

Automatic Road Extraction from High-Resolution Remote Sensing Images Using a Method Based on Densely Connected Spatial Feature-Enhanced Pyramid

Qiangqiang Wu^{ID}, Feng Luo, Penghai Wu^{ID}, Member, IEEE, Biao Wang^{ID}, Hui Yang^{ID}, and Yanlan Wu^{ID}

Abstract—Road extraction is an important task in remote sensing image information extraction. Recently, deep learning semantic segmentation has become an important method of road extraction. Due to the impact of the loss of multiscale spatial features, the results of road extraction still contain incomplete or fractured results. In this article, we proposed a deep learning model, which is called the dense-global-residual network that reduces the loss of spatial information and enhances context awareness. In the dense-global-residual network, the residual network is used to extract the features at different levels. To obtain more abundant multiscale features, a dense and global spatial pyramid pooling module based on Atrous Spatial Pyramid Pooling is built to perceive and aggregate the contextual information. The proposed method obtains better results on the GF-2 road dataset and public Massachusetts road dataset of aerial imagery. In order to prove the effectiveness of our method, we compared with four methods, such as DeepLabV3+, U-net, D-LinkNet, and coord-dense-global model, and found that the accuracy of our method is considerably better. Moreover, the dense-global-residual network can also effectively extract roads, especially trees and building shadows that occlude the road. In addition, our method can successfully extract roads in regions of different development levels in universality experiments. This indicates that the proposed method can effectively maintain the completeness and continuity of roads and improve the accuracy of road segmentation from high-resolution remote sensing images.

Index Terms—Deep learning, dense and global spatial pyramid pooling (DGSP) module, remote sensing image, road extraction.

Manuscript received July 18, 2020; revised September 14, 2020 and November 8, 2020; accepted November 23, 2020. Date of publication December 7, 2020; date of current version January 6, 2021. This work was supported in part by National Natural Science Foundation of China under Grant 41971311 and Grant 41901282, in part by Major Projects of Science and Technology of Anhui under Grant 18030801111, and in part by the Natural Science Foundation of Anhui Province under Grant 200805QD188. (Corresponding authors: Yanlan Wu; Hui Yang.)

Qiangqiang Wu, Penghai Wu, and Biao Wang are with the School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China (e-mail: x18301096@stu.ahu.edu.cn; wuph@ahu.edu.cn; wangbiao-rs@ahu.edu.cn).

Feng Luo is with CCCC Second Highway Consultants Company Ltd., Wuhan Economic & Technological Development Zone, Hubei 430056, China (e-mail: lfengxx@126.com).

Hui Yang is with the Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China (e-mail: yanghui@ahu.edu.cn).

Yanlan Wu is with the Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei 230601, China, with the School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China, and also with the Anhui Engineering Research Center for Geographical Information Intelligent Technology, Hefei 230601, China (e-mail: wuyanlan@ahu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3042816

I. INTRODUCTION

Road extraction from remote sensing images is an important topic in modern society. It is of great significance to traffic management, urban planning, and map updating [1]–[3]. However, due to the complex road background and the similarity of the spectrum, high-precision road extraction is still difficult to obtain.

The methods for remote sensing image road extraction are mainly divided into feature-based, classification-based, and deep learning approaches [4], [5]. Feature-based methods mainly extract roads from images by considering their features, such as road extraction methods based on shape [6], texture [7]–[9], and geometry [10], [11]. Feature-based methods have a good effect on simple and regular road extraction, but they have poor extraction effects on complex roads and require substantial post-processing to repair the initially extracted roads. Classification-based approaches such as the maximum likelihood methods (ML classification methods) [12], support vector machine methods (SVM classification methods) [13]–[15], Markov random fields classifier methods (MRF classification methods) [16]–[18], and mean shift-based methods [19] extract road fragments from the image and further refine them by customizing rules that are based on the spectral and spatial features of the road. Although the accuracy is better than that of the feature-based methods, the extraction results depend largely on the accuracy of classification rules, which require manual design. Due to the spectral similarity of roads, buildings, parking lots, and other objects, the extraction accuracy is not high.

Deep learning technology has made remarkable achievements in the fields of computer vision and artificial intelligence [20]–[22], and more researchers have applied this method to remote sensing image information extraction [23]–[29]. The fully convolutional network (FCN) proposed by Long *et al.* [30] greatly advanced image segmentation by using standard convolutional layers to replace the fully connected layers. There are increasingly more applications of fully convolutional neural networks based on FCN extensions, especially in road extraction [31]–[34]. However, due to the loss of spatial features caused by pooling in the downsampling process and the underutilization of features caused by simple convolution, FCN has difficulty in completely recovering the resolution of the input feature maps.

To solve this problem and to apply a deeper network to improve network performance, the residual network (ResNet) [35] and the densely connected convolutional network (DenseNet) [36] were proposed. They both strengthen the transmission of upper and lower information flow and improve the feature reuse rate. Zhang *et al.* [37] combined ResNet and U-net [38] to extract road networks, which are designed with few parameters but achieve good performance. Recently, Wang *et al.* [39] proposed coord-dense-global model (CDG), which uses DenseNet, the coordconv module, and the global attention module to construct the road map from remote sensing imagery. Although these FCN-based methods perform well in road extraction, they perform poorly when nonroad objects occlude the road. Because they are affected by a complex background, it is difficult to consider multiscale features in downsampling. Therefore, multiscale contextual features (semantic information) are critical to the integrity of road extraction. The local detail features help to accurately segment roads, while the global context features help reduce misclassification and maintain road connectivity. In the field of semantic segmentation, there are two mainstream methods used to gather contextual information by extracting multiscale features [40]–[43]. One method uses a U-shaped structure with skip connections; FCNs based on a U-shaped structure obtain contextual information by connecting multi-level information [38], [41]. The other method uses dilated convolutions (or atrous convolutions) to increase the receptive field and extract multiscale features. For example, Chen *et al.* [42] designed parallel dilated convolutions with different dilation rates to obtain more multiscale contextual information in DeepLabv3+. Zhou *et al.* [43] used dilated convolution with a combination of parallel and cascading modes to expand the receptive field and aggregate the contextual information and ultimately improved the accuracy of road result segmentation and achieved superior performance in the CVPR DeepGlobe 2018 Road Extraction Challenge. Although these two methods can extract multiscale features to obtain the context information, there are still some problems. The method based on the U-shaped structure is insufficient in extracting the target features, resulting in an incomplete consideration of context information. The dilated convolution-based methods cause a loss of spatial information due to the continuously dilated convolution, which leads to a “chessboard effect.” Moreover, because atrous spatial pyramid pooling (ASPP) is located at the bottom of the network, it is effective for large-scale target feature extraction but loses small-scale targets.

To further strengthen the spatial information, we propose a deep learning model, called the dense-global-residual network (DGRN), which reduces the loss of spatial information and enhances the context awareness. ResNet has powerful feature reuse capabilities. In the DGRN, a ResNet is used to extract the features at different levels. To obtain more abundant multiscale features, we built a dense and global spatial pyramid pooling (DGSPP) module based on ASPP to perceive and aggregate the contextual information. In addition, for the road extraction of GF-2, we constructed a standard semantic segmentation sample dataset of extracted road by the GF-2 remote sensing images.

II. METHODOLOGY

The road network of remote sensing image has the characteristics of rich local detailed information and large distribution span of the global road network. Moreover, it is affected by complex backgrounds, such as backgrounds that include the shadows of buildings. Relevant local details can effectively distinguish roads from the surrounding ground features. Therefore, it is very important to maintain detailed spatial information for road extraction. ResNet has effective feature extraction capabilities. The model in this article uses improved ResNet as the main network, which can make full use of road details. The ASPP-based method [42], [44] is an effective means to extract multiscale information; it has achieved state-of-the-art performance in target extraction tasks. In order to extract target features of different scales and levels, we designed DGSPP and introduced the coding part of the network to enhance the spatial information perception of ResNet and improve the accuracy of road extraction.

A. Dense-Global-Residual Network

The overall network is shown in Fig. 1. The network consists of three parts: encoding, bridge, and decoding. The encoder part consists of four residual blocks and a DGSPP. The DGSPP is located between the third residual block and the fourth residual block and is used to extract the multiscale features of the road, construct the spatial feature pyramid, and aggregate the context information, as shown in the green box in the coding part of Fig. 1. Because contextual information plays an important role in the quality of the extraction results [45], the bridge part is composed of downsampling, upsampling, and residual blocks to supplement the extraction of contextual information. There are five residual blocks in the decoding part. In the decoding process, the high-level feature information and the low-level features of the corresponding layer of the encoding path are concatenated to enhance the feature information. Eventually, the network output will obtain a segmentation map of the same size as the input image. In addition, to improve the generalization ability of the model, we have improved the residual unit.

B. Improved Residual Units

The residual unit consists of a batch normalization (BN) layer [46], a rectified linear unit (ReLU) [47], a convolution layer, and a shortcut connection layer. Because BN has strict requirements on batch size, when the batch size is small, the error of normalization will increase rapidly, weakening the model’s generalization ability. Due to memory limitations, the batch size often fails to meet the BN requirement during training. Compared to BN, the calculation of the group normalization (GN) [48] is independent of the batch size. When the batch size is small, the accuracy is stable, and the generalization ability is very strong. Therefore, we have improved the residual unit. In this article, we change the BN into the GN according to the actual situation. The formula of the residual unit is defined as follows:

$$y_l = h(x_l) + F(x_l, w_l) \quad (1)$$

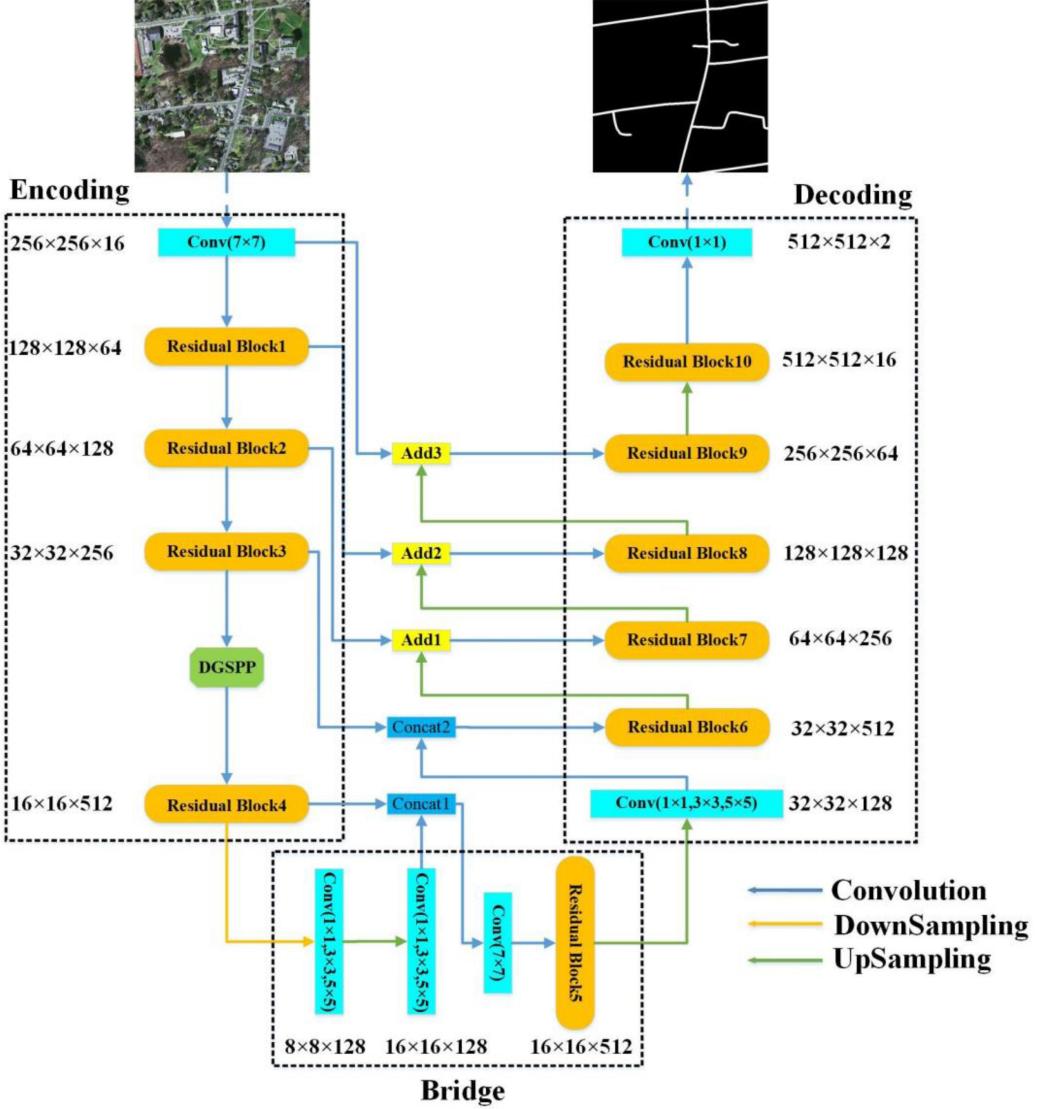


Fig. 1. Architecture of the dense-global-residual model proposed in this article.

$$x_{l+1} = f(y_l) \quad (2)$$

where x_l and x_{l+1} represent the input and output in residual unit l , $F(\cdot)$ represents the residual function, $f(\cdot)$ represents the activation function, and $h(\cdot)$ represents the identity mapping function.

C. Dense and Global Spatial Pyramid Pooling Module

To solve the problem of small-scale detailed feature loss of ASPP and further improve the perception ability of the spatial information, we proposed the DGSPP module to build a spatial feature pyramid and to perceive and extract the multiscale road features. Compared with ASPP, DGSPP can obtain a larger receptive field, more multiscale features, and richer context information. The specific improvements are as follows: The DGSPP replaces the parallel connections in the traditional ASPP with dense connections. For each layer, the output feature maps of all preceding layers are used as inputs, and its own output is

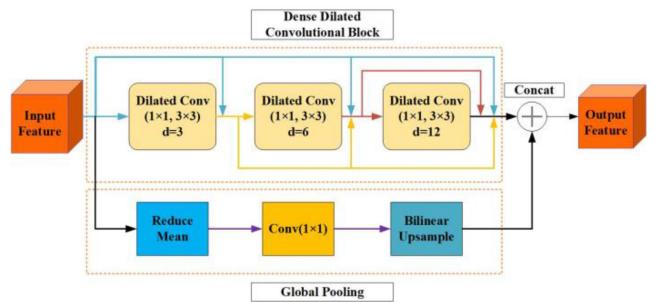


Fig. 2. Structure of dense global atrous spatial pyramid pooling based on atrous spatial pyramid pooling. $\text{Dilated Conv}(1 \times 1, 3 \times 3)$ represents a 1×1 dilated convolution and 3×3 dilated convolution, and d represents the dilation rate.

used as the input of all subsequent layers, as shown in Fig. 2. This design effectively connects the dilated convolution layers by combining the advantages of parallel and cascading modes,

which not only obtains larger receptive fields and rich multiscale information but also generates denser and larger feature pyramids, as shown in part of “Dense Dilated Convolutional Block” in Fig. 2. Therefore, this method is more adaptable for roads with multiscale features. The final output is a feature map generated by multirate and multiscale convolution that covers the global semantic information and local detailed information. Global context feature information [49]–[54] is very important for semantic segmentation, and it helps to classify pixels correctly and can improve the completeness and accuracy of the extraction results. Global pooling has the ability to improve global context awareness [55]; so we retain the global pooling branch to avoid the ASPP kernel degradation problem and supplement the global information with the output in another way, as shown in part of “Global Pooling” in Fig. 2. DGSPP is composed of a dense dilated convolutional block and global pooling, as shown in Fig. 2.

From the above, the DGSPP consists of densely connected dilated convolutions and global pooling. The two parts can be formulated as follows:

$$X_L = F_{K,d_l}([X_0, X_1, X_2, \dots, X_{L-1}]) \quad (3)$$

where X_L represents the received total feature maps in layer L , and $[X_0, X_1, X_2, \dots, X_{L-1}]$ represents the feature map formed by concatenating the outputs from all previous layers. F_{K,d_l} represents dilated convolution, d_l represents the dilation rate of layer l , and K stands for the size of dilated convolution

$$Y = X_L + F(X_0) \quad (4)$$

where Y represents the final output of DGSPP, X_L represents the feature map of densely connected atrous convolution, and $F(\cdot)$ is defined as a composite function of two consecutive operations: global average pooling and a 3×3 convolution layer.

D. Implementation Details of Network

The network consisted of three parts: 1) encoding, 2) bridge, and 3) decoding. These three parts are mainly composed of 10 residual blocks and DGSPP. At the top of the network, a 7×7 convolution layer with stride = 2 was used to extract the features as the initial input. In the encoder part, the residual block mainly contained a 3×3 convolution layer with stride = 2 and two 1×1 convolution layers with stride = 1. We designed DGSPP to obtain the multiscale features and contextual information. Correspondingly, the decoding path also comprises five residual blocks, in which the strides of the convolutions are all 1. In the decoding part, the low-level features are upsampled and connected to the feature maps of the corresponding encoding path. The middle part is the bridge connecting the encoding and decoding paths. At the bottom of the network, a 1×1 convolution layer with a ReLU activation function was used to output the final feature maps with the same resolution as the input image. The detailed parameters and output size of each unit are presented in Table I.

TABLE I
IMPLEMENTATION DETAILS OF DENSE-GLOBAL-RESIDUAL NETWORK

	Unit	Stride	Output size
Input			$512 \times 512 \times 3$
Encoding	Conv(7×7)	2	$256 \times 256 \times 16$
	Residual Block1	2	$128 \times 128 \times 64$
	Residual Block2	2	$64 \times 64 \times 128$
	Residual Block3	2	$32 \times 32 \times 256$
Bridge	Residual Block4	2	$16 \times 16 \times 512$
	Residual Block5	1	$16 \times 16 \times 512$
	Residual Block6	1	$32 \times 32 \times 512$
Decoding	Residual Block7	1	$64 \times 64 \times 256$
	Residual Block8	1	$128 \times 128 \times 128$
	Residual Block9	1	$256 \times 256 \times 64$
Output	Residual Block10	1	$512 \times 512 \times 16$
	Conv(1×1)	1	$512 \times 512 \times 2$



Fig. 3. Examples for GF-2 road dataset images.

III. EXPERIMENTS

A. Datasets

Gaofen-2 (GF-2) images are an important part of high-resolution remote sensing imaging. To verify the performance of the model, we made a GF-2 road dataset based on GF-2 remote sensing satellite images, as shown in Fig. 3. Among them, we produced 28 GF-2 images distributed in Hefei, China and Tianjin, China, including 16 images of size 4909×4672 and 12 images of size 4578×4442 with a resolution of 1 m. We use manual labeling to create the road labels. We divided these 28 images into a training dataset (20 images), a validation dataset (2 images), and a test dataset (6 images). The road labels of the training dataset are used by the model to learn and understand the features of the road. In order to effectively learn the road features, we train the model as fully as possible. This article crops the training set and validation set to increase the number of samples. Finally, the numbers of images used for training and validating are 5892 and 1480, respectively, and these images have a size of 512×512 . After the training, we tested four 4578×4442 remote sensing images of Hefei, China, and two 4909×4672 remote sensing images of Tianjin, China.

The Massachusetts roads dataset [56] is the largest publicly available road dataset in the world. The dataset contains



Fig. 4. Examples for Massachusetts roads dataset images.

1171 images, which are divided into 1108 training images, 14 validation images, 49 test images, and corresponding label images, as shown in Fig. 4. The size of each image is 1500×1500 pixels, and the resolution is 1.2 m/pixel. The dataset contains various features, such as roads, grasslands, forests, and buildings.

Because the number of images in the dataset is too small, it is not conducive to the full training of the model. In this article, the dataset is cropped to increase the number of samples. The remote sensing images of the training and validation sets and the corresponding label images are cut into 512×512 pixel scale image samples (the entire image is first cut in order and then randomly cropped on the image). According to the rules of dataset division, all the divided sample data are randomly divided into a new training set and a validation set with a 4:1 ratio. When examining samples, we delete a small number of interfering images and their corresponding label images. Finally, the training set contains 14 366 images of 512×512 pixels, and the validation set contains 3592 images of 512×512 pixels.

B. Implementation Details

In this article, the proposed model uses TensorFlow as a deep learning framework, the development platform uses JetBrains PyCharm 2017, and the development language is Python. All models are trained and tested on a computer configured with an Intel Core(TM) i9-7980XE CPU and an NVIDIA GeForce GTX 1080 Ti graphics card.

The model in this article uses a cross-validation training method, that is, the training set and the validation set are entered into the model at the same time. Each training randomly selects the batch size data of the validation set to calculate the loss and accuracy and optimize the training of the model.

Because training the model has large GPU memory requirements, the method in this article takes the image with the size of 512×512 as the input of the network. Adam [57] is an adaptive learning rate optimizer with high computational efficiency and low memory requirements. Therefore, this article uses the Adam optimizer to optimize the network and update the parameters. In addition, the network proposed in this article uses binary cross entropy (BCE) + dice coefficient loss [43] as the loss function, the batch size is set to 4, the number of epochs is 50,

the number of iterations per round is 4000, and the initial learning rate is set to 0.001. To better train the model, the learning rate will automatically adjust with an increase in training epochs to accelerate the convergence of the network combined with the optimization. In the test process, we directly test the 1500×1500 pixel images in the test dataset, and it takes approximately 2–3 s for each image.

C. Comparative Studies Using Different Networks and Evaluation Metrics

This article proposes an improved semantic segmentation model that implements automatic road extraction of high-resolution remote sensing images. To verify the performance of this method, we compared the DGRN with four representative deep learning network models of U-net [38], DeepLabV3+ [42], D-LinkNet [43], and CDG [39] on two datasets under the same conditions. The U-net network model is mainly used for medical image segmentation. It can use deep feature localization and shallow feature segmentation for accurate segmentation; so it has better performance for the extraction of slender road features. DeepLabV3+ is a new full convolution network model recently proposed by scholars, which further improves the performance of semantic segmentation tasks. D-LinkNet won the CVPR 2018: DeepGlobe Road Extraction Challenge. CDG is a novel road extraction algorithm.

To quantify the effect of road extraction, in this article, we use the most common evaluation metrics in the field of semantic segmentation: Recall, IoU, and F1 score. At the same time, the first two indicators correspond to the completeness and quality of classic road extraction [58], [59]. They can evaluate the geometric quality of the extracted road network [60], [61]. This article considers road extraction as a two-class classification problem, and the prediction results are divided into two categories: road and nonroad. For the binary classification problem, the sample data can be divided into true positive (TP), false positive (FP), true negative (TN), and false negative (FN) according to the combination of the real category and prediction category. The recall metric represents the correct pixels over the ground truth, while the precision metric represents the correct pixels over the prediction result. The F1 score is a powerful evaluation metric for the harmonic mean of the precision and recall metrics. IoU is the ratio of the intersection and union of real and predicted values in different categories and adopted to evaluate the shape and area [62]. The evaluation metric formulas are as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$F1 = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \quad (7)$$

$$IoU = \frac{\text{TP}}{\text{FN} + \text{TP} + \text{FP}}. \quad (8)$$

In order to comprehensively evaluate the quality of the road network, this article uses topology criteria (Topological

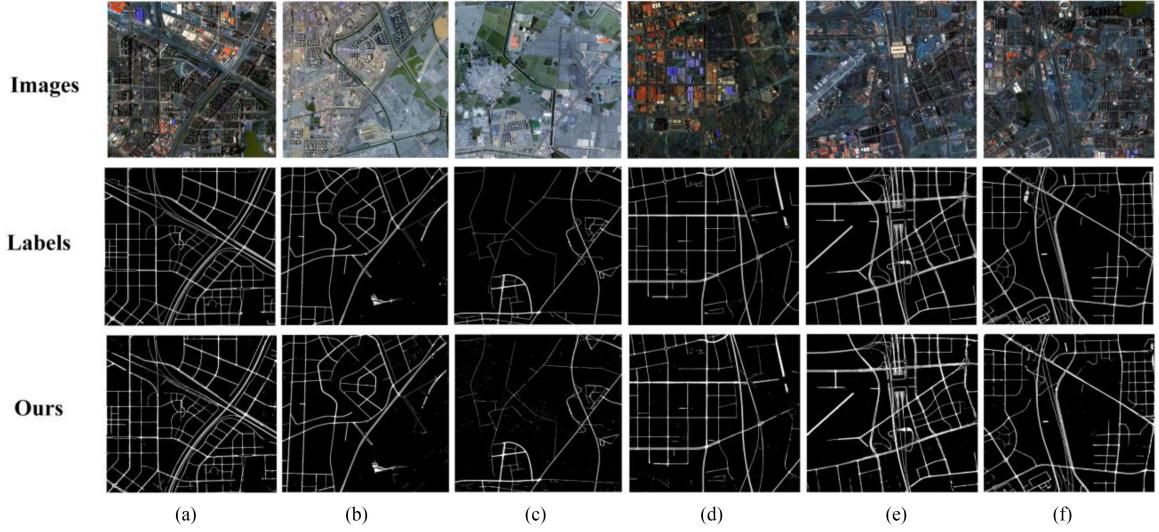


Fig. 5. Results of the GF-2 road test dataset.

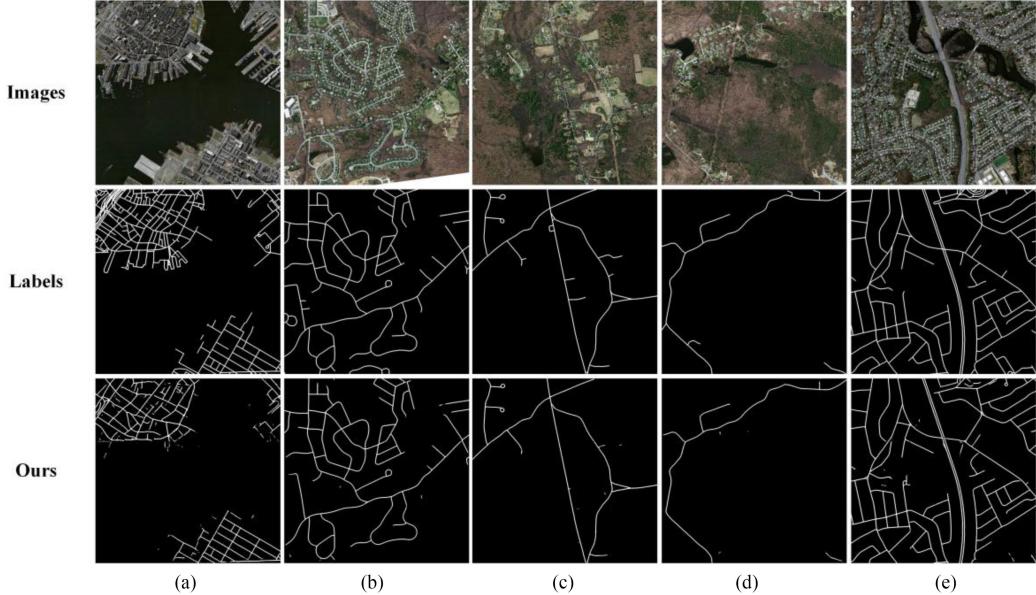


Fig. 6. Demonstration of the experimental results of the Massachusetts roads dataset. The accuracy of (a–e) decreases in turn.

Completeness and Topological Correctness) [60], [61] to measure the connectivity of the road.

Topological completeness's definition is shown in the following equation:

$$\text{Topological Completeness} = \frac{n_{LE}}{n_L} \quad (9)$$

where n_L represents all pairs of connected nodes in the label image and n_{LE} is the pairs of nodes that are connected at the same time in the road network of the two images.

Topological correctness's definition is shown in the following equation:

$$\text{Topological Correctness} = \frac{n_{EL}}{n_E} \quad (10)$$

where n_E represents all pairs of connected nodes in the label image and n_{EL} is the pairs of nodes that are connected at the same time in the road network of the two images.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Results

After the training, the test set data of the GF-2 dataset and Massachusetts roads dataset were used for testing. From the perspective of qualitative analysis, we can see that each image of the two test sets has a good result, and the road is completely extracted and is located very close to the real label of the ground, as shown in Figs. 5 and 6. To quantitatively verify the performance of the method in this article, Table II lists the accuracy of the test images of the GF-2 dataset, and Table III lists the

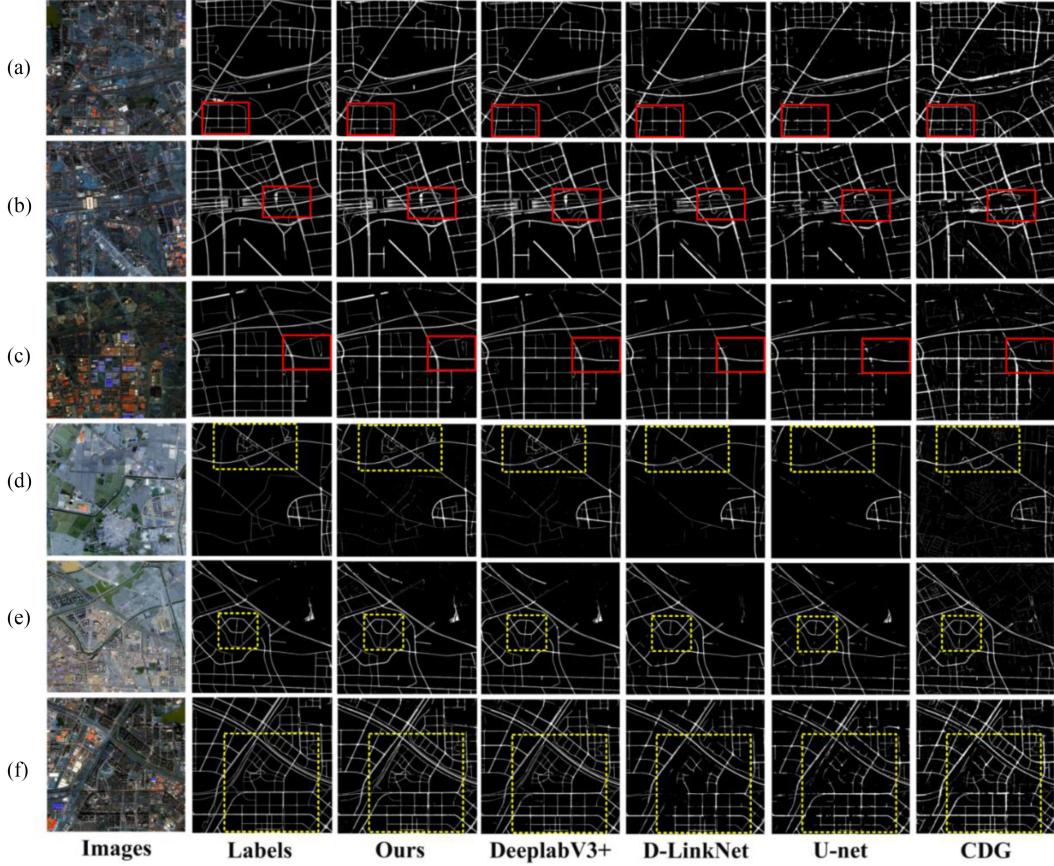


Fig. 7. Prediction results of CDG, U-net, D-LinkNet, DeepLabV3+, and the proposed dense-global-residual network (DGRN) on the GF-2 test images.

TABLE II
TEST RESULTS OF THE GF-2 ROADS DATASET

Test Datasets	IoU	Recall	F1
all images	80.39%	88.11%	89.09%
a	79.91%	87.03%	88.83%
b	84.10%	91.39%	91.36%
c	77.05%	85.05%	87.04%
d	82.77%	89.93%	90.57%
e	82.64%	90.02%	90.49%
f	75.85%	85.23%	86.27%

accuracy of the test images of the Massachusetts roads dataset. As we can see in Table II, the overall accuracy (IoU, Recall, and F1 score) of the test results are 80.39%, 88.11%, and 89.09%, respectively. For all test images of the GF-2 dataset, the recall and comprehensive evaluation index (F1 score) exceeds 85%, and the accuracy of image5 is even higher than 91%. Although the IoU is lower than the other two evaluation indicators, each image still achieves good accuracy. As seen from Table III, the overall accuracy (IoU, Recall, and F1 score) of the test results are 62.48%, 71.97%, and 76.59%, respectively. Among them,

the IoU index of the highest and lowest accuracy results are 75.29% and 38.36%, respectively. In summary, the results of the experiment preserved the continuity of the road, and there were no mistakes, which show that the proposed model performs well in the task of road extraction of high-resolution remote sensing images.

B. Comparisons and Analysis

Fig. 7 shows the test results of all the methods on the GF-2 road dataset. On the whole, the DGRN extraction results are very complete and consistent with the label; the results contain no incorrect extractions, and the method has a great advantage over the other four models, as shown in the yellow box in Fig. 7. It can be seen from the red marked box in Fig. 7 that the extraction results of the methods DeepLabV3+ and D-LinkNet are broken. In Fig. 7(c), the extraction result of the D-LinkNet method directly shows the phenomenon of extraction failure. Among the four models, the extraction results of the CDG model are the worst. The degree of fragmentation of the extraction results and the number of extraction failures are the largest among all images, which is also consistent with the results of our quantitative analysis. Table IV shows the accuracy statistics of all models. The topological quality index accuracy of our method is higher than other methods. This is basically consistent with our overall

TABLE III
TEST RESULTS OF THE MASSACHUSETTS ROADS DATASET

Test Datasets	IoU	Recall	F1
All images-average	62.48%	71.97%	76.59%

TABLE IV
COMPARISONS BETWEEN THE PROPOSED DGRN AND SEVERAL TYPICAL MODELS ON THE GF-2 DATASET

Model	IoU	Recall	F1	Topological Correctness	Topological Completeness
CDG	54.11%	74.61%	70.02%	66.00%	76.00%
U-net	56.32%	61.10%	71.86%	78.00%	63.00%
D-LinkNet	63.72%	68.76%	77.67%	80.00%	67.00%
DeepLabV3+	79.24%	83.71%	88.38%	83.00%	82.00%
Ours	80.39%	88.11%	89.09%	85.00%	87.00%

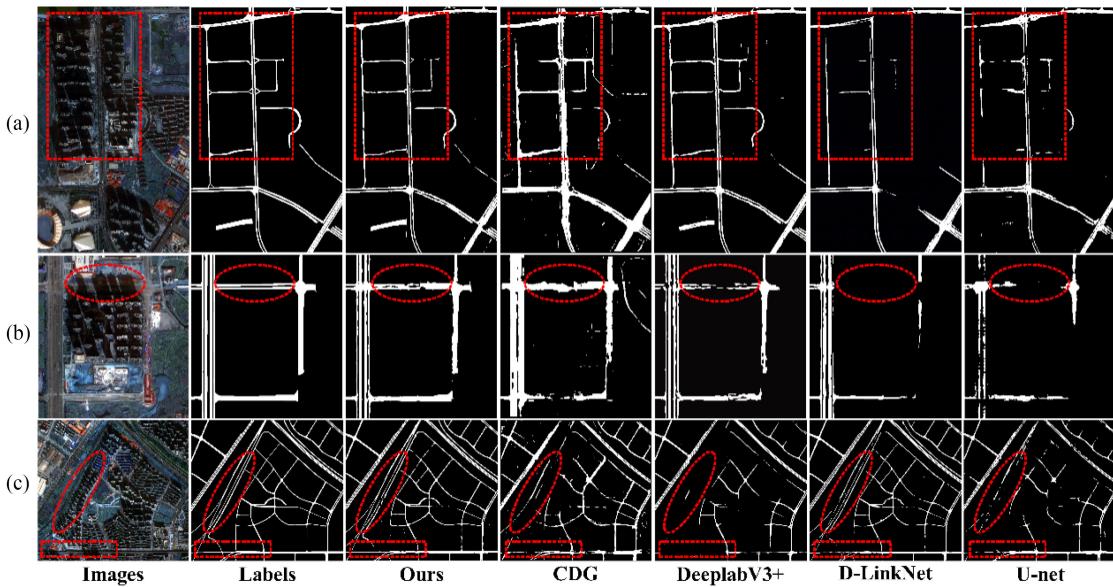


Fig. 8. Comparison of road extraction results under the influence of building shadows. From left to right are remote sensing images, road true value, our extraction results, DeepLabV3+ extraction results, D-LinkNet extraction results, and U-Net extraction results.

evaluation of road continuity. As seen from Table IV, the extraction accuracy of the DGRN reaches 80.39%, 88.11%, and 89.09% on the three evaluation indexes of IoU, Recall, and F1, respectively. The accuracy values of the DGRN given by each evaluation index are better than those of all the other comparison methods. Compared to the results of the high-precision DeepLabV3 + method, the IoU, Recall, and F1 results of the DGRN are 1.15%, 4.40%, and 0.71% better, respectively, and they are much better than those of the other three comparison methods.

In the process of urban road image extraction, the results are usually greatly affected by high-level shadows; so the extraction effect is very poor. However, our model can effectively alleviate such problems, as shown in Fig. 8. Fig. 8 is a partially enlarged image of the extraction results, reflecting the performance of all the extraction methods in overcoming the shadows of buildings. From Fig. 8, we can see that the DGRN method can well

solve the problem caused by shadow occlusion of high-rise buildings and can ensure the continuity and integrity of the road. Compared with the other four methods, the DGRN extraction results are better. Among them, the CDG and DeepLabV3+ have a certain extraction ability for the roads occluded by the shadows of high-rise buildings, but the extraction results are severely fragmented. U-net and D-LinkNet have poor resistance to shadow interference, and most roads occluded by shadows have failed to be extracted. The above qualitative analysis is consistent with our quantitative analysis results. As shown in Table V, our method surpassed other methods in all indicators. From the road extraction results of the five methods, we can see that due to the addition of DGSP to the DGRN network, the network can make good use of the feature information of the road in the coding process. At the same time, the feature reuse mechanism of DGSP can reduce the loss of information in the coding process, improve the utilization rate of the road feature

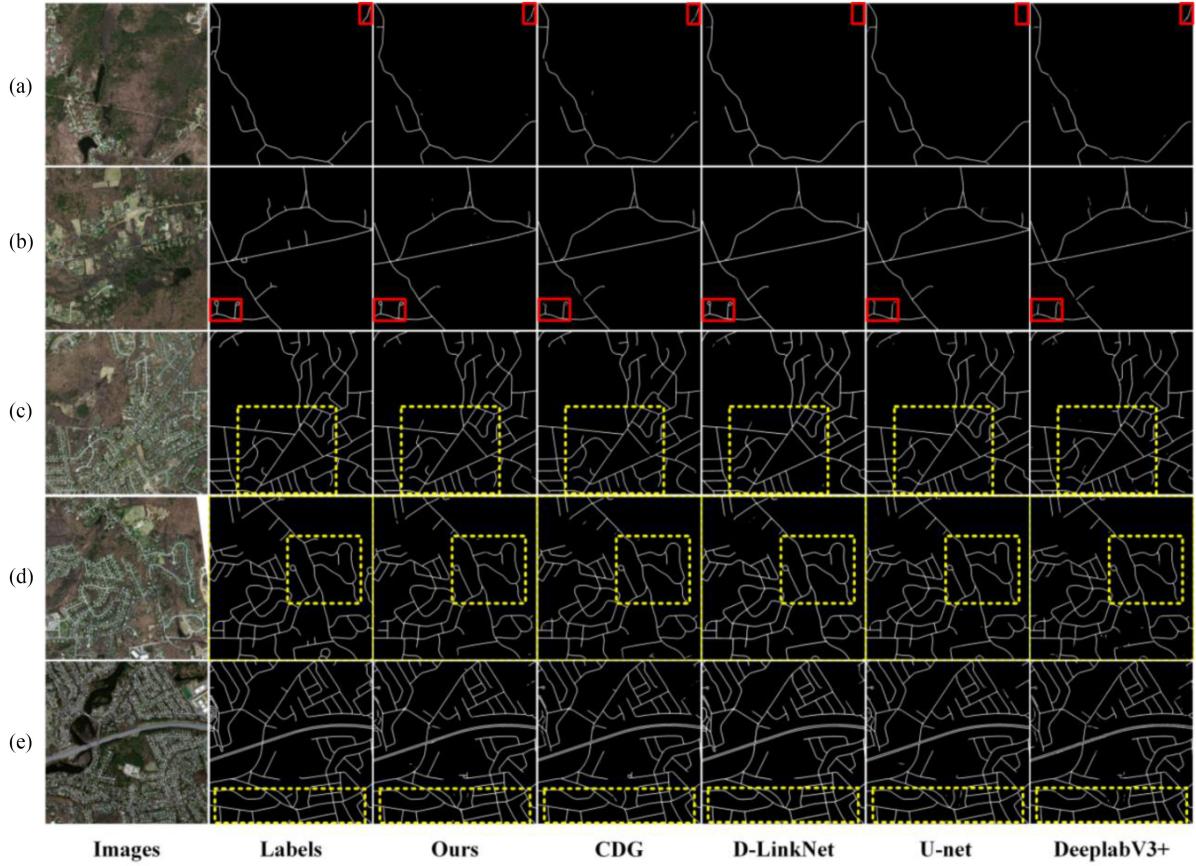


Fig. 9. Prediction results of all methods with the Massachusetts roads dataset.

TABLE V
COMPARISON OF ROAD EXTRACTION RESULTS UNDER THE INFLUENCE OF
BUILDING SHADOWS

Model	IoU	Recall	F1
U-net	40.00%	43.34%	56.93%
D-LinkNet	42.73%	45.32%	59.87%
CDG	47.50%	63.93%	64.27%
DeepLabV3+	66.23%	69.11%	79.66%
Ours	68.51%	74.26%	81.24%

information, and finally improve the accuracy of the extracted roads, which also shows that the DGRN is an advanced new method for road extraction.

The DGRN method has achieved superior performance on the GF-2 road dataset. To verify the effectiveness and advancement of the DGRN, we further verified the DGRN model on the Massachusetts roads dataset. The results of the Massachusetts test dataset are shown in Fig. 9. The overall extraction result of the DGRN is also the most complete and continuous, as shown in the yellow box in Fig. 9. Almost all roads in the test image are extracted, and there are few broken roads. From the red edit box in Fig. 9, we can see that the results of the other three methods have good performance overall and are able to extract a more complete road contour. However, the road detail extraction performance is insufficient, and there are many fracture

discontinuities in the results. Among them, the D-LinkNet method [see the red box in Fig. 9(a) and (c)] and U-net method [red box in Fig. 9(a)] extraction results exhibited partial road extraction failure phenomena. Although DeepLabV3+ did not fail to extract the roads, the degree of fragmentation of its extraction results was the most severe. The overall performance of CDG is already great, but there are still road breakage problems. At the same time, to better analyze the performance of the model, we conducted a quantitative analysis and listed all the statistical indicators in Table VI. The topological criteria of our method are higher than other methods. This is basically consistent with our qualitative analysis of road continuity. From Table VI, we can see that our accuracy is the highest on the three evaluation indicators IoU, Recall, and F1. Compared with the CDG method with the highest precision of the comparison methods, our method obtained result accuracies that are 0.58%, 0.17%, and 0.49%, higher than the result accuracies obtained by CDG; our result accuracies are also much higher than those of the other two road extraction methods. On the Massachusetts roads dataset, the DGRN method has certain advantages over other network models in road extraction tasks.

In addition, the model also has a great performance in solving the problem caused by tree occlusion. Fig. 10 is a partial magnification of the Massachusetts roads data test image, reflecting the resolution of the tree occlusion problem by all methods. As seen from Fig. 10, the DGRN method performs well in solving

TABLE VI
QUANTITATIVE COMPARISONS OF THE DIFFERENT METHODS WITH THE MASSACHUSETTS ROADS DATASET

Model	IoU	Recall	F1	Topological Correctness	Topological Completeness
DeepLabV3+	51.95%	60.22%	67.64%	79.00%	62.00%
U-net	56.91%	65.74%	72.11%	80.00%	66.00%
D-Linknet	60.71%	71.96%	75.15%	81.00%	72.00%
CDG	61.90%	71.80%	76.10%	82.00%	72.00%
Ours	62.48%	71.97%	76.59%	85.00%	74.00%

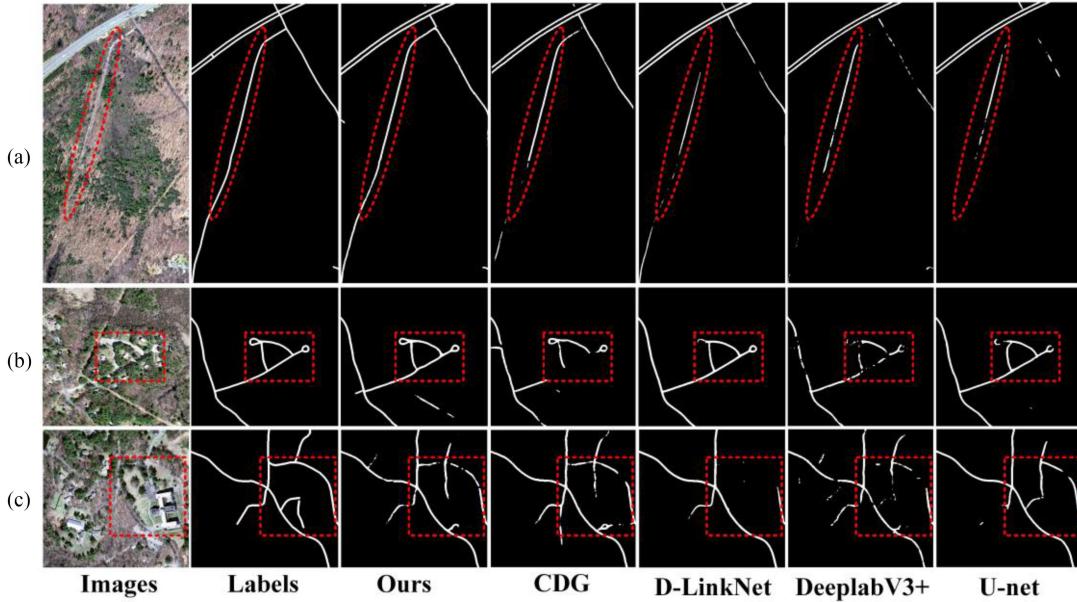


Fig. 10. Comparison of road extraction results under the influence of trees occlusion. From left to right are the remote sensing images, road true value, our extraction results, D-LinkNet extraction results, DeepLabV3+ extraction results, and U-Net extraction results.

TABLE VII
COMPARISON OF ROAD EXTRACTION RESULTS UNDER THE INFLUENCE OF TREES OCCLUSION

Model	IoU	Recall	F1
DeepLabV3+	48.54%	55.01%	65.19%
U-net	51.99%	59.91%	67.47%
D-LinkNet	59.07%	65.73%	73.62%
CDG	59.24%	68.89%	74.25%
Ours	66.32%	79.66%	79.55%

the problem of tree occlusion, and all roads blocked by trees are extracted. Although there is a discontinuity in the extraction result of Fig. 10(c), most roads have been extracted with high integrity. However, the other four methods are less capable of solving such problems, and all results in which trees occlude the roads are broken and discontinuous. In the images of Fig. 10(b) and (c), the CDG and D-LinkNet methods also directly show the loss of tree occlusion road extraction. The above qualitative analysis is consistent with our quantitative analysis results. As shown in Table VII, our method surpassed other methods in all indicators. Compared with the four classical deep learning methods, our DGRN method is also a new road extraction method with high precision and strong anti-tree shielding ability.

C. Effect of the DGSPP Module

To reflect the importance of the improved DGSPP layer in the road extraction process, the baseline model and the ASPP-Net model are compared under the same training conditions. In the baseline network model, we remove the improved DGSPP layer. In the ASPP-Net model, we used ASPP to replace the improved DGSPP layer at the same location in the DGRN network. In addition, we perform validation experiments on the GF-2 dataset and Massachusetts roads dataset. Fig. 11 shows the extraction results of the GF-2 data. As we can see from the marker box in Fig. 11, there are many discontinuities in the baseline model and ASPP-Net model, and the condition of road fragmentation is severe, which is consistent with the performance of the evaluation indicators in Table VI. As seen from Table VI, the extraction results of the DGRN in IoU, Recall, and F1 are far higher than those of the baseline network model. Compared with the traditional ASPP-Net network model, we also achieved 0.29%, 0.49%, and 0.18% improvements in IoU, Recall, and F1, respectively. The extraction results of the Massachusetts roads dataset (Fig. 12) show that the integrity of the DGRN extraction results is also far better than those of the other two comparison models (see the red box in Fig. 12). The road results extracted by the baseline network model and ASPP-Net network model inevitably exhibit fracture and loss

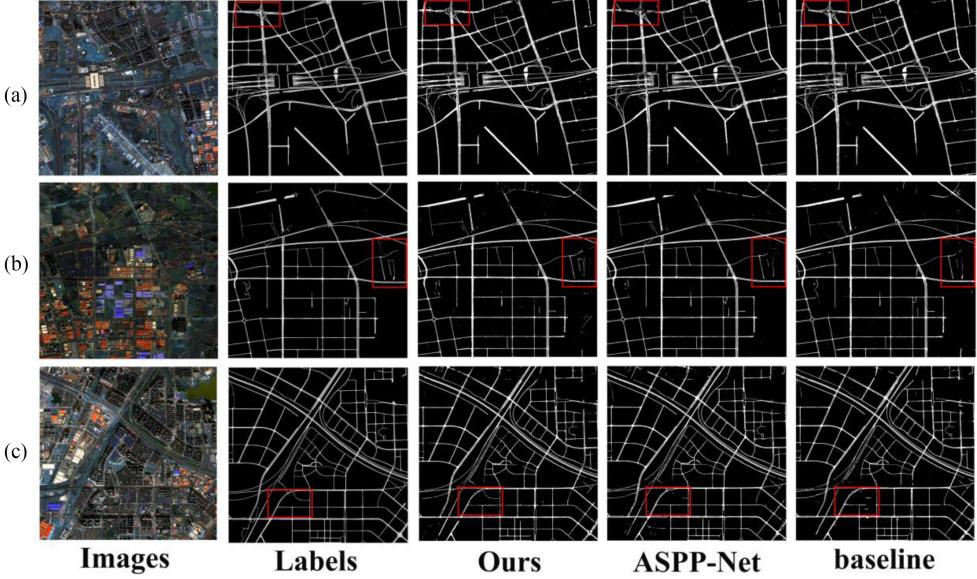


Fig. 11. Comparison of the three methods for road extraction on the GF-2 roads dataset.

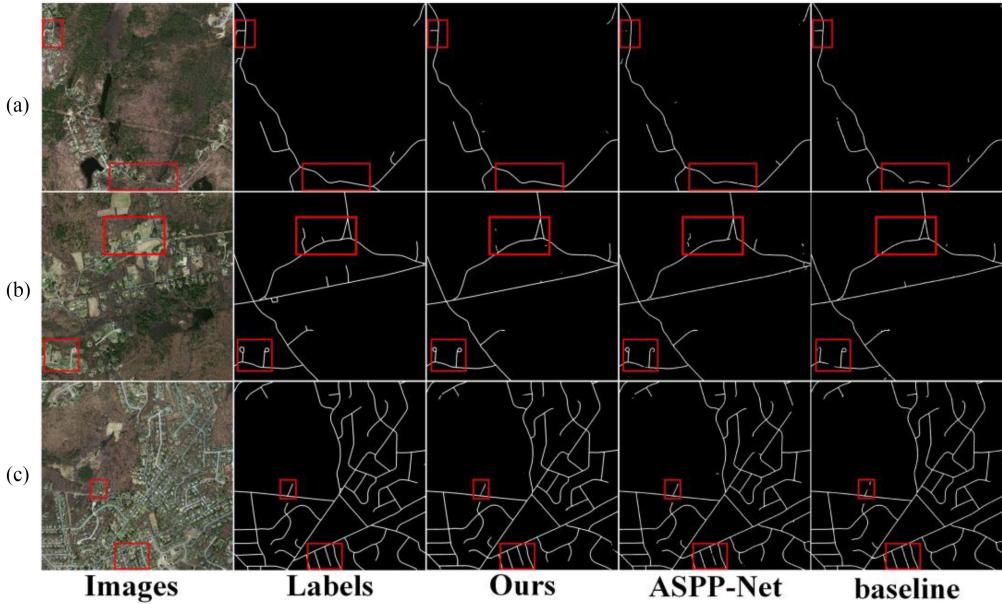


Fig. 12. Comparison of the three road extraction methods on the Massachusetts roads dataset.

TABLE VIII
COMPARISON OF EVALUATION INDICATORS OF THE THREE ROAD EXTRACTION
METHODS ON THE GF-2 ROADS DATASET

Model	IoU	Recall	F1
baseline	78.29%	86.60%	87.79%
ASPP-Net	80.10%	87.62%	88.91%
Ours	80.39%	88.11%	89.09%

TABLE IX
COMPARISON OF EVALUATION INDICATORS OF THREE ROAD EXTRACTION
METHODS ON THE MASSACHUSETTS ROADS DATASET

Model	IoU	Recall	F1
baseline	53.70%	64.24%	69.53%
ASPP-Net	54.35%	63.77%	70.03%
Ours	62.48%	71.97%	76.59%

problems. At the same time, to show the performance of the model more intuitively, we calculated the accuracy indexes of the model, which are listed in Table VIII. As seen from Table IX, in

the Massachusetts roads dataset, the DGRN leads the traditional ASPP-Net network model by 8.13%, 8.20%, and 6.56% in IoU, Recall, and F1, respectively, and far exceeds the baseline basic

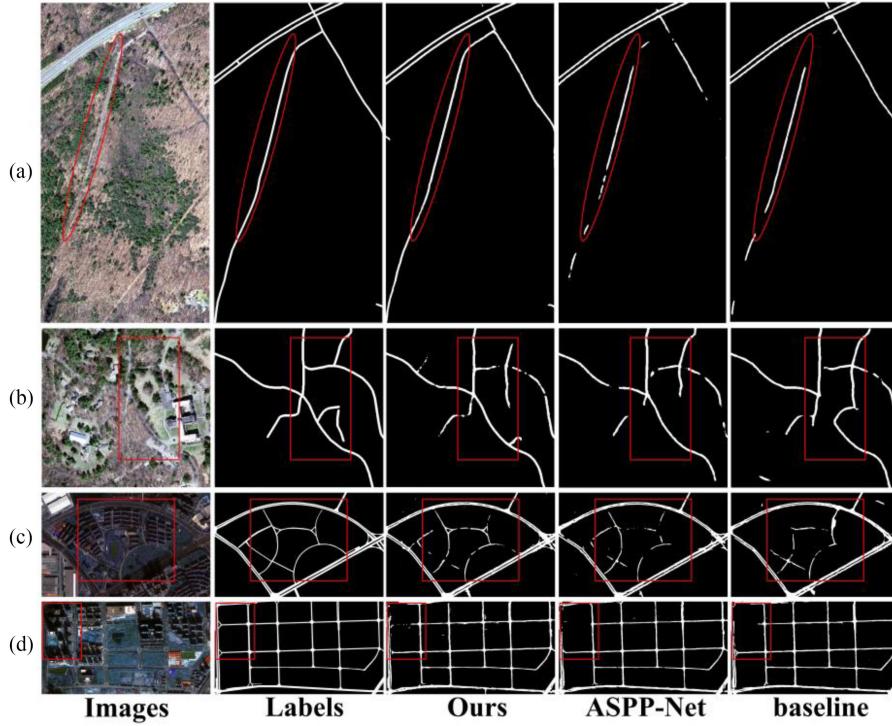


Fig. 13. Comparison of the extraction results of the three methods for road occlusion problems.

model. Compared with the statistical results of the GF-2 dataset, the statistical data of the Massachusetts roads dataset have more advantages.

In addition, the improved DGSPP layer can also adapt to the interference of complex background features such as trees and building shadows. Fig. 13 shows an enlarged view of the details of the extraction results. It can be seen from the red box in Fig. 13 that for the extraction results of the blocked road, the DGRN model has the best performance, far exceeding the other two models. It can be seen from Fig. 13(a)–(c) that the extraction results of the baseline model are better than those of the ASPP-Net model, indicating that the traditional ASPP module has poor performance in solving the problem of trees and shadow occlusion.

In conclusion, the improved DGSPP layer can reduce the loss of the spatial feature information and extract the global context feature information efficiently. It plays a key role in improving the precision of road extraction, ensuring the integrity of the road and resisting the interference of the background information. Experiments show that the improved DGSPP layer is an effective and advanced functional module.

D. Generalization Analysis of the DGRN

To verify the universality of the DGRN, GF-2 remote sensing images of Hefei, Tianjin and Wuhu were tested in this article, and the test results are shown in Fig. 14. Fig. 14 shows that the roads in the three cities have been completely extracted, and the Hefei image has the best extraction effect. As seen from the road results of Tianjin, the connectivity of the road is well maintained.

The extraction results of the DGRN model not only demonstrate good global performance but also grasp the details of the road well (see the red marker box in the Hefei image). It can be seen from the image extraction results of Wuhu that the DGRN model can effectively alleviate the interference phenomenon caused by the shadow of urban high-rise buildings on road extraction (see the red box in the Wuhu (c) image). In conclusion, the DGRN model is a road extraction model with strong extraction ability and good universality.

E. Analysis of Problems

The method in this article has achieved outstanding results in the above comparative experiments, but due to the influence of various factors, there are still a small number of roads that are missed during the extraction process. Deep convolutional neural network models also have these difficulties in road extraction. As shown in the blue box in Fig. 14, the presence of urban buildings, trees, and many other features around the road network, as well as the roof in the image that has an appearance similar to that of a road, contribute to the complex background. Therefore, in the process of road information extraction, the above factors interfere with the network model perception of the target features. Simultaneously, due to the multiscale features of the road network itself, the network model is required to have an efficient multiscale feature perception capability. In the case of complex backgrounds and multiscale road coexistence, the method in this article significantly improves the anti-occlusion ability compared with other methods (see Section IV-B for details), but there are also a few local road discontinuities.

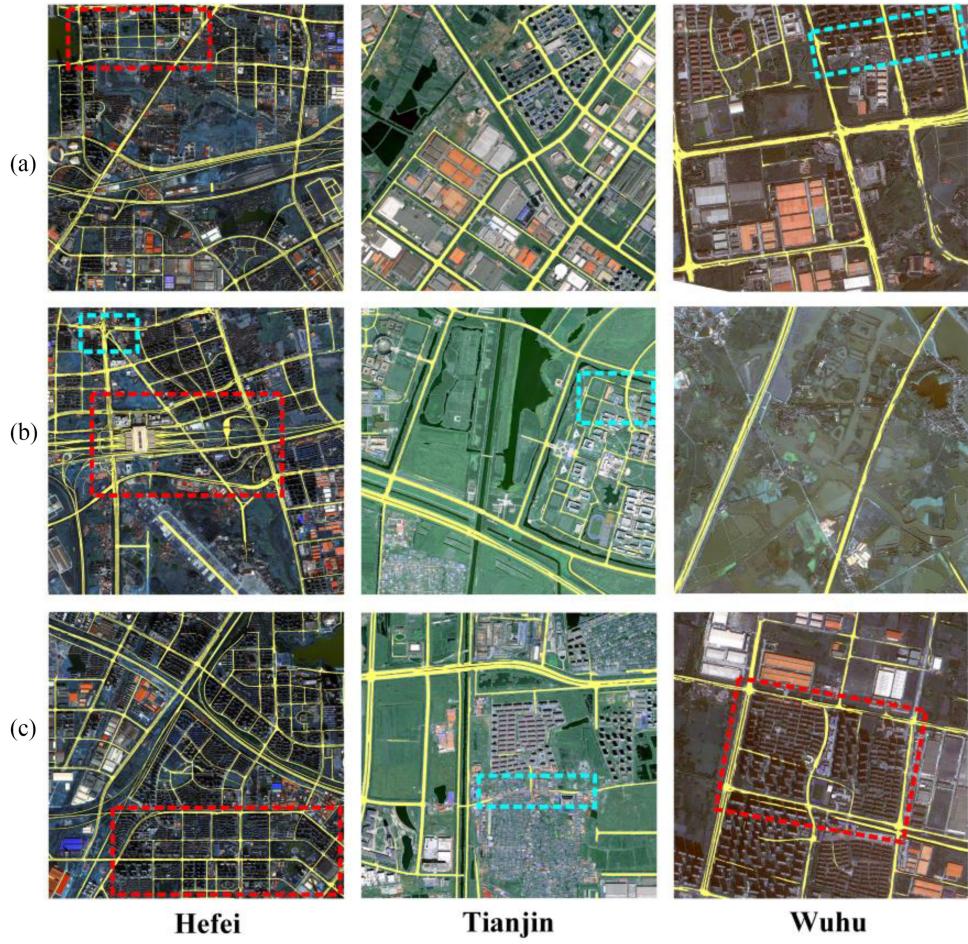


Fig. 14. GF-2 image extraction results from three different cities in China.

V. CONCLUSION

In this article, we analyze the problems of the FCN-based methods, consider the road features, and construct an FCN network called the DGRN that is suitable for road extraction from high-resolution remote sensing images. Based on the structure of ResNet, the network introduces the DGSPP module in the encoding process. The combination of the two not only improves the reuse of the road features, which reduces the loss of information, but also aggregates multiscale contextual information to mitigate the effects of shadows, which improves the consistency of the semantic segmentation. In addition, our method was compared with U-Net, D-LinkNet, DeepLabV3+, and CDG on the GF-2 dataset and Massachusetts roads dataset. The GF-2 dataset experiment showed that the results of the DGRN method were 1.15% higher in IoU, 4.4% higher in Recall, and 0.71% higher in F1 score than those in DeepLabV3+, which had the best effect on the four comparison models. On the published Massachusetts roads dataset, compared with the CDG method, which had the best performance of the four comparison models, our method also achieved 0.58%, 0.17%, and 0.49% improvements. In the case of occlusion, experiments on the two datasets also showed the advantages of the proposed method for road extraction. The better performance on the two datasets

verifies that the DGSPP module we designed has a strong advantage in the extraction of multiscale feature information. In the general experiment, our network also successfully extracted roads from images of Tianjin, China, Hefei, China, and Wuhu, China, and achieved good results. Therefore, the DGRN model is a road extraction model with superior extraction performance and strong generalization ability.

Although the DGRN model has achieved good extraction results, when shadows of trees and houses coexist and the roof is similar to the road, the road extraction results will still appear slightly broken, which is also a difficult point in the current road extraction process. In future research, we plan to use the geometric feature information of the road to supplement the missing feature information occluded by trees to ensure the integrity of the extracted road.

REFERENCES

- [1] S. Hinz *et al.*, “Modeling contextual knowledge for controlling road extraction in urban areas,” in *Proc. IEEE/ISPRS Joint Workshop Remote Sens. Data Fusion Over Urban Area*, Rome, Italy, Nov. 2001, pp. 40–44.
- [2] Q. Li, L. Chen, M. Li, S. Shaw, and A. Nüchter, “A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 540–555, Feb. 2014.

- [3] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3359–3372, Jun. 2014.
- [4] W. Wang *et al.*, "A review of road extraction from remote sensing images," *J. Traffic Transp. Eng.*, vol. 3, no. 3, 2016, pp. 271–282.
- [5] J. Qi *et al.*, "Spatial information inference Net: Road extraction using road-specific contextual information," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9478–9481.
- [6] L. J. Quackenbush, "A review of techniques for extracting linear features from imagery," *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 12, pp. 1383–1392, 2004.
- [7] G. Cheng, F. Zhu, S. Xiang, and C. Pan, "Road centerline extraction via semisupervised segmentation and multidirection nonmaximum suppression," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 545–549, Apr. 2016.
- [8] J. Senthilnath *et al.*, "Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method," *J. Indian Soc. Remote Sens.*, vol. 37, no. 3, pp. 351–361, 2009.
- [9] J. Dai *et al.*, "Lane-level road extraction from high-resolution optical satellite images," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2672.
- [10] J. Liu *et al.*, "Rural road extraction from high-resolution remote sensing images based on geometric feature inference," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 10, 2017, Art. no. 314.
- [11] Y. Wang and Q. Zheng, "Recognition of roads and bridges in SAR images," *Pattern Recognit.*, vol. 31, no. 7, pp. 953–962, 1998.
- [12] H. Jin, Y. Feng, and Z. Li, "Extraction of road lanes from high-resolution stereo aerial imagery based on maximum likelihood segmentation and texture enhancement," in *Proc. Digit. Image Comput.: Techn. Appl.*, 2009, pp. 271–276.
- [13] N. Yager and A. Sowmya, "Support vector machines for road extraction from remotely sensed images," in *Computer Analysis of Images and Patterns*, N. Petkov and M. A. Westenberg, Eds., Heidelberg, Germany: Springer, pp. 285–292, 2003.
- [14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [15] C. Simler, "An improved road and building detector on VHR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 507–510.
- [16] F. Tupin, H. Maitre, J. Mangin, J. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Application to road network extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 434–453, Mar. 1998.
- [17] Y. Li, R. Zhang, and Y. Wu, "Road network extraction in high-resolution SAR images based CNN features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1664–1667.
- [18] D. Zhu, X. Wen, and C. Ling, "Road extraction based on the algorithms of MRF and hybrid model of SVM and FCM," in *Proc. Int. Symp. Image Data Fusion*, Tengchong, Yunnan, China, 2011, pp. 1–4.
- [19] Z. Miao, B. Wang, W. Shi, and H. Zhang, "A semi-automatic method for road centerline extraction from VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1856–1860, Nov. 2014.
- [20] S. Christian *et al.*, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI*, 2017, pp. 12.
- [21] Y. Xu *et al.*, "Quality assessment of building footprint data using a deep autoencoder network," *Int. J. Geographical Inf. Sci.*, vol. 3, no. 10, pp. 1929–1951, 2017.
- [22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [23] Z. Zhang, Y. Wang, Q. Liu, L. Li, and P. Wang, "A CNN based functional zone classification method for aerial images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 5449–5452.
- [24] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
- [25] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3708–3712.
- [26] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
- [27] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3016820](https://doi.org/10.1109/TGRS.2020.3016820).
- [28] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and LiDAR data using coupled CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4939–4950, Jul. 2020.
- [29] R. Hang, F. Zhou, Q. Liu, and P. Ghamisi, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3003341](https://doi.org/10.1109/TGRS.2020.3003341).
- [30] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proc. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [31] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017.
- [32] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 230–241, Jan. 2018.
- [33] T. Panboonyuen *et al.*, "An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery," in *Proc. Recent Adv. Inf. Commun. Technol.*, 2017, pp. 191–201.
- [34] Y. Zhang, G. Xia, J. Wang, and D. Lha, "A multiple feature fully convolutional network for road extraction from high-resolution remote sensing image over mountainous areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1600–1604, Oct. 2019.
- [35] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [36] H. Gao *et al.*, "Densely connected convolutional networks," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [37] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [38] O. Ronneberger *et al.*, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [39] S. Wang *et al.*, "An improved method for road extraction from high-resolution remote-sensing images that enhances boundary information," *Sensors*, vol. 20, no. 7, 2020, Art. no. 2064.
- [40] F. Zhou, R. Hang, and Q. Liu, "Class-Guided feature decoupling network for airborne image segmentation," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3006872](https://doi.org/10.1109/TGRS.2020.3006872).
- [41] V. Badrinarayanan *et al.*, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [42] L. Chen *et al.*, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, pp. 833–851, 2018.
- [43] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 192–1924.
- [44] M. Yang *et al.*, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3684–3692.
- [45] C. Yu *et al.*, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 334–349.
- [46] I. Sergey *et al.*, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167v3*.
- [47] X. Glorot *et al.*, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, FL, USA, Apr. 2011, pp. 11–13.
- [48] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis.*, vol. 128, pp. 742–755, 2020, doi: [10.1007/s11263-019-01198-w](https://doi.org/10.1007/s11263-019-01198-w).
- [49] X. He, R. S. Zemel, and M. A. Carreira-Perpinan, "Multiscale conditional random fields for image labeling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. II–II.
- [50] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 1–8.
- [51] P. Kohli *et al.*, "Robust higher order potentials forenforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 302–324, 2009.
- [52] L. Ladický, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. 12th Int. Conf. Comput. Vis.*, 2009, pp. 739–746.
- [53] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 2–23, 2009.

- [54] J. Yao, S. Fidler, and R. Urtasun, "Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 702–709.
- [55] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1857–1866.
- [56] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 2013.
- [57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [58] J. Wegner *et al.*, "A higher-order CRF model for road network extraction," in *Proc. Comput. Vis. Pattern Recognit.*, 2013, pp. 1698–1705.
- [59] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [60] C. Wiedemann, "External evaluation of road networks international archives of photogrammetry," *Remote Sens. Spatial Inf. Sci.*, vol. 34, no. Part 3/W8, pp. 93–98, 2003.
- [61] M. Maboudi *et al.*, "Impact of gap filling on quality of road networks," in *Proc. ISPRS - Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, 2019, pp. 683–686.
- [62] X. Yao *et al.*, "Land use classification of the deep convolutional neural network method reducing the loss of spatial features," *Sensors*, vol. 19, no. 12, pp. 2792–2792, 2019.



Penghai Wu (Member, IEEE) received the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the School of Resources and Environmental Engineering, Anhui University, Hefei, China. His research interests include land surface temperature reconstruction and fusion, deep learning, and regional eco-environmental change.



Biao Wang received the Ph.D. degree in photogrammetry and remote sensing from Chungbuk National University, Cheongju, Korea, in 2015.

He is a Lecturer with the School of Resources and Environmental engineering, Anhui University, Hefei, China. His research interests include photogrammetry and remote sensing image processing, especially on change detection and object detection.



Qiangqiang Wu received the B.S. degree in geographic information science from School of Resources and Environmental Engineering, Anhui University, China, in 2018 and the post graduate's degree in environmental engineering at School of Resources and Environmental Engineering, Anhui University.

He majors in geographical information system and remote sensing image information extraction.



Yanlan Wu received the Ph.D. degree in cartography and geographic information system from Wuhan University, Wuhan, China, in 2004.

She is currently a Professor with the School of Resources and Environmental Engineering, Anhui University, Hefei, China. Her research interests include deep learning, remote sensing image data processing, and remote sensing image information extraction.



Feng Luo received the B.S. degree in geographic information system, the M.S. degree in cartography and geographic information system, and the Ph.D. degree in cartography and geographic information system from Wuhan University, Wuhan, China, in 2010, 2012, and 2017, respectively.

He is currently a Senior Engineer with the CCCC Second Highway Consultants Company Ltd., Hubei, China. His research interests include intelligent transportation, traffic data mining, and BIM.



Hui Yang received the B.S. degree in geography from Anqing Normal University, Anqing, China, in 2010, and the M.S. degree in surveying and the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2012 and 2019, respectively.

He is currently a Lecturer with Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University, Hefei, China. His research interests include remote sensing image semantic segmentation, artificial intelligence, and geographic information mining.