Research article

Satellite road extraction method based on RFDNet neural network

Weichi Liu, Gaifang Dong* and Mingxin Zou

College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010011, China

* **Correspondence:** Email: donggf@imau.edu.cn; Tel: +8615848906889.

Abstract: The road network system is the core foundation of a city. Extracting road information from remote sensing images has become an important research direction in the current traffic information industry. The efficient residual factorized convolutional neural network (ERFNet) is a residual convolutional neural network with good application value in the field of biological information, but it has a weak effect on urban road network extraction. To solve this problem, we developed a road network extraction method for remote sensing images by using an improved ERFNet network. First, the design of the network structure is based on an ERFNet; we added the DoubleConv module and increased the number of dilated convolution operations to build the road network extraction model. Second, in the training process, the strategy of dynamically setting the learning rate is adopted and combined with batch normalization and dropout methods to avoid overfitting and enhance the generalization ability of the model. Finally, the morphological filtering method is used to eliminate the image noise, and the ultimate extraction result of the road network is obtained. The experimental results show that the method proposed in this paper has an average F1 score of 93.37% for five test images, which is superior to the ERFNet (91.31%) and U-net (87.34%). The average value of IoU is 77.35%, which is also better than ERFNet (71.08%) and U-net (65.64%).

Keywords: remote sensing image; neural network; residual structure; morphological filtering; road extraction

1. Introduction

With the accelerated urbanization, cities are expanding and their traffic is getting busier and busier. The construction and maintenance of road networks have become a crucial task in the field of urban

planning and transportation [1]. Road extraction has been an important research direction in the field of computer vision, which is important for urban planning, traffic management, environmental monitoring and other fields [2,3]. In urban planning, road extraction can provide an important reference basis for urban traffic planning, road construction and greening planning; in traffic management, road extraction can improve the efficiency of traffic supervision and management and provide help for the prevention and treatment of traffic accidents; in environmental monitoring, road extraction can help monitor the environment around the road and provide strong support for urban environmental management. In addition, road extraction methods are also widely used in emerging technologies such as positioning and navigation, autonomous driving, etc. In positioning and navigation, road extraction methods can improve the accuracy and reliability of navigation systems by identifying and extracting road features to achieve precise positioning and path planning of vehicles. In autonomous driving, road extraction methods can achieve route planning and control of autonomous vehicles by recognizing and extracting road features, as well as improve the safety and reliability of autonomous driving systems.

However, the task of road extraction is not an easy one. Roads in remotely sensed images or images captured by cameras are usually hidden by the complex surrounding background [4], and the shape and direction of roads vary widely, thus putting high demands on the accuracy and robustness of the algorithm. In this context, researchers have proposed various algorithms and methods to solve the road extraction problem [5].

Traditional road extraction methods focus on information such as geometric and texture features of the road and mostly use pixel-level segmentation methods, which are usually based on basic image processing techniques such as morphology, threshold segmentation and edge detection. For example, the authors of [6] applied geometric a priori augmented road extraction methods to introduce geometric information into the model in order to generate an effective road extraction network. The authors of [7] proposed a light topological spatial network road extraction method based on knowledge distillation to extract road networks using topological features of roads. The authors of [7] proposed a semi-automatic road extraction method based on multiple descriptors to solve the problems of incomplete geometric information of road images and poor uniformity of internal road textures while ensuring the accuracy of road extraction. The authors of [8,9] proposed a semantic segmentation network to extract complex road information using the properties of neural networks. Although these algorithms can obtain better results in some cases, for road extraction in complex backgrounds, these methods have major limitations, specifically in terms of sensitivity to environmental factors such as lighting, shadows and noise, as well as difficulty in dealing with road crossings and road breaks.

Region-based road extraction algorithms, on the other hand, adopt a more advanced strategy. These algorithms usually extract the regions of interest of the road from the image, and then merge and filter these regions to finally obtain the boundary information of the road. Compared with pixel-level algorithms, region-based algorithms can better handle problems such as road shapes and complex backgrounds. For example, the authors of [10] proposed a road extraction network based on bidirectional spatial information inference, which captures spatial context dependence and extends the perceptual domain using neighborhood feature fusion; it also combines recurrent neural network structured information processing units to capture channel dependence. The authors of [11] proposed a graph-attention network for road extraction from remote sensing images, which uses environmental and spatial information of graph-attention module to extract roads; they also designed a channel fusion module to achieve the fusion of low-level and high-level features, which can ensure that the extracted road network presents rich detailed information. However, such algorithms also have some drawbacks;

for example, they are sensitive to road width and image quality, cannot handle complex road structures and have high computational complexity.

Thanks to recent advances in deep learning techniques, the tedious manual road segmentation can be automated. However, most of these models are computationally intensive and thus not suitable for resource-limited road extraction tasks. To alleviate this bottleneck, the authors of [12] proposed two lightweight models based on depth-separable convolution and ConvMixer initial blocks, both of which exploit the computational efficiency of depth-separable convolution and initial block multiscale processing and combine them into the codec architecture of a U-Net. The authors of [13] proposed a graffiti-based weakly supervised remote sensing road extraction network, which can extract roads from remote sensing images based on graffiti annotations. The authors of [14,15] proposed a semi-weakly supervised remote sensing image road extraction method based on adversarial learning to make full use of weak annotations, and it was trained using a small set of pixel-by-pixel annotated data and a large amount of weakly annotated data. The authors of [16,17] proposed a multi-task, multi-origin fusion network to extract road networks using a collection of remotely sensed images and trajectory data based on the fact that roads and intersections are two key elements for road network generation. Although neural network-based road extraction algorithms have made significant progress, they still have some drawbacks and limitations, specifically, their large data requirements, long training time and high requirements for computational resources.

In the field of semantic segmentation, the authors of [18] proposed the efficient residual factorized convolutional neural network (ERFNet) for the first time in 2018; this network model includes a redesigned residual layer to make the model achieve a good balance between reliability and speed, and it has been widely used in many fields, especially in autonomous driving. The authors of [19] designed a conditional generative adversarial network (GAN) to learn the distribution of roads present in official maps in an unsupervised environment, which is able to overcome the shortcomings caused by extracting roads through semantic segmentation; but, as with most deep learning models, the quality of the generated machine predictions is highly dependent on the quality of the conditional training data and the model is sensitive to the number of holes in the data species. The authors of [20] proposed a new architecture for real-time semantic segmentation tasks called DuFNet. This model is able to fuse spatial and contextual features more efficiently and contributes to the previous global representation; it has promising applications in the field of autonomous driving. The authors of [21] proposed a new U-net model that incorporates depthwise separable convolution and an attention mechanism. The improved model constituted a significant improvement in segmentation accuracy, especially for important segmentation targets in autonomous driving systems. While these models in the semantic segmentation domain are able to balance the accuracy and speed of extraction, they are designed mainly for roads in the autonomous driving class, which have a very different spatial resolution than the one used in our task.

In order to solve the problems of the above road extraction methods, we propose a residual neural network model combined with morphological filtering for road extraction from remote sensing images [22]. Our approach draws on the advantages of various models proposed in the field of semantic segmentation and makes targeted improvements for our task. First, we acquired and annotated 1270 images with high quality and used data enhancement techniques to create a dataset of moderate size and excellent quality. Then, we constructed an encode–decode-based residual network architecture for road extraction, in which the middle residual module can identify high-dimensional road features, and DoubleConv can compensate for the disadvantage of missing information extraction caused by the

residual structure., which can effectively solve the problems such as excessive burr phenomenon, incomplete road extraction and incorrect recognition of targets similar to road linear features such as rivers and railroads as roads. By setting the Exponential Linear Unit (ELU) activation function and Adam optimizer, the convergence speed of the model is accelerated and computational resources are saved. Finally, the morphological filtering method is used to smooth the image for the noise phenomenon existing in the extracted road images. The experimental results prove that the method used in this study can effectively and accurately identify various types of roads, with excellent performance on evaluation indexes such as the F1 score and Intersection over Union (IoU) score.

2. Materials and methods

2.1. Experimental process
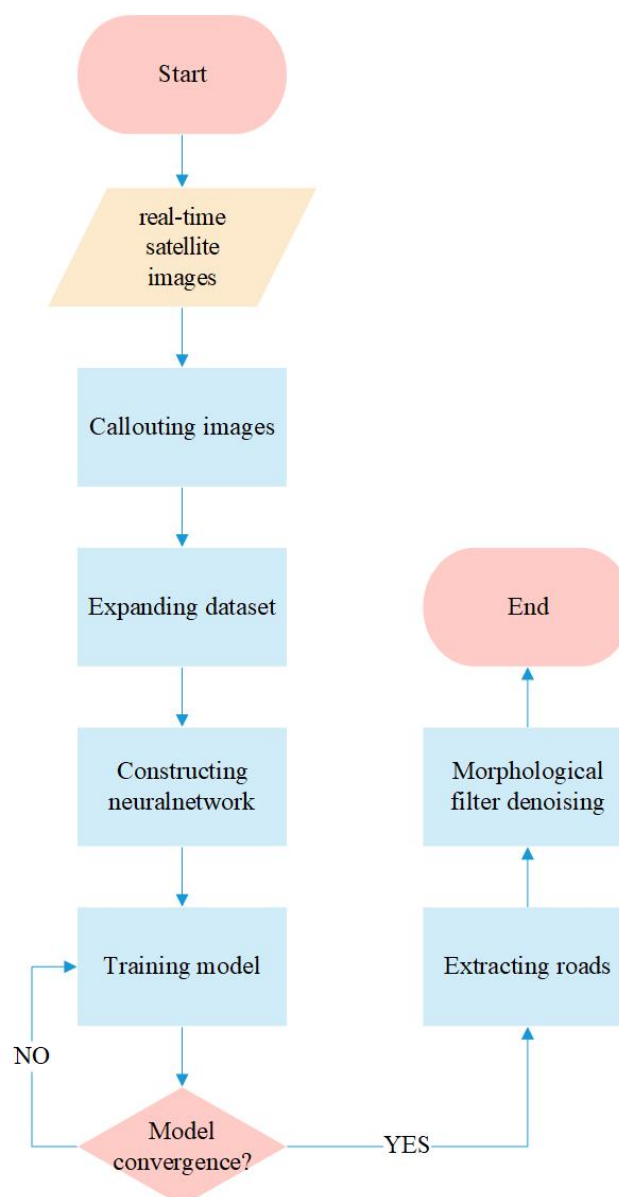
**Figure 1.** Road network extraction flowchart.

The experimental process of this study is as follows. In the first step, real-time satellite images were acquired from Bigemap remote sensing satellite software; then, the labelme image annotation tool was used to annotate the road network; after that, the corresponding labels of the images were manually interpreted to obtain the labels. In the second step, the dataset was expanded using the image flipping method. In the third step, the training dataset was adjusted to a size appropriate to the network structure for training, and the model with the best training results was saved and used for road network extraction. Finally, the morphological filtering open operation was used to eliminate the noise present in the road network extraction results, and the ultimate road network extraction results were obtained. The overall flow of the experiment is shown in Figure 1.

2.2. RFDNet architecture

The ERFNet is a semantic segmentation network based on residual connection and deep separable convolution. It is an improvement on the residual neural network and efficient neural network [23]. The design of the RFDNet has been improved on the basis of the ERFNet, as shown in Table 1. First, the RFDNet introduces the DoubleConv module, which uses two consecutive convolution operations, allowing the model to better extract edge features of images and distinguish regions with similar linear features but different textures. Second, we replaced the Rectified Linear Unit (ReLU) activation function used by ERFNet with the ELU because the ELU can take a negative value, unlike ReLU, which makes the average output value of the activation unit close to 0 so as to reduce the offset and make the gradient closer to the natural gradient; this better solves the overfitting problem that occurs in neural network training. Finally, in the selection of optimization strategies, the RFDNet adds the Adam optimizer, enabling the network to achieve better convergence performance. The RFDNet network architecture is shown in Figure 2. The "coder-decoder" approach is adopted. In this architecture, the fusion of feature maps of each layer can make the network output clearer. Layers 1–21 of the network make up the encoder, and layers 22–28 make up the decoder. The module types and related parameters of each layer are shown in Table 2.

Table 1. Comparison of implementation details of three models.

	RFDNet	ERFNet	U-net
Number of layers	28	23	46
Main operation types of modules	Downsample, Upsampling, DoubleConv, Non-bt-1D	Downsample, Upsampling, Non-bt-1D	Downsample, Upsampling
Activation function	ELU	ReLU	ReLU
Optimization strategy	BatchNorm, Dropout, Adam	BatchNorm, Dropout	BatchNorm
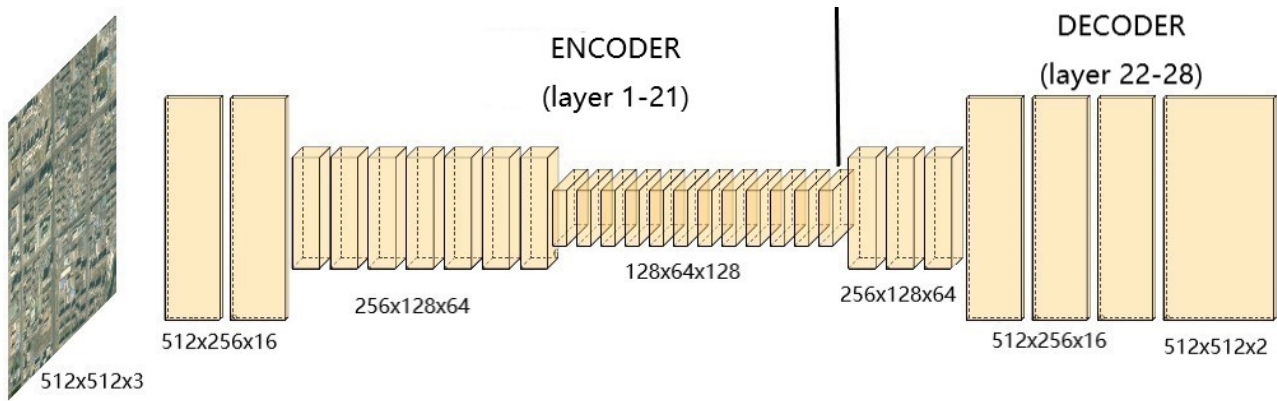
Figure 2. RFDNet architecture.

Table 2. Modules and parameters of each layer of the RFDNet.

	Layer	Type	out-F	out-Res
ENCODER	1	Downsampler	16	512×256
	2	DoubleConv	16	512×256
	3	Downsampler	64	256×128
	4	DoubleConv	64	256×128
	5–9	$5 \times$ Non-bt-1D	64	256×128
	10	Downsampler	128	128×64
	11	DoubleConv	128	128×64
	12–21	$10 \times$ Non-bt-1D	128	128×64
DECODER	22	Upsampling	64	256×128
	23–24	$2 \times$ Non-bt-1D	64	256×128
	25	Upsampling	16	512×256
	26–27	$2 \times$ Non-bt-1D	16	512×256
	28	Upsampling	2	512×512

The coding part of the network includes layers 1–21. Downsampler is used in layers 1, 3 and 10, and the specific implementation is to optimize the model with a convolutional kernel size of 3×3, step size of 2, maximum pooling operation and batch normalization (BN). The function of Downsampler enables the underlying unit to obtain more semantic information, which helps the network to extract texture and structure information, ignore useless features and improve the focus of the model. The DoubleConv module is used in layers 2, 4 and 11. DoubleConv uses successive double convolution operations to enhance the network for edge information extraction and texture information differentiation. Compared with the traditional convolution operation, it fully utilizes the information of feature channels and improves the overall performance of the network. In layers 5–9 and 12–21, Non-bt-1D operations are used to compensate for the information difference between low-level detail information and high-level information. The specific implementation has been designed with five parallel branches with dilation rates of 2, 4, 8, 16 and 32. Non-bt-1D is able to expand the feature acceptance field without sacrificing the feature spatial resolution, and to densely connect the cavity convolution with different dilation rates to obtain a larger range of perceptual fields. Non-bt-1D is also able to transfer the feature information from the bottom layer directly to the corresponding top layer and perform the feature information connection operation. This actually creates a channel for message

transmission between low-level and high-level information, which can greatly reduce the time it takes for the model to achieve convergence.

The decoding part of the network includes layers 22–28. Since the semantic segmentation task needs to keep the output results invariant in terms of spatial resolution, it requires the use of decoders to recover the size of the feature maps. Layers 22, 25 and 28 use Upsampling, and they are responsible for resizing the number of feature map channels to a specific size. In layers 23, 24, 26 and 27, the Non-bt-1D module continues to be used to improve the decoding performance of the model. Because Non-bt-1D can reduce the information loss in the road network extraction process and maximize the retention of key features in the images, which ensures that the final output of the model can achieve higher recognition accuracy and effectively complete the target of extracting roads.

2.3. Residual structure

The residual structure of Non-bottleneck-1D designed for the network presented in this paper is shown in Figure 3. This structure is different from the two-dimensional convolution of the traditional residual neural network's residual structure. It uses the more advanced one-dimensional dilated convolution at present. The advantage of this design is that, by reducing the number of parameters and increasing the nonlinear layer, the range of the receptive field can be expanded, and each convolutional output can also cover a wide range of semantic information to improve the accuracy of semantic recognition. Compared with ordinary convolution operations, dilated convolution has an additional expansion rate parameter. This parameter refers to the number of intervals between convolution cores, where the hole area is filled with 0 and the expansion rate of ordinary convolution operations is 1. The receptive field of cavity convolution is calculated as shown in formula (1):

$$F = k + (k - 1)(r - 1) \tag{1}$$

In the formula, k is the size of the convolutional kernel and r is the expansion rate. The dilated convolution in the residual structure used in this paper adopts the parallel way, as shown in Figure 4, where a, b, c and d respectively represent the receptive fields with sizes of 5, 9, 17 and 33 that can be received at the expansion rates of 2, 4, 8 and 16.
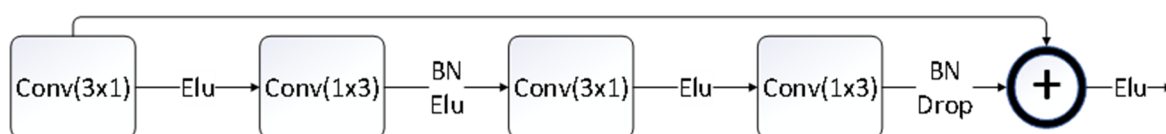
Figure 3. Residual structure.
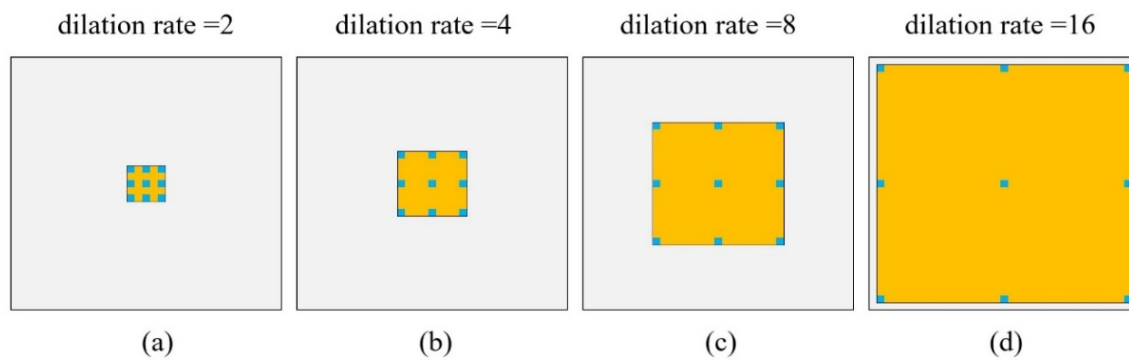
Figure 4. Dilated convolution.

2.4. BN and dropout

BN is a special structure in a neural network [24]. BN has the following advantages when building neural networks. First, after adding a BN layer, the training process of the network will be more stable and the training speed will be improved. BN adopts two methods of normalization and linear change, which can make the mean and variance of network input data more stable and thus allow the lower structure to eliminate the influence of network input changes in the lower structure, promote the independent learning of each layer and accelerate the convergence speed of the network. Second, BN can optimize parameter adjustment, reduce the dependence of the model on network parameters and improve the generalization ability of the model. After the BN layer is used, it effectively alleviates the problem that the fine-tuning of parameters will enlarge with the increase of network depth, and it makes the update status of parameters more stable. Finally, BN can effectively suppress the problem of gradient disappearance. It can transform the distribution of the eigenvalues of each layer of the network into a standard normal distribution so that the eigenvalues are maintained in the range where the activation function is sensitive to input, thus avoiding the gradient vanishingin the training process of the network.

Dropout is a method proposed to avoid overfitting of the model. In the process of network training, if the model is large and the data size is small, it is easy to cause overfitting. Dropout will randomly discard the training units of the network according to a certain probability, which can reduce the dependence of each neuron so that the network can learn more robust features. After the introduction of Dropout, each node of the neural network can play a role, effectively reducing the number of intermediate features and realizing the regularization of the neural network.

2.5. Loss function

The loss function used in this work is CrossEntropyLoss, and its calculation method is shown in formula (2):

$$C = -\frac{1}{n}\sum_x[y \ln \acute{y} + (1 - y) \ln(1 - \acute{y})] \tag{2}$$

In the formula, x represents the sample, n represents the total number of samples, y represents

the actual value and $ý$ represents the output value. The cross-entropy loss function is mainly used to evaluate the close degree between the output value of the neural network and the actual value. According to the difference between the "distribution of the learning model" and the "distribution of the training data", when the cross-entropy value is the lowest, the neural network can be considered to have learned the model with the best performance. At the same time, when using a sigmoid function to optimize the gradient descent problem, the cross-entropy loss function can also control the output error to avoid the decrease of the learning rate.

2.6. Morphological filtering

After using the network model to extract the road network initially, we applied the open operation method of morphological filtering to remove the noise in the image, so as to obtain a clearer road network extraction result.

The open operation is a filter based on a geometric operation whose equation is expressed as shown in formulas (3)–(5): If the set A and B satisfy formula (3), then A is said to be dilated by B; If formula (4) is satisfied, then A is said to be eroded by B; If formula (5) is satisfied, then the image B is said to do the open operation on the image A. The open operation adopts the processing method of erosion before dilation, whereby the erosion operation can make the target area in the image smaller and eliminate small and meaningless objects [25], while the dilation operation can fill the holes in the target and eliminate the existing particle noise. In general, the open operation can eliminate the tiny objects in the image and separate them in a thin place, and it can also smooth the boundary of the object without significantly changing the area, effectively solving the "burr" problem that often occurs in road extraction. The effects of open operation in this paper are shown in Figure 5.

$$A \oplus B = \{x: \hat{B}_x \cap A \neq \emptyset\} \tag{3}$$

$$A \odot B = \{x: B_x \subseteq A\} \tag{4}$$

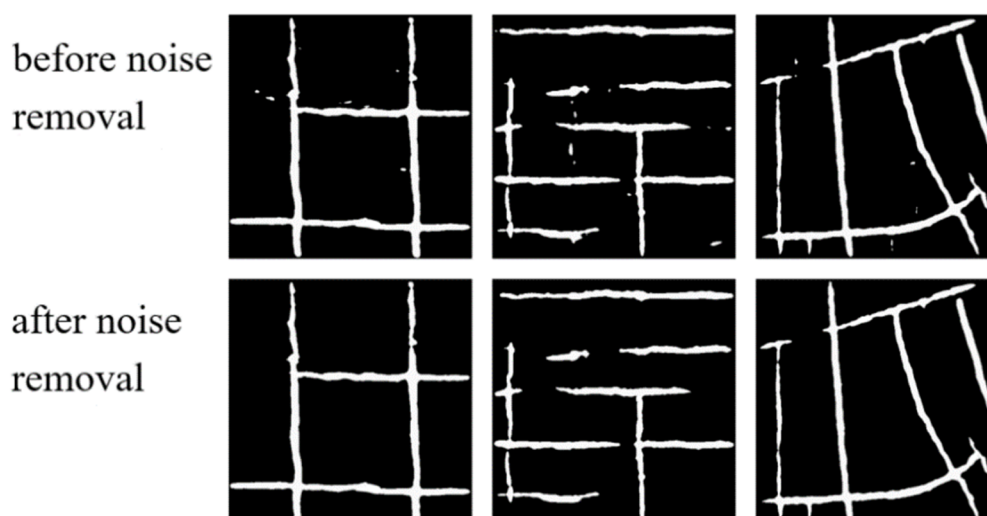$$A \circ B = (A \odot B) \oplus B \tag{5}$$

Figure 5. Filtering effect.

3. Experimental results and analysis

3.1. Dataset processing

The remote sensing dataset used in the experiment was selected from BigmapGIS, and it included 1250 images. The image resolution is 512 × 512, and each image has a corresponding binary label. The dataset covers large cities, small counties, urban suburbs and mountainous areas. The selected image backgrounds are complex, and they include interference factors such as buildings, rivers, bridges, vegetation and shadows. Because the training dataset wass relatively small, the dataset was expanded by applying vertical and horizontal image flipping.

3.2. Experimental configuration

Training and testing were conducted on a Windows 11 with a Python deep learning platform. The processor was AMD Ryzen 5800 H, the graphics card was NVIDIA GeForce RTX 3060 and the computer memory size was 16 GB. In the training phase, the epoch number was set to 100, the batch size was set to 5 and the optimizer was the Adam optimizer; the learning rate was dynamically set, with an initial value of 0.001, followed by decreases with the increase of the epoch number.
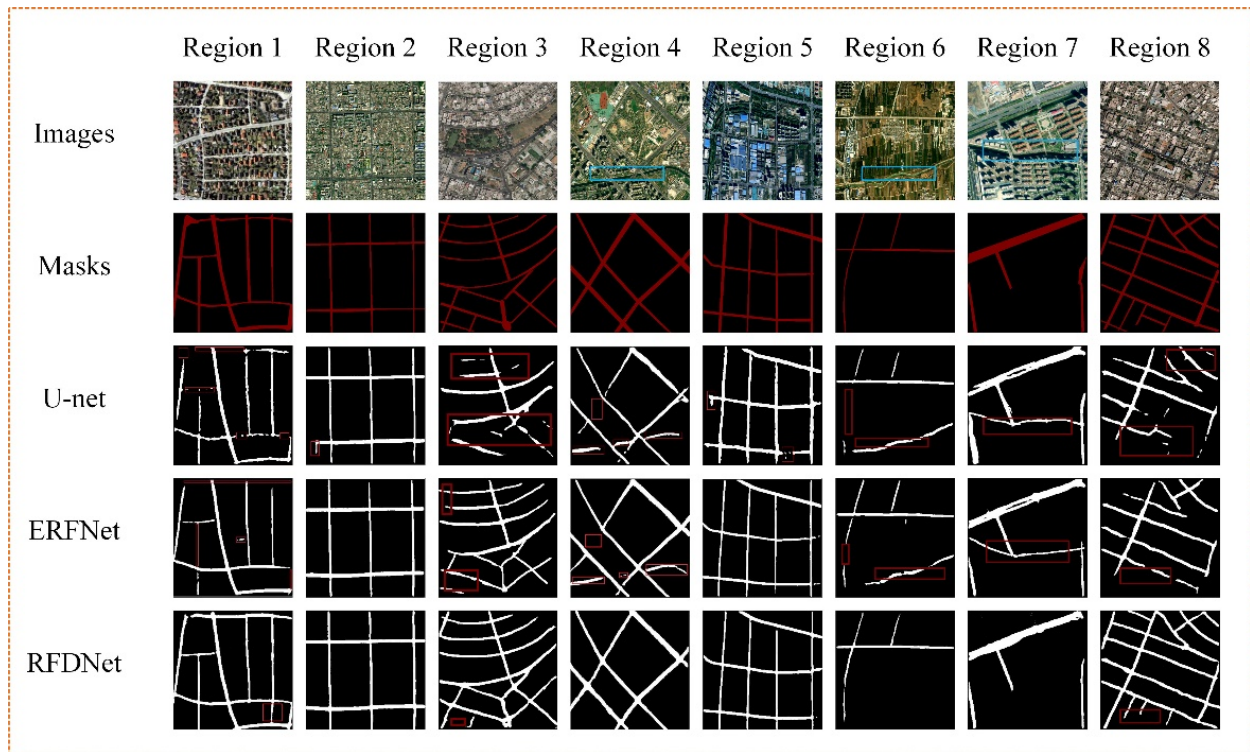
3.3. Experimental results

3.3.1. Visual comparison

To test the feasibility of our proposed method, the U-net and ERFNet were compared with the extraction performance of the RFDNet used in this work (using the same dataset and both denoised). The test images included both urban and rural areas, and the image backgrounds covered road targets such as urban main roads, highways and rural dirt roads, as well as distracting ground objects such as rivers, vegetation, shadows and buildings. In the satellite images, the rivers that appear are marked with blue wireframes, and, in the output result map, the areas with poor extraction results are marked with red wireframes.

From Figure 6, it can be found that our proposed method can extract complete and clear road areas from remote sensing images covering various complex scenes. In particular, RFDNet can extract road targets more completely in regions with dense road networks, such as Regions 1, 5 and 8, compared with the other two models, as shown by the accurate identification of edge regions and narrow road targets and the less occurrence of broken and isolated road regions in the extraction results. In the images with rivers, such as Regions 4, 6 and 7, RFDNet is able to avoid the incorrect extraction of rivers, unlike the other two models. The reason for this is that RFDNet increases the number of downsampling operations and designs the DoubleConv module to fully combine the texture information, edge information and deep semantic information of the shallow feature map so that the high-dimensional road feature information can be extracted more accurately.

For the poorly recognized road regions of RFDNet, such as the parts marked with red boxes in Regions 1, 3 and 8, the reason may be due to the different road levels and the different labeling standards used in the process of labeling the images, leading to the overfitting of the model training data, which thus caused the model to miss the extraction of lower-level roads.

Figure 6. Comparison of experimental results.

3.3.2. Quantitative comparison

Road extraction is essentially a binary problem. The pixel value of the road area is positive, and the pixel value of the background area is negative. In this work, the commonly used evaluation methods of semantic segmentation, F_1 score and IoU were used to quantitatively evaluate the extraction effect for roads. The calculation formula is as follows: the accuracy, P, represents the proportion of the real road area in the extracted road area; the recall rate R represents the proportion of real roads in extracted roads; F_1 represents the weighted average of the accuracy and recall rate; IoU represents the ratio of intersection to union between the extracted road area and the real road area. When the two areas overlap completely, the IoU is the largest, at 1. α_{TP} represents the area correctly identified as a road; α_{FP} represents the area where the background is recognized as a road; α_{FN} refers to the area where the road is misjudged as the background.

$$P = \frac{\alpha_{TP}}{\alpha_{TP} + \alpha_{FP}} \tag{6}$$

$$R = \frac{\alpha_{TP}}{\alpha_{TP} + \alpha_{FN}} \tag{7}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{8}$$

$$IoU = \frac{\alpha_{TP}}{\alpha_{TP} + \alpha_{FN} + \alpha_{FP}} \tag{9}$$

As shown in Table 3, RFDNet achieved the best performance among the three models, with an

average F1 score and IoU score of 93.37% and 77.35%, respectively, on the eight test results, and it achieved the highest score on six test results; although the scores for Regions 2 and 5 tests were not as good as ERFNet, the difference was small. In Regions 4, 6 and 7, the F1 and IoU scores of RFDNet (92.15%, 92.91%, 94.70 and 73.44%, 78.54%, 78.39, respectively) were much higher than those of U-net (88.29%, 88.15%, 91.86% and 66.42%, 59.48%. 62.17%, respectively) and ERFNet (89.46%, 88.63%, 88.83% and 68.34%, 61.33%, 62.85%, respectively), which is due to the fact that RFDNet is able to distinguish well from the other two models when there are interfering targets with linear features such as rivers in these three test images, once again demonstrating the superiority of RFDNet.

Table 3. Performance comparison for different models.

Model parameter	U-Net		ERFNet		RFDNet	
	F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
Region 1	78.64	57.32	90.08	65.15	**93.17**	**76.94**
Region 2	91.44	72.68	**95.86**	**79.04**	95.66	78.07
Region 3	81.05	60.24	92.28	76.18	**93.14**	**77.71**
Region 4	88.29	66.42	89.46	68.34	**92.15**	**73.44**
Region 5	91.27	71.14	**92.91**	**78.01**	92.56	77.87
Region 6	88.15	59.48	88.63	61.33	**92.91**	**78.54**
Region 7	88.04	62.17	88.83	62.85	**94.70**	**78.39**
Region 8	91.86	75.69	92.45	77.74	**92.66**	**77.82**
Mean	87.34	65.64	91.31	71.08	**93.37**	**77.35**

4. Application on public datasets

To indicate our method's application scope and efficiency for different image types, we selected three public datasets (CHN-CUG, GF-2 and Massachusetts) for testing. The CHN6-CUG (https://grzy.cug.edu.cn/zhuqiqi) dataset is a remote sensing image dataset released by China University of Geosciences (Wuhan, China), and it contains six bands, namely, blue, green, red, near-infrared, short-wave infrared and short-medium-wave infrared. The GF-2 (http://www.aircas.cas.cn/index_73758.html) dataset is provided by China National University of Defense Technology and China Key Laboratory of Remote Sensing Satellite Radiation Measurement and Control Technology, and it contains eight optical bands and one Synthetic Aperture Radar (SAR) band. The Massachusetts (https://www.cs.toronto.edu/~vmnih/data/) dataset is a high-resolution remote sensing image dataset collected by the city of Cambridge, Massachusetts, USA in cooperation with Massachusetts Institute of Technology (MIT), and it contains three bands, i.e., red, green and blue. For other specific information about these three datasets, please refer to Table 4. In satellite images, as shown in Figure 7, the rivers that appear are marked with blue wireframes, while, in masked images, the parts that are clearly missing in road extraction are marked with red wireframes. Through visual analysis of the road extraction results from nine satellite images, it can be seen that our method can clearly and completely extract road targets, and it can also effectively avoid false extraction for interference targets such as rivers. The application on different datasets shows that RFDNet exhibits good generalization ability for various region types.
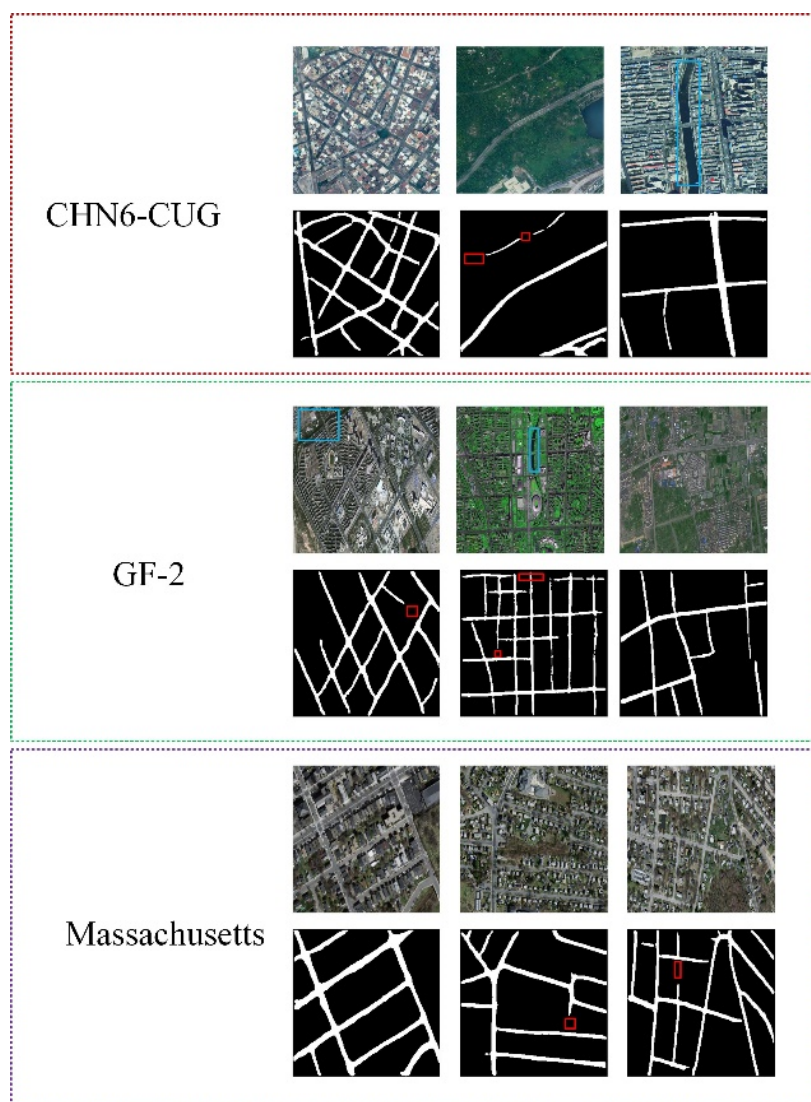
Figure 7. Public dataset test results.

Table 4. Dataset information.

Dataset	Band wavelength (μm)	Spatial resolution (m)	Launched country	Date
CHN6-CUG	0.45–2.35	1	China	2010
GF-2	0.45–0.90	0.8	China	2014
Massachusetts	0.52–0.90	1	America	1994

5. Conclusions

We constructed RFDNet to extract road information from remote sensing images based on a deep learning framework through a series of operations such as image acquisition, labeling, data enhancement, network structure design and optimization, model training and testing and morphological filtering to reduce noise and evaluate the extraction effect. In the selection of model type, considering the traditional full convolutional neural network (e.g., U-net) with more layers, which tends to cause slow model convergence, high computational cost consumption and poor ability

to recognize multi-scale information, we decided to use the more concise and efficient residual network. However, the residual neural network (e.g., ERFNet) relies too much on the residual structure and is prone to losing edge detail information and misclassifying linear features (e.g., identifying rivers as roads), which causes problems such as missing extraction and false extraction. For this reason, we have proposed RFDNet based on ERFNet. By designing the DoubleConv module and optimizing the network structure and learning strategy, RFDNet allows the model to learn more abstract deep semantic information so that the model can effectively fit roads with various complex backgrounds. In particular, the DoubleConv module increases the perceptual field of the model feature points without losing edge information, which improves the accuracy of road extraction; and, for similar linear features, DoubleConv identifies texture differences by applying successive double convolution operations, which can distinguish interference targets such as rivers. The output image of the network model usually has more noise points, and we use morphological filtering to perform the denoising operation to make the extracted roads clearer and smoother.

The final experimental results show that our method can achieve road extraction for different environmental backgrounds, and the accuracy of extraction is higher than that for U-net and ERFNet, as shown by it having the highest F1 and IoU scores in the evaluation results. Future research will focus on the design of an algorithm to extract road areas with severe occlusions; in addition, whether RFDNet can be applied in the field of medical image segmentation will also be our research direction.

Acknowledgments

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. F. Yi, R. Te, Y. Zhao, G. Xu, EUNetMTL: multitask joint learning for road extraction from high-resolution remote sensing images, *Remote Sens. Lett.*, **13** (2022), 258–268. https://doi.org/10.1080/2150704x.2021.2019344
2. Y. Li, H. Liang, G. Sun, Z. Yuan, Y. Zhang, H. Zhang, A land cover background-adaptive framework for large-scale road extraction, *Remote Sens.*, **14** (2022), 5114–5127. https://doi.org/10.3390/rs14205114

3. T. K. Behera, P. K. Sa, M. Nappi, S. Bakshi, Satellite IoT based road extraction from VHR images through superpixel-CNN architecture, *Big Data Res.*, **30** (2022), 100334–100346. https://doi.org/10.1016/j.bdr.2022.100334

4. D. Chang, Q. Wang, J. Yang, W. Xu, Research on road extraction method based on sustainable development goals satellite-1 nighttime light data, *Remote Sens.*, **14** (2022), 6015–6024. https://doi.org/10.3390/rs14236015

5. T. Alshaikhli, W. Liu, Y. Maruyama, Automated method of road extraction from aerial images using a deep convolutional neural network, *Appl. Sci.* **9** (2019), 4825–4840. https://doi.org/10.3390/app9224825

6. X. Chen, Q. Sun, W. Guo, C. Qiu, A. Yu, GA-Net: A geometry prior assisted neural network for road extraction, *Int. J. Appl. Earth Obs. Geoinf.*, **114** (2022), 103004–103015. https://doi.org/10.1016/j.jag.2022.103004

7. J. Dai, T. Zhu, Y. Wang, R. Ma, X. Fang, Road extraction from high-resolution satellite images based on multiple descriptors, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **13** (2020), 227–240. https://doi.org/10.1109/jstars.2019.2955277

8. H. Wang, F. Yu, J. Xie, H. Wang, H. Zheng, Road extraction based on improved Deeplabv3 plus in remote sensing image, *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **48** (2022), 67–72. https://doi.org/10.5194/isprs-archives-XLVIII-3-W2-2022-67-2022

9. Z. Zhang, X. Sun, Y. Liu, GMR-Net: Road-extraction network based on fusion of local and global information, *Remote Sens.* **14** (2022), 5476–5494. https://doi.org/10.3390/rs14215476

10. H. Tan, H. Xu, J. Dai, BSIRNet: A road extraction network with bidirectional spatial information reasoning, *J. Sens.* **2022** (2022), 1–11. https://doi.org/10.1155/2022/6391238

11. H. Huan, Y. Sheng, Y. Zhang, Y. Liu, Strip attention networks for road extraction, *Remote Sens.* **14** (2022), 4516–4533. https://doi.org/10.3390/rs14184516

12. F.Sultonov, J. H. Park, S. Yun, D. W. Lim, J. M. Kang, Mixer U-Net: An improved automatic road extraction from UAV imagery, *Appl. Sci.*, **12** (2022), 1953–1968. https://doi.org/10.3390/app12041953

13. G. Yuan, J. Li, X. Liu, Z. Yang, Weakly supervised road network extraction for remote sensing image based scribble annotation and adversarial learning, *J. King Saud Univ. Comput. Inf. Sci.*, **34** (2022), 7184–7199. https://doi.org/10.1016/j.jksuci.2022.05.020

14. H. Chen, S. Peng, C. Du, J. Li, S. Wu, SW-GAN: Road extraction from remote sensing imagery using semi-weakly supervised adversarial learning, *Remote Sens.*, **14** (2022), 4145–4160. https://doi.org/10.3390/rs14174145

15. K. Geng, X. Sun, Z. Yan, W. Diao, X. Gao, Topological space knowledge distillation for compact road extraction in optical remote sensing images, *Remote Sens.*, **12** (2020), 3175–3195. https://doi.org/10.3390/rs12193175

16. Y. Li, L. Xiang, C. Zhang, F. Jiao, C. Wu, A guided deep learning approach for joint road extraction and intersection detection from RS images and taxi trajectories, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **14** (2021), 8008–8018. https://doi.org/10.1109/jstars.2021.3102320

17. P. Li, Y. Li, J. Feng, Z. Ma, X. Li, Automatic detection and recognition of road intersections for road extraction from imagery, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **43** (2020), 113–117. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-113-2020

18. E. Romera, J. M. Álvarez, L. M. Bergasa, R. Arroyo, ERFNet: Efficient residual factorized convnet for real-time semantic segmentation, *IEEE Trans. Intell. Transp. Syst.*, **19** (2017), 263–272. https://doi.org/10.1109/tits.2017.2750080

19. C. I. Cira, M. Kada, M. Á. Manso-Callejo, R. Alcarria, B. Bordel Sanchez, Improving road surface area extraction via semantic segmentation with conditional generative learning for deep inpainting operations, *ISPRS Int. J. Geo-Inf.*, **11** (2022), 43–61. https://doi.org/10.3390/ijgi11010043

20. T. Duan, Y. Liu, J. Li, Z. Lian, Q. Li, DuFNet: Dual flow network of real-time semantic segmentation for unmanned driving application of internet of things, *Comp. Model. Eng. Sci.*, **136** (2023), 223–239. https://doi.org/10.32604/cmes.2023.024742

21. C. Sun, H. Zhao, L. Mu, F. Xu, L. Lu, Image semantic segmentation for autonomous driving based on improved U-Net. *Comp. Model. Eng. Sci.*, **136** (2023), 787–801. https://doi.org/10.32604/cmes.2023.025119

22. R. Xu, Y. Zeng, A method for road extraction from high-resolution remote sensing images based on multi-kernel learning, *Information*, **10** (2019), 385–398. https://doi.org/10.3390/info10120385

23. J. Zhang, Y. Li, Y. Si, B. Peng, F. Xiao, S. Luo, et al., A low-grade road extraction method using SDG-DenseNet based on the fusion of optical and SAR images at decision level, *Remote Sens.*, **14** (2022), 2870–2894. https://doi.org/10.3390/rs14122870

24. K. Zhou, Y. Xie, Z. Gao, F. Miao, L. Zhang, FuNet: A novel road extraction network with fusion of location data and remote sensing imagery, *ISPRS Int. J. Geo-Inf.*, **10** (2021), 39–57. https://doi.org/10.3390/ijgi10010039

25. G. P. Cardim, E. A. D. Silva, M. A. Dias, I. Bravo, A. Gardel, Statistical evaluation and analysis of road extraction methodologies using a unique dataset from remote sensing, *Remote Sens.*, **10** (2018), 620–636. https://doi.org/10.3390/rs10040620