Road Extraction in SAR Images Using Ordinal Regression and Road-Topology Loss

Xiaochen Wei 1,2,3, Xiaolei Lv 1,2,3,* and Kaiyu Zhang 1,2,3

Abstract: The road extraction task is mainly composed of two subtasks, namely, road detection and road centerline extraction. As the road detection task and road centerline extraction task are strongly correlated, in this paper, we introduce a multitask learning framework to detect roads and extract road centerlines simultaneously. For the road centerline extraction problem, existing works rely either on regression-based methods, or classification-based methods. The regression-based methods suffer from slow convergence and unsatisfactory local solutions. The classification-based methods ignore the fact that the closer the pixel is to the centerline, the higher our tolerance for its misclassification. To overcome these problems, we first convert the road centerline extraction problem into the problem of discrete normalized distance label prediction, which can be resolved by training an ordinal regressor. For the road extraction task, most of the previous studies apply pixel-wise loss function, for example, Cross-Entropy loss, which is not sufficient, as the road has special topology characteristics such as connectivity. Therefore, we propose a road-topology loss function to improve the connectivity and completeness of the extracted road. The road-topology loss function has two key characteristics: (i) The road-topology loss function combines road detection prediction and road centerline extraction prediction to promote the two subtasks to each other by using the correlation between the two subtasks; (ii) The road-topology loss can emphatically penalize gaps that often appear in road detection results and spurious segments that easily appear in centerline extraction results. In this paper, we select the AdamW optimizer to minimize the road-topology loss. Since there is no public dataset, we build a road extraction dataset to evaluate our method. State-of-the-art semantic segmentation networks (LinkNet34, DLinkNet34, DeeplabV3plus) are used as baseline methods to compare with two kinds of method. The first kind of method modifies the baseline method by adding the road centerline extraction task branch based on ordinal regression. The second kind of method uses the road topology loss and has the same network architecture as the first kind of method. For the road detection task, the two kinds of methods improve the baseline methods by up to 3.51% and 11.98% in $\text{IoU}$ metric on our test dataset, respectively. For the road centerline extraction task, the two kinds of methods improve the baseline methods by up to 8.22% and 10.9% in the Quality metric on our test dataset.

Keywords: road extraction; ordinal regression; road-topology loss; multitask learning; deep learning; SAR images

1. Introduction

Synthetic Aperture Radar (SAR) can penetrate clouds and collect data under any weather at any time. Therefore, it has been widely used in various fields. With the rapid development of SAR technology in decades, high-resolution synthetic aperture radar images have become an essential source of information due to their broader coverage and increased

Remote Sens. 2021, 13, 2080. https://doi.org/10.3390/rs13112080
ground details. Roads are an indispensable part of the modern transportation network, and they have a pivotal position both geographically and economically. Automatic road extraction from SAR images has become a research hotspot. It has a wide range of urban planning applications, disaster prevention and mitigation, and geographic information system updates.

In recent decades, many methods of road extraction from SAR images have been proposed [1–7]. The road extraction task can be divided into two subtasks: road detection and road centerline extraction. The road extraction methods in previous works separated these two tasks, which can be roughly summarized into three steps. Firstly, the road in the SAR image is detected according to the extracted image features. Then, the road centerline is obtained by skeletalizing the road detection result. Finally, the road centerline is converted into a graph, which then is topologically optimized to obtain the road centerline network. The methods that separate the two tasks ignore the correlation between the two tasks, and the errors of road detection affect the result of road centerline extraction. The work [8] achieved a better performance using a conditional random field model to simultaneously perform the three steps: road detection, skeletonization, and topology optimization.

Since deep convolutional neural network (DCNN) has shown an outstanding performance in the 2012 ImageNet image classification challenge, DCNN [9] has made great progress in various traditional computer vision tasks such as image classification [10–14] and image semantic segmentation [15–19]. Deep learning has been widely used in the fields of remote sensing [20–24]. However, the works that applied deep learning to SAR images [25,26] are relatively few. One of the reasons for this is that the unique characteristics of SAR images make labeling time-consuming and labor-intensive. Reference [27] proposed a method to automatically label buildings from SAR images, which significantly reduces labeling difficulty. At the same time, reference [27] also applied the architecture combined with FCN used to extract features with CRF-RNN that exploited spatial information to extract buildings from SAR images. Although there have been fewer minimal works that apply deep learning for SAR imagery to date, it is undeniable that deep learning has colossal application potential for SAR image processing. A DisDBN which combined ensemble learning with a deep belief network to learn discriminant features is proposed in [28] for SAR image classification. Reference [29] let extracted six-channel covariance matrix fed into DCNN for PolSAR image classification. Reference [30] proposed a deep supervised and contractive neural network for SAR image classification. Reference [31] extended the semantic segmentation network from the real-value domain to the complex-valued domain, which can exploit the unique information of SAR data. Reference [32] proposed a new fully convolutional neural network that can be trained in an end-to-end for semantic segmentation of polarimetric SAR imagery in the complex land cover ecosystem. Reference [33] realized smooth classification with a small training set with four classes by deep transfer learning. Reference [34] extracted multiscale trained features using a multiscale CNN model to detect built-up areas from SAR images. Reference [35] firstly applied a deep fully convolutional neural network to segment road from SAR images.

There are two challenges that restrict the accuracy of road extraction from SAR images: (1) the road in SAR images is usually modeled as dark elongated areas surrounded by bright edges which are easily confused with other objects such as rivers, shades of trees, shadows of buildings, etc. (2) Speckle noise in SAR images seriously degrades image quality and interferes with road extraction from SAR images. A deep convolutional neural network with a large receptive field can effectively extract contextual information. This characteristic can be used to distinguish the road from other similar objects. A deep convolutional neural network reduces the influence of speckle noise by continuously performing convolution operations on the input images. Therefore, in this paper, we will introduce a deep learning framework to extract roads from SAR imagery.

The road extraction task includes two subtasks: road detection task and road centerline extraction task. Most of the previous works relied on multistage-based learning methods to extract roads. These methods obtain road centerlines by a post-processing step of thin road
Remote Sens. 2021, 13, 2080 3 of 25
detection predicted by CNN. The disadvantage of these methods is that imperfect road detection results lead to road centerline extraction results with low connectivity. There is a symbiotic relationship between the road detection task and road centerline extraction task. The road detection task and road centerline extraction task can promote each other. The road detection task can provide detection cues for the road centerline extraction task to constraint road centerlines, which can avoid spurious parts. The road centerline extraction task can motivate the road detection task to pay more attention to the key points of roads, which can enhance the road connectivity. In order to make full use of this relationship, our proposed network learns the road detection task and road centerline extraction task simultaneously under the multitask learning scheme.

For road centerline extraction from raw image data, the previous works can be roughly divided into two categories: one is based on classification methods, and the other is based on regression methods. Reference [7,36,37] extracted road centerline by a classification-based method. The features of pixels on the road centerline are similar to features of pixels adjacent to the centerline, and the features of pixels far from the centerline of the road are completely different from the pixels in the centerline of the road. However, the errors caused by the misclassification of pixels adjacent to the centerline of the road are the same as those caused by the misclassification of pixels far from the centerline of the road, which is unreasonable. In practice, the closer the pixel is to the centerline of the road, the more we can tolerate its classification errors. This contradiction makes it difficult for the classification-based to cause the network converge to a better result. Reference [38] first proposed method-based regression, which learned a designed function whose returns value decreases with the distance from the pixel to the centerline. However, due to abnormal, i.e., annotation errors, a deep network for regression is relatively unstable, and the network trained by MSE loss will not converge to a satisfactory global solution. Reference [39] learned the map of distance from each pixel to the nearest boundary by training a multi-class classification network, which ignores the ranking relation between different distance classes. To avoid the above problems, we exploit the method based on ordinal regression to learn the discrete normalized distance labels. We use the ordinal loss, minimized to learn network parameters of the road centerline task.

In the real world, the roads have unique topology properties. Previous works usually applied topology prior to using variational and Markov random field-based methods [36,40–42]. Reference [40] imposed a topology constraint by high-order CRF, in which high-order cliques connect superpixels of the road network. Reference [36] represented the road network as a sequence of graph structures and found an optimal subgraph by integer programming. These previous works generally employed road topology optimization as post-processing, which cannot remove large spurious parts and connect large gaps. Recently, some approaches guaranteed a perfect topology of the extracted road by minimizing a topology-preserving loss function [43,44]. Reference [43] stated that only using pixel-wise loss functions is not enough for curvilinear detection and proposes the topology-aware loss function defined by the selection of the filter responses of pretrained VGG19 to penalty topology errors. Reference [44] adopted a loss function based on Persistent Homology, which is continuous. Neither VGG19 nor Betti are specifically designed for road extraction tasks, so their penalty for topology errors of extracted road network is limited. To solve the above-mentioned problems, a new road-topology loss is specially designed for the road extraction task, which can reduce the topology errors. Our main contributions are as follows:

1. Different from the previous methods for road extraction from SAR imagery, we detect the road and extract road centerline simultaneously. This multitask learning scheme can exploit the correlation between road detection task and road centerline extraction task;
2. For the road extraction task, we build our dataset with TerraSAR-X images, which cover urban, suburbs and rural areas. Our experiments are carried out based on this
Our experimental results show that our proposed framework can achieve a better road extraction performance;

3. For the road centerline extraction task, we first convert the road centerline extraction problem into the problem of discrete normalized distance label prediction, which can be solved by training an ordinal regressor;

4. Consider the special topology feature of road network, we propose a new road-topology loss which is designed to reduce the topology error of road extraction including spurious parts and gaps.

The remainder of our paper is organized as follows. We present the proposed method in detail in Section 2. In Section 3, we quantitatively and qualitatively analyzed the superior performance of our method compared with baseline methods. We discuss the stability of different methods with various binarization thresholds in Section 4. Finally, we conclude the whole paper in Section 5.

2. Materials and Methods

Figure 1 illustrates our road extraction framework. For road extraction, our proposed network learns the road detection task and the road centerline extraction task jointly under a multitask learning scheme. As shown in Figure 1, our framework has two branches: the road detection branch and the road centerline extraction branch. The encoder of the two branches is shared for feature extraction, which establishes a connection between the two branches. In this section, we first separately introduce how our network performs road detection tasks and road centerline extraction tasks. Next, the definition of our first proposed road-topology loss function is given. Finally, we introduced how our multitask learning framework simultaneously learns the road detection task and road centerline extraction task based on ordinal regression, using the road-topology loss that we first proposed.

In the following discussion, we let \( I \in \mathbb{R}^{H \times W} \) be the input image, let \( Y \in \{0, 1\}^{H \times W} \) be the corresponding ground truth, with 1 indicating pixels on the road and 0 indicating background pixels, and let \( \hat{Y} \in \{0, 1\}^{H \times W} \) be the predicted probability map of the road. We let the mini-batch be \( B \), let \( i \) be a pixel in \( I \), \( y_i \) is label of pixel \( i \). The predicted probability that pixel \( i \) is on the road is denoted by \( \hat{y}_i \).

### 2.1. Road Detection

The road detection task aims to detect the roads from SAR imagery. The output of the road detection task is the binary image, in which the pixels seen on the roads are 1 and the others are 0. In practice, most pixels of the SAR imagery belong to the non-road regions. As a result, there is a label-unbalancing problem in the road detection task. To overcome this problem, we use weighted cross-entropy loss proposed in [45]. The weighted cross-entropy loss of \( I \) is

\[
L_{rs} = - \frac{1}{H \times W} \sum_{i \in I} w_1 y_i \log \hat{y}_i + w_0 (1 - y_i) \log (1 - \hat{y}_i).
\]  (1)

Next, we will present the weights in the weighted cross-entropy loss. Let \( |B| \) be the number of the images. The loss weight for road pixels is \( w_1 = \frac{1}{|B| \times H \times W} \sum_{i \in B} \sum_{i \in I} 1(y_i = 0) \) and the loss weight for non-road pixels is \( w_0 = \frac{1}{|B| \times H \times W} \sum_{i \in B} \sum_{i \in I} 1(y_i = 1) \).
2.2. Road Centerline Extraction

For road centerline extraction, the classification-based approach learn a function $y(\cdot)$ such that

$$y(f_i) = \begin{cases} 1, & i \text{ is on centerline}, \\ 0, & i \text{ is not on centerline}, \end{cases}$$

where $f_i$ is the feature of pixel $i$. The method based on regression is to learn a regressor $y(\cdot)$ whose values decrease monotonically as the distance of $i$ to the centerline increase. Especially in [38], the regressor $y(\cdot)$ is such that

$$y(f_i) = \begin{cases} 1 - \frac{D_C(i)}{d_M}, & \text{if } D_C(i) \leq d_M, \\ 0, & \text{otherwise}, \end{cases}$$

where $D_C(i)$ is the metric distance from pixel $i$ to the closest the pixel on the centerline and $d_M$ is $s/2$, where $s$ is the size of local neighbourhoods that are used to compute feature vector $f_i$. Our proposed method is based on ordinal regression, which is different from both. In the remainder of this subsection, we first model the road centerline extraction problem as the discrete normalized distance label prediction problem. We then describe how to predict discrete normalized distance labels by learning an ordinal regressor.
Roads are surrounded by bright edges in high-resolution SAR images. As a result, we can predict the distance \(d_i\) from any pixel \(i\) to the nearest road edge. However, the probability that the pixel \(i\) is on the centerline of the road is not proportional to the distance from this pixel \(i\) to the nearest road edge \(d_i\). This is due to the width of the road is various, which is depicted in Figure 2. As a result, we predict the normalized distance from pixel \(i\) to the nearest road edge. The normalized distance \(dn_i\) is defined as follows

\[
dn_i = \begin{cases} 
2 \times \frac{d_i}{w_i}, & \text{if } i \text{ is on road,} \\
0, & \text{if } i \text{ is not on road,} 
\end{cases}
\]  

(4)

where \(w_i\) is the road width of pixel \(i\). In particular, \(dn_i\) is proportional to the probability of \(i\) on the centerline of the road. Meanwhile, if \(i\) is on the road centerline, \(dn_i\) is the local maxima value along the direction that is perpendicular to the direction of the road. We further quantize each \(dn_i\) using the thresholds \(\{t_0, t_1, \ldots, t_{K-1}\}\) into one of the \(K + 1\) intervals. The reason we quantize the normalized distance \(dn_i\) is that the direct training of the deep network for regression is relatively unstable, because outliers (annotation errors) cause large error terms, making it difficult for the network to converge and lead to unstable predictions [39]. After quantization, each \(i\) is given a discrete normalized distance label \(l_i\), such as

\[
l_i = \begin{cases} 
0, & d_{n_i} \leq t_0, \\
k + 1, & t_k \leq d_{n_i} < t_{k+1}, k = 0, 1, 2, \ldots, K - 2, \\
K, & t_{K-1} \leq d_{n_i}.
\end{cases}
\]  

(5)

We can predict discrete normalized distance label prediction by a typical method based on multi-class classification. However, the ordinal information between discrete normalized distance labels will be ignored. [46] first combines ordinal regression with DCNNs to address the age estimation, which transforms an ordinal regression problem into a series of binary classification sub-problems to achieve sequential age regression tasks, thus taking accounting for the fact that the set of ages is well-ordered. Therefore, we use ordinal regression in [46] to solve the discrete normalized distance label prediction problem and modify the ordered loss to adapt to the road centerline extraction task.

![Figure 2](image_url)

**Figure 2.** (a) SAR image \(I\). (b) ground truth \(Y\). (c) distance map \(D\). (d) normalized distance map \(D_N\). (e) discrete normalized distance label \(L\). (f) Centerline \(C\).

Next, we will introduce the ordinal regression and the ordinal loss used in this paper in detail. Let \(\varphi\) denote the feature extractor of network that is used to extract the road centerlines and the parameters of \(\varphi\) are denoted by \(\Phi\). \(I \in R^{H \times W}\) is input image. One of the pixels in \(I\) is denoted as \(i\). The feature map of \(I\) and the feature vector of pixel \(i\) for the road centerline extraction task are \(F = \varphi(I, \Phi)\) and \(f_i \in F\), respectively. \(\psi\) is the last layer of the network for the road centerline extraction task, which is used for ordinal regression. Its parameters are given by \(\Theta = \{\theta_0, \theta_1, \ldots, \theta_{2K-1}\}\) where \(\theta_j (j \in \{0, 1, \ldots, 2K - 1\})\) is the weight vector. Ordinal output of \(I\) and ordinal output vector of \(i\) are presented by \(O = \psi(F, \Theta)\) and \(o_i \in O (o_i = (o_i^0, o_i^1, \ldots, o_i^{2K-1}))\), respectively. With the softmax activation function
that is also used in [46], the probability \( \hat{p}_i^k = P(I_i > k|F, \Theta)(k \in \{0, 1, ..., K - 1\}) \) that the predicted label of \( i \) is greater than \( k \) is calculated as

\[
\hat{p}_i^k = \frac{e^{\omega_i^{2k+1}}}{e^{\omega_i^{2k}} + e^{\omega_i^{2k+1}}},
\]

(6)

\[
\omega_i^{2k} = \theta_i^{T} f_i,
\]

(7)

\[
\omega_i^{2k+1} = \theta_i^{T} f_i + 1.
\]

(8)

According to the method of calculating the ordinal loss in [46], the pixel-level ordinal loss of pixel \( i \) is then given by

\[
\Psi(f_i, \Theta) = -\left( \sum_{k=0}^{l_i-1} \omega_i^k \log(\hat{p}_i^k) + \sum_{k=l_i}^{K-1} \omega_i^k (\log(1 - \hat{p}_i^k)) \right),
\]

(9)

where \( \omega_i^k = \frac{1}{|B| \times H \times W} \sum_{l \in B} \sum_{l_i \in I} 1(l_i \leq k) \) and \( \omega_i^k = \frac{1}{|B| \times H \times W} \sum_{l \in B} \sum_{l_i \in I} 1(l_i > k) \) are used to solve the unbalanced classes problem.

The ordinal loss of \( I \) is defined as the sum of the ordinal loss of each \( i \) in image \( I \) and given by

\[
L_c(F, \Theta) = \frac{1}{H \times W} \sum_{i \in I} \Psi(f_i, \Theta),
\]

(10)

The advantage of the ordinal loss is that the greater the difference between the predicted label \( \hat{l}_i \) and the true label \( l_i \), the greater the ordinal loss. We use iterative optimization algorithm to minimize \( L_c(F, \Theta) \). The partial derivative of \( L_c(F, \Theta) \) with respect to \( \theta_j(j \in \{0, 1, ..., 2K - 1\}) \), is

\[
\frac{\partial L_c(F, \Theta)}{\partial \theta_j} = \frac{1}{H \times W} \sum_{i \in I} \frac{\partial \Psi(f_i, \Theta)}{\partial \theta_j},
\]

(11)

\[
\frac{\partial \Psi(f_i, \Theta)}{\partial \theta_{2k+1}} = -\frac{\partial \Psi(f_i, \Theta)}{\partial \theta_{2k}},
\]

(12)

\[
\frac{\partial \Psi(f_i, \Theta)}{\partial \theta_{2k}} = -\left( \omega_i^k f_i \eta(l_i > k)(\hat{p}_i^k - 1) \right) + \omega_i^k f_i \eta(l_i \leq k) \hat{p}_i^k,
\]

(13)

where \( k \in \{0, 1, ..., K - 1\} \) and \( \eta(\cdot) \) is the indicator function, where \( \eta(true) = 1 \) and \( \eta(false) = 0 \). We can update the parameters of the network for road centerline extraction task through backward propagation.

In the test phase, we calculate \( \hat{p}_i \), which is the mean of \( \hat{p}_i^k \) for each pixel \( i \). \( \hat{p}_i \) is given by

\[
\hat{p}_i = \frac{1}{K} \sum_{k=0}^{K-1} \hat{p}_i^k.
\]

(14)

We observe that \( \hat{p}_i \) is proportional to the normalized distance \( d_n_i \). We can regard \( \hat{p}_i \) as the predicted probability that the pixel \( i \) is on the road centerline. Let \( \hat{P} \) be the centerline probability map of \( I \), where \( \hat{P} \in \hat{P} \). According to Formula (4), we know that, from the normalized distance map \( D_N \), the necessary and sufficient condition for pixel \( i \) on the road centerline is that \( d_n_i \) is a local maximum along the direction perpendicular to the direction of the road. However, if only by judging whether \( p_i \) is a local maximum along the direction perpendicular to the direction of the road to infer whether pixel \( i \) is on the road centerline, some non-road regions will be extracted as the road centerline. As a result, we first set the value of \( \hat{P} \) that is less than \( T \) to zero, then we apply the canny-like non-maximum suppression algorithm to \( \hat{P} \) to obtain the road centerline.
2.3. Road-Topology Loss

In practice, the cross-entropy loss is widely used in various segmentation tasks, such as semantic segmentation and instance segmentation. Cross-entropy loss is a pixel-wise loss, which is completely local and does not take the special and complex topological characteristics of the road into account. Such a loss penalizes the mistake of each pixel equally and independently, regardless of the effect of the error on geometry. However, in practice, we find that the pixels closer to the centerline of the road are more important. Because the misclassification of these pixels will cause serious topology errors, such as gaps and spurious parts. To penalize the gaps in the prediction of road detection and the spurious parts in the road centerline extraction prediction, we propose a new road-topology loss \( L_T \).

Next, we will give the details defining the road-topology loss. To measure the connectivity of road detection prediction, we define the connectivity metric as

\[
T_{\text{con}} = \frac{|\hat{Y} \circ L|}{|L|},
\]

where \( \hat{Y} \in [0, 1]^{H \times W} \) is a prediction map of road detection, \( L \in \{0, 1, ..., K\}^{H \times W} \) (\( l_i \in L \)) is discrete normalized distance label of \( I \) and \(|| \) is an operation to calculate the sum of matrix. Similarly, we define the differentiable correctness metric to measure the correctness of road centerline extraction prediction as

\[
T_{\text{cor}} = \frac{|Y \circ \hat{P}|}{|\hat{P}|},
\]

where \( Y \in \{0, 1\}^{H \times W} \) is ground truth of road detection, \( \hat{P} \in [0, 1]^{H \times W} (\hat{p}_i \in \hat{P}) \) is predicted road centerline probability map of \( I \). We observe that the measure \( T_{\text{con}} \) is susceptible to gaps in the road detection prediction while the measure \( T_{\text{cor}} \) is susceptible to spurious parts in the road centerline extraction prediction. Finally, we define road-topology metric \( T_{\text{Road}} \) as the harmonic average between connectivity metric \( T_{\text{con}} \) and differentiable correctness metric \( T_{\text{cor}} \) as

\[
T_{\text{Road}}(\hat{Y}, \hat{P}, Y, L) = \frac{2 \cdot T_{\text{con}} \cdot T_{\text{cor}}}{T_{\text{con}} + T_{\text{cor}}}. \tag{17}
\]

The road-topology metric measures the connectivity and correctness of the road extraction result at the same time.

In order to maximize the road-topology metric in CNNs in an end-to-end manner, we define our road-topology loss \( L_T \) as

\[
L_T = -\log (T_{\text{Road}}) \tag{18}
\]

\( L_T \) is calculated directly from the raw prediction \( \hat{Y} \) and \( \hat{P} \) without thresholding. As a result, \( L_T \) is differentiable over the prediction \( \hat{Y} \) and \( \hat{P} \) which can be integrated into CNN. In this paper, we use AdamW optimizer to minimize the road-topology loss. The partial derivatives of loss \( L_T \) over network activation \( \hat{Y} \) and \( \hat{P} \) at the location of pixel \( i \) are

\[
\frac{\partial L_T}{\partial \hat{y}_i} = -\frac{1}{T_{\text{Road}}} \frac{\partial T_{\text{Road}}}{\partial \hat{y}_i} = \frac{2}{T_{\text{Road}} \left( \frac{T_{\text{con}}}{T_{\text{con}} + T_{\text{cor}}} - \frac{T_{\text{con}} \cdot T_{\text{cor}}}{(T_{\text{con}} + T_{\text{cor}})^2} \right)} \frac{\partial T_{\text{con}}}{\partial \hat{y}_i} = -\frac{1}{T_{\text{con}} \left( \frac{T_{\text{con}}}{T_{\text{con}} + T_{\text{cor}}} \right)} \frac{\partial T_{\text{con}}}{\partial \hat{y}_i} = -\frac{1}{T_{\text{con}} \left( \frac{T_{\text{con}}}{T_{\text{con}} + T_{\text{cor}}} \right)} \frac{l_i}{|L|}. \tag{19}
\]
\[ \frac{\partial L_T}{\partial \hat{p}_i} = - \frac{1}{T_{\text{Road}}} \frac{\partial T_{\text{Road}}}{\partial \hat{p}_i} \]
\[ = - \frac{2}{T_{\text{Road}}} \left( \frac{T_{\text{cor}}}{T_{\text{cor}} + T_{\text{con}}} - \frac{T_{\text{con}} \cdot T_{\text{cor}}}{(T_{\text{con}} + T_{\text{cor}})^2} \right) \cdot \frac{\partial T_{\text{cor}}}{\partial \hat{p}_i} \]
\[ = - \frac{1}{T_{\text{cor}}} \left( \frac{T_{\text{cor}}}{T_{\text{cor}} + T_{\text{con}}} \right) \cdot \frac{\partial T_{\text{cor}}}{\partial \hat{p}_i} \]
\[ = - \frac{1}{T_{\text{cor}}(T_{\text{con}} + T_{\text{cor}})} \cdot \frac{y_i \cdot |\hat{P}| - |Y \circ \hat{P}|}{|\hat{P}|^2}. \] (20)

2.4. Multitask Learning

In our road extraction framework, the input image \( I \) feeds into the shared encoder to extract features. The feature maps are, respectively, input into the decoders corresponding to the two tasks to obtain road detection prediction and road centerline extraction prediction. As shown in Figure 1, the prediction of the road detection task \( \hat{Y} \) and the ground truth \( Y \) are used to calculate the weighted cross-entropy loss that can be minimized to update the parameters of the road detection network. The prediction of the road centerline extraction task \( \hat{Y} \) and the discrete normalized distance label map \( L \) are used to calculate the ordinal loss that can be minimized to update the parameters of the road centerline extraction network. \( Y, \hat{Y}, \hat{L}, \) and \( L \) are used to calculate our proposed road-topology loss that combines the prediction of road detection and the prediction of road centerline extraction. Our proposed road topology loss makes full use of the correlation between the two tasks and can be minimized to make the two tasks promote each other. The entire loss function is the sum of the cross-entropy loss, the ordinal loss, and road-topology loss. By minimizing the entire loss function, the parameters of the road detection network and the parameters of the road centerline extraction network can be updated simultaneously, which means our framework learns the road detection task and road centerline extraction task. The loss of mini-batch is calculated by

\[ L = \frac{1}{|B|} \sum_{I \in B} (L_{rs} + L_c + L_T). \] (21)

3. Experiments

3.1. Dataset

In this subsection, we aim to present the dataset used in this paper. There is no public dataset that is applicable to our research. We create our dataset by using high-resolution TerraSAR-X images that are obtained by striped mode. As shown in Table 1, we label the roads in two SAR images, the coverage areas of which include urban, suburb, and rural areas. The Google Earth Maps of study area are shown in Appendix A. As the dataset is applied to extract roads, we only label all roads in the region where the road network is dense. Our labeled area was split into a training and test set as follows: the upper 80 % of the area \((435 \times 1024 \times 1024 \text{ pixels})\) was used for training, and the lower 20% \((104 \times 1024 \times 1024 \text{ pixels})\) for testing. We use the raw SAR intensity image without any preprocessing. There are speckle noise in SAR images. Our train set contains raw SAR patches, the ground truth of road, and the ground truth of discrete labels. Algorithm 1 describes how road centerline and road width ground truth can be outlined. Figure 3 shows a sample in the train set. There is the only ground truth of road and ground truth of road centerline in the test set.
Figure 3. (a): SAR image. (b): ground truth of road. (c): ground truth of discrete label.

Algorithm 1 Obtain Road Centerline and Road Width Ground Truth

**Input:** road ground truth Y
**Output:** Road centerline ground truth C, Road width ground truth W

1: Calculate the distance map $D$
2: From distance map $D$, calculate the horizontal gradient map $D_X = D \otimes f$, where $f = [-1, 2, -1]$ 
3: From distance map $D$, calculate the vertical gradient map $D_Y = D \otimes f^T$, where $f = [-1, 2, -1]$ 
4: Calculate the orientation map $OR = \tan^{-1}(D_Y, D_X)$ that is perpendicular to road direction.
5: for (do $i \in I$ and $y_i == 1$)
6: if $d_i$ is local maximum along the direction of $i$, then 
7: $i$ is on the road centerline: $c_i = 1$
8: the road width of pixel $i$: $w_i = 2 \cdot d_i$
9: else 
10: $i$ is not on the road centerline: $c_i = 0$
11: the road width of pixel $i$: $w_i = 2 \cdot d_j$, where $d_j$ is the local maximum starting from $i$ along the direction of $d_j$ in the distance map $D$.

Table 1. Metadata of the TerraSAR-X Images Used in Our Data Set.

|                | Beijing, China                          | Beijing, China                          |
|----------------|-----------------------------------------|-----------------------------------------|
| Size           | 21,800 × 15,500 px                      | 27,600 × 18,700 px                     |
| Range Sample Distance | 0.909627 m/px                            | 0.908790 m/px                           |
| Azimuth Sample Distance | 1.848561 m/px                            | 1.888833 m/px                           |
| Spatial Resolution | 3 m/px                                  | 3 m/px                                  |
| Center Coordinate | [39.8798466,116.4503446]                 | [39.9571641,116.6996268]                |
| Polarization    | HH                                      | HH                                      |
| Date           | 2013-05-09                               | 2013-03-04                              |
3.2. Evaluation Metrics

To assess the qualitative performance in both road detection task and road centerline extraction task, we apply the metrics that are introduced in [47]. In this paper, our framework can detect road and extract road centerline simultaneously. To highlight the characteristics of our method, we propose a series of metrics to evaluate the performance of road extraction task.

3.2.1. Road Detection

The metrics that are employed to evaluate the performance of our approach for road detection task are precision, recall, F1-score, and intersection over union (IoU) that is the ratio of the intersection of prediction and groundtruth to the union of prediction and groundtruth. Precision (P) measures the ratio of the number of the pixels which are labeled as road pixels in the ground truth and are predicted as road pixels to the number of pixels that are inferred as road pixels. Recall (R) calculates the ratio of the number of the pixels which are labeled as road pixels in the ground truth and are predicted as road pixels to the number of pixels that are labeled as road pixels. F1-score ($F_{1rd}$) is used to balance precision and recall, which is a harmonic average between precision and recall. IoU is the ratio of the intersection and union of the true label and predicted result, which can trade-off between recall and precision. Specifically, the four metrics are defined as:

$$P = \frac{TP}{TP + FP'},$$  \hspace{1cm} (22)

$$R = \frac{TP}{TP + FN'},$$  \hspace{1cm} (23)

$$F_{1rd} = \frac{2 \times Precision \times Recall}{Precision + Recall},$$ \hspace{1cm} (24)

$$IoU = \frac{TP}{TP + FN + FP'},$$ \hspace{1cm} (25)

where $TP$ is true positive, $FP$ is false positive and $FN$ is false negative.

As there is a deviation between the manually labelled roads and the real roads, we relax metrics using the buffer method given in [48]. Specifically, if the regions in the prediction result are within the $\rho$ pixels range, they are regarded as matching regions. In this paper, we set $\rho = 2$.

3.2.2. Road Centerline Extraction

Due to the differences between the road detection task and road centerline extraction task, it is better to use different metrics for the road centerline task. Next, we present the metrics that are used for the road centerline extraction task. We calculate completeness, correctness, quality, and F1-score to assess the performance of our approach for the road centerline extraction task. As it was difficult to directly compare the pixel difference between the extracted centerline and the ground truth, we introduced a buffer-based evaluation for the road centerlines. Completeness (COM) is the ratio of the length of reference road centerline that lies within the buffer around the extracted centerline to the length of reference centerline. Correctness (COR) is the ratio of the extracted road centerline that lies within the buffer around the reference centerline to the length of the extracted centerline. Quality (Q) is a comprehensive metric that combines the completeness and correctness. F1-score ($F_{1rec}$) is used to balance COM and COR, which is a harmonic average between COM and COR. The four metrics can be calculated as
COM = \frac{\text{length of matched reference}}{\text{length of reference}} \approx \frac{TP}{TP + FN'}, \quad (26)

COR = \frac{\text{length of matched extraction}}{\text{length of extraction}} \approx \frac{TP}{TP + FP'}, \quad (27)

Q = \frac{\text{length of matched extraction} + \text{length of unmatched reference}}{\text{length of extraction}} \approx \frac{TP}{TP + FN + FP'}, \quad (28)

F_{1re} = \frac{2 \times \text{COM} \times \text{COR}}{\text{COM} + \text{COR}} \approx \frac{2 \times P \times R}{P + R}. \quad (29)

### 3.2.3. Road Extraction

From the definition, we know that the metrics of road detection: precision \((P)\), recall \((R)\), IoU, F1-score \((F_{1rd})\) correspond to the metrics of road centerline extraction: correctness \((COR)\), completeness \((COM)\), quality \((Q)\), F1-score \((F_{1rc})\), respectively. As a result, we design four metrics: precision for road extraction \((P_{re})\), recall for road extraction \((R_{re})\), quality for road extraction \((Q_{re})\), F1-score for road extraction \((F_{1re})\) to evaluate the performance for road extraction. The metrics for road extraction are given as

\[
P_{re} = \frac{\alpha \times P + \beta \times COR}{\alpha + \beta}, \quad (30)
\]

\[
R_{re} = \frac{\alpha \times R + \beta \times COM}{\alpha + \beta}, \quad (31)
\]

\[
Q_{re} = \frac{\alpha \times IoU + \beta \times Q}{\alpha + \beta}, \quad (32)
\]

\[
F_{1re} = \frac{\alpha \times F_{1rd} + \beta F_{1re}}{\alpha + \beta}, \quad (33)
\]

where \(\alpha\) and \(\beta\) can be set according to the importance of the road detection task and the road centerline extraction task. If the evaluator pays more attention to road detection task, then \(\alpha > \beta\). Otherwise, we set \(\alpha < \beta\). If the road detection task and road centerline extraction task are equally important, then we set \(\alpha = \beta\).

### 3.3. Implementation Details

In this subsection, we will specifically present the values of all hyperparameters in our experiment. We adopt the Pytorch framework to implement networks trained on a single NVIDIA Tesla V100 with 16G memory using a batch size of one. We train the networks with AdamW optimizer with the initial learning rate of 0.001, and we drop the learning rate by the factor of 0.1 at every ten epochs. We applied data augmentation to the training set with image rotation and horizontal and vertical flips. The augmented training set is composed of 3480 SAR images, the size of which is \(1024 \times 1024\). For road centerline extraction, we set the threshold \(T\) as 0.25. As \(T\) only works in the preprocessing step of our test stage, this will not affect the final road centerline extraction result too much. If the parameter \(T\) is set too large, there will be more discontinuities in the road centerline extraction results. In practice, we do not know the optimal value of \(T\) for a test image without ground truth, and users can set \(T\) according to their needs. If users do not want spurious parts in the results, they can set a larger \(T\). Conversely, if the user prefers to ensure the completeness of the road centerline extraction result, \(T\) can be set to a smaller value. In this paper, we set the \(T\) to 0.25 that is neither too big nor too small to remove some spurious parts and guarantee the connectivity of the road centerlines.

### 3.4. Results

In this subsection, we first introduce the networks that are used as baseline methods in this paper. We choose three fully convolutional neural networks (FCNNs), including LinkNet34 [48], DLinkNet34 [49], and DeeplabV3plus [50] as baseline methods to study the performance of our method. These three networks were implemented to extract roads.
from optical remote sensing imagery. To adapt to the size of images in our dataset, we set the dilation rates of dilated conventional operations in DeeplabV3plus to $[2, 4, 8, 16]$. To verify that both the proposed multitask architecture and loss function are effective, we will perform three sets of comparative experiments. Each set of comparative experiments contains three methods, including the baseline method, the method I, the method II. As shown in Figure 4, (a) is the network architecture of the baseline method, while (b) and (c) are the network architectures of the method I and the method II, respectively, which both have two branches: road detection branch and road centerline extraction branch. The methods I in the three sets of comparative experiments are, respectively, abbreviated as LinkNet34+, DLinkNet34+, and DeeplabV3plus+, which are obtained by modifying the networks of the baseline methods with dual and identical decoders having shared encoders. The methods II in the three sets of comparative experiments are, respectively, abbreviated as LinkNet34++, DLinkNet34++, and DeeplabV3plus++, which use the road topology loss to update the network parameters and have the same network architecture as the methods I.

Figure 4. (a) baseline methods with only road detection branch; (b) the methods obtained by modifying baseline methods with road centerline extraction branch. (c) the methods obtained by modifying baseline methods with road centerline extraction branch and road-topology loss. ‘RD’ denotes the road detection; ‘RCE’ denotes the road centerline extraction.

Comparative Evaluation on Road Detection

To compare the performance of road detection, all the methods are evaluated based on the test samples in the test set for road detection. For qualitative comparison, we show the results produced by all methods based on example images depicted in Figure 5. The quantitative comparisons are reported in Table 2.
depicts the obvious advantage of our methods

Figure 5. Qualitative comparison of road detection results produced by different methods.

As shown in Figure 5, the methods with multitask learning architecture and using road-topology loss function generally perform better than the baseline methods. The baseline methods miss road regions in many places, i.e., the false negative (green) part is large, which leads to the poor connectivity of road detection result. With the learning road detection task and road centerline extraction based on ordinal regression task jointly, the false negative part is smaller. By further using the road-topology loss function that can penalize discontinuous parts of the detected road, the performance of road connectivity is improved. Table 2 presents the comparative quantitative evaluation measured in terms of $P$, $R$, IoU, $F_1$. In Table 2, the best value of performance is presented with bold style. We observe from the Table 2 that the proposed methods outperform baselines in terms of
three metrics, i.e., R, IoU, F1. Though the precision of proposed methods is less than the baseline, the slight decrease in P is insignificant compared to the large increase in the other three metrics.

Table 2. Comparative quantitative Evaluation Among Different Methods for Road Detection on our dataset. It should be noted that the results are the average performance of all images in the test set. With the best results marked in bold.

| Methods Names | P     | R     | IoU   | F1     |
|---------------|-------|-------|-------|--------|
| DeeplabV3plus | 0.9153| 0.6341| 0.5629| 0.7378 |
| DeeplabV3plus+| 0.9039| 0.6666| 0.5980| 0.7534 |
| DeeplabV3plus++| 0.8691| 0.7673| 0.6827| 0.8069 |
| DLinkNet34    | 0.9295| 0.6300| 0.5666| 0.7382 |
| DLinkNet34+   | 0.9314| 0.6478| 0.5899| 0.7507 |
| DLinkNet34++  | 0.8922| 0.7424| 0.6707| 0.8009 |
| LinkNet34     | 0.9252| 0.6565| 0.5918| 0.7554 |
| LinkNet34+    | 0.9097| 0.6897| 0.6216| 0.7724 |
| LinkNet34++   | 0.8948| 0.7415| 0.6693| 0.8007 |

3.5. Comparison of Road Centerline Extraction

In this subsection, we present the result of a comparative evaluation based on the centerline extraction task. For LinkNet34, DLinkNet34, DeeplabV3plus, we applied the morphological thinning algorithm [51] to the road detection results so as to extract road centerline. Figure 6 illustrates the centerlines identified by different methods. From Figure 6, we can see that more discontinuities are observed in road centerline extraction results of baseline methods. As shown in Figure 6, by learning the road centerline extraction task based on ordinal regression to extract road centerline, false negative parts in the road centerline extraction results decrease, and the connectivity of the centerline network is enhanced. Figure 6 also shows that by using road topology loss, the quality of road centerline extraction results has been further improved. The one reason is that our method can learn centerline extraction task that is based on ordinal regression and the road detection task simultaneously under a multitask learning scheme, whereas the prediction of one subtask can bootstrap the performance of solving another subtask. The other reason is that minimizing our firstly proposed road-topology loss helps eliminate topology errors of road extraction results, which include spurious parts and gaps. Table 3 summarized the results evaluated by COM, COR, Q,and F1 metrics. From Table 3, we can see that the methods with centerline extraction based on ordinal regression perform better than baseline methods. The methods with centerline extraction based ordinal regression and road-topology loss achieve the best performance.
Figure 6. Qualitative comparison of road centerline extraction results produced by different methods. For visualization, we perform the morphological dilation operation on the road centerline extraction results.

depicts the obvious advantage of our methods
Table 3. Comparative quantitative Evaluation Among Different Methods for Road Centerline Extraction on our dataset. It Should Be Noted That the Results are the Average Performance of all Images in Test Set. With the best results marked in bold.

| Methods Names | COR   | COM   | Q     | F1    |
|---------------|-------|-------|-------|-------|
| DeeplabV3plus | 0.9532| 0.7320| 0.6374| 0.8192|
| DeeplabV3plus+| 0.9244| 0.7860| 0.6880| 0.8379|
| DeeplabV3plus++| 0.9138| 0.8346| 0.7360| 0.8658|
| DLinkNet34   | 0.9670| 0.6929| 0.6187| 0.7955|
| DLinkNet34+  | 0.9350| 0.7839| 0.6976| 0.8456|
| DLinkNet34++ | 0.9150| 0.8276| 0.7277| 0.8618|
| LinkNet34    | 0.9668| 0.7102| 0.6363| 0.8076|
| LinkNet34+   | 0.9063| 0.8237| 0.7185| 0.8561|
| LinkNet34++  | 0.9171| 0.8265| 0.7255| 0.8623|

3.6. Comparison of Road Extraction

In this subsection, we present the result of a comparative evaluation based on the road extraction task. To compare the performance of road extraction qualitatively, we adapt our proposed metrics with different values of $\alpha$ and $\beta$ to evaluate different methods. For qualitative comparison, we show the results produced by all methods based on example images depicted in Figure 6. The quantitative comparisons are reported in Tables 4–6. From Figure 7, we can see that the poor road detection results of the baseline method lead to many discontinuities in the road centerline extraction results. As shown in Figure 7, by learning the road detection task and the road centerline extraction task based on ordinal regression to extract the road centerline, the performances of both tasks were improved. Figure 7 also shows that, by using our first proposed road-topology loss, the continuity of road detection result is enhanced, and the quality of road detection result is increased. This indicates that minimizing the road-topology loss helps eliminate topology errors in the road extraction results, which include spurious parts and gaps. Tables 4–6 summarized the results evaluated by $P_{re}$, $R_{re}$, $Q_{re}$, and $F1_{re}$ metrics. We can observe that whether the evaluator pays more attention to the road detection task or the road centerline extraction task, the method II that learns road detection task and road centerline extraction task simultaneously and adapt our proposed road-topology loss can achieve the best performance.

Table 4. Comparative quantitative Evaluation Among Different Methods for Road Extraction on our dataset. With the best results marked in bold. $\alpha = 1$, $\beta = 1$.

| Methods Names | $P_{re}$ | $R_{re}$ | $Q_{re}$ | $F1_{re}$ |
|---------------|----------|----------|----------|-----------|
| DeeplabV3plus| 0.9342   | 0.6831   | 0.6002   | 0.7785    |
| DeeplabV3plus+| 0.9141   | 0.7263   | 0.6430   | 0.7957    |
| DeeplabV3plus++| 0.8915   | 0.8010   | 0.7094   | 0.8364    |
| DLinkNet34   | 0.9483   | 0.6615   | 0.5926   | 0.7669    |
| DLinkNet34+  | 0.9332   | 0.7159   | 0.6437   | 0.7981    |
| DLinkNet34++ | 0.9036   | 0.7850   | 0.6992   | 0.8313    |
| LinkNet34    | 0.9460   | 0.6833   | 0.6141   | 0.7815    |
| LinkNet34+   | 0.9080   | 0.7567   | 0.6701   | 0.8143    |
| LinkNet34++  | 0.9060   | 0.7840   | 0.6974   | 0.8315    |
Figure 7. Qualitative comparison of road extraction results produced by different methods. For visualization, we perform the morphological dilation operation on the road centerline extraction results.
Table 5. Comparative quantitative Evaluation among Different Methods for Road Extraction on our dataset. With the best results marked in bold. $\alpha = 2, \beta = 1$.

| Methods Names   | $P_{re}$ | $R_{re}$ | $Q_{re}$ | $F_{1re}$ |
|-----------------|----------|----------|----------|-----------|
| DeeplabV3plus   | 0.9279   | 0.6668   | 0.5877   | 0.7649    |
| DeeplabV3plus+  | 0.9107   | 0.7064   | 0.6280   | 0.7816    |
| DeeplabV3plus++ | 0.8840   | 0.7897   | 0.7005   | 0.8265    |
| DLInNet34       | 0.9420   | 0.6510   | 0.5839   | 0.7573    |
| DLInNet34+      | 0.9326   | 0.6932   | 0.6258   | 0.7823    |
| DLInNet34++     | 0.8998   | 0.7708   | 0.6897   | 0.8212    |
| LinkNet34       | 0.9391   | 0.6744   | 0.6067   | 0.7728    |
| LinkNet34+      | 0.9086   | 0.7344   | 0.6539   | 0.8003    |
| LinkNet34++     | 0.9022   | 0.7699   | 0.6880   | 0.8213    |

Table 6. Comparative quantitative Evaluation Among Different Methods for Road Extraction on our dataset. With the best results marked in bold. $\alpha = 1, \beta = 2$.

| Methods Names   | $P_{re}$ | $R_{re}$ | $Q_{re}$ | $F_{1re}$ |
|-----------------|----------|----------|----------|-----------|
| DeeplabV3plus   | 0.9405   | 0.6994   | 0.6126   | 0.7921    |
| DeeplabV3plus+  | 0.9175   | 0.7462   | 0.6580   | 0.8098    |
| DeeplabV3plus++ | 0.8989   | 0.8122   | 0.7183   | 0.8462    |
| DLInNet34       | 0.9545   | 0.6720   | 0.6013   | 0.7764    |
| DLInNet34+      | 0.9338   | 0.7386   | 0.6617   | 0.8140    |
| DLInNet34++     | 0.9074   | 0.7992   | 0.7087   | 0.8415    |
| LinkNet34       | 0.9529   | 0.6923   | 0.6215   | 0.7902    |
| LinkNet34+      | 0.9075   | 0.7790   | 0.6862   | 0.8282    |
| LinkNet34++     | 0.9097   | 0.7982   | 0.7068   | 0.8418    |

4. Discussion

In this section, we will discuss the robustness of different methods when the binarization threshold is changed. When we set different binarization thresholds, the road detection results will be different. However, in practice, we do not know the optimal binarization threshold. Figure 8 reflects the statistical characteristics of IoU and F1 across different binarization thresholds. The IoU and F1-score results of baseline methods (LinkNet34, DLInNet34, DeeplabV3plus) are unstable with the vary of the binarization thresholds, as evidenced by the large variance in IoU and F1. In contrast, the F1 and IoU of our methods show better stability. Figure 8 also reports that, for F1 and IoU metrics, the means of the baseline methods are less than our proposed methods’ means. This is due to the use of a multitask learning scheme that can make two tasks mutually promote each other, and a road-topology loss that is minimized to eliminate topology errors. In conclusion, our methods can achieve a high performance under a wider range of binarization thresholds than baseline methods.
5. Conclusions

In this paper, we have learned the road detection task and the road centerline extraction task jointly with a multitask learning scheme to solve the problem of the road extraction from SAR imagery. To eliminate topology errors in road extraction results, we have specially designed a road-topology loss function for road extraction, which is differentiable. Different from the centerline extraction method based on regression method or classification method, we have adapted ordinal regression to learn discrete distance labels and trained the network by minimizing ordinal loss. By using the road centerline extraction method based on ordinal regression, the network is not sensitive to incorrect labels and can converge to a satisfactory result. Using multitask learning architecture, we have made full use of the correlation between the road detection task and the road centerline extraction task by learning the two tasks at the same time. The performance of one task has been improved under bootstrapping by the other task. The test result has shown that the networks, modified as multitask architecture, perform better than baseline methods. Considering the unique topological characteristics of the road, we have proposed a new road-topology loss function to penalize spurious parts in centerline extraction results and gaps in road detection results. The results show that the proposed road-topology loss function improved the connectivity and completeness of road networks. Finally, we
discussed the robustness of our method, and the result has shown that our method not only greatly improves the performance of road extraction, but is also more stable than the baseline methods.

6. Future Work

Although our proposed method improves the performance of road extraction task, there are still some false detections. There are two main reasons for the false detection. One reason is that the noisy speckle nature of sar images, the layover and the shadowing effects of imaging destroy the roads in the SAR image. This leads many road parts to be missed. The other reason is that the roads in sar images can often be confused with other targets, such as railway tracks, rivers, or even tree hedges. This leads some non-road regions to be detected as the roads. In the future, we will further improve our network to eliminate the influence of these two factors.

Author Contributions: Conceptualization, X.W.; methodology, X.W.; software, X.W.; validation, X.W.; formal analysis, X.W.; investigation, X.W.; resources, X.L.; data curation, X.W. and X.L.; writing—original draft preparation, X.W.; writing—review and editing, X.W., X.L.; visualization, X.W.; supervision, X.W.; project administration, X.L. and K.Z.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China grant number 2018YFC1505100.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Figures A1 and A2 are the Google Earth images of the study area.

Figure A1. The Google Earth image of the study area.
Figure A2. The Google Earth image of the study area.

References

1. Tupin, F.; Maitre, H.; Mangin, J.; Nicolas, J.; Pechersky, E. Detection of linear features in SAR images: Application to road network extraction. *IEEE Trans. Geosci. Remote Sens.* 1998, 36, 434–453. [CrossRef]

2. Negri, M.; Gamba, P.; Lisini, G.; Tupin, F. Junction-aware extraction and regularization of urban road networks in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* 2006, 44, 2962–2971. [CrossRef]

3. Lu, P.; Du, K.; Yu, W.; Wang, R.; Deng, Y.; Balz, T. A New Region Growing-Based Method for Road Network Extraction and Its Application on Different Resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2014, 7, 4772–4783. [CrossRef]

4. He, C.; Liao, Z.; Yang, F.; Deng, X.; Liao, M. Road Extraction From SAR Imagery Based on Multiscale Geometric Analysis of Detector Responses. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2012, 5, 1373–1382.

5. Jiang, M.; Miao, Z.; Gamba, P.; Yong, B. Application of Multitemporal InSAR Covariance and Information Fusion to Robust Road Extraction. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 3611–3622. [CrossRef]

6. Liu, J.; Sui, H.; Tao, M.; Sun, K.; Xin, M. Road extraction from SAR imagery based on an improved particle filtering and snake model. *Int. J. Remote Sens.* 2013, 34, 8199–8214. [CrossRef]

7. Cheng, J.; Guan, Y.; Ku, X.; Sun, J. Semi-automatic road centerline extraction in high-resolution SAR images based on circular template matching. In Proceedings of the 2011 International Conference on Electric Information and Control Engineering (ICECE), Wuhan, China, 15–17 April 2011; pp. 1688–1691. [CrossRef]

8. Xu, R.; He, C.; Liu, X.; Dong, C.; Qin, Q. Bayesian Fusion of Multi-Scale Detectors for Road Extraction from SAR Images. *Int. J. Geo-Inf.* 2017, 6, 26. [CrossRef]
9. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

10. Wu, J.; Yu, Y.; Huang, C.; Yu, K. Deep multiple instance learning for image classification and auto-annotation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3460–3469.

11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 770–778.

12. Ciregan, D.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3642–3649.

13. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual Attention Network for Image Classification. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6450–6458.

14. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85.

15. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

16. Ros, G.; Sellart, L.; Materzynska, J.; Vazquez, D.; Lopez, A.M. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 3234–3243.

17. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Learning a Discriminative Feature Network for Semantic Segmentation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1857–1866.

18. Huang, Z.; Wang, X.; Wang, J.; Liu, W.; Wang, J. Weakly-Supervised Semantic Segmentation Network with Deep Seeded Region Growing. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7014–7023.

19. Shen, F.; Gan, R.; Yan, S.; Zeng, G. Semantic Segmentation via Structured Patch Prediction, Context CRF and Guidance CRF. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5178–5186.

20. Pan, X.; Zhao, J.; Xu, J. Conditional Generative Adversarial Network-Based Training Sample Set Improvement Model for the Semantic Segmentation of High-Resolution Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 2020, 1–17. [CrossRef]

21. Li, A.; Jiao, L.; Zhu, H.; Li, L.; Liu, P. Multitask Semantic Boundary Awareness Network for Remote Sensing Image Segmentation. IEEE Trans. Geosci. Remote Sens. 2021, 1–14. [CrossRef]

22. Zheng, C.; Zhang, Y.; Wang, L. Semantic Segmentation of Remote Sensing Imagery Using an Object-Based Markov Random Field Model With Auxiliary Label Fields. IEEE Trans. Geosci. Remote Sens. 2017, 55, 3015–3028. [CrossRef]

23. Ding, L.; Zhang, J.; Bruzzone, L. Semantic Segmentation of Large-Scale VHR Remote Sensing Images Using a Two-Stage Multiscale Training Architecture. IEEE Trans. Geosci. Remote Sens. 2020, 58, 5367–5376. [CrossRef]

24. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote Sens. Lett. 2018, 15, 749–753. [CrossRef]

25. Kussul, N.; Lavreniuk, M.; Skakan, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geosci. Remote Sens. Lett. 2017, 14, 778–782. [CrossRef]

26. Zhang, Q.; Kong, Q.; Zhang, C.; You, S.; Li, L. A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network. Eur. J. Remote Sens. 2019, 52, 572–582. [CrossRef]

27. Shahzad, M.; Maurer, M.; Fraundorfer, F.; Wang, Y.; Zhu, X. Buildings Detection in VHR SAR Images Using Fully Convolutional Neural Networks. IEEE Trans. Geosci. Remote Sens. 2019, 57, 1100–1116. [CrossRef]

28. Zhao, Z.; Jiao, L.; Zhao, J.; Gu, J.; Zhao, J. Discriminant Deep Belief Network for High-Resolution SAR Image Classification. Pattern Recognit. 2016, 61, 686–701. [CrossRef]

29. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.Q. Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 2017, 13, 1935–1939. [CrossRef]

30. Geng, J.; Wang, H.; Fan, J.; Ma, X. Deep Supervised and Contractive Neural Network for SAR Image Classification. IEEE Trans. Geosci. Remote Sens. 2017, 55, 2442–2459. [CrossRef]

31. Chen, J.; Qiu, X. Equivalent Complex Valued Deep Semantic Segmentation Network For SAR Images. In Proceedings of the International Applied-Computational-Electromagnetics-Society Symposium China (ACES), Nanjing, China, 8–11 August 2019; Volume 1, pp. 1–2.

32. Mohammadimanesh, F.; Salehi, B.; Mandianpari, M.; Gill, E.; Molinier, M. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. ISPRS J. Photogramm. Remote Sens. 2019, 151, 223–236. [CrossRef]

33. Wu, W.; Li, H.; Li, X.; Guo, H.; Zhang, L. PolSAR Image Semantic Segmentation Based on Deep Transfer Learning—Realizing Smooth Classification With Small Training Sets. IEEE Geosci. Remote Sens. Lett. 2019, 16, 977–981. [CrossRef]
34. Li, J.; Zhang, R.; Li, Y. Multiscale convolutional neural network for the detection of built-up areas in high-resolution SAR images. In Proceedings of the 36th IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 910–913. [CrossRef]

35. Henry, C.; Azimi, S.M.; Merkle, N. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* 2018, 15, 1867–1871. [CrossRef]

36. Turetken, E.; Benmansour, F.; Andres, B.; Glowacki, P.; Pfister, H.; Fua, P. Reconstructing Curvilinear Networks Using Path Classifiers and Integer Programming. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016, 38, 2515–2530. [CrossRef]

37. Glowacki, P.; Pinheiro, M.A.; Mosinska, A.; Turetken, E.; Lebrecht, D.; Sznitman, R.; Holmaat, A.; Kybic, J.; Fua, P. Reconstructing Evolving Tree Structures in Time Lapse Sequences by Enforcing Time-Consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 40, 755–761. [CrossRef]

38. Sironi, A.; Turetken, E.; Lepetit, V.; Fua, P. Multiscale Centerline Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016, 38, 1327–1341. [CrossRef]

39. Wang, Y.; Wei, X.; Liu, F.; Chen, J.; Zhou, Y.; Shen, W.; Fishman, E.K.; Yuille, A.L. Deep Distance Transform for Tubular Structure Segmentation in CT Scans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 3832–3841.

40. Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A Higher-Order CRF Model for Road Network Extraction. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 8–14 December 2019; pp. 1698–1705. [CrossRef]

41. Han, X.; Xu, C.; Prince, J.L. A topology preserving level set method for geometric deformable models. *IEEE Trans. Pattern Anal. Mach. Intell.* 2003, 25, 755–768.

42. Stoica, R.; Descombes, X.; Zerubia, J. A Gibbs Point Process for Road Extraction from Remotely Sensed Images. *Int. J. Comput. Vis.* 2004, 57, 121–136. [CrossRef]

43. Mosinska, A.; Marquez-Neila, P.; Kozinski, M.; Fua, P. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3136–3145.

44. Hu, X.; Fuxin, L.; Samaras, D.; Chen, C. Topology-Preserving Deep Image Segmentation. arXiv 2019, arXiv:1906.05404.

45. Zhao, K.; Gao, S.; Wang, W.; Cheng, M.M. Optimizing the F-measure for Threshold-free Salient Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8848–8856.

46. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal Regression with Multiple Output CNN for Age Estimation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 4920–4928.

47. Heipke, C.; Mayer, H.; Wiedemann, C.; Jamet, O. Evaluation of Automatic Road Extraction. *Int. Arch. Photogramm. Remote Sens.* 1997, 34-4/W2, 151–156.

48. Mnih, V.; Hinton, G. Learning to Label Aerial Images from Noisy Data. In Proceedings of the 29th International Conference on Machine Learning (ICML), Edinburgh, UK, 26 June–1 July 2012; Volume 1, pp. 567–574.

49. Zhou, L.; Zhang, C.; Ming, W. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–196.

50. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

51. Zhang, T.Y.; Suen, C.Y. A Fast Parallel Algorithm for Thinning Digital Patterns. *Commun. ACM* 1984, 27, 236–239. [CrossRef]