Bi-HRNet: A Road Extraction Framework from Satellite Imagery Based on Node Heatmap and Bidirectional Connectivity

Ziyun Wu 1, Jinming Zhang 2,3,*, Lili Zhang 2,3, Xiongfei Liu 2,3 and Hailang Qiao 2,3

1 School of Remote Sensing and Information Engineering, 129 Luoyu Roud, Wuhan University, Wuhan 430079, China; 2019282130140@whu.edu.cn
2 Key Laboratory of Network Information System Technology, Institute of Electronic, Chinese Academy of Sciences, Beijing 100190, China; zhangll002569@aircas.ac.cn (L.Z.); liuxf004327@aircas.ac.cn (X.L.); qiaohl@aircas.ac.cn (H.Q.)
3 The Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: niyzjm@whu.edu.cn

Abstract: Today, with the rapid development of the geographic information industry, automatic road extraction from satellite imagery is a basic requirement. Most existing methods have been designed based on binary segmentation. However, these methods do not consider the topological features of road networks, which include point, edge, and direction. In this study, a topology-based multi-task convolution network is designed, namely Bi-HRNet, which can effectively learn the key features of nodes and their directions. First, the proposed network learns the node heatmap of roads, and then the pixel coordinates are extracted from the node heatmap via non-maximum suppression (NMS). At the same time, the connectivity between nodes is predicted. To improve the integrity and accuracy of connectivity, we propose a bidirectional connectivity prediction strategy, which can learn the bidirectional categories instead of direction angles. The bidirectional categories are designed based on “top-to-bottom” and “bottom-to-top” strategies, which can improve the accuracy of the connectivity between nodes. To illustrate the effectiveness of the proposed Bi-HRNet, we compare our method with several methods on different datasets. The experiments show that our method achieves a state-of-the-art performance and significantly outperforms various previous methods.

Keywords: road extraction; multi-task learning; node heatmap; bidirectional connectivity

1. Introduction

Automatic road extraction technology is of great significance for national condition monitoring. Accurate and complete road network data can provide convenience for map navigation and national mapping production. With the development of high-resolution satellite technology in recent years, high-resolution images have a relatively high spatial resolution, and they display more abundant surface information; therefore, the images provide more detailed information. High-resolution images can usually clearly show the structure of roads and lane lines, as well as vehicles and pedestrians on the roads. Although high-resolution image data provide rich and high precision information, they also include irrelevant interference noise around the roads. For example, green vegetation on both sides of multi-lane highways and shadows of buildings become obstacles to the automatic extraction of roads. Therefore, it is particularly important to use deep learning technology in artificial intelligence to assist with automatic road extraction in high-resolution images.

According to different standards, the classification of road extraction methods is diverse. On the basis of different object characteristics of roads, these methods can be subdivided into pixel-based [1], area-based, and knowledge-based [2] methods. There are many classic algorithms for road extraction based on linear features, which include several different algorithms, such as clustering [3,4], classification, active contour mode [5], dynamic programming [6], and Hough transforming [7,8]. Based on the output form, they are
also divided into road segmentation [9], centerline extraction [10], sideline extraction [11], road classification [12], and so on.

1.1. Heuristic Road Extraction Algorithm

According to the degree of interaction, heuristic road extraction algorithms can be divided into two categories, i.e., semi-automatic and automated. Using traditional high-resolution remote sensing satellite images to manually update road features has low efficiency, long cycle, and high cost. The semi-automatic extraction algorithm is offered as an improvement, in which it is necessary to preplace, after human judgment, some of the seed points, which are usually the road center point or the road contour line, as prior knowledge of road extraction, and use the prior road area to extract road features. The method based on an active contour model proposed by [13–15] for road network extraction can also be called the snake algorithm. This method generates an energy function at the contour points provided by the user, and then the algorithm obtains the minimum value of the energy function by adjusting the contour to achieve the final convergence. Ideally, the energy function can fit the edge of the road surface at its iterative minimum. However, the energy equation relies too much on the parameterization of the curve equation in the calculation process, and therefore it cannot handle the topological changes of curved roads. Among them, Ref. [16,17] proposed a sliding matching window of a specific size to search for a series of matching points based on some spectral features of the road, and then connected these matching points to a template matching method. However, the template matching method for road extraction was affected by scenes without road morphology exposure where some road occlusion situations were encountered, because no matching seed points could be found.

Semi-automatic algorithms still require human intervention. In order to further improve the efficiency of automated work, some road extraction algorithms can automatically extract roads in specific scenarios and achieve a degree of satisfactory results. Ref. [18] used support vector machines to solve the problem of road extraction in remote sensing images based on edge features, and performed nonlinear classification through a kernel binary classifier, but the accuracy was relatively low. Ref. [19] only used support vector machines to divide the images into two categories: roads and non-roads. Ref. [20] proposed a road extraction algorithm through SVM and level-setting strategy. Although the SVM classifier had good generalization ability and minimized the cost of structure loss, it had certain difficulties in selecting the kernel function and was more sensitive to training samples. Ref. [21] described a Bayesian framework classification method based on multiscale features, which was achieved by an iterative conditional pattern algorithm. The naive Bayes classifier showed better robustness than SVM, and SVM was relatively susceptible to noise samples. Ref. [22] used a region adaptive segmentation to automatically find the initial marker. Inspired by their method, [23] proposed a fast automatic road extraction algorithm. The algorithm first used the technique proposed in [22] to collect road areas, and then used linear features [24] to extract linear features. Finally, the algorithm could automatically identify road objects by fusing regional features, such as linear features and road prior information.

1.2. Deep-Learning-Based Road Automatic Extraction Algorithm

Early road network extraction methods mostly relied on the prior knowledge of road linear geometry or the spectral features of remote sensing images to optimize complex objective functions [25,26]; however, most of these have been replaced by deep learning technology. One deep learning method is to take the image as input to a fully connected neural network [27]; however, due to memory limitations, only limited content information can be used.

The emergence of convolutional neural networks (CNNs) has increased the range of receptive fields and greatly improved the accuracy of road extraction. Ref. [28] proposed a back-propagation neural network for road extraction. Ref. [29] found the best network struc-
ture to extract the road structure by designing networks with different hidden layer sizes, and trained the network in different periods. Ref. [30] mentioned that multi-hidden-layer neural networks had excellent feature representation capabilities, and the difficulty of training deep neural networks could be effectively overcome by “layer-by-layer initialization”.

Ref. [31] first attempted to build a GPU-based deep convolutional neural network DCNN (deep convolutional neural network, DCNN), which used spatial context information in remote sensing images to learn distinguishing features. In order to effectively utilize the correlation between neighboring pixels, DCNN also used a larger image as input, predicted a small portion of tags from the same context, and predicted the road probability of neighboring pixels. This operation improved the classification accuracy to a certain extent and reduced the calculation cost. Ref. [32] designed a deep convolutional neural network based on a single small block program to simultaneously extract roads and buildings from high-resolution remote sensing images (HRSI), and then used post-processing to improve the accuracy of the extracted roads. Ref. [33] used full convolutional networks (FCN) to extract roads and buildings in remote sensing images; however, the upsampling operation in FCN made the prediction noise increase. Ref. [34] designed SegNet to avoid excessive upsampling of network layers while limiting the number of pooling layer networks to reduce spatial context information. DeconvNets is a variant network structure in which FCN replaces the interpolation layer as the deconvolution layer, similar to SegNet, DeepLab [35], and U-Net [36]. The decoder maps the low-resolution features output by the encoder stack to the full input image size feature map. The authors of [37] used a network structure based on VGG to extract roads using proposed road cross entropy loss. The study [38] proposed an enhanced deep convolutional network based on convolutional linear units, which used SegNet as the backbone to segment aerial images. The study [39] designed a variant network based on FCN, which used part of the ResNet structure as an encoder with a complete deconvolution decoder to directly extract road topological features from remote sensing images. The study [40] introduced an iterative refinement method based on U-Net to extract topological relationships. Considering that the pixel loss is not suitable to reflect the topological influence of the prediction error, Ref. [41] proposed an end-to-end framework similar to the RSR-CNN multi-feature pyramid network. They took advantage of the multi-level semantic features of HRSI and designed a novel loss function to focus on the problem of category imbalance. Inspired by DenseNet [42] and U-Net, Ref. [43] proposed a GL-Dense-U-Net for extracting roads from aerial images. The study [44] combined dilated convolution [45] and Linknet [46] to expand the acceptance domain for extracting roads from high-resolution satellite images.

The segmentation method is sensitive to background noise. In order to eliminate background noise, the threshold in the post-processing method is used to binarize the road segmentation, and then the morphological refinement technology is applied to obtain a single-pixel wide road skeleton. In order to eliminate the redundancy of the graph, DeepRoadMapper used a lightweight CNN network with softmax loss in the first process to generate segmentation output. Ref. [47] introduced directed learning and deletion-refinement learning. Directional learning gives neural networks the ability to process the connections between pixels. In addition, deletion-refinement learning can learn the pattern of road connection and optimize the road segmentation output from the first step. The obtained road network had good connectivity on the employed average path length similarity (APLS) metric [48].

Ref. [49] used iterative exploration algorithms to directly generate road maps. Ref. [50] used polygons to adapt to the shape of roads and buildings. However, most existing deep learning network models have yielded discontinuous and incomplete results because of shadows and occlusions. To address this problem, a dual-attention road extraction network (DA-RoadNet) [51] with a certain semantic reasoning ability was proposed. Kai Zhou et al. [52] proposed a novel fusion network (FuNet) with fusion of remote sensing imagery and location data, which played an important role of location data in road connectivity reasoning. To increase the accuracy of road extraction from high-resolution remote
sensing images, [53] proposed a split depthwise (DW) separable graph convolutional network (SGCN). To improve the accuracy and connectivity of road extraction, [54] proposed an inner convolution integrated encoder–decoder network with the post-processing of directional conditional random fields. Motivated by the road shapes and connections in the graph network, Ref. [55] proposed a connectivity attention network (CoANet) to jointly learn the segmentation and pairwise dependencies. Z. Sun et al. [56] proposed a weak roads extraction approach under strong speckle interference based on shearlet, which can overcome the interference of speckle and completely detect road information.

The segmentation-based and graph-based methods both have obvious loopholes. The segmentation-based method suffers from small-scale topology errors due to lack of connectivity. Although the graph-based method has no obvious topological errors, it is prone to error propagation due to the iterative reconstruction strategy.

To solve the problem, we propose a road extraction convolutional neural network from satellite imagery based on node heatmap and bidirectional connectivity. First, a convolutional neural network (CNN) based on HRNet is used to learn the node heatmap and bidirectional categories of the road. The pixel coordinates of nodes can be obtained via the NMS algorithm, and the bidirectional connectivity between nodes can be obtained. The bidirectional connectivity prediction is based on a “top-to-down” and “down-to-top” strategy. In the feature space between the corresponding nodes on the bidirectional connectivity categories, the angle prediction is replaced by the angle classification, making it possible to train a simple, supervised node that predicts the key nodes and their bidirectional connectivity. Experiments on DeepGlobe, RoadTracer, and Google datasets demonstrate that our method outperforms other methods and achieves a state-of-the-art performance.

We explicitly state our original contributions as follows:

1. We propose a new way of predicting the direction of road networks, which classifies the importance of road topology connectivity according to different road nodes and converts the regression of the direction and angle into a regional classification problem to enhance the network on direction learning.

2. To improve the accuracy of node connectivity prediction, we propose a bidirectional connectivity prediction strategy, which is based on a “top-to-down” and “down-to-top” strategy.

3. We propose a framework for predicting key points of road networks based on a multiresolution road node heatmap, which can improve the precision of key nodes.

2. Materials and Methods

The workflow of the proposed road extraction method is shown in Figure 1. As shown in Figure 1, we first input the remote sensing image into the proposed Bi-HRNet, and obtain the node heatmap, “top-to-down road direction” and “down-to-top road direction”. Then, the pixel coordinates of nodes can be obtained from node heatmap via NMS method, and the connectivity between each node can be obtained according to the bidirectional connectivity map. Finally, the nodes can be connected according to the obtained connectivity map, and the final extracted road map can be obtained.

2.1. Overview of the Proposed Framework

In this section, we describe the details of the proposed network. As shown in Figure 2, the proposed Bi-HRNet contains two stages, i.e., road direction prediction and road node prediction. In the road direction prediction stage, we propose a “top-to-down” and “down-to-top” road direction prediction strategy, which constructs a bidirectional connection between two nodes. To simplify the complexity of road direction prediction by calculating angle, we convert angle regression to a classification problem. In the road node prediction stage, we propose a multiscale node heatmap prediction strategy. The small-scale prediction branch helps to improve the accuracy of normal scale prediction of road nodes. Finally, the multi-task learning strategy with multi-loss can better enhance the performance of road extraction on satellite imagery.
2.1. Overview of the Proposed Framework

In this section, we describe the details of the proposed network. As shown in Figure 2, the proposed Bi-HRNet contains two stages, i.e., road direction prediction and road node prediction. In the road direction prediction stage, we propose a “top-to-down” and “down-to-top” road direction prediction strategy, which constructs a bidirectional connection between two nodes. To simplify the complexity of road direction prediction by calculating angle, we convert angle regression to a classification problem. In the road node prediction stage, we propose a multiscale node heatmap prediction strategy. The small-scale prediction branch helps to improve the accuracy of normal scale prediction of road nodes. Finally, the multi-task learning strategy with multi-loss can better enhance the performance of road extraction on satellite imagery.

2.2. Bidirectional Road Graph Prediction

In satellite imagery, each road is connected by two nodes. Thus, each pair of nodes has two-directional information. Therefore, we propose a bidirectional road connectivity prediction strategy to predict the direction between two nodes.

Figure 3 shows the definition of the proposed bidirectional connectivity, where Figure 3a,b present the “top-to-down” and “down-to-top” strategy, respectively. As shown in Figure 3, the direction from node A to B in the road section differs from the direction from node B to A by an angle of 180°. The angle from node B to A is less than 180°, which can be placed into the “down-to-top” branch. The angle from node A to B is more than 180°, which can be placed into the “top-to-down” branch. Thus, the angle from the two node which is between 0° and 179° can be placed into the “down-to-top” branch, while others can be placed into the “top-to-down” branch.
Figure 3 shows the definition of the proposed bidirectional connectivity, where Figure 3a,b present the “top-to-down” and “down-to-top” strategy, respectively. As shown in Figure 3, the direction from node A to B in the road section differs from the direction from node B to A by an angle of 180°. The angle from node B to A is less than 180°, which can be placed into the “down-to-top” branch. The angle from node A to B is more than 180°, which can be placed into the “top-to-down” branch. Thus, the angle from the two node which is between 0° and 179° can be placed into the “down-to-top” branch, while others can be placed into the “top-to-down” branch.

![Diagram](image)

**Figure 3.** The definition of the proposed bidirectional graph: (a) represents the “top-to-down” direction prediction from node A to B; (b) represents the “down-to-top” direction prediction from node B to A.

To improve the accuracy of connectivity prediction, for each road, we design a method for angular classification, which is shown as Figure 4. Theoretically, setting the classification interval of the direction angle to 1 degree can most accurately assist the prediction of the angle. However, it can be a significant hindrance to classification predictions. Under comprehensive consideration, the angular interval is set to 15 degrees, and the classification of the road angles can be divided into 25 categories, defined as $R_{a_i}$:

$$R_{a_i} = \begin{cases} 
\text{floor}(a_i/15), & \text{if } a_i \text{ exists and } 0 \leq a_i < 180 \\
\text{floor}((a_i - 180)/15) + 13, & \text{if } a_i \text{ exists and } 180 \leq a_i < 360 \\
24, & \text{otherwise}
\end{cases} \quad (1)$$

where $a_i$ represents the road angle of $i^{th}$ pixel and floor is the floor math function. If the $i^{th}$ pixel belongs to a road, it can be calculated to a category between 0 and 23. However, if the $i^{th}$ pixel belongs to background, it is defined as category 24. From Equation (1), it can be seen that the “top-to-down” angle prediction in Figure 3a and the “down-to-top” angle prediction in Figure 3b are clearly calculated.

![Diagram](image)

**Figure 4.** An angular classification method. The interval angle is set to 15, and if the node direction exists, it can be divided into 24 classes.
2.3. Road Node Prediction

In this section, we introduce the generation method of road nodes in detail. For the road network, we obtain the pixel coordinates \( x_i \) of the road inflection points, and then use a Gaussian distribution function to calculate the heat map of the inflection points, defined as \( f(x_i) \):

\[
f(x_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)
\]  

where \( \mu \) is the mathematical expectation of \( x_i \) and \( \sigma \) is standard deviation.

However, the road nodes generated according to the inflection points are relatively sparse, which cause difficulty in prediction of the connectivity between nodes. To solve the problem, we encrypt the road nodes. We define \( V = \{ v_1, v_2, \ldots, v_n \} \), where \( v_n \) is the \( n^{th} \) nodes. The final encrypted heatmap \( P(u) \) can be defined as

\[
P(u) = \alpha \times \exp\left(-\frac{(u - v_k)^2}{2\sigma^2}\right)
\]

where \( \alpha \) is the coefficients which is set as 1.5 in our method, \( u = (x_u, y_u) \) represents the pixel coordinate, and \( (u - v_k)^2 \) is the squared distance between pixel \( u \) and the \( k^{th} \) node.

The difference of the heatmap generated only by inflection points and generated by encrypted points is shown in Figure 5. It can be seen that after encrypting, the basic structure of a road can be obtained.

![Figure 5](image.png)

(a) Visualization of the node heatmap: (a) represents inflection point in road heatmap; (b) represents the encrypted heatmap.

2.4. Training Bi-HRNet

In this section, we describe, in detail, how to train our Bi-HRNet, using cross-entropy loss and \( L_2 \) loss. We use cross-entropy loss to train the bidirectional angle graph, and the loss function \( \text{Loss}_A \) is defined as

\[
\text{Loss}_A = \mathcal{L}_{CE}(v'_1, v_1) + \mathcal{L}_{CE}(v'_2, v_2)
\]

where \( \mathcal{L}_{CE}(v'_1, v_1) \) and \( \mathcal{L}_{CE}(v'_2, v_2) \) are the bidirectional angle graph loss function, respectively; \( v'_1 \) and \( v_1 \) represent the ground truth and prediction of “top-to-down” road angle direction, while \( v'_2 \) and \( v_2 \) represent the ground truth and prediction of “down-to-top” road angle direction.

In order to obtain the heatmap of the nodes more accurately, we use the multiresolution node prediction method to predict the node heatmap of the original size \( S' \) and the node...
Heatmap of 1/4 size $S'_{1/4}$, and the corresponding ground truths are $S_{1/4}$ and $S$, respectively.

To train the parameters of the node heatmap, the loss function $Loss_p$ is defined as

$$Loss_p = \frac{1}{N_{1/4}} \left( \sum_i \left( S'_{1/4}(x_i) - S_{1/4}(x_i) \right)^2 \right) + \frac{1}{N} \left( \sum_i \left( S'(x_i) - S(x_i) \right)^2 \right)$$

(5)

where $N_{1/4}$ is the number of pixels of the original image and $N$ represents the number of pixels of the image scaled down by a factor of 4.

The final trained loss of Bi-HRNet can be defined as

$$Loss = Loss_p + \lambda \times Loss_A$$

(6)

where $\lambda$ is the weight, which is set to 1.5, because the value of $Loss_A$ is relatively low.

2.5. Implementation Details

We implemented the proposed Bi-HRNet using PyTorch. We trained the model on a RTX 3090 GPU for about 100 epochs with a learning rate starting from 0.001 and decreasing by $2 \times$ every 50,000 iterations. The RMSPROP optimizer is used, with a decay rate of 0.9 and decay step of 10,000. The batch size is set to 2.

3. Experimental Results
3.1. Experimental Datasets

To illustrate the effectiveness of the proposed framework, we tested it on three datasets: a DeepGlobe dataset, a RoadTracer dataset, and a Google dataset. Samples of the three datasets are shown in Figure 6.

Figure 6. Visualization of the three experimental datasets: (a) DeepGlobe dataset; (b) RoadTracer dataset; (c) Google dataset.
The DeepGlobe dataset was proposed in [57], which included three public competitions for segmentation, detection, and classification tasks on satellite images. The dataset covers three countries: Thailand, Indonesia, and India, covering an urban area of 220 square kilometers. The data scenarios include urban areas, villages, wilderness areas, seaside, tropical rain forests, and other scenarios. The resolution is 0.5 m. The DeepGlobe dataset contains 8570 images with $1024 \times 1024$ pixels. For the road extraction challenge, we selected part of the whole dataset and randomly choose 4226 images for training and 1600 images for testing.

The RoadTracer dataset was proposed in [49], which is a large corpus of high-resolution satellite imagery and ground truth road network graphs covering the urban core of forty cities across six countries. For each city, it covers a region of approximately 24 square km around the city center with a resolution of 0.5 m. We randomly selected 960 images for testing, and 3840 images for training.

The Google dataset considered in this study contains Berlin, Copenhagen, Frankfurt, and Belgrade. The resolution is 0.5 m. For the evaluation of our method, we selected one of the cities, Berlin, as our experimental dataset. We choose one third of this city as the testing dataset and the rest as the training dataset. After cropping and augmentation, we obtained 27,900 images with a size of $1024 \times 1024$ pixels for training and 1536 images with the same size for testing.

### 3.2. Metrics

To measure the accuracy of the extracted road, we employ the precision, recall, and F1-score as the evaluation metrics, which can be defined as

- Precision = \( \frac{TP}{TP + FP} \)  
- Recall = \( \frac{TP}{TP + FN} \)  
- \( F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)

where \( TP, FN, \) and \( FP \) are true positive, false negative, and false positive, respectively. True positive is the number of road pixels correctly identified; the false negative is the number of road pixels wrongly identified as non-road pixels; false positive is the number of non-road pixels identified as road pixels.

To further measure the difference between the predicted road net and ground truth, we employ the average path length similarity (APLS) metric which sums the differences in optimal path lengths between nodes in the ground truth path \( G \) and the predicted road net \( G' \). The APLS metric scales from 0 (poor) to 1 (perfect), which can be defined as

\[
\text{APLS} = 1 - \frac{1}{N} \sum \min \left\{ 1, \frac{|L(a,b) - L(a',b')|}{L(a,b)} \right\}
\]

where \( N \) is the number of unique paths, while \( L(a,b) \) is the length of path \( (a,b) \). The sum is taken over all possible source \((a)\) and target \((b)\) nodes in the ground truth graph. The node \( a' \) represents the node in the predicted graph closest to the location of ground truth node \( a \), and the node \( b' \) represents the node in the predicted graph closest to the location of ground truth node \( b \).

### 3.3. Experimental Results on DeepGlobe Dataset

In this section, we evaluate our proposed Bi-HRNet on the DeepGlobe dataset, and the visualization results are shown in Figure 7. From Figure 7, it can be seen that the proposed Bi-HRNet performs well in cities. In urban scenes, the road network is intertwined and complex, which is very challenging for road extraction tasks. Nevertheless, the proposed Bi-HRNet can still extract the complex road network completely.
3.2. Metrics 
To measure the accuracy of the extracted road, we employ the precision, recall, and F1-score as the evaluation metrics, which can be defined as

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $TP$ is the number of road pixels correctly identified; $FP$ is the number of non-road pixels wrongly identified as road pixels; $FN$ is the number of road pixels wrongly identified as non-road pixels.

The APLS metric scales from 0 (poor) to 1 (perfect), which can be defined as

$$\text{APLS} = 1 - \frac{\text{APA}}{\text{MAX APA}}$$

where $\text{APA}$ is the average path length of non-road pixels identified as road pixels.

To further measure the difference between the predicted road net and ground truth, we calculate the APLS metric.

To verify the superiority of Bi-HRNet as compared with other methods, we compared it with LinkNet, D-LinkNet, and RoadTracer. Figure 8 shows the four representative visual results of the compared methods on non-urban dense areas of the DeepGlobe dataset. Figure 9 shows the four representative visual results of the compared methods on urban dense areas of the DeepGlobe dataset. The rows are images and prediction results for various samples in test dataset. The columns consist of the input image, the corresponding ground truth, and the predicted results by the proposed Bi-HRNet, and three compared methods.

**Figure 7.** Visualization results of the proposed Bi-HRNet on DeepGlobe dataset.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
</tr>
</thead>
</table>

**Figure 8.** Visual comparisons of road extraction results with different models on non-urban dense areas of the DeepGlobe dataset: (a) input test image; (b) corresponding ground truth image; (c) results using the LinkNet; (d) results of D-LinkNet; (e) results of RoadTracer; (f) results of proposed Bi-HRNet.

It can be seen in Figure 8 that the greatest difficulty in the tested two images is that there are two-lane roads in the area. Since the road lines in a two-lane road are close together, and there are usually only a few pixels between them, this is a significant obstacle to the extraction of two-lane roads. In addition, there are many dirt roads which have texture...
features that are similar to the background. From the visual comparison results, it can be seen that our Bi-HRNet performs well in such abovementioned situations. The LinkNet model cannot deal with the two-lane roads, while the D-LinkNet method performs a poor extraction on dirt roads. RoadTracer is a one-way tracking road extraction network and, due to the lack of bidirectional judgment, there are some disconnections in the prediction results. The proposed Bi-HRNet can completely predict each road in the double-lane road scene, and, in the difficult-to-distinguish dirt road scene, the Bi-HRNet also has certain advantages compared with other methods.

To further illustrate the effectiveness of the proposed Bi-HRNet, we compare it with other methods quantitatively. Table 1 shows the accuracies of the different methods. As shown in Table 1, the proposed Bi-HRNet outperforms various compared methods in all metrics. The Bi-HRNet achieved the highest F1-score of 0.8651, indicating that our method has the best effect on road integrity prediction, while the achieved highest APLS of 0.5478 illustrates that the Bi-HRNet has the best effect on road connection prediction.

**Table 1.** Comparison of LinkNet, D-LinkNet, RoadTracer, and the proposed Bi-HRNet on the DeepGlobe dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>𝑅𝑒𝑐𝑎𝑙𝑙</th>
<th>𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛</th>
<th>𝑊</th>
<th>𝑨𝑷𝑳𝑺</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkNet</td>
<td>0.8076</td>
<td>0.8574</td>
<td>0.8318</td>
<td>0.5089</td>
</tr>
<tr>
<td>D-LinkNet</td>
<td>0.8177</td>
<td>0.8786</td>
<td>0.8471</td>
<td>0.5143</td>
</tr>
<tr>
<td>RoadTracer</td>
<td>0.8345</td>
<td>0.8721</td>
<td>0.8529</td>
<td>0.5324</td>
</tr>
<tr>
<td>Bi-HRNet</td>
<td>0.8439</td>
<td>0.8878</td>
<td>0.8651</td>
<td>0.5478</td>
</tr>
</tbody>
</table>

3.4. Experimental Results on RoadTracer Dataset

In this section, we evaluate our proposed Bi-HRNet on the RoadTracer dataset, and the visualization results are shown in Figure 10. From Figure 10, it can be seen that the proposed Bi-HRNet performs well. There are many inner roads in this dataset, which have a big difference to common roads.

Figure 9 shows the visual comparisons of road extraction results with different models on urban dense areas of the DeepGlobe dataset. In the urban scene, the buildings are denser, which causes some disturbance to the road extraction task. The results of LinkNet have many disconnections in dense building areas. In many areas, both the LinkNet and D-Linknet models do not predict roads, proving both of the methods lack robustness in such scenes. The RoadTracer method and proposed Bi-HRNet perform well in such urban scenes; however, it can be clearly seen that the results extracted by Bi-HRNet are more complete than those by RoadTracer, which has a better effect on the prediction of some short roads.

To further illustrate the effectiveness of the proposed Bi-HRNet, we compare it with other methods quantitatively. Table 1 shows the accuracies of the different methods. As shown in Table 1, the proposed Bi-HRNet outperforms various compared methods in all metrics.
metrics. The Bi-HRNet achieved the highest F1-score of 0.8651, indicating that our method has the best effect on road integrity prediction, while the achieved highest APLS of 0.5478 illustrates that the Bi-HRNet has the best effect on road connection prediction.

Table 1. Comparison of LinkNet, D-LinkNet, RoadTracer, and the proposed Bi-HRNet on the DeepGlobe dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
<th>APLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkNet</td>
<td>0.8076</td>
<td>0.8574</td>
<td>0.8318</td>
<td>0.5089</td>
</tr>
<tr>
<td>D-LinkNet</td>
<td>0.8177</td>
<td>0.8786</td>
<td>0.8471</td>
<td>0.5143</td>
</tr>
<tr>
<td>RoadTracer</td>
<td>0.8345</td>
<td>0.8721</td>
<td>0.8529</td>
<td>0.5324</td>
</tr>
<tr>
<td>Bi-HRNet</td>
<td><strong>0.8439</strong></td>
<td><strong>0.8878</strong></td>
<td><strong>0.8651</strong></td>
<td><strong>0.5478</strong></td>
</tr>
</tbody>
</table>

3.4. Experimental Results on RoadTracer Dataset

In this section, we evaluate our proposed Bi-HRNet on the RoadTracer dataset, and the visualization results are shown in Figure 10. From Figure 10, it can be seen that the proposed Bi-HRNet performs well. There are many inner roads in this dataset, which have a big difference to common roads.

Figure 10. Visualization results of the proposed Bi-HRNet on RoadTracer dataset.

To verify the superiority of Bi-HRNet as compared with other methods on the RoadTracer dataset, we compared it with LinkNet, D-LinkNet, and RoadTracer. Figure 11 shows the four representative visual results of the compared methods. The rows are images and prediction results for various samples in the test dataset. The columns consist of the input image, the corresponding ground truth, and the predicted results by the proposed Bi-HRNet, and three compared methods.

From Figure 11, disconnection among the compared LinkNet, D-LinkNet, and RoadTracer methods is obvious. LinkNet and D-LinkNet are segmentation-based methods, which do not take into consideration the geometric topological properties of road networks. Since the Bi-HRNet encrypts the nodes of the road, as compared with the RoadTracer method, it can predict the road nodes more densely, thereby reducing the probability of disconnection prediction. From the visualization results in Figure 11, the proposed Bi-HRNet outperforms the compared methods in integrity and connectivity.

To further illustrate the effectiveness of the Bi-HRNet, we compare it with other methods quantitatively. Table 2 shows the accuracies of the different methods.

Table 2. Comparison of LinkNet, D-LinkNet, RoadTracer, and the proposed Bi-HRNet on the RoadTracer dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
<th>APLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkNet</td>
<td>0.6143</td>
<td>0.6523</td>
<td>0.6327</td>
<td>0.5021</td>
</tr>
<tr>
<td>D-LinkNet</td>
<td>0.6038</td>
<td>0.6789</td>
<td>0.6392</td>
<td>0.5067</td>
</tr>
<tr>
<td>RoadTracer</td>
<td><strong>0.6356</strong></td>
<td>0.6612</td>
<td>0.6481</td>
<td>0.5203</td>
</tr>
<tr>
<td>Bi-HRNet</td>
<td>0.6337</td>
<td><strong>0.6634</strong></td>
<td><strong>0.6482</strong></td>
<td><strong>0.5317</strong></td>
</tr>
</tbody>
</table>
To verify the superiority of Bi-HRNet as compared with other methods on the RoadTracer dataset, we compared it with Link Net, D-LinkNet, and RoadTracer. Figure 11 shows the four representative visual results of the compared methods. The rows are images and prediction results for various samples in the test dataset. The columns consist of the input image, the corresponding ground truth, and the predicted results by the proposed Bi-HRNet, and three compared methods.

As shown in Table 2, the proposed Bi-HRNet obviously outperforms the LinkNet and D-LinkNet methods and improves the F1-score and APLS from 0.6327 to 0.6482 and from 0.5021 to 0.5317, respectively. The F1-score gap between the Bi-HRNet and RoadTracer is very low, which proves that the two methods have similar capabilities in extraction road integrity. However, the Bi-HRNet improves the APLS of RoadTracer from 0.5203 to 0.5317, and such significant improvement illustrates that the proposed Bi-HRNet has a better connectivity prediction than RoadTracer.

3.5. Experimental Results on Google Dataset

In this section, we test our methods on the Google dataset that we constructed. The visualization results are shown in Figure 12.

As shown in Figure 12, the Bi-HRNet proposed in this paper also has a certain robustness on the dataset we constructed. Table 3 shows the quantitative results of the proposed method and compared methods on the Google dataset. The proposed method achieves a recall of 0.8671, a precision of 0.9017, an F1 of 0.8841, and an APLS of 0.5615, indicating the effectiveness of the Bi-HRNet on different datasets. Compared with other methods, the proposed Bi-HRNet improves the APLS of RoadTracer from 0.5582 to 0.5615 and improves the F1-score from 0.8801 to 0.8841, which illustrates that the proposed Bi-HRNet has a better connectivity prediction than other previous methods. Figure 13 shows the visualization results of the proposed method and other compared methods. From Figure 13, it can be seen that the proposed Bi-HRNet performs a better connectivity than other previous methods.

As compared with the public DeepGlobe and the RoadTracer datasets, one of the most significant features of the Google dataset constructed for our experiments is that the inclination of the image is very large. The large inclination angle causes the entire road to be completely covered by vegetation, and therefore it is impossible to visually determine whether there is a road in the area covered by vegetation.
Table 3. Quantitative results of the proposed method and compared methods on the constructed Google dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>Precision</th>
<th>F₁</th>
<th>APLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LinkNet</td>
<td>0.8421</td>
<td>0.8648</td>
<td>0.8533</td>
<td>0.5418</td>
</tr>
<tr>
<td>D-LinkNet</td>
<td>0.8539</td>
<td>0.8802</td>
<td>0.8669</td>
<td>0.5523</td>
</tr>
<tr>
<td>RoadTracer</td>
<td>0.8650</td>
<td>0.8957</td>
<td>0.8801</td>
<td>0.5582</td>
</tr>
<tr>
<td>Bi-HRNet</td>
<td>0.8671</td>
<td>0.9017</td>
<td>0.8841</td>
<td>0.5615</td>
</tr>
</tbody>
</table>

Figure 12. Visualization results of the proposed Bi-HRNet on the constructed Google dataset.

Figure 14 shows the visualization results of the proposed Bi-HRNet on the vegetation-covered area of the constructed Google dataset. As shown in Figure 11, the vegetation on both sides of the road in this area is densely covered, and there are serious shadows, which have a significant impact on the results of road extraction. Nevertheless, the method proposed in this paper can still completely and accurately extract the roads in this area, which shows that the proposed method has a certain anti-interference ability for vegetation coverage.
To obtain the connection between the predicted nodes, we proposed a bidirectional angle graph prediction branch. We ignored the prediction of angle value and instead adopted the inflection point of the predicted heatmap, we use multiscale heatmap learning to enhance feature expression. Thus, the proposed bidirectional connectivity prediction strategy. Finally, to investigate the effect of the multi-scale road nodes prediction branch, we start this branch in our network. It can be seen that the Bi-HRNet with the three parts achieves the highest \( \text{F}_1 \) and APLS.

Table 3. Quantitative results of the proposed method and compared methods on the constructed Google dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( \text{Recall} )</th>
<th>( \text{Precision} )</th>
<th>( \text{F}_1 )</th>
<th>( \text{APLS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-HRNet</td>
<td>0.8671</td>
<td>0.9017</td>
<td>0.8841</td>
<td>0.5615</td>
</tr>
<tr>
<td>D-LinkNet</td>
<td>0.8539</td>
<td>0.8802</td>
<td>0.8669</td>
<td>0.5523</td>
</tr>
<tr>
<td>RoadTracer</td>
<td>0.8650</td>
<td>0.8957</td>
<td>0.8801</td>
<td>0.5582</td>
</tr>
</tbody>
</table>

4. Discussion

4.1. Main Goals of the Study

The main goals of this study were to extract roads using satellite imagery. By performing the proposed node heatmap extraction branch on satellite imagery, the inflection point heatmap and encryption point heatmap could be predicted. To improve the accuracies of the predicted heatmap, we use multiscale heatmap learning to enhance feature expression. To obtain the connection between the predicted nodes, we proposed a bidirectional angle graph prediction branch. We ignored the prediction of angle value and instead adopted the method of predicting the range of the bidirectional angle. Finally, the proposed method demonstrated better extraction accuracy.

4.2. Ablation Experiment

To investigate the behavior of the proposed top-to-down directional connectivity, down-to-top directional connectivity, and multi-scale road nodes, we conducted several...
ablation studies on DeepGlobe dataset. The ablation experimental results are shown in Table 4.

Table 4. Ablation experimental results of different part on DeepGlobe dataset.

<table>
<thead>
<tr>
<th>Top-to-Down Directional Connectivity</th>
<th>Down-to-Top Directional Connectivity</th>
<th>Multi-Scale Road Nodes Prediction</th>
<th>Recall</th>
<th>Precision</th>
<th>F₁</th>
<th>APLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>√</td>
<td></td>
<td></td>
<td>0.8211</td>
<td>0.8572</td>
<td>0.8388</td>
<td>0.5382</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
<td></td>
<td>0.8389</td>
<td>0.8781</td>
<td>0.8581</td>
<td>0.5449</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.8439</td>
<td>0.8878</td>
<td>0.8651</td>
<td>0.5478</td>
</tr>
</tbody>
</table>

First, we show the effect of the proposed bidirectional connectivity prediction. For this, we use top-to-down directional connectivity prediction and stop the down-to-top directional connectivity predication and multi-scale road nodes prediction. From Table 4, it can be seen that the method achieves an F₁ of 0.8388 and an APLS of 0.5382. Next, to investigate the effect of bidirectional connectivity prediction, we start down-to-top directional connectivity prediction branch. It can be seen that the F₁ improves from 0.8388 to 0.8581 and the APLS improves from 0.5382 to 0.5449, which illustrates the effectiveness of the proposed bidirectional connectivity prediction strategy. Finally, to investigate the effect of the multi-scale road nodes prediction branch, we start this branch in our network. It can be seen that the Bi-HRNet with the three parts achieves the highest F₁ and APLS.

4.3. Extended Experiment

The experimental results in Section 3 illustrate that the proposed Bi-HRNet has a better performance than other methods on each independent dataset; however, this does not show the transfer ability of the model. To prove the transfer ability, we used the model trained on the constructed Google dataset to test the public Massachusetts dataset. The visualization results are shown in Figure 15.

We trained an optimal model using our own labeled Google images and did not optimize the design on the Massachusetts road dataset. Figure 15 shows part of the road extraction results in the Massachusetts road data. It can be seen from the results that there are still some road fractures and missing connections in the extraction results. However, it has been proven by experiments that our algorithm model is highly transferable and has a good robust effect. The quantitative results are shown in Table 5. The proposed Bi-HRNet achieved F₁ of 0.8388 and APLS of 0.5170, which illustrates the transferability of our method.

Table 5. Quantitative results of the proposed method on the constructed Massachusetts dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>Precision</th>
<th>F₁</th>
<th>APLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-HRNet</td>
<td>0.8163</td>
<td>0.8626</td>
<td>0.8388</td>
<td>0.5170</td>
</tr>
</tbody>
</table>
4.3. Extended Experiment

The experimental results in Section 3 illustrate that the proposed Bi-HRNet has a better performance than other methods on each independent dataset; however, this does not show the transfer ability of the model. To prove the transfer ability, we used the model trained on the constructed Google dataset to test the public Massachusetts dataset. The visualization results are shown in Figure 15.

Figure 15. Visualization results of the proposed Bi-HRNet on the Massachusetts dataset with the model trained on the Google dataset.

5. Conclusions

This study presents a road extraction framework, namely, Bi-HRNet, for satellite imagery. The Bi-HRNet is a multi-task learning framework that contains three parts: the “top-to-down” road direction prediction branch, “down-to-top” road direction prediction branch, and node heatmap prediction branch. The “top-to-down” road direction graph and
“down-to-top” road direction graph, which are called bidirectional graphs, are the key for predicting road direction between nodes. To obtain the direction angle conveniently, we proposed a road direction angle classification method instead of road angle prediction. In the node heatmap prediction branch, to obtain the node heatmap, we proposed a multiscale heatmap prediction method, which enhanced the feature expression in this branch. In comparison with results from other road extraction methods, such as LinkNet, DLinkNet, and RoadTracer, the extraction accuracy using the proposed framework can satisfy practical applications. In addition, the proposed Bi-HRNet can provide a convenient way to extract road via satellite imagery.

For further work, we plan to focus on the following area: a new deep-learning-based framework for road extraction using semi-supervised or weakly supervised learning.

Author Contributions: Z.W. conceived and designed the algorithm and experiments and wrote the manuscript; J.Z. guided the algorithm design and revised the manuscript; L.Z. provided a part of the comparative experimental results; X.L. and H.Q. provided the experimental environment. All authors have read and agreed to the published version of the manuscript.

Funding: This manuscript received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

Remote Sens. 2022, 14, 1732


41. 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]


56. Sun, Z.; Lin, D.; Wei, W.; Wozniak, M.; Damasevicius, R. Road Detection Based on Shearlet for GF-3 Synthetic Aperture Radar Images. *IEEE Access* 2020, 8, 28133–28141. [CrossRef]