Article

Spaceborne SAR Data for Regional Urban Mapping Using a Robust Building Extractor

Juanjuan Li ^{1,2}, Hong Zhang ^{1,*} , Chao Wang ^{1,2} , Fan Wu ¹ and Lu Li ^{1,2}

¹ Key Laboratory of Digital Earth Science, Aerospace information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zkyljj@radi.ac.cn (J.L.); wangchao@radi.ac.cn (C.W.); wufan@radi.ac.cn (F.W.); lilushhf@sina.com (L.L.)

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zhanghong@radi.ac.cn; Tel.: +86-10-8217-8186

Received: 28 July 2020; Accepted: 24 August 2020; Published: 27 August 2020

Abstract: With the rapid development of urbanization, timely and accurate information on the spatial distribution of urban areas is essential for urban planning, environmental protection and sustainable urban development. To date, the main problem of urban mapping using synthetic aperture radar (SAR) data are that nonbuilding objects with high backscattering cause high false alarms, while small-scale buildings with low backscattering result in omission errors. In this paper, a robust building-area extraction extractor is proposed to solve the above problems. The specific work includes (1) building a multiscale and multicategory building area dataset to learn enough building features in various areas; (2) designing a multiscale extraction network based on the residual convolutional block (ResNet50) and a pyramid-based pooling module to extract more discriminative features of building areas and introducing the focal loss item as the object function of the network to further extract the small-scale building areas and (3) eliminating the false alarms using the Normalized Difference Vegetation Index (NDVI) and Modified Normalized Difference Water Index (MNDWI) index. GF-3 SAR data with a 10-m resolution of four regions in China are used to validate our method, and the regional building-area mapping results with overall accuracy above 85% and kappa coefficient not less than 0.73 are obtained. Compared with the current popular networks and the global human settlement layer (GHSL) product, our method shows better extraction results and higher accuracy in multiscale building areas. The experiments using Sentinel-1 and ALOS-2/PALSAR-2 data show that the proposed method has good robustness with different SAR data sources.

Keywords: building-area mapping; deep learning; ResNet50; pyramid-pooling module; GF-3 SAR

1. Introduction

Urbanization has become a global trend affecting most of the world's citizens. According to the World Urbanization Outlook (2018 revision) [1], approximately 55% of the world's population is actually living in urban areas and is expected to increase to 68% by 2050. With the acceleration of urbanization, accurate and timely global human presence information becomes essential for risk assessment, infrastructure planning, population estimation and economic construction as well as to support sustainable urban development [2].

Satellite remote sensing has been widely recognized as the most economical and feasible approach to derive human settlement information over large areas [3,4]. Most of the current regional or global land cover products are based on time-series optical satellite data [5–16], and have been formed by a series of data sets and algorithms [17–19]. These data have also been widely used in cities, but the extraction results of urban areas cannot meet the needs of urban change detection. Due to the data availability problems caused by rainy and cloudy weather—as well as the large amount of data and

computation involved—the mapping of regional or global city boundaries using optical data are still a major challenge.

Compared to optical data, synthetic aperture radar (SAR) data have the advantages of observing the land surface because of their all-weather/all-time capability and unique information content [20]. SAR data have been increasingly investigated for global urban extent extraction at various spatial resolutions with promising results. In 2012, Esch et al. [21] of DLR in Germany validated the potential of high-resolution X-band data for the automated mapping of settlement areas. Several years later, the group introduced the Urban Footprint Processor (UFP), a novel unsupervised and fully automated system from the TanDEM-X mission (TDM) [22]. Then, inspired by the findings of several studies [23–26], the UFP suite was optimized and modified by adding an automated postediting module aiming at the elimination of false positive alarms. The methodology was then applied at a global scale and resulted in the global urban footprint (GUF) dataset with a spatial resolution of 12 m [27]. Gamba and Lisini et al. [28] developed a fast and efficient method for global urban extent mapping using ENVISAT ASAR wide swath mode data with 75-m resolution, and the results are more accurate than the existing global datasets, including GlobCover 2009. To evaluate ENVISAT SAR data for global urban mapping, one group developed the KTH-Pavia urban extractor to effectively extract urban extent and small towns using ENVISAT ASAR 30 m data [29]. Since the Sentinel-1 images were published and shared globally, some scholars have attempted to extract global human settlements based on Sentinel-1 data. Ban et al. [30] used Sentinel-1A SAR data to evaluate the KTH-Pavia urban extractor for urban extent extraction and preliminarily determined that the Sentinel-1A stripmap mode is very suitable for urban extraction, reaching accuracy more than 83%. This attempt lays a foundation for the extraction of new generation human settlements. Cao et al. [31] introduced the spatial indicator and texture feature together with the intensity of Sentinel-1 SAR data to the seed selection procedure and successfully extracted the built-up areas of cities in China.

At present, the traditional algorithms applied to SAR urban area extraction are mainly the threshold method based on intensity and texture features [27,28,32], support vector machine (SVM) [33], neural network [34], etc. In recent years, deep learning has made remarkable achievements in the field of classification and target recognition [35–39]. Many scholars use deep-learning methods to solve the problem of multiscale distribution of building area in SAR data. Based on the abundant building feature information in SAR data, Li et al. [40] proposed a multiscale convolution neural network composed of CNN84, CNN42 and CNN14 which realized multiscale and multilevel feature extraction. Gao et al. [41] employed transfer learning to extract building areas in SAR images. To solve the problem of insufficient receptive field of convolutional neural network, the context network was used to enhance the original network, and the detection result was better than that of multiscale CNN method. Wu et al. [42] proposed a multiscale structured network to solve the problem of multiscale distribution of building areas in SAR data.

These methods solved the problem of multiscale distribution of buildings in SAR images to some extent, but they have not been applied to large-scale mapping of complex terrain and multicategory buildings. To date, there are two main problems in building-area extraction under the complex background of large regions: (1) Omission errors: Low-rise buildings—such as villages in cities and scattered villages in rural areas—present low backscattering values in SAR images, which are prone to be missed. Some flat-roofed buildings are dominated by single scattering, and the scattering echo is relatively weak. Under some imaging conditions, only linear outlines are displayed in SAR image, which are easy to be omitted; (2) Commission errors: Some roads, bridges and ships exhibit high backscattering and texture characteristics similar to building areas, which easily lead to misclassification.

Therefore, in order to accurately delineate building areas, it is necessary to obtain deeper texture feature information of buildings; a network must have the extraction performance of building areas with large-scale changes. Thus, in this paper, a pyramid multiscale building-area extraction network is designed based on a residual convolutional network and pyramid-pooling module. While

deepening the depth of the network, residual network (ResNet) [43] addresses the problem of gradient disappearance problem well; the pyramid structure takes into account the target features under multiple receptive fields in parallel [44], which is suitable for multiscale building-area extraction. In addition, to solve the problem of imbalanced distribution of positive and negative samples, focal loss [45] was introduced to the classifier to replace the original cross-entropy (CE) loss function. GF-3 SAR data of four regions in China were selected to verify the effectiveness and accuracy of the proposed method. The data robustness of the method is evaluated by Sentinel-1 and ALOS-2/PALSAR-2 data. The experimental results show that the method cannot only be applied to building-area extraction in a variety of complex environments, but also suitable for multi-source data. It will have a good application prospects for urban mapping in the future.

2. Methodology

The difference in scattering characteristics of buildings in SAR images is caused by differences in building properties, heights and scales. Not only is the scattering intensity of high-rise buildings very different from that of low-rise buildings, but the scale difference of building areas is also be characterized as multiscale in SAR images. In addition, the scattering characteristics of vegetation, roads and mountains are similar to those of buildings under certain imaging conditions. Therefore, it is necessary to extract deeper levels of information and multiscale features of buildings; the corresponding postprocessing for building-area extraction is also important.

Therefore, a method for building-area extraction using SAR images is proposed as shown in Figure 1. The main research work includes the following three parts: (1) building pixel-level-labeled samples for multicategory building areas, (2) construction of a pyramid multiscale building-area extraction network based on a residual-feature extraction network and multiscale pyramid-pooling module, (3) false-alarm removal based on non-building masks.

Figure 1. Flowchart of building-area extraction using synthetic aperture radar (SAR) data.

First, all SAR images were preprocessed by radiometric calibration, filtering and geocoding using ENVI SARscape software. Second, pixel-level-labeled samples of multicategory buildings were

produced based on the preprocessed SAR data and the auxiliary data using ArcGIS software and the sample images were divided into 256×256 patches covering the training set and test set. Then, the pyramid multiscale building-area extraction network based on the residual-feature extraction network and multiscale pyramid-pooling module were constructed. Next, the training set and test set were used to train and test the network, respectively. Finally, the optimized network was obtained. Finally, the preprocessed SAR image was input into the network to obtain the preliminary binary image of building areas, and the final results were obtained by removing non-building mask layers through logical operations.

2.1. Building Pixel-Level Labeled Samples for Multicategory Building Areas

As mentioned above, the urban distribution of each city has multiscale characteristics. Building areas in the central city are distributed in patches, while building areas in the suburbs are small and scattered. However, due to the influence of topography, history and other factors, the spatial layout of buildings in different cities is still different. In some cities, there are many factories and villages; in some cities, the terrain fluctuates greatly, and the buildings are mostly distributed along the roads or rivers; in some cities, there are many small settlements scattered on the plain.

Therefore, this paper designs a pixel-level-labeled samples strategy for multicategory building areas—including downtown built-up areas, mountain buildings, plain villages and urban villages. Figure 2 shows the production process of pixel-level-labeled samples. First, in order to ensure the accuracy of the building area samples, Sentinel-2 data with 10-m resolution were referenced to ensure the integrity of the building area, so as to reduce the omission error. Second, in the process of clipping, an overlap area of 30 pixels was set to ensure that the boundary of small buildings was not damaged. Then, the incomplete samples (SAR image edge containing less building information) were eliminated to ensure the quality of the samples. Each sample was divided into positive samples (effective samples) or negative samples according to whether the sample contains building area targets and then stored in different subsets. According to the actual situation of the study area, building area and nonbuilding area in plain area could be easily distinguished, so the samples in these areas were all positive samples; while building samples in mountainous areas include positive samples and negative samples, which ensured that the network could fully learn building features and mountain features to accurately distinguish them.

Figure 2. Production process of various samples.

2.2. Construction of Pyramid Multiscale Building Area Extraction Network

Based on residual network (ResNet) and the pyramid-pooling module proposed in the pyramid scene parsing network (PSPNet) model, a pyramid multiscale building-area extraction network was constructed. ResNet was mainly used for the extraction of deeper level texture features, and the

pyramid-pooling module was used for multiscale feature extraction and pixel-level classification. To solve the classification problem of the imbalanced pixel proportion between the building areas and the non-building areas in SAR images, focal loss was introduced to the classifier of the network to replace the original cross-entropy loss function.

2.2.1. Residual Feature Extraction Network

ResNet [43] is based on the convolutional neural networks (CNN) model by introducing a deep-residual-learning framework, and can effectively address the degradation problem with increasing network depth. In this paper, ResNet50 was designed as the feature extraction network of building areas, which is a stack of many identity blocks and convolutional blocks (CONV blocks). The framework of ResNet50 is shown in Figure 3. Figure 3b shows that the identity block is a residual module with a stack of 3 layers, and each layer is composed of CONV2D, Batch normalization (BN), and rectified linear unit (ReLU) with a shortcut connection, in which shortcut connections go across feature maps of two sizes. The shortcut connection, which also denotes identity mapping, is the core idea of ResNet for addressing the gradient degradation problem. The identity mapping sends the feature map of building areas extracted from the shallow network to the deeper network directly across several layers, which can ensure that the details of the extracted texture information do not degrade as the network depth increases. Figure 3c shows the structure of the CONV block. The difference between the CONV block and the identity block is that there is a 1×1 CONV2D layer in the shortcut connection, which is used to reduce the dimensions of the feature map so that the dimensions of the input match the output connected behind. The identity block and CONV block all use a stack of 3 layers with 1×1 , 3×3 and 1×1 convolutions, where the 1×1 layers are responsible for reducing and then increasing (restoring) dimensions, leaving the 3×3 layer a bottleneck with smaller input/output dimensions. Thus, not only is the accuracy of feature extraction increased, but also the number of calculations is reduced. However, ResNet can only obtain the multilevel texture information of buildings at one scale. For multiscale building distribution, the pyramid module is needed.

Figure 3. Structure of the feature extraction network (ResNet50).

2.2.2. The Pyramid Pooling Module

The pyramid-scene-parsing network (PSPNet) achieves state-of-the-art performance in embedding difficult scenery context features in an FCN-based pixel-prediction framework [44], in which the pyramid-pooling module has been empirically demonstrated to be an effective global contextual prior. In this paper, the pyramid pool module was introduced into the extraction of multiscale building areas. The pyramid-pooling module fuses features under four-level pyramid scales with

bin sizes of 1×1 , 2×2 , 3×3 and 6×6 (see Figure 4). The coarsest level highlighted in red is global pooling to generate a single bin output. The following pyramid level separates the feature map of building areas into different subregions and forms a pooled representation for different locations. The output of different levels in the pyramid-pooling module contains the feature map with varied sizes. To maintain the weight of the global feature, the 1×1 convolution layer after each pyramid level reduces the dimensions of the context representation to $1/N$ of the original one, if the level size of the pyramid is N . Then, the low-dimensional feature maps are directly upsampled to obtain the same size feature as the original feature map via bilinear interpolation. Finally, different levels of features are concatenated as the final pyramid pooling global feature, which was used for subsequent classification work. The pyramid-pooling module collects multilevel information of multiscale building areas and combines them with the original feature map extracted from ResNet50 to improve the accuracy of building-area mapping.

Figure 4. Structure of the feature extraction network (ResNet50).

2.2.3. Focal Loss

The building samples contained some small villages. In this kind of sample, the area of buildings accounted for a relatively small proportion, and there was a serious imbalance between positive and negative samples of buildings. To address this problem, focal loss was introduced to the classifier to replace the original cross-entropy (CE) loss function.

Focal loss is proposed based on CE loss function [45]. The formula of CE is as follows:

$$L_{ce} = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}) = \begin{cases} -\log \hat{y}, & y = 1 \\ -\log(1 - \hat{y}), & y = 0 \end{cases} \quad (1)$$

In the above, $y \in \{0, 1\}$ specifies the category of labels, in this paper, including buildings and non-buildings. $\hat{y} \in [0, 1]$ is the model's estimated probability for the class with label $y = 1$, which is obtained from the sigmoid activation function. It can be seen that for normal cross entropy, the higher the output probability is, the smaller the loss is for positive samples; for negative samples, the smaller the output probability is, the smaller the loss. The loss function is slow and may not be optimized optimally in the iterative process of a large number of samples.

Thus, focal loss is proposed by reshaping the loss function to downplay easy examples and thus focus training on hard negatives. The formula of focal loss is as follows:

$$L_{fl} = \begin{cases} -(1 - \hat{y})^\gamma \log \hat{y}, & y = 1 \\ -\hat{y}^\gamma \log(1 - \hat{y}), & y = 0 \end{cases} \quad (2)$$

A tunable focusing parameter $\gamma \geq 0$ was added to the CE loss function, which can reduce the loss of easily classified samples and make the classification focused more on difficult and misclassified

samples. In addition, a weighting factor $\alpha \in [0, 1]$ is added to balance the unbalanced proportion of positive and negative samples.

$$L_{fl} = \begin{cases} -\alpha(1-\hat{y})^\gamma \log \hat{y}, & y = 1 \\ -(1-\alpha)\hat{y}^\gamma \log(1-\hat{y}), & y = 0 \end{cases} \quad (3)$$

Through a series of parameter adjustments, we found $\alpha = 0.25$, $\gamma = 2$ to work best in our extraction task.

2.3. False-Alarm Removal

Finally, further postprocessing was needed to eliminate false alarms and improve the extraction accuracy. Common false alarms include vegetation, roads and mountains. Using auxiliary data to obtain several kinds of nonbuilding mask layers, a logic calculation was used to remove the intersection of initial extraction results and mask layers (See Figure 5). The specific steps were as follows: (1) Normalized Difference Vegetation Index (NDVI) and Modified Normalized Difference Water Index (MNDWI) images were obtained from Sentinel-2 optical data using the spectral index module of ENVI Software. Then, appropriate thresholds were set to obtain vegetation and water mask layers. In this paper, the threshold values of vegetation and water mask layers were 0.42 and 0.38, respectively. (2) Road data were obtained from OpenStreetMap open source data website. In this paper, highway, national highway and provincial highway layers were merged to one road layer and then polyline layer was converted to polygon layer; (3) The 30-m SRTM DEM was resampled to 10-m resolution. Then, the average slope was computed in a window around the pixel under test with the slope threshold value 15° [28]. Finally, false-alarm removal work was carried out to erase the mask layers from the initial results layer to obtain the final postprocessing layer.

Figure 5. Framework of postprocessing process.

3. Study Areas and Experimental Data

3.1. Study Areas and Experimental Data

Four cities in China, Beijing, Wuhan, Guangzhou and Shangqiu of Henan Province were selected to carry out the experiment. Beijing, Wuhan and Guangzhou are densely populated urban areas with different environmental conditions, while Shangqiu is located on the plain with many small villages. A total of 18 GF-3 data were used in the four experimental areas. The specific information is shown in Table 1. In addition, in Beijing, Sentinel-1 and ALOS-2/PALSAR-2 data were also used for data-applicability analysis; the detailed information is shown in Table 2.

Table 1. Basic information of the GF-3 SAR data.

City	Acquisition Date	Imaging Mode	Orbit Direction	Resolution (m)	Polarization	Num.
Beijing	17 July 2018	FSII	ASC	10	HH/HV	1
	15 June 2019	FSII	ASC	10	HH/HV	1
	15 June 2019	FSII	ASC	10	HH/HV	1
	14 July 2019	FSII	ASC	10	HH/HV	1
	14 July 2019	FSII	ASC	10	HH/HV	1
Wuhan	21 October 2018	FSII	ASC	10	HH/HV	1
	1 December 2018	FSII	ASC	10	HH/HV	1
	1 December 2018	FSII	ASC	10	HH/HV	1
	10 June 2019	FSII	ASC	10	HH/HV	1
	14 July 2019	FSII	ASC	10	HH/HV	1
Guangzhou	13 June 2019	FSII	ASC	10	HH/HV	1
	13 June 2019	FSII	ASC	10	HH/HV	1
	11 September 2019	FSII	ASC	10	HH/HV	1
	11 September 2019	FSII	ASC	10	HH/HV	1
Shangqiu, Henan Province	26 October 2018	FSII	ASC	10	HH/HV	1
	3 March 2019	FSII	ASC	10	HH/HV	1
	22 June 2019	FSII	ASC	10	HH/HV	1
	22 June 2019	FSII	ASC	10	HH/HV	1

Table 2. Basic information of Sentinel-1 and ALOS-2/PALSAR-2 data.

Data	City	Acquisition Date	Resolution (m)	Polarization	Num.
Sentinel-1	Beijing	3 June 2019	20	VV/VH	1
ALOS-2/PALSAR-2	Beijing	25 June 2019	10	HH/HV	2
		25 June 2019	10	HH/HV	3

3.2. Samples of Various Building Areas

The spatial layout and types of buildings were different under different terrain scenes in the four regions. A total of 8984 valid samples covering multicategory building areas were produced (see Figure 6). The ratio of training data to test data was 7:3; their specific information is shown in Table 3. In the experimental area, there were more buildings in the plain area than that in the mountainous area, and the proportion of buildings in the sample was more than that of the mountainous area. Therefore, negative samples (mountain samples) were added to the training data, so that the network could fully learn the difference between the building area and the mountain.

Figure 6. Multicategory-labeled samples. (a1–f1) Google Earth image slices; corresponding (a2–f2) SAR image slices and (a3–f3) pixel-level-labeled samples. Line (a) indicates downtown built-up areas; line (b) indicates mountain villages; line (c–e) indicates plain villages; line (f) indicates urban villages.

Table 3. Basic information of training data and test data.

Type	Category	Positive Samples	Negative Samples	Total
Training samples	Downtown built-up areas	3144	0	3144
	Mountain buildings	734	1153	1887
	Plain villages	1069	0	1069
	Urban villages	188	0	188
	Total			6288
Test samples	Downtown built-up areas	1347	0	1347
	Mountain buildings	809	0	809
	Plain villages	459	0	459
	Urban villages	81	0	81
	Total			2696

4. Experimental Results and Analysis

In order to evaluate the effectiveness of the proposed method, we designed two experimental schemes. First, GF-3 SAR data were used to evaluate the effectiveness of this method in different regions with multiscale and multicategory building areas, and then the robustness of the method in different SAR data sources (Sentinel-1, ALOS-2/PALSAR-2) was analyzed.

For the residual-feature extraction network, the main parameters used in our experiment are summarized as follows: the training environment of the network was Ubuntu 16.0.4 with a Core i7 CPU and NVIDIA GTX 1080Ti 12G GPU. The learning rate, epoch and batch size were 0.0005, 30 and 5, respectively.

4.1. Experimental Results of Four Regions

Figure 7 shows the mapping results of four regions based on the pyramid multiscale building-area extraction network proposed in this paper. The results indicate that the building area extracted by the proposed method was basically consistent with the building area representation in SAR images. We can see that the detection results were very good in the urban center, mountainous buildings, suburban villages, dense contiguous building areas and small-scale villages. Figure 8 shows the local region extraction results in detail. The green marks in the first and second columns are flat-roofed

buildings and low-rise buildings. In the SAR images, flat-roofed buildings were hollow buildings, and the backscattering values of low-rise buildings were low. The green mark in the third column indicates the village in the mountain area, and the backscattering value was very low, which was difficult to detect. In the last column, buildings close to paddy fields were densely distributed, which were difficult to distinguish from non-buildings in SAR images. As could be seen from figure, all the buildings were detected completely. The global and local extraction results of Figures 7 and 8 show that the proposed method is suitable for multicategory and multiscale regional building areas mapping.

Table 4 shows the accuracy evaluation of the extraction results of the four regions. The overall accuracy of the extracted building areas at different test sites in four regions was higher than 85%, and Kappa coefficient was greater than 0.7. These results show that the proposed method has high extraction accuracy of building areas in large and small cities, in plains, mountainous areas, especially in small-scale plain villages.

To give a quantitative assessment, the results were compared with the global human settlement layer (GHSL 2014) and finer resolution observation and monitoring-global land cover 10 (FROM-GLC10). The image data of GHSL products with resolutions ranging from 0.5 to 10 m were collected by satellite SPOT (2 and 5), CBERS 2B, RapidEye (2 and 4), WorldView (1 and 2), GeoEye 1, QuickBird2, Ikonos 2 and airborne sensors. The FROM-GLC10 product is a global 10-m resolution land cover map produced by the team of Professor Gong Peng of Tsinghua University using Sentinel-2 optical data. Our results are compared with GHSL and FROM-GLC10 as shown in Figure 9, the results of each region are a whole building area layer containing different buildings. It can be seen that the building area extracted by the proposed method was basically consistent with the other two products. However, GHSL was somewhat overfitting—in which the boundary range and density of urban areas were larger than those of the other two products. For small-scale villages, FROM-GLC10 product missed most of the villages, while our results show a good detection results in small villages. The GHSL product was better than the FROM-GLC10, but worse than our result. Table 5 shows the comparison results of our results and optical products in four regions. For Beijing, our results were consistent with FROM-GLC10 product, but the commission error of FROM-GLC10 was 6.55% and that of GHSL product was 30.40%. For Wuhan, the overall accuracy of our results and optical products was more than 90%, but GHSL product had the highest commission error, which amounts to 10.86%, while FROM-GLC10 product had the highest omission error, reaching 11.76%. For Guangzhou, GHSL product had the highest accuracy and FROM-GLC10 product had the highest commission error. For Shangqiu city in Henan Province, our result had the highest accuracy among the three products, followed by GHSL product and FROM-GLC10 product had the worst extraction effect with 31.3% omission error. On the whole, the commission error of our results was the lowest among the four regions compared with the other products, However, in four regions, the omission error of our results was 6.61% in Beijing and 5.11% in Wuhan, the omission error mainly occurs in the building area with low backscattering values. The texture and geometric structure of buildings in SAR image were incomplete, thus these building features were unstable in the convolution neural network, so as to be divided into nonbuilding areas in the classification layer. For Guangzhou, our result had the highest omission error, because there were many paddy fields in Guangzhou, and some of the ridges and crops had similar characteristics with building areas in SAR data, leading to high omission error. However, from the perspective of spatial scale, our method could still accurately map the regional building areas. In addition, our method could make up for the 10-m resolution optical mapping in a certain extent in terms of small-scale village extraction and provide technical support for the application of surface monitoring based on multi-source remote sensing data.

Figure 7. Extraction results based on the proposed model. (a–d) Extraction results of Beijing, Wuhan, Guangzhou and Shangqiu, respectively. Extraction results shown in red.

Figure 8. Detailed display of local extraction results. (a) GF-3 SAR image slices; (b) extraction results; and (c) corresponding Sentinel-2 image slices.

Table 4. Accuracy evaluation of building-area extraction results in four regions.

Test sites	Kappa	Overall Accuracy (%)	Producer's Accuracy (%)	User's Accuracy (%)
Beijing (a1)	0.70	85.18	76.82	92.78
Beijing (a2)	0.84	97.82	82.64	89.29
Beijing (a3)	0.75	88.87	81.16	86.21
Wuhan (b1)	0.78	94.49	82.63	81.45
Wuhan (b2)	0.80	91.92	89.36	83.94
Wuhan (b3)	0.81	96.33	82.90	84.69
Guangzhou (c1)	0.75	85.24	80.21	86.89
Guangzhou (c2)	0.84	96.65	85.72	86.20
Guangzhou (c3)	0.76	87.19	83.99	85.75
Shangqiu (d1)	0.78	89.48	87.74	93.65
Shangqiu (d2)	0.82	96.18	88.09	92.56
Shangqiu (d3)	0.80	90.13	91.23	88.62

Figure 9. Comparison of our results and optical products.

Table 5. Accuracy comparison of our results and optical products.

Test Regions	Extraction Results	Overall, Accuracy (%)	Commission Error (%)	Omission Error (%)
Beijing	Our result	94.97	3.82	6.61
	GHSL	86.42	30.04	2.31
	FROM-GLC 10	95.25	6.55	3.80
Wuhan	Our result	96.38	3.44	5.11
	GHSL	91.10	10.86	6.16
	FROM-GLC 10	93.12	5.80	11.76
Guangzhou	Our result	95.44	5.75	1.04
	GHSL	93.52	8.32	2.43
	FROM-GLC 10	91.06	10.96	5.15
Shangqiu	Our result	96.86	3.14	3.06
	GHSL	93.10	7.23	5.51
	FROM-GLC 10	88.57	6.85	31.30

At the same time, A total of 38 GF-3 FSII data in Henan Province from October 2018 to July 2019 were selected to extract building areas of Henan Province, as shown in Figure 10. We can see that the spatial distribution of small villages in the suburbs is highly consistent with that in Google Earth images. For a clearer quantitative analysis, we counted the area of the built-up area extracted in this paper and compared it with the official statistics. The results are shown in Figure 11. The comparison result indicates that our results were roughly consistent with the official statistics obtained from the Henan Province Bureau of Statistics (<http://www.ha.stats.gov.cn/>). Here, the predicted area and the official statistical area refer to the area of built-up areas, including only the central part of the city.

Figure 10. Extraction results of Henan Province and the corresponding Google Earth image.

Figure 11. Comparison between the predicted area and the official statistical area in Henan Province.

4.2. Comparison of the Proposed Method with Fully Convolutional Network (FCN) and Residual U-Net Methods

In order to verify the extraction performance of the proposed method, we compared it with FCN and residual U-Net. FCN and residual U-Net models were trained with the same training data. The comparison of building areas extracted by the three methods is shown in Figure 12; the accuracy evaluation results are shown in Table 6. Figure 12 shows that the visualization detection results of the proposed method in four regions were the best, and the visualization detection results of FCN method in Beijing, Wuhan and Guangzhou were the worst with high omission error. In Henan Province, the detection effect of residual U-Net was the worst. It can be seen from Table 6 that the overall detection accuracy of our method in the four regions was above 95% (the average value of Kappa was 0.88), while the average kappa values of FCN and residual U-Net were 0.61 and 0.70, respectively. Compared with the proposed method, FCN in Beijing, Wuhan and Guangzhou had the highest omission error, followed by the residual U-Net method, while in Henan Province, the omission error of residual U-Net method was the highest, reaching 28.32%, which was consistent with the visualization results in Figure 12. In terms of method analysis, the residual U-Net method was generally more accurate than FCN method in urban extraction. The main reason was that residual U-Net not only used the U-shaped structure to transfer the encoded features to the decoder, but also used residual connection to ensure the transmission of detailed features in each convolution block of the encoder, while FCN did not take advantage of the feature transfer advantage of U-shaped network. Our method had the highest extraction accuracy, which was mainly attributed to using ResNet50 as the feature extraction network. The network not only ensures the deepening of depth, but also solves the problem of gradient disappearance. The texture and geometric structure features of the building area were fully learned by the network, and then the pyramid-pooling module was introduced to further extract the characteristics of different sub regions of the building area. Finally, the deep level features and multiscale features were combined to realize the accurate extraction of multiscale building areas.

Figure 12. Extraction results of four regions based on three methods.

Table 6. Accuracy evaluation of extraction results based on three methods.

Region	Method	Kappa	Overall, Accuracy (%)	Commission Error (%)	Omission Error (%)
Beijing	Proposed method	0.88	96.17	4.32	1.73
	FCN	0.64	79.39	2.73	34.31
	Residual U-Net	0.72	82.58	3.63	18.08
Wuhan	Proposed method	0.89	95.86	4.92	1.58
	FCN	0.57	72.48	1.86	37.54
	Residual U-Net	0.73	86.32	3.24	15.36
Guangzhou	Proposed method	0.87	95.65	5.16	1.05
	FCN	0.47	76.42	1.61	35.24
	Residual U-Net	0.72	91.06	2.52	18.86
Shangqiu	Proposed method	0.89	96.99	2.91	3.46
	FCN	0.79	89.53	1.11	10.90
	Residual U-Net	0.64	88.53	2.79	28.32

4.3. Evaluating the Robustness of the Proposed Method in Different SAR Data Sources

In order to evaluate the robustness of proposed method using different SAR data, Sentinel-1 and ALOS-2/PALSAR-2 data were selected as experimental data. Table 7 shows the detailed information of training data and test data. Figures 13 and 14 show the extraction results using Sentinel-1 and ALOS-2/PALSAR-2 data. Table 8 shows the accuracy evaluation result. In central city and suburb, the overall extraction accuracy was higher than 93%. The omission error of the results based on

Sentinel-1 data in central city was the highest due to low backscattering value of some building areas. While the commission error of the results based on ALOS-2/PALSAR-2 data were higher because some farmland show similar characteristic as buildings. On the whole, the experimental results show that the multiscale building-area extraction network had good generalization ability in regional urban mapping using different SAR data.

Table 7. Basic information on training samples and test samples.

Data	Pol.	Training Samples	Test Samples
Sentinel-1	VV	686	70
ALOS-2/PALSAR-2	HH	686	70

Figure 13. Building area extraction results based on the proposed method using Sentinel-1 data.

Figure 14. Extraction results based on the proposed method using ALOS-2/PALSAR-2 data.

Table 8. Accuracy evaluation of the extraction results based on Sentinel-1 and ALOS-2 SAR data.

Data	Region	Overall, Accuracy (%)	Commission Error (%)	Omission Error (%)
Sentinel-1	Central city	93.32	2.54	6.71
	Suburb	96.21	2.50	1.52
ALOS-2/PALSAR-2	Central city	95.74	4.67	2.89
	Suburb	93.57	7.55	3.80

5. Discussion

It can be seen from the experimental results that all types of buildings under different terrain environment were well-detected, which shows that multicategory labeled samples are necessary. Compared with FCN-based methods, the ResNet50 and pyramid models can learn deeper and multiscale features of buildings. Using Sentinel-1 and ALOS-2/PALSAR-2 data for regional building-area mapping also achieved considerable accuracy, indicating that the pyramid multiscale building-area extraction network has good generalization performance. These results show that our building extractor is suitable for regional urban mapping using medium resolution SAR data, and it is feasible to apply deep-learning methods and SAR data to regional or global urban mapping.

To date, we have achieved promising building-area extraction results, but there are still some commission and omission error in the four regions. There is spatial correlation between SAR image pixels, thus the landscape near the building area, small villages in paddy fields, roads and ridge are easy to be detected as building targets. The postprocessing module is used to optimize the initial extraction results and the extraction accuracy is improved. The main reasons to omission error is that some GF-3 SAR data are difficult to distinguish the buildings from the background due to the low image contrast. In addition, the quantity and quality of pixel level labeled samples have certain limitations on the accuracy of building mapping and the error of manual marking will lead to omission error. Therefore, future work can focus on building-area extraction methods based on few-shot learning methods and improving SAR image contrast.

6. Conclusions

According to the application requirements of regional and global scale urban mapping, in this paper, a multiscale urban extraction network was designed, and a regional urban mapping extraction framework based on SAR data has been proposed. GF-3 SAR data with 10-m resolution were used for regional urban mapping, the experimental results of four different regions show that the proposed method could accurately extract building areas and villages in different terrain environments, with the overall accuracy higher than 85%. Compared with optical products, the building area extracted by this method was highly consistent with optical products in the central urban area, and the extraction accuracy in rural areas was significantly higher than that of optical products. In addition, compared with FCN and residual U-Net methods, the visualization and quantitative evaluation in four regions show that the proposed approach had stronger building detection ability than FCN and residual U-Net methods.

Moreover, Sentinel-1 and ALOS-2/PALSAR-2 data were used to test the applicability of the algorithm. The experimental results in Beijing city showed that the extraction accuracy of the proposed method was better than 93% in the two kinds of data, indicating that our method had good data robustness and was suitable for a variety of medium resolution SAR data.

These findings indicate that the proposed method is suitable for multiscale building-area extraction in different terrain environment and has strong robustness in the application of building-area extraction from different SAR data. In the future, we will study the sustainable urban development based on long time series data to support urban sustainable development.

Author Contributions: J.L. and H.Z. designed and performed the experiments, H.Z. and C.W. supervised and designed the research and contributed to the paper’s organization, F.W., L.L. and J.L. conducted experimental analysis, H.Z. and J.L. responsible for original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grants 41930110, and the JAXA Earth Observation Research Center under EORA2_A2 PI No. ER2A2N127.

Acknowledgments: The authors would like to thank China Center for Resources Satellite Data and Application for providing Gaofen-3 images, and ESA and EU Copernicus Program for providing the Sentinel-1A SAR data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Economic & Social Affairs. *World Urbanization Prospects—The 2018 Revision*; Technical Report; New York: United Nations: New York City, NY, USA, 31 December 2018.
2. Zhu, Z.; Zhou, Y.; Seto, K.; Stokes, E.; Deng, C.; Pickett, S.T.S.; Taubenböck, H. Understanding an Urbanizing Planet: Strategic Directions for Remote Sensing. *Remote Sens. Environ.* **2019**, *228*, 164–182. [\[CrossRef\]](#)
3. Donnay, J.P.; Barnsley, M.J.; Longley, P.A. *Remote Sensing and Urban Analysis: GISDATA 9*; CRC Press: Boca Raton, FL, USA, 2000.
4. Taubenböck, H.; Wegmann, M.; Berger, C.; Breunig, M.; Roth, A.; Mehl, H. Spatiotemporal analysis of Indian megacities. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, China, 3 July 2008; pp. 75–82.
5. Defries, R.S.; Townshend, J.R.G. NDVI-derived land cover classifications at a global scale. *Int. J. Remote Sens.* **1994**, *15*, 3567–3586. [\[CrossRef\]](#)
6. Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331–1364. [\[CrossRef\]](#)
7. Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [\[CrossRef\]](#)
8. Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [\[CrossRef\]](#)
9. Arino, O.; Bicheron, P.; Frédéric, A.; Latham, J.; Witt, R.; Weber, J.-L. GLOBCOVER: The most detailed portrait of Earth. *Eur. Space Agency Bull.* **2008**, *2008*, 24–31.
10. Friedl, M.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of new Datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [\[CrossRef\]](#)
11. Fritz, S.; See, L.; Rembold, F. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *Int. J. Remote Sens.* **2010**, *31*, 2237–2256. [\[CrossRef\]](#)
12. Schneider, A.; Friedl, M.; Potere, D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on ‘urban ecoregions’. *Remote Sens. Environ.* **2010**, *114*, 1733–1746. [\[CrossRef\]](#)
13. Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tseng-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaadeh, B.; et al. Production of global land cover data—GLCNMO. *Int. J. Digit. Earth* **2011**, *4*, 22–49. [\[CrossRef\]](#)
14. Yu, L.; Wang, J.; Gong, P. Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: A segmentation-based approach. *Int. J. Remote Sens.* **2013**, *34*, 5851–5867. [\[CrossRef\]](#)
15. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [\[CrossRef\]](#)
16. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370. [\[CrossRef\]](#)
17. Longley, P. Global Mapping of Human Settlement: Experiences, Datasets, and Prospects: Book Reviews. *Photogramm. Rec.* **2010**, *25*, 205–207. [\[CrossRef\]](#)

18. Pesaresi, M.; Ehrlich, D.; Caravaggi, I.; Kauffmann, M.; Louvrir, C. Toward Global Automatic Built-Up Area Recognition Using Optical VHR Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 923–934. [[CrossRef](#)]
19. Pesaresi, M.; Huadong, G.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2102–2131. [[CrossRef](#)]
20. Kobayashi, T.; Satake, M.; Masuko, H.; Manabe, T.; Shimada, M. CRL/NASDA airborne dual-frequency polarimetric interferometric SAR system. In Proceedings of the SPIE—The International Society for Optical Engineering, San Jose, CA, USA, 26–28 January 1998. [[CrossRef](#)]
21. Esch, T.; Taubenbck, H.; Roth, A.; Heldens, W.; Felbier, A.; Thiel, M.; Schmidt, M.; Mller, A.; Dech, S. Tandem-X Mission—New Perspectives for the Inventory and Monitoring of Global Settlement Patterns. *J. Appl. Remote Sens.* **2012**, *6*, 1702. [[CrossRef](#)]
22. Esch, T.; Marconcini, M.; Felbier, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M.; Taubenbck, H.; Mller, A.; Dech, S. Urban Footprint Processor-Fully Automated Processing Chain Generating Settlement Masks from Global Data of the TanDEM-X Mission. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1617–1621. [[CrossRef](#)]
23. Felbier, A.; Esch, T.; Heldens, W.; Marconcini, M.; Zeidler, J.; Roth, A.; Klotz, M.; Wurm, M.; Taubenbck, H. The Global Urban Footprint—Processing Status and Cross Comparison to Existing Human Settlement Products. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec, QC, Canada, 13–18 July 2014; pp. 4816–4819. [[CrossRef](#)]
24. Gessner, U.; Machwitz, M.; Esch, T.; Bertram, A.; Naeimi, V.; Kuenzer, C.; Dech, S. Multi-sensor mapping of West African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data. *Remote Sens. Environ.* **2015**, *164*, 282–297. [[CrossRef](#)]
25. Klotz, M.; Kemper, T.; Gei, C.; Esch, T.; Taubenbck, H. How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sens. Environ.* **2016**, *178*, 191–212. [[CrossRef](#)]
26. Esch, T.; Thiel, M.; Schenk, A.; Roth, A.; Mller, A.; Dech, S. Delineation of Urban Footprints from TerraSAR-X Data by Analyzing Speckle Characteristics and Intensity Information. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 905–916. [[CrossRef](#)]
27. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 30–42. [[CrossRef](#)]
28. Gamba, P.; Lisini, G. Fast and Efficient Urban Extent Extraction Using ASAR Wide Swath Mode Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2184–2195. [[CrossRef](#)]
29. Ban, Y.; Jacob, A.; Gamba, P. Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor. *ISPRS J. Photogramm. Remote Sens.* **2014**, *103*, 28–37. [[CrossRef](#)]
30. Jacob, A.; Ban, Y. Sentinel-1A SAR data for global urban mapping: Preliminary results. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 1179–1182. [[CrossRef](#)]
31. Cao, H.; Zhang, H.; Wang, C.; Zhang, b. Operational Built-Up Areas Extraction for Cities in China Using Sentinel-1 SAR Data. *Remote Sens.* **2018**, *10*, 874. [[CrossRef](#)]
32. Sun, Z.; Xu, R.; Du, W.; Wang, L.; Lu, D. High-Resolution Urban Land Mapping in China from Sentinel 1A/2 Imagery Based on Google Earth Engine. *Remote Sens.* **2019**, *11*, 752. [[CrossRef](#)]
33. Niu, X.; Ban, Y. Multitemporal RADARSAT-2 Polarimetric SAR Data for Urban Land Cover Classification using Object-based Support Vector Machine and Rule-based Approach. *Int. J. Remote Sens.* **2012**, *34*, 1–26. [[CrossRef](#)]
34. Geng, J.; Wang, H.; Fan, J.; Ma, X. Deep Supervised and Contractive Neural Network for SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2442–2459. [[CrossRef](#)]
35. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

36. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
37. Zhang, Z.; Liu, Q. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 749–753. [[CrossRef](#)]
38. Li, L.; Wang, C.; Zhang, H.; Zhang, B.; Wu, F. Urban Building Change Detection in SAR Images Using Combined Differential Image and Residual U-Net Network. *Remote Sens.* **2019**, *11*, 1091. [[CrossRef](#)]
39. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sens.* **2019**, *11*, 68. [[CrossRef](#)]
40. Li, J.; Zhang, R.; Li, Y. Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 910–913. [[CrossRef](#)]
41. Gao, D.-L.; Zhang, R.; Xue, D.-X. Improved Fully Convolutional Network for the Detection of Built-Up Areas in High Resolution SAR Images. In *Image and Graphics*; Zhao, Y., Kong, X., Taubman, D., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; Volume 10668. [[CrossRef](#)]
42. Wu, Y.; Zhang, R.; Li, Y. The Detection of Built-up Areas in High-Resolution SAR Images Based on Deep Neural Networks. In *Lecture Notes in Computer Science*; Zhao, Y., Kong, X., Taubman, D., Eds.; Springer: Cham, Switzerland, 2017; Volume 10668, pp. 646–655. [[CrossRef](#)]
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
44. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
45. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 2999–3007. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).