NIGAN: A Framework for Mountain Road Extraction Integrating Remote Sensing Road-Scene Neighborhood Probability Enhancements and Improved Conditional Generative Adversarial Network

Weitao Chen, *Member, IEEE*, Gaodian Zhou, Zhuoyue Liu, Xianju Li, Xiongwei Zheng, and Lizhe Wang, *Fellow, IEEE*

***Abstract*— Mountain roads are a source of important basic geographic data used in various fields. The automatic extraction of road images through high-resolution remote sensing imagery using deep learning has attracted considerable attention. But the interference of context information limited extraction accuracy, especially for roads in mountain area. Furthermore, when pursuing research in a new district, many algorithms are difficult to train due to a lack of data. To address these issues, a framework based on remote sensing road-scene neighborhood probability enhancement and improved conditional generative adversarial network (NIGAN) is proposed in this article. This framework can be divided into two sections: 1) road scenes classification section. A remote sensing road-scene neighborhood confidence enhancement method was designed for classifying road scenes of the study area to reduce the impact of nonroad information on subsequent fine-road segmentation and 2) fine-road segmentation section. An improved dilated convolution module, which is helpful in extracting small objects such as road, was added into the conditional generative adversarial network (CGAN) to increase the receptive field and pay attention to global information, and segment roads from the results of road scenes classification section. To validate the NIGAN framework, new mountain road-scene and label datasets were constructed, and diverse comparison experiments were performed. The results indicate that the NIGAN framework can improve the integrity and accuracy of mountain road-scene extraction in diverse and complex conditions. The results further confirm the validity of the NIGAN framework in small samples. In addition, the mountain road-scene datasets can serve as benchmark datasets for studying mountain road extraction.**

***Index Terms*— Deep learning, generative adversarial network (GAN), remote sensing, road extraction, scene classification, semantic segmentation, ZiYuan-3.**

Manuscript received 28 February 2022; revised 27 May 2022; accepted 19 June 2022. Date of publication 6 July 2022; date of current version 21 July 2022. This work was supported in part by the Fundamental Research Funds for the Natural Science Foundation of China under Grant U1803117, Grant 41925007, and Grant 42071430; and in part by the Department of Natural Resources of Hubei Provincial under Grant ZRZY2021KJ04. *(Corresponding author: Lizhe Wang.)*

The authors are with the Faculty of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: wtchen@cug.edu.cn; zhougaodian@cug.edu.cn; clay@cug.edu.cn; ddwhlxj@cug.edu.cn; zhengxiongwei@mail.cgs.gov.cn; lizhe.wang@gmail.com).

Digital Object Identifier 10.1109/TGRS.2022.3188908

I. Introduction

ROADS are widely considered to be important basic geographical information in many fields, such as land use and cover analysis, vehicle navigation [1]–[3], early warning of disasters and rescue, and path planning [4]–[6]. High-quality mountain road information supports map updates and emergency responses [7]–[10]. However, owing to the complex terrain of diversified landscapes, roads in mountainous areas display several complicated characteristics, such as irregular shapes; narrow, tiny, tortuous, and sparse distributions [11]–[14]; and terrain occlusions [1]. These features make it difficult to obtain highly precise mountain road data as compared with those in plains and urban areas [15]–[18].

High-spatial-resolution remote sensing imagery is relatively inexpensive and beneficially redundant. In mountain areas, remote sensing techniques are often the only feasible method of extracting road information at regional scales [19]–[21]. Road extraction using remote sensing technology has undergone four phases: from pixel-based [9], [22], [23], regional feature-based [19], [24]–[27], knowledge-based [28]–[31], to deep learning-based methods [19]–[21], [27], [32]–[35].

The pixel-based method does not perform well in regions with complex road features. The regional feature-based method is very dependent on preprocessing methods. The knowledge-based method relies on prior knowledge which leads to a weak generalization ability. Hence, these three mathematical morphology and texture analysis methods are less effective when handling the multiscale and complex features of mountain roads [19], [27], [32], [36]. Therefore, deep learning-based methods have increasingly attracted considerable attention in road extraction [20], [21], [33]–[35], [37]. Road extraction can be considered a semantic segmentation task when using DCNN-based methods [2], [3]. These DCNNs can be divided into three types: fully convolutional network (FCN)-based structures, encoder, and decoder structures [40]–[50], and generative adversarial network (GAN)-based structures [11], [27], [32], [51]–[57].

1558-0644 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

FCN-based and encoder and decoder structures have achieved good performance in road extraction; however, with the deepening of network layers, the input information is diluted, road details can be lost, and the extraction accuracy is reduced [7]. Some researchers try to solve the problem with boundary information. Shao *et al.* [17] used atrous convolutions and a pyramid scene parsing pooling module for road centerline extraction. Wei and Ji [41] detected the mask and prior road boundary information, then input this information into the proposed dual-branch encoder–decoder network. Zhou *et al.* [44] proposed a split depth-wise separable graph convolution network (SGCN) to capture global information to enhance the features of road, and the model performed best when compared with related road extraction models. Xu *et al.* [45] proposed a model, named IDANet, which adopts iterative D-LinkNets with attention modules for road extraction and the attention mechanism can be used to achieve a better fusion of features from different levels. Wan *et al.* [18] proposed a dual-attention road extraction network to minimize the loss of road structure information caused by multiple down-sampling operations. Abdollahi *et al.* [32] proposed using multilevel context gating UNet and bi-directional ConvLSTM UNet to maintain the boundary information, even in complicated conditions.

Nevertheless, these methods cannot completely fill the gap between weak supervision and full supervision, especially when the road information is less in the study district. For road extraction, GAN can effectively compensate for the scarcity of road feature information [58], [59]. The generative adversarial structure has high robustness and renders the network convergence stably [58], [60]. Some GAN-based methods performed well on public road datasets [2], [3], [60]–[64]. Zhang *et al.* [61] proposed a road extraction method using an improved GAN that only requires a few samples for training and achieving 87% $F1$-score and 98% pixel accuracy (PA) in the Massachusetts road dataset. A learning-based road extraction method using a multisupervised GAN was proposed and achieved a 96.7% $F1$-score in Cheng's dataset [62]. This method is jointly trained by the spectral and topological features of the road network. Varia *et al.* [58] applied GANs for road extraction from UAV images, and FCN was used in the generative part. It performed better than FCN-32 for extracting roads.

However, these GAN-based methods are still prone to gradient disappearance and mode collapse, which result in information loss. Additionally, while many research studies use public datasets for road extraction [17], [20], [37], [44], [59], [61], [63], there are few reports on mountain road extraction from remote sensing images [21]. Consequently, there is an increasing demand to extract mountain road information at the regional scale using deep learning. A discontinuous road result is still the key issue of popular road segmentation methods [35].

In this study, a mountain road extraction dataset was constructed and to address the problem of insufficient data and contextual information loss in mountain road extraction, we constructed a road extraction framework: improved conditional generative adversarial network (NIGAN). From scene level to pixel level, this framework allows us to improve the road extraction accuracy under small samples.

The contributions of this study are presented as follows.

1) To improve the connectivity of road extraction, we developed a road remote sensing road-scene neighborhood confidence enhancement strategy to enhance the confidence of scenes containing roads.
2) An improved ResNet34 model and dilated convolution were added to the proposed framework to reduce diameters while conserving the main road features. This model is beneficial to improve feature extraction ability and prevent overfitting.

II. Methods

Considering the complex features of mountain roads and the homogeneity of local areas at regional remote sensing images, a NIGAN framework was proposed to improve the extraction performance of mountain roads. It consists of two phases, road-scene extraction and fine-road extraction. The overall framework is shown in Fig. 1. In road scene extraction section, we extract scenes, which contain roads, with neighborhood enhancement method. While in fine-road extraction section, roads are segmented from road scenes outputted in road scene extraction section. The output of NIGAN is extracted fine road. The code is available on Github: https://github.com/cug103/NIGAN.

1) *Road-Scene Extraction:* In this section, the module analysis of the scene contains road. The main structure is based on conditional GANs (CGANs). The original CGAN expression is as follows [65]:

$$\min_G \max_D V(D,G) = E_{xp_{\text{data}(x)}}\left[\log D(x \mid y)\right] + E_{zP_z(z)}\left[\log(1 - D(G(z \mid y)))\right]. \quad (1)$$

In (1), G represents the generation network, and D represents the discrimination network. The condition variable, y, is added to the inputs of the generator and discriminator. Road-scene extraction can be considered a binary classification task. Thus, the use of the CGAN can better control the generator's output in this study. Moreover, road and nonroad scenes can obtain high-quality data augmentation using the CGAN. The generator network consists of five deconvolution layers with a kernel size of 4×4, and the discriminator network consists of five convolution layers with a kernel size of 4×4.

First, when training is initiated, part of the CGAN is trained while the classifier is temporarily frozen. The CGAN training is then stabilized, combining the generated samples and original image datasets to create a mixed dataset. Thereafter, the classifier is unfrozen, and the mixed datasets become fully engaged in training the classifier. Second, a classifier with strong feature extraction capabilities is selected because the road is a weak target, and the nonroad context is complex and volatile. Owing to its excellent residual structure, the ResNet-18 was selected as a classifier for road-scene extraction. Simultaneously, in the classification

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

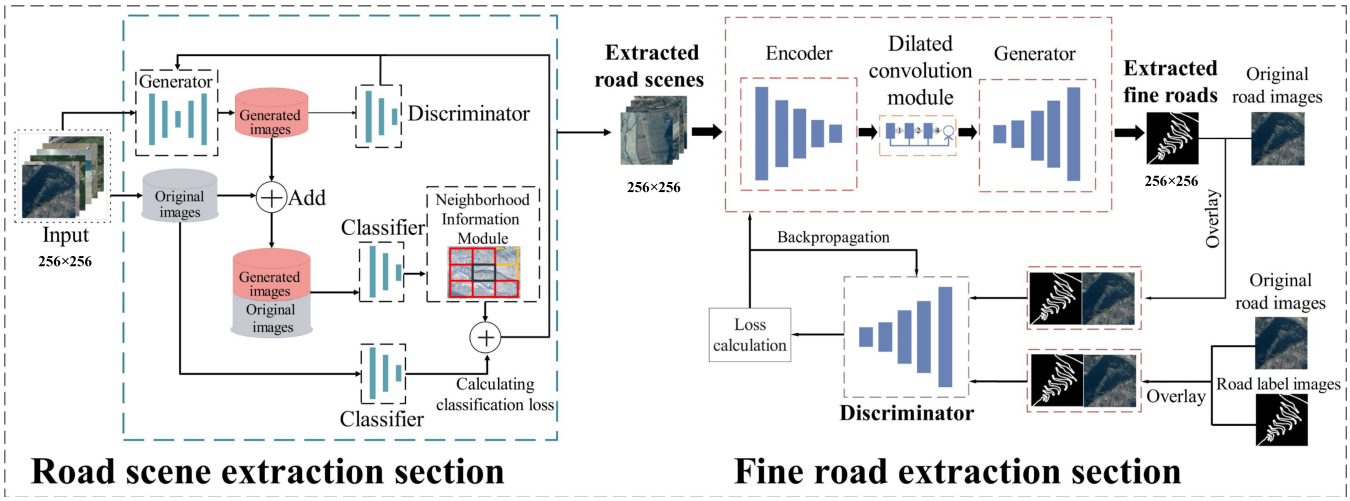


Fig. 1. Proposed NIGAN framework.

loss calculation, the label is abandoned as the positive value, but the classification results in the original dataset being used as the positive value of the loss calculation. As a result, the model in the classification loss calculation is changed from absolute to relative. Using mixed loss narrows the distributional differences between generated and real images, achieving a smooth generator optimization and a stable model convergence via reverse optimization. Finally, the road-scene neighborhood probability enhancement method is developed to improve the integrity and accuracy of road scenes.

2) *Fine-Road Extraction:* In this section, the module segment road contains the output road scene in road scene extraction section. The road extraction result obtained in this study is a binary image (the road value is "1," and the background information is "0"). Based on the results of the road-scene extraction, a fine-road extraction model based on CGAN, which integrates an SELU activation function and a dilated convolution, is proposed. The extracted road scene images serve as the input to the CGAN to perform fine-road extraction. In this study, the discriminant network accepts negative and positive sample pairs as the input of the network. The negative sample pair is composed of the fine-road extraction result and the corresponding images, and the positive sample pair is composed of the road label image and corresponding original image. The output result is the probability that the current input sample belongs to the positive sample. The generator network contains five decoder blocks, and a block consists of a deconvolution layer with a kernel size of 3×3 and two convolution layers with a kernel size of 1×1; the discriminator network consists of five convolution layers with a kernel size of 4×4.

A. Constructing the Road-Scene Neighborhood Probability Enhancement Method (Road-Scene Extraction)

The neighborhood information refers to the feature information regarding the neighboring areas. Generally, two calculation methods are used to describe the neighborhood information of remote sensing image pixels: four and eight neighborhoods. Roads are generally considered to be local consistencies that provide regional connectivity. Particularly, a single road can be distributed across adjacent patches in remote sensing images. These characteristics inspired us to extend pixel-neighborhood theory to remote sensing scene levels. Accordingly, we developed a remote sensing road-scene neighborhood probability enhancement method based on the local consistency criterion. This method could improve the connectivity and accuracy of road extraction.

In traditional remote sensing scene classification algorithms, road-scene neighborhood information cannot be constructed for discrete data with no geographic spatial coordinates. However, regional remote sensing images have spatial geographic information; hence, the road-scene neighborhood information can be constructed. The concept of this study is demonstrated in Fig. 2.

The black-framed patch in Fig. 2 is the scene to be classified. Based on the eight-neighborhood calculation rule, the calculation range is eight adjacent remote sensing scenes. Four-neighborhood refers to the following positions: up, bottom, left, and right. In Fig. 2, the central, black-framed patch only contains a very small portion of the road. They are very likely to be missed, resulting in a discontinuous road network. However, using the proposed road-scene neighborhood confidence enhancement strategy, the six road scene patches around the black frame are calculated, increasing the probability of classifying the black-frame scene as a road: $p_i + (6/8) \times p_i$.

B. Improving ResNet34 for Feature Extraction (Fine-Road Extraction)

The ResNet effectively reduces redundant information while maintaining a rather high convergence rate to avoid data collapse in deep networks [37], [66]. Considering information theory, owing to data processing inequalities, image information in the feature map decreases layer by layer as the number of network layers increases during a forward transmission. However, the participation of residual nets featuring a direct

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

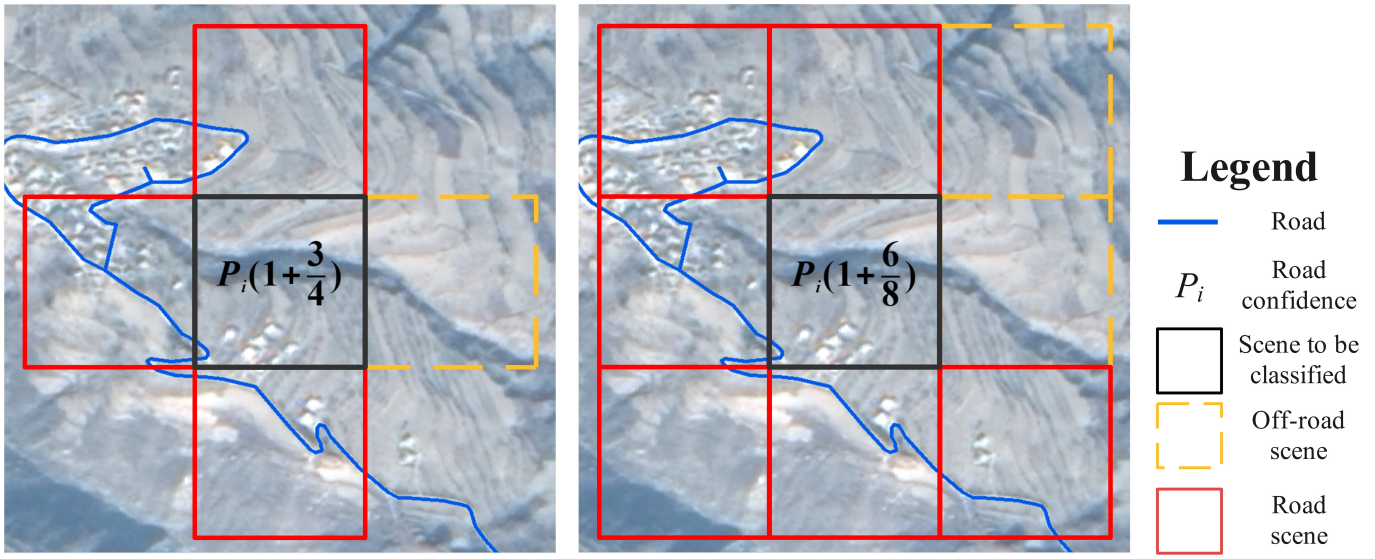


Fig. 2. Concept presentation of the confidence enhancement method of the remote sensing road-scene neighborhood. (Left) Four- and (Right) eight-neighborhood types, respectively.

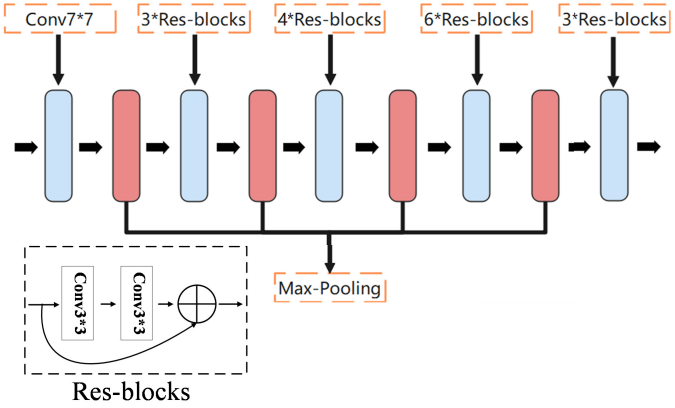


Fig. 3. Improved feature extractor structure based on ResNet34 in this study.

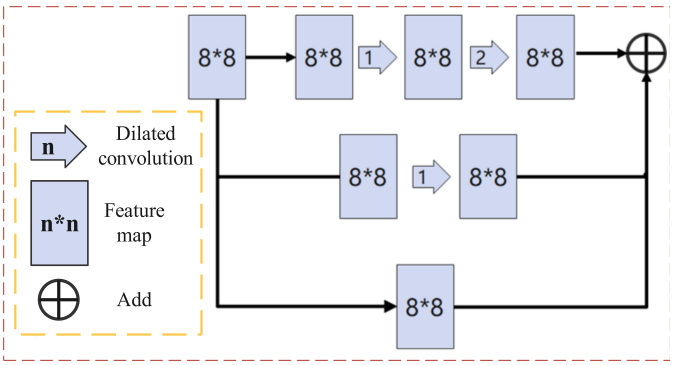


Fig. 4. Structure of the dilated convolution cascade used in this study.

mapping function ensures more image information in the $(n+1)$st layer of the network than that in the nth layer. Hence, it serves as a stabilizer during model training. The original ResNet34 model contains 33 convolutional layers and one fully connected layer. Because only the ResNet is employed as the image feature extractor in this study, we propose an improved method to prevent overfitting as follows: the fully connected layer is removed, and a max-pooling layer is inserted between all layers except for the first, which has a 7×7 convolution kernel size; the rest are all 3×3. A max-pooling layer is added behind the first, seventh, 15th, and 27th layers to reduce diameters while conserving the main road features. The improved ResNet34 network structure is shown in Fig. 3.

C. Improving the Dilated Convolution to Improve Receptive Field Enhancement (Fine-Road Extraction)

In a CNN, the overall receptive field slowly grows as the number of layers increases. Usually, the pooling function reduces the image size and increases the receptive field at the cost of information loss. However, few road features and information loss caused by the lower receptive field lead to broken or missed road extractions. An improved dilated convolution structure is used to enhance the receptive field in this study as follows: To obtain a filter, we skip the step size of a specific input value to apply it to an area larger than the convolution kernel. This method increases the receptive field of the network unit layer-wise and renders a receptive field growth from linear to exponential as the depth of the network layer increases. In this study, the dilated convolution is stacked under a cascade mode. If the expansion coefficients of the stacked dilated convolution are 1, 2, 4, 8, and 16, then the receptive fields accepted by each layer will be 3, 7, 15, 31, and 63, respectively. Following feature extraction, the pixel size of the input scene image changes from 256×256 to 8×8. The dilated convolutional layers having expansion coefficients of 1 and 2 were harnessed therein to better extend the input feature map to the dilated convolution coverage area. The improved cascade structure is shown in Fig. 4.

D. Improving the Activation Function in the Generator Convolution Structure (Fine-Road Extraction)

In the traditional generator convolution structure, the region value of the negative semi-axis in the RELU activation

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

function is simply reduced to zero, leading to potential gradient disappearance during network training. This condition renders the generator model unable to produce effective image, limiting the accuracy of road-scene extraction. However, a negative value exists in the negative semi-axis area of the SELU activation function. Its specific mathematical expression is as follows:

$$F(x) = \begin{cases} \lambda x, & x > 0 \\ \lambda(\alpha e^x - \alpha), & x \leq 0. \end{cases} \tag{2}$$

This function has three advantages: First, the output value can better control the mean value. Second, a saturation zone is used to suppress large variances at the lower level. Third, multiplying parameter λ in the function renders the slope of some of the areas greater than 1, enabling flexible adjustments to inflate the undersized low-level variance. Therefore, in this study, we replaced the RELU activation function in the traditional generator structure with the SELU function. Furthermore, owing to its capacity of batch normalization, there is a need to remove the batch normalization layer synchronously. During the data generation of the improved generator model, fewer mode collapses and gradient disappearances result in more authentic and diversified generated road scenes. The two hyperparameters λ and α are set to $\lambda \approx 1.67$ and $\alpha \approx 1.05$ in unison, retaining two decimal places.

E. Evaluation Criteria

The model evaluation criteria used in this study include the overall accuracy, recall rate, mean IOU (MIOU), PA, mean pixel accuracy (MPA), floating point operations (FLOPs), and number of parameters (Param). The overall accuracy enables a very intuitive evaluation of the results, with its value representing the number of correct classification samples divided by the total number of samples. PA represents the percentage share of well-classified pixels in the aggregate images. MPA refers to the percentage of correctly predicted pixels in each category and calculates the average number of all the categories. IOU is generally used to evaluate the overlapping degree between the predicted and targeted areas in the detection of the target. The MIOU incorporates the IOU calculation of each category, presenting the mean value of all intersection ratios. FLOPs are the floating-point operations performed by a model and are often used to describe the computational complexity of the given model

$$\text{PA} = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{3}$$

$$\text{MPA} = \frac{1}{k} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{4}$$

$$\text{MIOU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{5}$$

where k denotes the total categories; $p_{ij}$is the number of pixels, which belong to class i, predicted as class j; and p_{ii} is the number of pixels, which belong to class i, predicted as class i

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

where true positive (TP) denotes the number of scenes correctly classified as road scene, false positive (FP) is the number of other scenes classified as road scene, true negative (TN) denotes the number of scenes correctly extracted as background scene, and false negative (FN) is the number of road scenes classified as background scene.

III. Experimental Results and Analysis

A. Dataset Construction

Currently, road datasets mainly focus on cities (e.g., Massachusetts road dataset) [21], [67]. Roads in cities are insufficient for practical applications in which the images have high intraclass diversity and low interclass variation.

1) Remote Sensing Data Source and Preprocessing: Current datasets on road extraction select aerial and satellite images as the data source, with a spatial resolution of approximately 1 m [9], [36], [41]. They are mainly used to perform deep learning model tests. However, the proposed NIGAN model focuses on applications at a large regional scale. It is very difficult to obtain images with a spatial resolution higher than 1 m in a mountainous area covering 2589 km^2. Hence, we selected Ziyuan-3 with 2.1-m spatial resolution images to construct our mountain road dataset. Because of the edge effect of road remote sensing imaging, it is feasible to use Ziyuan-3 for road extraction. Ziyuan-3 is China's first high-resolution civilian stereo mapping satellite and was launched in 2012. The basic parameters of the satellite are listed in Table I.

Ziyuan-3-02 images taken on December 17, 2018, without cloud coverage, were used to generate our dataset in this study. First, we used ENVI (The Environment for Visualizing Images) 5.3 software to extract the digital terrain model (DTM) based on stereo pairs. Second, the DTM data were used to ortho-rectify the multispectral and panchromatic images. Thereafter, we fused the rectified multispectral image with a panchromatic one, using the pansharpening method [68] to produce a fused image having a resolution of 2.1 m.

2) Dataset Description: The datasets include road scenes and labels. The total patch number of the scene dataset was 8958, including 2585 road scenes and 6373 nonroad scenes, and we chose 878 representative images for the training (603 nonroad scenes and 275 road scenes). The size of the road scene image is 256 × 256 (road scene denotes that the image contains a road). There was a total of 225 road label datasets. These datasets can be downloaded from the following website: https://drive.google.com/drive/folders/1Hk9eOP_b4gfwQ2NHvNcFRMc95EimhfuG?usp=sharing. To examine dataset validity, we referred to Google Earth and compared images to the data of China's third national land survey. Classical scenarios are shown in Fig. 5.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

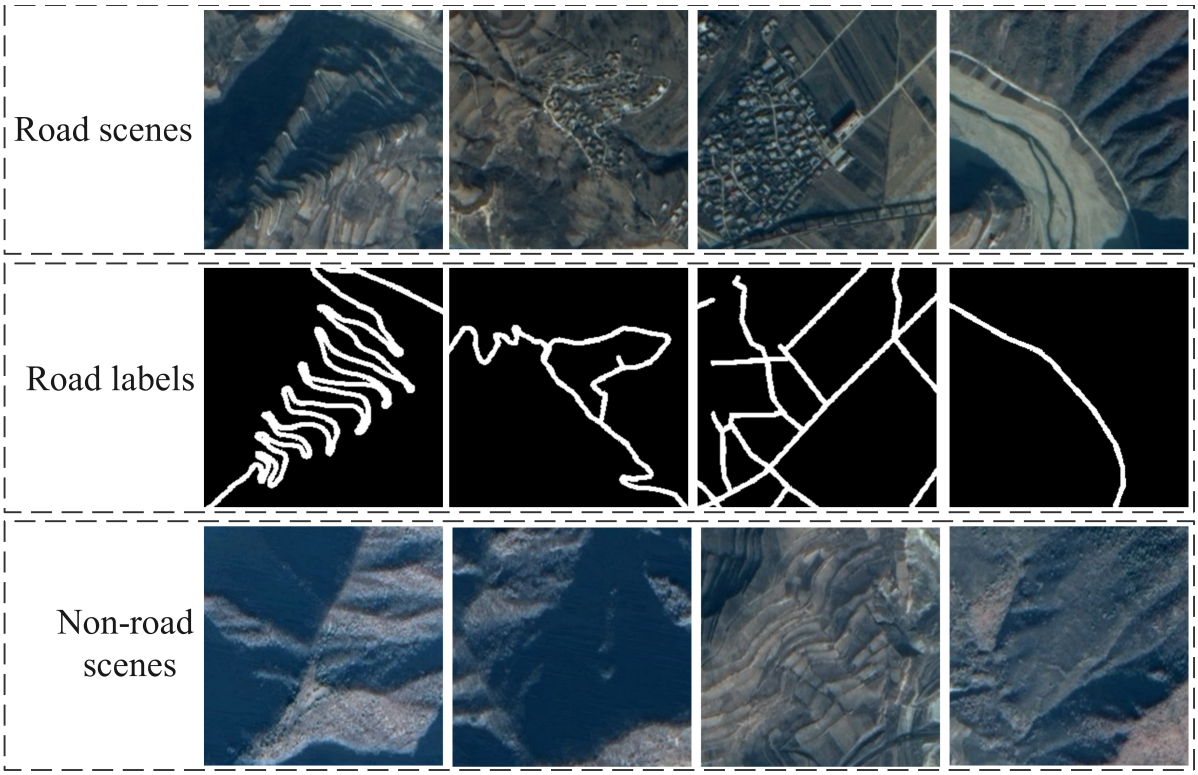


Fig. 5. Example of road scenes, nonroad scenes, and fine label datasets.

TABLE I
MAIN PARAMETERS OF ZIYUAN-3

Data types	Parameters
Camera model	Forward camera, nadir camera, backward camera
Resolution	The nadir camera has a spatial resolution of 2.1 m
Wavelength	PAN: 450–800 nm
	Blue: 450–520 nm
	Green: 520–590 nm
	Red: 630–690 nm
	NIR: 770–890 nm

TABLE II
EXPERIMENTAL PARAMETER SETTINGS

Parameter	Detailed description
Batch size	32
Number of training epochs	2000
Initial learning rate	0.0002
Lower limit on the learning rate	0.0000005
Upper epoch of unoptimized model	48
Attenuation coefficient of the learning rate	1.5
Attenuation period of the learning rate	24
Optimizer	Adam
Loss function	Cross-entropy and Dice coefficient

B. Experimental Parameter Settings

The experiments of this study were conducted using two E5-2620v4 processors and two GeForce RTX 2080T platforms. The detailed parameters based on the NIGAN training are listed in Table II. We conducted five rounds of experiments and averaged the results.

We chose Adam as the optimizer. Compared with other optimizers, the main advantage of Adam is that the learning rate of each iteration has a certain range with offset correction, which makes parameters relatively stable [16], [61]. Cross-entropy and the Dice coefficient are popular loss functions and perform well in many deep learning frameworks [69], [70], so we used them in our model. All discriminators except for those of the CGAN were migrated to the preliminary training parameters on ImageNet [71], [72]. This practice rendered low-level network parameters more stably with the basic image recognition capability and expedited training convergence to compensate for the lack of data caused using a small sample. The setting of the dataset was consistent with that of other methods [14], [73]. For training, 20% of the dataset were selected, and during training, the ratio of training versus verification was 5:5.

C. Experimental Results and Analysis

1) Road-Scene Extraction Results and Analysis: To verify the effect of the proposed road-scene extraction model, we

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

TABLE III
DESCRIPTION AND EXPERIMENTAL RESULTS OF THE ROAD-SCENE EXTRACTION

Model	Road-scene neighborhood enhancement method	Overall accuracy（%）	Recall rate（%）	Param	FLOPs
AlexNet	×	91.49± 0.91	85.93 ± 1.05	60.97M	7.27G
VGG-11	×	91.65± 0.87	86.43± 0.68	132.87M	7.77G
VGG-16	×	91.42± 0.75	87.14± 0.92	138.36M	15.61G
ResNet-18	×	92.77 ± 0.95	87.85± 0.72	11.69M	1.82G
BnInception	×	92.21± 1.09	85.71 ± 1.77	56.00M	13.00G
NasNetamobile	×	91.83 ± 0.88	85.60± 1.02	29.60M	32.14G
DenseNet-121	×	92.10 ± 0.71	87.50± 0.83	7.89M	2.90G
NIGAN	×	93.61 ± 0.78	87.24± 0.88	35.70M	99.84G
NIGAN	Four-neighborhood	93.74 ± 0.62	88.89 ± 0.91	35.70M	99.84G
NIGAN	Eight-neighborhood	**94.40 ± 0.51**	**90.28± 0.61**	**35.70M**	**99.84G**

used the NIGAN framework and performed training and verification on the constructed scene datasets to extract the road scenes. First, to verify the superiority of the neighborhood information enhancement, the NIGAN, excluding the road-scene neighborhood information function, was implemented for comparison. The four-neighborhood NIGAN was also added to verify whether the eight-neighborhood method was better. Training and testing were conducted on the same dataset using seven popular classification networks in the proportion of 5:5 for training versus validation. The networks used included AlexNet [74], ResNet-18 [75], VGG-11 [76], VGG-16 [76], BnInception [77], NasNetamobile [78], and DenseNet-121 [79]. The six networks are all widely recognized as baseline models for comparison studies. The results are summarized in Table III.

Comparing the overall accuracy and recall rates of the different models, the NIGAN generally performed best with an accuracy reaching 94.4% and recall rate accuracy approximating 90.28%. The main reason is that proposed road-scene neighborhood information enhancement method is beneficial to the classification of regional scenes. ResNet-18 and DenseNet-121 achieved good results, owing to the residual connection rendering the forward and backward communications smoother. However, ResNet-18 had a shallower network layer, outperforming the other networks based on small sample data.

Fig. 6 demonstrates the visualization results of the extracted road scenes in the study area. We chose ResNet-18, which performs best in comparison models, and three types of the NIGAN, which apply different neighborhood information enhancement strategies, for illustration. The road scenes were basically extracted, although they were prone to misjudgment in some details (e.g., parts of river valleys). Such a misjudgment is attributed to the sharp interclass similarity in image features among valleys, ridges, and roads. An increase of approximately 3% in the road-scene integrity was achieved using the confidence-enhancement method of neighborhood information. Nevertheless, the improvement of the overall accuracy was not as evident. Consequently, although the completeness of the road-scene extraction improved, the neighborhood information enhancement strategy misjudged

some features with high similarities to roads. To contract more complete road scenes in this section, this method performs better in terms of accuracy and connectivity for network training. Furthermore, the results of the eight-neighborhood NIGAN were better than that of the four-neighborhood method in terms of accuracy and recall rates. This finding confirms the efficacy of the road-scene neighborhood enhancement strategy designated for mountain road features in this study.

As shown in the red frame B of Fig. 6, the road-scene neighborhood enhancement strategy was more effective in enhancing the completeness of the road extraction compared with excluding enhancement. Red frames A and C show that the application of the eight-neighborhood mode generated more complete results for neighborhood information. This plays an important role in subsequent fine-road extractions with more integral road-scene data.

2) Results and Analysis of Fine-Road Extraction: To verify the superiority of the proposed NIGAN network, five state-of-the-art semantic segmentation models were widely used on the self-built dataset. These networks included UNet [80], Attention-UNet [81], PSPNet [10], Linknet34 [82], and SegNet [83]. The comparison results are summarized in Table IV.

The NIGAN's performance is the best in all the three indicators. Although we adopted ResNet34 as the main network for LinkNet34, the proposed NIGAN exhibited better results, owing to the larger receptive field and more stable CGAN training mode.

For another comparison of each network, the results of four road scenes with different complexities were analyzed, as shown in Fig. 7.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

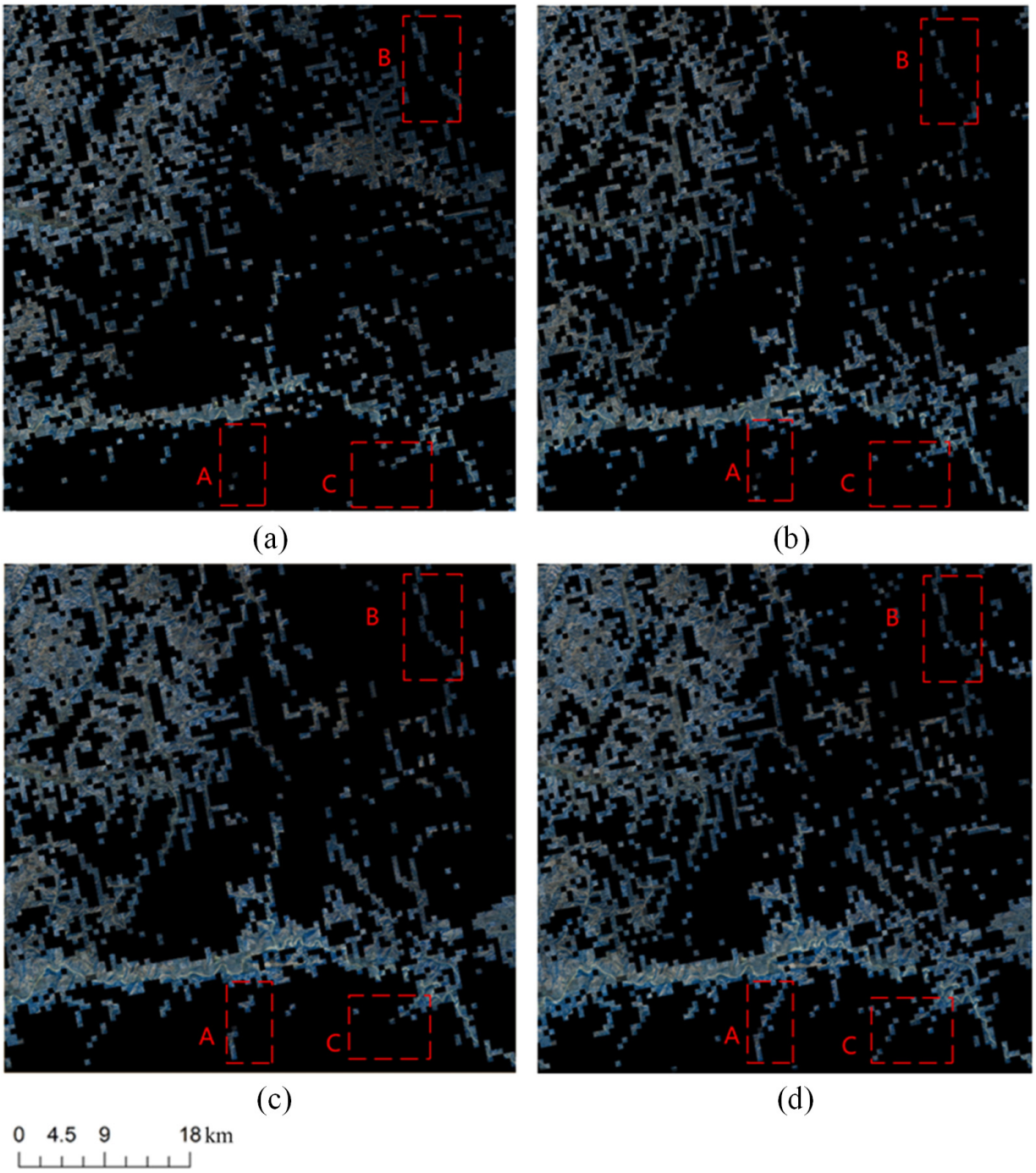


Fig. 6. Road-scene extraction results of four different conditions. (a) ResNet-18 model. (b) NIGAN excluding the neighborhood enhancement method. (c) NIGAN including the four-neighborhood enhancement method. (d) NIGAN including the eight-neighborhood enhancement method. Red frames A–C present local differences.

1) Comparison of the fine extraction effect on simple road-scenes. Fig. 7(a) presents a simple road scene where there is only one relatively straight road across. However, a small portion of the road is exposed at the upper left of the scene, and there exist cross-roads beneath the road. The NIGAN extracts the crossing road and the part of the road cropped at the upper-left corner in a relatively complete manner; however, it omits part of the triangular road. By contrast, Attention UNet and PSPNet only extracted the crossing road. Although all networks exhibited identical performance, the NIGAN extracted a more complete road.
2) Comparison of the fine extraction effect on multiple road scenes. Fig. 7(b) presents a scene of multiple roads where four roads have varying widths, an artifact, and a town. The NIGAN contracted the four roads while failing at the right end of the uppermost road. UNet merely contracted the partial outlines of the three roads above and completely omitted the roads in the towns

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
ACCURACY EVALUATION OF THE FINE-ROAD EXTRACTION RESULTS UNDER THE DIFFERENT MODELS

Models	PA	MPA	MIOU	Param	FLOPs
UNet	94.83 %± 0.44	73.51 %± 0.47	63.62 %± 0.55	31.04 M	24.00 G
Attention-UNet	96.16 %± 0.61	78.95 %± 0.60	69.12 %± 0.73	31.39 M	24.42 G
PSPNet	93.88 %± 0.46	73.84 %± 0.56	61.15 %± 0.78	48.70 M	19.17 G
LinkNet34	96.27 %± 0.52	79.30 %± 0.67	70.34 %± 0.91	32.80 M	13.09 G
SegNet	96.08 %± 0.73	75.22 %± 0.78	67.75 %± 0.81	29.45 M	15.47 G
NIGAN (Proposed)	96.34%± 0.57	81.15%± 0.63	71.65%± 0.71	35.70 M	99.84 G

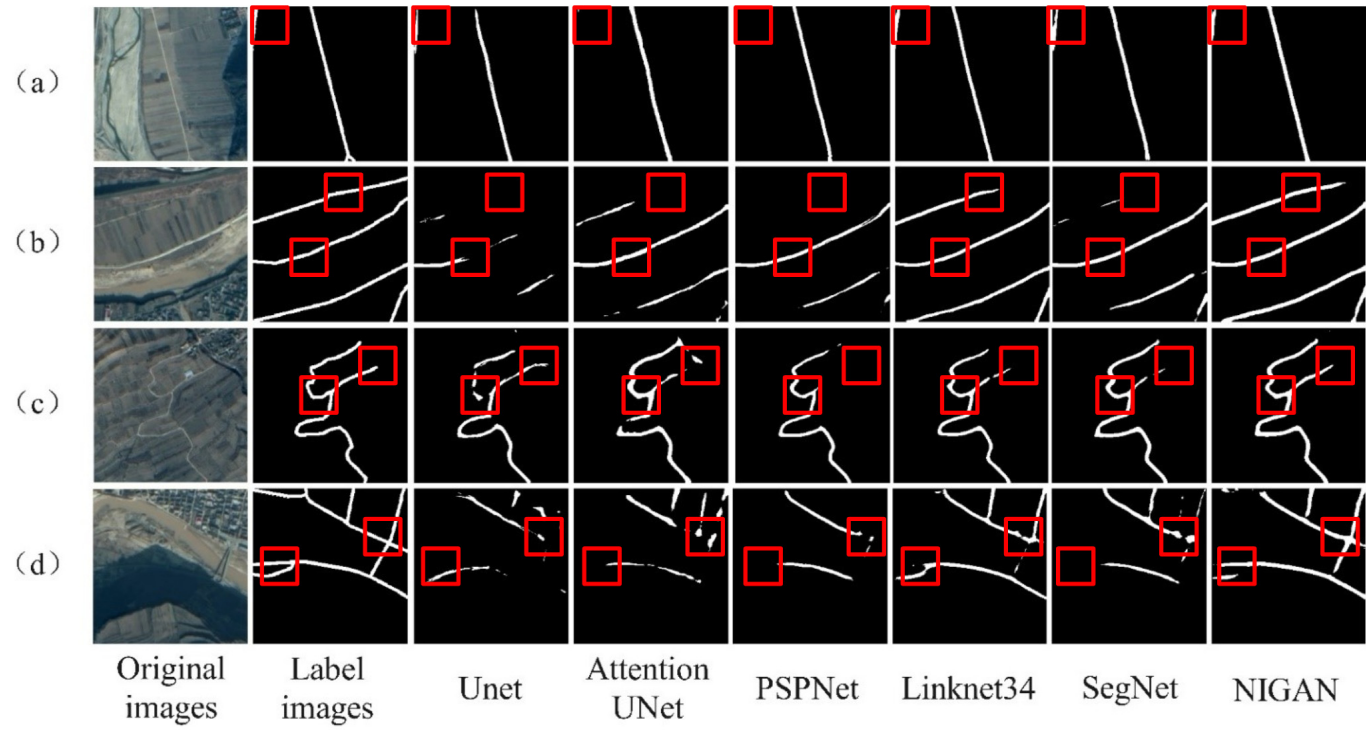


Fig. 7. Original and label images in four different road scenarios and extraction results from different models. (a) Simple road scene. (b) Multiple road scenes. (c) Tortuous road scene. (d) Complex road scene. Red frames circle the differences.

beneath. Considering that all networks are equipped with the ability to extract partial road outlines, LinkNet34 led because, as a feature extractor, ResNet-18 retained more shallow features during image extraction. Conversely, the rest of the models underperformed when extracting the small roads above and the roads in cities and towns beneath.

3) Comparison of fine extraction effects on tortuous road scenes. Fig. 7(c) presents a winding road with crossroads in a relatively simple context except for the presence of a village at the upper-right corner and a canal to the left of the village. The NIGAN completely extracted the roads in the scene; however, it failed to capture the details at the bends. The other networks extracted most roads; however, compared with the NIGAN, they were far from complete. Attention UNet even misidentified the canal as a road, and there were severe interrupted road extraction phenomena.
4) Comparison of the fine extraction effect on complex road scenes. The leftmost part of Fig. 7(d) presents a comparatively complex town street, where a bridge crosses a river. As a result of the imaging incidence angle and time, the road below the scene was covered with shadows, making it difficult to achieve fine-road extraction. As shown in Fig. 7(d), the NIGAN extracted most roads and ensured the continuity of the roads to the best extent possible, except for the bifurcated road junction at the left side of the shadow. With regard to the other models, LinkNet34 also extracted most of the roads, but there exists a severe extraction loss for the bridge roads. Other models missed most of the information of shadows and bridge roads.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

TABLE V
COMPARISON RESULTS OF THE FINE-ROAD EXTRACTION CONSIDERING INCLUDED AND EXCLUDED SCENE EXTRACTIONS

Model	Road-scene extraction	MPA	MIOU
NIGAN	×	81.15%± 0.63	71.65%± 0.71
NIGAN	✓	**81.82%**± 0.56	**72.10%**± 0.63

Generally, the proposed NIGAN was superior to the five state-of-the-art networks in terms of road integrity and fine-road extraction accuracy.

3) Fine-Road Mapping in the Study Area: To verify the mapping ability of the NIGAN model at a regional scale to extract road information, we conducted fine-road extractions mapping in the study area. Considering the experiment, LinkNet34 generally performs best with mountain road extractions; hence, it was selected as the comparison model. The detailed results are shown in Fig. 8. LinkNet34 and NIGAN can extract a relatively complete road network. However, compared to the red frame (1), the NIGAN was insusceptible to interferences from nonroad backgrounds; thereby, we eliminated false detections. Considering red frames (2)–(4), the NIGAN performed well in fine-road extraction in terms of integrity and comprehensiveness.

IV. DISCUSSION

A. Efficacy of the Road-Scene Extraction Strategy

To verify the performance of the proposed road-scene extraction strategy, we conducted a road network extraction, including and excluding a scene extraction module, as shown in Table V. The results are demonstrated in Fig. 9.

As shown in Table V, there is an increase in the MPA and MIOU. This is because, from the perspective of the scene scale, the NIGAN is accessible to more information to judge whether there is a road in the current scene.

Accordingly, the proposed strategy can avoid the interference of some objects having high similarities between the classes on the road network extraction.

Considering the three yellow frames in Fig. 9, the strategy of the scene-extraction-to-fine-road-extraction scheme can reduce the incidence of false detection for nonroad backgrounds. This shows that our proposed two-stage method has worked, and that the extraction accuracy can be improved by modifying the two networks

B. Effectiveness of the Road-Scene Neighborhood Enhancement, Improved Dilated Convolution, and SELU Function

Comparative experiments were conducted to verify the effectiveness of different modules against the NIGAN. The experiments were mainly divided into the following categories: 1) NIGAN with the backbone of ResNet50 (NIGAN-50); 2) NIGAN with the backbone of ResNet101 (NIGAN-101); 3) NIGAN excluding dilated convolution; 4) NIGAN excluding the SELU activation function; and 5) NIGAN excluding neighborhood enhancement.

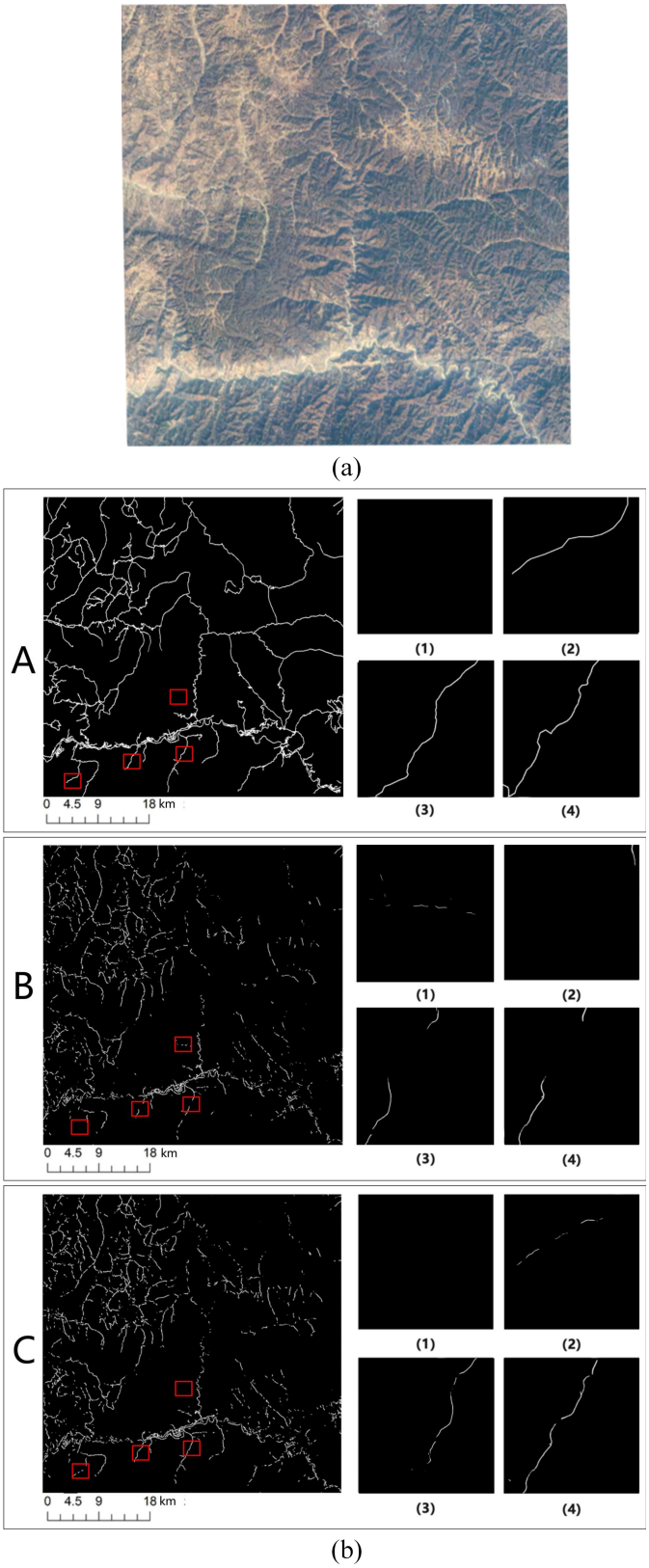


Fig. 8. Results of the fine extraction of roads in the entire area. (a) RGB image. (b) A: road labels; B: results of LinkNet34; C: results of the NIGAN red frames (1)–(4) are used to present local differences.

The results are summarized in Table VI. Theoretically, when the number of layers is deeper, deeper features can be extracted. However, there were only small road data samples

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
COMPARISON RESULTS OF THE FINE-ROAD EXTRACTION WITH DIFFERENT BACKBONES

Models	Backbone	Neighborhood information	SELU	Dilated Convolution	PA	MPA	MIOU
NIGAN-50	ResNet50	✓	✓	✓	96.25%± 0.51	77.09%± 0.67	69.51%± 0.81
NIGAN-101	ResNet101	✓	✓	✓	96.06%± 0.66	76.43%± 0.61	68.28%± 0.77
NIGAN	ResNet34	✓	×	×	95.96%± 0.91	74.55%± 1.01	67.04%± 0.98
NIGAN	ResNet34	✓	×	✓	96.09%± 0.82	74.95%± 0.92	67.68%± 0.91
NIGAN	ResNet34	✓	✓	×	96.37%± 0.73	77.04%± 0.76	69.92%± 0.86
NIGAN	ResNet34	×	✓	✓	96.34%± 0.57	81.15%± 0.62	71.65%± 0.76
NIGAN	**ResNet34**	✓	✓	✓	**96.47%**± 0.46	**81.82%**± 0.56	**72.10%**± 0.63

TABLE VII
OVERALL ACCURACY AND RECALL RATES OF EACH MODEL ON DIFFERENT RATIOS BETWEEN TRAINING AND VALIDATION. THE LISTED 9:1, 8:2, 5:5, 2:8, AND 1:9 ARE THE RATIOS OF THE TRAINING VERSUS VALIDATION DATASETS

Ratio Models	Index	9:1	8:2	5:5	2:8	1:9	Variance
AlexNet	Accuracy	93.68%	93.96%	91.49%	91.48%	91.40%	1.16%
	Recall	90.63%	89.83%	85.93%	86.88%	84.27%	2.39%
VGG11	Accuracy	93.68%	93.41%	91.65%	91.05%	91.00%	1.16%
	Recall	91.77%	89.83%	86.43%	81.48%	81.07%	4.31%
VGG16	Accuracy	94.89%	92.87%	91.42%	91.38%	90.43%	1.56%
	Recall	91.36%	89.83%	87.14%	83.26%	81.69%	3.71%
ResNet18	Accuracy	94.74%	93.96%	92.77%	92.05%	90.77%	1.40%
	Recall	90.63%	90.63%	87.86%	85.52%	79.83%	4.01%
BN-inception	Accuracy	94.74%	94.25%	92.21%	91.62%	91.02%	1.47%
	Recall	90.63%	89.81%	85.72%	85.97%	79.03%	4.11%
NasNetamobile	Accuracy	92.72%	92.31%	91.83%	91.24%	89.91%	0.98%
	Recall	90.12%	89.14%	85.60%	83.71%	81.45%	3.25%
DenseNet131	Accuracy	94.89%	93.96%	92.10%	91.48%	90.52%	1.61%
	Recall	90.53%	89.31%	87.50%	85.97%	83.87%	2.36%
NIGAN (Proposed)	Accuracy	**95.27%**	**95.11%**	**94.40%**	**93.50%**	**93.25%**	**0.82%**
	Recall	**92.59%**	**91.20%**	**90.28%**	**88.89%**	**87.89%**	**1.66%**

in this study. Excessive layers in the fine-road extraction mission delivered a less-than-perfect convergence effect, causing poor model prediction results related to insufficient training. When the proposed improved SELU and dilated convolution were excluded, the NIGAN achieved good results on the PA; however, it still had an approximately 7% gap with the NIGAN on MPA. This finding indicates that when excluding the improved dilated convolution, the framework failed to obtain global features. Accordingly, it had an insufficient ability to extract road information against complex features background.

Considering the inclusion of the improved RELU, it was possible for the NIGAN (excluding the improvement dilated convolution in the backpropagation process) to generate a

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

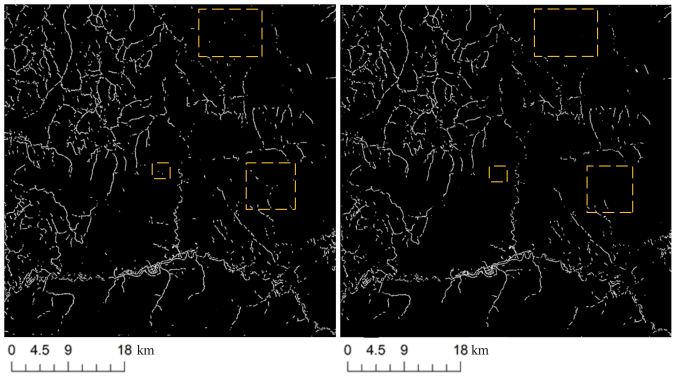


Fig. 9. Fine-road extraction results. (Left) NIGAN excluding the road-scene extraction. (Right) NIGAN including the road-scene extraction.

large gradient, which resulted in gradient disappearance for the input of the abnormal range of the neural network. Although an overridden historical weight can serve as a solution, the recognition result can be influenced by the gradient disappearance to generate less accurate road network extraction results. The NIGAN (excluding the neighborhood information) cannot effectively avoid the influence of high similarity between classes, which leads to low precision of road-scene extraction and affects the precision of fine-road extraction.

C. Efficacy of Adversarial Training on Road-Scene Extraction Under Small Samples

To further verify the ability of the NIGAN to extract road scenes under small samples, experiments were conducted on disparate ratio training and verification sets using the same parameter settings. In the proposed NIGAN, small sample features of CGAN were only used in the road-scene extraction section, so the following experiments verify the road-scene extraction task. The overall accuracy and recall rates of results are listed in Table VII.

The NIGAN achieved the highest accuracy under all five disparate ratios. As the proportion of the training set declined, the accuracy of each network correspondingly decreased. However, the NIGAN achieved the lowest reduction of accuracy. Moreover, the recall rate of the NIGAN was the best among all the five ratios, resulting in the highest integrity of road-scene extraction.

The traditional data enhancement methods, such as rotation, mirror image, and cutting, only increase the shallow features of the data, whereas the CGAN can make the generated data have more deep features by learning the distribution of the original data. Consequently, the generative adversarial training of the NIGAN played an important role in mitigating overfitting caused by small samples. Considering the training of the five varying ratios, the NIGAN had the lowest variance of accuracy, recall rate, and lower dispersion. This result further confirms the validity of the application of the NIGAN to the small samples.

D. Limitations

Although the proposed NIGAN framework in this article has achieved high-quality results, there are deficiencies in shadow areas, dense curved areas of roads, and mountainous winding areas. The reason is that the proportion of data is too low, so we did not pay enough attention to this part of the region in the process of model training. To address this issue, special training can be performed on these areas.

In addition, the spatial resolution of the constructed mountain road dataset was only 2.1 m in this study, which is a bit lower for road extraction. In the future, mountain road datasets with a resolution better than 1 m would be constructed based on GaoFen-2 satellite images to further verify the performance of the NIGAN model. Considering that there are fewer roads in mountainous and rural areas, the intraclass differences between roads, for example, earth roads and mountain roads, are clearly different in shape and color. The dataset needs to be further enhanced, such as by enhancing the patches that contain roads and selecting the area with more roads as the study area.

Furthermore, parameters such as loss function and optimizer are not well-adjusted. And some latest structures, such as transformer and graph convolution network, are not applied in the proposed model. Therefore, these problems need to be addressed in the follow-up work.

V. Conclusion

In this study, considering the complexity of mountain roads and the homogeneity of local spatial areas, to minimize the interference of context information and improve the integrity and accuracy of road extraction, we propose a hierarchical road extraction strategy from "road-scene classification" to "fine-road extraction" levels and construct a benchmark dataset including road scenes and label datasets for mountain road extraction. In the road-scene extraction section, a remote sensing road-scene neighborhood probability enhancement method is designed to be added to the CGAN for extracting road scenes in the study area. This method reduces the impact of high-level interclass similarities among feature images on subsequent fine-road extractions. In the fine-road extraction section, an improved dilated convolution module was added between the encoder and generator to improve the integrity and accuracy of road extraction. To optimize the performance of ResNet34 in extracting features, an improved SELU function was added to the generator to prevent gradient disappearance. The proposed framework was evaluated on our datasets and used to generate road scene and road network map. The main conclusions are as follows.

1) Compared with the five state-of-the-art models, the NIGAN improved road extraction accuracy in four mountain road scenes with diverse complexities.
2) The NIGAN framework achieved higher road extraction results regardless of whether the road-scene module was added. However, when the road-scene extraction section was added, the influence of surface features with high interclass similarity decreased. This finding reflects the effectiveness of the proposed road-scene extraction strategy first executed.
3) The improved dilated convolution cascade structure can further improve the feature-extraction ability of insignificant road features and enhance the integrity of road extraction.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

4) The improved SELU function effectively avoided mode collapse and gradient disappearance during the adversarial training in this study.
5) The NIGAN was equipped with excellent generalization ability for road-scene extraction under small samples.

Future work will pay attention to enhancing road extraction capacities in specific complex scenes by adding the attention mechanism.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

[1] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sens.*, vol. 12, no. 9, p. 1444, May 2020.

[2] J. Zhang, Q. Hu, J. Li, and M. Ai, "Learning from GPS trajectories of floating car for CNN-based urban road extraction with high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1836–1847, Mar. 2021.

[3] Y. Zhang, Y. Li, X. Zhou, X. Kong, and J. Luo, "Curb-GAN: Conditional urban traffic estimation through spatio-temporal generative adversarial networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 842–852.

[4] A. Broggi *et al.*, "PROUD-public road urban driverless test: Architecture and results," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 648–654.

[5] S. Wang, X. Mu, D. Yang, H. He, and P. Zhao, "Road extraction from remote sensing images using the inner convolution integrated encoder–decoder network and directional conditional random fields," *Remote Sens.*, vol. 13, no. 3, p. 465, Jan. 2021.

[6] J. Xu *et al.*, "Multi-scale network based on dilated convolution for bladder tumor segmentation of two-dimensional MRI images," in *Proc. 15th IEEE Int. Conf. Signal Process. (ICSP)*, Dec. 2020, pp. 533–536.

[7] F. Bastani *et al.*, "RoadTracer: Automatic extraction of road networks from aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4720–4728.

[8] 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.

[9] Z. Miao, W. Shi, A. Samat, G. Lisini, and P. Gamba, "Information fusion for urban road extraction from VHR optical satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1817–1829, May 2016.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[11] Y.-C. Chen, X. Xu, and J. Jia, "Domain adaptive image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5274–5283.

[12] L. Ma *et al.*, "Capsule-based networks for road marking extraction and classification from mobile LiDAR point clouds," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 1981–1995, Apr. 2021.

[13] R. Niu, X. Sun, Y. Tian, W. Diao, K. Chen, and K. Fu, "Hybrid multiple attention network for semantic segmentation in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2021.

[14] Z. Zhang and Y. Wang, "JointNet: A common neural network for road and building extraction," *Remote Sens.*, vol. 11, no. 6, p. 696, 2019.

[15] Z. Xu *et al.*, "Road extraction in mountainous regions from high-resolution images based on DSDNet and terrain optimization," *Remote Sens.*, vol. 13, no. 1, p. 90, Dec. 2020.

[16] S. P. Kearney, N. C. Coops, S. Sethi, and G. B. Stenhouse, "Maintaining accurate, current, rural road network data: An extraction and updating routine using RapidEye, participatory GIS and deep learning," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 87, May 2020, Art. no. 102031.

[17] Z. Shao, Z. Zhou, X. Huang, and Y. Zhang, "MRENet: Simultaneous extraction of road surface and road centerline in complex urban scenes from very high-resolution images," *Remote Sens.*, vol. 13, no. 2, p. 239, Jan. 2021.

[18] J. Wan, Z. Xie, Y. Xu, S. Chen, and Q. Qiu, "DA-RoadNet: A dual-attention network for road extraction from high resolution satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6302–6315, 2021.

[19] R. Lian, W. Wang, N. Mustafa, and L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5489–5507, 2020.

[20] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-deeplab model," *Remote Sens.*, vol. 12, no. 18, p. 2985, Sep. 2020.

[21] Q. Zhu *et al.*, "A global context-aware and batch-independent network for road extraction from VHR satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 353–365, May 2021.

[22] A. Baumgartner, C. Steger, H. Mayer, W. Eckstein, and H. Ebner, "Automatic road extraction based on multi-scale, grouping, and context," *Photogramm. Eng. Remote Sens.*, vol. 65, pp. 777–786, Jul. 1999.

[23] G. Cheng, Y. Wang, Y. Gong, F. Zhu, and C. Pan, "Urban road extraction via graph cuts based probability propagation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5072–5076.

[24] C. Zhang, S. Murai, and E. P. Baltsavias, "Road network detection by mathematical morphology," in *Proc. ISPRS Workshop 3D Geospatial Data Prod., Meeting Appl. Requirements*. Zürich, Switzerland: ETH-Hönggerberg, 1999.

[25] M. Amo, F. Martínez, and M. Torre, "Road extraction from aerial images using a region competition algorithm," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1192–1201, May 2006.

[26] M. Li, A. Stein, W. Bijker, and Q. Zhan, "Region-based urban road extraction from VHR satellite images using binary partition tree," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 44, pp. 217–225, Feb. 2016.

[27] R. Lian and L. Huang, "DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1905–1916, 2020.

[28] P. Gamba, F. Dell'Acqua, and G. Lisini, "Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 387–391, Jul. 2006.

[29] Q. Guo and Z. Wang, "A self-supervised learning framework for road centerline extraction from high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4451–4461, 2020.

[30] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, "Road network extraction and intersection detection from aerial images by tracking road footprints," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4144–4157, Dec. 2007.

[31] F. M. Porikli, "Road extraction by point-wise Gaussian models," *Proc. SPIE*, vol. 5093, pp. 758–764, Sep. 2003.

[32] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179424–179436, 2020.

[33] Z. Chen, C. Wang, J. Li, N. Xie, Y. Han, and J. Du, "Reconstruction bias U-Net for road extraction from optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2284–2294, 2021.

[34] Y. Wang, J. Seo, and T. Jeon, "NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[35] W. Zhou, Y. Wang, J. Chu, J. Yang, X. Bai, and Y. Xu, "Affinity space adaptation for semantic segmentation across domains," *IEEE Trans. Image Process.*, vol. 30, pp. 2549–2561, 2021.

[36] 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.

[37] L. Ding and L. Bruzzone, "DiResNet: Direction-aware residual network for road extraction in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10243–10254, Dec. 2021.

[38] Y. Li, Q. Wang, J. Zhang, L. Hu, and W. Ouyang, "The theoretical research of generative adversarial networks: An overview," *Neurocomputing*, vol. 435, pp. 26–41, May 2021.

[39] Y. Lu, Y. Chen, D. Zhao, and J. Chen, "Graph-FCN for image semantic segmentation," in *Proc. Int. Symp. Neural Netw.* Cham, Switzerland: Springer, 2019, pp. 97–105.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

[40] L. Ma, Y. Li, J. Li, J. M. Junior, W. N. Goncalves, and M. A. Chapman, "BoundaryNet: Extraction and completion of road boundaries with deep learning using mobile laser scanning point clouds and satellite imagery," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5638–5654, Jun. 2022.

[41] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021.

[42] Y. Chen, A. Dapogny, and M. Cord, "SEMEDA: Enhancing segmentation precision with semantic edge aware loss," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107557.

[43] D. Pan, M. Zhang, and B. Zhang, "A generic FCN-based approach for the road-network extraction from VHR remote sensing images—Using openstreetmap as benchmarks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2662–2673, 2021.

[44] G. Zhou, W. Chen, Q. Gui, X. Li, and L. Wang, "Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.

[45] B. Xu, S. Bao, L. Zheng, G. Zhang, and W. Wu, "IDANet: Iterative D-LinkNets with attention for road extraction from high-resolution satellite imagery," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2021, pp. 140–152.

[46] C. Miao, Z. Zhang, and Q. Tian, "TransLinkNet: LinkNet with transformer for road extraction," *Proc. SPIE*, vol. 12173, pp. 138–143, May 2022.

[47] Z. Zhang, C. Miao, C. Liu, and Q. Tian, "DCS-TransUperNet: Road segmentation network based on CSwin transformer with dual resolution," *Appl. Sci.*, vol. 12, no. 7, p. 3511, Mar. 2022.

[48] Z. Yang, D. Zhou, Y. Yang, J. Zhang, and Z. Chen, "TransRoadNet: A novel road extraction method for remote sensing images via combining high-level semantic feature and context," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[49] N. Wambugu *et al.*, "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102603.

[50] X. Wang, K. Tan, P. Du, C. Pan, and J. Ding, "A unified multiscale learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.

[51] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGAN: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019.

[52] P. Jian, K. Chen, and W. Cheng, "GAN-based one-class classification for remote-sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[53] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, Sep. 2021.

[54] S. Sun, L. Mu, L. Wang, P. Liu, X. Liu, and Y. Zhang, "Semantic segmentation for buildings of large intra-class variation in remote sensing images with O-GAN," *Remote Sens.*, vol. 13, no. 3, p. 475, Jan. 2021.

[55] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Multi-object segmentation in complex urban scenes from high-resolution remote sensing data," *Remote Sens.*, vol. 13, no. 18, p. 3710, Sep. 2021.

[56] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process.*, vol. 35, no. 1, pp. 53–65, Jan. 2017.

[57] M. Lee and J. Seok, "Regularization methods for generative adversarial networks: An overview of recent studies," 2020, *arXiv:2005.09165*.

[58] N. Varia, A. Dokania, and J. Senthilnath, "DeepExt: A convolution neural network for road extraction using RGB images captured by UAV," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 1890–1895.

[59] C. Yang and Z. Wang, "An ensemble Wasserstein generative adversarial network method for road extraction from high resolution remote sensing images in rural areas," *IEEE Access*, vol. 8, pp. 174317–174324, 2020.

[60] D. Costea, A. Marcu, M. Leordeanu, and E. Slusanschi, "Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2100–2109.

[61] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, and L. Jiao, "Aerial image road extraction based on an improved generative adversarial network," *Remote Sens.*, vol. 11, no. 8, p. 930, Apr. 2019.

[62] Y. Zhang, Z. Xiong, Y. Zang, C. Wang, J. Li, and X. Li, "Topology-aware road network extraction via multi-supervised generative adversarial networks," *Remote Sens.*, vol. 11, no. 9, p. 1017, Apr. 2019.

[63] A. Abdollahi, B. Pradhan, and N. Shukla, "Road extraction from high-resolution orthophoto images using convolutional neural network," *J. Indian Soc. Remote Sens.*, vol. 49, no. 3, pp. 569–583, Mar. 2021.

[64] A. Hu, S. Chen, L. Wu, Z. Xie, Q. Qiu, and Y. Xu, "WSGAN: An improved generative adversarial network for remote sensing image road network extraction by weakly supervised processing," *Remote Sens.*, vol. 13, no. 13, p. 2506, Jun. 2021.

[65] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[66] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.

[67] V. Mnih, *Machine Learning for Aerial Image Labeling*. Toronto, ON, Canada: Univ. of Toronto, 2013.

[68] M. Ehlers, S. Klonus, P. J. Åstrand, and P. Rosso, "Multi-sensor image fusion for pansharpening in remote sensing," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 25–45, Mar. 2010.

[69] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified U-Net architecture for semantic segmentation of diabetic retinopathy images," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1094–1109, Jul. 2020.

[70] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Perez, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2517–2526.

[71] B. Zoph *et al.*, "Rethinking pre-training and self-training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 3833–3845.

[72] K. He, R. Girshick, and P. Dollár, "Rethinking ImageNet pre-training," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4918–4927.

[73] 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 (CVPRW)*, Jun. 2018, pp. 182–186.

[74] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[75] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[76] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[77] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[78] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[79] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.

[80] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[81] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

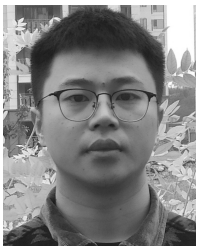
[82] S. Woo, D. Kim, D. Cho, and I. S. Kweon, "LinkNet: Relational embedding for scene graph," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1–11.

[83] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2015.

Weitao Chen (Member, IEEE) was born in Wugang, Henan, China. He received the B.Eng. degree from the Jiaozuo Institute of Technology, Jiaozuo, China, in 2003, and the M.E. and Ph.D. degrees from the China University of Geosciences, Wuhan, China, in 2006 and 2012, respectively.

He is currently a Professor with the School of Computer Science, China University of Geosciences. He has published more than 30 articles. His main research interests include machine learning and remote sensing of environment.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.

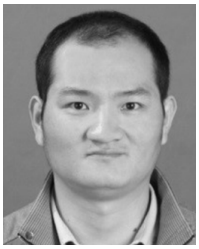


Gaodian Zhou received the B.Eng. and M.E. degrees from the China University of Geosciences, Wuhan, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the School of Computer Science.

His research interests are semantic segmentation, remote sensing image process, and big data.

Zhuoyue Liu was born in Chongqing, China. He received the B.Eng. degree from the Hefei University of Technology, Hefei, China, in 2018, and the M.E. degree from the China University of Geosciences, Wuhan, China, in 2021.

His research interests are data generation, semantic segmentation, and remote sensing image process.



Xianju Li received the B.S., M.S., and Ph.D. degrees from the China University of Geosciences, Wuhan, China, in 2009, 2012, and 2016, respectively.

Since 2016, he has been an Associate Professor with the School of Computer Science, China University of Geosciences. He has published more than ten articles. His main research fields include remote sensing image processing and analysis, computer vision, and machine learning.

Xiongwei Zheng was born in Tianmen, Hubei, China. He received the bachelor's degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009. He is currently pursuing the Ph.D. degree in geoscience information engineering with the China University of Geosciences, Wuhan.

He is currently a professor level senior engineer. He is also the Director of the Big Data Center of China Airborne Geophysical and Remote Sensing Center for Natural Resources. His research directions include data acquisition and processing of satellite multispectral, hyperspectral, laser sounding, and radar remote sensing.

Lizhe Wang (Fellow, IEEE) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the D.E. degree from the University of Karlsruhe, Karlsruhe, Germany, in 2007.

He is currently the ChuTian Chair Professor of the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include high performance computing (HPC), e-science, and remote sensing image processing.

Prof. Wang is a fellow of the Institution of Engineering and Technology (IET Fellow) and British Computer Society.

Authorized licensed use limited to: Wuhan University. Downloaded on May 09,2023 at 02:59:05 UTC from IEEE Xplore. Restrictions apply.