Article

Detecting Building Edges from High Spatial Resolution Remote Sensing Imagery Using Richer Convolution Features Network

Tingting Lu [1], Dongping Ming [1,*], Xiangguo Lin [2], Zhaoli Hong [1], Xueding Bai [1] and Ju Fang [1]

[1] School of Information Engineering, China University of Geosciences (Beijing), Beijing 10083, China; lutingtingGIS@163.com (T.L.); hongzhaoli9237@163.com (Z.H.); dztintin@163.com (X.B.); fangju_nanxia@163.com (J.F.)
[2] Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping, Beijing 100830, China; linxiangguo@casm.ac.cn
* Correspondence: mingdp@cugb.edu.cn

check for updates

Abstract: As the basic feature of building, building edges play an important role in many fields such as urbanization monitoring, city planning, surveying and mapping. Building edges detection from high spatial resolution remote sensing (HSRRS) imagery has always been a long-standing problem. Inspired by the recent success of deep-learning-based edge detection, a building edge detection model using a richer convolutional features (RCF) network is employed in this paper to detect building edges. Firstly, a dataset for building edges detection is constructed by the proposed most peripheral constraint conversion algorithm. Then, based on this dataset the RCF network is retrained. Finally, the edge probability map is obtained by RCF-building model, and this paper involves a geomorphological concept to refine edge probability map according to geometric morphological analysis of topographic surface. The experimental results suggest that RCF-building model can detect building edges accurately and completely, and that this model has an edge detection F-measure that is at least 5% higher than that of other three typical building extraction methods. In addition, the ablation experiment result proves that using the most peripheral constraint conversion algorithm can generate more superior dataset, and the involved refinement algorithm shows a higher F-measure and better visual effect contrasted with the non-maximal suppression algorithm.

Keywords: richer convolution features; building edges detection; high spatial resolution remote sensing imagery

1. Introduction

Buildings are one of the most important and most frequently updated parts of urban geographic databases [1]. As an important and fundamental feature for building description, the building edges detection plays a key role during building extraction [2,3]. Building edges detection has extensive applications in real estate registration, disaster monitoring, urban mapping and regional planning [4–6]. With the rapid development of remote sensing imaging technology, the number of high spatial resolution remote sensing (HSRRS) imagery has increased dramatically. HSRRS imagery have improved the spectral features of objects and highlighted information on the structure, texture, and other details of the objects. At the same time, they also brought severe image noise, "different objects with similar spectrum" and other problems [7]. In addition, due to the diversity of the structure of the buildings themselves and the complexity of the surroundings, the detection of building edges from HSRRS imagery is a challenge in the field of computer vision and remote sensing urban application.

In the rich history of edge detection, typically, the early edge detectors were designed by gradient and intensity. Later, researchers began to use artificial design features to detect edges. But these traditional edge detection algorithms mainly rely on handcrafted low-level features to detect edges, whose accuracy are difficult to guarantee and cannot adapt to application. However, with the rapid progress of artificial intelligence, deep learning has excellent performance in the field of natural image edge detection. N4-fields [8], DeepContour [9], DeepEdge [10], HFL [11], HED [12], and richer convolution features (RCF) network [13] were successively proposed. The accuracy of their test results on the BSDS500 [14] dataset has been continuously improved, while the accuracy of the newly proposed RCF network has even exceeded the human performance.

Lots of studies have shown that the deep-learning-based edge detection model can not only detect the edge of the image effectively, but also generate a higher accuracy than the traditional edge detection algorithm. However, it is not applicable to directly extract building edges from HSRRS imagery by using pre-trained deep learning network. The reasons come as follows:

- The dataset used in network training is natural image rather than remote sensing imagery. Remote sensing imagery has some features that natural images do not possess such as resolution information [15] and spatial autocorrelation.
- The remote sensing imagery has other superfluous objects in addition to the building. The network trained by the natural image cannot identify the edges of a certain object, so it is difficult to obtain the building edges directly through the pre-trained deep learning network.

Although it is difficult to acquire a high quality building edges dataset for deep learning, the limitation of the data can be overcome by modifying the existing datasets. Due to the special architecture of RCF and its excellent performance in the deep-learning-based edge detection, this paper presents a new method to detect building edges. Using the most peripheral constraint conversion algorithm, a high-quality HSRRS imagery building edges dataset for deep learning is built for the first time. This paper constructs a building edges detection model by fine-tuning the pre-trained RCF network with this self-build dataset, and the generated RCF-building model can exclusively detect the building edges. In the post-processing stage, this paper involves a geomorphological concept to refine the edge probability map generated by the RCF-building model and obtains accurate building edges. In particular, the advantage of the RCF network special architecture is exploited, which can make full use of all the convolution layers to improve the edge detection accuracy.

The rest of this paper is organized as follows. In Section 2, we briefly present the related work. RCF-based building edges detection model is described in Section 3. Section 4 presents the experiment and contrast results and analyzes the performance of the proposed methods. Finally, the discussion and conclusions are drawn in Sections 5 and 6, respectively.

2. Related Work

Although there are various edge detection algorithms and theories, but there is a great gap between theory and application, only considering the edge detection algorithm cannot directly extract buildings from imagery. As the edge detection algorithm does not have the function to distinguish what kind of object the edge belongs to, it is difficult to obtain the building edges directly by the edge detection. The previous building edges detection methods can be grouped into the following 3 categories:

- Edge-driven methods. This category usually extracts line segments by low-level edge detection algorithm first, then, groups the building edges from the line segments based on various rules [16–26]. Those rules, for example, can be perceptual grouping [16–19], Graph structure theory [20,21], Markov random field models [22], geometry theory [23], circle detection [24], heuristic approach [25], and dense matching [26]. Additionally, a series of models [27–30] have been set up to directly detect the building edges. This kind of method, in comparison to the classical methods, can detect building edges more accurately, and avoid the boundaries of features

in the building neighborhood such as streets and trees. Snake model [31], also called active contour model, was widely applied in fields of building edges detection [27–29]. The research in Garcin et al. [30] built a shape-model using Markov Object processes and a MCMC Algorithm, and this model used perspective of the whole building to detect the building.

- Region-driven methods. The building region feature and the edge feature are the important elements of the building description. Under certain circumstances, building edges can be converted from building region. Various classification strategies were utilized to extract building region, here are only a few classification strategies for HSRRS imagery:

  - ◆ Object-based image analysis (OBIA) extraction method has gradually been accepted as an efficient method for extracting detailed information from HSRRS imagery [7,32–39]. For example, references [7,32–37] comprehensively used object-based image segmentation and various features of objects such as spectrum, texture, shape, and spatial relation to detect buildings. Due to the scale parameter has an important influence on OBIA, Guo et al. [38] proposed a parameter mining approach to mine parameter information for building extraction. In addition, Liu et al. [39] adopted the probabilistic Hough transform to delineate building region which extract by multi-scale object oriented classification, and result showed that with the boundary constraint, most rectangular building roofs can be correctly detected, extracted, and reconfigured.

  - ◆ Extraction method based on deep learning is a long-standing problem in recent years [40–48]. References [40–45] designed an image segmentation using convolutional neural network, full convolutional network or other network, to effectively extract building region from imagery. The above research is still pixel-level-based, references [46–48] proposed superpixel-based convolution neural network (SML-CNN) model in hyperspectral image classification in which superpixels are taken as the basic analysis unit instead of pixels. Compared to other deep-learning-based methods, superpixel-based method gain promising classification results. Gao et al. [49] combined counter map with fully convolutional neural networks to offer a higher level of detection capabilities on image, which provided a new idea for building detection. In addition, constantly proposed theories, such as transfer hashing [50] and structured autoencoders [51] can also be introduced into this application field to solve problems, such as data sparsity and data mining.

  - ◆ Extraction method based on mathematical morphology [52–58]. Huang et al. and Rongming et al. [52,53] used morphological building index by differential morphological profile to extract buildings and optimized methods are proposed in references [54–58].

- Auxiliary-information-based methods. Due to the complexity of the structure and surrounding environment of the building, many scholars have proposed the method of extracting the building by the shadow, stereoscopic aerial image or digital elevation model (DEM) data to assist the building extraction. Liow et al. [59] pioneering proposed a new idea of using shadow to extract buildings. Later, research in [59–62] proposed to identify and extract buildings based on the shadow features and graph based segmentation in high-resolution remote sensing imagery. In addition, local contrast in the image where shadow and building interdepend will be increase. Based on this principle, references [63,64] proposed PanTex method with gray level co-occurrence matrix contrast features, which is practically used to identify buildings and build-up areas. Hu et al. [65] used the shadow, shape, color features, similarity of angle between shade lines and so on multiple cues to extract buildings. In addition, stereo information can provide great convenience for the extraction of buildings information [5,66–78].

Among the methods mentioned above, the first category normally used semantic analysis to grouping lines segments, and they have shown relatively good performance on moderate and low spatial resolution remote sensing imagery because of its high signal noise ratio (SNR). However, for

HSRRS imagery, the high spatial resolution and low SNR substantially increases the difficulties of locating and identifying the accurate building edges [39]. For the second category, they have many advantages, such as a comprehensive consideration of prior knowledge, image features, pattern recognition theory and other factors. However, the related methods still have the problems of cumbersome workflow, which requires more prior knowledge, and unable to meet the practical requirements of buildings extraction from high spatial resolution images with high scene complexity. The applicability is also limited by buildings type, density, and size. Moreover, the edge of extraction results is not ideal, so it is difficult to ensure the edge integrity of complex objects. For the last category, although the accuracy of building extraction can be improved based on stereo information, it is greatly limited by multiple data sources scarcity and data misalignment.

Therefore, to overcome these limitations of single data, building structure, surrounding complexity and prior knowledge, this paper tries to detect building edges using state-of-the-art method of edge detection with deep learning, which is only based on two-dimensional HSRRS imagery, also needs no prior knowledge once the deep supervision based dataset is perfectly built.

3. Methodology

As shown in Figure 1, the workflow of proposed method is mainly divided into three stages. In the dataset construction stage, the initial dataset is processed by conversion, clipping, rotation, and selection into a special dataset which can be dedicated to deep-learning-based edge detection. The second stage is network training. Based on the training set, the RCF network is retrained to generate a RCF-building edges detection model. The third stage is detecting and post-processing. The edge probability map is obtained by using RCF-building model. Subsequently, the edge probability map is refined by the involved algorithm, so that the building edges are obtained.
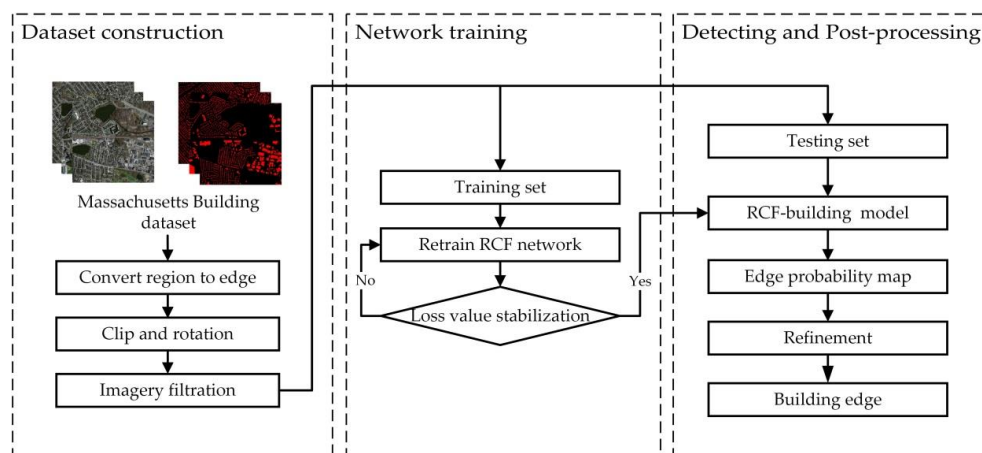

Figure 1. Workflow of fine-tuning RCF network.

3.1. Dataset Construction

As mentioned previously, in the field of deep learning, there is no experimental HSRRS imagery dataset available to building edges detection. Therefore, this paper builds an edge based sample dataset that satisfies the training and testing requirements of the RCF network by pre-processing Massachusetts Building Dataset [79]. The Massachusetts Building dataset is constructed by Mnih and publicly available at http://www.cs.toronto.edu/vmnih/data/. This dataset has a resolution of 1 m and sizes of 1500×1500 pixels. It contains 137 training images, 10 testing images, and four validation images between which has no intersection. Each set of data includes an original remote sensing image and a manually traced building region map, as shown in Figure 2a,b. Since the output of RCF network is based on the fusion of multi layers, RCF network is tolerable to slight overfitting. Thus, RCF network does not need validation sets.
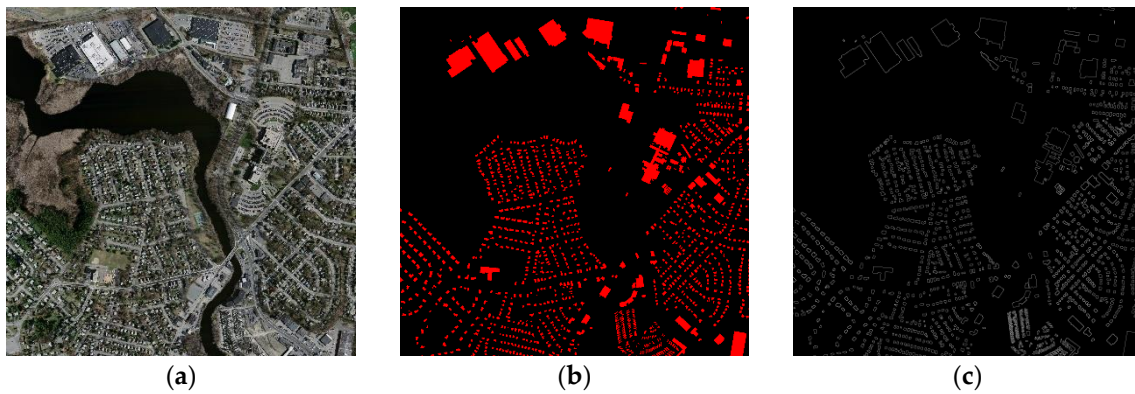
(**a**)　　　　　　　　　　　　　　　(**b**)　　　　　　　　　　　　　　　(**c**)

Figure 2. Dataset sample. (**a**) Original image; (**b**) building region map; and (**c**) building edges ground truth map.

Edge detection is different from region extraction, and the location shift of only one pixel may cause the model fail to extract features and reduce the overall precision. To ensure that there is no error occurred when convert building region to building edges, this paper proposes most peripheral constraint algorithm. With constraint of "most peripheral", it emphasizes on only extracting the outermost pixels of the building region features as building edges, and the width of edge is only one pixel. Figure 3 shows the diagram of this conversion algorithm. The steps come as follows:

(1)　Binarization of the building region map. Supposing the building pixel value is 1, and the non-building pixel is 0;

(2)　Generating an image with the same size as the original image, and all the pixel values are 0. Scanning the building region map row by row to find all pixels (marked as P_r) satisfying two conditions: the pixel value is 1, and the pixel value shifts from 1 to 0 or from 0 to 1. In the newly generated image, setting the pixel values at the same locations with P_r as 1. Thus, building edge pixels on each row are detected;

(3)　Generating an image with the same size as the original image, and all the pixel values are 0. Repeating step 2 to detect all building edge pixels on each column;

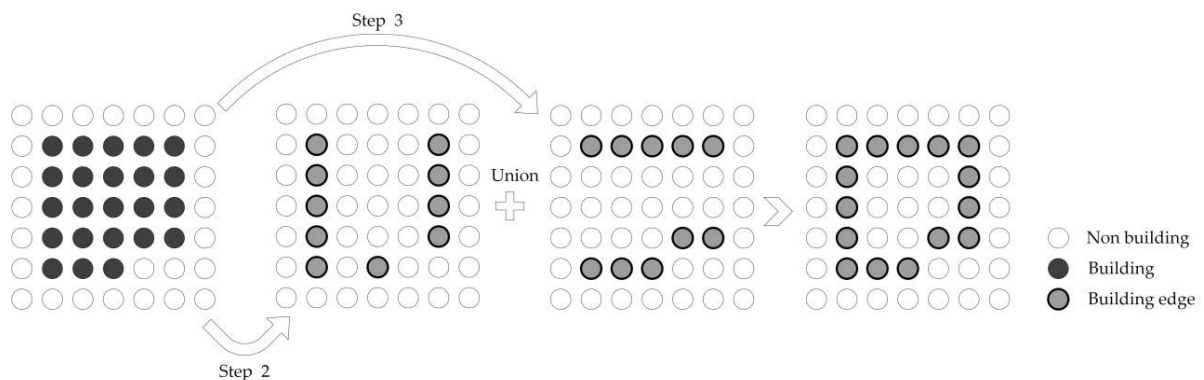(4)　All building edge pixels on each row and each column are combined. Thus, the building edge is finally detected.

Figure 3. Diagram of conversion from building region into building edges.

Figure 2c shows the conversion result of Figure 2b. After conversion, in order to improve the accuracy of the training network, we augment the data by rotating the imagery by 90, 180, and 270 degrees. Meanwhile, to avoid memory overflow and invalid imagery, this paper ultimately constructs the dataset after image clipping and choosing. The final dataset contains 1856 training images with size of 750 × 750 pixels and 56 testing images with size of 750 × 750 pixels, named Massachusetts Building-edge dataset.

3.2. RCF Network

The RCF network was originally proposed by Liu in 2017 [13]. It was optimized on the basis of VGG16 [80] network. The input of the RCF network is an RGB image with unlimited size, and the output is the edge probability map with the same size. Figure 4 shows the architecture of RCF network when the input image size is 224 × 224 pixels. The main convolutional layers in RCF (as shown in the red dashed rectangle) are divided into five stages and the adjacent two stages are connected through the pooling layer. After the down sampling of the pooling layer, different scales of features can be extracted, and useful information can be obtained while reducing the amount of data. Different from VGG16 network, the RCF network discards all the fully connected layers as well as the fifth pooling layer, and each main convolution layer is connected to a convolution layer with kernel size 1 × 1 and channel depth 21. Then, RCF network sets an element_wise layer for accumulation after each stage. Afterwards, each element_wise layer is connected to a convolution layer with kernel size 1 × 1 and channel depth 1. The difference between the RCF network and the traditional neural network lies in: for the boundary extraction, the previous neural networks only use the last layer as the output, and lose many feature details, while the RCF network fuses the convoluted element_wise layers of each stage (convoluted element_wise layers of 2, 3, 4, and 5 stages need to be restored to its original image size by deconvolution) with the same weights to get a fusion output. This special network architecture allows the RCF network to make full use of semantic information and detailed information for edge detection.
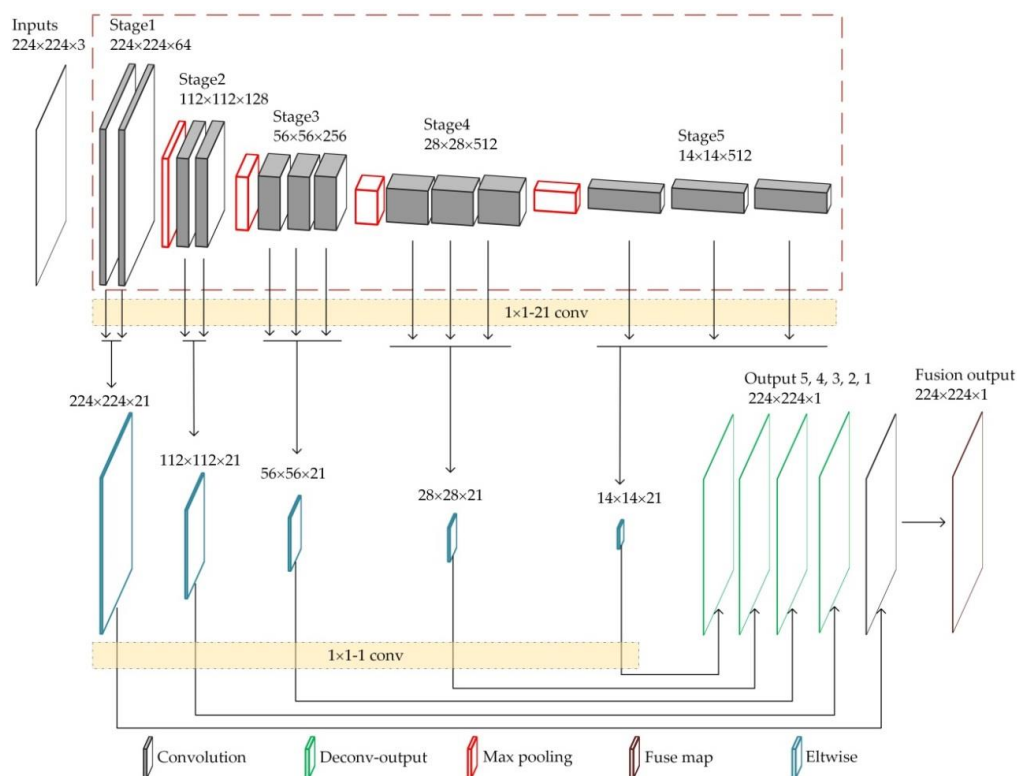

Figure 4. Overview of the RCF network architecture.

3.3. Refinement of Edge Probability Map

The test results of the RCF network are gray-scale edge probability map, on which the greater the gray value is, the higher the probability that the pixel is on an edge. To accurately detect the building edges, it is necessary to refine edge probability map. In computer vision filed, non-maximal suppression (NMS) algorithm is a commonly used refining method. However, as observed in Figure 5,

the results show that using NMS algorithm to refine building edges has the following problems: broken outliers, isolated points and flocculent noises.

Figure 5. The refinement results by NMS algorithm.

Therefore, this paper involves a geomorphological concept to refine edge probability map according to geometric morphological analysis of topographic surface. As illustrated in Figure 6, our basic idea is to regard the edge probability value as elevation, according to the principles of geometric morphology, and the points with maxima elevation (i.e., the watershed point) on the topographic profile curve are extracted as accurate edges.
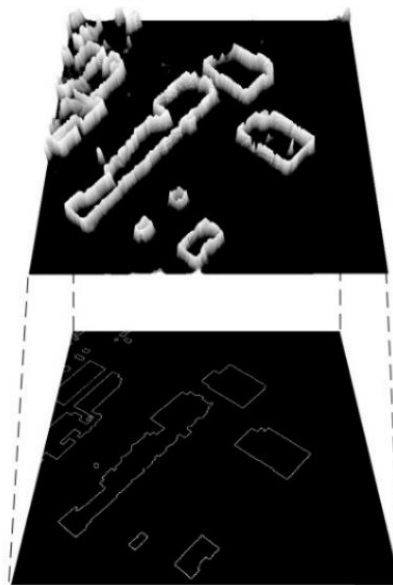

Figure 6. Diagram of edge refinement algorithm.

As described in Figure 7, the procedures of this refinement algorithm are as follows:

(1) Scanning from four directions (vertical, horizontal, left diagonal, and right diagonal) to find the local maxima points as candidate points;

(2) Setting a threshold to discard the candidate points whose probability is less than 0.5 (After many experiments, the highest accuracy is obtained under this threshold. For gray image, the threshold value is 120.);

(3) Calculating the times that each candidate point is detected out. When a candidate point is detected at least twice, it is classified as an edge point;

(4) Checking the edge points got by step (3) one by one. When there is no other edge point in an eight neighborhood, this point is determined as an isolated point and deleted;

(5) Generating edge mask map based on the edge point map got by step (4) to refine the edge probability map and obtain the final edge refinement map.
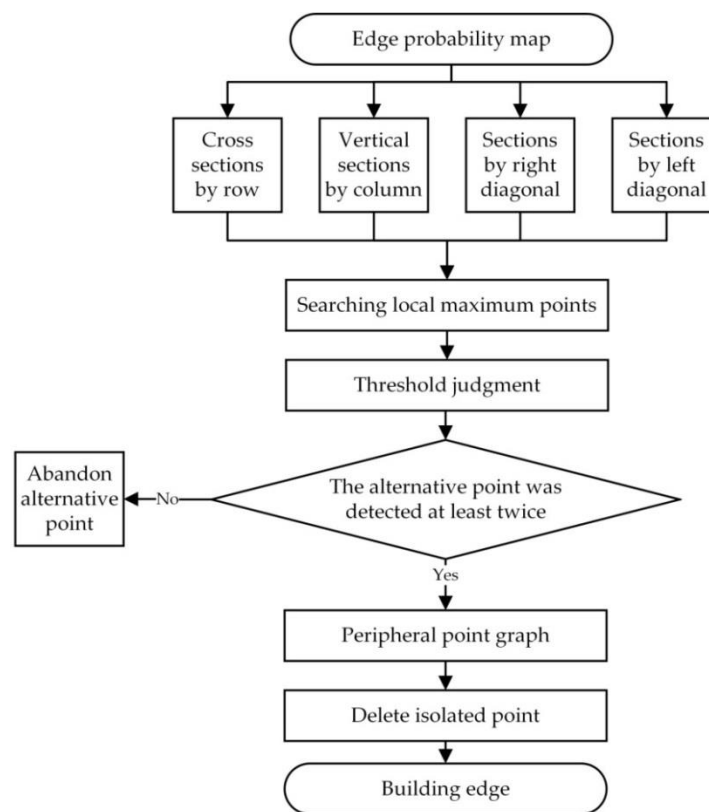
Figure 7. Workflow of edge probability map refinement.

4. Experiments and Analysis

The experimental environment for RCF network re-training and testing is the Caffe framework [81] in Linux system with support of NVIDIA GTX1080 GPU. The learning rate refers to the rate of descent to the local minimum of the cost function, and the initial learning rate is 1×10^{-7}. Every ten thousand iterations, the learning rate will be divided by 10 in training process. The experimental data are the self-processed Massachusetts Building-edge dataset which has been introduced in Section 3.1.

4.1. Experimental Results

In this paper, a trained model generated by 40,000 iterations is selected to extract the building edges. Some example of the building edges detection results are shown in Figure 8(e1–e3). From the visual perspective, the RCF-based building edges detection method adapts to the background very well. As can be seen from the third line of data (Figure 8(a3,b3,c3,d3,e3), which are highlighted by red rectangle, the fine-tuned RCF-building model can not only detect building edges correctly, but also extract building edges that the human unrecognized. Additionally, the refinement results of involved refinement algorithm (Figure 8(e1–e3) are experimentally compared with the results of NMS algorithm (Figure 8(d1–d3). There are less isolated points and flocculent noises in the building edges detection results by the involved refinement algorithm.
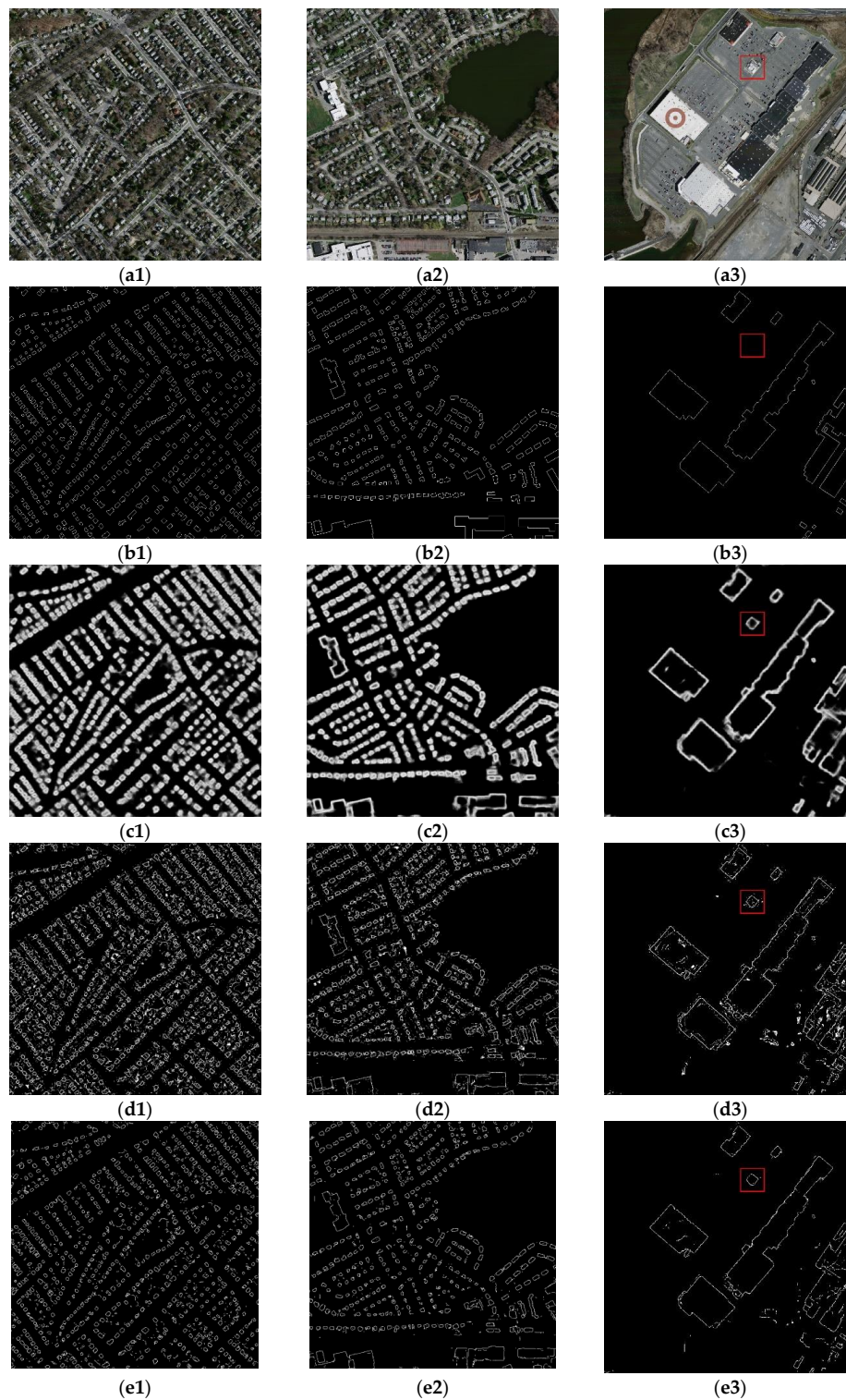
Figure 8. RCF based building edges detection results. (**a1**–**a3**) Original imagery; (**b1**–**b3**) building edges ground truth map; (**c1**–**c3**) building edges probability map generated by RCF-building; (**d1**–**d3**) building edges refinement map generated by NMS algorithm; and (**e1**–**e3**) building edges refinement map generated by involved algorithm.

4.2. Precision and Recall Evaluation

In this paper, inspired by references [82–84], we used recall, precision, and F-measure as the criteria for RCF-building model. The evaluation indices can be described by Equations (1)–(3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precison} + \text{Recall}} \tag{3}$$

where true positive (TP) represents the number of coincident pixel between detected edges and referenced building edges of ground truth. False positive (FP) represents the number of non-coincident pixel between detected edges and referenced building edges of ground truth. False negative (FN) represents the number of non-coincident pixel between detected non-building objects and non-building edges in the referenced ground truth. F-measure is a synthetic measurement of precision and recall. Actually, the precision and the recall are two contradictory measurements. Generally, they are negatively correlated [85,86]. Based on recall and precision, the precision-recall curve (P-R curve) can be drawn.

As shown in Figure 9, it can be noted that our RCF-building model has an F-measure of 0.89 on the test set, which is higher than the 0.51 from the original RCF network. In addition, compared with the original RCF network, the precision of RCF-building model increases at least 45%. It means that the retraining RCF network has the function of recognizing the edges of buildings. The generated RCF-building model can exclusively detect the building edges, and effectively avoid the superfluous objects edges.
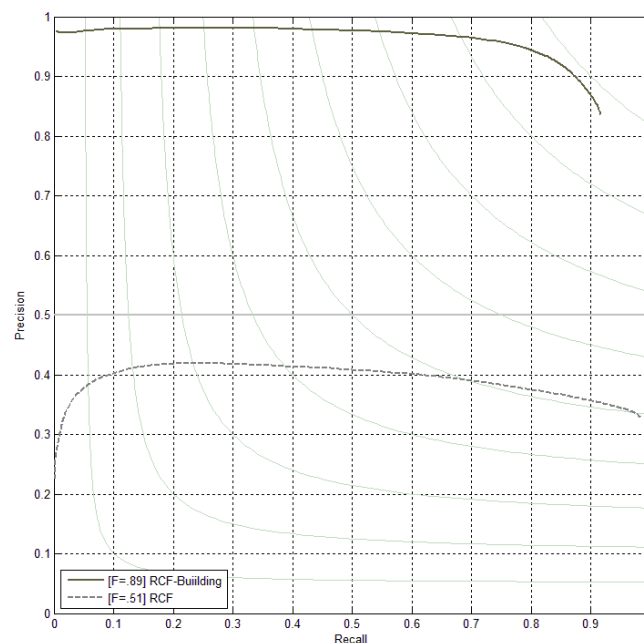

Figure 9. The P-R curves. The solid curve is the result of proposed RCF-building on the test set and the dotted one is the original RCF network.

4.3. Comparison with Other Building Extraction Methods

In this paper, four remote sensing images with different characteristics from the testing set are selected to compare the performance of our method with other three representative building detection methods. Figure 10 illustrates the visual results of our method, OBIA-based ENVI

Feature Extraction [87], Superpixel-based SML-CNN [47] and CNN-based Saito's Method [43]. The ENVI Feature Extraction was implemented through ENVI's Example-Based Classification [88]. The segmentation scale and merge level parameters were set respectively as 40, 30, 30, 40 and 50, 40, 50, 80. Classification was accomplished by training the nearest neighbor classifier with the selected samples of building and non-building objects point. The scale parameters of the SML-CNN are set to 15. In addition, image1 has similar characteristics to the image3, so the model generated by image1 was selected to classify image3. The sample sizes of the data are shown in Table 1. The results of the Saito's Method are derived from the experimental results in the reference [43] which uses the same dataset as this paper. To ensure fair comparison, this paper cuts the related image data into the same size with those used in this paper.

Table 1. Number of sample points marked on the Figure 10 original images.

Sample Category	Image1	Image2	Image3	Image4
Building	1200	194	1200	1954
Non building	2088	1298	2088	1774

It can be clearly seen from Figure 10 that the method used in this paper has better visual effects than ENVI Feature Extraction and SML-CNN. Compared with the overall view results of Saito's Method, although our results have more broken line segments inside the building, as can be seen from the last row in Figure 10, the detailed image shows that the method we used can maximize the integrity of the building edges. In the corner part of the building, the angular characteristics are preserved better by our method.

Table 2 shows the evaluation results of building edges detected by ENVI Feature Extraction [87], SML-CNN [47], Saito's Method [43] and the proposed method in four images. It can be seen from the comparison of the F-measure values that the RCF network has the best performance regardless of whether the building group is high-density (image1) or low-density (image2), or the structure of the building is simple (image3) or complicated (image4). ENVI Feature Extraction is a traditional module for extracting information from high-resolution panchromatic imagery by spatial, spectral, and texture characteristics. Although we manage to cover all types of buildings in the selection of samples, the classification results of buildings still have serious noises and misclassifications, and the building edges extracted by this method would be mixed with more non-architectural edges and closed noise lines inside the building. Compared to traditional building edge detection methods based on image processing, RCF-building is more robust and it is applicable in complicated environment because this model depends on not only image but also supervised dataset. Manually labeled building samples implement deep supervision of each layer of network to achieve optimal fitting of building edge information at different scales, and enhance the saliency-guided building feature learning. Thus the method of ENVI Feature Extraction has similar Recall as the proposed method but much lower Precision. SML-CNN first divides the image into superpixels, and then uses CNN network in classification. Therefore, SML-CNN can extract building edges completely, but at the same time, it might have misclassification. This method has a slightly higher recall and much lower precision than the method we proposed. Saito's Method is a CNN network which simultaneously extracts multiple kinds of objects. Due to the limitation of network architecture, only region features are emphasized while line features are ignored. Although this method can roughly locate buildings in the imagery, the boundaries between the buildings and the non-buildings are not accurate, and present lower Recall value and higher precision value. The method proposed in this paper has a good performance on both precision and recall. Compared to deep-learning-based building extraction methods, RCF-building could better retain building edges angular characteristics.
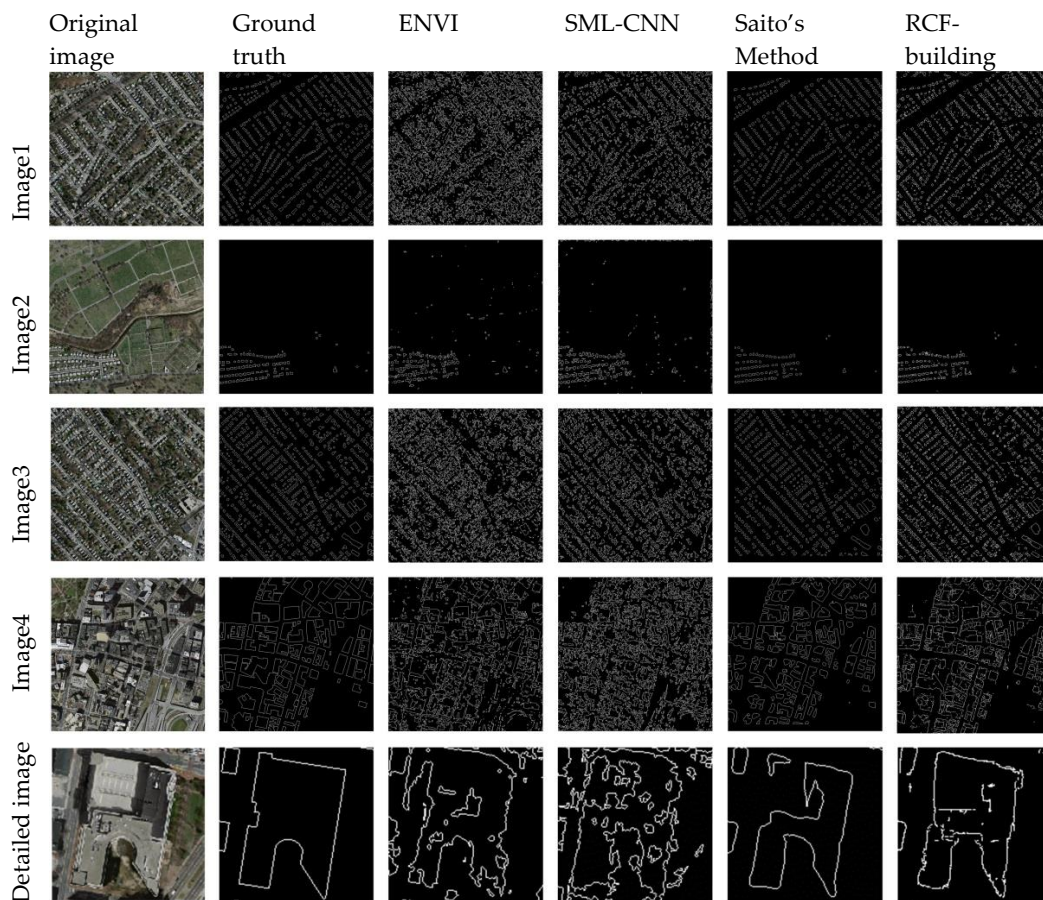
Figure 10. Building edges detection results on Massachusetts Building Dataset. The last row shows the details image of one building.

Table 2. Evaluation results of four different methods on four typical images.

Approach	Index	Image1	Image2	Image3	Image4	Mean
ENVI Feature Extraction	Precision	0.35	0.71	0.44	0.45	0.49
	Recall	0.97	0.90	0.96	0.87	0.93
	F-measure	**0.52**	**0.80**	**0.61**	**0.60**	**0.63**
SLIC-CNN	Precision	0.51	0.54	0.57	0.35	0.49
	Recall	0.99	0.97	0.97	0.96	0.97
	F-measure	**0.68**	**0.70**	**0.72**	**0.52**	**0.65**
Saito's Method	Precision	0.99	1.00	0.99	0.78	0.94
	Recall	0.55	0.72	0.50	0.75	0.63
	F-measure	**0.70**	**0.84**	**0.67**	**0.77**	**0.74**
RCF-building	Precision	0.85	0.96	0.88	0.74	0.86
	Recall	0.94	0.82	0.93	0.94	0.91
	F-measure	**0.89**	**0.89**	**0.91**	**0.82**	**0.88**

5. Discussion

5.1. Ablation Experiment

　　To verify the effectiveness of different steps of the proposed method, this paper compares the performance of RCF model trained by the self-processing dataset (Massachusetts Building-edge dataset) with the RCF model trained by Canny algorithm [89] converted dataset on all testing set. We also quantitatively compare the performance of the involved edge refining algorithm with NMS edge refining algorithm. Table 3 lists the evaluation results of different pre-processing and post-processing methods. Our methods present the best performance in the Precision, Recall and F-measure. The experimental results verify the effectiveness of proposed conversion algorithm for dataset pre-processing, which proves that the superior dataset has positive influence on RCF network. Furthermore, comparison results also reveal that the good performance of our approach takes the advantage of the involved refinement algorithm. For all testing set, the refining algorithm presented in this paper has better performance.

Table 3. The performance of training set generated by different conversion methods and performance comparison of different refinement algorithms.

Conversion Algorithm	Refinement Algorithm	Precision	Recall	F-Measure
Canny algorithm	NMS	0.46	0.99	**0.63**
Canny algorithm	Our refinement algorithm	0.70	0.94	**0.80**
Our conversion algorithm	NMS	0.60	0.98	**0.75**
Our conversion algorithm	Our refinement algorithm	0.85	0.89	**0.87**

5.2. Influence of the RCF Fusion Output

　　To explore why RCF-building can recognize the edge of building, this paper compares the average Precision, Recall and F-measure values of all testing set imagery at each stage of network. As shown in Figure 11, with the deepening of the network, the precision and recall value rises gradually during the first three stages, and then the precision and recall value descend (or roughly descend) during the fourth and fifth stage. During the first three stages, the network gradually learns the characteristics of the building edge, so the precision and recall of the detected building edges increase gradually. However, during the fourth and fifth stages, the network is overfitting and regards the characteristics of one training sample as the general nature of all the potential samples. This phenomenon of reduced generalization performance eventually leads to the failure of detecting some parts of the building edges. On the other hand, the overfitting of edge detection is different from the overfitting in other fields, which means after overfitting, if one pixel is judged as edge, the probability of actually being edges is higher. Above all, to make full use of the information generated at each stage, the RCF network utilizes a special architecture that the traditional neural networks do not have: the fusion output layer. The fusion output layer fuses all the output of each stage with the same weight, so that it can perfectly inherit the advantages of each stage and suppress the useless information at first two stages. Thus, the fusion output guarantees the highest precision and recall value.
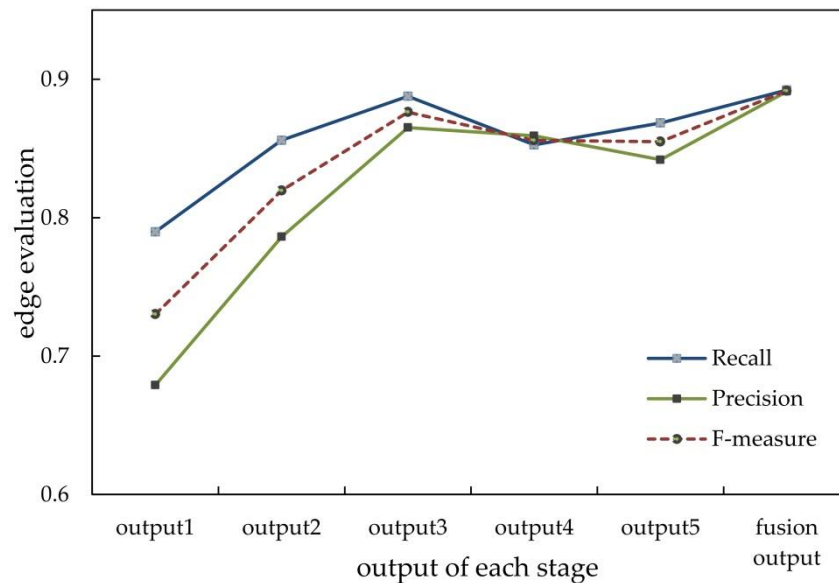
Figure 11. Comparison of precision, recall and F-measure of the output maps at different stages.

Take a test image as an example, the output of the each stage and fusion output images are shown in Figure 12. It is clear that with the deepening of the network stages, the model can gradually extract the edge of the building and eliminate the edges of other superfluous objects, but in the fourth or fifth stage, the edge of the building cannot be completely extracted in the image. The visual result of the fusion output image has the best performance, and the edge of the building can be extracted completely and accurately compared with other stages output. Therefore, RCF's special fusion output architecture makes it suitable for building edges extraction from high resolution remote sensing images.
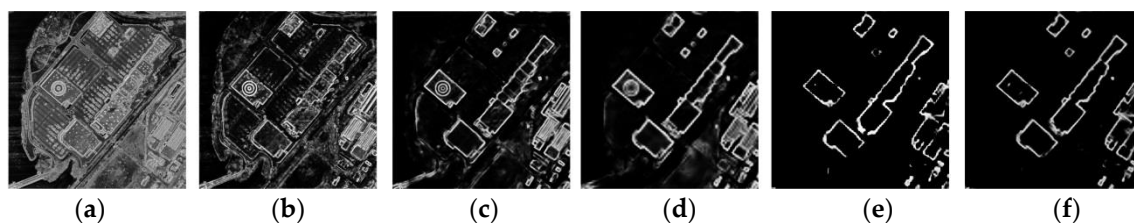

(**a**)	(**b**)	(**c**)	(**d**)	(**e**)	(**f**)

Figure 12. Output images of each stage and fusion output image. From (**a**–**f**): stage1 output, stage2 output, stage3 output, stage4 output, stage5 output, and fusion output.

6. Conclusions

This paper proposes a method for detecting building edges from HSRRS imagery based on the RCF network. The highlights of this work are listed as follows:

- The RCF network is firstly combined with HSRRS imagery to detect building edges and then an RCF-building model that can accurately and comprehensively detect the building edges is built. Compared to the traditional building edge extraction method, the method used in this paper can make use of high-level semantic information and can get a higher accuracy evaluation value and better visual effects. Compared to deep-learning-based building extraction methods, RCF-building could better retain the corner part building edges. In addition, this paper also analyzes the influence of the RCF fusion output architecture on the building edges detection accuracy, and the precision and recall lines affirm that this unique architecture of RCF can perfectly inherit the advantages of each stage and has a strong applicability to the detection of building edges.

- In the preprocessing stage, on the basis of Massachusetts Building dataset, we proposed the most peripheral constraint edge conversion algorithm and created the Massachusetts Building-edge dataset specifically for deep-learning-based building edges detection. The comparison result shows that the dataset produced by the most peripheral constraint algorithm can effectively improve the performance of RCF-building model, and affirms the positive impact of accurately labeled data on network training. This Massachusetts Building-edge dataset makes the foundation for future research on deep-learning-based building edges detection.

- In the post-processing stage, this paper involves a geomorphological concept to refine edge probability map according to geometric morphological analysis of topographic surface. Compared to the NMS algorithm, the involved refinement algorithm could balance the precision and recall value, and get a higher F-measure. It can preserve the integrity of the building edges to the greatest extent and reduce noise points. However, there are still some broken lines, as well as some discontinuities in the detected building edges results after the post-processing.

Additionally, it is worth noting that building edges detection is not the terminal goal of building extraction from HSRRS imagery. The future work will include: (1) connection of the broken edges of the building; (2) vectorization of building edges features; (3) the improvement of RCF network architecture; and (4) using various strategies to ensure that large images can be processed in memory [90].

References

1. Du, S.; Luo, L.; Cao, K.; Shu, M. Extracting building patterns with multilevel graph partition and building grouping. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 81–96. [CrossRef]

2. Li, Y.; Wu, H. Adaptive building edge detection by combining lidar data and aerial images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 197–202.

3. Hu, X.; Shen, J.; Shan, J.; Pan, L. Local edge distributions for detection of salient structure textures and objects. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 4664–4670. [CrossRef]

4. Yang, H.-C.; Deng, K.-Z.; Zhang, S. Semi-automated extraction from aerial image using improved hough transformation. *Sci. Surv. Mapp.* **2006**, *6*, 32.

5. Siddiqui, F.U.; Teng, S.W.; Awrangjeb, M.; Lu, G. A robust gradient based method for building extraction from lidar and photogrammetric imagery. *Sensors* **2016**, *16*, 1110. [CrossRef] [PubMed]

6. Wu, G.; Guo, Z.; Shi, X.; Chen, Q.; Xu, Y.; Shibasaki, R.; Shao, X. A boundary regulated network for accurate roof segmentation and outline extraction. *Remote Sens.* **2018**, *10*, 1195. [CrossRef]

7. Ming, D.-P.; Luo, J.-C.; Shen, Z.-F.; Wang, M.; Sheng, H. Research on information extraction and target recognition from high resolution remote sensing image. *Sci. Surv. Mapp.* **2005**, *30*, 18–20.

8. Ganin, Y.; Lempitsky, V. N 4-fields: Neural network nearest neighbor fields for image transforms. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; Springer: Berlin, Germany, 2014; pp. 536–551.

9. Shen, W.; Wang, X.; Wang, Y.; Bai, X.; Zhang, Z. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3982–3991.

10. Bertasius, G.; Shi, J.; Torresani, L. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4380–4389.

11. Bertasius, G.; Shi, J.; Torresani, L. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 504–512.

12. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.

13. Liu, Y.; Cheng, M.-M.; Hu, X.; Wang, K.; Bai, X. Richer convolutional features for edge detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5872–5881.

14. Martin, D.R.; Fowlkes, C.C.; Malik, J. Learning to detect natural image boundaries using brightness and texture. Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–13 December 2003; pp. 1279–1286.

15. Chen, Z.; Zhang, T.; Ouyang, C. End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sens.* **2018**, *10*, 139. [CrossRef]

16. Lin; Huertas; Nevatia. Detection of buildings using perceptual grouping and shadows. In Proceedings of the IEEE Computer Vision & Pattern Recognition, Seattle, WA, USA, 21–23 June 1994.

17. Jaynes, C.O.; Stolle, F.; Collins, R.T. Task driven perceptual organization for extraction of rooftop polygons. In Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 152–159.

18. Mohan, R.; Nevatia, R. Using perceptual organization to extract 3d structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 1121–1139. [CrossRef]

19. Turker, M.; Koc-San, D. Building extraction from high-resolution optical spaceborne images using the integration of support vector machine (SVM) classification, hough transformation and perceptual grouping. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 586–589. [CrossRef]

20. Kim, T.; Muller, J.P. Development of a graph-based approach for building detection. *Image Vis. Comput.* **1999**, *17*, 31–34. [CrossRef]

21. Tao, W.B.; Tian, J.W.; Liu, J. A new approach to extract rectangle building from aerial urban images. In Proceedings of the 2002 6th International Conference on Signal Processing, Beijing, China, 26–30 August 2002; Volume 141, pp. 143–146.

22. Krishnamachari, S.; Chellappa, R. Delineating buildings by grouping lines with mrfs. *IEEE Trans. Image Process.* **2002**, *5*, 1641–1668. [CrossRef] [PubMed]

23. Croitoru, A.; Doytsher, Y. Right-angle rooftop polygon extraction in regularised urban areas: Cutting the corners. *Photogramm. Rec.* **2010**, *19*, 3113–3141. [CrossRef]

24. Cui, S.; Yan, Q.; Reinartz, P. Complex building description and extraction based on hough transformation and cycle detection. *Remote Sens. Lett.* **2012**, *3*, 1511–1559. [CrossRef]

25. Partovi, T.; Bahmanyar, R.; Krauß, T.; Reinartz, P. Building outline extraction using a heuristic approach based on generalization of line segments. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 9339–9347. [CrossRef]

26. Su, N.; Yan, Y.; Qiu, M.; Zhao, C.; Wang, L. Object-based dense matching method for maintaining structure characteristics of linear buildings. *Sensors* **2018**, *18*, 1035. [CrossRef] [PubMed]

27. Rüther, H.; Martine, H.M.; Mtalo, E.G. Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas. *ISPRS J. Photogramm. Remote Sens.* **2002**, *56*, 269–282. [CrossRef]

28. Peng, J.; Zhang, D.; Liu, Y. An improved snake model for building detection from urban aerial images. *Pattern Recognit. Lett.* **2005**, *26*, 5875–5895. [CrossRef]

29. Ahmadi, S.; Zoej, M.J.V.; Ebadi, H.; Moghaddam, H.A.; Mohammadzadeh, A. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 1501–1557. [CrossRef]

30. Garcin, L.; Descombes, X.; Men, H.L.; Zerubia, J. Building detection by markov object processes. In Proceedings of the International Conference on Image Processing, Thessaloniki, Greece, 7–10 October 2001; Volume 562, pp. 565–568.

31. Kass, A. Snake: Active contour models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331. [CrossRef]

32. Zhou, J.Q. Spatial relation-aided method for object-oriented extraction of buildings from high resolution image. *J. Appl. Sci.* **2012**, *30*, 511–516.

33. Tan, Q. Urban building extraction from vhr multi-spectral images using object-based classification. *Acta Geod. Cartogr. Sin.* **2010**, *39*, 618–623.
34. Wu, H.; Cheng, Z.; Shi, W.; Miao, Z.; Xu, C. An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. *Nat. Hazards* **2014**, *71*, 151–174. [CrossRef]
35. Benarchid, O.; Raissouni, N.; Adib, S.E.; Abbous, A.; Azyat, A.; Achhab, N.B.; Lahraoua, M.; Chahboun, A. Building extraction using object-based classification and shadow information in very high resolution multispectral images, a case study: Tetuan, Morocco. *Can. J. Image Process. Comput. Vis.* **2013**, *4*, 1–8.
36. Mariana, B.; Lucian, D. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 67–75.
37. Tao, C.; Tan, Y.; Cai, H.; Du, B.; Tian, J. Object-oriented method of hierarchical urban building extraction from high-resolution remote-sensing imagery. *Acta Geod. Cartogr. Sin.* **2010**, *39*, 394–395.
38. Guo, Z.; Du, S. Mining parameter information for building extraction and change detection with very high-resolution imagery and gis data. *Mapp. Sci. Remote Sens.* **2017**, *54*, 38–63. [CrossRef]
39. Liu, Z.J.; Wang, J.; Liu, W.P. Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic hough transform. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'05), Seoul, Korea, 25–29 July 2005; pp. 250–253.
40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
41. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 2481–2495. [CrossRef] [PubMed]
42. Huang, Z.; Cheng, G.; Wang, H.; Li, H.; Shi, L.; Pan, C. Building extraction from multi-source remote sensing images via deep deconvolution neural networks. In Proceedings of the Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 1835–1838.
43. Saito, S.; Yamashita, T.; Aoki, Y. Multiple object extraction from aerial imagery with convolutional neural networks. *Electron. Imaging* **2016**, *2016*, 1–9.
44. Zhong, Z.; Li, J.; Cui, W.; Jiang, H. Fully convolutional networks for building and road extraction: Preliminary results. In Proceedings of the Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 1591–1594.
45. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sens.* **2018**, *10*, 144. [CrossRef]
46. Cao, J.; Chen, Z.; Wang, B. Deep convolutional networks with superpixel segmentation for hyperspectral image classification. In Proceedings of the Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 3310–3313.
47. Zhao, W.; Jiao, L.; Ma, W.; Zhao, J.; Zhao, J.; Liu, H.; Cao, X.; Yang, S. Superpixel-based multiple local cnn for panchromatic and multispectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4141–4156. [CrossRef]
48. Liu, Y.; Cao, G.; Sun, Q.; Siegel, M. Hyperspectral classification via deep networks and superpixel segmentation. *Int. J. Remote Sens.* **2015**, *36*, 3459–3482. [CrossRef]
49. Gao, J.; Wang, Q.; Yuan, Y. Embedding structured contour and location prior in siamesed fully convolutional networks for road detection. In Proceedings of the IEEE International Conference on Robotics and Automation, Singapore, 29 May–3 June 2017; pp. 219–224.
50. Zhou, J.T.; Zhao, H.; Peng, X.; Fang, M.; Qin, Z.; Goh, R.S.M. Transfer hashing: From shallow to deep. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *PP*, 1–11. [CrossRef] [PubMed]
51. Peng, X.; Feng, J.; Xiao, S.; Yau, W.Y.; Zhou, J.T.; Yang, S. Structured autoencoders for subspace clustering. *IEEE Trans. Image Process.* **2018**, *27*, 5076–5086. [CrossRef] [PubMed]
52. 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*, 1611–1672. [CrossRef]
53. Rongming, H.U.; Huang, X.; Huang, Y. An enhanced morphological building index for building extraction from high-resolution images. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 514–520.

54. Huang, X.; Yuan, W.; Li, J.; Zhang, L. A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 654–668. [CrossRef]

55. Lin, X.; Zhang, J. Object-based morphological building index for building extraction from high resolution remote sensing imagery. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 724–733.

56. Jiménez, L.I.; Plaza, J.; Plaza, A. Efficient implementation of morphological index for building/shadow extraction from remotely sensed images. *J. Supercomput.* **2017**, *73*, 482–489. [CrossRef]

57. Ghandour, A.; Jezzini, A. Autonomous building detection using edge properties and image color invariants. *Buildings* **2018**, *8*, 65. [CrossRef]

58. Cardona, E.U.; Mering, C. Extraction of buildings in very high spatial resolution's geoeye images, an approach through the mathematical morphology. In Proceedings of the Information Systems and Technologies, Nashville, TN, USA, 12–13 November 2016; pp. 1–6.

59. Liow, Y.T.; Pavilidis, T. Use of shadows for extracting buildings in aerial images. *Comput. Vis. Graph. Image Process.* **1989**, *49*, 242–277. [CrossRef]

60. Shi, W.Z.; Mao, Z.Y. Building extraction from high resolution remotely sensed imagery based on shadows and graph-cut segmentation. *Acta Electron. Sin.* **2016**, *69*, 11–13.

61. Wang, L. Development of a multi-scale object-based shadow detection method for high spatial resolution image. *Remote Sens. Lett.* **2015**, *6*, 596–598.

62. Raju, P.L.N.; Chaudhary, H.; Jha, A.K. Shadow analysis technique for extraction of building height using high resolution satellite single image and accuracy assessment. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-8*, 1185–1192. [CrossRef]

63. 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.* **2009**, *1*, 180–192. [CrossRef]

64. Pesaresi, M.; Gerhardinger, A. Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 162–166. [CrossRef]

65. Hu, L.; Zheng, J.; Gao, F. A building extraction method using shadow in high resolution multispectral images. In Proceedings of the Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 1862–1865.

66. Fraser, C. *3D Building Reconstruction from High-Resolution Ikonos Stereo-Imagery*; Automatic Extraction Of Man-Made Objects From Aerial And Space Images (iii); Balkema: London, UK, 2001.

67. Gilani, S.; Awrangjeb, M.; Lu, G. An automatic building extraction and regularisation technique using lidar point cloud data and orthoimage. *Remote Sens.* **2016**, *8*, 27. [CrossRef]

68. Uzar, M.; Yastikli, N. Automatic building extraction using lidar and aerial photographs. *Boletim De Ciências Geodésicas* **2013**, *19*, 153–171. [CrossRef]

69. Awrangjeb, M.; Fraser, C. Automatic segmentation of raw lidar data for extraction of building roofs. *Remote Sens.* **2014**, *6*, 3716–3751. [CrossRef]

70. Shaker, I.F.; Abdelrahman, A.; Abdelgawad, A.K.; Sherief, M.A. Building extraction from high resolution space images in high density residential areas in the great cairo region. *Remote Sens.* **2011**, *3*, 781–791. [CrossRef]

71. Sportouche, H.; Tupin, F.; Denise, L. Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and sar spaceborne images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3932–3946. [CrossRef]

72. Grigillo, D.; Fras, M.K.; Petrovič, D. *Automated Building Extraction from Ikonos Images in Suburban Areas*; Taylor & Francis, Inc.: London, UK, 2012; pp. 5149–5170.

73. Hu, X.; Ye, L.; Pang, S.; Shan, J. Semi-global filtering of airborne lidar data for fast extraction of digital terrain models. *Remote Sens.* **2015**, *7*, 10996–11015. [CrossRef]

74. Pang, S.; Hu, X.; Wang, Z.; Lu, Y. Object-based analysis of airborne lidar data for building change detection. *Remote Sens.* **2014**, *6*, 10733–10749. [CrossRef]

75. Siddiqui, F.U.; Awrangjeb, M. A novel building change detection method using 3d building models. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, Sydney, Australia, 29 November–1 December 2017; pp. 1–8.

76. Yang, B.; Huang, R.; Li, J.; Tian, M.; Dai, W.; Zhong, R. Automated reconstruction of building lods from airborne lidar point clouds using an improved morphological scale space. *Remote Sens.* **2016**, *9*, 14. [CrossRef]

77. Tian, J.; Cui, S.; Reinartz, P. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosc. Remote Sens.* **2013**, *52*, 406–417. [CrossRef]

78. Siddiqui, F.U.; Awrangjeb, M.; Teng, S.W.; Lu, G. A new building mask using the gradient of heights for automatic building extraction. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, Gold Coast, Australia, 30 November–2 December 2016; pp. 1–7.

79. Mnih, V. *Machine Learning for Aerial Image Labeling*; University of Toronto: Toronto, ON, Canada, 2013.

80. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

81. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.

82. Hermosilla, T.; Ruiz, L.A.; Recio, J.A.; Estornell, J. Evaluation of automatic building detection approaches combining high resolution images and lidar data. *Remote Sens.* **2011**, *3*, 1188–1210. [CrossRef]

83. Zhang, Z.; Zhang, X.; Xin, Q.; Yang, X. Combining the pixel-based and object-based methods for building change detection using high-resolution remote sensing images. *Acta Geod. Cartogr. Sin.* **2018**, *47*, 102–112.

84. Lin, X.; Zhang, J. Extraction of human settlements from high resolution remote sensing imagery by fusing features of right angle corners and right angel sides. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 838–839.

85. Buckland, M.; Gey, F. The relationship between recall and precision. *J. Am. Soc. Inf. Sci.* **1994**, *45*, 12–19. [CrossRef]

86. Zhou, Z. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016.

87. Envi Feature Extraction Module User's Guide. Available online: http://www.harrisgeospatial.com/portals/0/pdfs/envi/Feature_Extracyion_Module.pdf (accessed on 1 December 2008).

88. Deng, S.B.; Chen, Q.J.; Du, H.J. *Envi Remote Sensing Image Processing Method*; Higher Education Press: Beijing, China, 2014.

89. Canny, J. A computational approach to edge detection. In *Readings in Computer Vision*; Elsevier: New York, NY, USA, 1987; pp. 184–203.

90. Zhang, Z.; Schwing, A.G.; Fidler, S.; Urtasun, R. Monocular object instance segmentation and depth ordering with cnns. In Proceedings of the The IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 2614–2622.