remote sensing

MDPI

Article

BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images

Zhenfeng Shao [1], Penghao Tang [2,*], Zhongyuan Wang [3], Nayyer Saleem [1] and Sarath Yam [1] and Chatpong Sommai [1]

1 The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; shaozhenfeng@whu.edu.cn (Z.S.); saleemnayyer@whu.edu.cn (N.S.); yamsarath@me.com (S.Y.); 2018276190034@whu.edu.cn (C.S.)
2 The School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, China
3 The National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430072, China; wzy_hope@whu.edu.cn
* Correspondence: tangpenghao@whu.edu.cn

Received: 26 February 2020; Accepted: 21 March 2020; Published: 24 March 2020

Abstract: Building extraction from high-resolution remote sensing images is of great significance in urban planning, population statistics, and economic forecast. However, automatic building extraction from high-resolution remote sensing images remains challenging. On the one hand, the extraction results of buildings are partially missing and incomplete due to the variation of hue and texture within a building, especially when the building size is large. On the other hand, the building footprint extraction of buildings with complex shapes is often inaccurate. To this end, we propose a new deep learning network, termed Building Residual Refine Network (BRRNet), for accurate and complete building extraction. BRRNet consists of such two parts as the prediction module and the residual refinement module. The prediction module based on an encoder–decoder structure introduces atrous convolution of different dilation rates to extract more global features, by gradually increasing the receptive field during feature extraction. When the prediction module outputs the preliminary building extraction results of the input image, the residual refinement module takes the output of the prediction module as an input. It further refines the residual between the result of the prediction module and the real result, thus improving the accuracy of building extraction. In addition, we use Dice loss as the loss function during training, which effectively alleviates the problem of data imbalance and further improves the accuracy of building extraction. The experimental results on Massachusetts Building Dataset show that our method outperforms other five state-of-the-art methods in terms of the integrity of buildings and the accuracy of complex building footprints.

Keywords: building residual refine network; convolutional neural network; building extraction; high resolution; remote sensing images

1. Introduction

With the development of imaging sensor technology, the imaging quality of remote sensing images is constantly improving, which makes obtaining high-resolution remote sensing images more convenient. High-resolution remote sensing images play an important role in many aspects, not only in marine, agricultural, and ecological protection but also in urban planning [1–4]. With the increasing amount of remote sensing images, automatic and accurate extraction of useful information from images has become valuable topic. Among them, the automatic extraction of buildings from high-resolution images is of great significance in the establishment and updating of urban geographic information

databases and urban planning. However, the phenomenon of "synonyms spectrum" or "foreign objects with the same spectrum" often appear in the remote sensing images, such as inconsistent hues and textures of the same building and similar spectra of the building and the bare land, which brings great challenges to the automatic extraction of buildings. The improvement of the spatial resolution of remote sensing images makes the information of ground features more abundant, but it also brings greater challenges to the extraction of buildings [5–8]. For example, the difference in hues and textures of the same building increases, which aggravates the problem of incomplete building extraction, and the footprints of buildings on high-resolution images are more complicated, which makes accurate extraction of them difficult for a neural network-based method. Many researchers are working on the automatic extraction of buildings from remote sensing images, and have proposed some effective building extraction methods. These methods can roughly be classified into two categories: one is based on artificially designed features and the other is based on deep learning.

The artificial features-based methods mainly make use of geometric, spectra, and contextual information design features of buildings in remote sensing images for building extraction. Lin and Nevatia [9] used the edge detection algorithm for the first time to extract buildings by detecting roof, walls, and shadows of buildings. After the edge detection, geometric constraint processing for line segment detection, or the combination of region segmentation, region growing, and region merging can contribute to the accurate extraction of buildings [10]. Katartzis et al. [11] combined the edge detection with the Markov model and used the onboard image to extract the buildings. However, the method of edge detection is vulnerable to many objects close to the building such as roads, so that the result of the extraction contains a large number of non-building areas. In addition to the use of edge or corner detection methods, researchers also use image segmentation methods for building extraction. Baatz [12] presented a fractal network algorithm to segment the image in multiple scales, and combined the image features, texture, and other features to extract the building. In the study of image feature extraction, image segmentation by a certain threshold of the index is often used for extracting features. Zhang et al. [13] proposed a pixel shape index, which extracts the buildings by clustering homogeneous pixels with similar shape and contour information. Shao et al. [14] proposed a new index to extract built-up areas from high-resolution remote sensing images by visual attention model. Liu et al. [15] used the texture, shape, spectrum, and structural features of the extracted image to construct the feature vector; introduced the machine learning method to take feature vector as input; and classified each pixel to distinguish the building. Although the traditional image segmentation methods for high-resolution images used to extract buildings have made some achievements, the extraction accuracy is not high at present. The main reason is that the spatial resolution of remote sensing images is improved, which makes the information of ground features more abundant. Overall, methods based on artificial design features often only make use of the shallow features of ground targets rather than deep features being able to effectively distinguish different ground features, leading to the low extraction accuracy. Additionally, they also require various rules to predefine features, which is thus laborious.

With the rapid development of deep learning, especially the convolutional neural network (CNN), deep learning methods have deserved great breakthrough in tasks such as natural image classification, object detection, and semantic segmentation. At present, commonly used convolutional neural networks mainly include AlexNet [16], VGGNet [17], GoogleNet [18], and ResNet [19]. Convolutional neural networks mainly consist of convolutional layers, non-linear activation functions, and pooling layers. Convolutional layers use a large number of convolution kernels to extract local features of the input image. The introduction of non-linear activation functions improves the network's extraction of non-linear features of the input image. The pooling layers can further improve the receptive field to extract more global features. Compared with the traditional artificial design feature methods, the convolutional neural network can automatically extract the features of the input image, which has gradually replaced the traditional artificial feature methods due to its powerful feature representation ability. As convolutional neural networks show strong advantages in the field of natural images, more and more researchers have tried to apply convolutional neural networks to the field of remote sensing

images, with some progress made in the segmentation and recognition of remote sensing images. Lv et al. classified remote sensing images with SEEDS-CNN and scale effectiveness analysis [20]. Chen et al. applied multi-scale CNN and scale parameter estimation in land cover classification [21]. Zhou et al. proposed So-CNN for urban functional zone fine division with VHR remote sensing images [22]. Lv et al. proposed a new method for region-based majority voting CNNs for very high-resolution image classification [23]. As an important issue of ground feature extraction, automatic building extraction has also obtained many results in the application of convolutional neural networks. Minh first applied the convolutional neural network to the building extraction of remote sensing images and proposed a building block extraction method based on image blocks. This method can directly obtain the building extraction results in the middle region of the input image [24]. To further improve the accuracy, Minh proposed using CRF or post-processing to refine the extraction results of the network. However, because these methods can only obtain the results of the middle region of the input image rather than the entire image, there are still discontinuous image blocks in the results.

In 2015, the Fully Convolutional Neural Network (FCN) was proposed to achieve pixel-level dense prediction of images [25]. FCN restores the size of the input image by using an upsampling operation and finally obtains the prediction result of each pixel of the input image. Most of the methods proposed later for image semantic segmentation were improved based on FCN, mainly including SegNet [26], DeconvNet [27], U-Net [28], and DeepLab [29]. Many researchers have also applied FCN-based methods to the automatic extraction of buildings from remote sensing images. Huang et al. proposed an improved DeconvNet, adding upsampling and dense connection operations to the deconvolution layer to get the results of building extraction [30]. Based on FCN, Maggiori et al. proposed a two-stage network that comprehensively considers the problems of identification and precise positioning [31]. Wu et al. used multi-constraint FCN to automatically extract buildings from aerial images [32]. Aiming at the problem that the use of the pooling layer will lose the information of the original image [33], Shariah et al. used FCN without the pooling layer to make the network retain as much important and useful information of the original image as possible [34]. Xu et al. combined the fully convolutional neural network and guided filtering to further optimize the extraction results of buildings [35]. Although FCN-based methods has achieved many results on the building extraction of remote sensing images, there are still some problems when applying them to the automatic extraction of buildings with high-resolution remote sensing images. On the one hand, due to the different hues and textures of the same building on the image, the extraction is incomplete or the extracted buildings are partially missed. On the other hand, these methods are not accurate enough for the building footprint extraction of buildings with complex shapes.

To better address the problems of incomplete building extraction and inaccurate building footprint extraction of complex buildings in high-resolution remote sensing images, this paper proposes a new network named Building Residual Refine Network (BRRNet). BRRNet is composed of a prediction module and a residual refinement module. The prediction module outputs the preliminary building extraction results of the input image and then the residual refinement module takes the output of the prediction module as an input. By correcting the residual between the result of the prediction module and the real result, it finally further improves the accuracy of building extraction. In the task of building extraction from high-resolution images, there is a problem of data imbalances due to the large difference in the total number of pixels between the building and the background, which affects the accuracy of building extraction. Therefore, we adopt Dice loss as the loss function during training, which effectively solves the problem of data imbalance and further improves the accuracy of building extraction.

The contributions of this paper mainly include the following three aspects.

(1) This paper proposes a new network named Building Residual Refine Network (BRRNet) composed of the prediction module and the residual refinement module for accurate and complete building extraction in remote sensing images. The extensive experimental results on Massachusetts

Building Dataset [24] show that our method outperforms other five state-of-the-art methods in terms of the integrity of buildings and the accuracy of complex building footprints.

(2) It was verified on Massachusetts Building Dataset that using Dice loss as the loss function during training is more conducive to alleviating the problem of data imbalance in building extraction tasks than BCE (Binary Cross Entropy) loss and thus boosts the accuracy of building extraction.

(3) This paper proposes a new residual refinement module named RRM_Bu equipped with deeper layers and atrous convolution and confirms that it can be readily migrated to other fully convolutional neural networks and improve the performance of the basic network.

The following sections present the proposed method and the experiments performed. Section 2 elaborates on the proposed network BRRNet. Section 3 gives the experiments and analysis, which explains in detail our comparative experiments and analysis of experimental results. In Section 4, we discuss the total parameters of different networks and the generalization ability of BRRNet. Finally, Section 5 concludes the paper.

2. Methodology

This section describes the method proposed in this paper. In particular, Section 2.1 introduces the atrous convolution. Section 2.2 presents the overall architecture of BRRNet. Section 2.3 details the network's prediction module. Section 2.4 details the network's residual refinement module. Section 2.5 describes the loss function used in the training process.

2.1. Atrous Convolution

Convolutional neural networks usually use pooling layers to expand the receptive field during feature extraction to extract more global information. Moreover, the use of pooling layers reduces the size of feature maps, which can reduce network parameters. However, because of the use of pooling layers, the resolution of the obtained feature maps are reduced so that the information of the input image is lost, which finally affects the positioning accuracy in image semantic segmentation. To further expand the receptive field during feature extraction without serious loss of input image information, researchers introduce atrous convolution into convolutional neural networks [29,36]. The atrous convolution kernel is obtained by performing a certain amount of zero paddings between the adjacent weight values of the ordinary convolution kernel. The distance between adjacent weights is called dilation rate. Compared with ordinary convolution, atrous convolution can expand the receptive field of feature extraction without increasing parameters to extract more global features. Figure 1a shows a single 3×3 ordinary convolution, whose number of parameters is 9. Figure 1b shows an atrous convolution with the size of 3×3 and the dilation rate of 2. Since the atrous convolution is obtained by filling zeros between the adjacent weights of the ordinary convolution, the atrous convolution keeps same with the number of parameters of the corresponding ordinary convolution, but the receptive field of the atrous convolution becomes larger so that it can extract more global features. We incorporate atrous convolution into the proposed network which will be explained in detail below.

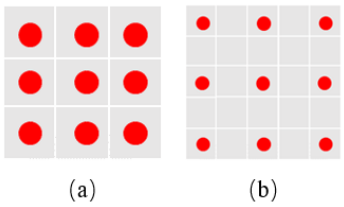

Figure 1. Examples of ordinary convolution and atrous convolution. Red dots in the grid represent weights, and non-red dots indicate that the grid has a value of 0: (**a**) an ordinary convolution with the size of 3×3; and (**b**) an atrous convolution with the size of 3×3 and the dilation rate of 2.

2.2. BRRNet

To better address the problems of incomplete building extraction and inaccurate building footprint extraction of complex buildings in high-resolution remote sensing images, we propose a new deep learning network Building Residual Refine Network (BRRNet). The network consists of two parts: the prediction module and the residual refinement module. The prediction module is based in an encoder–decoder structure, which gradually increases the receptive field during feature extraction by introducing atrous convolution with different dilation rate to extract more global features. The prediction module finally outputs the preliminary building extraction results of the input image. The residual refinement module takes the output of the prediction module as an input. By correcting the residual between the result of the prediction module and the real result, it finally further improves the accuracy of building extraction. The network's structure is shown in Figure 2.

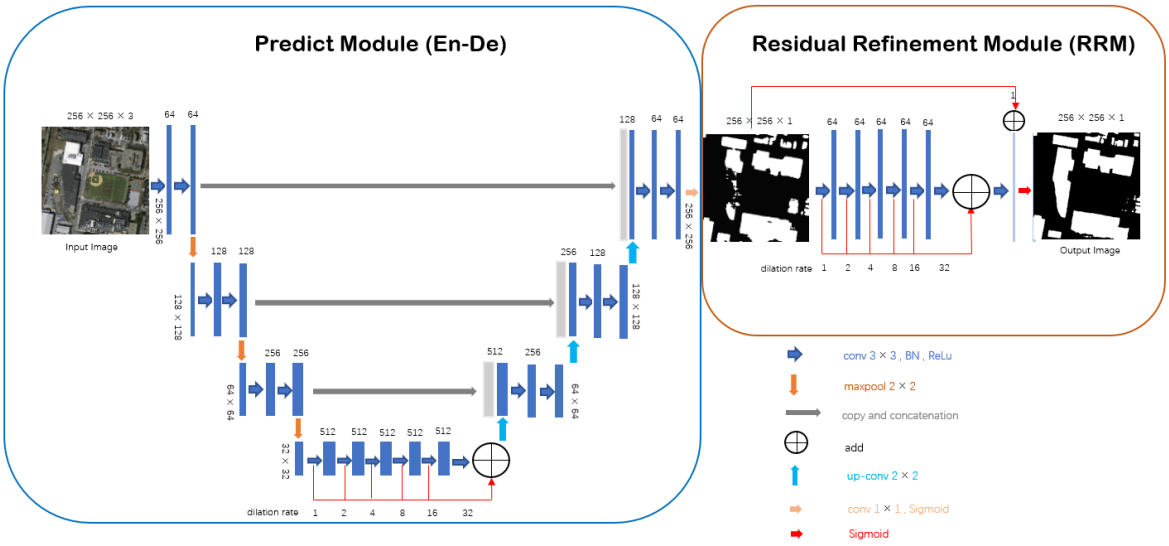

Figure 2. The structure of Building Residual Refine Network (BRRNet). BRRNet consists of two parts: the prediction module and the residual refinement module. The prediction module is based on an encoder–decoder structure, which gradually increases the receptive field during feature extraction by introducing atrous convolution with different dilation rate to extract more global features. The prediction module finally outputs the preliminary building extraction results of the input image. The residual refinement module takes the output of the prediction module as an input. By correcting the residual between the result of the prediction module and the real result, it finally further improves the accuracy of building extraction.

2.3. Prediction Module

The design of the prediction module is inspired by U-Net and the atrous convolution structure. It consists of five parts: input, encoder, bridge connection, decoder, and output. The inputs in our experiments are remote sensing image tiles. The encoder uses the first three blocks of U-Net's encoder, where each block is a two-layer convolution and a max-pooling layer with a window size of 2×2 and a step size of 2. After the convolution operation, Batch Normalization and ReLu activation functions are performed. The convolution kernel size used in these blocks is 3×3, and the number of the kernel is 64, 128, and 256, respectively. The encoder of our prediction module removes the fourth block in the U-Net's encoder because the fourth block has a large number of parameters. The previous deep learning methods applied to building extraction from high-resolution images generally have problems of incomplete building extraction and inaccurate building footprint extraction. An important reason for these problems is that, to avoid serious loss of input image's information, which affects the final positioning accuracy, the previous networks only use a small number of pooling operations. For remote

sensing images, the number of pixels occupied by buildings will increase relatively rapidly as the resolution increases. We call buildings with more pixels in the image as large buildings, and many large buildings have inconsistent hue and texture or complicated shape in high-resolution remote sensing images. The use of only a small number of pooling layers causes the receptive is insufficient to cover the entire building and the surrounding background. Therefore, it is not possible to extract the global features of the entire building, which causes problems such as incomplete building extraction and inaccurate building footprint extraction in the obtained results. To further increase the receptive field to extract more global information, and at the same time to avoid losing the information of the image as much as possible, we use a dilated convolutional series structure by successively increasing dilation rate in the bridge connection part of the prediction module. The size of the convolution kernel is 3×3 and the number of convolution kernels is 512. The dilation rate is set to 1, 2, 4, 8, 16, and 32 in turn. These hyper parameters are tuned on the validation set in experiments. After each convolution operation, Batch Normalization and ReLu activation function are used. To produce more rich information, we fuse different scales of feature maps from various dilated convolution outputs. The decoder part corresponding to the encoder part has three blocks to restore the size of the original image. Each block includes a deconvolution layer and two convolution layers. The function of the deconvolution operation is to upsample the feature maps obtained in the previous stage by two times, and then the upsampled feature maps and the corresponding encoder maps are concatenated. Then, two convolution layers are followed to extract features. Each convolution layer is followed by the Batch Normalization and ReLu activation function. Following the decoder, we use a convolution kernel of size $1 \times 1 \times 64$ to convert the number of channels of the feature maps to 1, and then the Sigmoid activation function is used to finally obtain the prediction probability map of the prediction module. To further correct the residual between the result of the prediction module and the real result, the prediction probability map is inputed to the residual refinement module.

2.4. Residual Refinement Module

At present, most of the methods for automatic building extraction of remote sensing images based on deep learning generate the building extraction results in one step without further correcting the obtained results in the model. For example, when using networks such as encoder–decoder for building extraction, the input remote sensing image is first subjected to feature extraction by the encoder, and then the size of the original image is restored by the upsampling operation of the decoder. Finally, the building extraction results can be obtained. However, the building extraction results obtained in this way may be significantly different from the real results. To further correct the residuals between the results obtained from the prediction module and the real results, we propose a new residual refinement module RRM_Bu. It takes the single-channel probability map output by the prediction module as the input, and automatically learns the residual between the input image and the corresponding real result during the training process to further refine the input image, producing more accurate building extraction results.

The residual refinement module RRM_Lc based on local context information was originally proposed by Peng et al. [37], and is used to further refine the boundary. This structure is shown in Figure 3a. Although RRM_Lc can improve the accuracy of the boundary to a certain extent, due to the small number of network layers, it is impossible to extract deeper features of the input image. On the other hand, this structure does not increase the receptive field during feature extraction so that it cannot extract more global features. In high-resolution remote sensing images, there are a large number of large buildings with inconsistent tone and textures or complex shapes. When using RRM_Lc to extract features from the input image, the receptive field is not enough to contain the entire building and the surrounding background. Consequently, it is impossible to extract sufficient global features, which fails to handle the problems of incomplete building extraction and inaccurate building footprint extraction. Therefore, we propose a new residual refinement module RRM_Bu which has more layers and larger receptive field when extracting features.

The structure of RRM_Bu is shown in Figure 3b. It is similar to the structure of the bridge connection part in the prediction module, which uses six atrous convolutions with a dilation rate of 1, 2, 4, 8, 16, and 32 to extract features and then fuses feature maps of different scales in an additive manner. After each convolution operation, the Batch Normalization and ReLu activation function are used. The number of kernels of the atrous convolution is 64. Then, a convolution kernel of size $3 \times 3 \times 64$ and a step size of 1 is used to convert the number of channels of feature map to 1. Since the input image of RRM_Bu contains the preliminary information of the prediction module, we fuse the input image with the feature map obtained at this stage in an additive manner and then input the fused result into the Sigmoid function to obtain the final probability map. Compared with RRM_Lc, the residual refinement module proposed in this paper has deeper layers, which can further extract the deep features of the input image. In addition, the use of atrous convolution gradually extracts more global information and fuse multi-scale information, which is not only beneficial to improving the building footprint extraction accuracy of the building but also beneficial to obtaining more complete building extraction results. The experiments presented in Section 3 demonstrate that our residual refinement module performs better than RRM_Lc.

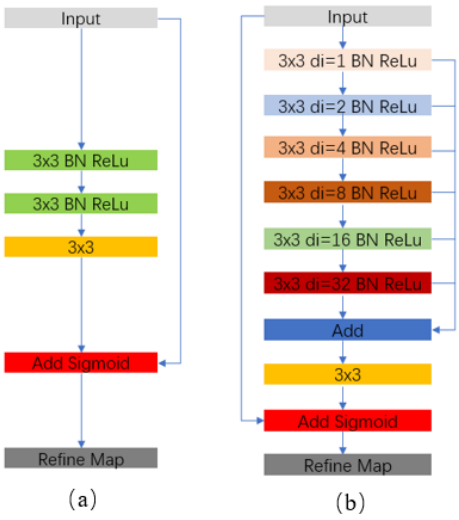

Figure 3. Comparison of: (**b**) the structure of RRM_Bu proposed in this paper; and (**a**) the structure of RRM_Lc proposed in [37], and *di* represents dilation rate.

2.5. Loss Function

The loss function adopted in the training process of the network is Dice loss, which is calculated as:

$$l_{dice} = 1 - Dice \tag{1}$$

where l_{dice} represents Dice loss and $Dice$ is Dice coefficient.

Dice coefficient, first proposed in [38], demonstrates that maximizing the Dice coefficient in training networks can solve the problem of data imbalance in medical image segmentation. Since minimizing Dice loss when training a network is essentially consistent with maximizing the goal of Dice coefficient, we use the Dice loss as the loss function. The calculation of Dice coefficient is shown in Equation (2):

$$Dice = \frac{\sum_i^N p_i \times g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2}$$

where p_i is the predicted probability value of the ith pixel of the image, g_i is the true value of the ith pixel of the image, and N is the total number of pixels of the image.

In the two-class image segmentation task, the Binary Cross Entropy loss (BCE loss) [39] is the most commonly used loss function, which is shown in Equation (3).

$$l_{bce} = -\frac{1}{N}\sum_{i}^{N}(g_i log p_i + (1 - g_i) log(1 - p_i)) \tag{3}$$

where l_{bce} is BCE loss, p_i is the predicted probability value of the ith pixel of the image, g_i is the true value of the ith pixel of the image, and N is the total number of pixels of the image.

We compare BCE loss and Dice loss in Section 3 to determine the loss function of BRRNet.

3. Experiments and Analysis

This section presents the experimental evaluation of the effectiveness of BRRNet proposed in this paper on the automatic building extraction of high-resolution remote sensing image, by comparing BRRNet with five other state-of-the-art methods. Section 3.1 describes the dataset used in the experiments. Section 3.2 presents the evaluation metrics of the experiments. Section 3.3 illustrates the details of the experiments. Section 3.4 shows the comparative experimental results along with analysis.

3.1. Dataset

The dataset used in the experiments is the Massachusetts Building Dataset [24]. This dataset contains 151 aerial images of the Boston area. The size of each image is 1500 × 1500 pixels and the resolution is 1 m. This dataset is already divided into three parts. The training set has 137 images, the validation set has 4 images, and the testing set has 10 images. The buildings in this dataset are mainly villas and commercial center buildings. There are many buildings with different tones and textures or complex shapes. Two types of buildings are shown in Figure 4.

(a) (b)

Figure 4. Two typical types of buildings in the Massachusetts Building Dataset: (**a**) the building with different tones and textures; and (**b**) the building with complex shapes.

3.2. Evaluation Metrics

We adopted two evaluation metrics to measure the effectiveness of our method: Intersection over Union (IoU) and F1-Score.

Mostly four indicators are used to evaluate the effectiveness in image pixel-level prediction tasks: true positive case (TP), false positive case (FP), true negative case (TN) and false negative case (FN). TP refers to the positive sample that is predicted to be a positive example. FP refers to the negative sample that is predicted to be a positive example. TN refers to the negative sample that is predicted to be a negative example. FN is the positive sample that is predicted to be a negative example. The evaluation metrics used to measure the effectiveness of our method are calculated based on these four indicators.

IoU is a common evaluation metric in image semantic segmentation. The degree of similarity between the predicted result and the ground truth is measured by calculating the IoU of them. The larger is the IoU value, the higher is the similarity between predicted results and ground truths. The calculation of IoU is shown in Equation (4).

$$IoU = \frac{TP}{FN + TP + FP} \tag{4}$$

F1-Score is also a commonly used evaluation metric and is the harmonic mean of precision and recall. Precision means the proportion of positive cases that are predicted to be true among all predicted positive cases. The calculation of precision is shown in Equation (5). Recall means the proportion of positive examples that are predicted to be true among all positive examples. The calculation of recall is given by Equation (6). These evaluation metrics relate to the effectiveness of the model from different aspects. To balance the results of precision and recall, we used F1-Score as another measure, as expressed in Equation (7).

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

3.3. Implementation Details

Due to memory limitations, remote sensing images needed to be cropped into tiles and then input into the network for training. We used a sliding window with a size of 256 × 256 pixels and set the step size in both the horizontal and vertical directions to 64 for image cropping. Therefore, the size of image tiles was 256 × 256 pixels. The same parameter initialization method and optimizer were used in the training process in all experiments. We used the glorot_uniform to initialize the parameters of the convolution operation. The batch size was set to 8. We used Adam as the optimizer and initialize the learning rate to 10^{-3}. We calculated the loss of validation set after each epoch. If the loss of validation set did not decrease after three epochs, the previous learning rate was multiplied by the attenuation coefficient of 0.1, and used as the new learning rate. We used a GPU to train our model and the entire training process took about 20 h.

The implementation of our model was based on TensorFlow v1.14. The experimental environment was Ubuntu16.04. The CPU model was Intel(R) Xeon(R) CPU E5-2683 v3 @ 2.00GHz and the GPU model is TITAN XP.

3.4. Comparisons and Analysis

To evaluate the effectiveness of BRRNet proposed in this paper, several sets of comparative experiments were carried out on the Massachusetts Building Dataset. Firstly, we compared different network structures and different loss functions to determine the structure and loss function of BRRNet, then compared BRRNet with five state-of-the-art methods, and finally performed migration experiments of RRM_Bu.

Comparative experiments of different network structures. We took U-Net as the baseline. Firstly, we compared our prediction module with U-Net to evaluate its effectiveness, and then we applied our residual refinement module RRM_Bu to U-Net for validation. Then, we combined the prediction module and the residual refinement module RRM_Bu to form our network BRRNet and compared it with the network combining the prediction module and the residual refinement module RRM_Lc. To ensure fairness, we used BCE loss in all comparative experiments of the network structure. Table 1 shows the results of the comparative experiment. As shown in Table 1, our prediction module for building extraction is significantly better than U-Net, with IoU/F1-Score increased by 0.0271/0.019. Applying our residual refinement module RRM_Bu to U-Net also significantly improve the result, with IoU/F1-Score increased by 0.0358/0.025. Therefore, the prediction module and the residual refinement module proposed in this paper are effective. Moreover, the combination of our prediction module and the residual refinement module RRM_Bu is better than the combination of our prediction module and RRM_Lc and obtained the highest scores in IoU and F1-Score in the comparative experiment of network structure. To further evaluate the effectiveness of our method, we analyzed the building extraction results of different methods from a qualitative perspective. Figure 5 shows

the building extraction results of the baseline U-Net and the combination of our prediction module and residual refinement module RRM_Bu. Rows 1 and 2 of Figure 5 show that, for the buildings with inconsistency of tone and textures, there is an obvious problem of missing part of the building in the building extraction results obtained using U-Net. In contrast, the buildings of this type extracted by our method are more complete. As shown in Row 3 of Figure 5, for the buildings with complex shape, the building footprints obtained by U-Net differ greatly from the ground truth, but the building footprints obtained by our method are very similar to the ground truth, which demonstrates that our method can obtain much more accurate building footprints.

Table 1. Experimental results of different network structures.

Method	IoU	F1-Score
Baseline(U-Net [28]) + bce loss	0.6743	0.8054
En-De + bce loss	0.7014	0.8245
U-Net+RRM_Bu + bce loss	0.7101	0.8304
En-De+RRM_Lc + bce loss	0.7240	0.8399
En-De + RRM_Bu(BRRNet) + bce loss	**0.7325**	**0.8456**

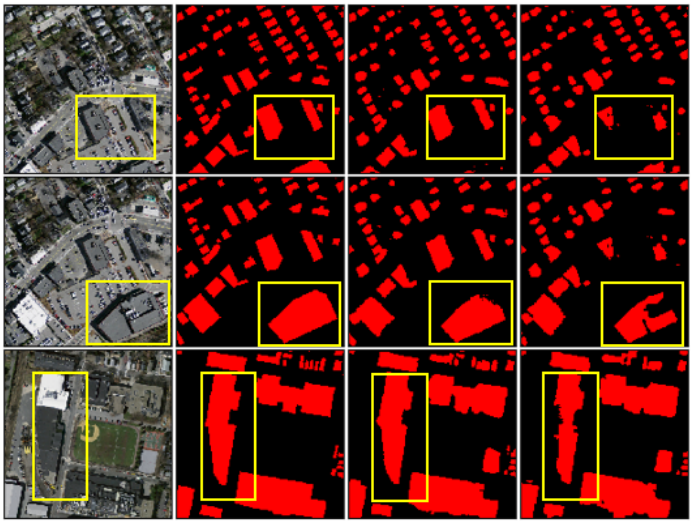

Figure 5. Typical results of the baseline U-Net and our network BRRNet, which is the combination of En-De and RRM_Bu. We used BCE loss as the loss function in these experiments: (**a**) input image; (**b**) ground truth; (**c**) the results of our network BRRNet; and (**d**) the results of the baseline U-Net. The areas in the yellow boxes are the more obvious results.

Comparative experiments of different loss functions. The number of pixels in background is usually much larger than that in the buildings in the remote sensing images, which leads to a serious data imbalance problem in the building extraction task. This problem causes the extraction results to be more biased towards the background, which causes problems with missing building parts in the results. The Dice loss proposed by Milletari et al. [38] can effectively solve the problem of data imbalance in medical image segmentation tasks. Therefore, we attempted to use Dice loss as the loss function in training to solve the data imbalance problem in the task of building extraction from remote sensing images. To validate the effectiveness of Dice loss, we used different loss functions including BCE loss, Dice loss, and the sum of BCE loss and Dice loss to train our network BRRNet. As shown in Table 2, when we used BCE loss as the loss function, the result is not so good. The result obtained by using the sum of BCE loss and Dice loss as the loss function is slightly better. Moreover, the result obtained by using Dice loss as the loss function achieves the highest score on both IoU and F1-Score. On the other hand, from a qualitative perspective, Figure 6 shows that the integrity of

building extraction in the results obtained by using Dice loss is better than that of the others. Therefore, we used Dice loss as the loss function of our network BRRNet.

When BCE loss was used as the loss function in training, the penalty weights for the pixels in buildings and background are the same. However, because the number of pixels in background is much larger than that in buildings, it is easy to fall into the local optimum instead of the global optimum in training, which causes the extraction result to be biased to the background, so that the extraction result of buildings may be partially missing. The difference from BCE loss is that the essence of Dice loss is to make the intersection over union of the predicted result and the real result continuously increase during the training process, which is conducive to approaching the global optimal. Therefore, Dice loss can better solve the problem of data imbalance and improve the integrity of building extraction.

Table 2. Experimental results of different loss functions.

Method	IoU	F1-Score
BRRNet + bce loss	0.7325	0.8456
BRRNet + bce_dice loss	0.7354	0.8475
BRRNet + dice loss	**0.7446**	**0.8536**

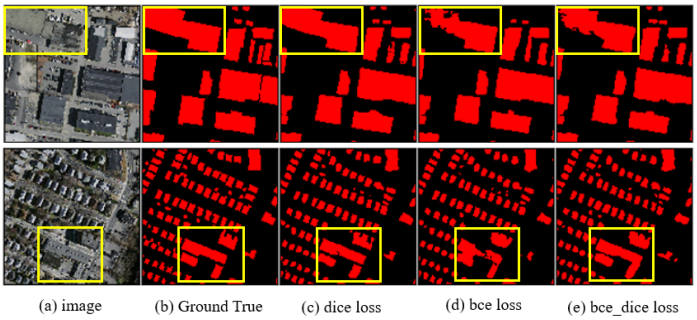

Figure 6. Typical results of different loss functions: (**a**) input image; (**b**) ground truth; (**c**) the result of Dice loss; (**d**) the result of BCE loss; and (**e**) the result of the sum of BCE loss and Dice loss. The areas in the yellow boxes are the more obvious results.

Comparative experiments with state-of-the-art methods. To further evaluate the effectiveness of our method, we compared BRRNet with five state-of-the-art methods including SegNet [26], Bayesian-SegNet [26], RefineNet [40], PSPNet [41], and DeepLabv3+ [42]. Table 3 shows the experimental results. It can be seen that our method gets higher scores than other methods in both IoU and F1-Score. Figure 7 shows the results of some typical buildings. The buildings in Rows 1 and 2 of Figure 7 are large buildings with inconsistent hue and texture. The results of the other five methods generally have the problem of incomplete building extraction. However, our method obtains much more complete results when extracting such buildings. The buildings in Rows 3 and 4 of Figure 7 have complex shape. In the results of the other five methods, the building footprints are generally different from the real results, and the details are not prominent. However, the results of our method show higher accuracy in terms of the building footprints, and the details are similar to the real results. Therefore, our method can effectively solve the problems of incomplete building extraction and inaccurate building footprint extraction of the buildings with complex shapes.

Table 3. Experimental results of our method and five other state-of-the-art methods.

Method	IoU	F1-Score
PSPNet (Dilated ResNet50) [41]	0.5847	0.7379
DeepLabv3+ (Xception) [42]	0.5913	0.7431
RefineNet (ResNet50) [40]	0.5949	0.7460
SegNet [26]	0.6798	0.8094
Bayesian-SegNet [26]	0.7003	0.8237
Ours	**0.7446**	**0.8536**

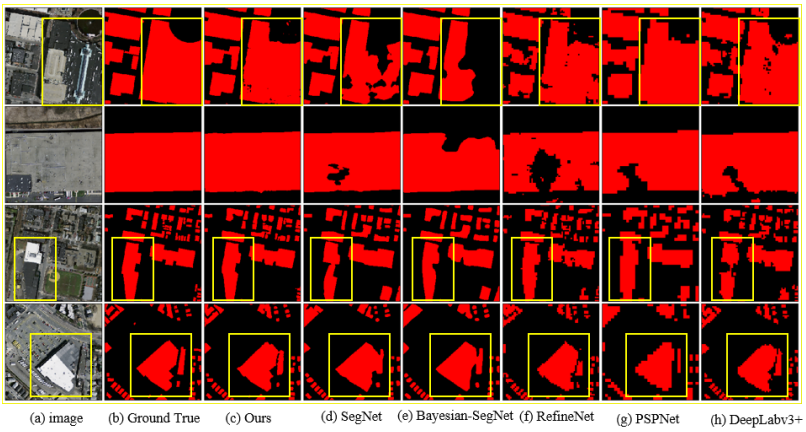

Figure 7. Typical results of our method and five other state-of-the-art methods: (**a**) Input image; (**b**) ground truth; (**c**) the result of our method BRRNet; (**d**) the result of SegNet; (**e**) the result of Bayesian-SegNet; (**f**) the result of RefineNet; (**g**) the result of PSPNet; and (**h**) the result of DeepLabv3+. The areas in the yellow boxes are the more obvious results.

We analyzed the reasons for the above results in detail. Most of the previous fully convolutional neural networks use pooling layers to increase the receptive field during feature extraction to extract more global information. The introduction of the pooling layers reduces the size of the feature maps and causes the loss of information of the original image, which eventually affects the positioning accuracy. To avoid serious loss of input image information, these models do not use too many pooling layers, and the number of pooling layers is generally less than 5. However, there are many large buildings with inconsistent tones and textures and complex shapes in high-resolution remote sensing images. This type of building occupies many pixels; thus, the methods using a small number of pooling layers to expand the receptive field cannot contain the whole building and the surrounding background. These methods can only extract local features of this type of building, which leads to incomplete extraction of such buildings and inaccurate building footprint extraction for buildings with complex shapes. Besides, because these networks are based on encoder–decoder single-stage structures, it is not possible to further refine the results.

The BRRNet proposed in this paper has two stages: preliminary prediction and residual refinement. We use our prediction module to output the preliminary results of input image. The prediction module first uses three pooling layers to expand the receptive field during feature extraction. To avoid the loss of image information due to the continued reduction in the size of the feature maps, we use an atrous convolution series structure with a gradually increasing dilation rate to further increase the receptive field, so that it can contain as much as possible the whole building and the surrounding background to extract more global information. Then, it fuses the multi-scale feature maps of different atrous convolutions to obtain richer information, which better solves the problem of incomplete extraction of such buildings. Furthermore, we use the output of the prediction module as the input of our residual refinement module RRM_Bu, so that our network can further learn the residuals between the output of the prediction module and the real results. Particularly, when the difference between the preliminary result and the real label becomes large, our network will further

refine the residuals to produce more complete buildings and accurate building footprints. In addition, we use Dice loss as the loss function in training to handle the data imbalance that biases the extraction results to the background, which also improves the accuracy of buildings.

Migration experiments of RRM_Bu. The prediction module and the residual refinement module proposed in this paper are two separate parts of our network BRRNet. We use the residual refinement module to further refine the residuals between the output of the prediction module and the ground truth so as to obtain more accurate building extraction results. To confirm that our residual refinement module can be readily migrated to other networks and improve the performance of the basic networks, we migrated it to three different networks: FCN-32s [25], SegNet [26], and Bayesian-SegNet [26]. The experimental results are shown in Table 4. As shown in the table, migrating our residual refinement module to these networks significantly improves the effectiveness of the corresponding counterparts, especially in the case of poor results given by the basic network. For example, the result of FCN-32s has been greatly improved after migrating RRM_Bu to it.

Table 4. Experimental results in migrating RRM_Bu to other networks.

Method	IoU	F1-Score
FCN-32s [25]	0.4203	0.5918
FCN-32s + RRM_Bu	**0.5039**	**0.6702**
SegNet [26]	0.6798	0.8094
SegNet + RRM_Bu	**0.6888**	**0.8157**
Bayesian-SegNet [26]	0.7003	0.8237
Bayesian-SegNet + RRM_Bu	**0.7026**	**0.8253**

4. Discussions

4.1. Total Parameters of Different Networks

As shown in Section 3.4, we compared the efficiency of our proposed network BRRNet with five state-of-the-art methods on the Massachusetts Building Dataset, and experimentally demonstrated that BRRNet obtains the highest scores on IoU and F1-Score. In addition to IoU and F1-Score, the total parameters of the network is also an important metric for evaluating the efficiency of a network. The more the parameters the network has, the larger the memory it needs to occupy during the training and testing process. Therefore, we further analyzed the total parameters of the proposed network and the five other state-of-the-art methods. Figure 8 shows the total parameters of different networks and the F1-Scores evaluated on the Massachusetts buildings dataset. Figure 8 shows that the total parameters of Bayesian-SegNet and SegNet are the least. The total parameters of our proposed BRRNet are slightly larger than those of the two methods while it obtains significantly higher F1-Score. In contrast, RefineNet, DeepLabv3+, and PSPNet have much more parameters and require a lot of memory during training and testing. Our proposed BRRNet is based on the improvement of U-Net. The total parameters of U-Net is 31.0M, while the total parameters of BRRNet is only 17.3M because BRRNet removes the fourth block of the encoder and the first block of the decoder of U-Net, which greatly reduces total parameters. It can be seen that our proposed BRRNet can not only obtain good results in the task of building extraction, but also has a relatively few parameters.

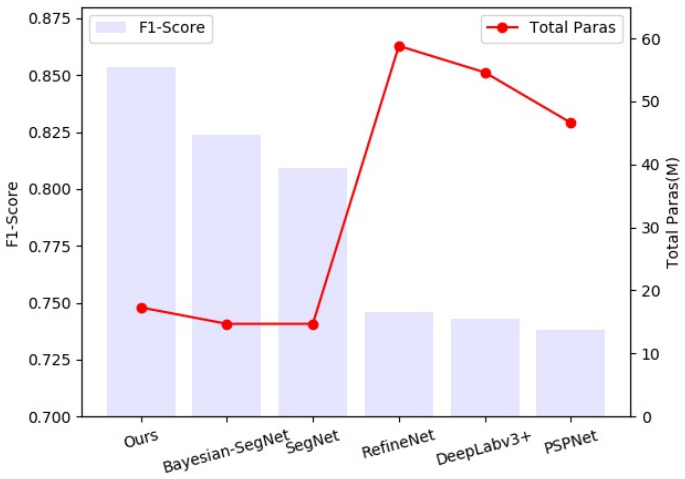

Figure 8. Segmentation accuracy and total parameters comparison of different networks. The bar diagram represents the F1-Scores evaluated on the Massachusetts Building Dataset and the line chart represents total parameters of different networks. The unit is Million (M).

4.2. Generalization Ability of BRRNet

To further discuss the generalization ability of our proposed BRRNet, we performed experiments on transfer learning. We trained several networks on the training set of the Massachusetts Building Dataset, and then tested the trained networks on the new test set which consists of 20 images randomly selected from the Inria Aerial Image Labeling (IAIL) Dataset [43]. This dataset is made up of 360 ortho-rectified aerial RGB images of 5000 $\times$ 5000 in size, with a spatial resolution of 0.3 m/pixel. Since the Massachusetts Building Dataset and the Inria Aerial Image Labeling Dataset cover different regions and consist of images with different spatial resolutions, the features of the buildings in these two datasets are quite different. Table 5 shows the test results of different networks. As shown in Table 5, when different networks were trained on the training set of the Massachusetts Building Dataset and then tested on the new test set, our proposed BRRNet obtained the highest scores on IoU and F1-Score compared with the other three networks, which demonstrates that BRRNet has better generalization ability. However, since the features of the buildings in these two datasets are quite different, the IoUs and F1-Scores evaluated on the Inria Aerial Image Labeling Dataset are much lower than those evaluated on the Massachusetts Building Dataset. In future studies, we will try to use a small number of images from the training set of the Inria Aerial Image Labeling Dataset to fine-tune these networks to obtain better test results.

Table 5. Transfer learning results of different networks.

Method	Massachusetts		IAIL	
	IoU	**F1-Score**	**IoU**	**F1-Score**
SegNet	0.6798	0.8094	0.2118	0.3196
DeepLabv3+	0.5847	0.7379	0.2322	0.3769
U-Net	0.6743	0.8054	0.2587	0.4110
Ours	**0.7446**	**0.8536**	**0.2790**	**0.4363**

The column "Massachusetts" represents the results of the networks that were trained on the training set of the Massachusetts Building Dataset and then tested on the test set of the Massachusetts Building Dataset. The column "IAIL" represents the results of the networks that were trained on the training set of the Massachusetts Building Dataset, and then tested on the new test set which consists of 20 images randomly selected from the Inria Aerial Image Labeling (IAIL) Dataset.

5. Conclusions

In this paper, we propose a new end-to-end deep learning network BRRNet to address the problems of incomplete building extraction and inaccurate building footprint extraction of the buildings with complex shapes in high-resolution remote sensing images. BRRNet consists of two parts: the prediction module and the residual refinement module. The prediction module uses the atrous convolution with different dilation rates for gradually increasing the receptive field during feature extraction so as to extract more global information, and then fuses multi-scale features from different layers to gain richer information. Then, the residual refinement module further refines the residuals between the preliminary results from prediction module and the ground truth, in order to improve the accuracy of building extraction. In addition, we also experimentally validated that the Dice loss-based loss function in training can effectively alleviate the problem of data imbalances and improve the accuracy of building extraction from remote sensing images. Experimental results on Massachusetts Building Dataset show that BRRNet is significantly superior over five other state-of-the-art methods in terms of building integrity and building footprint accuracy.

Author Contributions: Z.S. wrote the manuscript and designed the comparative experiments; P.T. designed the architecture and performed the comparative experiments; Z.W. supervised the study and revised the manuscript; N.S. revised the manuscript; and S.Y. and C.S. gave comments and suggestions to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National key R&D plan on strategic international scientific and technological innovation cooperation special project under Grant 2016YFE0202300; the National Natural Science Foundation of China under Grants 61671332, 41771452, 51708426, 41890820 and 41771454; and the Natural Science Fund of Hubei Province in China under Grant 2018CFA007.

Acknowledgments: We would like to thank the anonymous reviewers for their constructive and valuable suggestions on the earlier drafts of this manuscript.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Lo, C.P.; Quattrochi, D.A.; Luvall, J.C. Application of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *Int. J. Remote Sens.* **1997**, *18*, 287–304. [CrossRef]
2. Nichol, J.E.; Shaker, A.; Wong, M.S. Application of high-resolution stereo satellite images to detailed landslide hazard assessment. *Geomorphology* **2006**, *76*, 68–75. [CrossRef]
3. Yang, Q.H.; Qi, J.W.; Sun, Y.J. The Application of High Resolution Satellite Remotely Sensed Data to Landuse Dynamic Monitoring. *Remote Sens. Land Resour.* **2001**, *13*, 20–27.
4. Wulder, M.A.; Hall, R.J.; Coops, N.C.; Franklin, S.E. High spatial resolution remotely sensed data for ecosystem characterization. *BioScience* **2004**, *54*, 511–521. [CrossRef]
5. Schiewe, J. Segmentation of high-resolution remotely sensed data-concepts, applications and problems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2002**, *34*, 380–385.
6. Ehlers, M.; Gähler, M.; Janowsky, R. Automated analysis of ultra high resolution remote sensing data for biotope type mapping: New possibilities and challenges. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 315–326. [CrossRef]
7. Benediktsson, J.A.; Chanussot, J.; Moon, W.M. Very high-resolution remote sensing: Challenges and opportunities [point of view]. *Proc. IEEE* **2012**, *100*, 1907–1910. [CrossRef]
8. Mahabir, R.; Croitoru, A.; Crooks, A.; Agouris, P.; Stefanidis, A. A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Sci.* **2018**, *2*, 8. [CrossRef]
9. Lin, C.; Nevatia, R. Building Detection and Description from a Single Intensity Image. *Comput. Vis. Image Underst.* **1998**, *72*, 101–121. [CrossRef]
10. Fan, R.; Chen, Y.; Xu, Q.; Wang, J. A high-resolution remote sensing image building extraction method based on deep learning. *Acta Geodaetica et Cartographica Sinica* **2019**, *48*, 34.

11. Katartzis, A.; Sahli, H.; Nyssen, E.; Cornelis, J. Detection of Buildings from a Single Airborne Image using a Markov Random Field Model. In Proceedings of the IGARSS 2001. Scanning the Present and Resolving the Future. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, Australia, 9–13 July 2001.
12. Baatz, M. Object-oriented and multi-scale image analysis in semantic networks. In Proceedings of the 2nd International Symposium on Operationalization of Remote Sensing, Enschede, The Netherlands, 16–20 August 1999.
13. Zhang, L.; Huang, X.; Huang, B.; Li, P. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2950–2961. [CrossRef]
14. Shao, Z.; Tian, Y.; Shen, X. BASI: A new index to extract built-up areas from high-resolution remote sensing images by visual attention model. *Remote Sens. Lett.* **2014**, *5*, 305–314. [CrossRef]
15. Liu, Z.; Wang, J.; Liu, W. 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, Seoul, Korea, 25–29 July 2005; Volume 4, pp. 2250–2253.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Szegedy, C.; Liu, W.; Jia, Y.; 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, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
20. Lv, X.; Ming, D.; Chen, Y.; Wang, M. Very high resolution remote sensing image classification with SEEDS-CNN and scale effect analysis for superpixel CNN classification. *Int. J. Remote Sens.* **2019**, *40*, 506–531. [CrossRef]
21. Chen, Y.; Ming, D.; Lv, X. Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation. *Earth Sci. Inform.* **2019**, *12*, 1–23. [CrossRef]
22. Zhou, W.; Ming, D.; Lv, X.; Zhou, K.; Bao, H.; Hong, Z. SO–CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sens. Environ.* **2020**, *236*, 111458. [CrossRef]
23. Lv, X.; Ming, D.; Lu, T.; Zhou, K.; Wang, M.; Bao, H. A new method for region-based majority voting CNNs for very high resolution image classification. *Remote Sens.* **2018**, *10*, 1946. [CrossRef]
24. Mnih, V. *Machine Learning for Aerial Image Labeling*; Citeseer: Princeton, NJ, USA, 2013.
25. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
26. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
27. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 818–833.
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin, Germany, 2015; pp. 234–241.
29. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
30. 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 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1835–1838.
31. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 645–657. [CrossRef]

32. Wu, G.; Shao, X.; Guo, Z.; Chen, Q.; Yuan, W.; Shi, X.; Xu, Y.; Shibasaki, R. Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. *Remote Sens.* **2018**, *10*, 407. [CrossRef]
33. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. High-resolution aerial image labeling with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7092–7103. [CrossRef]
34. Sherrah, J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv* **2016**, arXiv:1606.02585.
35. 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]
36. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
37. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large Kernel Matters–Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
38. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
39. De Boer, P.T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [CrossRef]
40. Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
41. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
42. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
43. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3226–3229.