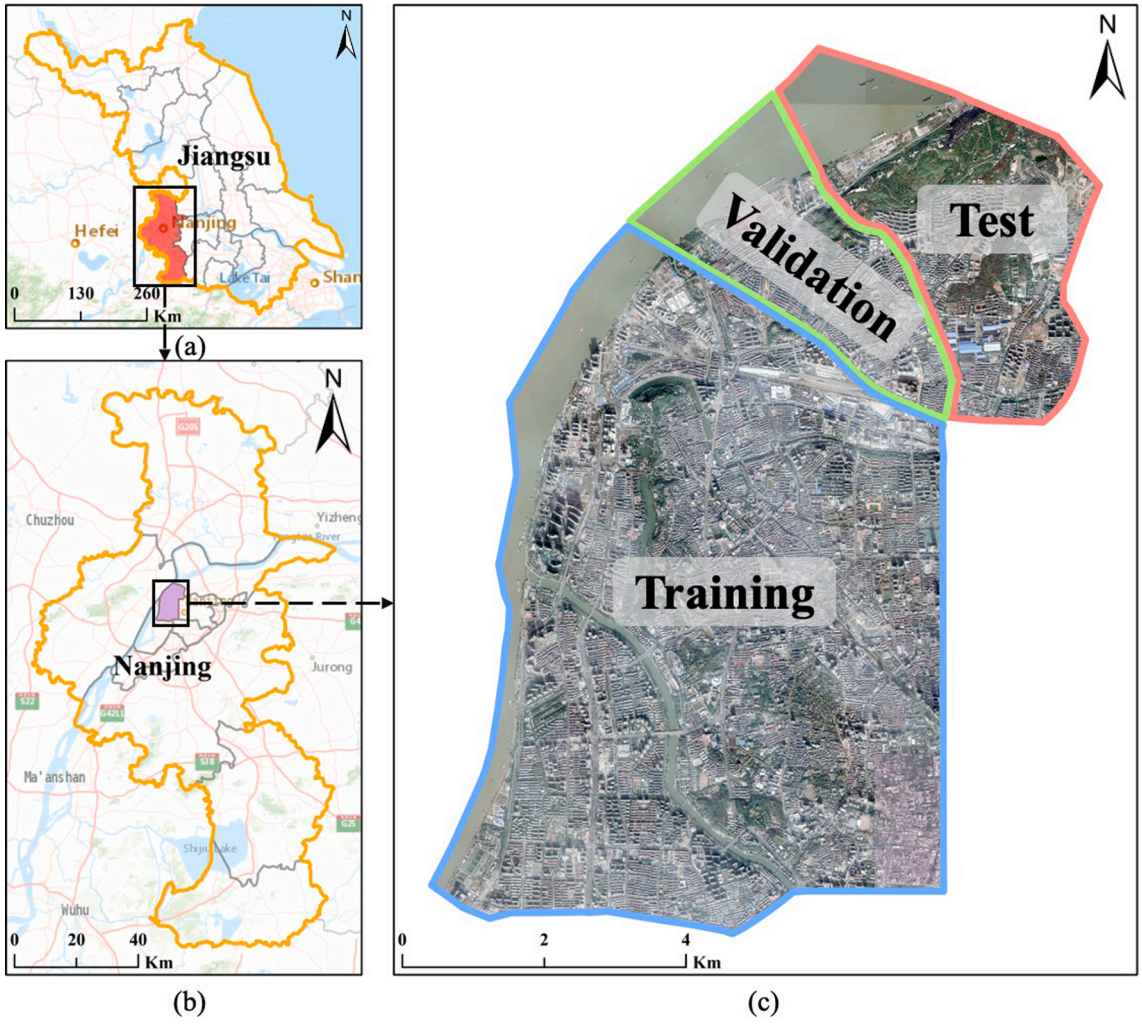


Fig. 2. Study area. (a) Jiangsu Province, (b) Nanjing, and (c) satellite imagery of Gulou District and area division.

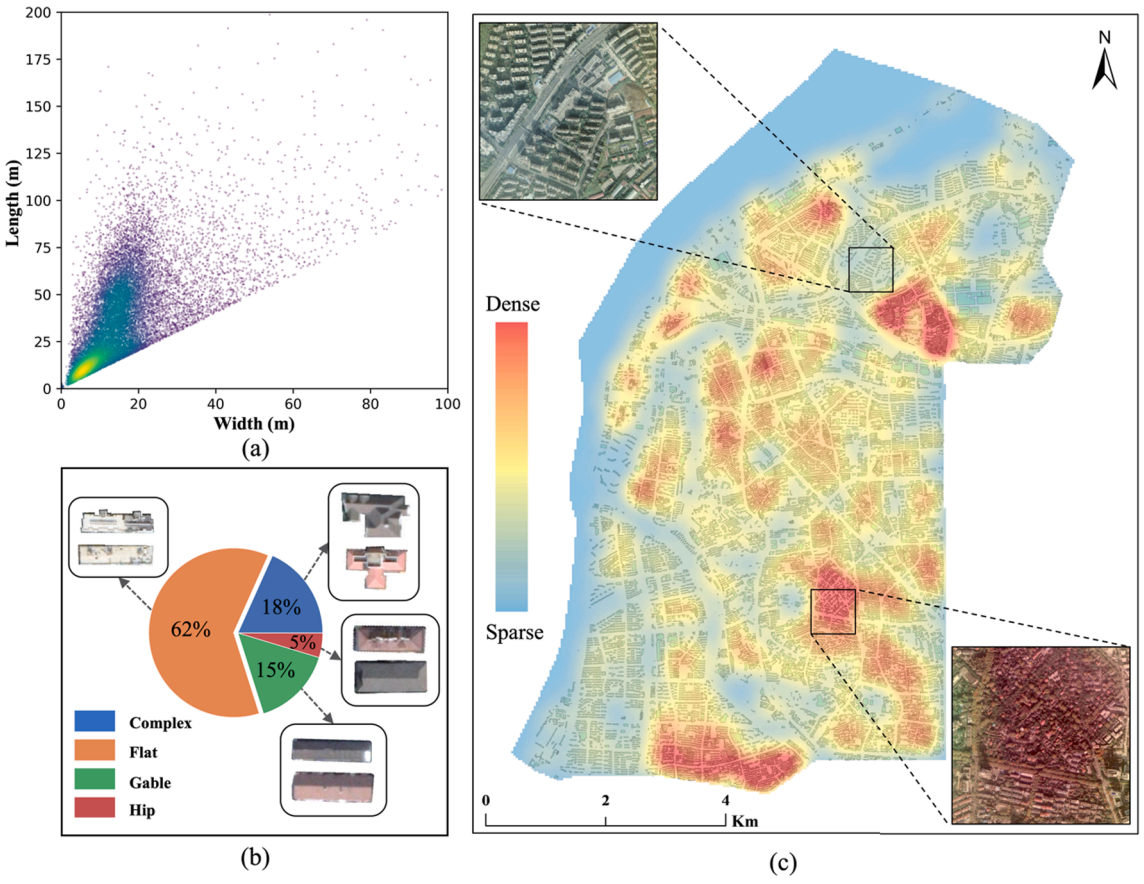


Fig. 3. Illustration of the first challenge. (a) Scatter chart of roof sizes, where the x-label and y-label refers to the width and length of roofs' bounding rectangle based on minimum area principle, (b) structure types and forms statistic of roofs, and (c) spatial distribution of buildings.

terms of roof sizes, shapes, and spatial distribution, which may confound the deep learning network and make it difficult to extract essential features.

2.2.2. Challenge II: Significant class imbalance between foreground objects and background context

Motivated by the findings of class imbalance from satellite imagery by Li et al. (2021), this study further explores the custom RSL dataset.

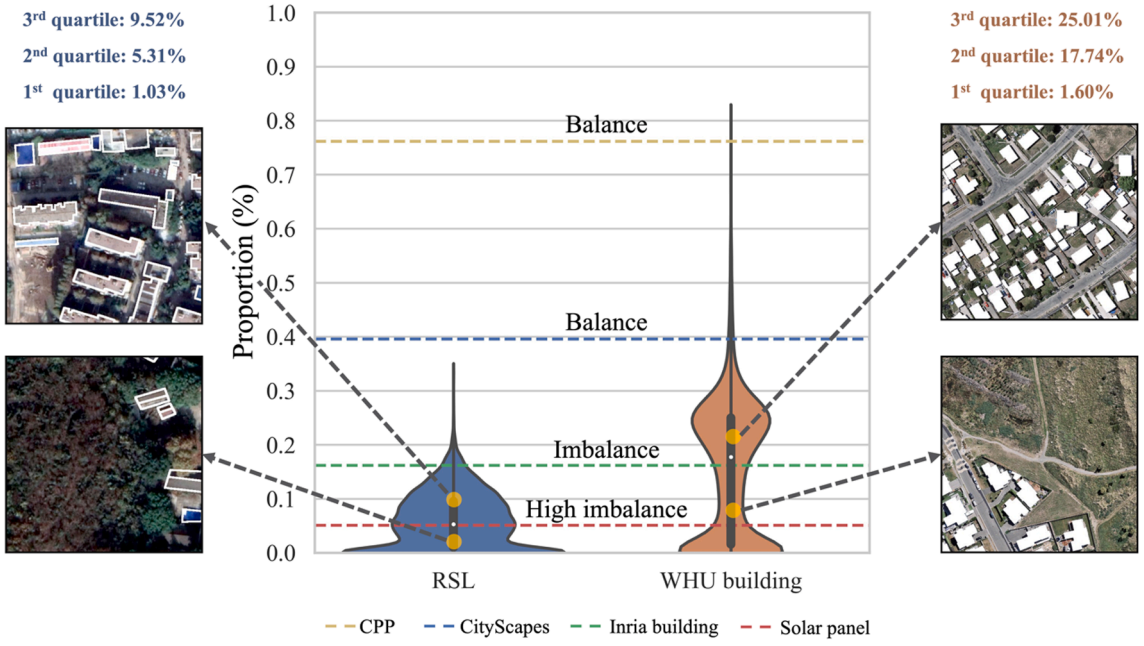


Fig. 4. Illustration of the second challenge. Foreground/background object proportion, where the white markers denote the median, and the black bars denote the interquartile range for the violin plot.

The foreground object proportion distributions in different datasets are shown in Fig. 4, where RSL proportion distribution in the custom RSL dataset and building proportion distribution in the Wuhan University (WHU) building dataset (Ji et al., 2018) are detailed presented. The results indicate that the Clothing Co-Parsing (CCP) dataset (Yang et al., 2014) and the Cityscapes dataset (Cordts et al., 2016) remain relatively class-balanced, with median proportions of 76.2% and 39.6%. On the contrary, the Inria building dataset (Maggiori et al., 2017) and the WHU building dataset presents class imbalances, with a median proportion of 16.2% and 17.7%. Moreover, the results show that there is a significant class imbalance issue in the custom RSL dataset and the Solar panel dataset. The median proportions for RSLs and solar panels are 5.9% and 5.1%. However, the majority of images (75% of images) in the RSL custom dataset have less than 9.5% foreground objects, which is more severe than the 12.7% in the Solar panel dataset (Li et al., 2021). The significant class imbalance issue between foreground objects and background context brings great challenges for network training and RSL location detection.

3. Methodology

This section introduces the components of the DRR, synthetic strategy and evaluation metrics, starting with a general overview of the RSL delineation workflow.

3.1. Overview

The workflow for RSL delineation is depicted in Fig. 5. To begin, training, validation, and test datasets are constructed using GES images. Furthermore, the detail-oriented deep learning network, DRR, is proposed for refined delineation of RSLs, as well as the synthetic strategy containing a series of advanced techniques is adopted to enhance the performance of the DRR. Specifically, the hybrid loss function and the transfer learning strategy are adopted during the training stage. When the DRR has converged, the ensemble learning strategy and the morphological post-processing are adopted during the inference stage. Finally, the quality of the delineated RSLs is evaluated quantitatively and qualitatively.

3.2. Network architecture

3.2.1. Structure of the Deep Roof Refiner (DRR)

The structure of the DRR is shown in Fig. 6. This study uses DeepLabv3+ (Chen et al., 2018) as the baseline, which uses ASPP to capture multi-scale features from the images. Furthermore, in the encoder phase, the powerful SAB is adopted to generate a comprehensive feature map from the raw images, and DAM is connected in parallel with ASPP to weight the spatial-wise and channel-wise information extracted in the large receptive field. In the decoder phase, the encoder feature map is amplified to twice its size and concatenated with low-level features from the backbone. After calculation in the 3×3 convolutions, the DRM is integrated to produce a more refined feature map that is later amplified to generate the RSL map. Based on this novel deep learning network structure, DRR can capture the essential RSL features efficiently and prevent being distracted by confusable objects, thus concentrating more on the RSL locations.

3.2.2. Split-attention backbone (SAB)

The SAB focuses on feature map attention and multipath representation, which can improve information sensing capability of DRR and are crucial for RSL feature extraction from roofs regardless of the diverse sizes, forms, and spatial distribution. The core module is shown in Fig. 7. In a previous study, a split-attention backbone network outperformed other SOTA DCNNs (e.g., ResNet, ResNeXt, and SE-ResNet) on image classification, object detection, and semantic segmentation tasks (Zhang et al., 2020). With the increase in the number of stacking module layers, the feature extraction capability of a network is significantly enhanced while the latency of inference increases, and 50, 101, and 200 layers are commonly used, which are referred to as ResNeSt-50, ResNeSt-101 and ResNeSt-152, respectively. Considering the trade-off between accuracy and speed of inference, ResNeSt-101 is used in the current study.

3.2.3. Dual-attention mechanism (DAM)

To alleviate the effects of confusable background context and the complex object forms, the interim feature maps need to be weighted spatial-wise and channel-wise to get essential representations. A dual-attention mechanism selectively aggregates the similar features of

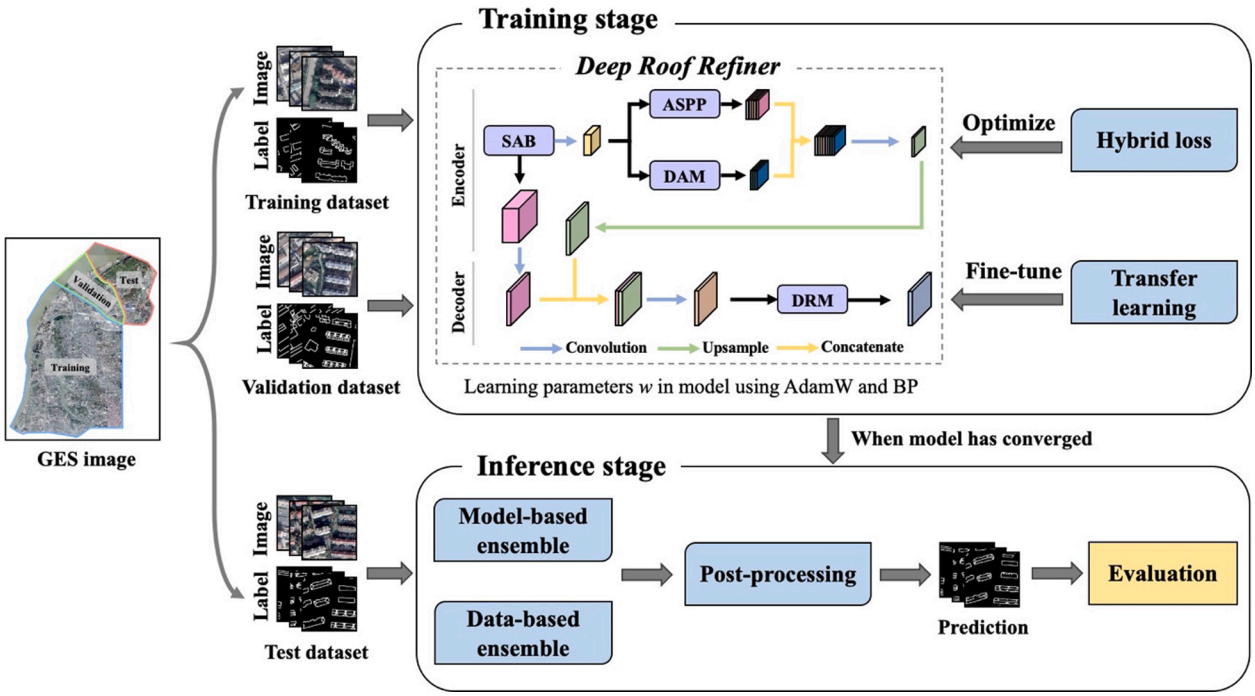


Fig. 5. The overall workflow of RSL delineation using our proposed method.

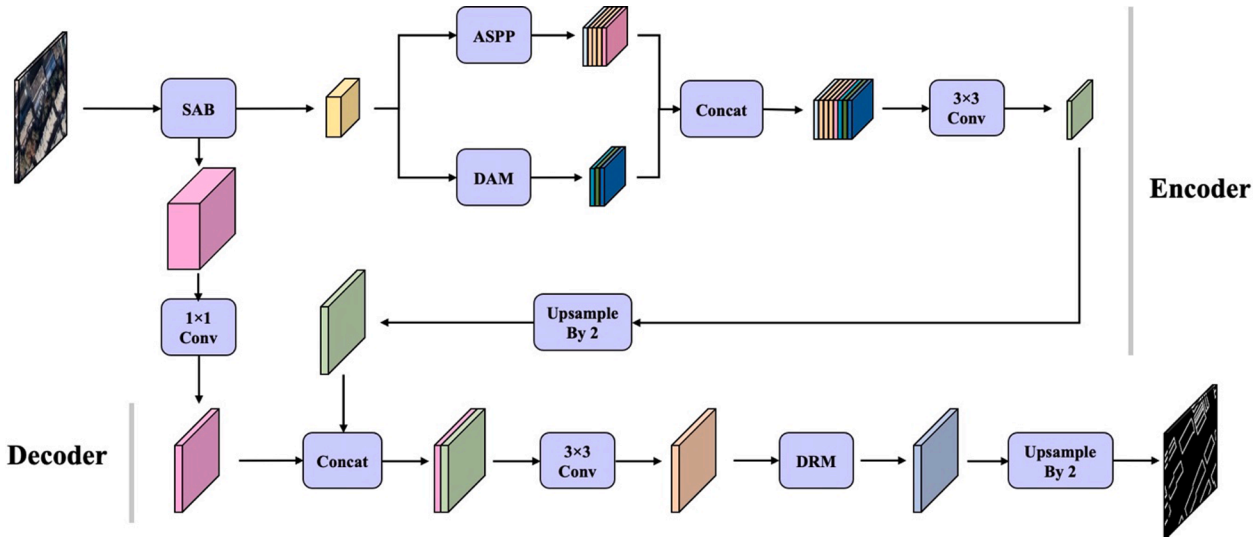


Fig. 6. The structure of the DRR.

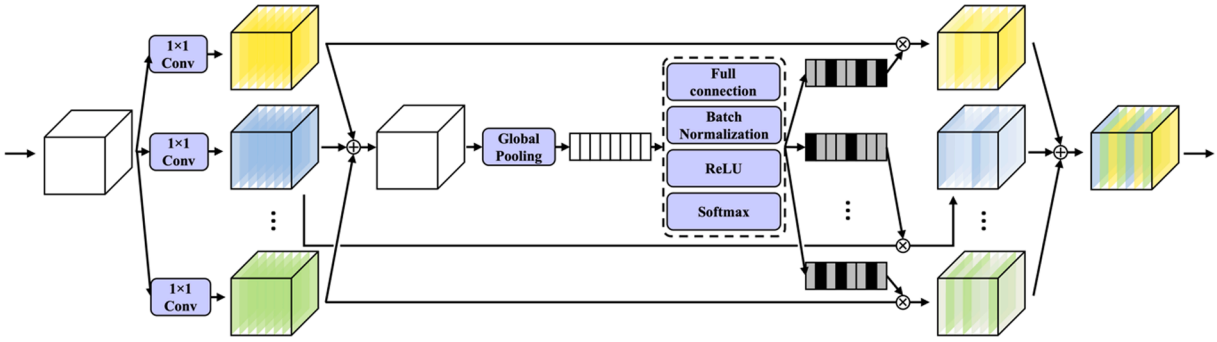


Fig. 7. Illustration of the core module in the SAB.

inconspicuous objects to highlight their feature representations and avoid the influence of spatial-wise and channel-wise salient objects (Fu et al., 2019). As shown in Fig. 8, unlike directly summing the spatial and channel attention results proposed as by Fu et al. (2019), this study further integrates their concatenation. With double-wise constraints on the feature maps, the RSL representation will be more pertinent.

3.2.4. *Detail-refinement module (DRM)*

DRM can be used to enhance the perception of detailed information in some hard samples, such as the transition zone between foreground objects and background context. There are different efficient types of DRM in related deep learning studies, such as multibranch back-propagation (Liu et al., 2017), coarse-to-refine mapping (Qin et al., 2019) and sampling-based refinement (Kirillov et al., 2020), as shown in

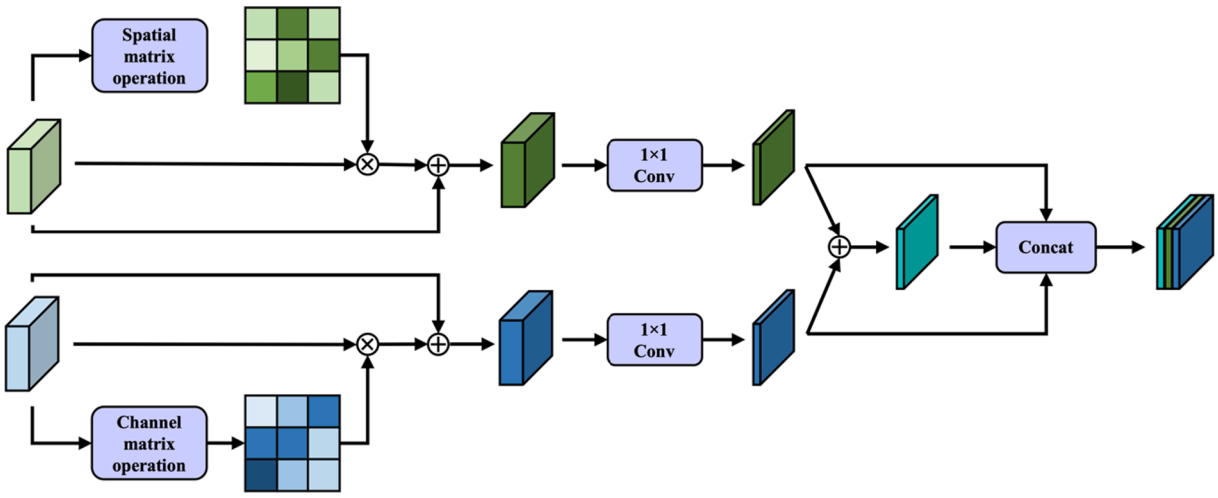


Fig. 8. Illustration of the DAM.

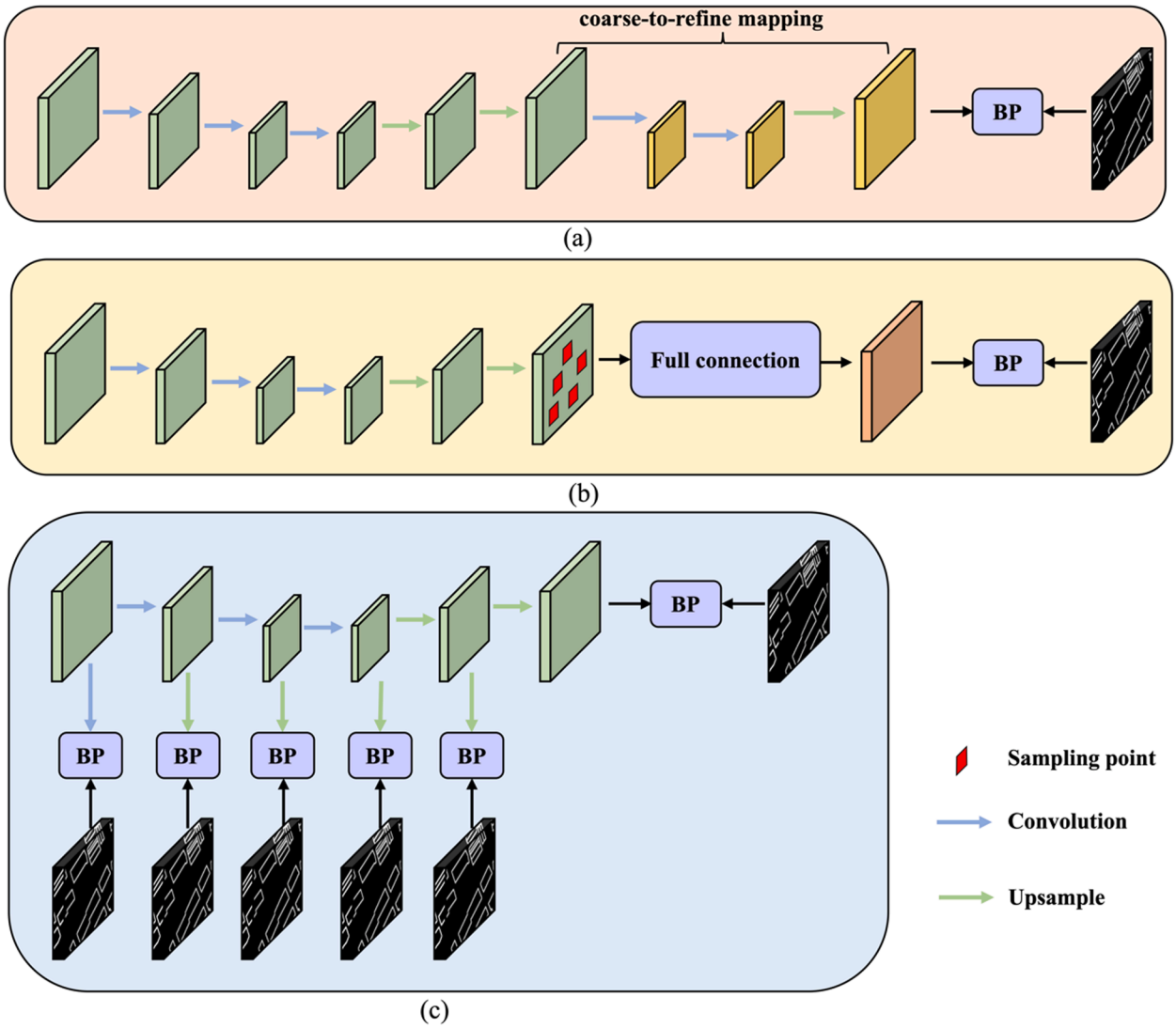


Fig. 9. Illustration of three types of DRM. (a) Coarse-to-refine mapping refinement, (b) sampling-based refinement, and (c) refinement using multibranch backpropagation.

Fig. 9. Multibranch backpropagation introduces multiple loss functions in different branches of a DCNN, whereas coarse-to-refine mapping presents a tiny encoder-decoder DCNN inserted at the end of a network. In comparison to these two methods, sampling-based refinement optimizes only a few hard-sampling points in feature maps, which has an acceptable trade-off between the network accuracy and efficiency. This study uses a typical sampling-based refinement method, Pointrend, proposed by Kirillov et al. (2020).

3.3. Synthetic strategy

Aiming to improve the performance of the DRR, this study adopts a synthetic strategy containing a series of techniques, where a hybrid loss function and transfer learning strategy are used during the training stage, and an ensemble learning strategy and morphological post-processing are used during the inference stage.

3.3.1. Hybrid loss function

To obtain high-quality delineation and alleviate the class imbalance issue, this study designs a hybrid loss, which can optimize the prediction of a feature map from three-level hierarchies: pixel-level, patch-level and map-level. The hybrid loss ℓ is defined as:

$$\ell = \ell_{bce} + \ell_{dice} + \ell_{iou} \tag{1}$$

where ℓ_{bce}, ℓ_{dice}, and ℓ_{iou} denote the BCE loss (De Boer et al., 2005), DICE loss (Milletari et al., 2016) and IoU loss (Máttyus et al., 2017), respectively.

Specifically, the BCE loss (De Boer et al., 2005) is commonly used in binary classification and segmentation tasks and is defined as:

$$\ell_{bce} = -\sum_{(r,c)}[G(r,c)\log(P(r,c)) + (1-G(r,c))\log(1-P(r,c))] \tag{2}$$

where $G(r,c) \in \{0,1\}$ is the ground truth label of pixel (r,c) and $P(r,c)$ is the predicted probability of pixel (r,c).

The DICE loss was originally proposed for medical image segmentation, which is suitable for class imbalance between foreground and background objects (Milletari et al., 2016) and is defined as:

$$\ell_{dice} = 1 - \frac{2\sum_{r=1}^{H}\sum_{c=1}^{W}P(r,c)G(r,c) + \varepsilon}{\sum_{r=1}^{H}\sum_{c=1}^{W}[P(r,c) + G(r,c)] + \varepsilon} \tag{3}$$

where $G(r,c) \in \{0,1\}$ is the ground truth label of pixel (r,c), $P(r,c)$ is the predicted probability of pixel (r,c) and ε is an infinitesimal number called the smoothing factor which smooths the gradient of ℓ_{dice}.

The IoU has been used as a training loss (Rahman and Wang, 2016). This study follows the definition for the IoU loss defined by Máttyus et al. (2017):

$$\ell_{iou} = 1 - \frac{\sum_{r=1}^{H}\sum_{c=1}^{W} P(r,c)G(r,c)}{\sum_{r=1}^{H}\sum_{c=1}^{w}[P(r,c) + G(r,c) - P(r,c)G(r,c)]} \quad (4)$$

where $G(r,c) \in \{0,1\}$ is the ground truth label of pixel (r,c) and $P(r,c)$ is the predicted probability of pixel (r,c).

3.3.2. Transfer learning strategy

Transfer learning is an improvement strategy for a new task that uses the transfer of knowledge from a related task that has already been learned (Torrey and Shavlik, 2010). Owing to the limited amount and diversity of RSL datasets with annotations, similar domain or data distribution datasets can be added to the pretraining DRR based on a transfer learning strategy. This study uses the WHU building dataset (Ji et al., 2018) to transfer knowledge by fine-tuning. The dataset contains approximately 22,000 independent buildings with annotations and can be obtained from http://gpcv.whu.edu.cn/data/building_dataset.html. The knowledge transfer of the building dataset enables DRR to learn the fundamental features of building roofs and understand complex scenes for building in satellite imagery, while also facilitating the capture of RSL essential features.

3.3.3. Ensemble learning strategy

The generalization capability of a single deep learning network is limited for RSL delineation, and ensemble strategies can be adopted to improve the final accuracy beyond that of a single network. Deep learning ensembles can be classified into two types: model-based and data-based (Cao et al., 2020; Ganaie and Hu, 2021). This study chooses one method each from the network-based and the data-based ensemble. The details are as follows.

1) *Snapshot*

Snapshot (Huang et al., 2017) is a typical model-based ensemble approach that integrates a single network. During the network training stage, periodic decay algorithms such as cosine annealing algorithm (Loshchilov and Hutter, 2016) are used to determine the hyperparameters, allowing the network to attain several local optimum states in a short amount of time. Snapshot selects several locally optimal checkpoints and integrates them based on the defined rules (e.g., voting and averaging).

2) *Test-time augmentation*

Test-time augmentation (TTA) is a data-based ensemble approach that augments test data during the inference stage (Wang et al., 2019). Specifically, several augmentation methods are adopted to generate duplicates of each sample in the test set and then then integrate the predictions for each duplicate. In this study, horizontal and vertical flipping and rotation are chosen as augmentation methods.

3.3.4. Morphological post-processing

The predicted RSL objects are frequently fragmented and disjointed, and the prediction maps contain some noise. This study applies a morphological post-processing approach to enhance the connectivity and topological properties of RSLs. To begin, the fractures of RSLs are connected by a closure operation. Furthermore, the area is used to filter isolated pixel clusters. Finally, the skeleton connectivity enhancement approach is designed. Specifically, the skeletons of RSLs are extracted and the vertices at the end of the skeletons are connected using eight neighborhoods to assure connectivity while remaining inside the original prediction zone.

3.4. Accuracy assessment

To quantify the effectiveness of the proposed methods, this study selects two metrics based on the annotations of RSLs: the optimal dataset scale F1-score (ODS-F1) and optimal image scale F1-score (OIS-F1) (Liu et al., 2017). The ODS-F1 is globally optimal on the dataset scale, where a fixed threshold is applied to all images to maximize the F1-score on the whole dataset. The OIS-F1 is optimal on the image scale, where a changeable threshold is applied to each image to maximize the F1-score on every image sample. The F1-score is defined as:

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

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

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

Where FN, TN, TP, and FP refer to numbers of false-negative, true-negative, true-positive, and false-positive pixels, respectively. TP means the pixels are positively predicted with positive labels in the prediction maps. In contrast, FN means negative labels with false predictions in the prediction maps.

4. Experiments

4.1. Experimental setup

The experiments are implemented using PyTorch (Paszke et al., 2019) and MMSegmentation (Xu et al., 2020) frameworks on two NVIDIA Tesla V100 GPUs. The dataset is generated by using sliding windows of size 512 × 512 pixels. The window step size of the training set and validation set is 384 × 384 pixels and that of the test set is 512 × 512 pixels. The training, validation, and test image patches are 4242, 720, and 680, respectively. Before training the network, this study applies data augmentation methods to the training set, including random crop, random flip, and photometric distortion, to avoid overfitting issue. During the training stage, the configurations for the networks are set as presented in Table 1.

4.2. Ablation study

In this section, this study validates the effectiveness of each proposed component and technique. The ablation study contains two parts: Network architecture ablation and synthetic strategy ablation.

4.2.1. Network architecture ablation

In this experiment, we test the effectiveness of our DRR architecture, starting by deploying DeepLabv3 + as the baseline and then progressively extend it with SAB, DAM and DRM. Table 2 and Fig. 10(a) illustrate the results of the architecture ablation study. As we can see, with

Table 1
Details of experiment configuration.

Item	Configuration
Batch size	4 per GPU
Optimizer	AdamW (Loshchilov and Hutter, 2018)
Learning rate	0.01
Weight decay rate	0.001
Learning rate scheduler	Polynomial warm restart (Mishra and Sarawadekar, 2019)
Power of learning rate scheduler	0.9
Minimum learning rate	0.0001
Loss function	DICE (Milletari et al., 2016)

Table 2
Quantitative results of the DRR architecture ablation.

Ablation	Baseline	SAB	DAM	DRM	ODS-F1 (%)	OIS-F1 (%)
I	✓				55.11	56.74
II	✓	✓			56.14	57.72
III	✓	✓	✓		56.47	58.05
IV	✓	✓		✓	58.48	60.10
DRR	✓	✓	✓	✓	58.57	60.32

different components integrated into the baseline, the performance increases gradually. The proposed DRR architecture achieves the best performance among these configurations, which outperforms the baseline by 3.46% and 3.58% for the ODS-F1 and OIS-F1, respectively.

4.2.2. Synthetic strategy ablation

We gradually adopt different techniques based on the DRR, including a hybrid loss function, a transfer learning strategy, ensemble learning strategy, and a post-processing procedure. The ablation experimental results are shown in Table 3 and Fig. 10(b), indicating that each technique can significantly improve the DRR's performance. Although the post-processing procedure boosts the ODS-F1 and OIS-F1 by only 0.04% and 0.18%, respectively, it enhanced the connectivity of the RSLs. With all techniques adopted, the result outperforms the DRR by 2.13% and 2.63% for the ODS-F1 and OIS-F1, respectively.

4.3. Comparison with SOTA methods

To evaluate the performance of the DRR, we report the quantitative and qualitative results compared to other approaches. Six methods from conventional computer vision to deep learning, including Canny (McIlhagga, 2011), Canny-Mask (Mainzer et al., 2017), PSPNet (Zhao et al., 2017), DeepLabv3+ (Chen et al., 2018), DMNet (He et al., 2019), and OCRNet (Yuan et al., 2020), were selected as the methods to compare with the DRR. These methods have been proven effective in edge detection, semantic segmentation, or building extraction. Due to lack of building footprint data from Bing Map for China, this study uses building footprint data published by the Resource and Environment Science and Data Center as substitutes in Canny-Mask, which is acquired from https://www.resdc.cn/.

4.3.1. Quantitative evaluation

The comparison deep learning networks are configured with the same universal hyperparameters listed in Table 1 with an integration of the ResNet-101 backbone network. As shown in Table 4 and Fig. 11, The accuracy of RSL delineation is greatly improved when it is applied as a data-driven task based on deep learning. Due to the poor feature extraction capability of Canny and the limited accuracy of auxiliary building footprint data, Canny-Mask is unable achieve the expected

Table 4
Comparison of the proposed method and six other methods.

Type	Method	ODS-F1 (%)	OIS-F1 (%)
Conventional method	Canny	25.10	26.47
	Canny-Mask	22.53	22.78
Deep learning method	PSPNet	53.90	55.65
	DeepLabv3+	55.11	56.74
	DMNet	54.53	56.18
	OCRNet	56.02	57.25
	DRR	58.57	60.32
	DRR-SS	**60.89**	**63.48**

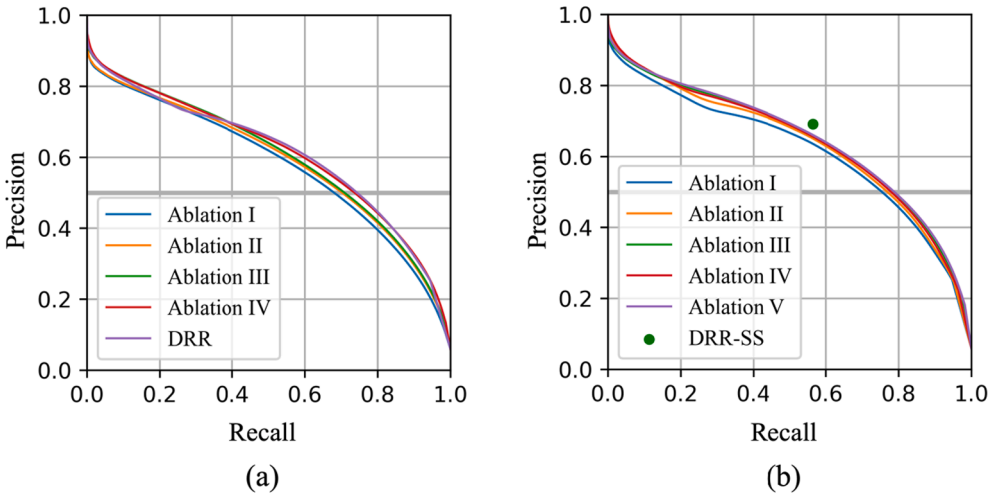


Fig. 10. Precision-recall curves of the ablation study. (a) Precision-recall curves of network architecture ablation, (b) precision-recall curves of synthetic strategy ablation. Precision-recall curve of DRR-SS presents a dot because the post-processing procedure maps the prediction results to 0 or 1 values, while the Precision-recall curves depend on the probability results.

Table 3
Quantitative results of the synthetic strategy ablation.

Ablation	Hybrid Loss	Transfer learning	Ensemble learning		Post-processing	ODS-F1 (%)	OIS-F1 (%)
			Snapshot	TTA			
I	✓					58.76	60.85
II	✓	✓				59.87	61.96
III	✓	✓	✓			60.44	62.83
IV	✓	✓		✓		60.36	62.65
V	✓	✓	✓	✓		60.85	63.30
DRR-SS	✓	✓	✓	✓	✓	**60.89**	**63.48**

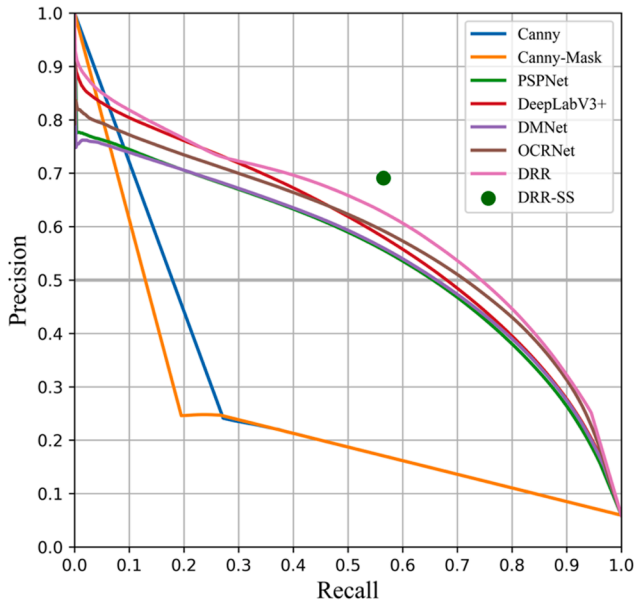


Fig. 11. Precision-recall curves of the different methods.

performance demonstrated in a related study (Mainzer et al., 2017). Although the comparison deep learning networks performed well on open datasets, they show significantly lower performance than DRR at RSL delineation. The proposed DRR with synthetic strategy achieves the best performance among the comparison methods, which outperforms SOTA deep learning architectures by over 4.87% and 6.23% and the related conventional research method by 38.36% and 40.70% for the ODS-F1 and OIS-F1 scores, respectively.

4.3.2. Qualitative evaluation

To further illustrate the performance of the proposed methods, Fig. 12 shows the qualitative comparison of the results with other methods. Compared with deep-learning-based methods, conventional methods (i.e., Canny and Canny-Mask) cannot delineate continuous RSLs and introduce extensive messy noise. With the support of building footprints, Canny-Mask can delineate clear eave roof lines. However, it has limitations in delineating the internal portions of roofs. Canny-Mask results depend on the quality of the building footprint data, resulting in numerous errors when building footprints do not match the most recent urban building layout. Additionally, the RSLs obtained from the compared deep learning methods are fuzzy without a clear boundary in most of the yellow rectangles, especially for rooftops with complex structures. Networks with large and multi-scale reception fields, such as PSPNet, DeepLabv3 + and DMNet, can sense the global features of images and delineate eave lines well, but the interior lines are blurred because of the lack of attention concentration on the feature maps. With an improvement in object-contextual representation, OCRNet has more sufficient delineation of the interior area of building roofs, but the interior lines still present pixel clusters. Compared with other deep learning networks, DRR has better RSL delineation results. With the support of the synthesis strategy, DRR achieves significant refinement results on different building roof forms, focusing on the location and edge perception of foreground RSL objects. This can be explained by the sufficient combination of high-power components in the DRR and the effectiveness of the synthetic strategy.

4.4. Robustness and limitation assessment

The robustness and limitations of proposed methods are evaluated using randomly selected locations from ten additional districts in Nanjing. Fig. 13 shows that building roofs present a variety of architectural sizes and forms. Nonetheless, the proposed DRR with synthetic strategy is capable of effectively distinguishing between foreground objects and complicated background context and delineating RSLs further. As can be seen from Fig. 13, when confronted with buildings with varying spatial distribution, the proposed methods can clearly delineate not only the RSLs of sparsely distributed buildings, but also enumerate and delineate each roof of densely distributed buildings. As a result, it demonstrates that the proposed methods are robust across a range of roof sizes, forms, and spatial distribution.

As shown in Fig. 14, the delineated RSLs based on satellite imagery in suburban areas, such as those in Gaochun and Lishui Districts, are somewhat cluttered. This is because, while level-18 GES imagery generally has a resolution of 0.6 m/pixel, there is a noticeable difference in resolution between urban and suburban areas, with suburban imagery having a lower resolution. Additionally, some GES imagery is not orthorectified, resulting in the DRR being confused between rooftops and walls that exhibit similar features, such as images in Liuhe District. Therefore, this indicates that the proposed methods have some limitations in terms of low-resolution and non-orthorectified satellite imagery.

5. Discussion

5.1. Challenge solving capabilities

Combining high-resolution satellite imagery with data-driven methods based on deep learning presents two distinct challenges for RSL delineation. To respond to these challenges, we develop a detail-oriented deep learning network with a synthetic strategy and prove its effectiveness through quantitative and qualitative experiments. To overcome the first challenge, the SAB and transfer learning strategy are utilized to capture essential features of RSLs with a variety of roof sizes, forms, and spatial distribution. As illustrated in Fig. 15, the RSLs of roofs exhibit a high degree of confidence. Due to the great capability for feature extraction, RSLs on roofs with clutter and roofs with complicated physical forms can be identified efficiently. In terms of addressing the second challenge, DAM and hybrid loss functions are used to tackle the issue of class imbalance. The confidence maps in Fig. 15 demonstrate that the proposed methods can effectively concentrate on RSL objects and distinguish them from a variety of backgrounds. Additionally, other measures such as the network's encoder-decoder design, DRM, ensemble strategy, and post-processing methods also contribute to the network's performance improvement. As a result, it implies that the proposed solutions are capable of handling both challenges.

The study demonstrates that the proposed methods have limits when used with low-resolution and non-orthorectified satellite imagery, escalating the severity of these two challenges. When imagery has a low resolution, the RSLs on it have hazy boundaries and ambiguous features, which can easily cause confusion for deep learning networks. Additionally, non-orthogonal satellite imagery captures the walls of structures, which might be mistaken for flat roofs due to their physical similarities. In this situation, addressing these challenges just through model modification is insufficient. Improving the data source appears feasible, for as by integrating more high-resolution satellite images and applying advanced data augmentation techniques based on generative models.

5.2. Potential usage of RSLs

RSLs delineated from satellite imagery can be used to conduct fine-grained urban studies. For instance, RSLs enable the division of roofs into numerous components, each of which can be mapped with additional properties to provide more precise data via geographic information systems. This study illustrates a potential usage for RSLs by creating fine-grained building footprints with azimuth angles. Based on related investigations have demonstrated that RSL may be transformed to building footprints (Brédif et al., 2013), the delineated RSLs in Gulou

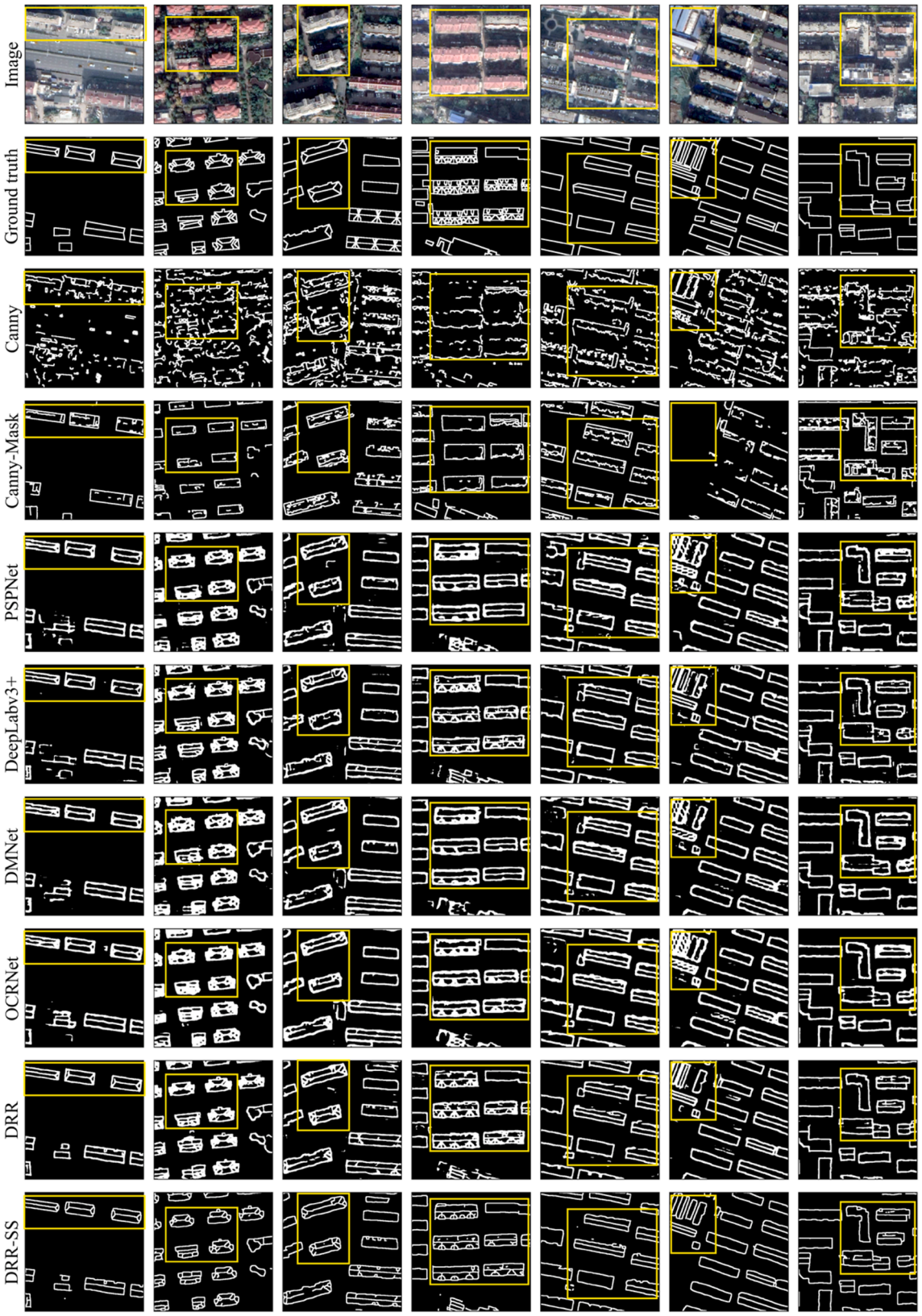


Fig. 12. RSL delineation results of the comparison methods.

District have been filled to generate building footprints. To partition building footprints by RSLs, both the RSLs and the building footprints are vectorized for topology analysis. As indicated in Fig. 16, the azimuth angles of the components are determined by their dominant direction. The azimuth angles of building roofs in Gulou District follow a normal distribution, with the majority of angles falling between -15° and 15°. Based on these generated data, it is possible to estimate the fine-grained solar energy of individual buildings and evaluate the urban layout.

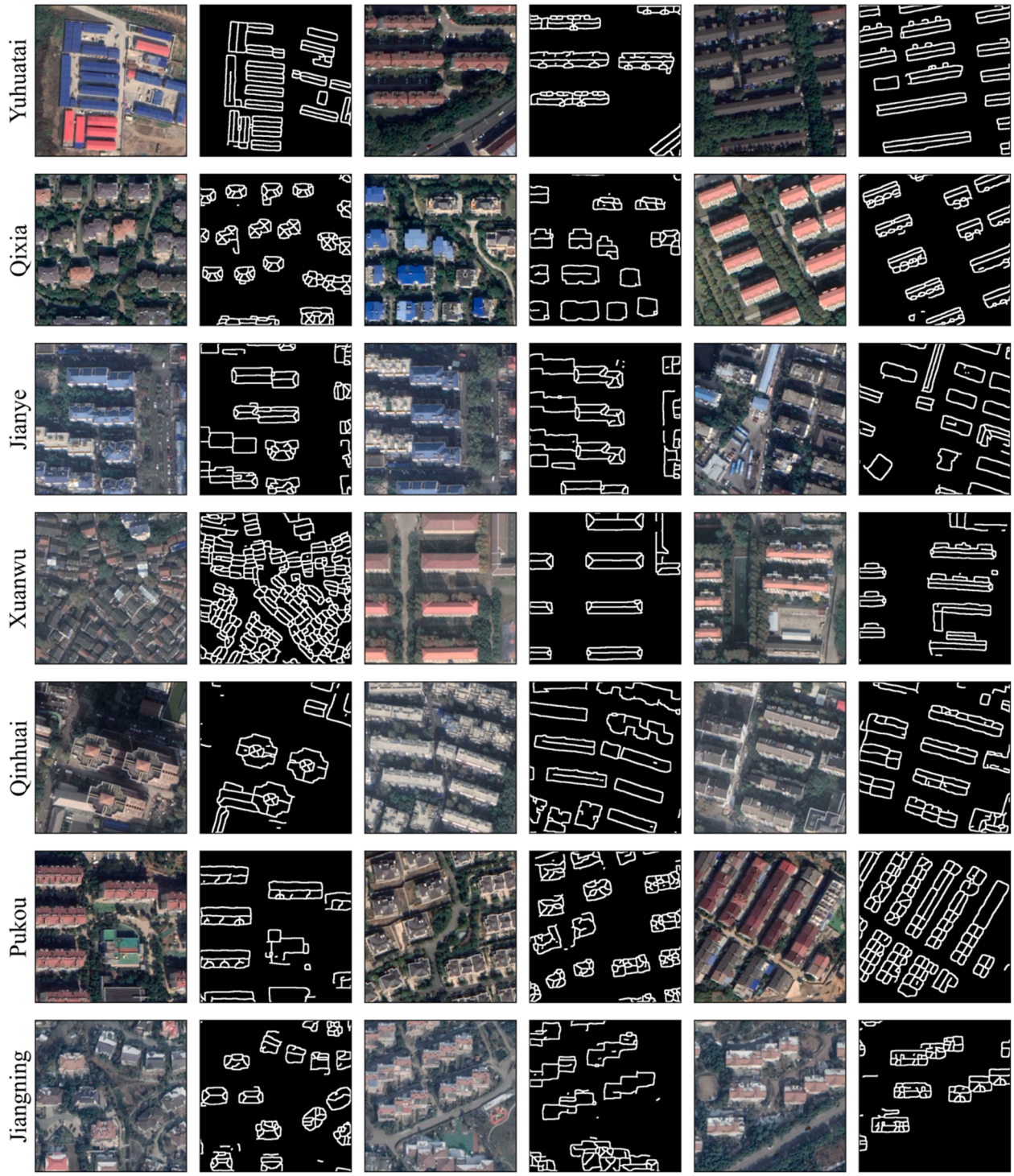


Fig. 13. RSL delineation results for robustness assessment.

6. Conclusion

This study highlights two challenges in utilizing deep learning networks to create RSLs from satellite imagery. To address these challenges, a novel detail-oriented deep learning network, DRR, and a synthetic strategy are designed. Quantitative and qualitative comparison experiments demonstrate the effectiveness and capability of DRR with synthetic strategy. The robustness assessment study shows that the proposed methods have a high generalization ability to different regions. Finally, the model's capability for resolving two challenges is discussed. Moreover, the potential usage of RSLs is discussed that the delineated RSLs can provide more detailed information for urban building-related research, such as urban layout mapping and solar energy estimation. In the future, in order to obtain more refined roof information, we will further conduct on detecting for detailed RSLs classes, including the identification of ridge lines, valley lines, hip lines, and eave lines.

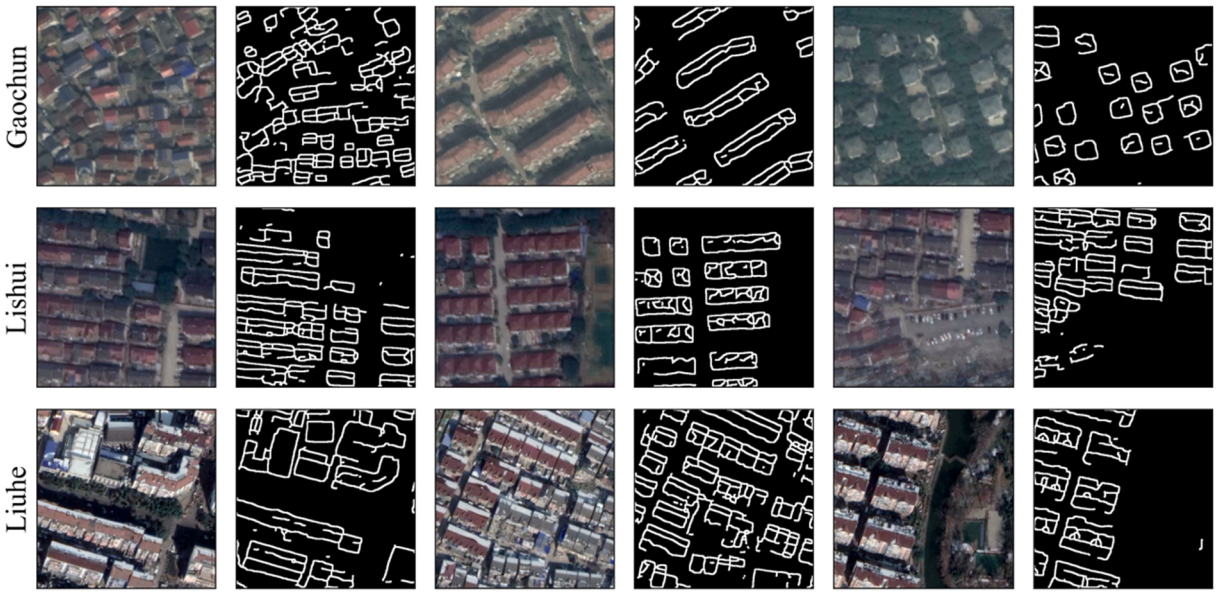


Fig. 14. RSL delineation results for limitation assessment.

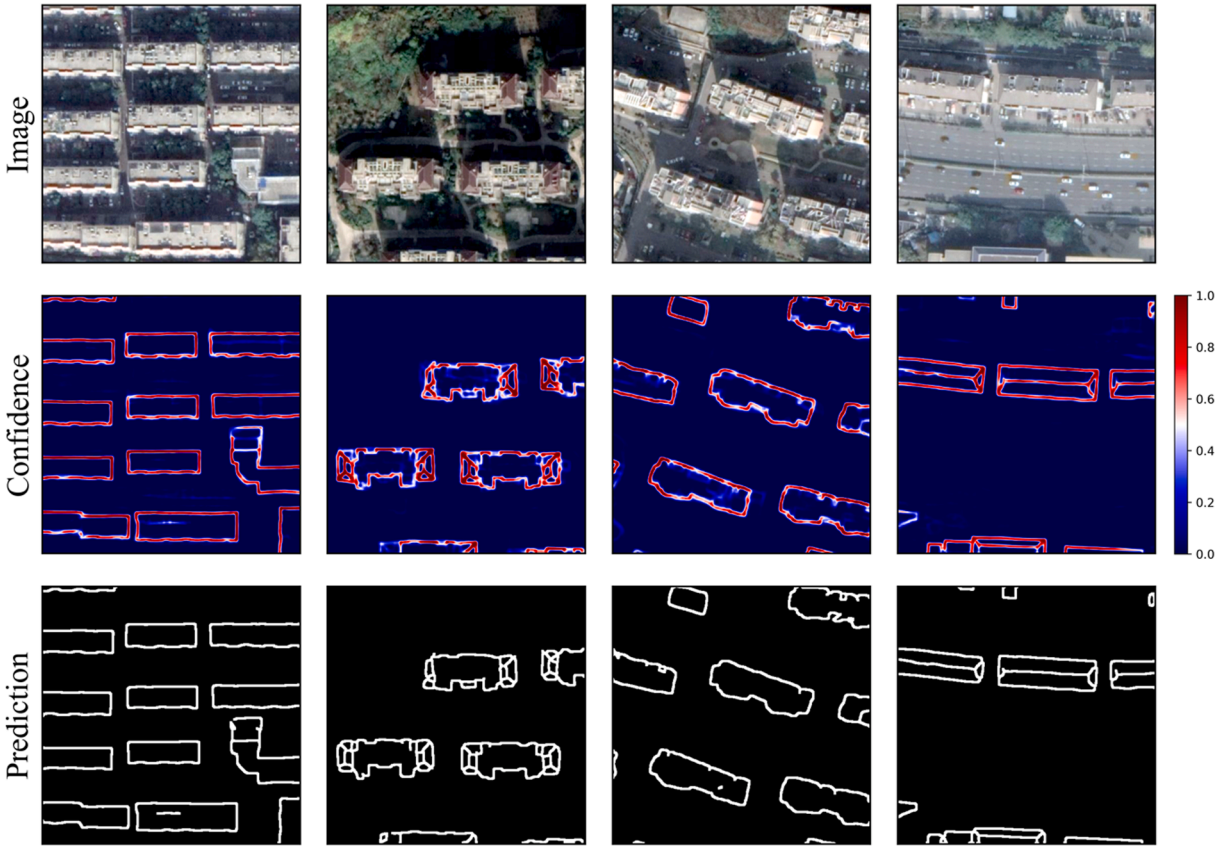


Fig. 15. Illustration of RSL delineation process based on the proposed methods. Confidence maps are calculated by DRR with synthetic strategy (without morphological post-processing procedure). Prediction maps are calculated based on the confidence map and post-processing procedure, where the confidence value greater than 0.3 is an RSL, and the opposite is the background.

CRediT authorship contribution statement

Zhen Qian: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Min Chen:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition. **Teng Zhong:** Methodology, Writing – review & editing. **Fan Zhang:** Methodology, Writing – review & editing. **Rui Zhu:** Formal analysis, Writing – review & editing. **Zhixin Zhang:** Formal analysis, Validation, Writing – review & editing. **Kai Zhang:** Investigation, Writing – review & editing. **Zhuo Sun:** Data curation, Resources. **Guonian Lü:** Conceptualization, Supervision.

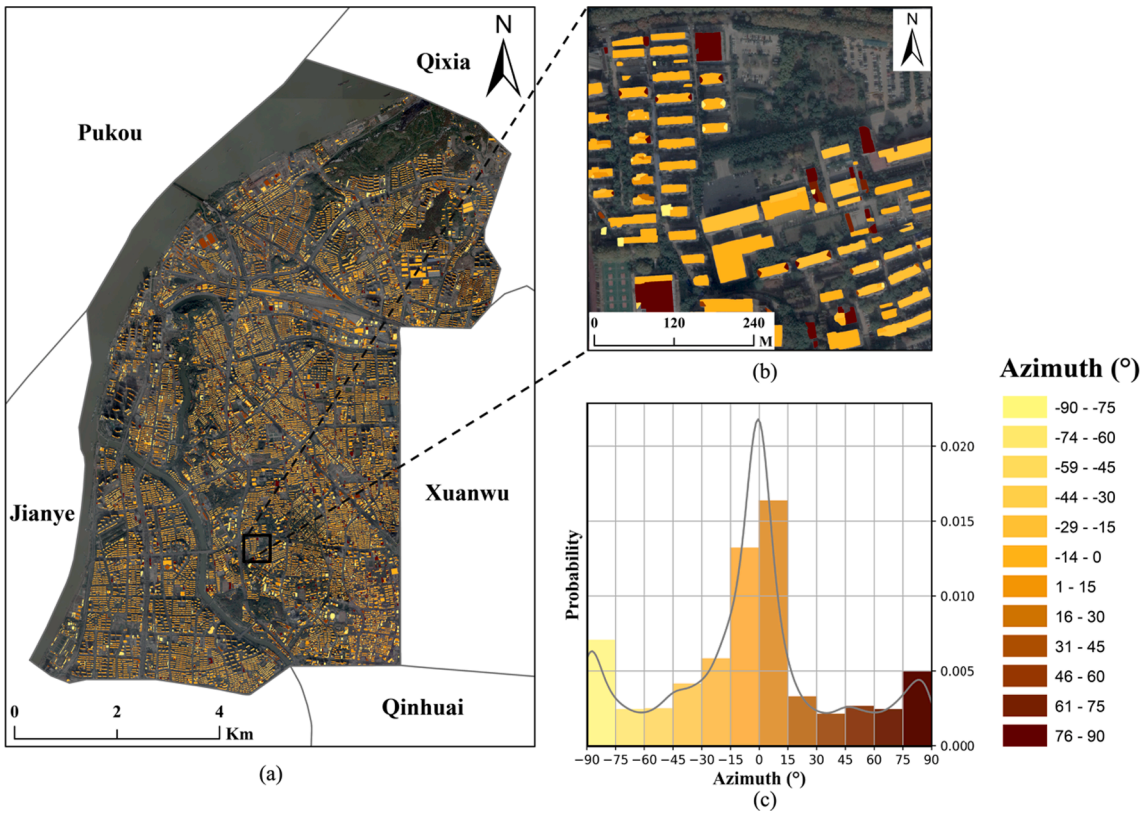


Fig. 16. Visualization results of the azimuth angles of roof units in Gulou District. (a) Azimuth angles of roof units in Gulou District, (b) details of the azimuth angles of roof units, (c) statistics of the azimuth angles in Gulou District. The azimuth angle is calculated based on the dominant angles of the basic units, which is the angle of the longest collection of segments and clockwise with zero at north.

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

We would like to thank the editors and the anonymous reviewers for their meticulous comments and suggestions, which greatly helped us to improve the manuscript quality. This work was supported by Joint Fund Project of National Natural Science Foundation of China (Grant U 1811464), Strategic Hiring Scheme (Grant No. P0036221) at the Hong Kong Polytechnic University, and the General Research Fund (Grant No. 15602619).

References

Alidoost, F., Arefi, H., Hahn, M., 2020. Y-Shaped convolutional neural network for 3D roof elements extraction to reconstruct building models from a single aerial image. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-2-2020, pp. 321–328.

Brédif, M., Tournaire, O., Vallet, B., Champion, N., 2013. Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework. ISPRS J. Photogramm. Remote Sens. 77, 57–65. https://doi.org/10.1016/j.isprsjprs.2012.11.007.

Cao, R., Zhang, Y., Liu, X., Zhao, Z., 2017. 3D building roof reconstruction from airborne LiDAR point clouds: a framework based on a spatial database. Int. J. Geogr. Inform. Sci. 31 (7), 1359–1380. https://doi.org/10.1080/13658816.2017.1301456.

Cao, Y., Geddes, T.A., Yang, J.Y.H., Yang, P., 2020. Ensemble deep learning in bioinformatics. Nat. Mach. Intell. 2 (9), 500–508. https://doi.org/10.1038/s42256-020-0217-y.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

Cheng, C., Zhu, R., Costa, A.M., Thompson, R.G., 2021. Optimisation of waste clean-up after large-scale disasters. Waste Manage. 119, 1–10. https://doi.org/10.1016/j.wasman.2020.09.023.

Xu, J., Chen, K., Lin, D., 2020. MMSegmenation. https://github.com/openmmlab/mmsegmentation.

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. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.

Dal Poz, A.P., Fernandes, V.J.M., 2016. Extraction of roof lines from high-resolution images by a grouping method. In: ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B3, pp. 853–857.

de Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y., 2005. A tutorial on the cross-entropy method. Ann. Oper. Res. 134 (1), 19–67. https://doi.org/10.1007/s10479-005-5724-z.

Demir, N., 2018. Automated detection of 3D roof planes from lidar data. J. Indian Soc. Remote Sens. 46 (8), 1265–1272. https://doi.org/10.1007/s12524-018-0802-2.

Deng, W., Shi, Q., Li, J., 2021. Attention-gate-based encoder-decoder network for automatical building extraction. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2611–2620. https://doi.org/10.1109/JSTARS.2021.3058097.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.

Ganaie, M., Hu, M., 2021. Ensemble deep learning: A review. arXiv preprint arXiv: 2104.02395.

Guo, H., Shi, Q., Marinoni, A., Du, B.o., Zhang, L., 2021. Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images. Remote Sens. Environ. 264, 112589. https://doi.org/10.1016/j.rse.2021.112589.

Peng, G., Lin, Z., 2006. A study on gamma correction and its implementation in image processing. Electron. Eng. 2.

He, J., Deng, Z., Qiao, Y., 2019. Dynamic multi-scale filters for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3562–3572.

Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q., 2017. Snapshot ensembles: Train 1, get m for free. arXiv preprint arXiv:1704.00109.

Ioannidou, A., Chatzilari, E., Nikolopoulos, S., Kompatsiaris, I., 2017. Deep learning advances in computer vision with 3D data. ACM Comput. Surv. 50 (2), 1–38. https://doi.org/10.1145/3042064.

Ji, S., Wei, S., Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Trans. Geosci. Remote Sens. 57 (1), 574–586. https://doi.org/10.1109/TGRS.2018.2858817.

Kayhan, O.S., Gemert, J., 2020. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14274–14285.

Kirillov, A., Wu, Y., He, K., Girshick, R., 2020. Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9799–9808.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. 25, 1097–1105. https://doi.org/10.1145/3065386.

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. https://doi.org/10.1016/j.adapen.2021.100057.

Li, W., Batty, M., Goodchild, M.F., 2020. Real-time GIS for smart cities. Taylor & Francis.

Liu, Y., Cheng, M.M., Hu, X., Wang, K., Bai, X., 2017. Richer convolutional features for edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3000–3009.

Loshchilov, I., Hutter, F., 2016. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983.

Loshchilov, I., Hutter, F., 2018. Fixing weight decay regularization in adam. arXiv preprint arXiv:1711.05101.

Lü, G., Chen, M., Yuan, L., Zhou, L., Wen, Y., Wu, M., Hu, B., Yu, Z., Yue, S., Sheng, Y., 2018. Geographic scenario: a possible foundation for further development of virtual geographic environments. Int. J. Digital Earth 11 (4), 356–368. https://doi.org/10.1080/17538947.2017.1374477.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) IEEE, pp. 3226–3229.

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. Sol. Energy 155, 561–573. https://doi.org/10.1016/j.solener.2017.06.065.

Máttyus, G., Luo, W., Urtasun, R., 2017. Deeproadmapper: Extracting road topology from aerial images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3438–3446.

McIlhagga, W., 2011. The Canny edge detector revisited. Int. J. Comput. Vision 91 (3), 251–261. https://doi.org/10.1007/s11263-010-0392-0.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016 fourth international conference on 3D vision (3DV). IEEE 565–571. https://doi.org/10.1109/3DV.2016.79.

Mishra, P., Sarawadekar, K., 2019. Polynomial learning rate policy with warm restart for deep neural network. In: TENCON 2019–2019 IEEE Region 10 Conference (TENCON) IEEE, pp. 2087–2092.

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., Scartezzini, J.-L., 2018. A city-scale roof shape classification using machine learning for solar energy applications. Renew. Energy 121, 81–93. https://doi.org/10.1016/j.renene.2017.12.096.

Niu, T., Chen, Y., Yuan, Y., et al., 2020. Measuring urban poverty using multi-source data and a random forest algorithm: A case study in Guangzhou. Sustain. Cities Soc. 54 (102014) https://doi.org/10.1016/j.scs.2020.102014.

Nouvel, R., Zirak, M., Coors, V., Eicker, U., 2017. The influence of data quality on urban heating demand modeling using 3D city models. Comput. Environ. Urban Syst. 64, 68–80. https://doi.org/10.1016/j.compenvurbsys.2016.12.005.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inform. Process. Syst. 32, 8026–8037.

Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K., 1987. Adaptive histogram equalization and its variations. Computer Vis. Graphics Image Process. 39 (3), 355–368. https://doi.org/10.1016/S0734-189X(87)80186-X.

Qian, Z., Liu, X., Tao, F., Zhou, T., 2020. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. Remote Sens. 12 (15), 2449. https://doi.org/10.3390/rs12152449.

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019. In: Basnet: Boundary-aware salient object detection, pp. 7479–7489.

Rahman, M.A., Wang, Y., 2016. Optimizing intersection-over-union in deep neural networks for image segmentation. International symposium on visual computing. Springer 234–244.

Rau, J.-Y., Lin, B.-C., 2011. Automatic roof model reconstruction from ALS data and 2D ground plans based on side projection and the TMR algorithm. ISPRS J. Photogramm. Remote Sens. 66 (6), S13–S27. https://doi.org/10.1016/j.isprsjprs.2011.09.001.

Tian, J., Krauß, T., D'Angelo, P., 2017. Automatic rooftop extraction in stereo imagery using distance and building shape regularized level set evolution. Int. Arch. Photogrammetry, Remote Sens. Spatial Inform. Sci. - ISPRS Arch. 42, 393–397. https://doi.org/10.5194/isprs-archives-XLII-1-W1-393-2017.

Torrey, L., Shavlik, J., 2010. Transfer learning, Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. In: Olivas, E. S., Guerrero, J.D.M., Martinez-Sober, M., Magdalena-Benedito, J.R., Serrano López, A.J. (Eds.), Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, pp. 242–264. https://doi.org/10.4018/978-1-60566-766-9.ch011.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep Learning for Computer Vision: A Brief Review. Comput. Intell. Neurosci. 2018, 1–13. https://doi.org/10.1155/2018/7068349.

Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., 2019. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing 338, 34–45. https://doi.org/10.1016/j.neucom.2019.01.103.

Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E., 2020. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. Remote Sens. Environ. 238, 111558. https://doi.org/10.1016/j.rse.2019.111558.

Yang, J., Guo, A., Li, Y., Zhang, Y., Li, X., 2019. Simulation of landscape spatial layout evolution in rural-urban fringe areas: a case study of Ganjingzi District. GIScience Remote Sens. 56 (3), 388–405. https://doi.org/10.1080/15481603.2018.1533680.

Yang, W., Luo, P., Lin, L., 2014. Clothing co-parsing by joint image segmentation and labeling. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3182–3189.

Yuan, Y., Chen, X., Wang, J., 2020. Object-contextual representations for semantic segmentation, Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. Springer 173–190.

Zhang, C., Lin, H., Chen, M., Zheng, X., Li, R., Ding, Y., 2017. A modelling system with adjustable emission inventories for cross-boundary air quality management in Hong Kong and the Pearl River Delta, China. Comput. Environ. Urban Syst. 62, 222–232. https://doi.org/10.1016/j.compenvurbsys.2016.12.004.

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., 2020. Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955.

Zhang, Z., Zhou, Y., Cui, J., Liu, H., 2014. An Improved Method of Building Rapid 3D Modeling Based on Digital Photogrammetric Technique. Chinese Conf. Image Graphics Technol. Springer 175–180. https://doi.org/10.1007/978-3-662-45498-5_20.

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, pp. 2881–2890.

Zhao, X., Wei, H., Wang, H., Zhu, T., Zhang, K., 2019. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. Sol. Energy 181, 510–518. https://doi.org/10.1016/j.solener.2019.01.096.

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. https://doi.org/10.1016/j.apenergy.2021.117132.

Zhong, Y., Han, X., Zhang, L., 2018. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 138, 281–294. https://doi.org/10.1016/j.isprsjprs.2018.02.014.

Zhu, R., Wong, M.S., Guilbert, É., Chan, P.-W., 2017. Understanding heat patterns produced by vehicular flows in urban areas. Sci. Rep. 7, 1–14. https://doi.org/10.1038/s41598-017-15869-6.

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. https://doi.org/10.1016/j.renene.2020.02.050.