remote sensing

MDPI

Article

Efficient Occluded Road Extraction from High-Resolution Remote Sensing Imagery

Dejun Feng, Xingyu Shen, Yakun Xie *, Yangge Liu and Jian Wang

Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; djfeng@swjtu.edu.cn (D.F.); sxyu@my.swjtu.edu.cn (X.S.); lyg99520@my.swjtu.edu.cn (Y.L.); wjian1025@163.com (J.W.)

* Correspondence: yakunxie@my.swjtu.edu.cn

Abstract: Road extraction is important for road network renewal, intelligent transportation systems and smart cities. This paper proposes an effective method to improve road extraction accuracy and reconstruct the broken road lines caused by ground occlusion. Firstly, an attention mechanism-based convolution neural network is established to enhance feature extraction capability. By highlighting key areas and restraining interference features, the road extraction accuracy is improved. Secondly, for the common broken road problem in the extraction results, a heuristic method based on connected domain analysis is proposed to reconstruct the road. An experiment is carried out on a benchmark dataset to prove the effectiveness of this method, and the result is compared with that of several famous deep learning models including FCN8s, SegNet, U-Net and D-Linknet. The comparison shows that this model increases the IOU value and the F1 score by 3.35–12.8% and 2.41–9.8%, respectively. Additionally, the result proves the proposed method is effective at extracting roads from occluded areas.

Keywords: remote sensing imagery; occluded road extraction; convolutional neural network; attention mechanism

check for updates

Citation: Feng, D.; Shen, X.; Xie, Y.; Liu, Y.; Wang, J. Efficient Occluded Road Extraction from High-Resolution Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4974. https://doi.org/10.3390/rs13244974

Academic Editor: Saeid Homayouni

Received: 29 October 2021
Accepted: 4 December 2021
Published: 7 December 2021

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

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Road information has received substantial attention due to its significance in geography [1–5]. At present, with the increase of the number and resolution of remote sensing images, the automatic extraction of road information by computers has become a hot research topic [6–8]. It is both theoretically and practically significant for improving the automation level of remote sensing data applications, reducing manual labor and unifying data production specifications [9,10].

Road extraction methods mainly include traditional methods based on remote sensing image features and deep learning methods, combined with computer vision. The traditional pixel-based methods focus on the difference in radiation features between roads and other objects. Researchers usually use spectral information to segment the images first, and then extract roads according to shape features [11], texture features [12], spectral characteristics [13] or pixel footprints [14,15]. Alternatively, threshold segmentation based on the gray value of images is the other common pixel-based method [16–18]. Its algorithm is not complicated and is easy to implement. However, its effect is limited to a certain type of image in the dataset, and the update of the threshold takes much calculation [19]. Finally, some pixel-based edge detection algorithms are also suitable for road extraction, such Canny arithmetic [20,21], the morphological method [22,23] and so on [24,25]. However, this approach has some limitations in the direction of the roads. Additionally, pixel-based methods will inevitably extract more noisy spots and require more cumbersome post-processing procedures. The object-based method performs better in suppressing noise [26]. It segments the images into multiple objects, each composed of pixels with similar spectral features, and then filters the road objects [27–31], but this method is only effective for

images with a single road type, such as urban images. In conclusion, the accuracy of traditional road extraction methods is not high enough. The versatility of these methods on different types of remote sensing images is poor, and it is hard to overcome the problem of roads being occluded by other ground objects such as buildings, vegetation and shadows.

In the early years, the deep learning theory had been proved to have great potential in computer vision [32,33]. The convolutional neural network (CNN) shows a great ability in digging deep features and performs well in the fields of image classification, object detection and image semantic segmentation [34–40].

On this basis, as an important research direction of image semantic segmentation, the effect and accuracy of road extraction are greatly improved. In 2010, Mnih et al., first used a CNN to extract road information in remote sensing images [41]. After that, Gao et al., proposed a siamesed fully convolutional network based on VGG-net architecture, which can learn more discriminative features and improve training speeds [42]. Cheng et al., proposed an end-to-end convolutional neural network to extract roads and recover the center line [43]. Gao et al., designed a tailored pyramid pooling module and embedded it into an end-to-end framework to solve the class imbalance problem caused by the sparseness of roads. The network performed well in narrow rural roads [44]. Wei et al., better improved the road extraction quality of aerial images via a road structure-refined convolutional neural network [45]. Zhou et al., proposed D-Linknet, based on Linknet, which utilized dilated convolution to enlarge the receptive field of the kernel and shortened the training time [46,47]. In recent years, the encoder-decoder based U-Net has been widely used in road extraction. This network can extract the deep features of images in the encoder stage and recover a pixel-wise segmentation image by up-sampling in the decoder stage [48]. Wang et al., used VGG16 as the network coding structure, and expanded the training dataset to enhance the ability of the network to extract roads [49]. Yang et al., embedded the residual block which enables the learning of very deep neural networks into the U-net network, and effectively extracted and fused abundant local features [50,51]. Zhang et al., proposed a semantic segmentation neural network by combining residual learning and U-Net to extract road information from high-resolution remote sensing images [52]. Sofla et al., improved U-Net with spatially squeezing and exciting channel-wise to highlight useful channels of the images [53]. Ren et al., proposed a dual-attention capsule U-Net and better used the context features of the images, but this performed less effectively when roads were severely occluded [54]. Lin et al., incorporated medium-resolution optical and SAR data and effectively improved the accuracy of road extraction in non-urban areas, but the applicability of this method is limited by the acquisition of datasets [55]. These approaches all verify the feasibility of extracting roads with deep learning.

These deep learning methods well combine the characteristics of remote sensing images with computer vision and greatly improve the accuracy and the completeness of the road extraction results. There are still some problems: (1) Insufficient extraction capability caused by network structure and dataset limitations; (2) The occlusion of other objects on the image changes the shape of the road, making the network unable to recognize the road; and (3) Many road extraction methods based on deep learning do not pay much attention to post-processing, but appropriate post-processing can effectively compensate for network defects and solve obvious errors in the network extraction results, such as error patches and broken road lines.

To solve these problems and refine on the road extraction methods, we select a well-performing U-Net basic model and a dataset with both good quality and data volume after some experiments. Finally, this paper proposes a network structure with strong feature extraction ability and a post-processing method with a simple algorithm and obvious effect. The main contributions of this paper are as follows.

(1) We propose a novel MAU-Net method to extract roads from high-resolution remote sensing images. With the use the of the attention mechanism module in the feature extraction and feature fusion, the network can adaptively adjust its attention to different regions, and weaken the influence of the interference area while highlighting

the key areas. Thus, the road extract ability of network is improved. Experiments show that the proposed method has distinct advantages over existing methods on a benchmark dataset.

(2) A data enhancement method aiming at occlusions is applied. Beyond the conventional method, this paper focuses on the situation of a road occluded by other ground objects, and the quality of the dataset is improved by generating an occlusion mask. This method can improve the network's extraction ability towards occluded roads.

(3) After analyzing the problems existing in the road semantic segmentation results, a geometric topology reconstruction algorithm based on connected domain analysis is established, and it can further weaken or eliminate the fracture phenomenon and remove the errors of extracted spots in the road extraction results. Compared with other post-processing methods, this algorithm is easy to realize and can achieve a better visual effect.

The structure of this paper is as follows: Section 2 details the methods of this paper, including convolutional neural networks with attention mechanisms and a resulting optimization method based on connected domain. The dataset, experimental environment, hyper-parameter settings and evaluation criteria are described in Section 3. The experimental results and discussion are presented in Section 4. Finally, Section 5 summarizes the method of this paper and presents future research directions.

2. Methods

The road extraction method in this paper includes the preliminary extraction stage and post-processing stage, as shown in Figure 1. The network for road extraction consists of two parts: encoding and decoding. The post-processing method based on the connected domain is useful for the restoring of roads' topological information.

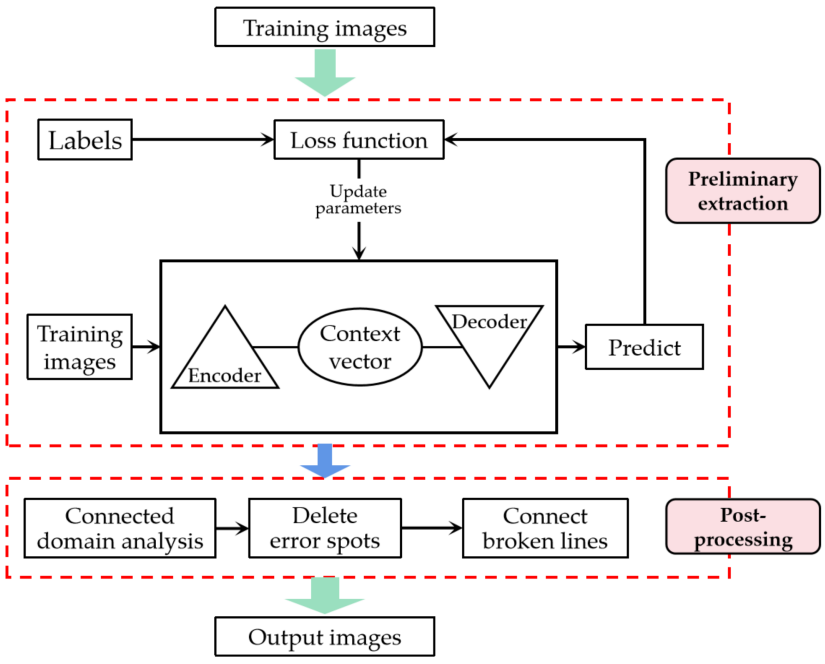

Figure 1. The overall framework of the proposed method: the preliminary extraction stage and the post-processing stage.

To improve the feature extraction capability of the deep convolutional neural network, this paper introduces modules combining channel attention mechanism and spatial attention mechanism into the encoding network [56], in which the channel attention mechanism enhances attention to useful features in the channel dimension while suppressing useless background information; thus, the efficiency of the network is improved. The spatial

attention mechanism concentrates more on important pixels on the feature map, which highlights the key areas by modeling the correlation of the individual point feature and the regional point feature, so as to improve the accuracy of the network. Secondly, in the feature fusion stage, this paper designs a method incorporating the attention mechanism, and connects the feature maps based on their correlation, which will further improve the attention given to key areas. Finally, the output of this paper aggregates information at multiple scales, which enhances the generalization of the model.

2.1. Road Extraction Using Attention Convolutional Neural Network

Aiming at the low accuracy in road extraction caused by the limitations of neural networks, this paper proposes a convolutional neural network model based on an enhanced attention mechanism. As shown in Figure 2, the network model consists of three parts. Firstly, the enhanced self-attention mechanism in the coding process establishes an effective context autonomous learning module by combining a space attention mechanism and channel attention mechanism, with which the feature extraction ability of the network will be improved. Secondly, the feature fusion algorithm, being based on the attention mechanism in the decoding process, which improves the fusion ability of feature maps with the same size in the encoding and decoding process, further highlights the key areas and suppresses the interference features. Finally, multi-scale aggregation is used to obtain better road extraction results.

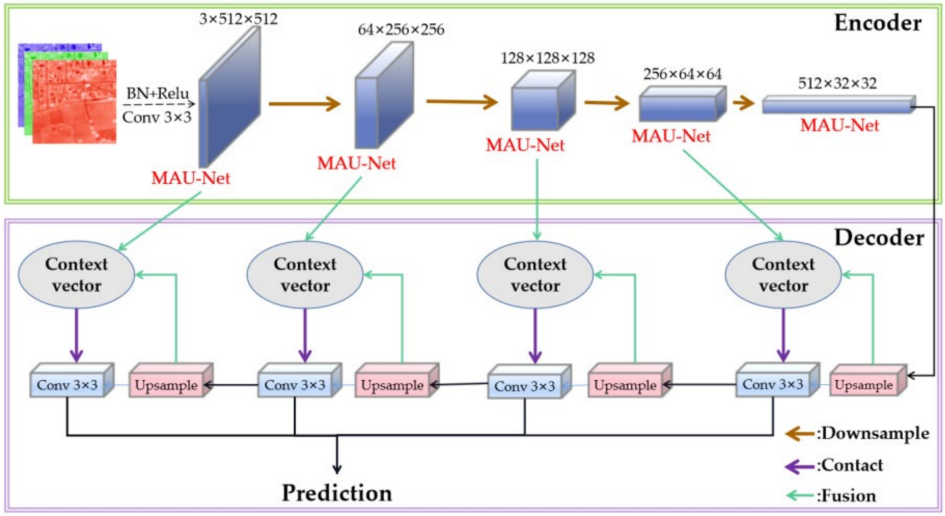

Figure 2. Structure of the mask attention convolutional neural network.

2.1.1. Enhanced Attention Module

The attention mechanism enables the network to selectively focus on certain important features. Given the small proportion of roads in the image, improving the attention given to the target area can greatly improve the extraction effect. Commonly used attention mechanisms include spatial attention mechanisms and channel attention mechanisms, where the spatial dimension helps the network identify important points on the image, while the channel dimension allows the network to focus on the most important channels. However, traditional attention mechanisms often require the manual design of convolution and pooling, which constantly improves the complexity of the network. As such, in this paper, with the calculation of the autocorrelation matrix, an attention mechanism module combining the spatial attention mechanism and the channel attention mechanism is constructed to improve the dependent manner of global features in the spatial and channel dimensions, as shown in Figure 3.

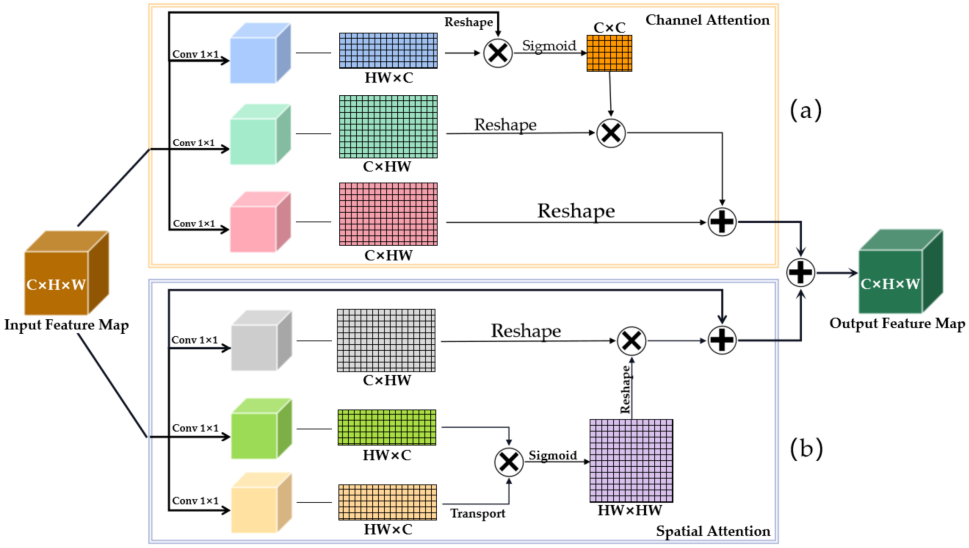

Figure 3. Structure of the attention mechanism module; (**a**) the channel attention module is designed by autocorrelation matrix operation and (**b**) the autocorrelation matrix of pixels is calculated to model a stronger correlation of individual point features and point features in local areas, which will further highlight the key areas.

Firstly, to simulate the interdependence between channels, the channel attention module is designed by autocorrelation matrix operation, as shown in Figure 3a. Different channels of the feature map correspond to different combinations of convolutional kernels, and channel attention gives different attention to different convolution processes, while the interdependency of channels is further strengthened by the autocorrelation matrix. Secondly, the spatial attention module is constructed to mine the correlation of pixels. Figure 3b shows that the autocorrelation matrix of pixels is calculated to model a stronger correlation of individual point features and point features in local areas, which will further highlight the key areas. Finally, the weighted feature maps obtained via channel attention and spatial attention are added to fully use the context information, which significantly improves the segmentation results.

2.1.2. Feature Fusion Based on Attention Mechanism

Furthermore, the convolutional modules of the same size during coding tend to be multiple, and connecting the decoding process to a given size graph during encoding also results in partial loss of features. Further, there are often multiple convolution modules of the same size in the encoding process, and the feature tends to partially lose if simply connecting maps of the same size in the encoding and decoding process. Aiming at these problems, this paper proposes a feature fusion algorithm based on the attention mechanism to fully use the feature map extracted by each convolution layer in the encoder, as shown in Figure 4.

Firstly, the inner product of the up-sample output at the decoder and each output after two convolution layers (conv1, conv2) after the down-sampling at the encoder are made, and the correlation degree between the convolutions of each layer in the process of coding and decoding is calculated, as shown in Equation (1). Secondly, the correlation degree is normalized to obtain the attention score, which can effectively highlight the key areas and suppress the influence of interference areas, as shown in Equation (2). Finally, the fused context vector is obtained through the weighted calculation of layers, as shown in Equation (3), so as to realize the feature fusion considering the context information. Compared with traditional fusion methods, the proposed method fully uses the feature map extracted by each convolution module of the encoder and improves the network's ability to extract the morphological features and the network's attention to key areas.

$$e_{nchw} = Output_{upsample} * key_{nchw} \quad (1)$$

where $Output_{upsample}$ is the outcome of upsampling; key_{nchw} represents the pixel values of the h row w column of c channel of n-th key tensor, that is, Conv1 and Conv2; e_{nchw} represents the weight of the h row w column of c channel of n-th key tensor; and $n = \{1, 2\}$.

$$score_{nchw} = \frac{exp(e_{nchw})}{\sum_{i=1}^{n} exp(e_{ichw})} \quad (2)$$

in which the $score_{nchw}$ and e_{nchw} are both tensors of $c \times h \times w$.

$$V_{chw} = \sum_{i=1}^{n} key_{ichw} * score_{ichw} \quad (3)$$

where tensor V has the same size with $Output_{upsample}$, and the concat of them will be input to the next convolution layer in the decoding phase

2.1.3. Multi-Scale Aggregate Output

In general, the scale of each type of road varies greatly. Traditional methods of taking the output of the last single layer of the network as the extraction result will lose features on some scales. To solve this problem, the multi-scale output structure of FPN [57] is used as a reference, and a deep multi-scale aggregation structure is established to better fit the multi-scale road information. This aggregation method uses information from all scales in a separate and integrated manner. Compared with single multi-scale weighted output, the physical combination can aggregate all multi-scale features and improve the optimization degree of the network. Additionally, as the proposed method only optimizes the output based on the original network, it hardly increases the computing time.

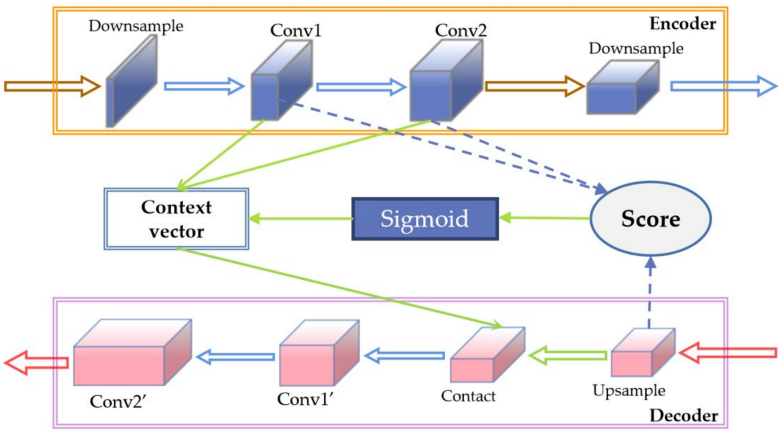

Figure 4. Structure of the feature fusion module based on attention mechanism.

2.2. Post-Processing Based on Connected Domain Analysis

2.2.1. Problems in Road Extraction

As the resolution of the remote sensing image improves, the details and background information of the image increase. Various problems will occur in the process of road extraction. The most typical problems are extracting background as road, as shown in the red ellipse in Figure 5b, and the disconnection of the road line caused by the occlusion of other ground features, as shown in the green ellipse in Figure 5b, both of which greatly reduce the integrity and accuracy of road extraction.

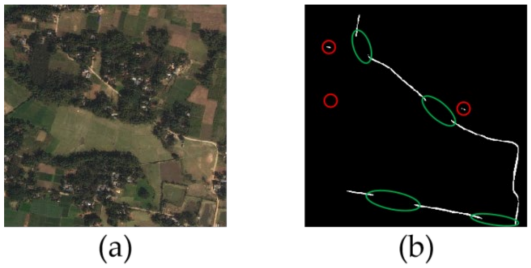

Figure 5. Problems in road extraction. (**a**) Original image; (**b**) Interference and disconnection of road extraction result.

2.2.2. Post-Processing Algorithm

Aiming at these problems, the algorithm of mathematical morphology dilation and erosion is commonly used. However, this method will change the shape of the road to a certain extent, with many steps and a large amount of calculation. In contrast, this paper proposes a post-processing method that is easy to operate and does not change all extraction results, but only restores the topological property of the road.

Roads occupy a certain area in remote sensing images and usually show connectivity. With these properties, the road extraction results can be optimized to solve the problem of error extraction and broken road lines. This paper proposes a post-processing method based on connected domains. First, the connected domains are labeled from the preliminary extraction result, and then the error extracted non-road blocks are removed based on the area. After that, the broken road can be reconstructed according to the distance of the connected domain. The process is shown in Figure 6.

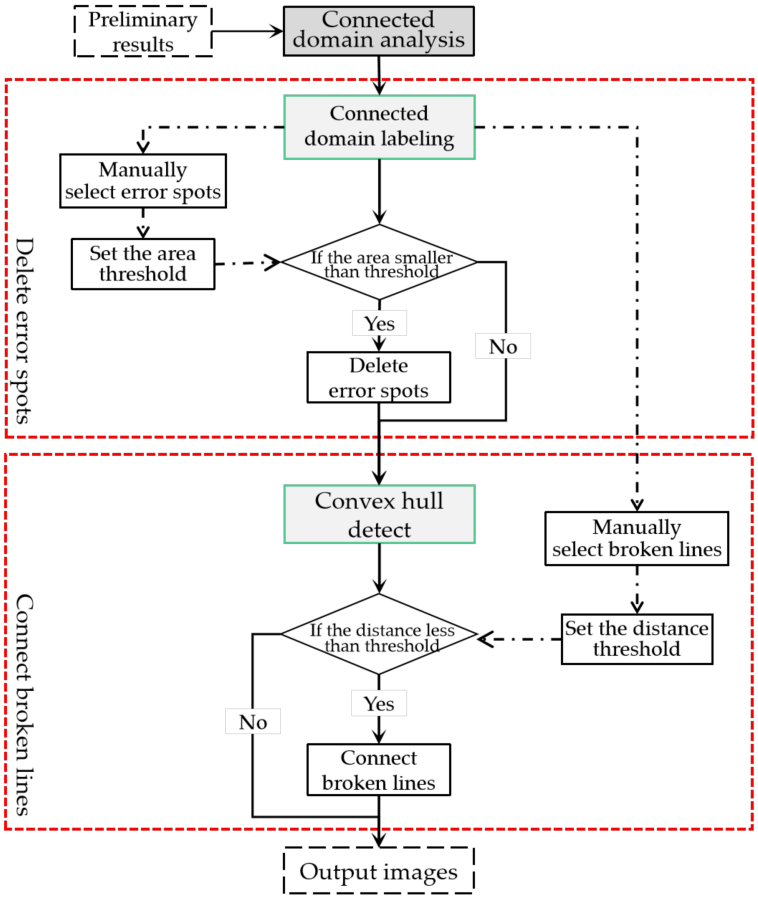

Figure 6. Flow chart of the road post-processing algorithm.

(1) Connected domain labeling

Connected domain labeling is one of the most basic methods for the analysis of binary images. It identifies each individual connected region by tracking the pixels, and then various geometric parameters of these blocks, such as contour and bounding rectangles, can be obtained. Figure 7 shows a four-connected domain and an eight-connected domain; two common adjacency relations.

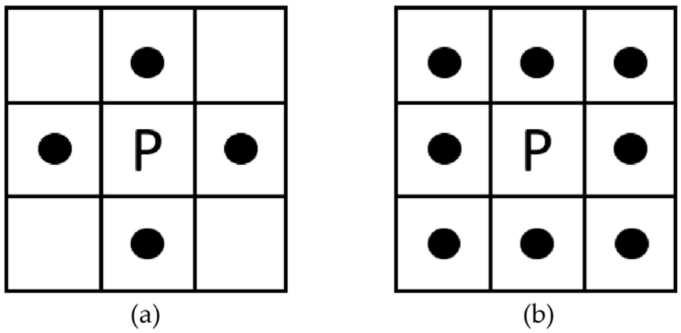

Figure 7. Connected domains. (**a**) Four-connected domain; (**b**) Eight-connected domain.

The binary image extracted by the network can be regarded as a two-dimensional matrix, and extracting the connected domain of an image is to label pixels with the same value under the domain relationship. In this paper, connected domains are extracted by the common Seed-Filling method. It takes a pixel with a value of 1 as the seed point, and constantly integrates the pixels with the same value in the neighborhood into a set to form a separate connected domain. Traversing the image can obtain all the connected domains in the image.

(2) Connected domain-based post-processing

Preliminary experiments indicate that the error extracted non-road spots are usually scattered and small, making them suitable for processing by threshold segmentation. Therefore, taking the connected domain as the basic unit, this paper manually selects typical non-road spots in the connected domain extraction results, and then the mean size of these spots is set as the threshold of image segmentation. Finally, these non-road spots can be identified and corrected to background information.

Secondly, the occlusion of trees, buildings and other ground objects will inevitably lead to the problem of road disconnection in road extraction. For this case, this paper tries to reconstruct the broken road lines based on the geometric distance of the connected domain contour. Firstly, the convex envelopes of each connected domain are extracted, then the shortest distance between the convex envelopes of each connected domain is calculated, and if it is less than the threshold, the convex envelopes will be connected by an appropriate line. In this way, the broken road can be restored to some extent without changing the original extraction results, making the morphology of roads more complete.

(3) Schematic diagram of post-processing

According to the above method, the flow of post-processing is shown in Figure 8. Firstly, the connected domains are labeled from the network extraction results, as shown in Figure 8b. Secondly, the error extracted spots are removed using the area threshold and the broken road lines are reconstructed according to the minimum distance of the convex envelopes, as shown in Figure 8c,d. Finally, the results obtained by the post-processing method in this paper are shown in Figure 8e.

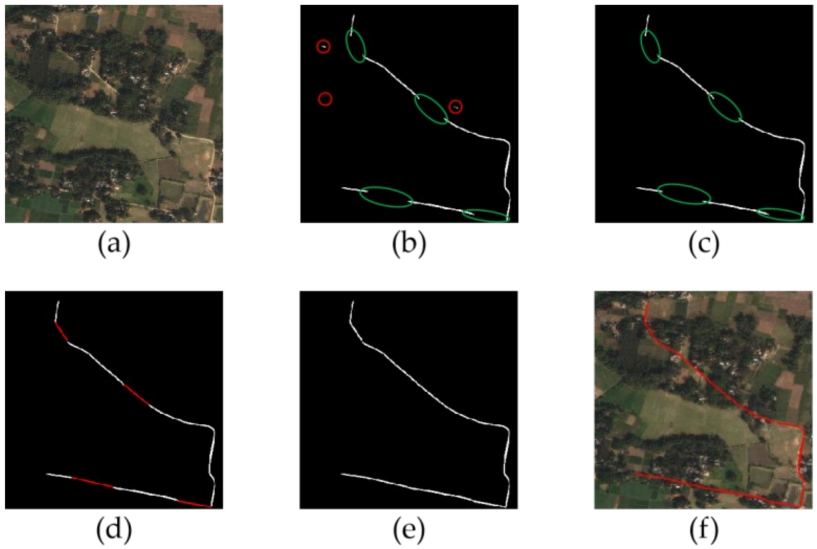

Figure 8. The flow of post-processing. (**a**) Original image; (**b**) Detect interference and disconnection; (**c**) Delete error extracted spots; (**d**) Connect disconnected road lines; (**e**) The result of post-processing; (**f**) The morphology of roads.

3. Dataset Descriptions and Experimental Configuration

This section describes the dataset, experimental conditions and evaluation indicators of the experiments in this paper in detail.

3.1. Dataset Descriptions

In this paper, the DeepGlobe Road Extraction dataset is utilized to carry out the experiment [58]. It contains 6226 high-resolution remote sensing images with pixel-level road labels with the size of 1024 × 1024. The resolution of the images is 0.5 m^2/pixel, as shown in Figure 9.

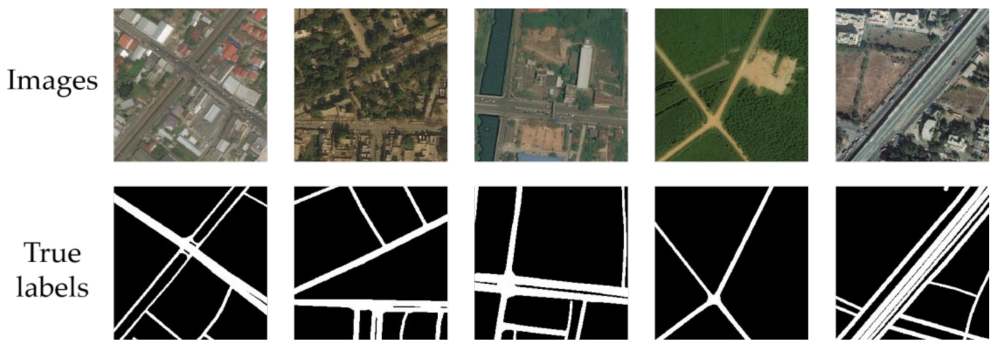

Figure 9. Examples of images from the dataset and their corresponding labels.

3.2. Mask Based Data Enhancement against Occluded Information

Preliminary experiments reveal that the effectiveness of the network is greatly reduced by the occlusion. As such, this paper utilizes an active data enhancement method to enhance the network's learning ability of occluded images. First, the method generates a series of discontinuous regular patches at a certain size, and then simulates the real occlusion situation by controlling the parameters of the patches. As shown in Equation (4):

$$\widetilde{x} = x \times M(\text{distance}, \text{ratio}, \text{mode}) \quad (4)$$

where x represents the input image; M is the mask for the pixels to be deleted and is determined by three parameters for the grid distance (d), occlusion ratio and occlusion mode (if occlude or not), as shown in Figure 10; and $\widetilde{x}$ is the output image with mask.

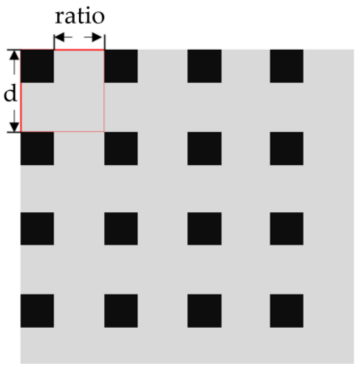

Figure 10. Demonstration of data augment.

Moreover, considering the performance of the GPU, the original images are cut to the size of 512 × 512 to improve the efficiency of model training. Further, the data is also augmented by vertical flip, horizontal flip, clockwise rotation of 90 and random color dithering in the HSV space to prevent overfitting. Finally, a dataset containing 14,400 training images, 4800 validation images and 4800 test images is obtained.

3.3. Experimental Configuration

3.3.1. Training Environment Description

The size of images for all experiments in this paper is 3 × 512 × 512. All training and tests are implemented by using PyTorch on the Windows 10 platform with Tesla V100 16 GB GPU.

3.3.2. Hyper-Parameter Settings

In this paper, the Adam optimizer is selected for training, whose initial learning rate is set to 2×10^{-4}, and the learning rate will adjust adaptively during training. The value of Val loss indicates that the learning rate is reduced by 0.7 after five consecutive epochs in which the performance does not improve. The Batch size is set to 16.

As for the loss function, because of the small proportion of roads in remote sensing images, there is an imbalance between positive and negative samples, and thus the model will tend to predict negative samples and the performance of it decreases. Thus, this paper introduced a category balance factor w into the BCE (Binary Cross Entropy) loss function, as shown in Equation (5), which can strengthen the attention given to the positive samples and reduce the attention given to the negative samples by changing the weight of samples.

$$L = -\frac{1}{mn}\sum_{i=1}^{mn}[wy_i log p_i + (1-w)(1-y_i)log(1-p_i)] \quad (5)$$

where L presents the value of average loss function; w is set to 0.4; y_i = {0,1} and $p_i \in (0,1)$ represent the true value of the pixel category and the prediction probability value of the pixel class, respectively; $m \times n$ is the sample size of the current batch; m is 512 × 512 pixels; and n is the number of input images.

To prevent the network from overfitting, the training needs to be stopped at an appropriate point. A common method is to introduce a validation set in the network training process to detect the network state where the loss function of the validation set is reduced to a certain range or successive multiple rounds are not optimized. In this paper, IOU is added as an additional condition for the training stop to intuitively connect the loss function with the road extraction results. As such, the training will stop when the

loss function value on the validation set does not drop for ten consecutive rounds and, meanwhile, the IOU value does not rise.

3.4. Evaluation Metrics

To objectively evaluate the results of the experiments in this paper, the quantitative evaluation criteria are focused on IOU (Equation (6)). It can take the impact of both error detection and omission detection into account, and has been the standard of semantic segmentation. Further, commonly used criteria including accuracy (Equation (7)), precision (Equation (8)) and recall (Equation (8)) are also considered. In addition, F1-score (Equation (9)), which combines two metrics of precision (P) and recall (R) to reflect the performance of the model, is also calculated in this paper.

$$IOU = \frac{TP}{TP + FP + FN} \tag{6}$$

$$Accuracy = \frac{TP + FP}{TP + FP + TF + TN} \tag{7}$$

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{9}$$

where *TP*, *FN*, *FP* and *TN* are pixel results classified by comparing the extracted shadow pixels with ground-truth points.

TP: true positives, i.e., the correct extraction of road pixels;
FN: false negatives, i.e., the omissive extraction of road pixels;
FP: false positives, i.e., the erroneous extraction of road pixels;
TN: true negatives, i.e., the correct extraction of non-road pixels.

4. Experimental Results and Discussion

Due to the stringent stopping conditions, a stable model is obtained after training for nearly 500 epochs. The loss and accuracy on the validation dataset are shown in Figure 11. This section mainly displays the test results in detail, and the proposed method is compared with several other common methods, including U-Net, D-Linknet, FCN and SegNet. Then, the proposed post-processing method is conducted.

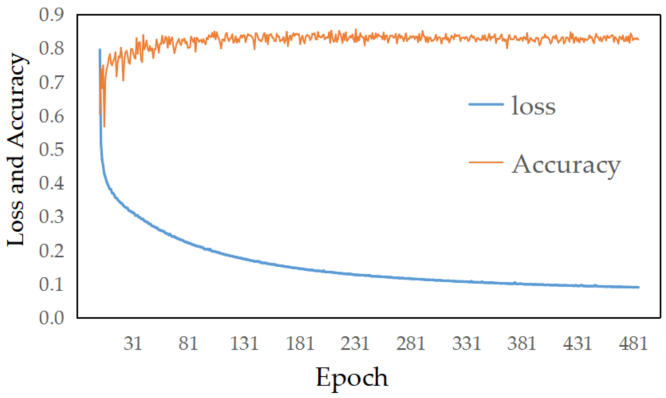

Figure 11. The loss and accuracy of the validation dataset during the training.

4.1. Qualitative Analysis

To fully reflect the advantages and disadvantages of the network in this paper, we select several representative images for comparison in consideration of road density, background information and road types, as shown in Figure 12, in which the increasing density

and complexity of the road from (a) to (d) rows and the increasing similar spectral features of the road and background make the extraction of the roads more difficult.

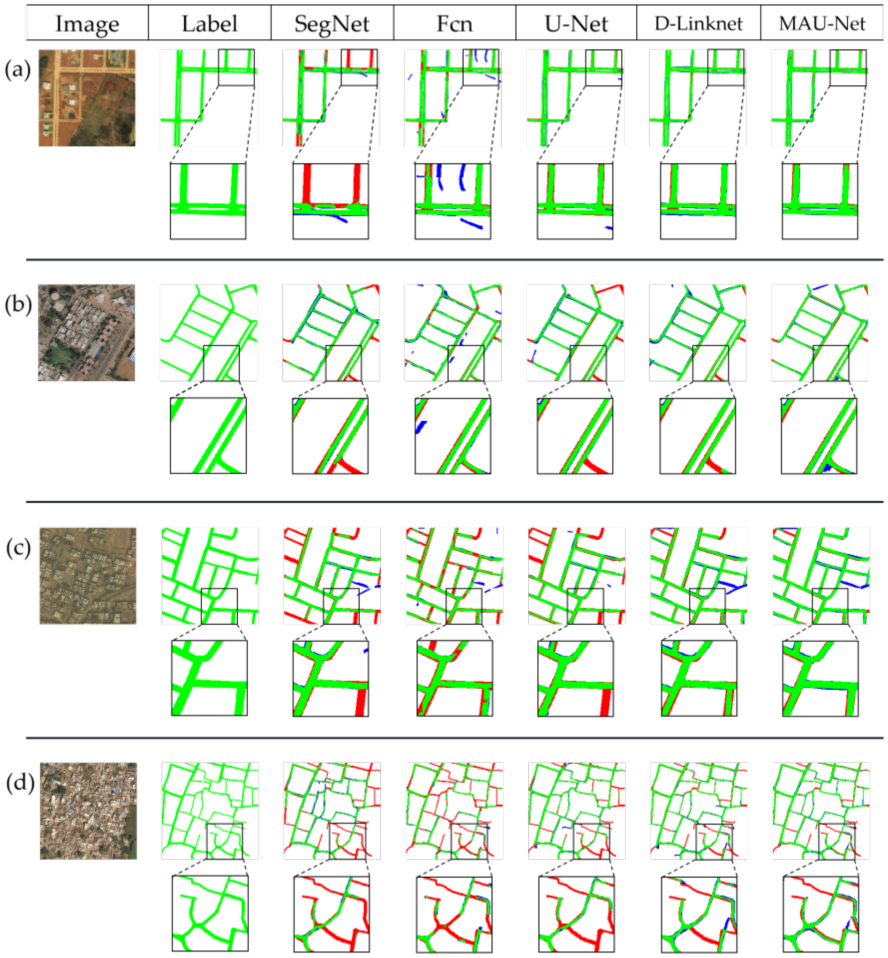

Figure 12. The visual comparison of different road extraction methods. (**a**) Sparse rural roads with obvious characteristics; (**b**) Sparse urban roads with obvious features; (**c**) Complex rural roads without obvious features; (**d**) Complex urban roads without obvious features.

From rows (a) and (b) of the visualization result, most methods effectively extract road information when road features are prominent. However, D-Linknet and our method still have obvious advantages. Their results show higher integrity and accuracy. There are almost no error or omission extractions. Further, they performed better in the details, which can be seen from the edge and the middle barrier of the roads.

With the increase of the density and complexity of roads, the extraction effect of each method decreases significantly, mainly manifested in error extraction, omission extraction and the change of road shape. As shown in the (d) row, both the SegNet, FCN and U-Net methods missed substantial road information, and the D-Linknet and MAU-Net methods still performed better. The (c) row shows similar results. The background and the road in Figure 12c are similar in spectral features which causes networks to tend to extract the background as road. In this respect, the method in this paper performs better than D-Linknet, with less erroneous extractions.

Moreover, experiments also show that when the background interference is strong and the scale of the road is large, some methods tend to change the shape of the road, which in turn affects the reliability of the extraction results and increases the difficulty of post-processing. As shown in Figure 12a,c, the roads extracted by SegNet and FCN are significantly thinner than true labels and there are many fragmented extract results, indicating that these two methods are insufficient in dealing with road edge information.

U-Net and D-Linknet have improved in this respect, with almost no road fragmentation, and the present method performed best, not only with no road fragmentation but also demonstrating a more complete extraction of road morphology. There is no phenomenon of road edge depression in the extraction results of other networks.

The mean IOU, mean accuracy and mean F1-score of these images under different methods are also presented in Figure 13, respectively. The chart also demonstrates the effectiveness of our method. Each indicator has a significant improvement.

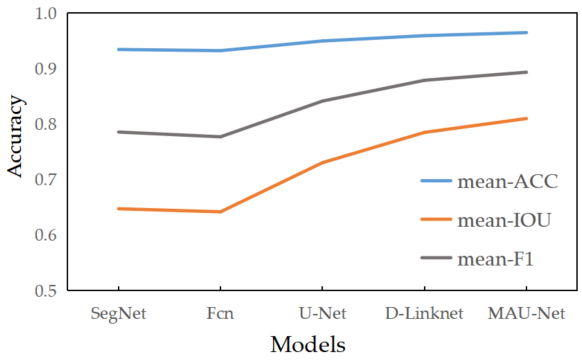

Figure 13. The performance line chart of different road extraction methods.

4.2. Quantitative Analysis

Quantitative analysis is an effective method of making an experiment more accurate and reliable. In this paper, various accuracy indicators of the results extracted by several methods are calculated, and a comparison is made to show the advantages of the proposed method. The results are shown in Table 1. It can be seen that the accuracy and IOU of our method are both higher than that of other methods, especially the IOU, which has increased by 3.35–12.8%, which demonstrates the superiority of this method. In addition, the precision does not seem to improve much due to the restriction of recall; however, the balance level, which is presented on the F1 score, is a more important factor of the model. The proposed method increases the F1 score by 2.41–9.8. As such, the method in this paper is synthetically concluded to perform best in this experiment.

Table 1. Comparison among the results of different methods.

Methods	Accuracy (%)	P (%)	R (%)	F1 (%)	IOU (%)
SegNet	97.61	83.41	64.05	72.46	56.81
FCN	97.32	75.10	67.80	71.26	55.36
U-Net	97.76	83.22	67.99	74.84	59.79
D-Linknet	97.94	80.06	77.29	78.65	64.81
MAU-Net	98.14	80.97	81.15	81.06	68.16

4.3. Post-Processing Analysis

For the common problems of spots caused by error extraction and road breaks caused by occlusion in the extraction results, this paper post-processed them based on connected domains. As it is visualized in Figure 14, the proposed post-processing method has effectively used the area feature of the connected domain and most error spots are removed, so the noise in the results is significantly reduced and the results are more concise. Additionally, the distance feature of the connected domain is also fully utilized to reconstruct road lines. The topological characteristics of the road can be recovered, which is one of the main information components of the road.

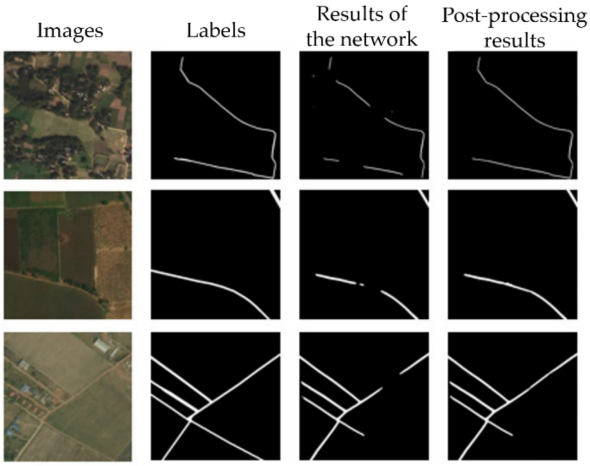

Figure 14. The visual result of post-processing.

Finally, a quantitative analysis is performed to verify the effect of the post-processing method. Several defective network extraction results are selected and post-processed. The Accuracy, IOU and F1-score of them are calculated in Table 2. These evaluation metrics are all seen to be improved, which demonstrates the effectiveness of the post-processing method in this paper.

Table 2. Evaluation of post-processing.

	Accuracy (%)	IOU (%)	F1 (%)
Results of the network	98.68	61.53	75.95
Post-processing results	98.83	69.69	82.11

4.4. Discussion of the Proposed Method

In conclusion, the use of attention mechanisms and mask-based dataset augmentation in this paper effectively improves the extraction ability of the network. A relatively simple and complete post-processing process is established, which makes the extraction effect of this method further improved. The proposed method achieves good results in remote sensing images in different scenes, and has a better performance on occluded roads.

However, there are also some deficiencies in the present approach, as shown in Figure 15. The proposed method performs poorly on some rural roads. This may be caused by the special road shape and the high similarity of pixel information between the background and roads.

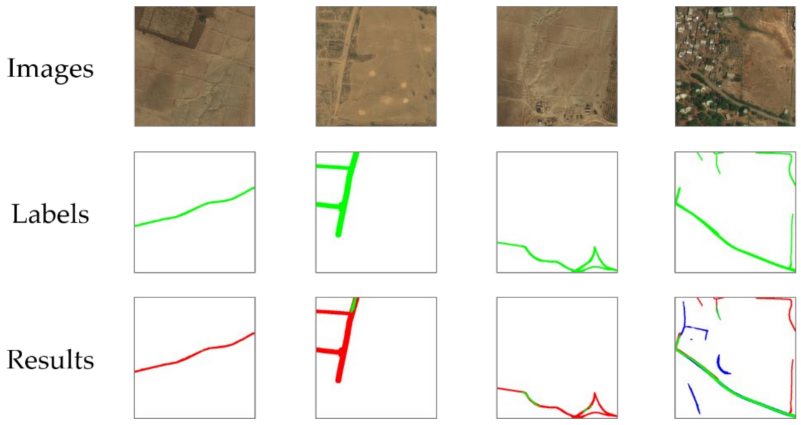

Figure 15. The visualization of some deficient extraction results.

5. Conclusions and Future Works

To improve the accuracy and efficiency of road information extraction from remote sensing images, a convolution neural network-integrated attention mechanism is established in this paper, and a method based on the connected domain is proposed to optimize the extraction results. Firstly, considering the road features in remote sensing images, a module combining spatial attention and channel attention is added to the decoding process, which effectively uses the context information in the process of network transmission, improves the expression of road features and suppresses the interference of background. Secondly, the traditional feature fusion method based on channel contact is replaced by the attention score weighted algorithm, which fully uses the feature map of each module in the network and promotes the extraction ability of the network and the model's attention to key areas. Finally, in terms of the output, considering the difference of image scale, this paper establishes a deep multi-scale aggregation structure, which uses all the information in the decoding process, so as to preliminarily optimize the output of the network and obtain output that is more fit to the scale of the image. In addition, according to the extraction results of the network and the topological characteristics of the road itself, a connected domain-based road post-processing method is proposed, which makes the extraction results more concise and makes the road structure more complete. Further, the experiment on the public dataset shows that our method performed best when compared with the four common road extraction networks of FCN, SegNet, U-Net and D-Linknet, which can be seen from the evaluation indexes and the visualization results. There are fewer error extractions and the morphology of the road is more complete. After that, the post-processing method in this paper is applied to the extraction results of the network, and its effectiveness is proved from both qualitative and quantitative aspects.

Although the proposed method obviously improves the accuracy and effect of road extraction over other methods, there are still some deficiencies. In the future, we will further optimize our network to strengthen its feature extraction ability and improve training efficiency, and we will search for a more appropriate threshold adjustment mode to improve the effectiveness of the proposed post-processing method. In addition, we are investigating the possibility of applying this semantic segmentation method to handle other remote sensing tasks, such as shadow extraction.

Author Contributions: Conceptualization, D.F., X.S. and Y.X.; data curation, D.F., X.S. and Y.L.; formal analysis, X.S.; funding acquisition, D.F.; investigation, D.F.; methodology, D.F., X.S. and Y.X.; project administration, D.F.; software, J.W. and Y.L.; supervision, D.F.; validation, J.W. and Y.L.; visualization, J.W.; writing—original draft, X.S. and Y.X.; writing—review and editing, D.F., X.S. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by the National Natural Science Foundation of China (Grant Nos. U2034202 and 41871289), the Sichuan Youth Science and Technology Innovation Team (Grant No. 2020JDTD0003).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in [DeepGlobe Road Extraction Dataset] at [https://doi.org/10.1109/CVPRW.2018.00031], reference number [58].

Acknowledgments: All authors would sincerely thank the reviewers and editors for their beneficial, careful, and detailed comments and suggestions for improving the paper.

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

References

1. Liu, Z.; Feng, R.; Wang, L.; Zhong, Y.; Cao, L. D-Resunet: Resunet and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3927–3930. [CrossRef]
2. Zhang, X.; Ma, W.; Li, C.; Wu, J.; Tang, X.; Jiao, L. Fully convolutional network-based ensemble method for road extraction from aerial images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1777–1781. [CrossRef]
3. Li, D.; Deng, L.; Cai, Z.; Yao, X. Intelligent transportation system in Macao based on deep self-coding learning. *IEEE Trans. Ind. Inf.* **2018**, *14*, 3253–3260. [CrossRef]
4. Chen, Z.; Wang, C.; Li, J.; Xie, N.; Han, Y.; Du, J. Reconstruction bias U-Net for road extraction from optical remote sensing images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 2284–2294. [CrossRef]
5. Abdollahi, A.; Pradhan, B. Integrated technique of segmentation and classification methods with connected components analysis for road extraction from orthophoto images. *Expert Syst. Appl.* **2021**, *176*, 114908. [CrossRef]
6. Chen, Z.; Fan, W.; Zhong, B.; Li, J.; Du, J.; Wang, C. Corse-to-fifine road extraction based on local Dirichlet mixture models and multiscale-high order deep learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4283–4293. [CrossRef]
7. Soni, P.K.; Rajpal, N.; Mehta, R. Semiautomatic road extraction framework based on shape features and LS-SVM from high-resolution images. *J. Indian Soc. Remote Sens.* **2020**, *48*, 513–524. [CrossRef]
8. Cheng, J.; Liu, T.; Zhou, Y.; Xiong, Y. Road junction identification in high resolution urban SAR images based on SVM. In *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*; Barolli, L., Xhafa, F., Hussain, O., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 994, pp. 597–606. [CrossRef]
9. Xu, Y.; Xie, Z.; Feng, Y.; Chen, Z. Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sens.* **2018**, *10*, 1461. [CrossRef]
10. Shao, Z.; Zhou, Z.; Huang, X.; Zhang, Y. MRENet: Simultaneous extraction of road surface and road centerline in complex urban scenes from very high-resolution images. *Remote Sens.* **2021**, *13*, 239. [CrossRef]
11. Luo, Q.; Yin, Q.; Kuang, D. Research on Extracting Road Based on Its Spectral Feature and Shape Feature. *Remote Sens. Technol. Appl.* **2011**, *22*, 339–344. [CrossRef]
12. Wang, J.; Qin, Q.; Yang, X.; Wang, J.; Ye, X.; Qin, X. Automated road extraction from multi-resolution images using spectral information and texture. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 533–536. [CrossRef]
13. Barzohar, M.; Cooper, D.B. Automatic finding main roads in aerial image by using geometric-stochastic models and estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 707–721. [CrossRef]
14. Hu, J.; Razdan, A.; Femiani, J.C.; Cui, M.; Wonka, P. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4144–4157. [CrossRef]
15. Shi, W.; Miao, Z.; Debayle, J. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3359–3372. [CrossRef]
16. Mu, H.; Zhang, Y.; Li, H.; Guo, Y.; Zhuang, Y. Road extraction base on Zernike algorithm on SAR image. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 1274–1277. [CrossRef]
17. Ghaziani, M.; Mohamadi, Y.; Bugra Koku, A.; Konukseven, E.I. Extraction of unstructured roads from satellite images using binary image segmentation. In Proceedings of the 2013 21st Signal Processing and Communications Applications Conference, Haspolat, Turkey, 24–26 April 2013; pp. 1–4. [CrossRef]
18. Ma, H.; Cheng, X.; Wang, X.; Yuan, J. Road information extraction from high resolution remote sensing images based on threshold segmentation and mathematical morphology. In Proceedings of the 2013 6th International Congress on Image and Signal Processing, Hangzhou, China, 16–18 December 2013; pp. 626–630. [CrossRef]
19. Shanmugam, L.; Kaliaperumal, V. Junction-aware water flow approach for urban road network extraction. *IET Image Process.* **2016**, *10*, 227–234. [CrossRef]
20. Ma, S.; Xu, Y.; Yue, W. Numerical solutions of a variable-order fractional financial system. *J. Appl. Math.* **2012**, *2012*, 1–16. [CrossRef]
21. Sirmacek, B.; Unsalan, C. Road network extraction using edge detection and spatial voting. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3113–3116. [CrossRef]
22. Gaetano, R.; Zerubia, J.; Scarpa, G.; Poggi, G. Morphological road segmentation in urban areas from high resolution satellite images. In Proceedings of the 2011 17th International Conference on Digital Signal Processing, Corfu, Greece, 6–8 July 2011; pp. 1–8. [CrossRef]
23. Anil, P.N.; Natarajan, S. Road extraction using topological derivative and mathematical morphology. *J. Indian Soc. Remote Sens.* **2013**, *41*, 719–724. [CrossRef]
24. Wang, Z.; Liu, X.; Liu, L.; Shi, J. A method of road extraction from high resolution remote image based on delaunay algorithms. In Proceedings of the 2018 International Conference on Robots & Intelligent System, Changsha, China, 26–27 May 2018; pp. 127–130. [CrossRef]
25. Yager, N.; Sowmya, A. Support vector machines for road extraction from remotely sensed images. In Proceedings of the Computer Analysis of Images and Patterns: 10th International Conference, Groningen, The Netherlands, 25–27 August 2003; pp. 285–292. [CrossRef]

26. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Queiroz Feitosa, R.; van der Meer, F.; van der Werff, H.; van Coillie, F.; et al. Geographic object-based image analysis–towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [CrossRef] [PubMed]
27. Kumar, M.; Singh, R.K.; Raju, P.L.N.; Krishnamurthy, Y.V.N. Road network extraction from high resolution multispectral satellite imagery based on object oriented techniques. *Remote Sens. Spat. Inf. Sci.* **2014**, *2*, 107–110. [CrossRef]
28. Herumurti, D.; Uchimura, K.; Koutaki, G.; Uemura, T. Urban road extraction based on hough transform and region growing. In Proceedings of the 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, Incheon, Korea, 30 January–1 February 2013; pp. 220–224. [CrossRef]
29. Lu, P.; Du, K.; Yu, W.; Wang, R.; Deng, Y.; Balz, T. A new region growing-based method for road network extraction and its application on different resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 4772–4783. [CrossRef]
30. Miao, Z.; Wang, B.; Shi, W.; Zhang, H. A semi-automatic method for road centerline extraction from VHR images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1856–1860. [CrossRef]
31. Li, M.; Stein, A.; Bijker, W.; Zhan, Q. Region-based urban road extraction from VHR satellite images using binary partition tree. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *44*, 217–225. [CrossRef]
32. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
34. Luus, F.P.S.; Salmon, B.P.; Van Den Bergh, F.; Maharaj, B.T.J. Multiview deep learning for land-use classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2448–2452. [CrossRef]
35. Huang, G.; Liu, Z.; Laurens, V.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
36. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
37. 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.* **2018**, *40*, 834–848. [CrossRef]
38. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards realtime object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
39. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *Comput. Sci.* **2017**. Available online: https://arxiv.org/abs/1706.05587v3 (accessed on 29 October 2021).
40. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lect. Notes Comput. Sci.* **2018**, *11211*, 833–851. [CrossRef]
41. Mnih, V.; Hinton, G.E. Learning to detect roads in high-resolution aerial images. In Proceedings of the 11th European Conference on Computer Vision, Part VI. Heraklion, Crete, Greece, 5–11 September 2010; pp. 210–223. [CrossRef]
42. Gao, J.; Qi, W.; 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. [CrossRef]
43. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [CrossRef]
44. Gao, X.; Sun, X.; Zhang, Y.; Yan, M.; Xu, G.; Sun, H.; Jiao, J.; Fu, K. An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. *IEEE Access* **2018**, *6*, 39401–39414. [CrossRef]
45. Wei, Y.; Wang, Z.; Xu, M. Road structure refined CNN for road extraction in aerial image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 709–713. [CrossRef]
46. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for effificient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing, St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [CrossRef]
47. Zhou, L.; Zhang, C.; Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186. [CrossRef]
48. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
49. Wang, Z.; Yan, H.; Lu, X. High-resolution remote sensing image road extraction method for improving U-Net. *Remote Sens. Technol. Appl.* **2020**, *35*, 741–748.
50. 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, 27–30 June 2016; pp. 770–778. [CrossRef]
51. Yang, X.; Li, X.; Ye, Y.; Lau, R.Y.K.; Zhang, X.F.; Huang, X.H. Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *59*, 7209–7220. [CrossRef]
52. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]

53. Sofla, R.A.D.; Alipour-Fard, T.; Arefi, H. Road extraction from satellite and aerial image using SE-Unet. *J. Appl. Remote Sens.* **2021**, *15*, 014512. [CrossRef]
54. Ren, Y.; Yu, Y.; Guan, H. DA-CapsUNet: A dual-attention capsule U-Net for road extraction from remote sensing imagery. *Remote Sens.* **2020**, *12*, 2866. [CrossRef]
55. Lin, Y.; Wan, L.; Zhang, H.; Wei, S.; Ma, P.; Li, Y.; Zhao, Z. Leveraging optical and SAR data with a UU-Net for large-scale road extraction. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102498. [CrossRef]
56. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. *Lect. Notes Comput. Sci.* **2018**, *11211*, 3–19. [CrossRef]
57. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]
58. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–181. [CrossRef]