

Article

Rural Built-Up Area Extraction from Remote Sensing Images Using Spectral Residual Methods with Embedded Deep Neural Network

Shaodan Li ^{1,2,*} , Shiyu Fu ^{1,2} and Dongbo Zheng ^{1,2} 

¹ School of Geographical Sciences, Hebei Normal University, Shijiazhuang 050024, China; fusy201909@163.com (S.F.); dongbozheng@hebtu.edu.cn (D.Z.)

² Hebei Technology Innovation Center for Remote Sensing Identification of Environmental Change, Shijiazhuang 050024, China

* Correspondence: lishaodan@hebtu.edu.cn

Abstract: A rural built-up area is one of the most important features of rural regions. Rapid and accurate extraction of rural built-up areas has great significance to rural planning and urbanization. In this paper, the spectral residual method is embedded into a deep neural network to accurately describe the rural built-up areas from large-scale satellite images. Our proposed method is composed of two processes: coarse localization and fine extraction. Firstly, an improved Faster R-CNN (Regions with Convolutional Neural Network) detector is trained to obtain the coarse localization of the candidate built-up areas, and then the spectral residual method is used to describe the accurate boundary of each built-up area based on the bounding boxes. In the experimental part, we firstly explored the relationship between the sizes of built-up areas and the kernels in the spectral residual method. Then, the comparing experiments demonstrate that our proposed method has better performance in the extraction of rural built-up areas.

Keywords: rural built-up area extraction; remote sensing; spectral residual; deep neural network



Citation: Li, S.; Fu, S.; Zheng, D. Rural Built-Up Area Extraction from Remote Sensing Images Using Spectral Residual Methods with Embedded Deep Neural Network. *Sustainability* **2022**, *14*, 1272. <https://doi.org/10.3390/su14031272>

Academic Editors: Jing Wei, Naisen Yang and Hong Tang

Received: 6 January 2022

Accepted: 21 January 2022

Published: 24 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

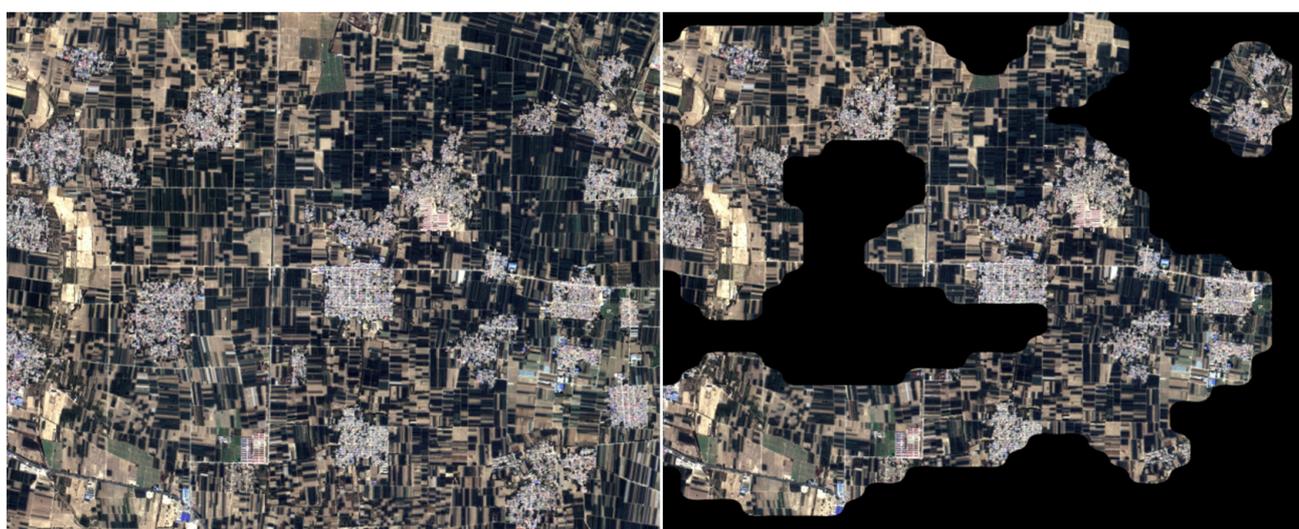
A built-up area is one of the most obvious targets in remote sensing images, and a rural built-up area is a very important part that cannot be ignored. The rural built-up area contains not only the rural buildings but also the area among buildings, such as roads, trees, vegetation and other man-made objects. Rural built-up areas play an extremely important role in agricultural production and the lives of farmers. Rapid and accurate extraction and mapping of rural built-up areas has great significance in various fields, such as rural development and planning, rural land management and monitoring, rural population estimates and rural modernization process.

In recent years, a number of automatic built-up area detection methods have been proposed. These methods can fall into the following three categories: typically index-based methods, image classification, and visual attention models. For the index-based methods, Zha et al. [1] presented a method based on normalized different built-up index (NDBI), which was successfully applied to extract the urban built-up areas using Landsat TM imagery. Similar various indices with NDBI (such as PanTex [2], IBI [3], NBI [4], and MBI [5]) are also proposed to extract the built-up areas. The index-based methods are simple and rapid, however, the formula of the index must be changed with different sensors and the optimal threshold is difficult to determine.

The image classification method is the most common method for extracting built-up areas. This method of image classification can be roughly divided into two categories: unsupervised classification and supervised classification. For the unsupervised classification, Tao et al. [6] first located candidate built-up regions with an improved Harris corner

detection, then detected built-up areas using an unsupervised classification with texture histogram modeling, spectral clustering, and graph-cuts. Sirmacek and Unsalan [7] first extracted the edges and corners of buildings in different orientations using Gabor filters, and then used these local feature points to vote for candidate urban areas. Chen et al. [8] extracted building edges with the Canny operator and configured them into several straight lines, and then formed a spatial voting matrix to extract the built-up areas. As for the supervised classification, they need a set of specific training samples to learn features for detection. Pang et al. [9] used a support vector machine (SVM) to detect built-up areas with efficient textural features extracted from the contourlet transform domain. Zhong et al. [10] presented an ensemble model to merge multiple features and learn their contextual information to obtain an urban area perspective. Other machine learning techniques are also utilized for extracting built-up areas, for examples, decision trees [11] and random forests [12,13].

In terms of visual attention models, it is a technique to derive important and prominent information from a scene in natural pictures. In recent years, visual attention methods are employed in remote sensing images for built-up area detection. Li et al. [14] proposed an improved Itti model to detect the salient targets in remote sensing images. Zhang et al. [15] detected the built-up areas in frequency domain with the Fourier transform. Li et al. [16] used spectral residual (SR) method to extract rural residential regions in Gaofen-1 images. The performance of the SR method is perfect in small-scale regions, but the result is unsatisfactory in a large-scale image, which is shown in Figure 1.



(a) Satellite image

(b) The extracted built-up areas

Figure 1. Extracted result of built-up areas in a large-scale satellite image. (a) Satellite image; (b) The extracted built-up areas.

Fortunately, the rapid development of the deep neural network (DNN) within the ImageNet contest [17–20] in recent years has given an opportunity for its use in remote sensing images and many other fields [21–24]. The traditional methods mostly focus on small regions, and DNN gives a good prospect for the extraction of built-up areas in large-scale images. Tan et al. proposed a segmentation framework based on deep feature learning and graph models to extract built-up areas [25,26]. Zhang et al. extracted the built-up areas based on the convolutional neural network (CNN), and chose Beijing, Lanzhou, Chongqing, Suzhou and Guangzhou of China as the experimentation sites [27,28]. Iqbal et al. proposed a weakly-supervised adaptation strategy and designed a built-up area segmentation network to extract built-up areas in diverse built-up scenarios in Rwanda [29]. Ma et al. proposed a new fusion approach for accurately extracting urban built-up areas based on the use of multisource remotely sensed data, i.e., the DMSP-OLS nighttime light

data, the MODIS land cover product and Landsat 7 ETM+ images [30]. However, these studies mostly are concentrated in urban areas, while there are relatively few studies in rural areas. Different from urban areas, rural built-up areas are smaller and more scattered, and they are often staggered with vegetation and farmland. Therefore, the results may be not satisfactory by using the method based on urban areas directly.

In this paper, the spectral residual (SR) method is embedded into the deep neural network to extract the built-up areas from large-scale satellite images, and the study area is focused on rural regions. Our proposed method is composed of two processes: coarse localization and fine extraction. Firstly, an improved Faster R-CNN detector is trained to obtain the coarse localization of the candidate built-up areas, and then the SR method is used to extract the accurate boundary of each built-up area based on the bounding boxes. It should be noted that the experimental part is another highlight of this paper. In the experimental part, we firstly explored the relationship between the sizes of built-up areas and discussed the kernels in the SR method. Then, the proposed method is compared to other methods, and the experiments demonstrate that our proposed method has a higher accuracy. The remainder of this paper is organized as follows: Section 2 introduces the overall architecture of our approach. Section 3 shows our experiments and results. Section 4 then discusses the results. Finally, a conclusion is drawn in Section 5.

2. Methods

As illustrated in Figure 2, the proposed method contains two processes to extract the rural built-up areas: coarse localization and fine extraction. Firstly, in the coarse localization, an improved Faster R-CNN detector is utilized to produce the bounding box of the candidate built-up areas in large-scale satellite images; meanwhile the indicated possibility of being a built-up area is also recorded. Then, in the fine extraction, the SR method is used to describe the accurate boundary of each built-up area based on the bounding boxes.



Figure 2. The flowchart of the proposed method for rural built-up area extraction.

2.1. Coarse Localization

In our proposed method, an improved Faster R-CNN model with a ResNet-FPN backbone is used to obtain the coarse localization of the candidate built-up areas in satellite images, and meanwhile the indicated possibility of being a built-up area is also recorded. The framework of detecting the bounding boxes of built-up areas is shown in Figure 3. The Faster R-CNN module consists of three steps, which is described as follows:

1. ResNet-FPN is chosen as the backbone in the Faster R-CNN. The residual networks (ResNet) are easier to optimize by residual learning in deeper neural network, and a 50-layer ResNet [20] is adopted in our framework. However, if the size of built-up areas to be detected is small, the information on the final feature map may disappear due to continuous down-sampling. In order to solve the problem, the feature pyramid network (FPN) [31] is adopted to the ResNet. The different scales in FPN can merge low-level location with high-level semantics in order to detect the built-up areas with different sizes in large-size satellite images.
2. In the original region proposal networks (RPN) design, a small subnetwork performs built-up area/non built-up area classification and bounding box regression on a

single scale convolution feature map. In the proposed framework, we adapt RPN by replacing the single-scale feature map with the FPN. As shown in Figure 3, the feature maps with different scales provided by the ResNet-FPN backbone are fed in the RPN, respectively, to obtain more potential built-up proposals, which can increase the accuracy of different sizes of built-up areas.

3. In the proposed framework, the RoIAlign [32] is adapted by replacing RoIPooling. The RoIAlign leads to large improvements by using bilinear interpolation to compute the exact values of the input features at four regularly sampled locations in each RoI bin, and aggregating the result. The RoIAlign processes the proposals with different sizes into a fixed size, and then they are input to the full connection layers for final classification and location refinement. Finally, the bounding box of each built-up area is detected; meanwhile, the indicated possibility of being a built-up area is also recorded.

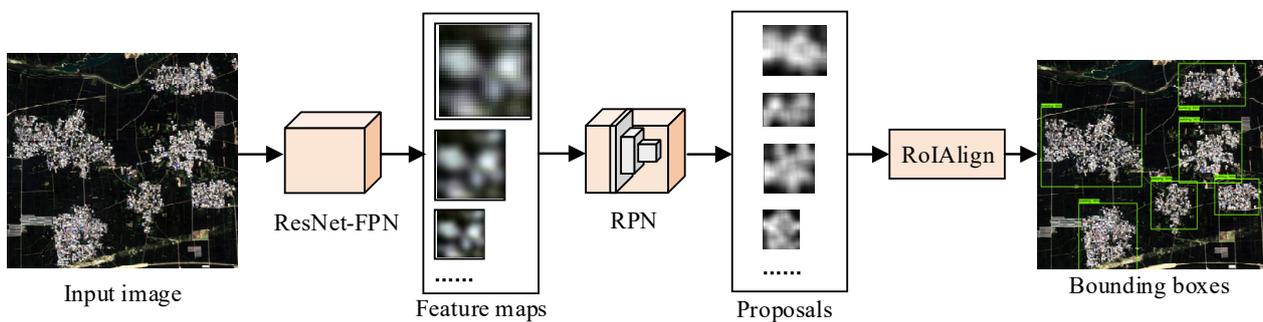


Figure 3. The framework of detecting the bounding boxes of built-up areas.

2.2. Fine Extraction

Based on the bounding boxes provided by the coarse localization, the accurate boundary of each built-up area is extracted by using the SR method. The SR method is one of the visual attention models that detects salient objects in the pictures. Hou et al. discovered the relationship between the human visual system and the log spectrum. By analyzing the log spectrum of a large number of natural pictures, they found that they shared similar trends. Therefore, it is assumed that the part which jumps out of the average trend indicates the salient information of the image [33]. For a large-scale satellite image, the built-up area is the salient information of the image. The framework of the SR for built-up area extraction from the bounding boxes is shown in Figure 4. Given an input image $I(x)$ with the detected bounding boxes, the built-up area map is computed as follows:

1. The image f in frequency domain is computed by the formula

$$f = F(I(x)). \quad (1)$$

where F refers to the Fourier transform.

2. The spectral residual $R(f)$ of the image is defined by

$$R(f) = L(f) - h(f) \times L(f), \quad (2)$$

where $L(f) = \log(A(f))$, and $A(f)$ denotes the amplitude spectrum of the frequency f . $h(f)$ is a local average filter, and $h(f) \times L(f)$ is the average log amplitude spectrum, which indicates the general shape of log spectra. Thus, the spectral residual $R(f)$ indicates the salient objects (i.e., the built-up areas) from the image.

3. The final saliency map in spatial domain is computed by the formula

$$S(x) = F^{-1}[\exp((R(f) + P(f)))]^2, \quad (3)$$

where F^{-1} refers to the inverse Fourier transform, and $P(f)$ denotes the phase spectrum of the frequency f .

4. Based on the saliency map, the Otsu threshold is used to obtain the binary image of the built-up areas.

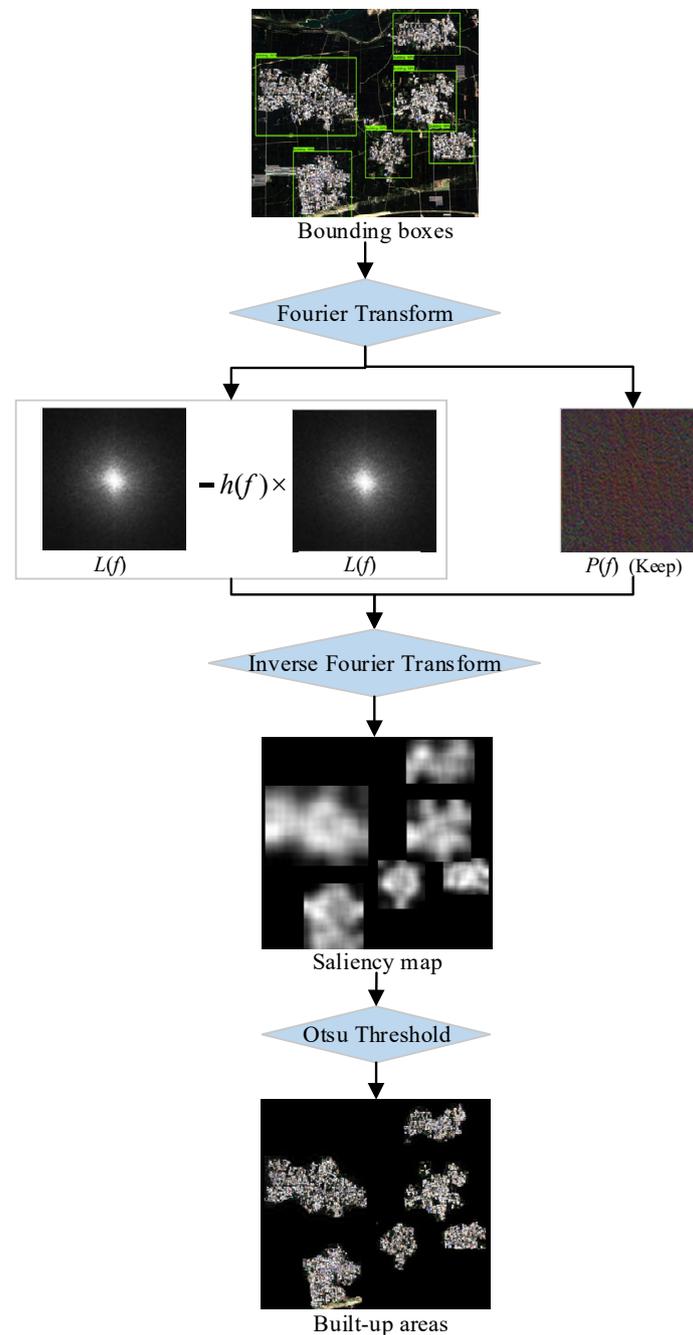


Figure 4. The flowchart of the SR to extract built-up areas from the bounding boxes.

3. Experiments and Results

3.1. Experimental Data

In this paper, experimental datasets are captured by Gaofen-1 (GF-1), Ziyuan-3 (ZY-3) and WorldView-2 (WV-2) remote sensing satellite images, mainly located in Hebei province in China. The GF-1 remote sensing satellite was successfully launched on 26 April 2013, which can provide panchromatic images with 2 m resolution and multispectral images with resolution of 8 m. The ZY-3 remote sensing satellite was launched on 9 January 2012, which

can provide panchromatic images with 2.1 m resolution and multispectral images with resolution of 5.8 m. WV-2 satellite can provide panchromatic images with 0.5 m resolution and multispectral images with resolution of 1.8 m, and the parameters of three sensors are shown in Table 1.

Table 1. The parameters of GF-1, ZY-3, and WV-2 sensors.

Sensors	Shooting Time	Panchromatic		MultiSpectral	
		Resolution (m)	Spectrum (μm)	Resolution (m)	Spectrum (μm)
GF-1	2017.04.22	2	0.45–0.90	8	0.45–0.52 0.52–0.59 0.63–0.69 0.77–0.89
ZY-3	2018.04.07	2.1	0.50–0.80	5.8	0.45–0.52 0.52–0.59 0.63–0.69 0.77–0.89
WV-2	2018.08.25	0.5	0.45–1.04	1.8	0.45–0.51 0.51–0.58 0.63–0.69 0.77–0.89

The creation of the dataset in this experiment takes the VOC2012 dataset as a reference, and the dataset consists of three parts: images, annotations, and index file. Data augmentation is widely used for preventing overfitting in deep neural networks, and it is essential to train the network on the desired invariance and robustness properties, when only few training samples are available. Abundant spectral information, which is underutilized by the deep neural networks, is contained in remote sensing images, and it is the most notable characteristic that differs from other image datasets. In this paper, in addition to the most common forms of data augmentation, such as flipping, cropping, and rotation, a new form DropBand is employed [34]. This method executes this operation by all the bands of an input image. With dropping a band of images out, the error rate of the deep neural networks can be reduced.

The framework of automatic detection of the bounding box of built-up areas from satellite images in a large scale depends on deep neural network, whose performance relies on the training data. In order to achieve the satisfactory result, a total of 47,088 samples are collected from GF-1, ZY-3, and WV-2 sensors, and the sample set is randomly divided into the training set, validation set and testing set with the proportion 6:2:2, details are shown in Table 2. There are no overlaps among the two data sets. Each sample contains only one built-up area, and the size of each built-up area is inconsistent.

Table 2. The division of sample dataset.

	Images	Built-Up Areas
Training set	789	29,608
Validation set	256	8662
Testing set	270	8818

The framework is implemented by pytorch framework and Python. The hardware platform includes 64 G memory, Intel Xeon E5-2643 v4 CPU and a NVIDIA Quadro P4000 GPU. In the training stage, the corresponding parameters are summarized as follows: the warm up strategy is used in the training process, and the initial learning rate is set to 0.005. After 5 epochs, the learning rate decreases by 0.33 time, optimization algorithm as momentum + SGD (learning rate = 0.005, momentum = 0.9, weight_decay = 0.0005).

3.2. Impact of the Sizes of Built-Up Areas

In the proposed method, the accurate boundary of each built-up area is extracted using the SR method after getting the bounding boxes. In this stage, there are two kernels to be specified. The first kernel refers to the local average filter $h(f)$ in the formula (2) in Section 2.2. The second kernel refers to a Gaussian filter in order to better visual effects in [33].

Based on the observation of the rural built-up areas, it is found that the size of each rural built-up area is diverse. Some built-up areas are larger, and some are smaller. Figures 5a and 6a give two different sizes of the built-up areas. Figure 5 shows the results of three different sizes of the second kernel for a smaller built-up area with the size of 256×239 , which the first kernel is specified as 3. Experimental results indicate that: (1) the size of 35 gives the best result among the three sizes of the second kernel. (2) When the size of the kernel is 7, the extracted result is more fragmented (Figure 5b); when the size is 101, the result is smoother (Figure 5d). The results of the two sizes are not satisfactory. As shown in Figure 6, the results of the same three sizes of the second kernel for a larger built-up area with the size of 688×1069 are given, which the first kernel is also specified as 3. Experimental results indicate that the size of 101 gives the best result, and the other two sizes do not perform so well.

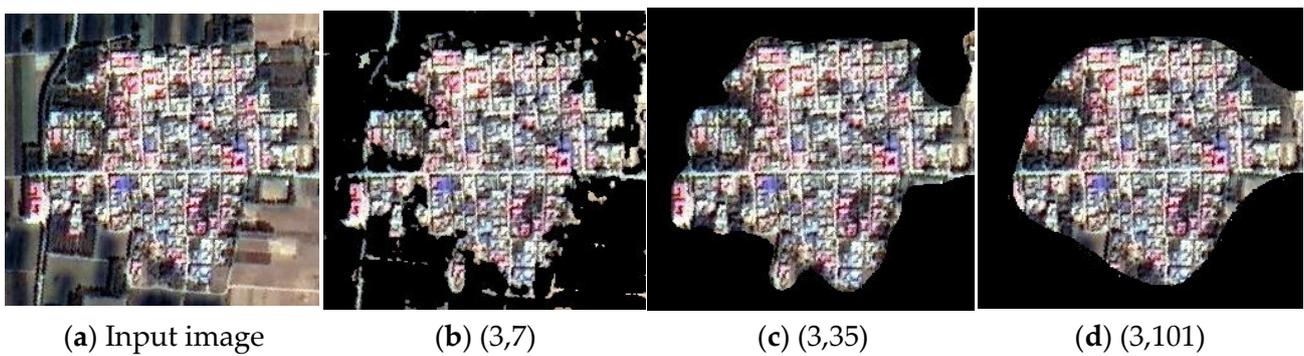


Figure 5. The results of the different kernels for a built-up area with smaller size.

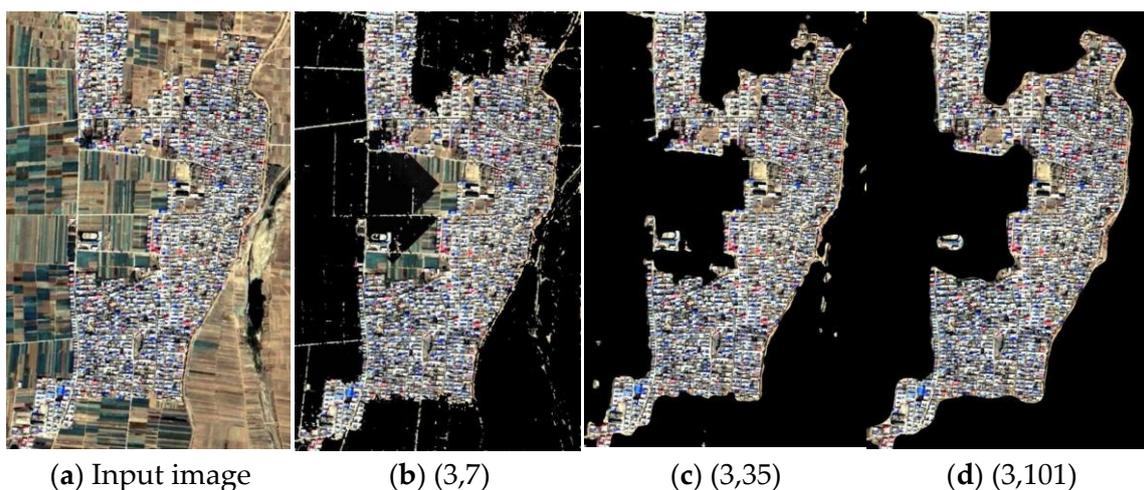


Figure 6. The results of the different kernels for a built-up area with larger size.

The two experiments indicate that the size of the two kernels directly affects the results of built-up area extraction, and the size of the kernel should be different for the different sizes of the built-up areas. In order to choose the optimal sizes for the different sizes of the built-up areas, the experiments are carried out in this subsection.

According to the sizes of the built-up areas, the built-up areas are divided into four groups. The range of the first group is about 150×150 pixels. The range of the second group is about 250×250 pixels. The range of the third group is between 500×500 pixels and 700×700 pixels. The range of the fourth group is greater than 900×900 pixels. For each group, we have tried to find the optimal sizes of two kernels.

In the experiments, the overall precision (P), recall (R) and F -Measure (F) [35] are used to evaluate the performance of the algorithm to extract built-up areas. P , R , and F are defined as:

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \tag{6}$$

where TP is correctly detected pixels by using algorithm among the ground truth. FP is the pixels detected using algorithm but not in the ground truth, and FN is the pixels which are not detected using algorithm but in the ground truth. The β^2 is a positive parameter for weighting the precision and recall (β^2 is chosen as 2 in this paper). The F -measure is the harmonic mean of the precision and recall.

Firstly, the size of the first kernel is fixed, such as 7, the optimal sizes of the second kernel are tried to find for each group. Figure 7 gives the precision (i.e., F -measure) curves of different sizes of the second kernel for four groups. The precision curves of different sizes of the second kernel show that (1) the precision increases first and then decreases, and the optimal sizes of the second kernel for four groups are about 35, 55, 85 and 125, respectively. (2) As the size of the built-up areas increases, the optimal size of the corresponding kernel also increases.

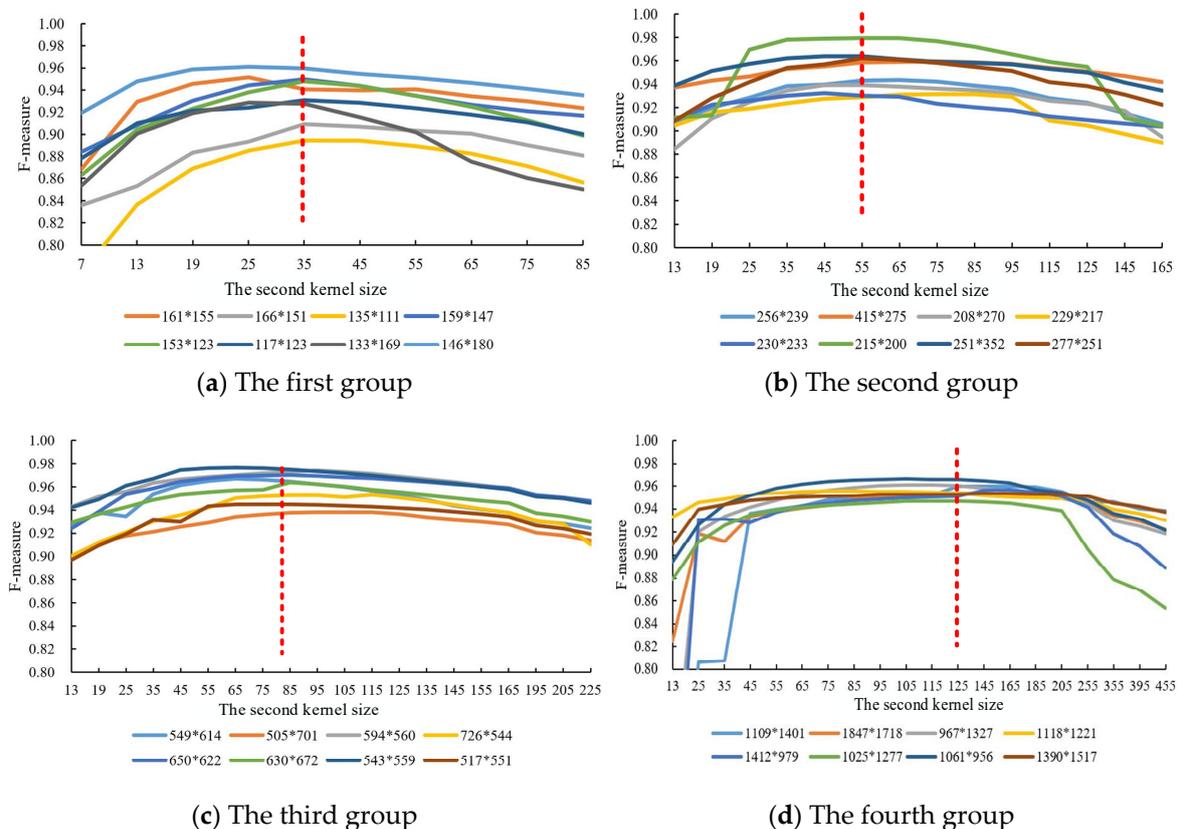


Figure 7. The precision curves of different sizes of the second kernel (the first kernel is set as 7).

Then, the size of the second kernel is fixed, the optimal sizes of the first kernel are tried to find for four groups. It should be noted that the size of the second kernel of each group is set to 35, 55, 85 and 125, respectively. Figure 8 gives the precision curves of different sizes of the first kernel for four groups. Experimental results reveal that the precision curves of different sizes of the first kernel first stays stable and then decreases. The optimal sizes of the first kernel for the first three groups is about 3–7, and the fourth group is about 5–11. The optimal sizes of the first kernel do not increase as the size of the built-up area increases, which is consistent with the conclusion in paper [30]. When the size of the built-up area is moderate, the built-up area extraction results are the best with the optimal size of 3–7. However, if the size of the built-up area is very large, the optimal size of the kernel can be increased slightly, such as 9 or 11.

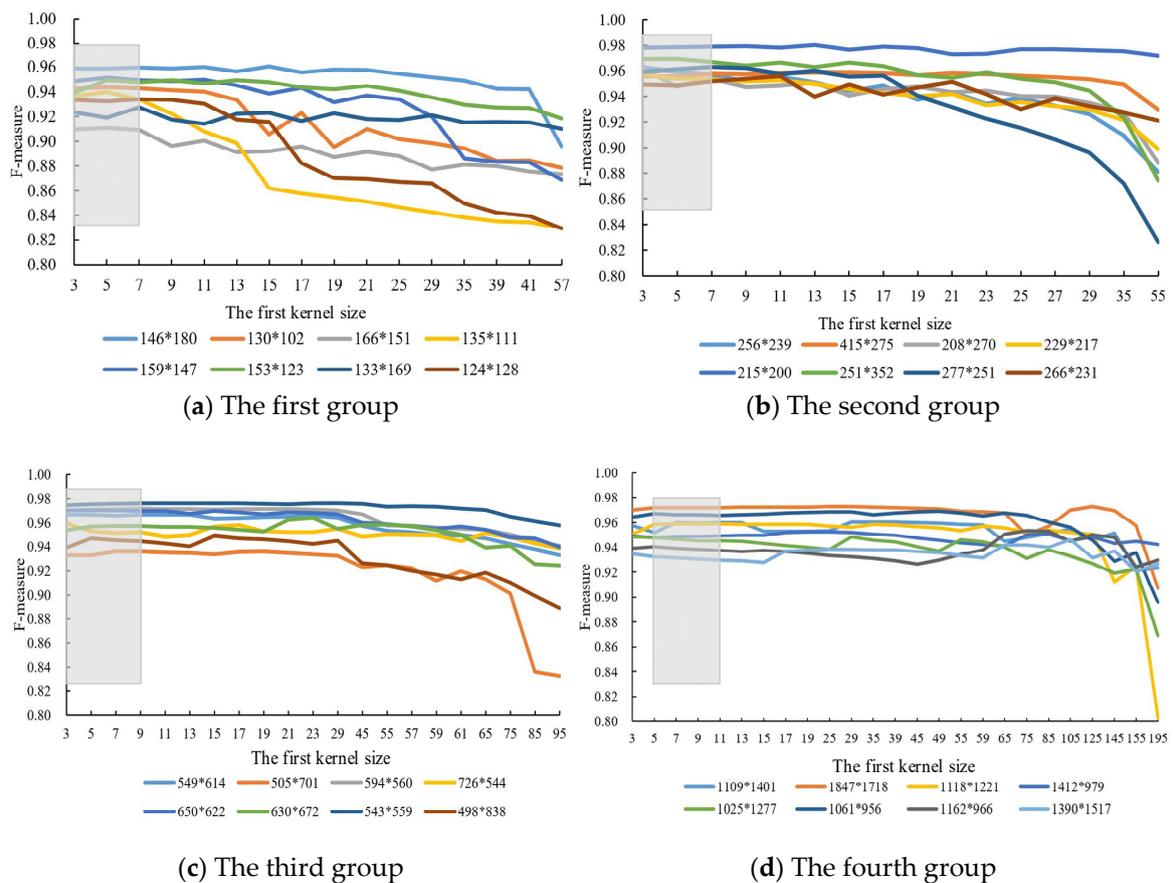


Figure 8. The precision curves of different sizes of the first kernel.

In the proposed method, after getting the bounding boxes of the built-up area, we determine which group the length and width of each built-up area belongs to. That is, the optimal values of the two kernels are dynamically applied in the algorithm.

3.3. Comparison with Other Algorithms

We compare the proposed method with several built-up area extraction algorithms, and they are anisotropic rotation-invariant textural measure (PanTex) [2], Gabor [7] and morphological building index (MBI) [5]. We test the above algorithms in two images which are located in Hebei province, and the comparison of results are shown in Figures 9 and 10. As we can see, our proposed method has the best performance. The visual result of Gabor is better than PanTex, and The MBI misses most of the built-up areas, since the MBI was originally designed to extract buildings. The 8 image, which are selected from the above two images, are enlarged to show the details of the built-up areas more clearly, as illustrated in Figure 11.

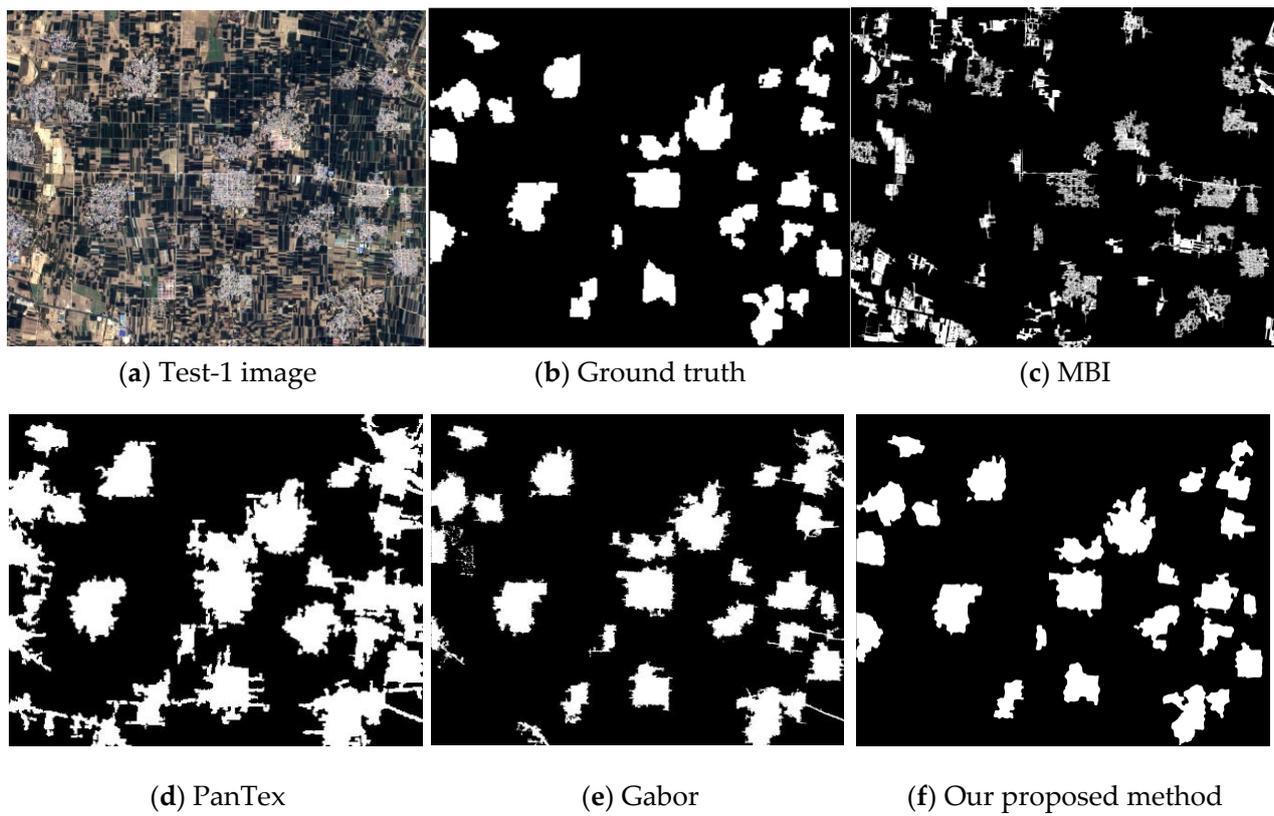


Figure 9. Comparison of results for test-1 image.

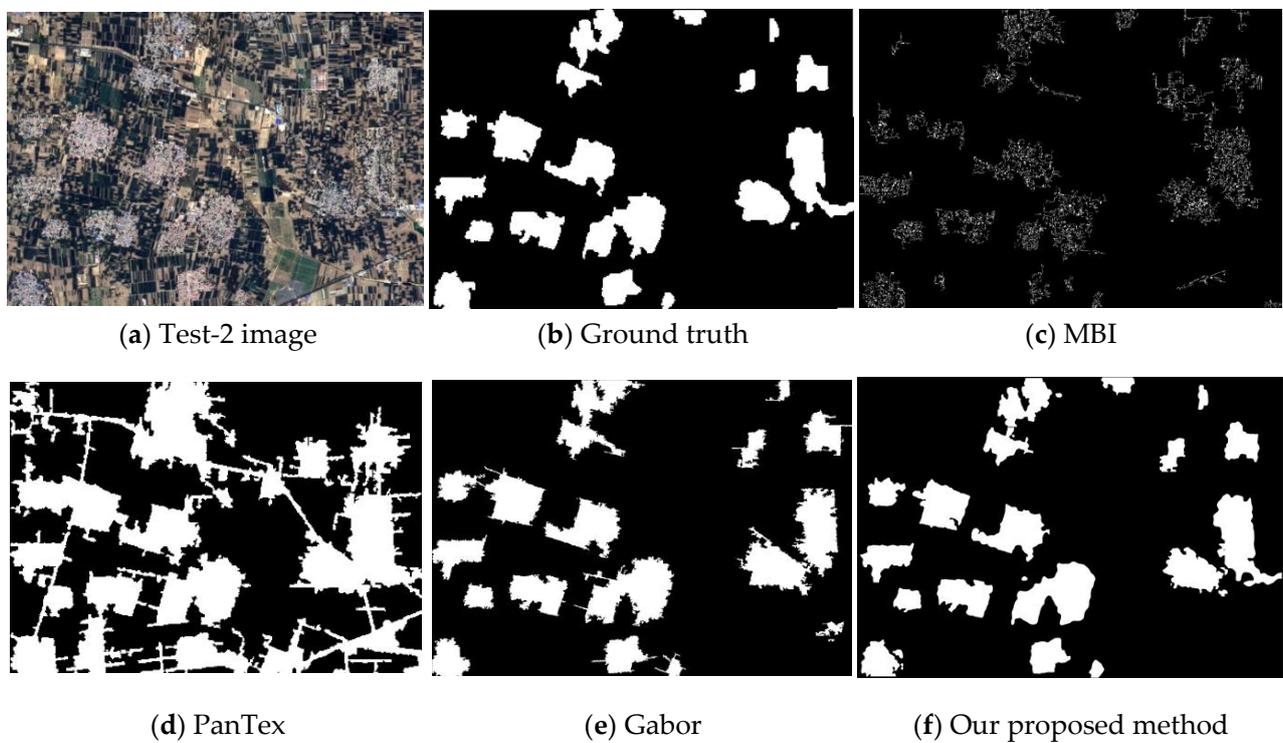


Figure 10. Comparison of results for test-2 image.

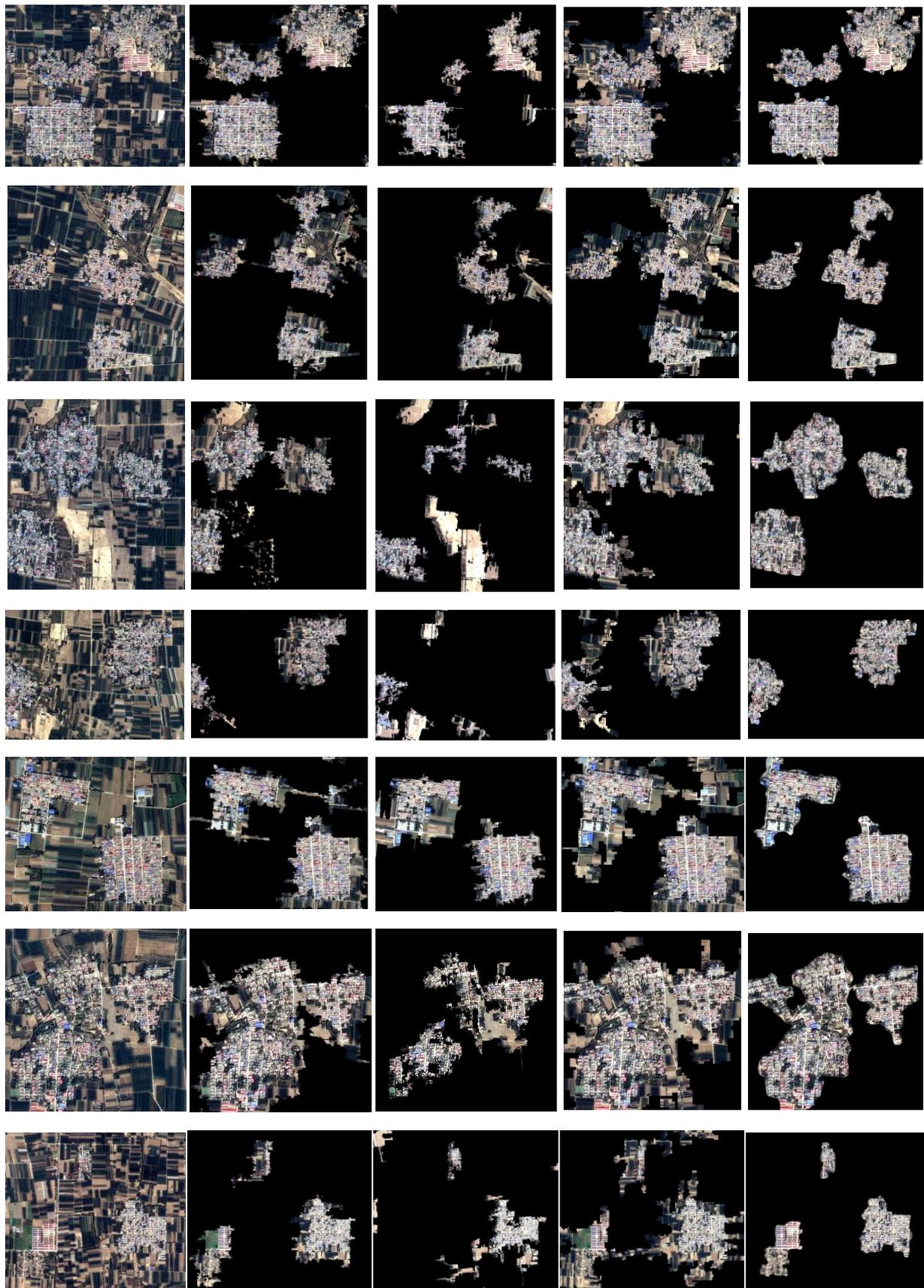


Figure 11. Cont.

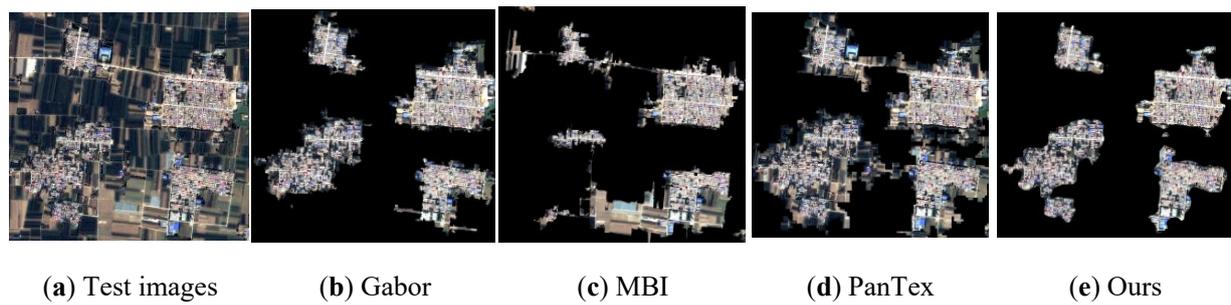


Figure 11. The details of the comparing algorithms from the two test images.

We calculate the evaluation indexes and show them in Figure 12. For the test-1 image, our proposed method shows the superiority to others in the indexes of P , R and F . In terms of the test-2 image, considering the R , the PanTex behaves best and our method is the second. As for the F value, our method has the best performance.

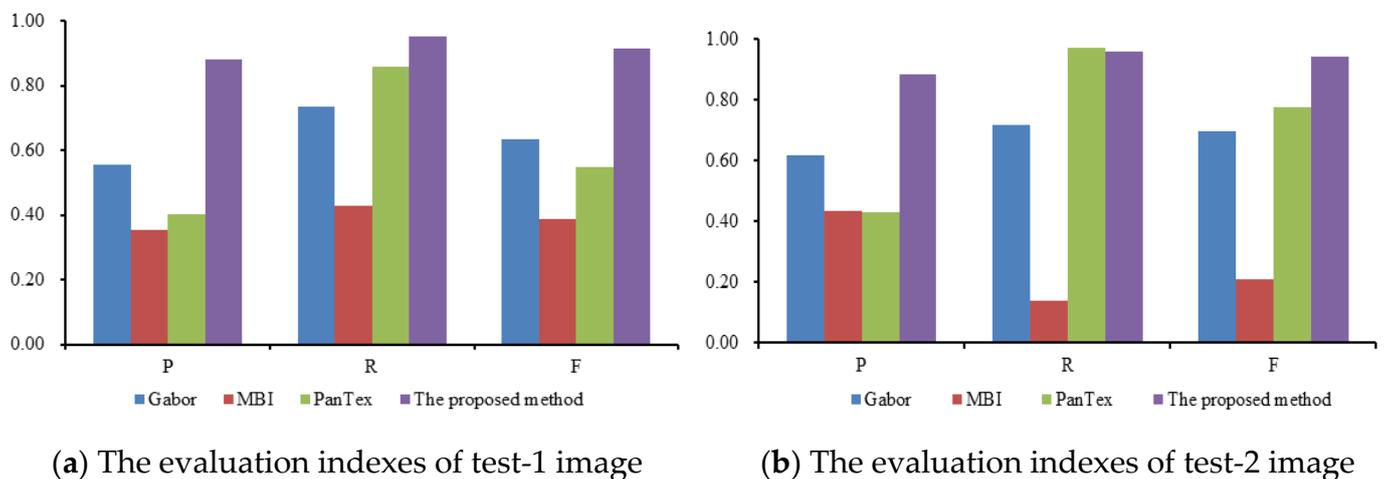


Figure 12. The evaluation indexes of the comparing algorithms.

3.4. Results on Large-Scale Satellite Images

In order to prove the performance of the proposed method for the rural built-up areas in large-scale images, we test 3 images from GF-1, ZY-3 and WV-2 sensors, respectively. As shown in Figure 13, the test image with 2 m resolution from GF-1 sensor has the size of 8240×8580 pixels, and the detected result shows that almost all built-up areas are detected successfully. According to the ground truth, the evaluation indexes of the test GF-1 image is calculated. We get the P value of 88.08%, R value of 95.08%, and the F value of 91.45%. In addition, for the ZY-3 image, the P , R , and F value are 89.58%, 91.50%, and 90.85%, respectively. In terms of WV-2 image, the P , R , and F value are 87.97%, 91.24%, and 90.12%, respectively.

To further verify the generalization ability, a Geoye-1 image with 8392×8392 pixels located in Sichuan province is prepared for the experiment. For the new type of the built-up area, which never appears in the training set, we still get the P value of 87.81%, R value of 90.44%, and the F value of 89.55%.

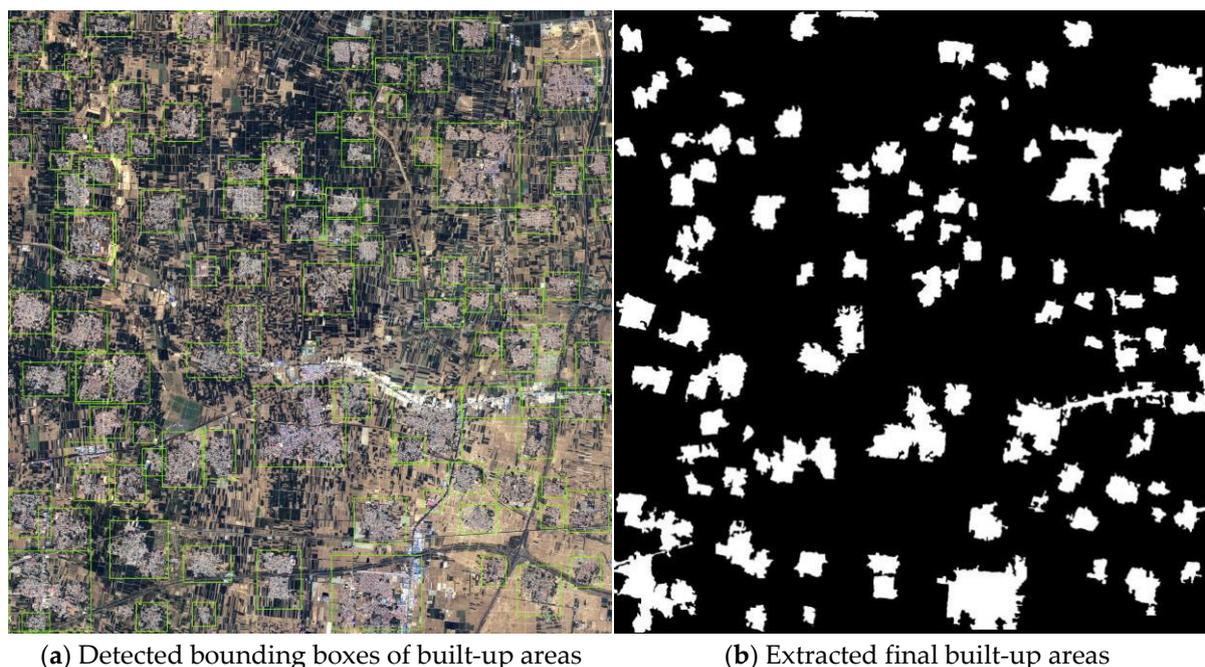


Figure 13. The detected result of built-up with a large-scale GF-1 image.

4. Discussion

In this section, we discussed the proposed method from the viewpoint of two aspects: built-up area extraction in large-scale satellite images and the impact of the sizes of built-up areas.

4.1. Built-Up Area Extraction in Large-Scale Satellite Image

Spectral residual method, one of visual attention methods, has a perfect performance in extracting built-up areas from small-scale images, but on large-scale images, the performance is not satisfactory. In this paper, the SR method is embedded into the deep neural network to extract the rural built-up areas from large-scale images. In the proposed method, coarse localization of the candidate built-up areas is firstly obtained through the Faster R-CNN framework, and then the accurate boundary of each rural built-up based on the bounding boxes is extracted by using the SR method. Therefore, as long as the detected bounding box completely contains the built-up area, it does not need to be perfectly matched with the accurate boundary. In addition, the evaluation index of the IoU has been employed to verify the rationality. The IoU of the bounding boxes of the built-up areas is 91.46%, and almost all bounding boxes of the built-up areas meet our requirements in the experiment.

In the experiment, our proposed method is compared with several built-up area extraction algorithms, such as PanTex, MBI, and Gabor, and experiments demonstrate that our proposed method has better performance. In addition, it is effective and accurate by testing the performance of our method on GF-1, ZY-3, and WV-2 images with a large-scale. We also verify a large-scale Geoeye-1 image, which never appears in the training set, and the result is still satisfactory.

4.2. The Impact of the Sizes of Built-Up Areas

In the experiment, the relationship between the sizes of built-up areas and the kernels in the SR method is mainly discussed. In the original SR method, the size of local average filter is specified as 3 due to the small size of the detected objects from natural pictures. However, when the SR method is used to extract the built-up areas from remote sensing images, we need to pay attention that the size of the built-up areas is diverse. Therefore, we explored how the size of the kernels affects the extracted results of different sizes of

built-up areas in the experiment. In Section 3.2, experimental results show that the size of the second kernel (i.e., visual scales) increases as the size of the built-up area increases, and the first kernel (i.e., local average filter) does not change significantly with the size of the built-up area. The found is instructive when using the SR method in the future.

In the proposed method, after getting the bounding boxes of the built-up area, we determine which group the length and width of each built-up area belongs to. That is, the optimal values of the two kernels are dynamically applied in the algorithm.

5. Conclusions

In this paper, the spectral residual, one of visual attention methods, is embedded into the deep neural network to rapidly and accurately extract the rural built-up areas from large-scale remote sensing images. In the proposed method, an improved Faster R-CNN framework is applied to coarse localization of the candidate rural built-up areas. Based on the bounding box of the built-up areas, the SR method is employed to extract the accurate boundary of each built-up area. In the experiment, the comparing experiments demonstrate that our proposed method is effective and accurate in extracting the rural built-up areas. In addition, three large-scale images from GF-1, ZY-3 and WV-2 sensors are tested and evaluated, and their F values are above 90%.

Another important contribution of this paper is that the relationship between the sizes of rural built-up areas and the kernels in the SR method is discussed. The result shows that the size of the visual scales increases as the size of the built-up area increases, and the local average filter does not change significantly with the size of the built-up area. This conclusion is instructive when using the SR method.

In the future, the extraction of rural buildings inside the built-up areas would be a subject of further research. In order to improve the accuracy of building extraction, the built-up areas could be further divided into different scenes. The relationship between the diversity of geometric shapes of buildings and the complexity of the scene may be further researched.

Author Contributions: Conceptualization, S.L.; software, S.F.; resources, D.Z.; writing—original draft preparation, S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant 41801240; the Natural Science Foundation of Hebei under grant D2019205067; the Science and Technology Project of Hebei Education Department under grant QN2018133; and the Doctoral Foundation of Hebei Normal University under grant L2018B19.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original satellite images in this study are available on request from the corresponding author.

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

References

1. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [[CrossRef](#)]
2. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [[CrossRef](#)]
3. Xu, H. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 4269–4276. [[CrossRef](#)]
4. Chen, X.L.; Zhao, H.M.; Li, P.X.; Yin, Z.Y. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sens. Environ.* **2006**, *104*, 133–146. [[CrossRef](#)]
5. Huang, X.; Zhang, L. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 161–172. [[CrossRef](#)]
6. Tao, C.; Tan, Y.H.; Zuo, Z.R.; Tian, J.W. Unsupervised detection of built-up areas from multiple high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1300–1304. [[CrossRef](#)]

7. Sirmacek, B.; Unsalan, C. Urban area detection using local feature points and spatial voting. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 146–150. [[CrossRef](#)]
8. Chen, H.; Tao, C.; Zou, Z.R.; Shao, L. Extraction of built-up areas extraction from high-resolution remote-sensing images using edge density features. *J. Appl. Sci.* **2014**, *32*, 537–542. [[CrossRef](#)]
9. Pang, Z.F.; Li, C.; Wang, S.G.; Zhao, B.J. Texture-based urban detection using contourlet coefficient on remote sensing imagery. In Proceedings of the IET International Radar Conference 2015, Hangzhou, China, 15–16 October 2015. [[CrossRef](#)]
10. Zhong, P.; Wang, R. A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3978–3988. [[CrossRef](#)]
11. 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](#)]
12. Goldblatt, R.; Stuhlmacher, M.F.; Tellman, B.; Clinton, N. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remote Sens. Environ.* **2018**, *205*, 253–275. [[CrossRef](#)]
13. Zhang, X.; Liu, L.Y.; Wu, C.S.; Chen, X.D.; Gao, Y.; Xie, S.; Zhang, B. Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform. *Antroposphere Land Cover. Land Use* **2020**, 1–27. [[CrossRef](#)]
14. Li, Z.Q. Research on Visual Attention Models and Application on Imagery Processing. Ph.D. Thesis, Shanghai JiaoTong University, Shanghai, China, 2009.
15. Zhang, L.B.; Yang, K.N. Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 916–920. [[CrossRef](#)]
16. Li, S.D.; Tang, H.; Yang, X. Spectral Residual Model for Rural Residential Region Extraction from GF-1 Satellite Images. *Math. Probl. Eng.* **2016**, 1–13. [[CrossRef](#)]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Int. Conf. Neural Inf. Processing Syst.* **2012**, *60*, 1097–1105. [[CrossRef](#)]
18. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.H.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
19. Szegedy, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [[CrossRef](#)]
20. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
21. Azizi, A.; Pleimling, M. A cautionary tale for machine learning generated configurations in presence of a conserved quantity. *Sci. Rep.* **2021**, *11*, 6395. [[CrossRef](#)]
22. Roshani, M.; Sattari, M.A.; Ali, P.J.M.; Roshani, G.H.; Nazemi, B.; Corniani, E.; Nazemi, E. Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Flow Meas. Instrum.* **2020**, *75*, 101804. [[CrossRef](#)]
23. Rafiee, P.; Mirjalily, G. Distributed Network Coding-Aware Routing Protocol Incorporating Fuzzy-Logic-Based Forwarders in Wireless Ad hoc Networks. *J. Netw. Syst. Manag.* **2020**, *28*, 1279–1315. [[CrossRef](#)]
24. Sanaat, A.; Zaidi, H. Depth of interaction estimation in a preclinical PET scanner equipped with monolithic crystals coupled to SiPMs using a deep neural network. *Appl. Sci.* **2020**, *10*, 4753. [[CrossRef](#)]
25. Tan, Y.H.; Xiong, S.Z.; Li, Y.S. Automatic Extraction of Built-Up Areas From Panchromatic and Multispectral Remote Sensing Images Using Double-Stream Deep Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3988–4004. [[CrossRef](#)]
26. Tan, Y.H.; Xiong, S.Z.; Yan, P. Multi-branch convolutional neural network for built-up area extraction from remote sensing image. *Neurocomputing* **2020**, *396*, 358–374. [[CrossRef](#)]
27. Zhang, T.; Tang, H. Built-Up Area Extraction from Landsat 8 Images Using Convolutional Neural Networks with Massive Automatically Selected Samples. In *Pattern Recognition and Computer Vision. PRCV 2018*; Springer: Cham, Switzerland, 2018. [[CrossRef](#)]
28. Zhang, T.; Tang, H. Evaluating the generalization ability of convolutional neural networks for built-up area extraction in different cities of China. *Optoelectron. Lett.* **2020**, *16*, 52–58. [[CrossRef](#)]
29. Iqbal, J.; Ali, M. Weakly-supervised domain adaptation for built-up region segmentation in aerial and satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 263–275. [[CrossRef](#)]
30. Ma, X.L.; Li, C.M.; Tong, X.H.; Liu, S.C. A New Fusion Approach for Extracting Urban Built-up Areas from Multisource Remotely Sensed Data. *Remote Sens.* **2019**, *11*, 2516. [[CrossRef](#)]
31. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.M.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [[CrossRef](#)]
32. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *42*, 386–397. [[CrossRef](#)]
33. Hou, X.D.; Zhang, L.Q. Saliency detection: A spectral residual approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8. [[CrossRef](#)]

34. Yang, N.S.; Tang, H.; Sun, H.Q.; Yang, X. DropBand: A Simple and Effective Method for Promoting the Scene Classification Accuracy of Convolutional Neural Networks for VHR Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 257–261. [[CrossRef](#)]
35. Ok, A.; Senaras, C.; Yuksel, B. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1701–1717. [[CrossRef](#)]