# RoadSeg-CD: A Network With Connectivity Array and Direction Map for Road Extraction From SAR Images

Fei Gao, Jun Tu, Jun Wang, Amir Hussain, and Huiyu Zhou

***Abstract*—Road extraction from synthetic aperture radar (SAR) images has attracted much attention in the field of remote sensing image processing. General road extraction algorithms, affected by shadows of buildings and trees, are prone to producing fragmented road segments. To improve the accuracy and completeness of road extraction, we propose a neural network-based algorithm, which takes the connectivity and direction features of roads into consideration, named RoadSeg-CD. It consists of two branches: one is the main branch for road segmentation; the other is the auxiliary branch for learning road directions. In the main branch, a connectivity array is designed to utilize local contextual information and construct a connectivity loss based on the predicted probabilities of neighboring pixels. In the auxiliary branch, we proposed a novel road direction map, which is used for learning the directions of roads. The two branches are connected by specific feature fusion process, and the output from the main branch is taken as the road extraction result. Experiments on real radar images are implemented to validate the effectiveness of our method. The experimental results demonstrate that our method can obtain more continuous and more complete roads than several state-of-the-art road extraction algorithms.**

***Index Terms*—Connectivity array, direction map, neural network, road extraction, synthetic aperture radar (SAR).**

## I. Introduction

ROADS are valuable transportation facilities. Real-time access to road network information is essential for many practical applications, such as traffic management and map update. Synthetic aperture radar (SAR) has become an important tool for obtaining road network information, because it can achieve the imaging of the earth's surface in all-weather and all-time [1]–[3]. Therefore, in the field of remote sensing, it has attracted considerable attention on how to accurately and automatically extract roads from SAR images in recent years.

In early research works, people mainly focused on the linear features of road edges or other man-made features. The extraction methods can be divided into pixel-based and object-based methods. In pixel-based methods, road features are extracted at the pixel level. Typical pixel-based methods are the edge analysis methods, where road extraction process generally consists of three steps. First, edge detectors are used to extract the pixels from road edges, such as ratio of averages (ROA) operator [4], [5], Duda operator [6], etc. Second, the high gradient pixels are connected to generate road segments based on Hough transform [7], [8], Radon transform [9], [10], etc. Finally, these segments are globally connected to form a complete road network. The global connection is usually the most critical step, where there are many algorithms such as Markov random field (MRF) model [11], [12], genetic algorithm [13], perceptual grouping [14], etc. However, the performance is limited by the background complexity in SAR images because of the speckle noise and the excessive dependence on the road edges. Meanwhile, the object-based methods are focused on the ribbon-like structures of roads. In these methods, adjacent pixels with similar features are merged into objects [15]. For example, fuzzy clustering algorithms were used for road area segmentation in [16]. Grinias *et al.* [17] utilized a MRF-based algorithm to achieve segmented regions and classified these regions by a random forest algorithm to detect roads. Superpixel-based methods can also achieve merged regions [18]. In [19], an entropy rate superpixel algorithm was used to segment the image into small homogeneous regions. Then multiple features including intensity and texture of superpixels were extracted, and the superpixels with similar features were merged into large objects based on graph model. To classify these objects into different types, Gaussian mixture model was applied to cluster objects and extract the road layer finally. Nevertheless, when using these traditional methods with artificially extracted features, it is not only difficult to give appropriate parameters related to manually designed road features, but also requires enormous calculations. Consequently, the accuracy of the traditional methods is limited, and it is difficult to guarantee the real-time performance.

Manuscript received September 25, 2021; revised December 22, 2021, March 4, 2022, April 19, 2022, and May 8, 2022; accepted May 13, 2022. Date of publication May 19, 2022; date of current version May 26, 2022. This work was supported by the National Natural Science Foundation of China under Grant 61771027, Grant 61071139, Grant 61471019, Grant 61501011, and Grant 61171122. The work of A. Hussain was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/M026981/1, Grant EP/T021063/1, and Grant EP/T024917/1. The work of H. Zhou was supported by the Royal Society-Newton Advanced Fellowship under Grant NA160342, the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska Curie Grant 720325. *(Corresponding author: Jun Wang.)*

Fei Gao, Jun Tu, and Jun Wang are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: feigao2000@163.com; tujun@buaa.edu.cn; wangj203@buaa.edu.cn).

Amir Hussain is with the Cyber and Big Data Research Laboratory, Edinburgh Napier University, EH11 4BN Edinburgh, U.K., and also with Taibah Valley, Taibah University, Medina 30001, Saudi Arabia (e-mail: a.hussain@napier.ac.uk).

Huiyu Zhou is with the Department of Informatics, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: hz143@leicester.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2022.3175594

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Recently, with the development of deep learning technology, deep convolutional neural network (DCNN)-based algorithms are widely applied in image classification and segmentation tasks [20]–[23]. In terms of semantic segmentation, it is particularly common to apply the fully convolution network (FCN)-based algorithms [24], such as U-Net [25], SegNet [26], and LinkNet [27]. For instance, in SegNet, a symmetric encoder–decoder structure is constructed, and the upsampling process in the decoding layer is simplified by the max-unpooling layer. In LinkNet, channel reduction is introduced to reduce the parameters in the decoder. Since the road extraction task can be regarded as segmenting roads from surrounding background, the above mentioned networks are widely used to extract roads from remote sensing images. For example, Li *et al.* [28] resplit the various layers of U-Net and build an ensemble learning model to extract roads from satellite images. Zhang *et al.* [29] introduced the residual block based on U-Net to form ResUnet for road extraction tasks. ResUnet promotes the dissemination of information in the network by skip connections, which is in favor of the extraction of road features. In [30], a cascaded convolutional neural network is built based on FCN and deconvolution layer to perform road extraction and road centerline extraction simultaneously. Based on LinkNet, Zhou *et al.* [31] introduced dilated convolution layers to form D-LinkNet. Because the additional dilated convolution can enlarge the receptive field without decreasing the resolution of feature maps, the continuity of the road extracted by this network is improved. In addition to using a single road segmentation network, Batra *et al.* [32] designed a road direction branch based on the vector data of road centerlines to form a double branch network to improve the road extraction performance. However, it assumes that all roads are of the same widths when constructing the direction maps, thus it is not suitable for extracting diverse kinds of roads in the large-scene SAR images. Ding *et al.* [33] also introduced road direction supervision, but only four main directions are considered and the network performance is limited. Zhou *et al.* [34] proposed a method that focus on the boundary quality but not on regional accuracy, and designed a spatial context module to obtain complete roads. Similar to the function of the spatial context module, a dense dilated spatial pyramid pooling module is designed by Abdollahi *et al.* [35] to produce more scale features over a broader range. Chen *et al.* [36] used an atrous spatial pyramid pooling module to aggregate contextual information in DeepLabv3+. However, these spatial modules enlarge the receptive field, which introduces some invalid context information and lead to errors in the extracted slender roads. Tao *et al.* [37] designed a spatial information inference structure with the three-dimensional convolution to capture the local topology of narrow roads and enhance the completeness of extracted roads.

Compared with traditional methods, the above mentioned deep learning technologies for road extraction can automatically extract more hierarchical features of roads from SAR images and ensure higher efficiency. Abdollahi *et al.* [38] presented a review of deep learning approaches applied to road extraction from remote sensing images and emphasized the considerable efficiency. Henry *et al.* [39] applied three FCN-based algorithms for road extraction in SAR images and demonstrated the potential of deep learning techniques. Nevertheless, there are still certain challenges in applying deep learning technology on road extraction, and the main difficulties are as follows. First, the widths of roads in various areas are not the same. It is necessary to extract road features on multiple scales. Second, compared to the surrounding environment and common objects, roads are generally slender. It is difficult to design a network architecture to take advantage of the linear feature. Third, since the roads in SAR images are easily occluded by the shadows of surrounding buildings and other obstacles, there is a lack of guarantee on the continuity of roads extracted by ordinary neural networks. Therefore, it is worthy of further research to extract more complete roads with fewer fractures from SAR images.

In view of the above challenges, to improve the continuity of road extraction results, a novel neural network-based algorithm with connectivity array and direction map for road extraction from SAR images is proposed in this article, namely RoadSeg-CD. First, based on the predicted probabilities of neighboring pixels, a connectivity array is designed to construct connectivity loss. Next, the connectivity loss and cross-entropy loss are integrated in the main branch of the network for supervision. For the main branch, the SAR images are used as input and the road networks are the output. In the meantime, the auxiliary branch is constructed to learn road directions. For the auxiliary branch, the local directions of SAR images are taken as input and the road direction maps are the output. Last, the two branches are connected by a specifically designed feature fusion process and shared encoder. The main branch output is adopted as the final road extraction result. The main contributions of this article are summarized as follows.

1) A connectivity array is designed to construct a new multiscale connectivity loss. This loss focuses on the connections between pixels at multiple scales because of various road widths. After introducing the loss into the network, the extraction of local contextual information can be enhanced.
2) In order to learn road directions to enhance the road extraction performance, we have designed a novel road direction map. It refines the road network topology based on tortuous roads with complex structure. As for linear targets like roads, the learning of directions can substantially deepen the use of linear features to improve the extraction accuracy.
3) A dual-branch network with dual inputs and dual outputs is proposed. By well-designed information sharing, the auxiliary branch is able to improve the road extraction performance of the main branch.

The rest of this article is organized as follows. Section II illustrates the proposed RoadSeg-CD network in detail. In Section III, our method is evaluated on real SAR images. Finally, Section IV concludes this article.

## II. Methodology

This section details the proposed road extraction method. First, the specific structure of RoadSeg-CD network is

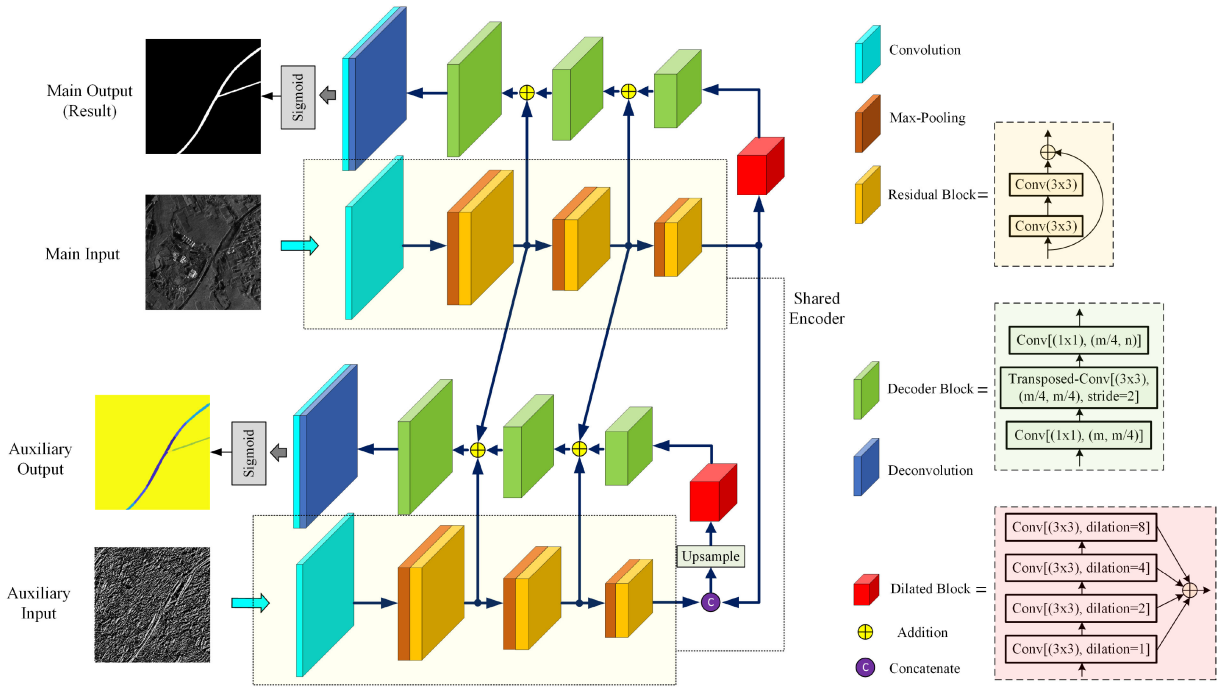


Fig. 1. Structure of the proposed RoadSeg-CD network. It consists of two branches with similar structures. The main branch takes the road network as output to learn the main road features including topological information. The auxiliary branch takes the road direction map as output to learn the road directions and improve the feature extraction performance in the main branch.

introduced. Second, the calculation process of the connectivity loss in the main branch and the direction map in the auxiliary branch are, respectively, described. Finally, we will explain the overall loss of this network.

### A. *Network Architecture*

Road directions can help network learn the geometric features of the linear targets like roads. To learn road directions, a network branch that takes road directions as the output is indispensable. Thus, we construct a dual-branch network RoadSeg-CD, whose structure is shown in Fig. 1. On one side, the main branch is a road extraction network based on segmentation, which takes SAR images as input and road network as output. This branch is responsible for learning the main road features such as topological relationship, texture, context, etc. On the other side, the auxiliary branch learns the road directions. Similar to [32], the auxiliary branch takes road direction maps as the output, where each pixel represents the direction of corresponding road segment. However, the direction branch in [32] takes SAR images as input, which is lack of pertinence to the directions. Our network takes the local direction maps of SAR images as the input, where each pixel value represents a direction, which is perpendicular to the gradient direction. At the road edges, most of local directions are relatively close to the directions of corresponding road segments. These similar local directions are obviously clustered in space because of the linear features of slender roads. Thus, the auxiliary branch can learn the linear features of roads. On the whole, by means of the feature fusion process and shared encoder between the two branches, the auxiliary branch can fine-tune the learning process of the main branch during training, thereby improving the road extraction performance.

In order to facilitate the feature fusion between two branches, similar structures are applied for these two branches. First of all, in terms of encoders, both branches adopt the lightweight and high-performance ResNet as the feature extraction backbone. ResNet is widely applied in road extraction networks as it has the advantage of balancing performance and the complexity of the model by controlling the depth of the network [31]–[34], [40]. To connect the two branches, the parameters of both encoders are shared with each other. Thus, the parameters in the main branch can be influenced by the training of the auxiliary branch. Second, a dilation module is inserted between the encoder and decoder for each branch. The dilation module is composed of four dilated convolutional layers, in which the dilation rates are different so that they have diverse receptive fields [31]. Next, we apply the same decoder as in LinkNet to decode the features. The decoder uses $1 \times 1$ convolution for channel reduction to decrease the training parameters, so as to constrain the model complexity to alleviate overfitting on a small dataset of SAR images. Since the outputs of two branches are different, the decoder parameters are not shared. Finally, both branches use the sigmoid layers to normalize the outputs. For the main branch, the outputs represent the predicted probabilities of being road of all pixels. For the auxiliary branch, the output data are mapped to the full angle range to represent road direction maps. In addition, for the connection of two branches, the decoder of the auxiliary branch is fused with the multiscale features from the encoder of the main branch. Different from the addition operation between feature maps of the encoder and decoder within each branch, the

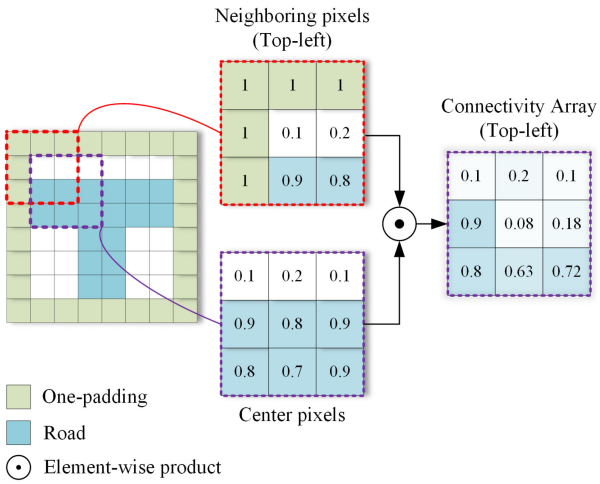


Fig. 2. Illustration of the calculation process of the connectivity array. The boundary of the original prediction probability matrix is padded with ones. The adjacent probabilities are multiplied to obtain a component of the connectivity array. For example, the top-left component of the connectivity array indicates the connectivity strength between each pixel in the road prediction result and its top left neighbor.

concatenation operation is adopted in the feature fusion between two branches because the extracted features of two branches are quite different. Due to the feature fusion and the shared encoder, during the back propagation of the auxiliary branch training, the learning of the encoder parameters in the main branch can be fine-tuned by the auxiliary branch to improve the road extraction performance during joint training.

### B. Connectivity

In addition to semantic segmentation, road extraction can also be viewed as a problem of finding connected regions. Thus, from the perspective of connected regions, a connectivity loss is introduced in the main branch to strengthen the learning of local contextual information. Similar to the method of designing connectivity loss in [41], a connectivity array is constructed in this article. However, the connectivity array in [41] is only composed of 0 and 1, so that the connectivity loss is too rough to accurately reflect the difference in connectivity between the extraction result and the ground truth (GT). Therefore, a new type of connectivity array is designed in this article. The array is calculated according to the predicted probabilities of being road of neighboring pixels. In addition, a connectivity array is divided into eight components based on eight-neighbor connectivity. Fig. 2 illustrates how a component of the connectivity array is calculated using a small example.

In Fig. 2, the numbers in the matrix on the left of the element-wise product operator represent the predicted probabilities of being road. In other word, these numbers are the output of the main branch. First, the boundary of each output probability map is padded with ones, which represent roads. This operation is called "one-padding," which can ensure the continuity of the road at the edge of the slice. Then, the product of a pair of neighboring pixels is defined as a component of connectivity. Finally, the normalized sum of all components is called the connectivity between the pixel and its neighbor. The entire matrix, which contains all connectivity values is called the connectivity array, which is calculated by

$$C(x,y) = \frac{1}{8}\sum_{i=-1}^{1}\sum_{j=-1}^{1} P(x+i,y+j)P(x,y),\ i^2+j^2 \neq 0 \tag{1}$$

where $P(x,y)$ is the predicted probability of being road of pixel $(x,y)$. For the GT, the probabilities only contain binary values 0 and 1. When a pixel and its neighbors are both labeled as road, the connectivity is the maximum value 1. When a pixel is an isolated point, the connectivity is the minimum value 0. For all background pixels in GT, the connectivity calculated by (1) is zero, which reduces the influence of noises in the learning process.

Considering the diversity of road widths in SAR images, the connectivity under different receptive fields should not be neglected. Therefore, the connectivity loss is calculated based on the connectivity arrays in multiple scales, as shown in Fig. 3.

First of all, the output of the main branch of RoadSeg-CD and the corresponding GT are both down-sampled by maximum pooling in multiple scales, and then the connectivity arrays are calculated by (1). Next, the MAEs of the connectivity arrays in different scales are taken as the connectivity loss in the corresponding scale. Finally, the weighted sum of all MAEs is the entire connectivity loss. The connectivity loss is specifically defined as

$$\mathcal{L}_{\text{conn}}^{k} = \frac{1}{N_k}\sum_{i=1}^{N_k}\left|C_i^k(T) - C_i^k(P)\right|$$

$$\mathcal{L}_{\text{conn}} = \frac{1-\alpha}{1-\alpha^m}\sum_{k=0}^{m-1}\alpha^k\mathcal{L}_{\text{conn}}^{k}$$

where $T$ is the GT of roads and $P$ is the output of the main branch. $k$ identifies the scale. The receptive field is smaller with a small value of $k$, when calculating the corresponding loss. $C_i^k$ is the $i_{th}$ value of the connectivity array at the $k_{th}$ scale. $N_k$ is the number of elements in the connectivity array at the $k_{th}$ scale. $\mathcal{L}_{\text{conn}}^{k}$ is the connectivity loss at the $k_{th}$ scale. $\mathcal{L}_{\text{conn}}$ is the entire connectivity loss. $m$ is the number of scales. $\alpha$ is a weight parameter with the range of $0 < \alpha < 1$. It represents the influence of the connectivity loss at a scale with a larger $k$ on the entire connectivity loss. The value of the entire connectivity loss reaches the maximum 1, if and only if the connectivity loss at each scale equals the maximum 1. Therefore, the connectivity loss defined by (2) is a normalized loss.

### C. Direction Learning

The direction learning task is performed by the auxiliary branch of RoadSeg-CD. The inputs are the local directions, which are calculated by

$$G_x(a,b) = \log\left(\frac{I(a+1,b)+I(a+1,b+1)}{I(a,b)+I(a,b+1)}\right)$$

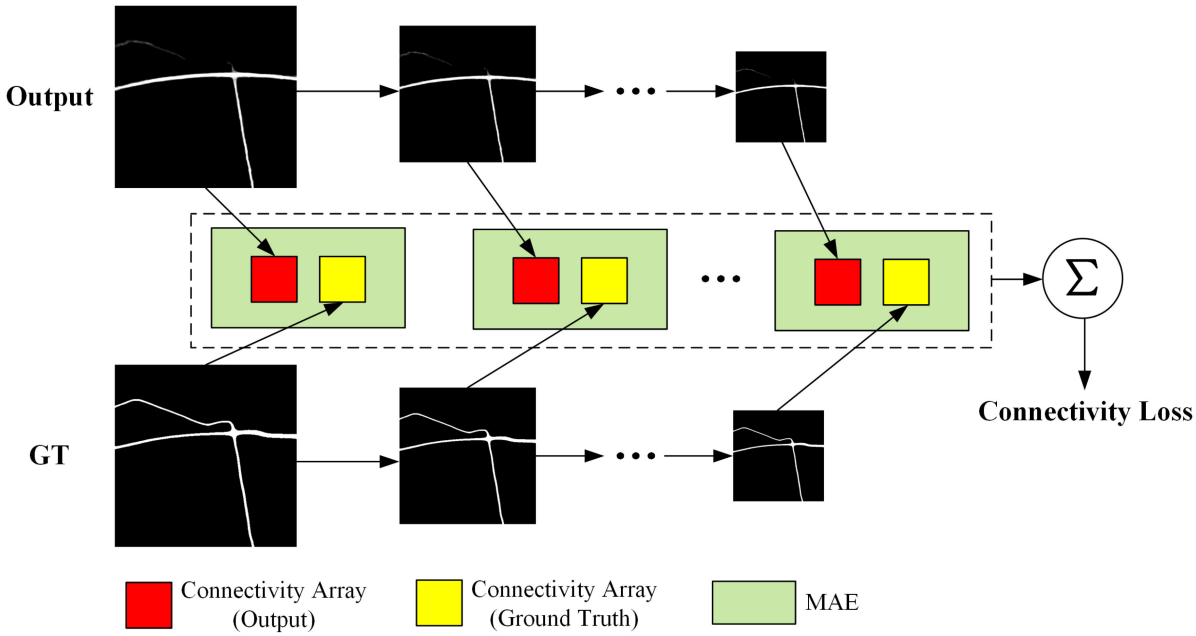


Fig. 3. Calculation method of the connectivity loss. The mean absolute error (MAE) of the connectivity arrays of the output and GT in multiple scales is calculated first. Then the weighted sum of MAEs is the connectivity loss.

$$G_y(a,b) = \log\left(\frac{I(a,b+1)+I(a+1,b+1)}{I(a,b)+I(a+1,b)}\right)$$

$$G_t(a,b) = \arctan\left(\frac{G_x(a,b)}{-G_y(a,b)}\right)$$

where $I$ is the gray value of the input SAR images. $G_x(a,b)$ and $G_y(a,b)$ denote the horizontal and vertical gradients of the pixel $(a,b)$, respectively. $G_t(a,b)$ is the local direction. The method for calculating local directions was proposed by Liu *et al.* [42] when researching the extraction of linear targets in SAR images. As the SAR images are affected by inherent multiplicative speckle noise, the calculation of each value in (2) is realized by the ratio-based method. At the same time, to distinguish the positive and negative of local directions, the logarithmic function is introduced.

The output of the auxiliary branch is the road direction map. Regarding the acquisition of the GT of the direction map, a method relying on road vector data is given in [32]. It assigns the direction of road centerline to neighboring pixels lying within a fixed distance. However, this method not only requires the assumption that the widths of all parts of roads are the same, but also has difficulty obtaining orderly directions near the road intersections. Therefore, a new algorithm for calculating of the road direction map is designed based on the GT of roads. The overall process is shown in Fig. 4.

The overall calculation process of road direction map consists of three steps. First, the road skeletons are obtained by skeletonizing the road GT. Second, the road skeletons are approximated with a series of line strings by the Ramer–Douglas–Peucker (RDP) algorithm, which is a classical and effective algorithm for the polygonal approximation of plane curves

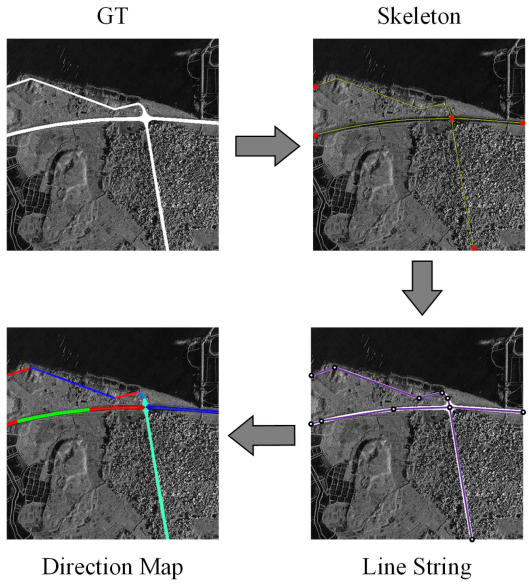


Fig. 4. Calculation process of road direction map. The process starts from the road GT, and the results of each stage are road skeleton, road line strings, and road direction map, respectively. In the direction map, various colors visually distinguish adjacent road line strings with different directions.

[43], [44]. Finally, to obtain the direction map, the direction of each line string is computed and assigned to the road pixels according to the distances to the line string. By approximating the road directions as the directions of numerous line strings connected end to end, the proposed method can not only accurately reflect the road directions on large receptive field, but also ensure the continuity of the road directions. The skeletonization in the first step is a classical morphological operation, which reduces

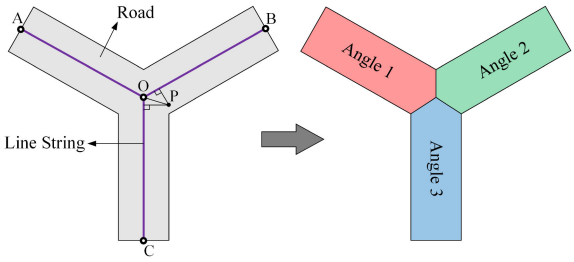


Fig. 5. Example of computing the road direction map. The purple solid lines are the road line strings. The gray area is the road. The direction value of each point in the road area is assigned based on the nearest line string. Different colors are used to distinguish different directions in the road direction map.

roads to one-pixel-wide representations. Because the process of skeletonization is relatively simple, and the RDP algorithm is well-known. The following mainly introduces the final step. The road direction map can be calculated as shown in Fig. 5.

Fig. 5 shows a typical "Y"-shaped intersection. To be specific, first, for a point $P$ located in the road area, the distances between it and the nearby line strings are calculated. If the distance between $P$ and the line string $I$ is the shortest, the direction of $I$ is taken as the direction at point $P$. Finally, after traversing all points in the road area, this area is divided into three areas with different directions to obtain a road direction map. The algorithm not only effectively avoids the confusion of the directions at the road intersection, but also makes the effective area of direction map exactly the same as the road area.

### D. Loss Function

The proposed RoadSeg-CD contains three different supervisions: segmentation supervision, connectivity supervision, and direction supervision. Thus, a hybrid loss is calculated on these supervisions and defined as

$$\mathcal{L} = w_1\mathcal{L}_{\text{seg}} + w_2\mathcal{L}_{\text{conn}} + w_3\mathcal{L}_{\text{direct}} \tag{2}$$

where $\mathcal{L}_{\text{seg}}$, $\mathcal{L}_{\text{conn}}$, and $\mathcal{L}_{\text{direct}}$ are the segmentation loss, the connectivity loss, and the direction loss, respectively. The first two are the losses of the main branch of RoadSeg-CD. The last one is the loss of the auxiliary branch. $w_1$, $w_2$, and $w_3$ are three weight parameters for different losses.

*1) Segmentation Loss:* In the main branch, the road segmentation results are probability maps, while the GT of roads are binary. Therefore, the common binary cross entropy loss is used as a measure of the difference between predictions and targets, calculated by

$$\mathcal{L}_{\text{seg}} = -\sum_{i=1}^{N}\left[t_i \log p_i + (1-t_i)\log(1-p_i)\right] \tag{3}$$

where $t_i \in \{0,1\}$ is the target value of the $i_{th}$ pixel in SAR images. $p_i$ is the predicted probability of being road of this pixel. $N$ is the total number of pixels in SAR images [45].

*2) Connectivity Loss:* This loss is part of the main branch together with the segmentation loss. It has been discussed in Section II-B and is calculated by (2).

*3) Direction Loss:* As shown in Fig. 1, the output of the auxiliary branch is the road direction map. The direction range is $[0,\pi)$. Due to the continuity of directions, the prediction of road directions is actually a regression problem rather than a classification problem. Therefore, we build the direction loss based on the commonly used $L_1$ loss. The direction loss is defined as

$$\mathcal{L}_{\text{direct}} = \frac{1}{N_r}\sum_{i=1}^{N_r} Diff\left(A_i(T), A_i(P)\right) \tag{4}$$

where $A_i(T)$ and $A_i(P)$ are the road direction GT and the output direction on the $i_{th}$ pixel, respectively. $N_r$ is the number of pixels in roads. To avoid the influence of noises in nonroad areas when learning of directional linear features, pixels in nonroad areas are not considered in (4). $Diff(\cdot,\cdot)$ is the included angle between the two directions. Since the included angle should not be greater than $\pi/2$, it is calculated by

$$\text{Diff}(x,y) = \min\left(|x-y|, \pi - |x-y|\right). \tag{5}$$

It should be noted that this is not exactly the same as the conventional $L_1$ loss. Instead of a simple numerical difference, the included angle between the lines with two directions is regarded as the difference.

## III. Experiments

In this section, a series of related experiments and analyses are shown in four parts. The first part introduces the dataset used in the experiments. In the second part, the specific details of experiments are given. The evaluation metrics are elaborated in the third part. The specific experimental results and analyses are in the fourth and fifth parts. The last part discusses the influence of parameters on the experiments.

### A. Dataset

The dataset used in our experiments is the SAR images taken from Zhanjiang and Shantou in Guangdong Province, China. The resolutions of SAR images in Zhanjiang dataset and Shantou dataset are 3 and 1.5 m, respectively. The polarization mode of these images is VV. In view of the difficulty of extracting roads from urban areas when the resolution is not high enough, we conduct experiments in the nonurban areas of Zhanjiang dataset and the urban areas of Shantou dataset. And then we can explore the performance and sensitivity of our methods in different topographic areas. Several roads in different typical scenes are shown in Fig. 6. As shown in Fig. 6(a) and (b), the rural and suburban roads belong to the nonurban areas in Zhanjiang dataset. As shown in Fig. 6(c), the urban roads are from the Shantou dataset. Zhanjiang dataset contains 144 images with a size of $512 \times 512$ pixels. Among them, 109 images are selected as the training set, and 35 images are used for testing. Shantou dataset contains 125 images with a size of $512 \times 512$ pixels. Among them, 105 images are selected as the training set, and 20 images are used for testing. All GT of roads are manually annotated according to Google Maps. Considering the distinguish ability of roads in SAR images and the feasibility of

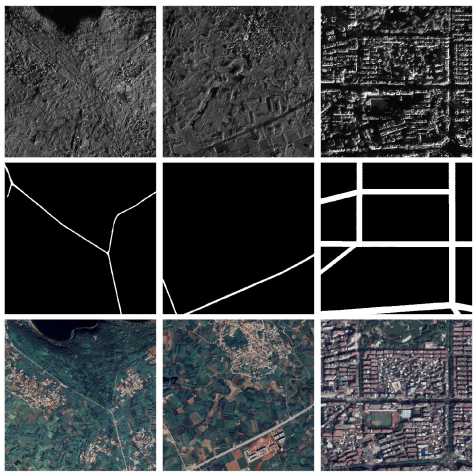


Fig. 6. Example of dataset used in this article. (a) rural road, (b) suburban road, and (c) urban road. The top row shows the SAR images. The middle row shows the corresponding GT of roads. The bottom row shows the optical images from Google Earth.

all experiments, the quite narrow roads like dirt paths are not annotated. We focus on the extraction of major roads in SAR images. In different areas, the widths of roads vary greatly as shown in Fig. 6. In addition, there are complex backgrounds and shadows occluding roads in many areas. Therefore, it is quite challenging to perform road extraction on these datasets.

### B. Implementation Details

In terms of the data processing, for deep learning network, the original small number of SAR images is likely to cause overfitting of the network. Therefore, data augmentation is used to increase the number of training images. The augmentation includes six operations: horizontal flip, vertical flip, 90° rotation, 180° rotation, 270° rotation, and transposition.

In terms of the parameter setting of the network, the batch size is 4. Refer to [31], the Adam optimizer is chosen. The learning rate is initially set 0.0002, and divided by 5 whenever the training loss does not decrease for three consecutive times [31]. Under the above training parameter settings, it takes about 120 epochs for our network to converge. In addition, the loss weights $w_1$, $w_2$, and $w_3$ in (2) are empirically set to 1, 10, and 10, respectively. The segmentation loss is assigned with a lower weight w1 because its values are larger.

In terms of the experimental platform, all experiments in this article are performed on a NVIDIA GeForce GTX 1070 GPU with 8 GB RAM. All networks are implemented using the PyTorch library [46].

### C. Evaluation Metrics

In the field of road extraction from SAR images, the following three metrics are generally used to evaluate the performance of methods: completeness (COM), correctness (COR), and quality $(Q)$ [47]. COM is the portion of matched road areas with respect to the GT of roads. COR is the portion of matched road areas in the extracted roads. $Q$ is a comprehensive index reflecting both COM and COR. Specifically, the three metrics are defined as

$$\mathrm{COM} = \frac{TP}{TP + FN}$$

$$\mathrm{COR} = \frac{TP}{TP + FP}$$

$$Q = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

where TP, FP, and FN denote the true positive, the false positive, and the false negative, respectively. For each metric, the larger the value is, the better performance of the method achieve.

Since the resolution of the dataset is 3 m, the width of roads corresponds to a few pixels in SAR images. Under this harsh condition, considering the noises of road edges, the accuracy of manually labeling all road pixels is limited. Thus, a tolerance $\rho$ is usually set up when calculating road extraction metrics [48], [49]. Specifically, if an area predicted as road are within $\rho$ pixels of a pixel labeled as road, the area is regarded as a matching road area. In our experiments, the tolerance $\rho$ is set to 3 as in [48].

In addition to the evaluation of the geometric quality of roads using the above three metrics, the topological quality evaluation is also important. To evaluate the topological correctness and connectivity of roads, the average path length similarity (APLS) is used in our experiments [50]. This metric captures the similarity in shortest path distances between all pair of nodes in the GT graph $G$ and the estimated graph $G'$. APLS is defined as

$$\mathrm{APLS} = 1 - \frac{1}{N_p} \sum \min \left\{ 1, \frac{|L(a,b) - L(a,' b')|}{L(a,b)} \right\} \quad (6)$$

where $N_p$ is the number of unique paths. $L(a,b)$ is the length of path between the nodes $a$ and $b$. $a', b'$ denote the nodes in the estimated graph $G'$ nearest the location of nodes $a, b$ in $G$, respectively.

### D. Ablation Study

To verify the effectiveness of the connectivity supervision and the direction supervision in the proposed network, related ablation experiments are designed in Zhanjiang dataset. According to the different losses introduced in Section II-D, the network that only contains the segmentation supervision in the main branch is called RoadSeg, which uses only $\mathcal{L}_{\text{seg}}$ as the loss. The entire main branch with a loss of $\mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{conn}}$ is denoted by RoadSeg-C. The network which uses $\mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{direct}}$ as the loss is represented by RoadSeg-D. The network that contains all losses is called RoadSeg-CD, which is the proposed network.

Part of the road extraction results of the ablation study are shown in Fig. 7. The main differences between the different results are marked with yellow circles. From the comparison of Fig. 7(b) and (c), it can be observed that, with the addition of connectivity supervision, the continuity of extracted roads of RoadSeg-C is better. As shown in Fig. 7(b) and (d), RoadSeg-D achieves better performance with the additional directional supervision. Especially for the long straight roads with obvious directionality in suburban areas, the completeness of extracted roads using RoadSeg-D is significantly improved

TABLE I
EVALUATION METRICS OF ROAD EXTRACTION RESULTS IN ABLATION STUDY

| Methods | Supervisions | | | Metrics | | | |
|---|---|---|---|---|---|---|---|
| | Segmentation | Connection | Direction | COM(%) | COR(%) | $Q$(%) | APLS(%) |
| RoadSeg | √ | | | 86.51 | 85.63 | 68.44 | 63.37 |
| RoadSeg-C | √ | √ | | 89.45 | 86.60 | 70.18 | 68.00 |
| RoadSeg-D | √ | | √ | 87.74 | 86.41 | 69.41 | 66.03 |
| RoadSeg-CD | √ | √ | √ | **89.95** | **87.07** | **70.71** | **69.65** |

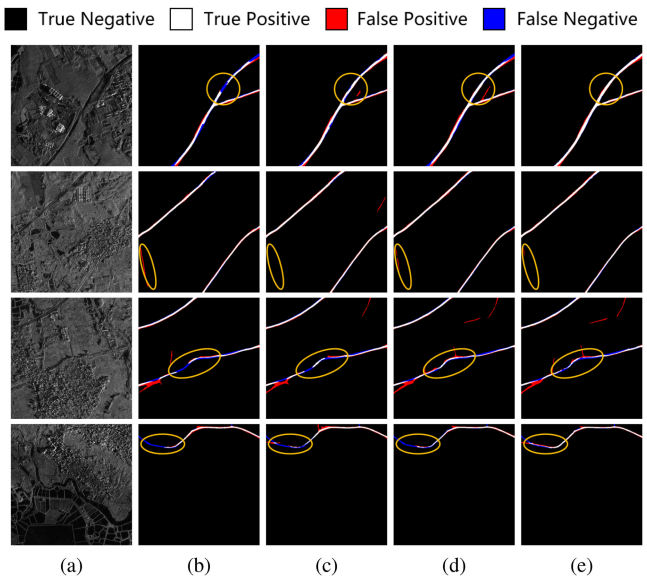


Fig. 7. Comparison of the road extraction results in ablation study. (a) SAR images, (b) RoadSeg, (c) RoadSeg-C, (d) RoadSeg-D, (e) RoadSeg-CD. White denotes the correct extraction parts. Red denotes the false alarms. Blue denotes the missed alarms.

compared with RoadSeg, as shown in the first image of Fig. 7(d). Observing Fig. 7(e), it can be observed that the RoadSeg-CD, which comprehensively considers connectivity supervision and direction supervision achieves the best overall performance for road extraction.

Table I quantitatively reports the performance of the ablation experiments. The $Q$ value of RoadSeg-C is 1.74% higher than that of RoadSeg, while the $Q$ value of RoadSeg-D is 0.97% higher than that of RoadSeg. In addition, the values of COM and COR of RoadSeg-D are greater than that of RoadSeg-C. It can be observed that the extraction performance of RoadSeg-C is better than that of RoadSeg-D, which indicates that the proposed connectivity supervision can bring more significant improvement to the original network than the direction supervision. As for the APLS, RoadSeg-C improves the RoadSeg by 4.63%, which proves that the connectivity supervision improves the topological quality of extracted roads. RoadSeg-CD is the best network in all metrics, and the $Q$ value of it is 2.27% higher than that of RoadSeg. It shows that the ingeniously designed dual-branch connection in RoadSeg-CD enables the auxiliary direction branch to improve the road extraction performance of the main branch.

### *E. Comparative Experiments*

To compare the performance of road extraction, different methods are evaluated on the test images. As an intuitive comparison, Fig. 8 qualitatively shows the road extraction results with various methods. Roads in nonurban areas and urban areas are shown separately. As shown in Fig. 8(b), the classic FCN is the most inferior, and the extracted roads are jagged and rough. This is mainly caused by the following two reasons: 1) In the architecture of FCN, the final extraction results are obtained by up-sampling the feature map with low resolution, which is relatively rough; 2) The spatial context information of images are hardly considered by FCN. As shown in Fig. 8(c) and (d), SegNet and LinkNet achieve better performance than FCN. It is mainly because the symmetrical "U"-shaped structure in these networks can map the extraction results from the low-resolution feature maps, when utilize the detailed information of the high-resolution feature maps as much as possible. In the "U"-shaped networks, the multiscale feature maps obtained by encoders are fused into the up-sampling results of decoders, so that the multiscale spatial context information of roads is utilized. Among the two types of "U"-shaped networks, the road extraction performance of SegNet is as poor as that of FCN. As shown in Fig. 8(d), the road extraction results of LinkNet are obviously better than that of SegNet. In particular, as shown in the extraction results of the second image, the number of road parts extracted correctly are increased significantly. Moreover, the channel reduction adopted in LinkNet makes the decoder have fewer parameters, which is beneficial to the training with small number of SAR images and reduces the possibility of overfitting. However, there are still many fragments in the road extraction results of LinkNet, due to the shadows and the winding of some roads. As shown in Fig. 8(e), for the ResUnet, residual units are introduced in the decoders, so the learning ability of the network is improved. As shown in Fig. 8(f) and (g), DeepLabv3+ and D-LinkNet can obtain more complete roads, which are similar to the results of ResUnet. The atrous spatial pyramid pooling module in DeepLabv3+ and the dilated convolutional block in D-LinkNet are able to aggregate contextual information in SAR images, so as to improve the completeness of extracted roads. Nevertheless, there are still some gaps in the results of the DeepLabv3+ and D-LinkNet. These gaps are especially obvious in urban areas. As shown in Fig. 8(h), compared to other comparison methods, the proposed network RoadSeg-CD shows two major advantages: 1) It can extract more continuous and complete roads; 2) It can obtain fewer false alarms. This is because the connectivity supervision and direction supervision

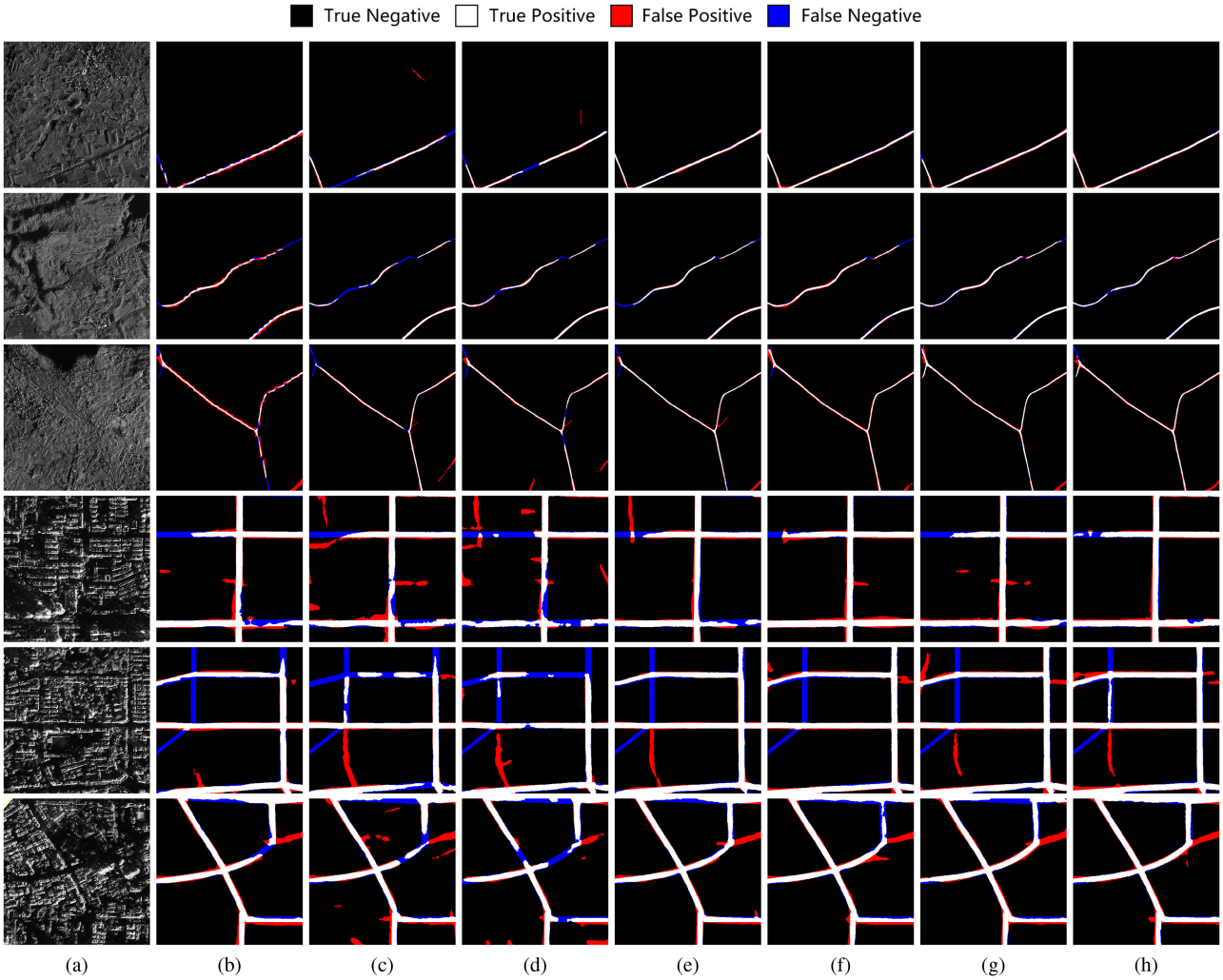


Fig. 8. Comparison of the road extraction results obtained by different methods. (a) SAR images, (b) FCN, (c) SegNet, (d) LinkNet, (e) ResUnet, (f) DeepLabv3+, (g) D-LinkNet, (h) RoadSeg-CD. White denotes the correct extraction parts. Red denotes the false alarms. Blue denotes the missed alarms. The first three rows are from the non-urban areas of Zhanjiang dataset. The last three rows are from the urban areas of Shantou dataset.

TABLE II
EVALUATION METRICS OF ROAD EXTRACTION RESULTS OF DIFFERENT METHODS IN ZHANJIANG DATASET

| Methods | Metrics(%) | | | | Time |
|---|---|---|---|---|---|
| | COM | COR | $Q$ | APLS | (ms) |
| FCN | 75.07 | 69.31 | 51.70 | 43.83 | **181** |
| SegNet | 79.50 | 78.09 | 59.14 | 47.04 | 353 |
| LinkNet | 83.31 | 82.87 | 63.52 | 49.58 | 229 |
| ResUnet | 88.16 | 85.21 | 68.50 | 64.66 | 234 |
| DeepLabv3+ | 89.83 | 85.61 | 69.83 | **69.74** | 228 |
| D-LinkNet | 88.05 | 85.67 | 68.45 | 61.05 | 343 |
| RoadSeg-CD | **89.95** | **87.07** | **70.71** | 69.65 | 622 |

introduced in RoadSeg-CD can enhance the continuity and accuracy of the road extraction results.

Table II quantitatively reports the results obtained by difference methods in Zhanjiang dataset. The performance of FCN and SegNet is relatively poor. For the two networks, the values of COM and COR are both below 80%, and the $Q$ value is below 60%. The poor performance makes it difficult for them to provide effective services for the practical application about road extraction. The results of many missed roads in Fig. 8(b) and (c) also prove the poor performance. The performance of LinkNet is generally better compared with the above methods. However, the APLS is below 50% like the previous two methods, and there are still fracture problems as shown in Fig. 8(d). ResUnet, DeepLabv3+ and D-LinkNet are three state-of-the-art methods, whose values of COM and COR are above 85% and APLS are above 60%. The performance of these networks still have room for improvement. The proposed network RoadSeg-CD is superior to the other six methods in all metrics except APLS. The values of COM and COR have reached more than 87%, indicating that it can not only extract most roads correctly, but also has fewer false alarms. Its $Q$ value reaches 70.71%, which proves that the proposed method is the best among the above methods. Although the APLS of our method is 69.65% which

TABLE III
EVALUATION METRICS OF ROAD EXTRACTION RESULTS OF DIFFERENT METHODS IN SHANTOU DATASET

| Methods | Metrics(%) | | | | Time |
|---|---|---|---|---|---|
| | COM | COR | $Q$ | APLS | (ms) |
| FCN | 81.08 | 81.37 | 63.31 | 31.08 | **180** |
| SegNet | 76.66 | 76.28 | 56.91 | 26.38 | 345 |
| LinkNet | 75.88 | 77.64 | 56.84 | 24.39 | 224 |
| ResUnet | 83.22 | 84.40 | 66.22 | 40.16 | 232 |
| DeepLabv3+ | 84.26 | 83.23 | 66.44 | 39.89 | 222 |
| D-LinkNet | 82.75 | 83.93 | 65.19 | 32.76 | 344 |
| RoadSeg-CD | **86.52** | **86.61** | **69.79** | **42.30** | 636 |

is 0.09% lower than the APLS of DeepLabv3+, the topological quality of roads extracted by our method is admirable.

Table III quantitatively reports the results obtained by difference methods in Shantou dataset. There are urban roads in the dataset. Although the road extraction from urban areas are difficult, the performance in Shantou dataset are not too worse than that in Zhanjiang dataset because of the higher resolution. In addition, for our method, the improvement of urban road extraction is obviously higher than that of nonurban road extraction. For urban road extraction, the values of COM and COR of RoadSeg-CD are above 86%, while these values of DeepLabv3+ and D-LinkNet are about 84%. As for the APLS, RoadSeg-CD improves other methods by at least 2%. It is because the roads in urban areas are straight and the road network topology is easier to be refined to further improve the quality. It is worth noting that the values of APLS in this dataset are generally low, indicating that it is difficult to maintain good topological performance in urban road extraction.

To discuss the computational cost of different methods, the average road extraction time per image on the test set is also given in Tables II and III. It can be observed that RoadSeg-CD takes the longest time. This is because that RoadSeg-CD consists of two branches, which take the connectivity and direction features of roads into consideration and slow down the test speed. However, it is acceptable because the high computational loss brings a substantial improvement in road extraction performance.

### *F. Parameter Setting*

In addition to the training parameters of networks shown in Section III-B, the experiments in this article also involve several other important parameters. It is very necessary to explain the relevant specific parameter settings. $\alpha$ is the key parameter in the connectivity loss, as shown in (2). The connectivity loss designed in this article is able to evaluate the road connectivity from a multiscale perspective. $\alpha$ determines the influence of the connectivity loss under the large receptive field on the entire connectivity loss. Furthermore, it determines the influence of the large fractures in extraction results on the entire connectivity loss. Fig. 9 shows the $Q$ values of road extraction results under different values of $\alpha$. It can be observed that the performance is poor when $\alpha$ is too large or too small. When $\alpha$ is too small, the connectivity loss hardly considers the connectivity between the road segments which are far apart from each other under the large receptive field, so that the large gaps easily appear in the road extraction results, as shown in Fig 10(a). When $\alpha$ is too large, the connectivity between the nearby road segments has too little influence on the overall connectivity loss, so that the extracted roads are too fragmented, as shown in Fig. 10(c). Therefore, $\alpha$ must be assigned a balanced value. In our experiments, the specific $\alpha$ value is taken as 0.5. Fig. 10(b) shows the relatively higher extraction performance under this value.

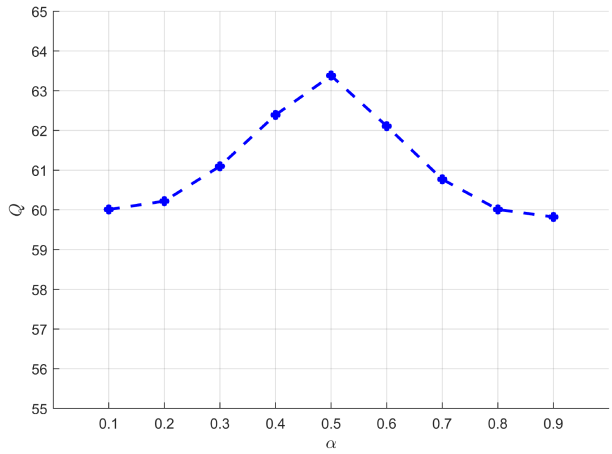


Fig. 9. $Q$ values of road extraction results under different values of $\alpha$. The performance is best when $\alpha$ is 0.5.


Fig. 10. Road extraction results under different values of $\alpha$. (a) $\alpha$=0.1, (b) $\alpha$=0.5, (c) $\alpha$=0.9. The continuity of extracted roads is best when $\alpha$ is 0.5.

The parameter $m$ is the total number of scales in the overall connectivity loss in (2). The value of $m$ is determined according to the ratio of the image size to the road width. Since the size of a single image in our dataset is $512 \times 512$ pixels and the average width of roads is 8 pixels, the value of $m$ is taken as $\log_2 \frac{512}{8} = 6$. In addition, all comparison methods use the default parameters settings recommended by the origin papers.

## IV. CONCLUSION

Aiming at the road extraction from SAR images, we propose a novel DCNN-based algorithm with connectivity array and direction map. For one thing, in order to utilize the road connectivity information, a connectivity array is proposed to construct a novel multiscale connectivity loss. For another, to exploit the linear features of roads, a new kind of the road direction map is designed to introduce the unique direction supervision of linear targets. In addition, the dual-branch network is built by

well-designed feature fusion process, so that the connectivity supervision and the direction supervision can simultaneously bring positive effects on road extraction. The experimental results on the measured SAR images show that the proposed network can not only extract more complete roads, but also obtain lower false alarms compared with other comparison algorithms. Both the values of completeness and correctness of our method reach more than 86% in different topographic areas.

## ACKNOWLEDGMENT

The SAR dataset used in the experiments are the courtesy of Beijing Institute of Radio Measurement and Aerospace Information Research Institute in Chinese Academy of Sciences, the authors would like to thank them for their support.

## REFERENCES

[1] Q. Zhang *et al.*, "A new road extraction method using sentinel-1 SAR images based on the deep fully convolutional neural network," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 572–582, 2019.

[2] F. Xiao, L. Tong, and S. Luo, "A method for road network extraction from high-resolution SAR imagery using direction grouping and curve fitting," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2733.

[3] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A new algorithm for SAR image target recognition based on an improved deep convolutional neural network," *Cogn. Comput.*, vol. 11, no. 6, pp. 809–824, 2019.

[4] R. Fjørtoft, A. Lopes, P. Marthon, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Trans. Geosci. Remote. Sens.*, vol. 36, no. 3, pp. 793–802, May 1998.

[5] Y. Chen, Q. Yang, Y. Gu, and J. Yang, "Detection of roads in SAR images using particle filter," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 2337–2340.

[6] F. Xiao, Y. Chen, L. Tong, L. He, L. Tan, and B. Wu, "Road detection in high-resolution SAR images using duda and path operators," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1266–1269.

[7] C. Jia, K. Ji, Y. Jiang, and G. Kuang, "Road extraction from high-resolution SAR imagery using hough transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2005, pp. 336–339.

[8] W. Liu, Z. Zhang, S. Li, and D. Tao, "Road detection by using a generalized hough transform," *Remote Sens.*, vol. 9, no. 6, 2017, Art. no. 590.

[9] C. R. D. Silva, J. A. Silva Centeno, and M. J. Henriques, "Automatic road extraction in rural areas, based on the radon transform using digital images," *Can. J. Remote Sens.*, vol. 36, no. 6, pp. 737–749, 2010.

[10] X. Li, S. Zhang, X. Pan, P. Dale, and R. Cropp, "Straight road edge detection from high-resolution remote sensing images based on the ridgelet transform with the revised parallel-beam radon transform," *Int. J. Remote Sens.*, vol. 31, no. 19, pp. 5041–5059, 2010.

[11] T. Perciano, F. Tupin, R. Hirata, and R. M. C. Junior, "A hierarchical Markov random field for road network extraction and its application with optical and SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 1159–1162.

[12] F. Tupin, H. Maître, 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.

[13] Y. Guo, Z. Bai, Y. Li, and Y. Liu, "Genetic algorithm and region growing based road detection in SAR images," in *Proc. IEEE 3rd Int. Conf. Natural Comput.*, 2007, pp. 330–334.

[14] 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.

[15] 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, Sep. 2020.

[16] F. Dell'Acqua and P. Gamba, "Detection of urban structures in SAR images by robust fuzzy clustering algorithms: The example of street tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2287–2297, Oct. 2001.

[17] I. Grinias, C. Panagiotakis, and G. Tziritas, "MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 145–166, 2016.

[18] W. Zhang, D. Xiang, and Y. Su, "Fast multiscale superpixel segmentation for SAR imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2020.

[19] J. Li, Q. Hu, and M. Ai, "Unsupervised road extraction via a gaussian mixture model with object-based features," *Int. J. Remote Sens.*, vol. 39, no. 8, pp. 2421–2440, 2018.

[20] Z. Yue *et al.*, "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cogn. Comput.*, vol. 13, no. 4, pp. 795–806, 2021.

[21] F. Ma, F. Zhang, Q. Yin, D. Xiang, and Y. Zhou, "Fast SAR image segmentation with deep task-specific superpixel sampling and soft graph convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Sep. 2021.

[22] W. He, H. Song, Y. Yao, and X. Jia, "A multiscale method for road network extraction from high-resolution SAR images based on directional decomposition and regional quality evaluation," *Remote Sens.*, vol. 13, no. 8, 2021, Art. no. 1476.

[23] Y. He, F. Gao, J. Wang, A. Hussain, E. Yang, and H. Zhou, "Learning polar encodings for arbitrary-oriented ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3846–3859, Mar. 2021.

[24] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

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

[26] 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. 2017.

[27] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.

[28] J. Li, Y. Meng, D. Dorjee, X. Wei, Z. Zhang, and W. Zhang, "Automatic road extraction from remote sensing imagery using ensemble learning and postprocessing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10535–10547, Jul. 2021.

[29] 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.

[30] 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.

[31] 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 Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 182–186.

[32] A. Batra, S. Singh, G. Pang, S. Basu, C. V. Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10385–10393.

[33] 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. 2020.

[34] M. Zhou, H. Sui, S. Chen, J. Wang, and X. Chen, "BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 288–306, 2020.

[35] A. Abdollahi, B. Pradhan, and A. Alamri, "RoadVecNet: A new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up," *GISci. Remote Sens.*, vol. 58, no. 7, pp. 1151–1174, 2021.

[36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[37] C. Tao, J. Qi, Y. Li, H. Wang, and H. Li, "Spatial information inference net: Road extraction using road-specific contextual information," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 155–166, 2019.

[38] 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, 2020, Art. no. 1444.

[39] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[41] M. C. Kampffmeyer, N. Dong, X. Liang, Y. Zhang, and E. Xing, "ConnNet: A long-range relation-aware pixel-connectivity network for salient segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2518–2529, May 2019.
[42] N. Liu, Z. Cui, Z. Cao, Y. Pi, and S. Dang, "Airport detection in large-scale SAR images via line segment grouping and saliency analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 434–438, Mar. 2018.
[43] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica Int. J. Geographic Inf. Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.
[44] U. Ramer, "An iterative procedure for the polygonal approximation of plane curves," *Comput. Graph. Image Process.*, vol. 1, no. 3, pp. 244–256, 1972.
[45] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, 2005.
[46] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 8026–8037, 2019.
[47] C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet, "Empirical evaluation of automatically extracted road axes," *Empirical Evaluation Techn. Comput. Vis.*, vol. 12, pp. 172–187, 1998.
[48] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 210–223.
[49] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks," *Electron. Imag.*, vol. 2016, no. 10, pp. 1–9, 2016.
[50] A. Van Etten, J. Shermeyer, D. Hogan, N. Weir, and R. Lewis, "Road network and travel time extraction from multiple look angles with spacenet data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 3920–3923.

**Fei Gao** received the B.S. degree in industrial electrical automation, and the M.S. degree in electromagnetic measurement technology and instrument from Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in signal and information processing from the Beihang University, Beijing, China, in 2005.

He is currently a Professor with the School of Electronics and Information Engineering, Beihang University, Beijing, China. His research interests include target detection and recognition, image processing, deep learning for applications in remote sensing.

**Jun Tu** received the B.S. degree in electronic and information engineering from the Beihang University, Beijing, China, in 2019. He is currently working toward the M.E. degree in information and communication engineering with Beihang University.

His current research activities are in object detection and synthetic aperture radar image processing.

**Jun Wang** received the B.S. degree in communication engineering from the North Western Polytechnical University, Xi'an, China, in 1995, and the M.S. and Ph.D. degrees in signal and information processing from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA, Beijing, China. He is interested in signal processing, DSP/FPGA real-time architecture, target recognition and tracking, etc. His research has resulted in more than 40 papers in journals, books, and conference proceedings.

**Amir Hussain** received the B.Eng. and Ph.D. degrees in electronic and electrical engineering from the University of Strathclyde, Scotland, U.K., in 1992 and 1997, respectively.

Following Postdoctoral and Senior Academic Positions at West of Scotland (1996-1998), Dundee (1998-2000) and Stirling Universities (2000-2018), respectively, he joined Edinburgh Napier University as founding Head of the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab and the Centre for AI and Data Science. His research interests include cognitive computation,machine learning and computer vision.

**Huiyu Zhou** received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, the M.S. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the Doctor of Philosophy degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently a Professor with the School of Informatics, University of Leicester, Leicester, U.K. He has published widely in medical image processing, computer vision, intelligent systems, and data mining.