Building Height Extraction From High-Resolution Single-View Remote Sensing Images Using Shadow and Side Information

Wanqi Xu, Zhangyin Feng, Qian Wan, Yakun Xie, Dejun Feng, Jun Zhu, and Yangge Liu

***Abstract*—Extracting building heights from single-view remote sensing images greatly enhances the application of remote sensing data. While methods for extracting building height from single-view shadow images have been widely studied, it remains a challenging task. The main reasons are as follows: 1) the traditional method for extracting shadow information exhibits low accuracy; and 2) the use of only shadow information to extract building height results in limited application scenarios. To solve the above problems, this article introduces building side and shadow information to complement each other, and proposes a building height extraction method from high-resolution single-view remote sensing images using shadow and side information. First, we propose the RMU-Net method, which utilizes multiscale features for the extraction of shadow and side information. This method aims to address issues related to pixel detail loss and imprecise edge segmentation, which result from significant scale differences within segmentation targets. Additionally, we employ the area threshold method to optimize the segmentation results, specifically to tackle small stray patches and holes, enhancing the overall integrity and accuracy of shadow and side information extraction. Second, we propose a method for building height extraction that integrates shadow and side information based on an enhanced proportional coefficient model. The accuracy of measuring building side and shadow lengths is improved by incorporating the fishing net method, informed by our analysis of the geometric relationships among buildings. Finally, we establish a dataset containing building shadow and side information from remote sensing images, and select multiple areas for experimental analysis. The results demonstrate a shadow extraction accuracy of 91.03% and a side extraction accuracy of 90.29%. Additionally, the average absolute error (MAE) for building height extraction is 1.22, while the average root mean square error is 1.21. Furthermore, the proposed method's validity and scalability are affirmed through experimental analyses of applicability and anti-interference performance in extensive areas.**

Manuscript received 9 November 2023; revised 14 January 2024; accepted 19 February 2024. Date of publication 1 March 2024; date of current version 18 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42301473, Grant 42171355, Grant and U22A20565; in part by the Postdoctoral Innovation Talents Support Program under Grant BX20230299; in part by the China Postdoctoral Science Foundation under Grant 2023M742884; and in part by the Natural Science Foundation of Sichuan Province of China under Grant 24NSFSC2264. *(Corresponding author: Yakun Xie.)*

Wanqi Xu, Yakun Xie, Dejun Feng, Jun Zhu, and Yangge Liu are with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China (e-mail: xwq1207@my.swjtu.edu.cn; yakunxie@my.swjtu.edu.cn; djfeng@swjtu.edu.cn; zhujun@swjtu.edu.cn; lyg99520@my.swjtu.edu.cn).

Zhangyin Feng is with the Faculty of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China (e-mail: fzy798@my.swjtu.edu.cn).

Qian Wan is with the Faculty of Foreign Languages, Southwest Jiaotong University, Chengdu 611756, China (e-mail: wanqian@my.swjtu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3372113

***Index Terms*—Building height extraction, deep learning, high-resolution remote sensing images, shadow, side.**

I. Introduction

In RECENT years, due to the rapid urbanization process, especially with the proposal and development of "Smart Cities," urban structures have been expanding both horizontally and vertically [1], [2], [3]. Building height is a crucial metric for vertical development [4], enabling various urban development applications, such as urban development assessment [5], [6], digital city construction [7], [8], and 3-D urban form analysis [9], [10], [11]. Furthermore, the development of remote sensing technology provides a more convenient method for obtaining surface information, and how to accurately obtain building height from remote sensing images constitutes a crucial component in the field of remote sensing research.

There are three primary methods for extracting building heights using remote sensing images. The first method involves building height extraction based on synthetic aperture radar (SAR) images, which includes single high-resolution SAR image extraction [12], [13], [14], measurement based on SAR data [15], [16], [17], and quantitative analysis of electromagnetic backscattering intensity [18], [19], [20]. For example, Tian et al. [21] extracted the heights of urban buildings using a backscattering model combined with multipolarization information from SAR images; Chen et al. [22] proposed a method for extracting building heights based on VHR SAR images; and Li et al. [23] achieved improved accuracy in building height estimation by iterating simulated InSAR images and matching them with measured SAR images. While using SAR data for building height extraction offers certain advantages, the postprocessing is intricate and requires the involvement of professionals to handle the data.

Second, as stereo imaging technology has advanced, methods for building height extraction based on stereo image pairs have been continuously proposed [24]. For instance, Liu et al. [25] employed the morphological reconstruction method to eliminate topographic height from the DSM to derive height above ground, or they normalized the DSM to represent building heights; Zhang et al. [26] introduced a method that relies on a stereo matching algorithm to impose roof contour constraints, and uses a DSM-based bottom elevation, to estimate building height from Gaofen-7 (GF-7) satellite images; and Wen et al. [27] proposed

a novel approach for generating building heights based on high-resolution multiview ZY-3 satellite images. However, the complex surface environment often leads to mutual obstruction among buildings, resulting in matching failures. Furthermore, the widespread utilization of multiview remote sensing sensors is hindered by their limited applicability and data acquisition challenges.

The third category comprises building height extraction methods based on single images [28], [29], [30]. For example, Xie et al. [31] proposed a multiscene building height extraction method based on shadows from high-resolution remote sensing images. Additionally, Liu et al. [32] introduced an automatic building height extraction framework based on high-resolution satellite remote sensing images, demonstrating that shadow rectification significantly enhances building height estimation accuracy. Moreover, Hao et al. [33] employed building shadows from a single RGB satellite image and satellite image metadata to estimate building height. However, many of the existing shadow detection methods are often reliant on specific assumptions and have limited applicability [34], [35], [36]. The single building height calculation method based on shadows still exhibits certain limitations [37], including its inability to calculate building height when shadows are obscured or obstructed. Furthermore, the intricate surface environment and the relatively low accuracy of existing shadow extraction methods directly contribute to the restricted accuracy of building height extraction.

The development of artificial intelligence technology has led to a continuous influx of deep-learning-driven building height extraction methods [38], [39], [40], [41], [42], [43]. For example, Liu et al. [44] introduced the convolutional neural network architecture IM2ELEVATION, which is dedicated to the task of extracting building height information from a single airborne or satellite optical image in a fully automatic manner; Chen et al. [47] devised an effective building height estimation method by combining photogrammetry and deep learning, leveraging Chinese multiview GF-7 images for high-resolution building height estimation; Lu et al. [45] developed a simple and efficient height-hierarchy guided dual-decoder network (HGDNet) for building height estimation; Yan and Huang [46] introduced a method for estimating building height from a single street view images using single-view metrology and deep learning; and Tang et al. [48] proposed the use of the Otsu algorithm and learning vector quantization neural network to extract building side information from high-resolution remote sensing images, calculate its length, and subsequently estimate the building's height based on a physical model of the images.

While deep learning [49], [50], [51], [52] can enhance certain auxiliary information, its accuracy is constrained. For instance, the accuracy of HGDNet is 80.12%, and the accuracy of building extraction based on instance segmentation [53] is 70%. Building height cannot be accurately obtained through a single prediction method alone. Furthermore, in remote sensing images, the sides of buildings are important sources of information, in addition to the tops of the buildings. However, existing methods do not effectively utilize this information. Finally, most high-resolution building height information extraction is limited to a small scale, and an effective building height extraction method on a large scale is lacking. Therefore, to address these issues, this article introduces a building height extraction method that incorporates shadow and side information. The main contributions are as follows.

1) To address the limited application scenarios of the shadow method, this article introduces a building height extraction method that combines shadow and side information to enable multiscene building height calculation. The experimental results demonstrate that the proposed method can achieve high-precision extraction not only in small areas but also at a large scale for building height extraction.
2) To address the challenges posed by complex ground object types and the low accuracy of shadow and side extraction, this article presents a building shadow and side extraction method that combines the RMU-Net model with multiscale features and the area threshold method. This approach enhances the network's ability to capture scale features and pixel details, resulting in improved accuracy and integrity in the extraction of shadow and side information. Additionally, a dataset of shadow and side information is established, which can facilitate the advancement of this field.

The rest of this article is organized as follows. Section II introduces the methods employed in this article, which encompass the shadow and side extraction method, extraction results optimization, shadow and side length calculation, and the building height inversion model. Section III presents the experiments and results, including data and evaluation criteria, and provides a detailed analysis of the accuracy, applicability, and anti-interference of the proposed method. Finally, Section IV concludes this article.

II. Methods

This article presents a method for extracting building height using a combination of shadow and side information, as illustrated in Fig. 1. Initially, we perform feature analysis of building shadows and sides. Based on this analysis, we establish an RMU-Net model with multiscale features to extract shadow and side information. Subsequently, we optimize the removal of stray patches and internal holes in the segmentation results by employing the area threshold method. Additionally, we use the fishing net method to enhance the accuracy of building shadow and side length calculation. Finally, leveraging the nonlinear relationship between shadow and side information, we propose an improved proportional coefficient model for building height inversion with shadow and side information complementing each other.

A. RMU-Net With Multiscale Features

1) Feature Analysis of Building Shadow and Side: Shadow refers to the dark areas created when objects on the ground block a light source due to their opacity, and any ground objects with height can form shadows. Particularly, in urban areas, the characteristics of building shadows in remote sensing images are notably intricate and variable, encompassing diverse shapes,

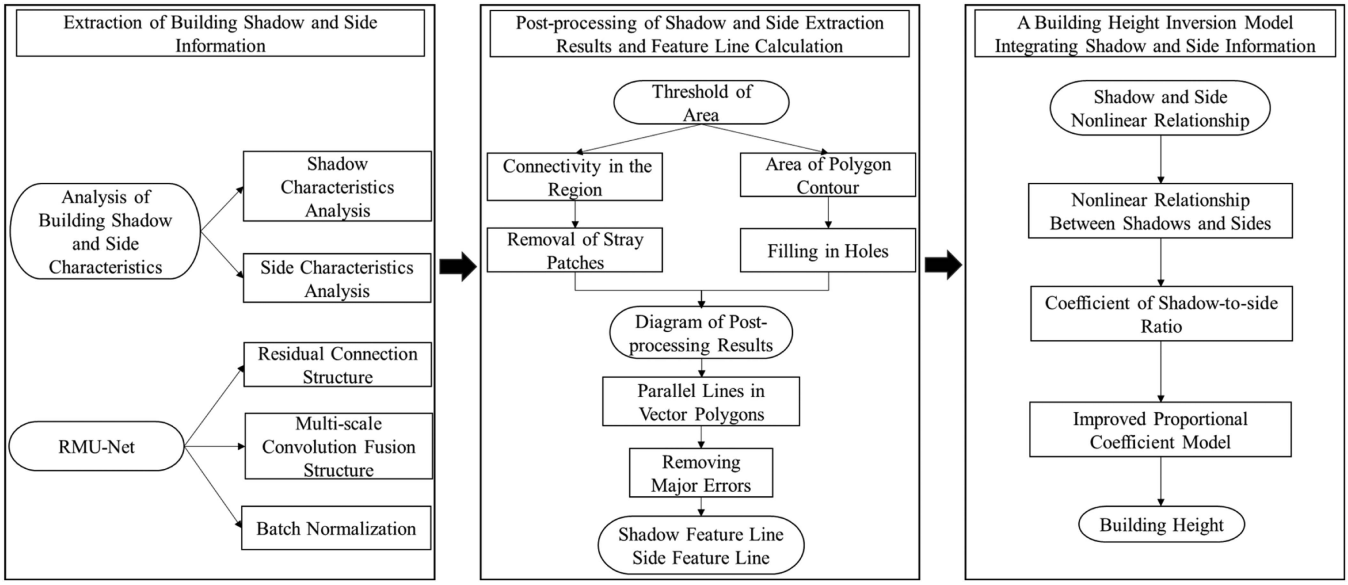

Fig. 1. Overall framework of the method in this article.

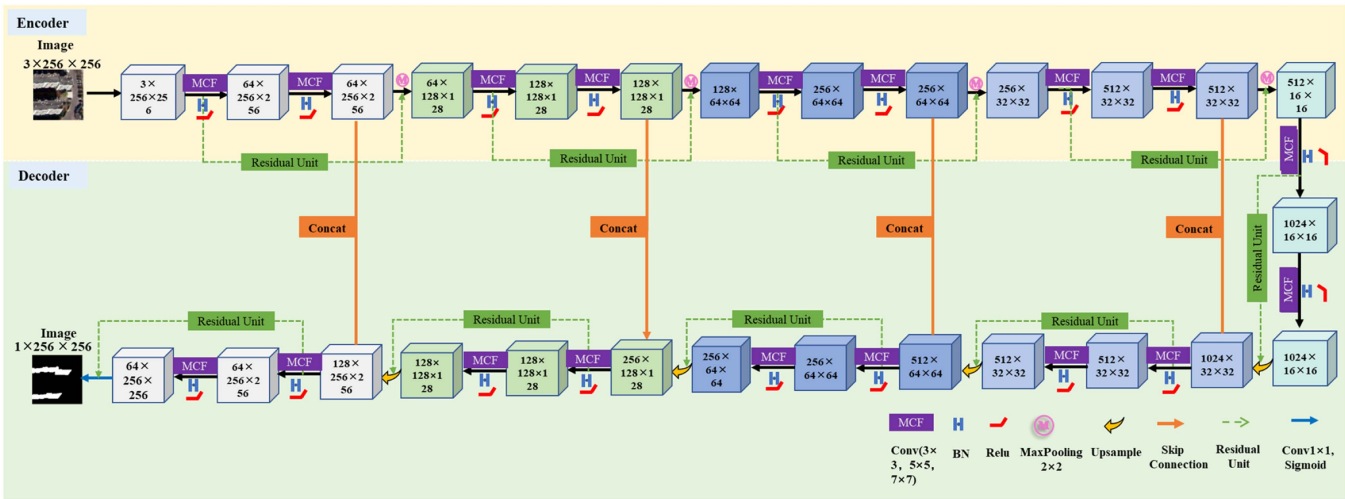

Fig. 2. Structure of RMU-Net.

intricate backgrounds, and significant scale disparities. The building side represents the facade of a real-world structure, typically having a polygonal shape that resembles its shadow. Moreover, the spectral characteristics of the building side exhibit greater complexity and variability, with noticeable variations in scale.

2) RMU-Net Design: This module enhances the network's capacity for global and local feature extraction through an aggregation structure, enabling the extraction of multiscale features. This article aims to enhance the precision of building shadow and side information extraction by initially focusing on the spectral characteristics of complex terrains, edge details, and scale variations. We introduce the RMU-Net, an advancement over the U-Net, which integrates residual connection and multiscale convolution fusion modules, depicted in Fig. 2. RMU-Net is structured as an end-to-end symmetrical model, comprising both encoding and decoding components. During the encoding phase, the skip-layer connections of the residual network facilitate a comprehensive capture of building shadow and side features, thus bolstering the network's feature extraction capabilities. Conversely, in the decoding phase, the multiscale fusion module (MCF) adeptly retrieves both shallow-level fine details and deep-level semantic information, augmenting the network's global and local feature extraction efficiency. Ultimately, the network's final aggregation structure ensures effective multiscale feature assimilation.

a) Residual connection structure: In this article, a residual connection [54] is employed in each encoder block of RMU-Net, and a direct connection channel is established between the input layer and the output layer. This channel directly conveys the input feature information to the output layer for activation and pooling, thereby ensuring the preservation of input data features' integrity. The inclusion of skip connections simplifies the network's training process.

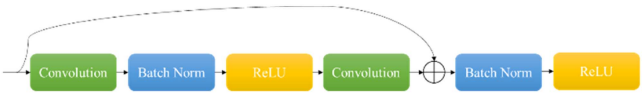

Fig. 3. Residual module.

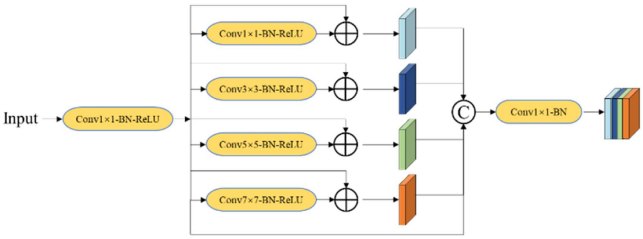

Fig. 4. Operation of multiscale convolution fusion.

Residual blocks typically require more than two layers to be effective. The residual concept can be understood as the summation of the input and the nonlinear transformation of the input, which enhances the network's feature extraction capability. The residual module is illustrated in Fig. 3.

b) MCF structure: As the network layers become deeper, the downsampling process can lead to the loss of fine-grained details, and the low-dimensional features of the images may be overlooked, resulting in a loss of detail in the extraction process and imprecise edge segmentation. To better aggregate the details of building shadow/side information and semantic information in remote sensing images, this article establishes a multiscale convolution fusion structure, as shown in Fig. 4. The structure convolves the output results of each layer with different receptive fields. Through the use of different receptive fields, semantic and detailed features at the same scale are obtained and aggregated to improve the accuracy of shadow/side information extraction from buildings.

B. Postprocessing Method Based on Area Threshold

Given the issues of stray patches and internal holes in the shadow/side extraction results obtained from the deep-learning model, this article proposes a postprocessing method based on the area threshold to enhance the completeness of the results. The method employs area connectivity to conduct initial post-processing on the segmentation results, calculates the size of each connected region, assesses the areas, and applies threshold constraints to eliminate stray spots. The contour area filling method is employed to fill the holes in the patches, with the main process depicted in Fig. 5.

1) Removal Stray Patches Based on Area Connectivity: A method for area connectivity removal is established to eliminate stray patches in the shadow and side images. First, all closed polygonal connected regions in the predicted binary images are examined, and the valid connected regions are labeled. Subsequently, the number of connected regions is counted, and the polygon area, maximum length, and width of each connected region are calculated. A dynamic area threshold is set based on these detection results. Finally, all polygonal areas are traversed to determine whether they are less than the area threshold. If an area is smaller than the threshold, it is considered a stray

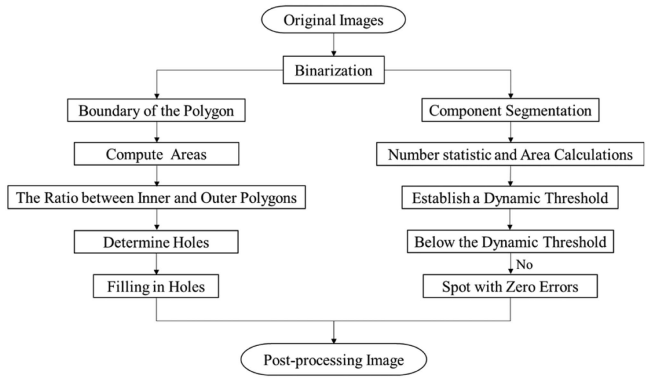

Fig. 5. Postprocessing of extraction results.

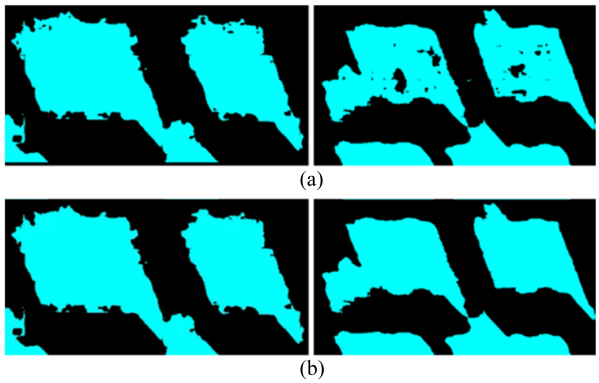

Fig. 6. Comparison before and after the postprocessing of extraction results. (a) Shadow exhibits holes and stray patches. (b) Post-processing of shadows.

patch and its pixel value is set to zero; otherwise, it is deemed an inerrant patch.

2) Filling in Holes Based on Contour Area: To fill holes in the contour, this article detects the contour of the target polygon based on the binary image, determines whether the contour contains one or more smaller polygons, calculates the area of the inner polygon and the outer polygon, and computes the ratio of the interior area to the exterior area as well as the size of the internal area. If the area of the inner polygon is small and the ratio to the area of the outer polygon is minimal, it can be determined that the polygon is a hole. Subsequently, we can trace the position of the hole and set its pixel value to 1 according to its contour. The comparison diagrams before and after the postprocessing of the extraction results are depicted in Fig. 6.

C. Calculation and Optimization of Shadow and Side Length

The length of the shadow/side feature line is a prerequisite for height calculation. Currently available methods for calculating the length of feature lines primarily include the pixel method, back calculation of area and circumference, computation method using the closest corner points, and the fishing net method. Among these methods, the fishing net method is the most advantageous; however, it faces challenges in handling complex building shadow and side shapes through traditional approaches. To address this issue, this article introduces an optimized algorithm for shadow and side length calculation, as depicted in Fig. 7.

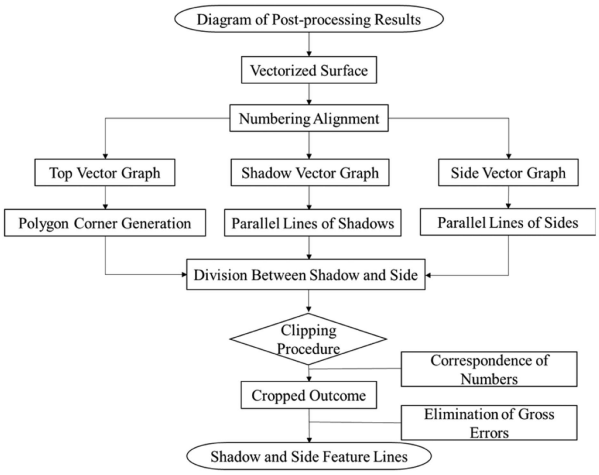

Fig. 7. Feature line extraction.

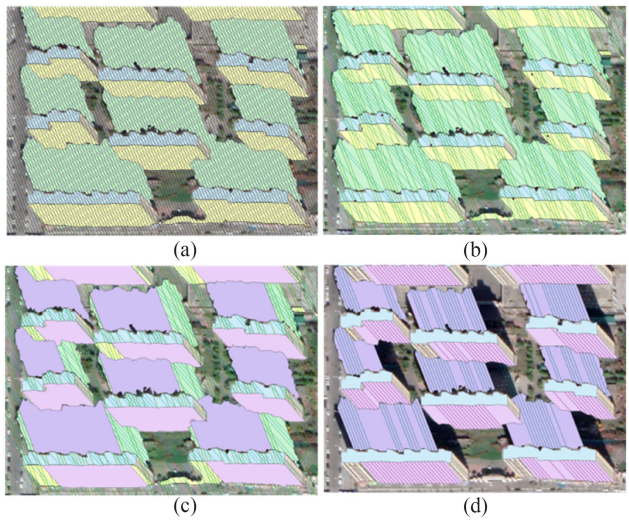

Fig. 8. Diagram of steps of the feature line extraction process. (a) Create parallel lines. (b) Generate corner points and segmentation lines. (c) Polygon segmentation. (d) Generated line feature.

1) *Feature line generation:* The shadow and side of the building exhibit a unified inclined direction, with each type of pixel sharing the same angle. In consideration of this, two sets of lines are generated in parallel with the inclined direction of the shadow or side vector surface. These lines are then overlaid and intersected with the shadow/side contour shapes to determine the length of feature lines. The primary process is as follows.
   a) The shadows create regularly spaced parallel lines originating from the direction of the sunlight, while the building's side produces a cluster of equally spaced parallel lines corresponding to the building's projection in the images.
   b) The polygon at the top of the building generates the corresponding corner point. By examining the pixel's external contour, we can determine the pixel width and obtain the segmentation line for the building's shadow or side.
   c) The vector contour is clipped along the boundary by the segmentation lines after overlaying pixel contours, corner points, and segmentation lines of the shadow/side layer. We then assess whether the polygon surpasses its corresponding corner points. If the polygon does not exceed these corner points, a new target contour and corresponding feature line are created, and any portion that exceeds the corner points is removed.
2) *Gross error elimination:* Gross errors may occur in the clipped line as a result of pixel contour deformation or segmentation result errors. The probabilistic interval for the value distribution is calculated based on the 3σ principle of normal distribution. If the feature line clipped along the pixel contour exceeds this interval, it is classified as a gross error and should be eliminated. If there are holes in the shadow/side vector surface of any building, it is advisable to remove these holes before extracting the feature lines to prevent interference from broken lines during the feature line length extraction process. Fig. 8 illustrates the steps of feature line extraction.

D. Dual-Feature Building Height Inversion Model

When shadows in the images are absent, damaged, obstructed by other ground objects, or overlap with each other, accurate information extraction becomes challenging based on semantic segmentation results, leading to exceptions in the length of the corresponding shadow feature line of the building. Similarly, similar issues may also arise on the building's sides. Therefore, in cases where information is missing or abnormal, the building height calculated using normal data can be automatically used as a supplement by searching for an abnormal proportional coefficient. This integration of shadow and side information helps reduce the error rate in building height estimation resulting from information gaps and enhances the universality of height estimation.

By conducting theoretical calculations and empirical analyses on the relationship between a building's shadow/side, the sun, and sensor azimuth, we observe a nonlinear relationship between a building's height, shadow, and side. As an example, when the sun and satellite are on the same side, model analysis results illustrate the complementary nature of the shadow and side. Through an analysis of the relationship between shadows and building sides, we can introduce a shadow/side proportional coefficient and propose an enhanced proportional coefficient model, which integrates shadow and side information for building height estimation. The key steps are illustrated in Fig. 9.

When the satellite and the sun are positioned on the same side, the geometric relationship among building height, building shadow, building side, sun, and satellite is illustrated in Fig. 10.

In Fig. 10, α represents satellite elevation, β represents solar elevation, θ represents satellite azimuth, γrepresents solar azimuth, ω is the angle between the shadow and the bottom of the building, CE is parallel to the building's horizontal line at its base, FH and AB represent building height H, FG corresponds to the length of the side feature line (L_{side}) in the image, BD represents the actual shadow length, and ED is the length of the

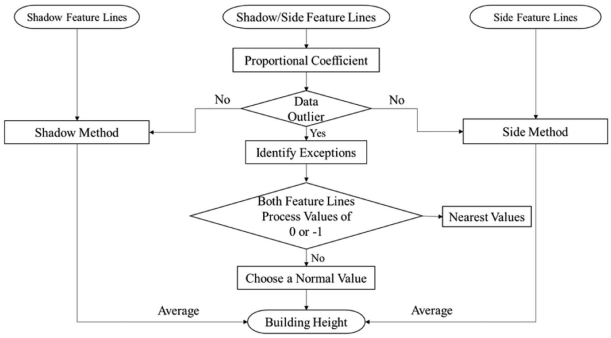

Fig. 9. Improved proportional coefficient model.

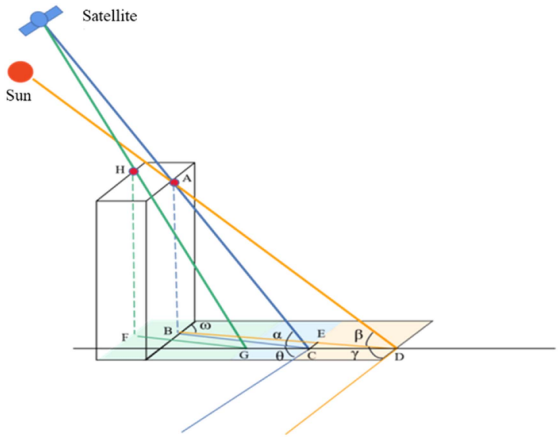

Fig. 10. Geometric relationship of building, sun, and satellite.

shadow feature line (L_{shadow}) in the image. According to the geometric relationship:

According to the trigonometric functions as follows:

$$AB = BC \times \tan\alpha \tag{1}$$

$$AB = BD \times \tan\beta = (BE + ED) \times \tan\beta \tag{2}$$

and as

$$\angle CBD = \theta - \gamma \tag{3}$$

according to the nature of parallel lines as

$$\angle BEC = \omega \tag{4}$$

according to the Sine Law as follows:

$$\frac{BE}{\sin(180^\circ - \omega - \theta + \gamma)} = \frac{BC}{\sin\omega} \tag{5}$$

$$FG = BC + \delta\,(\delta \to 0) \tag{6}$$

the building height H as

$$H = \frac{L_{\text{side}} \sin(180^\circ - \omega - \theta + \gamma)\tan\beta}{\sin\omega} + L_{\text{shadow}} \tan\beta. \tag{7}$$

Based on the formula and geometric relationship described above, it can be inferred that there is also a fixed proportional coefficient between the shadow and the side of the building, and the calculation formula is as follows.

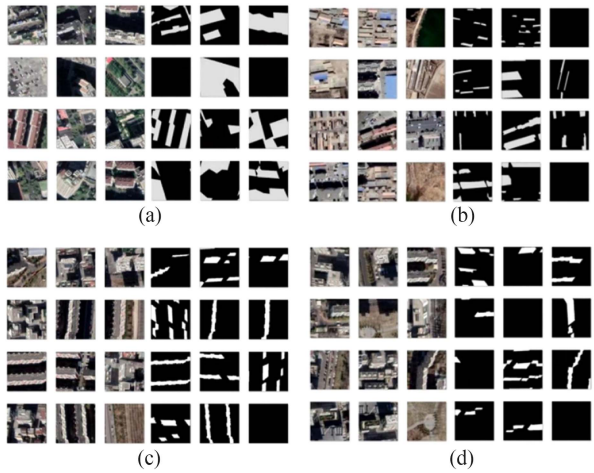

Fig. 11. Images and labels. (a) Training set for shadow analysis. (b) Test set for shadow evaluation. (c) Training set for side analysis. (d) Test set for side evaluation.

Substituting (5) and (6) into (1) and (2), we obtain

$$BC \times \left(\tan\alpha - \frac{\sin(180^\circ - \omega - \theta + \gamma)\tan\beta}{\sin\omega}\right) = ED\tan\beta \tag{8}$$

$$\frac{ED}{FG} = \frac{\sin\omega\tan\alpha - \sin(180^\circ - \omega - \theta + \gamma)\tan\beta}{\sin\omega\tan\beta}. \tag{9}$$

The proportional relationship between the shadow and the side of the building as

$$\frac{L_{\text{shadow}}}{L_{\text{side}}} = \frac{\tan\alpha}{\tan\beta} - \frac{\sin(180^\circ - \omega - \theta + \gamma)}{\sin\omega} = k. \tag{10}$$

Therefore, k is a constant coefficient that can be calculated when the angle is known. It is evident that within the same image, a fixed proportional coefficient exists between the building's shadow and its side.

III. Experimental Results and Analysis

A. Experimental Data and Evaluation Criteria

1) Experimental Area and Dataset: All the images used in this article are submeter-level high-resolution remote sensing images downloaded from Google Earth, specifically worldview-III satellite images. The dataset is extensive, encompassing various domestic and international areas. The shadow datasets include Alxa, Chongqing, Changchun, Hangzhou, Naqu, Chengdu, New York, Tokyo, Xinjiang, and Beijing. However, there are no open-source side datasets available; all side datasets were manually created and include Chengdu, Changchun, Shanghai, Hangzhou, Busan, Daejeon, and Seoul. In this article, we conducted experiments in the areas of Changchun, Shanghai, and Chengdu. All images in the dataset were cropped to 256 × 256 pixels. The dataset is divided into a training set and a verification set. The training set comprises 24 000 image sets with corresponding labels, while the test set consists of 6000 image sets with corresponding labels. A portion of the images is displayed in Fig. 11.

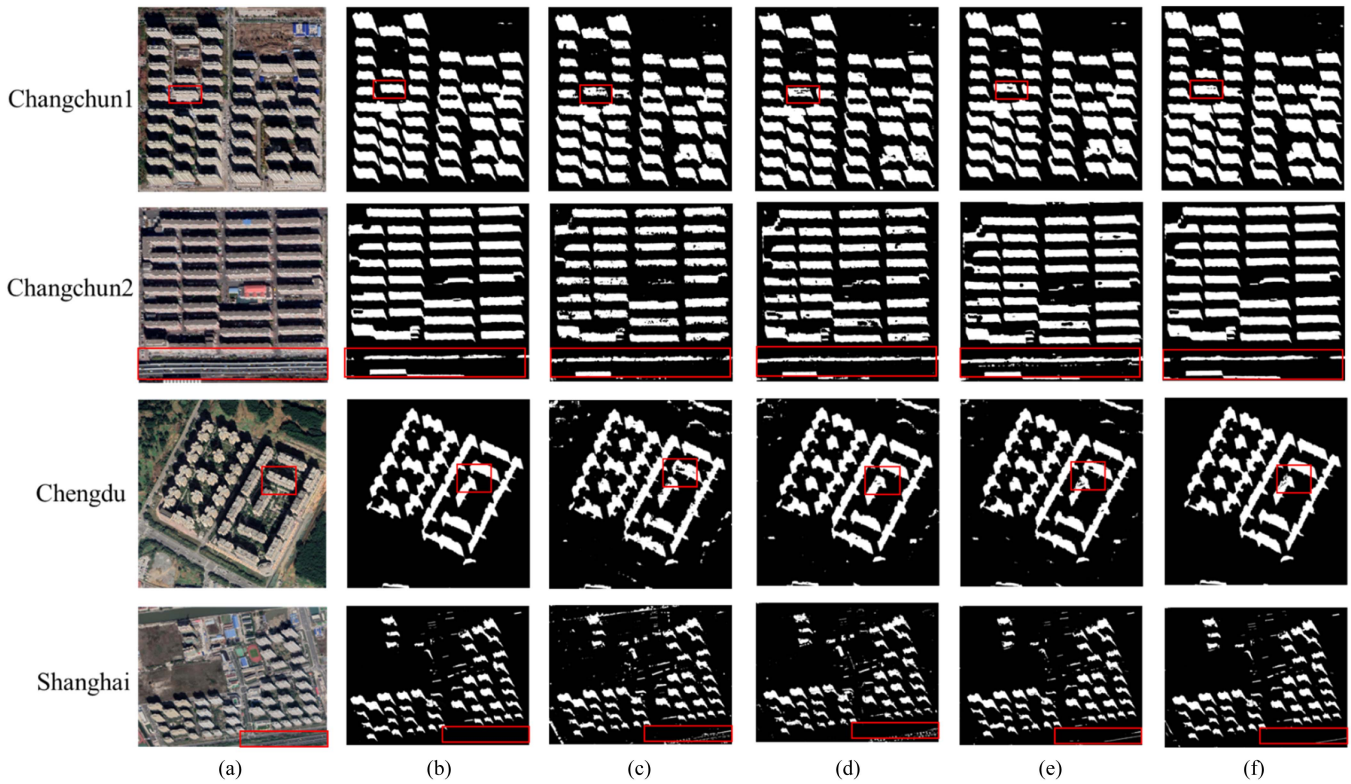

Fig. 12. Shadow extraction results on different networks. (a) Original image. (b) RMU-Net. (c) Res-Net. (d) U^2-Net. (e) ECANet. (f) MSASDNet.

2) Evaluation Indexes: In this article, we have selected four evaluation indexes to assess the accuracy of shadow and side extraction, namely Accuracy (Acc) (11), Intersection over Union (IoU) (12), Kappa coefficient (13), and F1 score (14). For building height extraction, we have chosen absolute error (15), relative error (16), and Root Mean Square Error (RMSE) (17), as the evaluation indexes. Absolute error quantifies the deviation between the observed value and the true value, relative error expresses the absolute error as a percentage of the true value, and RMSE measures the deviation between the predicted value and the true value.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{12}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{13}$$

$$F1_{\text{score}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

$$E_1 = b - a \tag{15}$$

$$E_2 = \frac{b - a}{a} \times 100\% \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{n} \left(Y_i - f\left(x_i\right)\right)^2}. \tag{17}$$

Regarding IoU and Acc, True Positive is defined as instances where the building shadow/side is correctly identified as such. False Positive (FP) occurs when the background class is mistakenly identified as building shadow/side. True Negative refers to correctly identifying the background class as background, while False Negative (FN) occurs when the building shadow/side is erroneously classified as background. In the context of Kappa statistics, p_o denotes the overall classification accuracy, whereas p_e represents the expected accuracy. For error metrics E_1 and E_2, b symbolizes the predicted height of buildings and a denotes the actual height. In the RMSE calculation, N signifies the total number of buildings, Y_i the actual height of the i building, and $f(x_i)$ is predicted height.

B. Analysis of Shadow/Side Extraction Results

1) Qualitative Analysis: Fig. 12 presents building shadow extraction results for three experimental areas. RMU-Net outperforms the other four networks in prediction accuracy, characterized by minimal FPs and FNs, smooth textures, distinct boundaries, and detailed precision. In contrast, Res-Net, U^2-Net, and ECANet [55] results are marked by multiple holes, mottled textures, increased FPs, fragmented patterns, and general disarray. MSASDNet's [56] outcomes feature fragmented patterns and less nuanced detail. Specifically, in the highlighted Changchun 1 and Chengdu areas, while Res-Net, U^2-Net, ECANet, and MSASDNet demonstrate numerous holes, RMU-Net maintains surface integrity with hole-free extractions. In similar road and shadow spectral areas like Changchun 2 and

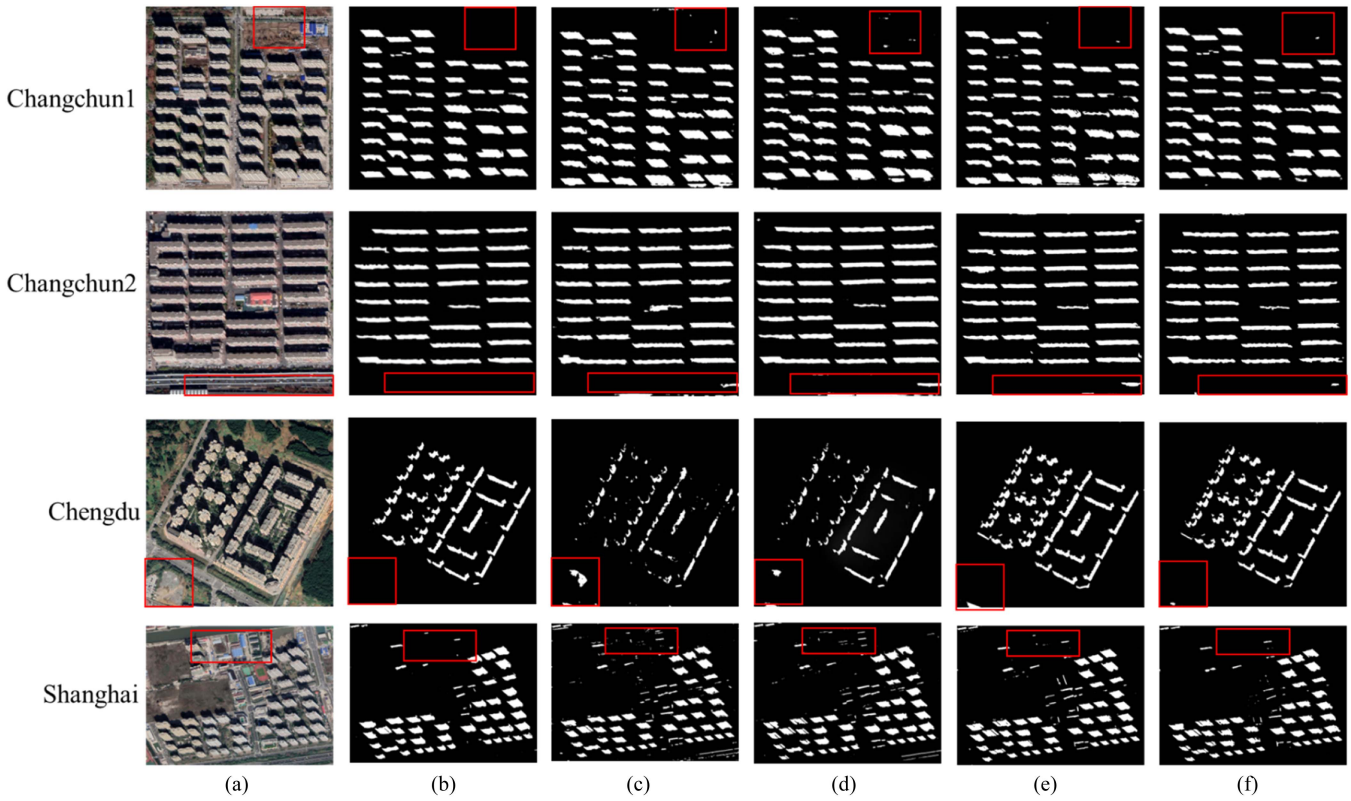

Fig. 13. Side extraction results on different networks. (a) Original image. (b) RMU-Net. (c) Res-Net. (d) U^2-Net. (e) ECANet. (f) MSASDNet.

Shanghai, all networks, except RMU-Net, exhibit FPs. Overall, RMU-Net shows superior performance in addressing FPs, negatives, holes, and fragmented shadow patterns. Its enhanced modules effectively reduce interference from nonshadow areas, preserving shadow region integrity.

Fig. 13 illustrates the side extraction results obtained by different networks across three experimental areas. RMU-Net's predictions exhibit a clean, coherent arrangement with minimal fragmentation, shapes closely resembling the originals, regular boundaries, a pristine surface, and complete interior patterns. In the Changchun area, Res-Net and U^2-Net provide relatively comprehensive predictions. However, in Chengdu, Res-Net's predictions lack integrity for irregular buildings, featuring disjointed sides, unclear boundaries, and scattered patterns. U^2-Net offers improved pattern completeness but shares similar issues with unclear boundaries and increased fragmentation. Moreover, Res-Net and U^2-Net introduce more fragmented small areas with mispredictions in the Shanghai region. ECANet and MSASDNet yield better predictions in the three experimental areas compared to Res-Net and U^2-Net, but they tend to produce more fragmented patterns and holes. Notably, within the highlighted regions (Changchun 1 and Changchun 2), only RMU-Net avoids mispredictions and provides accurate results, distinguishing it from other networks. In the highlighted Chengdu region, Res-Net, U^2-Net, ECANet, and MSASDNet incorrectly classify trees as building sides, while RMU-Net demonstrates superior discrimination without such misidentifications. In the highlighted Shanghai region, Res-Net and U^2-Net produce excessive fragmented small areas, resulting in clutter and reduced clarity. ECANet and MSASDNet exhibit unclear boundaries and fewer fragmented patterns, whereas RMU-Net maintains clear boundaries and explicit details, facilitating subsequent processing. These findings underscore the superiority of RMU-Net in side information extraction.

2) Quantitative Analysis: To better quantify and analyze the segmentation accuracy of the network proposed in this article, we conducted predictions on three experimental zones using Res-Net, U^2-Net, ECANet, and MSASDNet. The effectiveness was evaluated using accuracy metrics, and the results are presented in Table I. According to the table, our method achieved the highest segmentation accuracy among the four evaluation metrics. In terms of Acc, IoU, Kappa, and $F1$ for building shadow segmentation results, our method outperformed Res-Net, U^2-Net, ECANet, and MSASDNet by 0.17% to 1.65%, 0.22% to 4.49%, 0.45% to 3.68%, and 0.34% to 3.48%, respectively. For building side segmentation results, our method surpassed Res-Net and U^2-Net by 0.03% to 12.60%, 0.44% to 9.25%, 0.07% to 8.45%, and 0.21% to 8.89%, respectively. Based on the quantified accuracy assessment, it is evident that our network exhibits higher accuracy in various metrics compared to Res-Net, U^2-Net, ECANet, and MSASDNet, demonstrating superior feature extraction capabilities and greater applicability.

C. Analysis of the Building Height Extraction Results

1) Height Extraction in Changchun Area: In the Changchun experimental area, the original vector superposition images of the building tops and the extraction of feature lines are depicted

TABLE I
EVALUATION OF SEMANTIC SEGMENTATION ACCURACY

Experimental design	Experimental site	Network	Acc (%)	IoU (%)	Kappa (%)	*F*1 (%)
Building shadows	Changchun	Res-Net	91.07%	82.55%	82.79%	83.23%
		U^2-Net	90.74%	83.31%	84.08%	84.77%
		ECANet	91.32%	83.85%	84.96%	85.11%
		MSASDNet	91.45%	84.92%	85.82%	86.37%
		RMU-Net	**91.62%**	**85.14%**	**86.47%**	**86.71%**
	Chengdu	Res-Net	88.53%	79.64%	80.02%	81.65%
		U^2-Net	89.08%	81.17%	81.83%	82.19%
		ECANet	89.45%	81.73%	82.17%	82.96%
		MSASDNet	89.86%	82.04%	82.96%	83.74%
		RMU-Net	**90.15%**	**82.49%**	**83.41%**	**84.10%**
	Shanghai	Res-Net	89.68%	80.38%	82.26%	83.53%
		U^2-Net	90.23%	82.40%	83.34%	84.69%
		ECANet	90.75%	82.97%	84.01%	84.95%
		MSASDNet	90.94%	83.62%	84.65%	85.42%
		RMU-Net	**91.33%**	**84.87%**	**85.25%**	**86.01%**
Building sides	Changchun	Res-Net	91.86%	85.83%	86.32%	87.75%
		U^2-Net	92.07%	86.44%	87.89%	89.48%
		ECANet	92.11%	86.94%	88.15%	89.85%
		MSASDNet	92.15%	87.02%	88.92%	90.08%
		RMU-Net	**92.18%**	**87.63%**	**89.91%**	**90.23%**
	Chengdu	Res-Net	75.71%	70.37%	71.93%	72.35%
		U^2-Net	78.49%	74.93%	75.06%	76.63%
		ECANet	84.51%	78.68%	78.84%	78.59%
		MSASDNet	87.56%	79.18%	79.22%	80.15%
		RMU-Net	**88.31%**	**79.62%**	**80.38%**	**81.24%**
	Shanghai	Res-Net	88.47%	79.76%	81.25%	82.37%
		U^2-Net	89.15%	82.62%	84.74%	85.11%
		ECANet	89.36%	82.95%	84.77%	85.36%
		MSASDNet	89.97%	83.08%	84.96%	85.91%
		RMU-Net	**90.38%**	**83.56%**	**85.03%**	**86.12%**

The bold values are used to highlight the best-performing method, making it easier to compare with other methods.

Fig. 14. Vector superposition and feature line extraction of the Changchun area.

in Fig. 14. The pink areas represent the building tops, the regions with green parallel lines above the pink areas indicate the shadows of each building, and the areas with purple parallel lines below the pink regions represent the sides of each building. The numbers of the building tops, the shadows of the buildings, and the sides of the buildings correspond one-to-one.

The data for the Changchun experimental area were captured on October 27, 2020, with a resolution of 0.22 m. The results were computed using the improved proportional coefficient model of the combination method based on the shadow/side feature lines. We examined the published data for actual building heights, and found that the height of building 1 is 60.15 m. Subsequently, we calculated the proportional coefficient based on the feature line length, identified any abnormal coefficients, and computed the heights for each building.

As shown in Fig. 15, through the combination method's calculations, in the area, there were 26 buildings with absolute height errors between 0 and 1 m, 15 buildings between 1 and 2 m, 8 buildings between 2 and 3 m, 3 buildings between 3 and 5 m, and no buildings above 5 m. Out of the 51 buildings, 46 had height relative errors of less than 5%, accounting for 90.20% of the total. The maximum relative error value was not more than 7%, and the RMSE was 1.41. Given the above results, the overall accuracy meets the required standards, with a stable error distribution and no significant deviations, indicating the method's practical value.

2) Height Extraction in Shanghai Area: The distribution of buildings in the Shanghai experimental area is relatively regular,

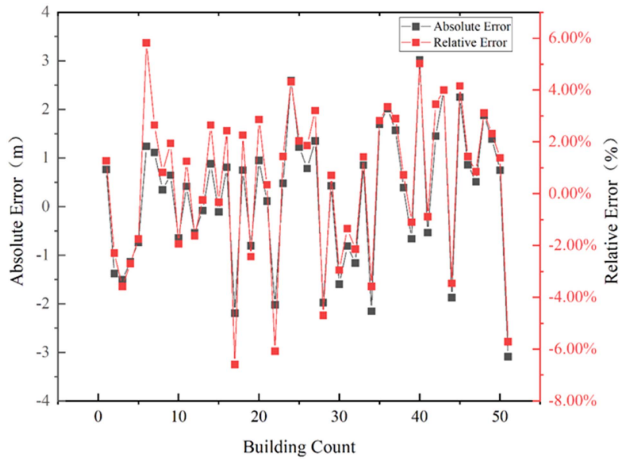

Fig. 15. Building height estimation accuracy of Changchun area.

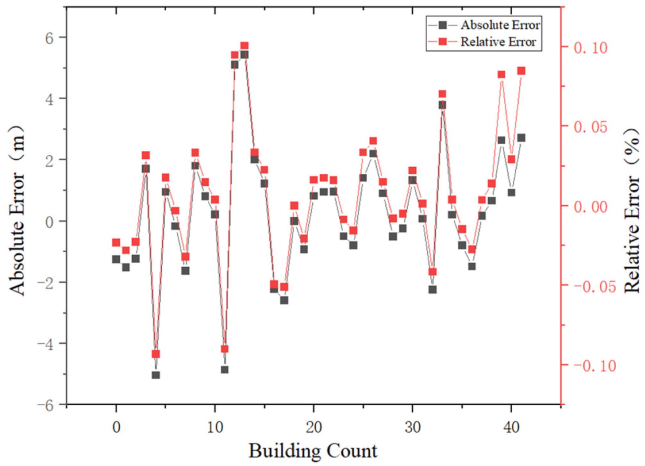

Fig. 17. Building height estimation accuracy of Shanghai area.

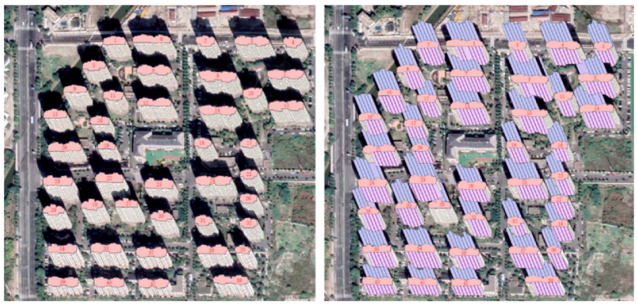

Fig. 16. Vector superposition and feature line extraction of Shanghai area.

Fig. 18. Vector superposition and feature line extraction of Chengdu area.

with low density, making it easy to extract building heights. Fig. 16 displays the original images of the building tops' vector superposition and the feature line extraction in the Shanghai experimental area.

The data for the Shanghai experimental area were collected on 14 November 2021, with a resolution of 0.12 m. Building heights were determined using the combination method in conjunction with shadow/side feature lines. Real-world data were obtained by referencing the actual site, and the true height of building 1 was measured at 54.20 m.

The accuracy of the estimated building height in this area, as shown in Fig. 17, is such that the absolute value of the absolute error is between 0 and 1 m for 20 buildings, 1 and 2 m for 10 buildings, 2 and 3 m for 7 buildings, and 3 and 5 m for 5 buildings, with no buildings exceeding 5 m in height. A remarkable 93.33% of height results exhibit relative errors below 5%, and the RMSE is measured at 1.12. The error curve displays minimal fluctuations, without any erratic jumps. Notably, for building number 27, while the side feature lines are absent, the shadow feature line length remains normal. Consequently, following the methodology outlined in this article, the shadow-based height result is deemed the final outcome for this specific building, effectively mitigating the impact of missing information on the overall results. The buildings in this area are meticulously arranged in a regular distribution, contributing to a high degree of semantic segmentation accuracy. Particularly in scenarios where buildings are evenly spaced, and shadow side information remains intact, and the precision of building height extraction is significantly improved.

3) Height Extraction in Chengdu Area: The Chengdu area is characterized by a multitude of buildings with irregular exterior shapes. In Fig. 18, the original remote sensing image and the vector superposition images of the building tops, each labeled with a building number, are presented on the left. On the right, you can see the image showing the extraction of feature line lengths during data processing.

The imagery for the Chengdu experimental area was captured on 29 February 2020, featuring a resolution of 0.51 m. Ground truth data, acquired by onsite positioning and floor count verification, revealed that the height of building 1 is 33.50 m. These heights were calculated using the refined scaling factor model.

As shown in Fig. 19, the results obtained for building heights in the region indicate that 24 buildings exhibit absolute errors within the 0–1 m range, while 8 buildings have errors ranging from 1 to 2 m. Importantly, all absolute errors in this area remain below 3 m. The percentage of height results displaying relative errors below 5% stands at 87.9%, accompanied by an RMSE value of 1.10. A closer examination of the accuracy plot reveals a consistently stable distribution of errors for the joint method.

D. Comparative Analysis of Combination Method and Single Method

To verify the improvement in building height extraction achieved by the combination method, and to compare it with the performance of the shadow method or side method alone, we compared the building height obtained using the combination

TABLE II
COMPARISON OF THE BUILDING HEIGHT ESTIMATION ACCURACY THROUGH THE COMBINATION METHOD AND THE SINGLE METHOD

Method	MAE (m)	MRE (%)	RMSE (m)
Combination method	0.84	3.34%	1.06
Shadow	19.08	26.88%	41.64
Side	6.41	31.03%	13.91

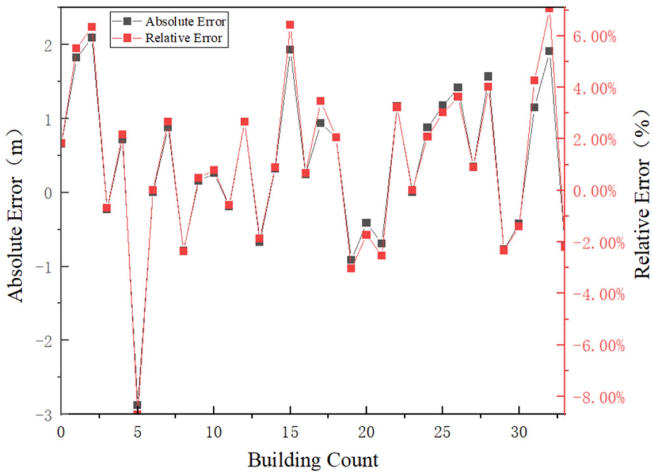

Fig. 19. Building height estimation accuracy of Chengdu area.

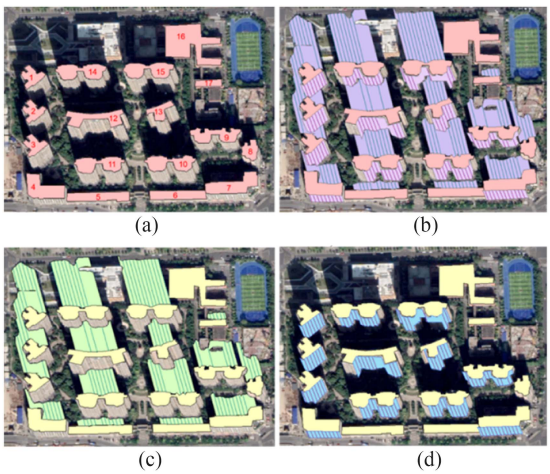

Fig. 20. Feature line extraction of Chengdu area. (a) Images and quantitative data for Chengdu. (b) Shadow and side feature lines. (c) Shadow feature line. (d) Side feature lines.

method in a specific area with that obtained using either the shadow method or side method separately. As shown in Fig. 20, the selected experimental area was a residential area in Chengdu characterized by high building density, diverse building types, and irregular facades, making it a representative example of complex scene scenarios.

The Chengdu experimental area was imaged on 29 February 2020, with a resolution of 0.51 m. Abnormal coefficients were identified through the experimental steps of the model proposed in this article, and the abnormal buildings were examined to determine whether there was a normal value in either the shadow or side information. If a normal value was found in either dataset, that value was used. If neither dataset contained a normal value, a nearby value was assigned. As seen in Fig. 20, neither shadow nor side information was available for building 16. Therefore, the height of building 17 was assigned to building 16 based on the area index calculation. The actual height data for the Chengdu experimental area were obtained through field measurements, with the height of building 1 measured at 98.04 m. Table II presents a comparison of building height estimation accuracy between the combination method and the single method.

It can be observed from Table II that the building height results obtained by the method proposed in this article are stable in this area, with absolute errors consistently below 3 m. In contrast, the shadow method or the side method alone leads to larger errors. This is because exceptions arise in the feature lines of the corresponding buildings due to the longitudinal adhesion of shadows or the absence of shadow and side information. Consequently, essential foundational information is missing when using the traditional height inversion model, resulting in incorrect building heights and fluctuating errors.

In summary, the proposed method is suitable for buildings in complex scenes. The accuracy of the combination method is 82.35%, surpassing that of the shadow method is 64.71% and the side method is 70.59%. This indicates that the height inversion accuracy of the combination method in this article is superior to that of the shadow method or the side method used independently. By employing optical remote sensing images, we can attain higher accuracy and applicability in the field of building height extraction.

E. Applicability Analysis

To assess the applicability of the proposed method in large areas, we conducted experiments in the expansive region of Changchun and analyzed the statistical results of height extraction within this large area. Fig. 21 illustrates images from each stage of the experimental process.

Given that it is not possible to obtain the true height of all buildings within the extensive area, the following section comprises a statistical analysis of the results from the height extraction experiments conducted in this large area, displayed at various levels. The large area contains a total of 10 691 buildings, and their distribution by height level is illustrated in Fig. 22.

The building heights are categorized into five levels, with corresponding colors ranging from light green to dark green. Table III presents a statistical summary of the building height extraction. It is evident from both the figure and the table that buildings taller than 100 m constitute a smaller proportion, accounting for only 4.17%. Those with heights between 60 and 100 m make up 16.77%, while buildings with heights ranging

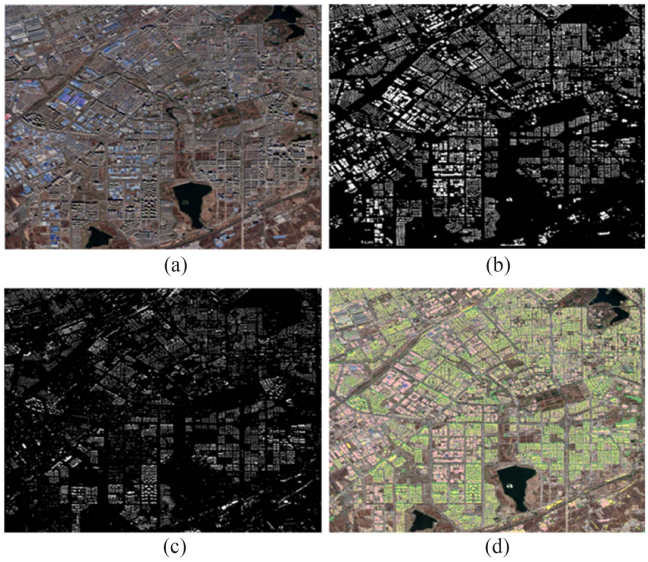

Fig. 21. Each stage of experimental process in the large area. (a) Original image. (b) Chart for predicting shadows. (c) Chart for predicting sides. (d) Diagram of vectorized superposition.

Fig. 22. Building height level distribution in the large area.

TABLE III
BUILDING HEIGHT EXTRACTION LEVELS

Classification of building heights (m)	Building count	Account (%)
0–10	1615	15.11%
10–30	3976	36.25%
30–60	2861	26.76%
60–100	1793	16.77%
>100	446	4.17%

from 0 to 60 m represent the majority at 78.12%. In terms of height distribution, the likelihood of incorrect extraction is higher within the 0–10 m range and for buildings exceeding 100 m in height. The instances of 0 m occur when certain buildings lack shadow or side information, have a small top area, and upon inspection of the exception, it becomes impossible to assign a height value. Therefore, in such cases, building height remains unextracted. Heights exceeding 100 m may be attributed to exceptionally tall structures or heights calculated using abnormal values derived from the feature lines. Assuming

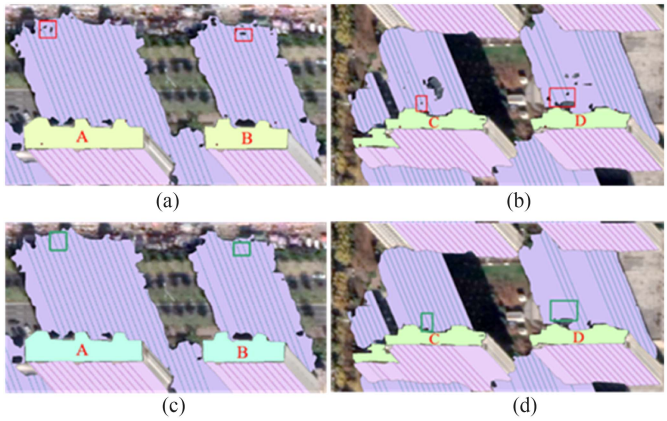

Fig. 23. Influences of holes on feature lines. (a) Exhibits holes. (b) Filling in holes.

that both the 0–10 m and greater than 100 m height categories are considered as part of the error rate, the accuracy rate in the large area reaches 79.78%. When we compare the height results in the sampled area to the actual building heights, the quantitative analysis indicates that even in areas with complex building distributions, the accuracy of the proposed method remains at 82.35%. Therefore, it is reasonable to estimate that the accuracy in larger areas will be at least 80%.

F. Anti-Interference Analysis

To further validate the anti-interference performance of the combination method, we utilize the approach described in Section II-B to enhance the initial segmentation outcomes for shadows and sides. Given that the preliminary semantic segmentation results for the sides are relatively comprehensive with few gaps, we will focus exclusively on the shadow areas for display and analysis in the following sections. We have selected the postprocessed results of preliminary building shadow extraction for a specific region in Changchun, emphasizing the improvements made to address any incomplete portions. As evident from the images in both the upper and lower sections of Fig. 23, the presence of gaps results in incomplete feature line extraction. In particular, the broken feature lines at these gaps fail to accurately connect to the intended contour boundary. Such incomplete feature lines can significantly affect the precision of building height calculations.

In the figure above, the red boxes mark the locations of the holes. The corresponding feature lines are not entirely connected to the entire vector area boundary. The green boxes indicate the positions where the holes have been repaired, and the feature lines are now seamlessly connected to the boundary. Consequently, the height measurements in this area are considerably more reliable. Table IV presents a comparison of building heights before and after the restoration process in the respective area.

In the table, the second and third columns display the building height and absolute error obtained without hole filling, while the last two columns show the building height and absolute error after hole filling. Since the displayed area contains only shaded

TABLE IV
BUILDING HEIGHT EXTRACTION LEVELS

	Actual height (m)	Building height (m)	Absolute error (m)	Building height (m)	Relative error (m)
A	72.04	69.73	−2.31	71.36	-0.68
B	72.04	70.52	−1.52	71.21	-0.83
C	60.15	57.39	−2.76	58.89	-1.26
D	60.15	58.26	−1.89	59.64	-0.51

regions with holes, building height is calculated solely using the traditional proportional coefficient model of the shadow method and compared with the actual building heights. The table illustrates that the absolute error in building height estimation is reduced after hole filling, resulting in estimates that are closer to the true values and a higher level of height extraction accuracy.

IV. CONCLUSION

Although there has been extensive research on building height extraction based on shadow information from single-view remote sensing images, achieving accuracy and applicability poses significant challenges due to the complexity of the scenes. In this article, we propose a method for extracting building height from high-resolution remote sensing images, which combines shadow and side information. The effectiveness of this framework is verified through experiments in various areas, along with detailed experimental analysis. First, to address the issue of low accuracy in shadow information extraction by traditional methods, we propose a building shadow and side extraction method that combines multiscale features from the RMU-Net and utilizes area thresholds. Additionally, we establish a semantic segmentation dataset for remote sensing images to enhance the accuracy of shadow and side information extraction. Second, to address the limited applicability of building height extraction using shadow information alone, we develop a building height calculation model that combines both shadow and side information. Additionally, we introduce a fishing net method to calculate the lengths of building shadows and sides, significantly enhancing the accuracy of building height extraction. Finally, the effectiveness of the method presented in this article is demonstrated through several experimental analyses. The superiority of the proposed method is established by comparing it with the shadow or side methods individually. Furthermore, applicability and anti-interference analyses confirm the scalability of the proposed method. However, despite advancements in deep-learning technology, there is still room for further improvement in the accuracy of shadow and side information extraction. Given the complexity of remote sensing image environment, this article represents an initial breakthrough from traditional shadow methods. In the future, we aim to explore broader application scenarios and develop real 3-D digital city models using this technology.

REFERENCES

[1] Y. Cao and X. Huang, "A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities," *Remote Sens. Environ.*, vol. 264, 2021, Art. no. 112590.

[2] M. Qomita and S. R. Giyarsih, "Smart city assessment using the Boyd Cohen smart city wheel in Salatiga, Indonesia," *Geo J.*, vol. 88, pp. 479–492, 2023.

[3] H. Huang et al., "Estimating building height in China from ALOS AW3D30," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 146–157, 2022.

[4] K. Jiang, X. Guan, L. Sun, and H. Wan, "Research on the design of urban three-dimensional reconstruction system for smart cities," in *Proc. IEEE 20th Int. Conf. High Perform. Comput. Commun.; IEEE 16th Int. Conf. Smart City; IEEE 4th Int. Conf. Data Sci. Syst.*, 2018, pp. 1656–1660.

[5] S. Shareef, "The impact of urban morphology and building's height diversity on energy consumption at urban scale: The case study of Dubai," *Build Environ.*, vol. 194, 2021, Art. no. 107675.

[6] P. S. Prakash and B. H. Aithal, "Digital building-height preparation from satellite stereo images," *Photogrammetry Eng. Remote Sens.*, vol. 87, pp. 557–566, 2021.

[7] S. Qian, Z. Chang, X. Liu, and Y. Wang, "Three-dimensional landscape modeling of Quancheng square in Jinan city," *IOP Conf. Ser. Mater. Sci. Eng.*., vol. 563, 2019, Art. no. 042013.

[8] M. Zhao and J. Wang, "A new method of feature line integration for construction of DEM in discontinuous topographic terrain," *Environ. Earth Sci.*, vol. 81, 2022, Art. no. 397.

[9] Q. Cao, Q. Luan, Y. Liu, and R. Wang, "The effects of 2D and 3D building morphology on urban environments: A multi-scale analysis in the Beijing metropolitan region," *Build Environ.*, vol. 192, 2021, Art. no. 107635.

[10] L. Yang, X. Yang, H. Zhang, J. Ma, and H. Zhu, "Urban morphological regionalization based on 3D building blocks—A case in the central area of Chengdu, China," *Comput. Environ. Urban Syst.*, vol. 94, 2022, Art. no. 101800.

[11] B. Wu et al., "A first Chinese building height estimate at 10m resolution (CNBH-10 m) using multi-source earth observations and machine learning," *Remote Sens. Environ.*, vol. 291, 2023, Art. no. 113578.

[12] S. Yao, S. Montazeri, Y. Wang, and X. Zhu, "Automatic registration of a single SAR image and GIS building footprints in a large-scale urban area," *ISPRS J. Photogrammetry Remote Sens.*, vol. 170, pp. 1–14, 2020.

[13] Y. Sun, L. Mou, Y. Wang, and X. Zhu, "Bounding box regression network for building height retrieval using a single SAR image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 56–59.

[14] Y. Sun, L. Mou, Y. Wang, S. Mobtazeri, and X. Zhu, "Large-scale building height retrieval from single SAR imagery based on bounding box regression networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 79–95, 2022.

[15] Y. Shi, R. Bamler, Y. Wang, and X. Zhu, "Sar tomography at the limit: Building height reconstruction using only 3-5 tandem-x bistatic interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8026–8037, Nov. 2020.

[16] L. Liu, W. Zhou, and M. Gutierrez, "Mapping tunneling-induced uneven ground subsidence using Sentinel-1 SAR interferometry: A twin-tunnel case study of downtown Los Angeles, USA," *Remote Sens.*, vol. 15, 2022, Art. no. 202.

[17] L. Newman, F. Weissgerber, A. Plyer, and É. Colin, "Advantages of polarimetry and interferometry for semantic segmentation of urban SAR images with consideration of the layover," in *Proc. Joint Urban Remote Sens. Event*, 2023, pp. 1–4.

[18] X. Li et al., "Fourfold bounce scattering-based reconstruction of building backs using airborne array TomoSAR point clouds," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1937.

[19] V. Letsios, T. Faraslis, and D. Stathakis, "Monitoring building activity by persistent scatterer interferometry," *Remote Sens.*, vol. 15, 2023, Art. no. 950.

[20] M. M. Omati and M. H. Bastani, "Building height reconstruction based on propagator and generalized maximum entropy—TomoSAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Feb. 2023, Art. no. 4002304.

[21] Y. Tian, S. Wang, Y. Zhou, W. Liu, and C. Lin, "Urban building height estimation from radarsat 2 imagery, a case study in Beijing, China," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1066–1069.
[22] J. Chen, C. Wang, H. Zhang, B. Zhang, and F. Wu, "Geometrical characteristics based building height extraction from VHR SAR imagery," in *Proc. Prog. Electromagn. Res. Symp.*, 2017, pp. 519–523.
[23] W. Li et al., "An improved iterative simulation and matching scheme for building height retrieval from SAR image," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Jul. 2023, Art. no. 4007905.
[24] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.
[25] C. Liu, X. Huang, D. Wen, H. Chen, and J. Gong, "Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery," *Remote Sens. Lett.*, vol. 8, no. 9, pp. 907–916, 2017.
[26] C. Zhang, Y. Cui, Z. Zhu, S. Jiang, and W. Jiang, "Building height extraction from GF-7 satellite images based on roof contour constrained stereo matching," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1566.
[27] D. Wen, X. Huang, A. Zhang, and X. Ke, "Monitoring 3D building change and urban redevelopment patterns in inner city areas of Chinese megacities using multi-view satellite imagery," *Remote Sens.*, vol. 11, 2019, Art. no. 763.
[28] G. Zhou and H. Sha, "Building shadow detection on ghost images," *Remote Sens.*, vol. 12, 2020, Art. no. 679.
[29] B. Wu, H. Huang, and Y. Zhao, "Utilizing building offset and shadow to retrieve urban building heights with ICESat-2 photons," *Remote Sens.*, vol. 15, 2023, Art. no. 3786.
[30] X. Zhou and S. W. Myint, "Shadow pattern-enhanced building height extraction using very-high-resolution image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 180–190, Nov. 2022.
[31] Y. Xie, D. Feng, S. Xiong, J. Zhu, and Y. Liu, "Multi-scene building height estimation method based on shadow in high resolution imagery," *Remote Sens.*, vol. 13, no. 15, 2021, Art. no. 2862.
[32] R. Liu, H. Zhang, K. H. A. Yip, J. Ling, Y. Lin, and H. Huang, "Automatic building height estimation with shadow correction over heterogeneous compact cities using stereo Gaofen-7 data at sub-meter resolution," *J. Building Eng.*, vol. 69, 2023, Art. no. 106283.
[33] H. Hao, S. Naireddy, and E. Bartusiak, "Building height estimation via satellite metadata and shadow instance detection," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 11729, pp. 175–190, 2021.
[34] H. Zhang, "Research on buildings shadow detection method and height inversion with high resolution sensed image," Ph.D. dissertation, Fac. Geosci. Environ. Eng., Southwest Jiaotong Univ., Chengdu, China, 2017.
[35] B. Cai, Z. Shao, X. Huang, X. Zhou, and S. Fang, "Deep learning-based building height mapping using Sentinel-1 and Sentienl-2 data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 122, 2023, Art. no. 103399.
[36] Q. Li et al., "3DCentripetalNet: Building height retrieval from monocular remote sensing imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 120, 2023, Art. no. 103311.
[37] Y. Zhao et al., "Combining ICESat-2 photons and Google Earth Satellite images for building height extraction," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 117, 2023, Art. no. 103213.
[38] Y. Xie et al., "An omni-scale global-local aware network for shadow extraction in remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 193, pp. 29–44, 2022.
[39] Y. Xie et al., "Damaged building detection from post-earthquake remote sensing imagery considering heterogeneity characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 4708417.
[40] L. Li, X. Huang, and J. Gong, "Deep neural network for remote-sensing image interpretation: Status and perspectives," *Nat. Sci. Rev.*, vol. 6, pp. 1082–1086, 2019.
[41] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.
[42] A. AI-Habashna, "Building height estimation using street-view images, deep-learning, contour processing, and geospatial data," in *Proc. 18th Conf. Robots Vis.*, 2021, pp. 103–110.
[43] Z. Xu, F. Zhang, Y. Wu, Y. Yang, and Y. Wu, "Building height calculation for an urban area based on street view images and deep learning," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 4, pp. 892–906, 2023.
[44] C. Liu, V. Krylov, P. Kane, G. Kavanagh, and R. Dahyor, "IM2ELEVATION: Building height estimation from single-view aerial imagery," *Remote Sens.*, vol. 12, 2020, Art. no. 2719.
[45] C. Lu et al., "HGDNet: A height-hierarchy guided dual-decoder network for single view building extraction and height estimation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 758–761.
[46] Y. Yan and B. Huang, "Estimation of building height using a single street view image via deep neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, pp. 83–98, 2022.
[47] P. Chen et al., "Leveraging Chinese GaoFen-7 imagery for high-resolution building height estimation in multiple cities," *Remote Sens. Environ.*, vol. 298, 2023, Art. no. 113802.
[48] H. Tang, L. Du, C. Wang, P. Li, S. Ge, and D. Meng, "Research on building height extraction method from high-resolution image," *J. Phys. Conf. Ser.*, vol. 1961, 2021, Art. no. 012062.
[49] Y. Xie et al., "An enhanced relation-aware global-local attention network for escaping human detection in indoor smoke scenarios," *ISPRS J. Photogrammetry Remote Sens.*, vol. 186, pp. 140–156, 2022.
[50] Y. Xie et al., "Clustering feature constraint multiscale attention network for shadow extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 4705414.
[51] Y. Xie et al., "Refined extraction of building outlines from high-resolution remote sensing imagery based on a multifeature convolutional neural network and morphological filtering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1842–1855, Apr. 2020.
[52] H. Chen et al., "Slice-to-slice context transfer and uncertain region calibration network for shadow detection in remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 203, pp. 166–182, 2023.
[53] F. B. Bagic, A. A. Kindriroglu, M. Yalcin, U. Uyan, and M. U. Ozturk, "Building height prediction with instance segmentation," in *Proc. 14th Int. Conf. Elect. Electron. Eng.*, 2023, pp. 1–4.
[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[55] X. Fang, X. He, L. Wang, and J. Shen, "Robust shadow detection by exploring effective shadow contexts," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 2927–2935.
[56] D. Liu, J. Zhang, Y. Wu, and Y. Zhang, "A shadow detection algorithm based on multiscale spatial attention mechanism for aerial remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Aug. 2021, Art. no. 6003905.

Wanqi Xu received the B.S. degree in surveying and mapping engineering from the School of Civil Engineering and Transportation, Shandong Jiaotong University, Jinan, China, in 2022. She is currently working toward the master's degree in surveying and mapping engineering with the Southwest Jiaotong University, Chengdu, China.

Her research interests include remote sensing image processing and GIS.

Zhangyin Feng received the B.S. degree in electrical and intelligent control engineering of science and technology from the Heilongjiang University, Harbin, China, in 2021. He is currently working toward the master's degree in electronic and information engineering with the Southwest Jiaotong University, Chengdu, China.

His research interests include robotics and deep learning.

Qian Wan received the B.S. degree in english of foreign languages from the Xihua University, Chengdu, China, in 2022. She is currently working toward the master's degree in english translation with the Southwest Jiaotong University, Chengdu, China.

Her research interests include English translation and deep learning.

Yakun Xie received the B.S. degree in survey engineering from the School of Survey Engineering, Henan University of Urban Construction, Pingdingshan, China, in 2015, and the M.S. degree in geodesy and survey engineering from the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, in 2018, and the Ph.D. degree in surveying and mapping from the Southwest Jiaotong University, Chengdu, China, in 2022.

He is currently an Assistant Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. His research interests include computer vision, 3-D GIS technology, and virtual geographic environments.

Dejun Feng received the M.S. and Ph.D. degrees in geodesy and survey engineering from the Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively.

He is currently an Associate Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University. His research interests include geographic information systems and remote sensing image processing.

Jun Zhu received the M.S. degree in geodesy and survey engineering from the Southwest Jiaotong University, Chengdu, China, in 2003, and the Ph.D. degree in cartography and geographic information systems from the Chinese Academy of Sciences, Beijing, China, in 2006.

From 2007 to 2008, he was a Postdoctoral Research Fellow with the Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University. His research interests include computer vision, 3-D GIS technology, and virtual geographic environments.

Yangge Liu received the B.S. degree in surveying and mapping engineering from the School of Mapping and Planning, Shangqiu Normal University, Shangqiu, China, in 2020, and the master's degree in civil and hydraulic engineering from the School of Earth Science and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, in 2022.

Her research interests include remote sensing image processing and deep learning.