Extracting Building Contour and Level by Coupling U-net and Single-View High-Resolution Remote Sensing Images

Kaihu Du[1], Boyang Cui[1], Yao Yao[1,2](✉), Yuyang Cai[1], Yaqian Zhai[1], and Qingfeng Guan[1]

[1] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China
yaoy@cug.edu.cn
[2] Alibaba Group, Hangzhou 311121, China

Abstract. Researches on the urban development and urban planning have an urgent need for building geographic data. Traditional methods of extracting buildings from high-resolution remote sensing images need multi-view images, and have a high cost but a low degree of automation. Thus, these methods are not applied in many fields at large-scale. This study couples U-net and single-view high-resolution remote sensing images to propose a low-cost and simple method for the extraction of the contour and level of the buildings in the remote sensing image. This study adopts the central urban area of Wuhan, Hubei, China as the case study. The results show that the proposed method obtains high accuracies both in identifying building height level (OA = 0.823, Kappa = 0.502) and contour. Compared with the method based on the normalized digital surface model (nDSM), the proposed method obtained a higher overall accuracy of height level extraction increased by 23.4%. The overall quality of building contour extraction is high, and 78.87% of the grids covered by buildings have a building completeness index above 0.4. In addition, we detected and analyzed the changes in buildings in the Nanhu district in the study area based on the proposed method. The results indicated that the height levels of newly added buildings are mainly low and middle levels. The above results have demonstrated the validity of proposed method for extracting buildings contours and level. Moreover, the proposed method can provide scientific supports and reliable help for urban management and renewal.

Keywords: Semantic segmentation · High-resolution remote sensing image · Building height · Contour extraction · Change detection

1 Introduction

As one of the most important features in remote sensing images, buildings are closely related to human production and life [1]. Extracting building height and contours by using remote sensing images can accelerate the update of geographic databases, increase the degree of automated collection, and evaluate the compactness of urban spatial forms,

G. Pan et al. (Eds.): SpatialDI 2021, LNCS 12753, pp. 236–252, 2021.
https://doi.org/10.1007/978-3-030-85462-1_21

which can provide important references for urban land change monitoring and urban three-dimensional modeling [2–5]. Therefore, building height and contours extraction is of great significance to urban planning and development.

Nevertheless, the previous studies are mostly limited by some simple features of buildings, such as the geometry and spectrum, which cannot reflect the relationship between the building and other features. Deep learning semantic segmentation can capture the deep-level features of buildings in remote sensing images, which has shown good performance in extracting building contour and height. Maggiori *et al.* created a remote sensing semantic segmentation model based on Fully Convolutional Networks (FCN) [6]. But when classifying independent pixels, they did not fully consider the relationship with other pixels, causing the lack of spatial consistency and insufficient extraction results. In order to make up for the shortcomings of FCN, Liu *et al.* used the FCN model including Spatial Residual Interception (SRI) to extract buildings [7]. The model uses a volume integral solution to reduce parameters and improve efficiency. It captures and aggregates multi-scale context information by fusing the semantic features layer by layer. Although these semantic segmentation methods have high pixel accuracy, they lack morphological and regular processing of the building extraction.

To tackle these problems, this study proposes a method for extracting building contour and height by coupling U-net and single-view high-resolution remote sensing images. This study uses monocular high-resolution remote sensing images and building vector data to make a semantic segmentation data set, and constructs a U-net semantic segmentation network to train and extract the contours and height levels of the building. We combined the structural differences between buildings and non-buildings in remote sensing images and the differences in materials of different height levels, shadow lengths and building spacing's, and performed multiple classification and contour extraction of building height levels. Moreover, the semantic segmentation results are simplified to obtain the regular building extraction results. Also, we apply the proposed method to building contour extraction, and perform building change detection based on the two-period extracting results.

Our main contributions in this paper can be summarized as follows:

1) This study proposes a set of building extraction methods that couple U-net and single-view high-resolution remote sensing images, which solves the problems of high data threshold, time-consuming and labor-consuming in existing methods.
2) We adopt a building contour regularization method based on grid filling and polygon simplification to solve the problem of irregular edges in the semantic segmentation results.
3) Compared with previous studies on building height estimation using nDSM, the OA and Kappa of building height level classification in this study have increased by 0.234 and 0.261 respectively. Moreover, 78.87% of the grids covered by buildings have the building completeness index above 0.4, which indicates that the overall quality of building extraction is high.

2 Related Studies

In recent years, scholars have used optical remote sensing and Synthetic Aperture Radar (SAR) images to extract the height of buildings. The extraction methods based on optical remote sensing mainly include shadow altimetry measurement, edge detection, and stereo image pairing. The shadow altimetry method is based on the geometric relationship of the sun, satellites, and building imaging with the length of building shadow to explore the height of urban buildings [8, 9]. However, the application of the shadow altimetry method for concentrated buildings with complex structure needs to be improved.

The edge detection method uses isogradient operators to extract the edges of the building itself and its shadow, and then calculates the height of the building based on the geometric relationship [10, 11]. Although this method is simple and easy to operate, it is hard to meet the need for edge enhancement and noise reducing when applied in practical.

The stereo pair method takes a pair of photos of the same area from two orthogonal angles. The stereoscopic observation method and special tools can be used to establish a stereoscopic model of the ground object in the overlapping image part of the image pair. Licciardi *et al.* used the characteristics of image geometric moment invariants to obtain building height from multi-angle high-resolution remote sensing images [12]. Liu *et al.* used remote sensing stereo images to generate a normalized digital surface model (nDSM) to match the building contour to estimate the height of the building [13]. However, it is difficult to obtain multi-angle images data, which is not conducive to large-scale application.

SAR is an active remote sensing that can directly obtain three-dimensional terrain. It is also an all-weather, multi-polarized earth observation method that complements the advantages of optical remote sensing. SAR-based building height extraction methods mainly include SAR stereo image pairing method and SAR and optical remote sensing fusion method. The SAR stereo image pairing method uses orthogonal SAR stereo image pairs to obtain the height of the point with the same name after matching, and combines the DEM data to remove the surface undulations to obtain the actual height of the building [14–16]. Chen *et al.* proposed a building height extraction method integrating the mean shift and AP clustering algorithm with SAR stereo pairs [16]. The fusion method of SAR and optical remote sensing matches the position of the building extracted by optical remote sensing with the three-dimensional data extracted by SAR to obtain a more accurate height of the building [17]. Sportouche *et al.* proposed a "height hypothesis-partitioning generation-criterion optimization" building height estimation method with the fusion of SAR images and optical images [18], and the extraction results are matched with the existing building contour. However, it is difficult and expensive to obtain the multi-view optical remote sensing and high-precision SAR stereo image pair required by such methods, which is why the method cannot be applied for extracting building height in large-scale.

In practice, remote sensing satellites have external azimuth elements when shooting, so most remote sensing images shoot ground objects from an oblique direction [19]. Visual interpreters use the principle of oblique perspective and can judge the approximate contour and height of ground objects based on information such as texture, shadow, and spacing [20]. Computer vision can simulate the visual observation process of the human eye [21, 22]. Semantic segmentation is widely used in remote sensing image pixel classification tasks [23]. The neural network methods (such as FCN) have been applied in building contour extraction in previous studies, which can prove the feasibility of using remote sensing images and deep learning models to extract the contour and height of buildings. Compared with FCN, U-net is a convolutional neural network with a symmetrical structure. It has a contraction path for capturing contextual information and a symmetrical expansion path for precise target positioning and can propagate contextual information to a higher resolution, and obtain building features of different scales, which can effectively solve the problem of insufficient FCN segmentation results, and achieve more accurate segmentation results [24]. Therefore, the U-net semantic segmentation method is adopted to extract buildings from remote sensing images in this study.

3 Methodology

As shown in Fig. 1, the method for extracting building contour and height proposed in this study involves three main parts. First, we use single-view high-resolution remote sensing images and building vector data as the semantic segmentation dataset and adopt U-net neural networks for training. Second, the building contours and height are extracted by the calibrated semantic segmentation model, and then the building contours are regularized based on raster filling and polygon simplification. Finally, we carry out building change detection with an overlay analysis and a raster operation based on the building extraction results.

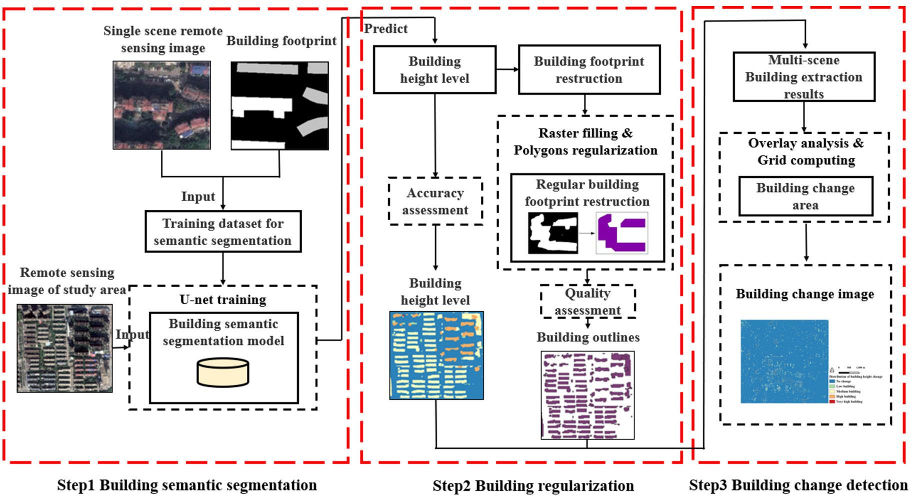

Fig. 1. Technical route of building extraction.

3.1 Building Contour and Height Extraction Based on U-net Semantic Segmentation

In this study, the semantic segmentation method is used to segment the target into several categories. The building height needs to be divided into corresponding categories according to the number of floors. The height of buildings is divided into four height levels, including low-level buildings (<3 floors), middle-level buildings (4–9 floors), high-level buildings (10–32 floors), and very high-level buildings (≥33 floors) [25].

Convolutional neural network U-net is one of the efficient algorithms in deep learning, which combines low-resolution information that provides a basis for building hierarchy classification and the high-resolution information that provides the basis for building contour segmentation and positioning. It can obtain building features of different scales [24], so as to accurately obtain the results of building semantic segmentation. Therefore, this paper uses the U-net semantic segmentation method to extract buildings from remote sensing images.

The U-net built in this study takes high-resolution remote sensing images as input and building classification images as output. The middle part is composed of a contraction path and an expansion path. In the contraction path, each step contains two repeated 3×3 unpadded convolutional layers. Both layers are followed by the modified linear unit (ReLU) activation function and a 2×2 max pooling operation with stride 2 for downsampling. In each downsampling step, the number of feature channels will be doubled. Every step in the expansive path consists of an upsampling of the feature map. A 2×2 convolution kernel is used for the up-convolution operation to reduce the number of feature channels by half. Then the process is cascaded with the corresponding cropped feature map in the contraction path, and two 3×3 convolution kernels are used for convolution operation, each followed by a ReLU. At the final layer, a 1×1 convolution kernel is used to map each 64-dimensional feature vector to the output layer of the network, corresponding to the category of the target.

3.2 Regularization of Building Contours

Semantic segmentation is the process of classifying each pixel into its category, which is easy to ignore the spatial correlation between adjacent pixels, and the extracted building contours will be rough. Therefore, we adopt a simplification method to regularize the building contours. This method can make the building contours obtained by semantic segmentation more standardized and orthogonal. Moreover, it can also repair the rough boundary phenomenon caused by semantic segmentation to a certain extent [26–28].

This study uses four steps to regularize the building contours: (1) We establish a two-dimensional coordinate system, then use the ray judgment method to find all the grids inside the polygon and fill these grids. (2) We apply morphological expansion and erosion calculations to eliminate the broken points in the grid filling result. (3) We follow the chain coding rule to track the grid in four directions to determine the vector boundary of the target result. (4) The BEND-SIMPLIFY algorithm is used to simplify the polygon, and only the important bends with the curvature radius coefficient r greater than the threshold $\boldsymbol{W}$ are retained to simplify the polygon [29]. The calculation formula of the radius of curvature is shown in formula (1), where y' and y'' represent the first derivative and the second derivative at the bend, respectively. This process retains only small bending points to achieve the effect of polygon simplification (Fig. 2).

$$r = \frac{(1 + y')^{\frac{3}{2}}}{y''} \tag{1.1}$$

$$y' = \frac{dy}{dx} \tag{1.2}$$

$$y'' = \frac{d^2y}{dx^2} \tag{1.3}$$

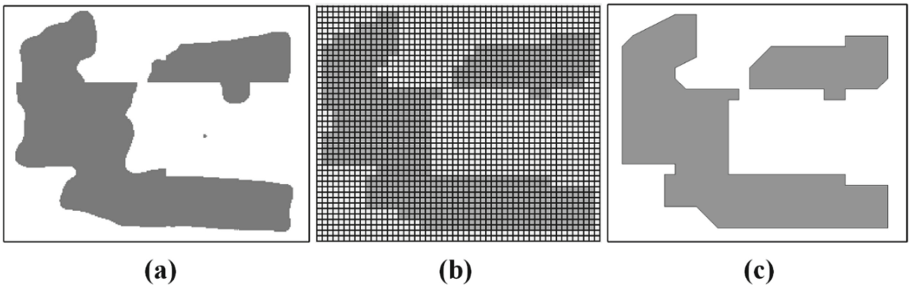

Fig. 2. Building contour regularization process: (a) Original building semantic segmentation results, (b) The result of (a) through morphology and grid filling algorithm, (c) The final result of (b) through the BEND-SIMPLIFY algorithm.

3.3 Accuracy and Quality Evaluation

We use each building as the basic unit to sample one by one and extracts the height at the geometric center of each building. The confusion matrix is constructed according to the extracted height levels and the real levels. We calculate recall rate (Recall), overall accuracy (OA), and Kappa coefficient to quantitatively evaluate the accuracy of the building level extraction results. The calculation method of Kappa is shown in formula (2), where $\boldsymbol{n}$ is the total number of classified samples, $\boldsymbol{t}$ is the number of correctly classified samples, a1, a2… are the actual number of samples for each type, and b1, b2…bn are the number of predicted samples for each type.

$$\text{Kappa} = \frac{P_o + P_e}{1 - P_e} \tag{2.1}$$

$$P_o = \frac{t}{n} \tag{2.2}$$

$$P_e = \frac{a1 \times b1 + a2 \times b2 + \ldots + ak \times bk}{n \times n} \tag{2.3}$$

In this study, the building contour reconstruction quality evaluation is based on building vector data. We use the common methods of OSM data quality evaluation and the building completeness index (Completeness) to evaluate the reconstruction quality [30, 31]. The completeness index calculation method is shown in formula (3), where ***TP*** represents the part of the reconstructed building profile that matches the reference building profile, and ***FN*** represents the part that belongs to the reconstructed building profile but not the reference building profile.

$$\text{Completeness} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

Finally, building completeness is evaluated by dividing a grid of 50 m $\times$ 50 m and calculating the quality of the reconstructed building vector. This study has formulated 5 quality levels: building completeness > 0.8, $0.6 \le$ building completeness < 0.8, $0.4 \le$ building completeness < 0.6, $0.2 \le$ building completeness < 0.4, building completeness < 0.2, and building reconstruction Mass distribution map (Fig. 3).

(a) (b)

Fig. 3. Evaluation method of building reconstruction quality: (a) The offset relationship between the reconstructed building (white) and the real building (grey). FN is the part that belongs to the reconstructed building contour but not the reference building

3.4 Building Change Detection

Building change detection is the process of comparing and analyzing multi-temporal remote sensing images to discover, determine, and obtain building change information [32, 33]. In the process of building change detection, the contour and height of the buildings are the primary factors to be investigated [34]. The basic idea of the building change detection method adopted in this study is to use the two phases of high-resolution remote sensing images at a certain time interval to extract the contour and height of the building according to the remote sensing image building extraction method proposed above. We obtain two sets of building extraction results, and use the methods of overlay analysis and raster operation to obtain both the buildings whose contours and height levels have changed, and the contours of new buildings. Then, we process the contours of the new buildings regularly and add statistics. The area occupied by the building and the level distribution of new buildings are compared with the real building distribution in the remote sensing image, which provides a valuable reference for urban planning and other fields.

4 Results and Analysis

4.1 Study Area and Data

As shown in Fig. 4, the research area is the central urban area of Wuhan, Hubei Province, China. The study data required for the experiment include the high-resolution remote sensing images and the building vector data. The high-resolution remote sensing images are from Google Earth in 2004 and 2019, with a resolution of 0.59 m. We use the contour map of Wuhan buildings in 2018 provided by Baidu Maps as the building vector data, including the number of floors and height of each building. This study creates a building semantic segmentation dataset based on the above data, in which remote sensing images are cropped to obtain training images, and building vectors are converted to raster data and cropped to obtain training labels. Considering the training efficiency, the actual size of the detection target, and the structural characteristics of the U-net, the image and the label need to be cropped to a multiple of an in the range of 200–500 pixels (a is the size of the convolution kernel, n is related to the network depth). Therefore, we choose 224

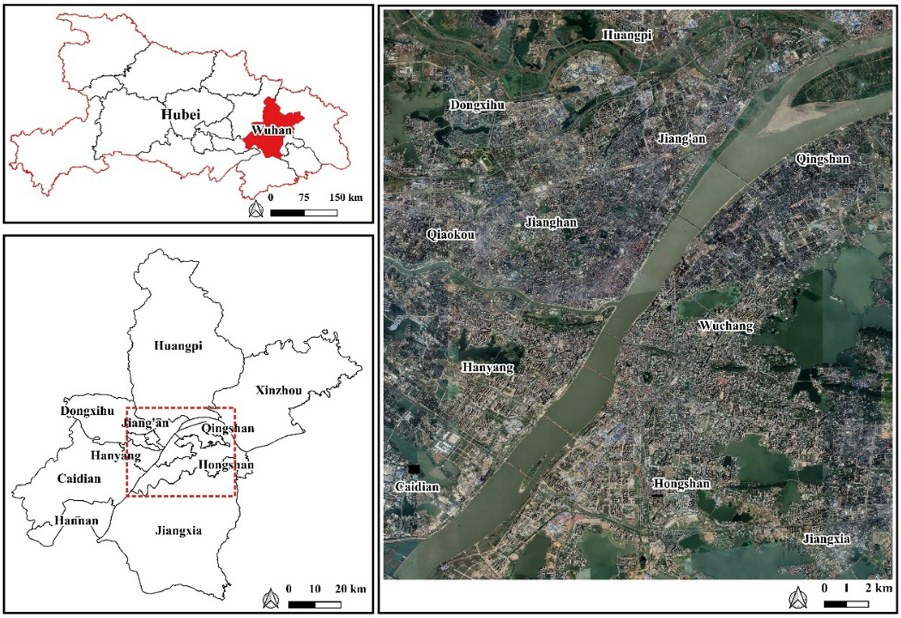

Fig. 4. Building extraction study area and its orientation.

$\times$ 224 pixels (between 200–300 pixels, a multiple of) as the size of the training image and the label. The above dataset samples are enhanced by the means of turning left and right and turning up and down, and a training data set containing a total of 16,800 pairs of images and labels is obtained.

4.2 Building Extraction Results and Accuracy Evaluation

The confusion matrix of height level classification in this study is shown in Fig. 5, and the accuracy indicators are shown in Table 1. The Recall of low-level buildings, middle-level buildings, high-level buildings, and very high-level buildings are 0.659, 0.909, 0.684, and 0.556 respectively. The Recall of various categories can reach more than 0.5, the OA of height level classification is 0.823, and the overall Kappa coefficient is 0.502. In the method [7] proposed by using nDSM to estimate height matching with building contours, the classification Recall rates of low-level buildings, middle-level buildings, high-level buildings, and very high-level buildings were 0.840, 0.396, 0.154, and 0.142, respectively. The OA reached 0.589 and the Kappa coefficient was 0.241. In addition to the classification Recall rate of low-level building, which is lower than the result of the method proposed in this study, the classification Recall rates of all the other building height levels are all higher than that of the proposed method. The OA and overall Kappa coefficients of the building height level classification in this study are also compared with Liu et al. The results increase by 0.234 and 0.261 respectively.

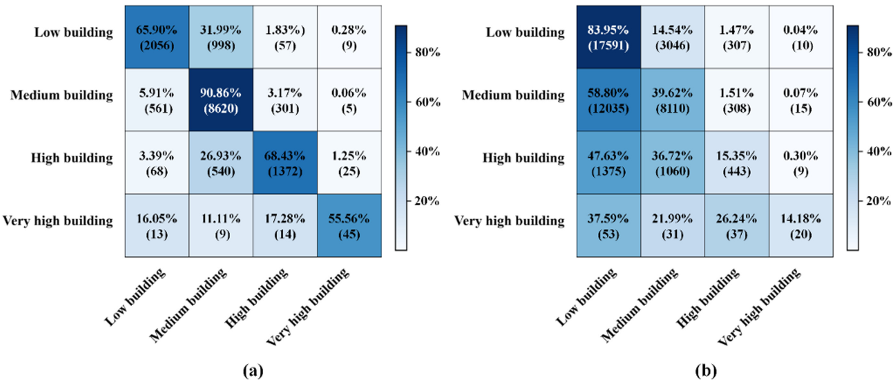

Fig. 5. Confusion matrix of building height level classification: (a) The confusion matrix of this experiment, the numbers in the squares indicate the classification ratio and the number of buildings, (b) The confusion matrix of Liu et al. experiment.

This study uses real building contour as a benchmark to evaluate the quality of the reconstructed building contour. The quality distribution diagram of the reconstructed building contour is shown in Fig. 6. Among the grids covered by buildings, grids with building completeness greater than 0.8 reach 42.02%; grids with building completeness between 0.6 and 0.8 reach 21.37%; grids with building completeness between 0.4 and 0.6 account for 15.47%; and grids with building completeness less than 0.4 account for

Table 1. Comparison of building height level accuracy between the method in this study and Liu et al.

Evaluation indicators	Method of coupling U-net and single-view high-resolution remote sensing image	nDSM method for estimating height and fitting with building contour
Recall rate of low-level building	0.695	0.840
Recall rate of middle-level building	0.909	0.396
Recall rate of high-level building	0.684	0.154
Recall rate of very high-level building	0.556	0.142
OA	0.823	0.589
Kappa	0.502	0.241

21.13%. Moreover, there are 78.87% of the grids have a completeness index above 0.4, which indicates that the building contour reconstruction has a high overall quality.

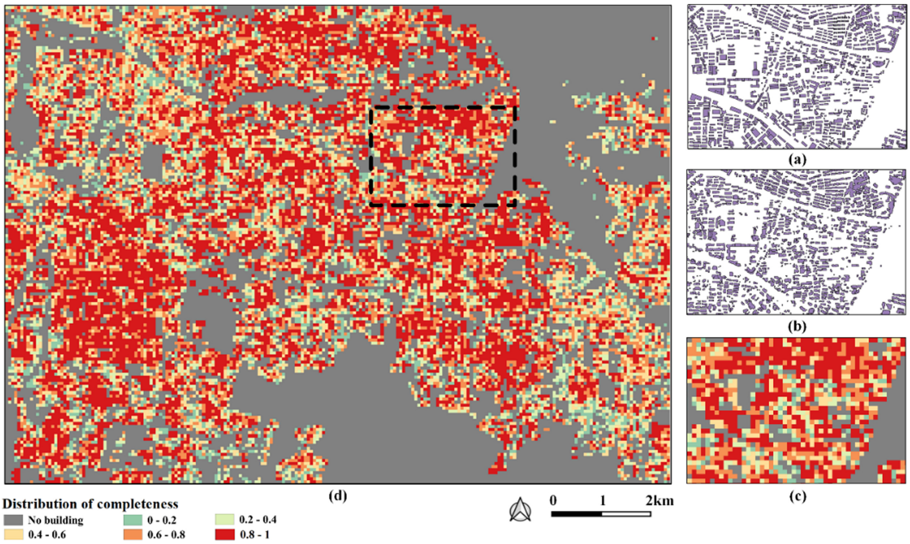

Fig. 6. Building reconstruction quality distribution map: (a) Real building vector, (b) Reconstructed building vector, (c) Building vector reconstruction quality distribution, (d) Vector reconstruction quality distribution of all buildings in the study area.

Figure 7(a) and (b) are the results of the regularization of building contours and the semantic segmentation of buildings, respectively. Figure 7(a1)–(a3) and (b1)–(b3) are the comparison of the local details of remote sensing images and the extracted results. We find that the edges and corners of the regularized buildings are distinct and have the general geometric features of the real building contour, which is more beautiful and regular than the result extracted directly from the semantic segmentation (Fig. 7(a)). Figure 7(b) shows that the boundary of the semantic segmentation of buildings is clear, the predicted building height level is consistent with the real height level in the remote sensing image, and the distribution of mixed building height levels can also be correctly distinguished.

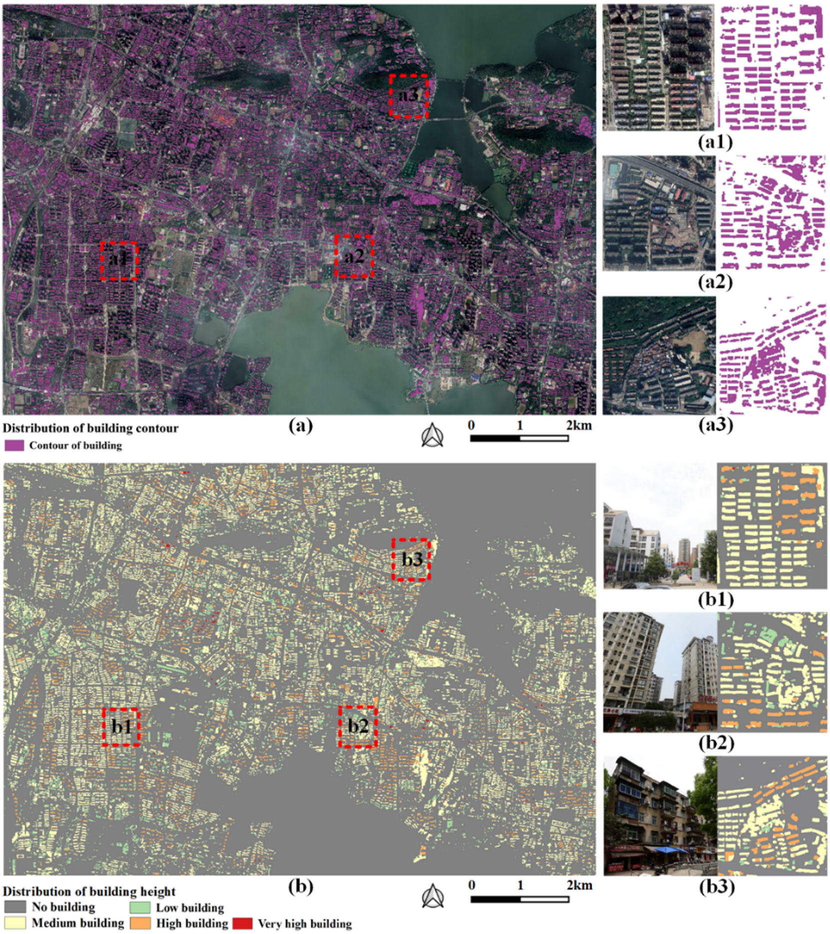

Fig. 7. Building height level and contour extraction results: (a) Result of building contour regularization, (a1)–(a3): Local details; (b) Result of building height level extraction, (b1)–(b3): Local details

4.3 Building Change Detection

This study takes Nanhu Lake area in Wuhan as the experimental area for building change detection, and the buildings within this area in 2004 and 2019 are extracted and compared. The experiment results are shown in Fig. 8 and Fig. 9. The results show that the construction land in this area had a clear trend from north to south during this period, and a piece of farmland in the area was gradually changed to construction land. The new buildings detected in the experiment are mainly concentrated in the original farmland in the south-central part of the area. In addition, there are some new buildings irregularly distributed in various places outside the original farmland, which shows that the area opens up new building land for intensive development, rebuilds and expands buildings on existing plots. The new buildings detected in the experiment occupies an increase to about 3.48% of the total area. Among the new buildings, low-level buildings

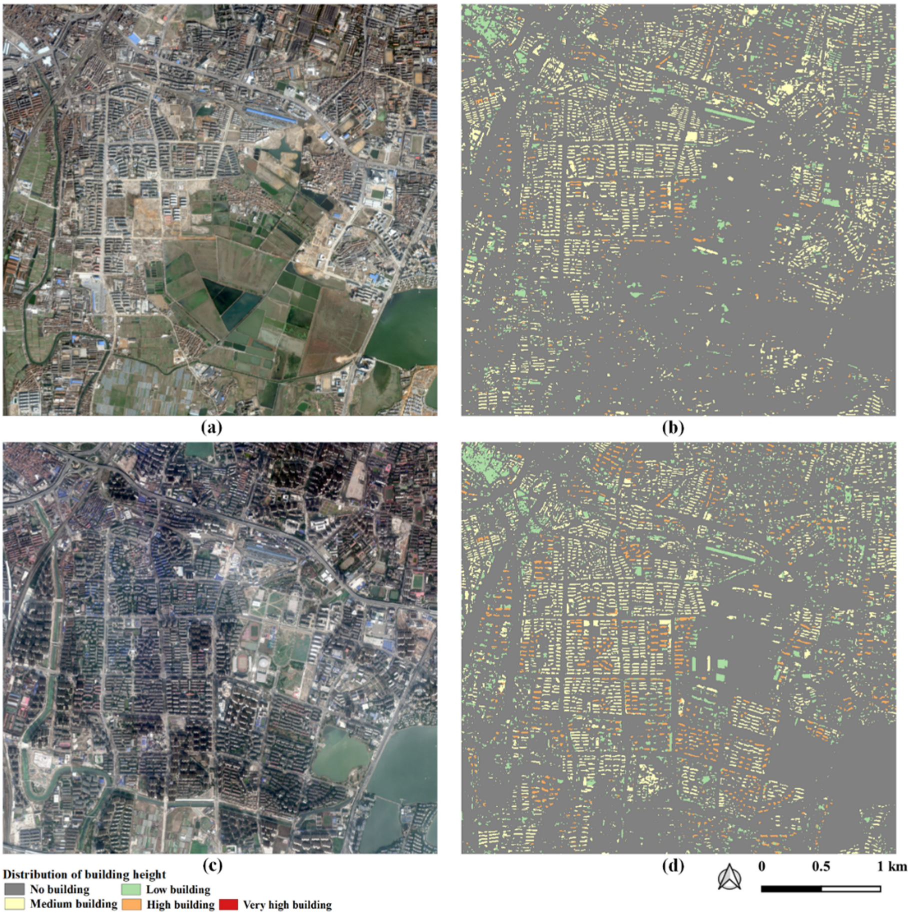

Fig. 8. Building height levels and contour extraction results of Wuhan Nanhu in 2004 and 2019: (a) Remote sensing image in 2004, (b) Building extraction result in 2004, (c) Remote sensing image in 2019, (d) Building extraction result in 2019.

account for about 30.08%, middle-level buildings account for about 43.67%, and high-level buildings account for about 26.25%. The above statistics show that other land-use types in Wuhan Nanhu were transformed into construction land from 2004 to 2019. The distribution of new buildings was mainly low-level and middle-level buildings, mixing with high-level buildings. The distribution of the newly added buildings in the detection results basically coincides with the farmland in the remote sensing image of the area in 2004, indicating that the detection results of the newly added buildings are reasonable.

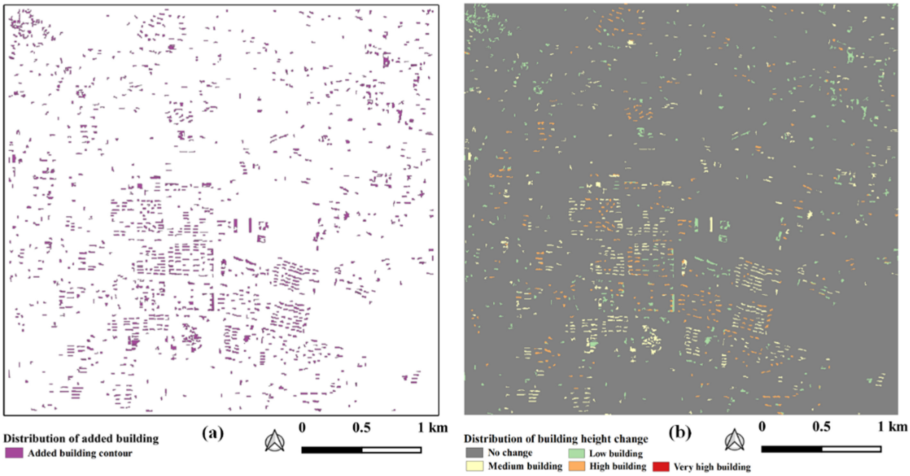

Fig. 9. Distribution of new building contours and height level changes in 2019 compared with 2004: (a) Added building contour, (b) Changes in building height level.

5 Discussion

The accurate extraction of buildings is of great significance to the research of the urban development process and urban planning. However, the traditional high-resolution remote sensing image building extraction research has problems such as high threshold for multi-view remote sensing and SAR data, low automation, and difficulty in large-scale application. In addition, the latest methods such as semantic segmentation also have defects such as irregular building contour extraction and noise points. Therefore, this study designs a set of building contour and height extraction methods that couple U-net and single-view high-resolution remote sensing images. The central urban area of Wuhan, Hubei Province was used as the research area, and the public monocular high-resolution remote sensing images were combined with semantic segmentation and regularization of building contours to enable accurate extraction of building height levels and contours in the research area. Results of the above experiments have proved that the proposed method is both of high accuracy and great efficiency.

This study uses single-view high-resolution remote sensing images for training and experiments, which is free and easily available compared to the expansive professional remote sensing data (e.g., SAR and DSM). The building extraction method proposed in this study has high efficiency. It can extract urban buildings within a range of tens of square kilometers at one time. It takes only about 54s to extract a building in a 5000*5000pixel size remote sensing image. The efficiency is much higher than that of the manually outlined supervised classification method. In summary, this method has the advantages of low cost, high degree of automation, and easy to apply on a large scale.

Since the U-net semantic segmentation model selected in this study takes the image features at multiple scales into account, it has strong ground object classification and analysis capabilities [24]. The building regularization method based on grid filling and polygon simplification make the expression of the building contour more accurate. The combination of the U-net semantic segmentation model with the building regularization method make the extraction results accurately. However, the analytical capabilities of semantic segmentation and building regularization are limited. There is some confusion in the results of building-level classification, and a few grids with completeness indicators less than 0.4 in the building contour quality map. In general, the overall building extraction results are considerable (Fig. 10).

Fig. 10. The effect of building contour regularization based on grid filling and polygon simplification: (a) Original remote sensing image, (b) Convert directly to vector graphics, (c) Convert to vector graphics after regularization.

The proposed method can continuously provide large-scale urban building height level data and contour vector data with the update of high-resolution remote sensing images, thereby speeding up the update frequency of the geographic database. Therefore, it has a high practical value in the field of urban planning and urban land use analysis. The results of the building change detection experiment indicate that a piece of farmland was newly added as construction land from 2004 to 2019, accounting for about 3.48% of the total area. As it is shown in Table 2, the newly added buildings were mainly low-level buildings and middle-level buildings. High-level buildings accounted for only a small percentage. These statistics indicate that during this period, there was an obvious process of concentrated urbanization from suburbs to towns in Nanhu, Wuhan. This conclusion can be verified in the comparison of remote sensing images in 2004 and 2019. In addition, the automatic detection of building changes based on remote sensing images is vital to the sustainable development of urban land use and has broad prospects in the field of disaster assessment [35].

The proposed method has high accuracy and quality, which has been verified to be used in the research of building change detection. However, the small part of the vector

Table 2. Statistics of the proportion of different building height levels.

Detection indicators	New buildings detected by change detection
Proportion of low-level building	30.08%
Proportion of middle-level buildings	43.67%
Proportion of high-level buildings	26.25%

data referenced in the experiment is rough and cannot be matched with the buildings in the remote sensing image, resulting in certain systematic errors in the semantic segmentation dataset. Therefore, the accuracy of this method still has room for improvement. The Recall of such buildings is only 0.556 due to the relatively small proportion of very high-level buildings in urban buildings and insufficient training. In addition, in the building change detection based on building extraction, we have only summarized the overall urban development situation. In-depth internal analysis of specific changes using individual buildings as a unit requires the support of higher extraction accuracy. In future researches, we can consider increasing the number of sampling areas when creating the dataset, and balancing the proportion of samples of buildings at all height levels to train each type of building fully. In addition, we can consider studying the matching algorithm of building vectors, associating buildings with the same name extracted in different periods, and analyzing the changes of buildings in detail to help researches related to land use and urban planning.

6 Conclusion

This study proposes a set of building extraction methods that couple U-net and single-view high-resolution remote sensing images, which solves the problems of high data threshold, time-consuming and labor-consuming in existing methods. We adopt a building contour regularization method based on grid filling and polygon simplification to solve the problem of irregular edges in the semantic segmentation results. Compared with previous studies on building height estimation using nDSM, the OA and Kappa of building height level classification in this study have increased by 0.234 and 0.261 respectively. Moreover, 78.87% of the grids covered by buildings have the building completeness index above 0.4, which indicates that the overall quality of building extraction is high. In addition, we use Wuhan Nanhu as the research area to detect building changes and obtained the distribution of new buildings in the area. The results are consistent with the actual development of the area in the remote sensing image. In conclusion, the building extraction method proposed in this study can help to promote the discovery of building changes in the process of urbanization. The results of this study can also be used in urban land intensity monitoring and urban planning.

References

1. Sirmaek, B., Unsalan, C.: Urban-area and building detection using SIFT keypoints and graph theory. IEEE Trans. Geosci. Remote **47**(4), 1156–1167 (2009)

2. Chen, Y., Tang, L., Yang, X., Bilal, M., Li, Q.: Object-based multi-modal convolution neural networks for building extraction using panchromatic and multispectral imagery. Neurocomputing **386**, 136–146 (2019)
3. Huang, J., Zhang, X., Xin, Q., Sun, Y., Zhang, P.: Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. ISPRS J. Photogramm. Remote Sens. **151**, 91–105 (2019)
4. Huang, X., Zhang, L.: A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. Photogramm. Eng. Remote Sens. **77**(7), 721–732 (2011)
5. Zhao, J., Song, Y., Shi, L., Tang, L.: Study on the compactness assessment model of urban spatial form. Acta Ecol. Sin. **31**(21), 6338–6343 (2011)
6. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: Fully convolutional neural networks for remote sensing image classification. In: Geoscience & Remote Sensing Symposium (2016)
7. Liu, C., Huang, X., Zhu, Z., Chen, H., Tang, X., Gong, J.: Automatic extraction of built-up area from ZY3 multi-view satellite imagery: analysis of 45 global cities. Remote Sens. Environ. **226**, 51–73 (2019)
8. Thiel, K.H.: Delimiting the building heights in a city from the shadow in a panchromatic SPOT-image—Part 1. Test of forty-two buildings. Int. J. Remote Sens. **16**, 409–415 (1995)
9. Shao, Y., Taff, G.N., Walsh, S.J.: Shadow detection and building-height estimation using IKONOS data. Int. J. Remote Sens. **32**(22), 6929–6944 (2011)
10. Jahagirdar, A., Patil, M., Pawar, V., Kharat, V.: Comparative study of satellite image edge detection techniques. Int. J. Innov. Res. Comput. Commun. Eng. **3297**(5) (2016)
11. Katartzis, A., Sahli, H.: A Stochastic framework for the identification of building rooftops using a single remote sensing image. IEEE Trans. Geosci. Remote **46**(1), 259–271 (2008)
12. Licciardi, G.A., Villa, A., Mura, M.D., Bruzzone, L., Chanussot, J., Benediktsson, J.A.: Retrieval of the height of buildings from worldview-2 multi- angular imagery using attribute filters and geometric invariant moments. IEEE J. STARS **5**(1), 71–79 (2012)
13. Liu, P., et al.: Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. Remote Sens. **11**(7), 830 (2019)
14. Soergel, U., Michaelsen, E., Thiele, A., Cadario, E., Thoennessen, U.: Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions. ISPRS J. Photogramm. **64**(5), 490–500 (2009)
15. Xu, F., Jin, Y.Q.: Automatic reconstruction of building objects from multiaspect meter-resolution SAR images. IEEE Trans. Geosci. Remote Sens. **45**(7), 2336–2353 (2007)
16. Chen, J., Chao, W., Hong, Z., Bo, Z., Fan, W.: Geometrical characteristics based building height extraction from VHR SAR imagery. In: 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL) (2017)
17. Tupin, F.: Merging of SAR and optical features for 3D reconstruction in a radargrammetric framework. In: IEEE International Geoscience and Remote Sensing Symposium (2004)
18. Sportouche, H., Tupin, F., Denise, L.: Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images. IEEE Trans. Geosci. Remote Sens. **49**(10), 3932–3946 (2011)
19. Wierzbicki, D., Krasuski, K.: Determining the elements of exterior orientation in aerial triangulation processing using UAV technology. Komunikacie **22**(1), 15–24 (2020)
20. Van Coillie, F.M.B., Gardin, S., Anseel, F., Duyck, W., Verbeke, L.P., De Wulf, R.R.: Variability of operator performance in remote-sensing image interpretation: the importance of human and external factors. Int. J. Remote Sens. **35**(2), 754–778 (2014)
21. Ahissar, M., Hochstein, S.: The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. **8**(10), 457–464 (2004)
22. Bar, M.: A cortical mechanism for triggering top-down facilitation in visual object recognition. J. Cogn. Neurosci. **15**(4), 600–609 (2003)

23. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. **39**(4), 640–651 (2015)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
25. China, M.O.C.O.: Code for Design of Civil Building (GB50352–2005)
26. Matei, B.C., Sawhney, H.S., Samarasekera, S., Kim, J., Kumar, R.: Building segmentation for densely built urban regions using aerial LIDAR data. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition (2008)
27. Goes, F.D., Cohen-Steiner, D., Alliez, P., Desbrun, M.: An optimal transport approach to robust reconstruction and simplification of 2D shapes. In: Computer Graph Forum (2011)
28. Wang, W., et al.: A Grid Filling Based Rectangular Building Outlines Regularization Method. Geomatics and Information Science of Wuhan University (2018)
29. Wang, Z., Muller, J.C.: Line generalization based on analysis of shape characteristics. Cartogr. Geogr. Inf. Syst. **25**, 3–15 (1998)
30. Fan, H., Zipf, A., Fu, Q., Neis, P.: Quality assessment for building footprints data on OpenStreetMap. Int. J. Geogr. Inf. Sci. **28**(3–4), 700–719 (2014)
31. Girres, J.F., Touya, G.: Quality assessment of the French OpenStreetMap dataset. Trans. GIS **14**(4), 435–459 (2010)
32. Champion, N.: 2D building change detection from high resolution aerial images and correlation digital surface models (2007)
33. Tian, J., Cui, S., Reinartz, P.: Building change detection based on satellite stereo imagery and digital surface models. IEEE Trans. Geosci. Remote 56, 406–417 (2013)
34. Huang, X., Zhang, L., Zhu, T.: Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(1), 105–115 (2013)
35. Chen, B., Chen, Z., Deng, L., Duan, Y., Zhou, J.: Building change detection with RGB-D map generated from UAV images. Neurocomputing **208**, 350–364 (2016)