A Self-Supervised Learning Framework for Road Centerline Extraction From High-Resolution Remote Sensing Images

Qing Guo, *Member, IEEE*, and Zhipan Wang

***Abstract*—Road extraction from the high-resolution remote sensing image is significant for the land planning, vehicle navigation, etc. The existing road extraction methods normally need many preprocessing and subsequent optimization steps. Therefore, an automatic road centerline extraction method based on the self-supervised learning framework for high-resolution remote sensing image is proposed. This proposed method does not need to manually select training samples and other optimization steps, such as the nonroad area removing. First, the positive sample selection method combining the spectral and shape features is proposed to extract the road sample. Then, the one-class classifier framework is introduced and the random forest positive unlabeled learning classifier is constructed to get the posterior probability of the pixel belonging to road. The shape feature and the posterior probability are combined to form the final road network in the object-oriented way. Finally, the road centerline is obtained through the tensor voting algorithm. In order to verify the effectiveness of the proposed algorithm, high-resolution remote sensing images and benchmark datasets are used to do experiments. The indexes of the completeness ratio, the correctness ratio, and the detection quality are used for the quantitative accuracy evaluation. Compared with the supervised, the unsupervised, and the one-class classification road extraction algorithms, this proposed algorithm achieves high accuracy and efficiency. For the deep learning method comparison, the deep learning method performs well in most cases especially in the complex urban area. However, the deep learning method needs a large number of samples and a long training time, and our self-supervised learning framework does not need the training samples.**

***Index Terms*—High-resolution remote sensing image, one-class classifier, road centerline, road extraction, self-supervised learning.**

I. Introduction

As AN important part of the basic geographic information, the road network plays a very significant role in the urban planning, vehicle navigation, urban spatial analysis, etc [1]. Road extraction based on the remote sensing image has the characteristic of wide coverage. Especially, with the popularity of high-resolution remote sensing images in recent years, it is possible to accurately obtain the spatial distribution of road network. However, most of the existing road extraction algorithms based on high-resolution remote sensing images rely on the manual participation, which is time-consuming and laborious. So, the automatic road extraction algorithm has become a hot topic [2].

For roads in high-resolution remote sensing images, the spectral characteristics are often uniform and the geometric structure is normally with narrow-long shape. Based on these characteristics, scholars have designed many road extraction algorithms, including the feature level methods [3]–[5], the object level methods [6], the knowledge methods [7], [8], and the machine learning methods [9]–[13]. Among the feature level methods, the template matching [14] and the Snake model [15] are typical methods. The template matching method gets the road area through the local matching based on the spectral and texture features with rectangular or circular templates. The Snake model method generally identifies the road through the road edge line. This kind of methods usually use shallow features to design algorithms. Hence, it is difficult to extract the high-quality road when the image scene is complex.

The road extraction algorithm based on machine learning theory has been widely used [16]–[20] with the development of machine learning. These methods effectively improve deficiencies of the traditional road extraction methods and have good noise immunity. The multistage framework based on salient features of roads, namely the distinct spectral contrast and the locally linear trajectory, is designed to extract roads using the probabilistic support vector machine and the dominant singular measure, which is a powerful method [9]. The deeply excavated morphological profile feature [21] combining the support vector machine (SVM) classifier is used to effectively get roads in different scenes [17]. In the literature [16], the multiscale segmentation is performed. Then, the geometric structures (such as the object-level length) of different segmentation layers are extracted. Finally, combining the spectral and the geometric structure features, the satisfactory road area is recognized through SVM. The machine learning road extraction algorithm has high accuracy comparing to the traditional methods. However, most machine learning road extraction algorithms often need to

Manuscript received January 5, 2020; revised May 5, 2020 and July 17, 2020; accepted July 29, 2020. Date of publication August 4, 2020; date of current version August 20, 2020. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19060103, in part by the National Natural Science Foundation of China under Grant 61771470 and Grant 41590853, and in part by the Key Research Program of Frontier Sciences, Chinese Academy of Sciences under Grant QYZDY-SSWDQC026. *(Corresponding author: Qing Guo.)*

Qing Guo is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: guoqing@aircas.ac.cn).

Zhipan Wang was with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. He is now with the School of Urban-Rural Development, Shaoyang University, Shaoyang 5306188, China (e-mail: wang749195@outlook.com).

Digital Object Identifier 10.1109/JSTARS.2020.3014242

manually select the positive and the negative samples, which is time-consuming and laborious. This becomes the key problem to restrict the development of the supervised learning method. The deep learning as a subclass of the machine learning, the new research topic performs well in the road extraction. However, in general, deep learning methods require large amounts of data and a lot of time to train.

Moreover, most of the existing traditional road extraction algorithms need many subsequent refinements to obtain the accurate road network, such as by extracting the length, the aspect ratio, the area and other geometric features, through one or more combinations to filter the nonroad area to get the final road network [22]–[25]. The subsequent processing steps achieve desirable results, but the addition of these steps affects the automation process of the road extraction. Furthermore, the parameter setting of different geometric features also affects the efficiency of the algorithm to a certain extent.

Through the abovementioned analysis, it can be concluded that most road extraction algorithms based on machine learning need to manually get the positive and negative training samples. Aiming at the problem of the supervised machine learning, a self-supervised learning framework method is proposed [26]. The self-supervision aims to generate training data automatically without human interaction, which allows us to use the advantage of the supervised classifier in a fully automated framework. However, it still needs to obtain negative samples, and the accuracy of the negative sample has a certain impact on the final accuracy.

Therefore, this article proposes a simple and effective method of road extraction under the framework of the self-supervised learning. This method automatically obtains the positive road sample and does not need to get the negative sample, only combining the designed positive sample one-class classifier for training. Moreover, no subsequent optimizing steps are needed. On the contrary, the road network is directly obtained. From the implementation point of view, this algorithm is an automatic method, which greatly improves the efficiency of road extraction.

The structure of this article is as follows. Section II is the design principle and procedure of the proposed algorithm. The experiments and result analysis are presented in Section III. Section IV concludes this article.

II. Road Extraction Algorithm Under SSLF

In this article, a self-supervised learning framework (SSLF) is proposed for road extraction. The overall procedure of the designed algorithm is shown in Fig. 1. The first part is the automatic road sample acquisition. The road samples are obtained by the joint constraints of the automatic spectral clustering and the road shape feature through the image segmentation. In the second part, the probability of road pixel is achieved by the classification method which only needs positive samples. The third part is the object-oriented road network extraction, which combines the shape feature from the first part and the posterior probability of road object from the second part, to automatically obtain the final road network. The fourth part uses the tensor voting algorithm to automatically connect the broken line to get the final road centerline.

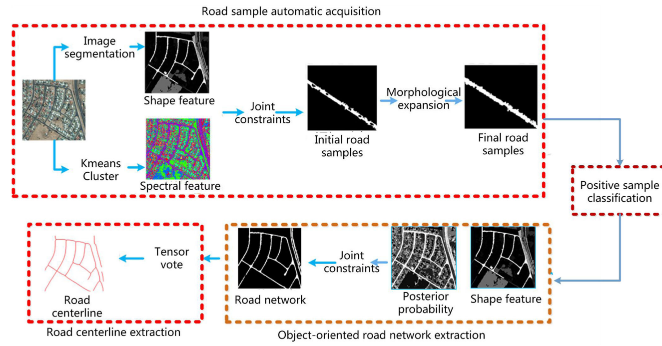


Fig. 1. Flowchart of the proposed SSLF road extraction algorithm.

A. Road Sample Acquisition

For the road extraction algorithm based on the machine learning, the positive and negative training samples are usually needed to be selected manually, which reduces the efficiency and the automation degree. Moreover, the sample choice has a crucial role in the final accuracy. In order to solve these problems, this article proposes an automatic road sample acquisition method to improve the efficiency of road extraction.

Roads in high-resolution remote sensing images are normally with the narrow-long linear shape feature and the uniform consistent spectral feature. Based on these two kinds of features, the road sample automatic acquisition method is constructed in this article. The road sample acquisition method flow is as follows.

1) The *K*-means clustering algorithm [27] is used to automatically get the initial clustering results.
2) The mean-shift segmentation algorithm [28] is applied to over segment the image, then the second-order moment shape features [9] of different objects are obtained.
3) Combining the results from step1 and step 2, the road samples are achieved by setting the simple threshold T1, as follows:

$$\begin{aligned}&\text{if } (P_{\text{Label}} = \text{Road and } S_m \geq T_1), L = \text{Road}\\&\text{else}, L = \text{nonRoad}.\end{aligned} \tag{1}$$

Herein, P_{Label} means the *K*-means initial cluster results, S_m means the second-order shape feature, L means whether the pixel is labeled as the road sample.

Through the simple but efficient *K*-means clustering algorithm, the initial road clustering results are automatically obtained. Then, one road seed point is determined manually visually, combining other feature constraints to get the initial road samples. Although, this algorithm is not a perfect unsupervised algorithm, it only needs to manually determine one pixel point, which has greatly improved the efficiency. Since the phenomenon of different things with the same spectrum is serious in the high-resolution remote sensing image, other effective feature should be considered to further optimize the road samples. Due to the narrow-long linear shape feature of road, the linear shape description operator [9] formulated by the second-order moment is combined to obtain accurate road samples. The more obvious the linear feature of target is, the greater the operator value is. The second-order moment feature is described by the object. In order to get the road object, the mean-shift image segmentation is used in our method. Meanwhile, the over-segmentation way is adopted in order to ensure the homogeneity of the road object.

B. Positive Sample Classification

In the road extraction algorithm based on the machine learning, it is usually necessary to join both the positive and negative samples in order to effectively obtain the identification results [16], [17], [22]–[24], [29]. However, the acquisition of positive and negative samples often requires the manual intervention, which greatly limits the automation degree. If only based on the positive samples, the classification results still can be obtained that will be good. The positive sample classification framework based on the Bayesian rule can solve the problem of one-class classification [30].

Setting $y \in \{-1, 1\}$indicates whether the pixel x is a target class, where 1 denotes the target class. Supposing $s \in \{0, 1\}$expresses whether the pixel x is predicted as a target class, where 1 indicates that the pixel is predicted as a target class. The positive sample is divided into two parts $\{T, V\}$, in which T represents the training data set and Vrepresents the validation data set. The negative samples are automatically randomly obtained in global, and the number of the negative sample is automatically determined by setting the ratio R of the positive sample to the negative sample. In the actual implementation process, only the positive samples need to be determined, the negative samples are obtained automatically by the proposed algorithm itself. From this perspective, the proposed method does not require any negative samples. However, inside the method, the negative samples are still needed. The classifier in the traditional machine learning is selected to train all the pixels. Then, the posterior probabilities $P(s = 1|x)$ of all the pixels predicted as the target class are achieved based on the following formula:

$$P(y = 1|x) = \frac{P(s = 1|x)}{C} \tag{2}$$

$$C = \frac{1}{n} \sum_{x \in V} P(s = 1|x) \tag{3}$$

Here, n is the number of pixels in the validation data set. It can be seen from (2) that the positive sample is divided into two parts–the training data set and the validation data set. According to the literature [31], in our experiments, the ratio is set to R = 1.0, 75% of the positive samples are used for training and 25% for predicting the posterior probability.

In addition, the influence of the sample purity on the classification accuracy should be considered. When the purity of the positive sample reaches a certain threshold value, the higher the sample purity is, the more obvious the classification accuracy reduces [31]. Therefore, in this article, a simple and effective method to solve the over pure sample is designed by using the morphological image processing technique. The morphological expansion is applied to fill the holes in the initial road samples to modify the over pure samples. In order to ensure that the initial road samples are evenly expanded, the circular structural element is adopted and the expanded range does not exceed the road boundary. In other words, the radius of the structural element should not be much large. Therefore, in our high-resolution remote sensing image experiments, the radius is set to two pixels.

For the one-class classification framework proposed in [30], the choice of classifiers can be any classifier or a combination of them. Random forest (RF) classifier is an effective classification algorithm. Compared with the SVM and other machine learning classifiers, RF classifier has the typical advantages of robustness to feature dimension, fast training convergence speed, and few parameter setting [31], [32]. Therefore, based on the RF, this article constructs a new one-class classifier–positive and unlabeled learning of random forest classifier (PULRF) to achieve the one-class classification. Herein, the number of trees is 500, the number of tree's depth is $\text{Depth} = \sqrt{\text{num}_{\text{feature}}}$, $num_{feature}$ is the number of features, which can be the number of the used remote sensing image bands. The obtained samples are put into the PULRF classifier. After the training and predicting, the posterior probability of each pixel belonging to road is obtained.

C. Object-Oriented Road Network Extraction

The traditional one-class classification method directly gets the classification result by setting the threshold based on the obtained pixel probability [31]. However, the result is only the label-classification result with serious "salt and pepper" noise [33], [34]. Meanwhile, the follow-up processes are needed, which adds the complexity and affects the automation degree of algorithm. Hence, in this article, from the object-oriented point of view [35], the way of combining the probability of object with the shape feature is proposed to directly get the road extraction result. The proposed way has the following advantages. First, it is accurate to describe the classification attribution of pixel by using the probability. Second, no follow-up optimization steps are needed, which improves the automation of algorithm. Third, the object-oriented way effectively reduces the "salt and pepper" phenomenon [24]. In addition, since the over-segmentation manner is adopted in this article, the "adhesion" phenomenon is effectively inhibited.

The main flow of the object-oriented road network extraction is in the following.

1) The label matrix and the second-order moment shape feature in the image segmentation (Section II-A), and the probability of each pixel belonging to road (Section II-B) are acquired.
2) All the labeled objects acquired by the image segmentation in step 1 are traversed. For each labeled object, the probabilities of all pixels in the object are considered as the mean probability of the object belonging to road according to the (4):

$$P_{\text{mean}} = \frac{\sum_{i=1}^{n} P_i}{n} \tag{4}$$

Where n expresses the number of pixels in the object, P_i expresses the probability of the pixel belonging to road.

3) Combining the second-order moment shape feature and the probability of each object belonging to road, the final road network is gotten by setting two thresholds T2 and P1, as follows:

$$\text{if } P_{\text{mean}} \geq P1 \text{ and } S_m \geq T2,\ \text{Pixel} = \text{Road}$$
$$\text{else, Pixel} = \text{nonRoad} \tag{5}$$

Herein, $Pixel$ indicates whether the pixel is detected as road.

D. Road Centerline Extraction

The above mentioned obtained road network inevitably produces some broken phenomenon due to some factors such as the building shadow, which affects the uniqueness and completeness of the road. The tensor voting algorithm includes tensor coding, tensor transfer, and tensor decomposition, which detects the intersection and significant area in the road network for voting [36]. The tensor voting is used to automatically connect the broken roads and achieve complete roads [17]. Moreover, only one parameter σ is needed to set in the tensor voting algorithm, which is a very important advantage. Therefore, this article uses the tensor voting algorithm to extract the road centerline.

III. Experimental Analysis and Discussion

From the view of the algorithm implementation, this proposed SSLF method belongs to the unsupervised algorithm without manual intervention. In order to verify the effectiveness and feasibility of the SSLF algorithm, the state-of-the-art supervised and unsupervised road extraction algorithms are compared, respectively. It is noted that road samples are manually selected in the supervised road extraction algorithm. In addition, the state-of-the-art deep learning algorithm is also compared. Moreover, in order to validate the robustness and applicability of our method, different images (IKONOS, Beijing-2) with different sizes (small size 675 × 507, 513 × 477, large size 2693 × 2409) are used in our experiments.

In order to quantitatively assess the validity of methods, three commonly used indicators are adopted, which are the completeness ratio, the correctness ratio, and the detection quality [37]. The formulas are as follows, respectively:

$$E_1 = \frac{TP}{TP + FN}\ , \tag{6}$$

$$E_2 = \frac{TP}{TP + FP}, \tag{7}$$

$$E_3 = \frac{TP}{TP + FP + FN}. \tag{8}$$

Herein, TP means the road length (the number of pixels) extracted correctly, FN means the road length that is not extracted, and FP means the road length extracted wrongly. The completeness ratio E_1 is to describe the integrity of road extraction. The correctness ratio E_2 is to assess the correct degree of road extraction. The detection quality E_3 is to evaluate the total quality of the road extraction results, which takes into account both the completeness and the correctness. The optimal values of all three indexes are equal to 1. The accuracy of road extraction result is not just dependent on one index, but relies on all three indexes at the same time.

Fig. 2. IKONOS image in experiment 1.

A. Comparing With Unsupervised Road Extraction Algorithm

As the proposed SSLF algorithm does not need the manual intervention, the unsupervised algorithm has a great comparability. The road extraction algorithm in literature [22] is chosen as a contrast. First, based on edge features, the road and the background are distinguished by setting thresholds. Then, the area and the length-width ratio are combined to filter the nonroad area. Finally, the road centerline is extracted based on the multiple linear regression method. According to the suggestions in literature [22], the optimal edge segmentation threshold is 0.075, the area filtering parameter is 200, the length-width ratio is 30, and the centerline extraction parameter is 30 in this experiment.

In our proposed SSLF method, the experimental parameters are set as follows. The shape constraint parameter to obtain samples is T1 = 0.85.The positive and negative sample ratio in one-class classifier is R = 1.0.The parameters of shape feature and the posterior probability to get the final road network are T2 = 0.1 and P1 = 0.6.The centerline extraction parameter σ is 20.

The experimental IKONOS image with 675 × 507 pixels is shown in Fig. 2. The spectral band is as the true color composite RGB with the spatial resolution 1 m after the multispectral and panchromatic image fusion. The reference road binary image and the reference road centerline are obtained by using the manual visual detection. The road extraction results are given in Fig. 3.

From Figs. 2 and 3, it can be seen that the road binary image extracted by our SSLF method removes the nonroad area and has accurate and complete road geometric shape, which effectively reduces the phenomenon of adhesion. For the road centerline, the centerline extracted by our method is smooth and compact. The Miao method is not satisfactory for the curve section, because the road centerline is gotten by using the fitting method. The fitting method effectively reduces glitches in the traditional centerline extraction algorithms. But, the over-fitting problem is easy to be produced for the curve or ring road.

In order to accurately verify the validity of the SSLF algorithm, the quantitative evaluation is shown in Table I. From the values, it can be seen that our SSLF method has higher accuracy

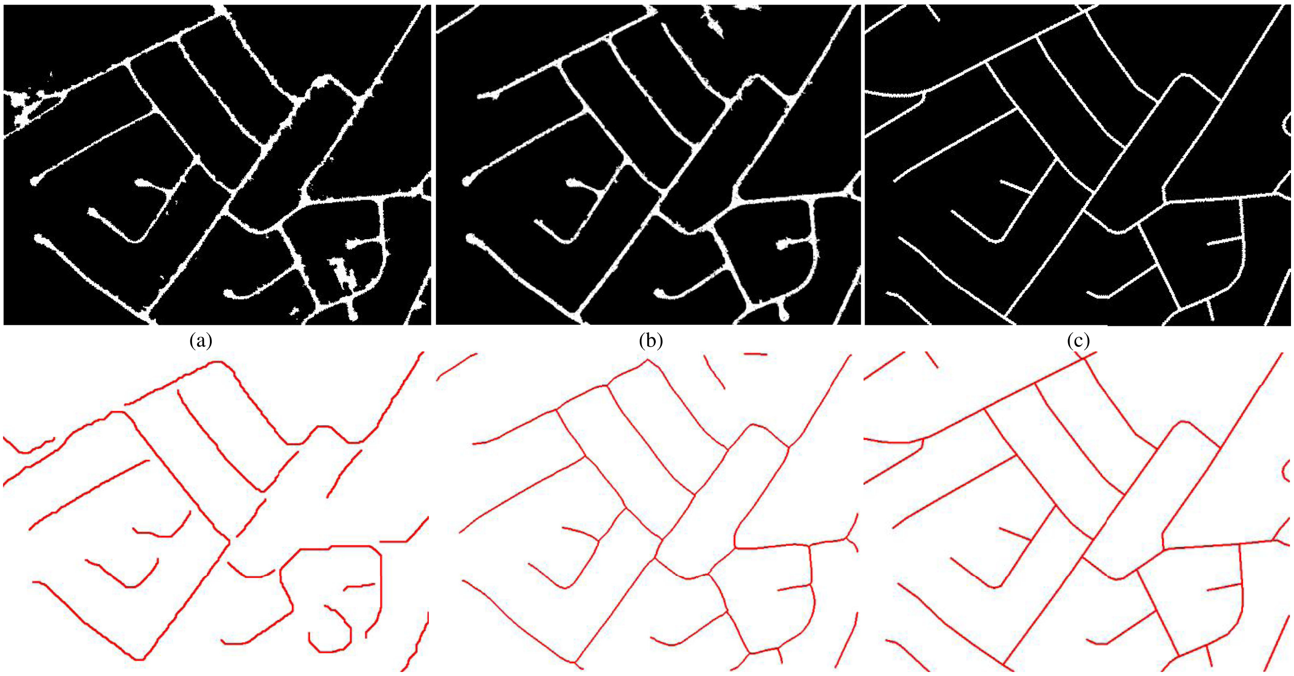


Fig. 3. Result comparison with unsupervised road extraction algorithm in experiment 1. (a) Road binary image by Miao method. (b) Road binary image by our SSLF method. (c) Reference road binary image. (d) Road centerline by Miao method. (e) Road centerline by our SSLF method. (f) Reference road centerline.

TABLE I
COMPARISON RESULT WITH THE UNSUPERVISED METHOD (%)

Method	Completeness	Correctness	Detection quality
Our SSLF	**94.42**	**96.57**	**92.18**
Miao	84.89	85.76	80.63

than that of the Miao method, which validates the efficiency of our method.

B. Comparing With Supervised Road Extraction Algorithm

In order to compare with the supervised road extraction algorithm, the algorithms in literature [17], [16] are chosen. For the Shi method, the morphological profile features of pixels are extracted from the pixel level. Then, the spectral-spatial feature set is formed by adopting the vector-stacking way. The centerline extraction method is the local weighted regression. The road extraction is finally completed through the SVM classifier. The optimal parameters of the morphological profile feature in this experiment are designed as 10, 15, 20, and 25. Moreover, the 5% ground truth data for each class is randomly selected as the training sample [17].

For the Huang method, the multiscale features of ground objects are described through the multiscale segmentation, then the object is formed. According to the narrow-long geometric feature of road, three spatial structural features such as the compactness, density, and shape of the object, combining with the spectral feature of the object are used to extract road under each scale. The multiscale road areas are then voted to get the candidate road area. Finally, the nonroad area is removed from the candidate road area to obtain the final road network. According to the reference [16], the optimal parameters of this experiment are as follows. The number of segmentation scale is 4 and the scales are 10, 20, 30, and 40, respectively. The area filtering parameter is set as 200. The centerline is achieved using the morphological refinement algorithm. According to the true distribution of ground objects, road and nonroad samples are selected to complete the road extraction in two-class way through the SVM classifier.

In experiment 2, the relevant parameters of our SSLF method are in the following. The shape constraint parameter for obtaining sample is $T1 = 0.85$. The ratio of positive to negative sample in one-class classifier is $R = 1.0$. The parameters of shape feature and the posterior probability to get the final road network are $T2 = 0.2$ and $P1 = 0.6$. The centerline extraction parameter is 25.

The 1 m spatial resolution IKONOS image with 513×477 pixels and true color composite RGB is shown in Fig. 4. Fig. 5 shows the road extraction results of different methods, the reference centerline and binary image.

From the visual point of view in Figs. 4 and 5, the road shape of the SSLF method is complete and smooth. For the Huang method, road "spur" phenomenon is serious, because the traditional morphological refinement algorithm is difficult to get the ideal centerline for the nonsmooth edge. The Shi method has similar binary image as the proposed SSLF method. Hence, the validity of the SSLF algorithm is verified from this experiment 2.

The quantitative evaluation of different methods is shown in Table II. The SSLF method has the best values of three indices among three different methods. The Shi method has the higher correctness ratio and detection quality than the Huang method;

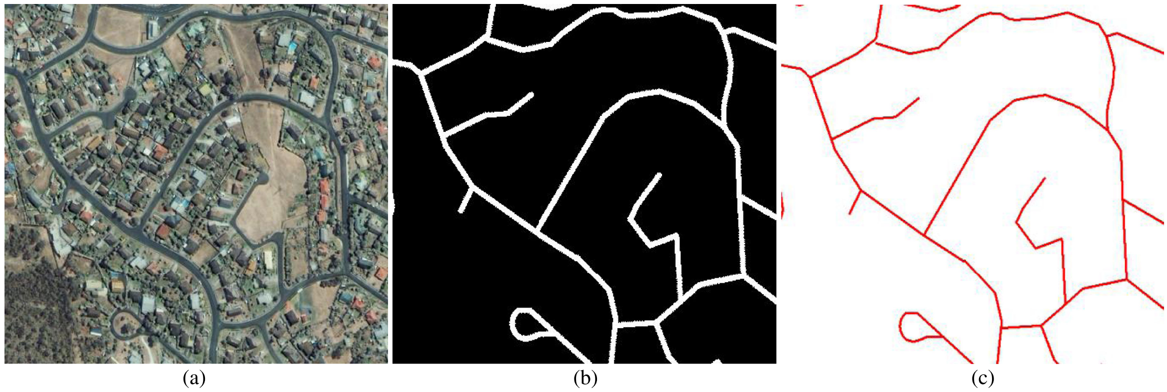


Fig. 4. IKONOS image and references in experiment 2. (a) IKONOS image. (b) Reference road binary image. (c) Reference road centerline.

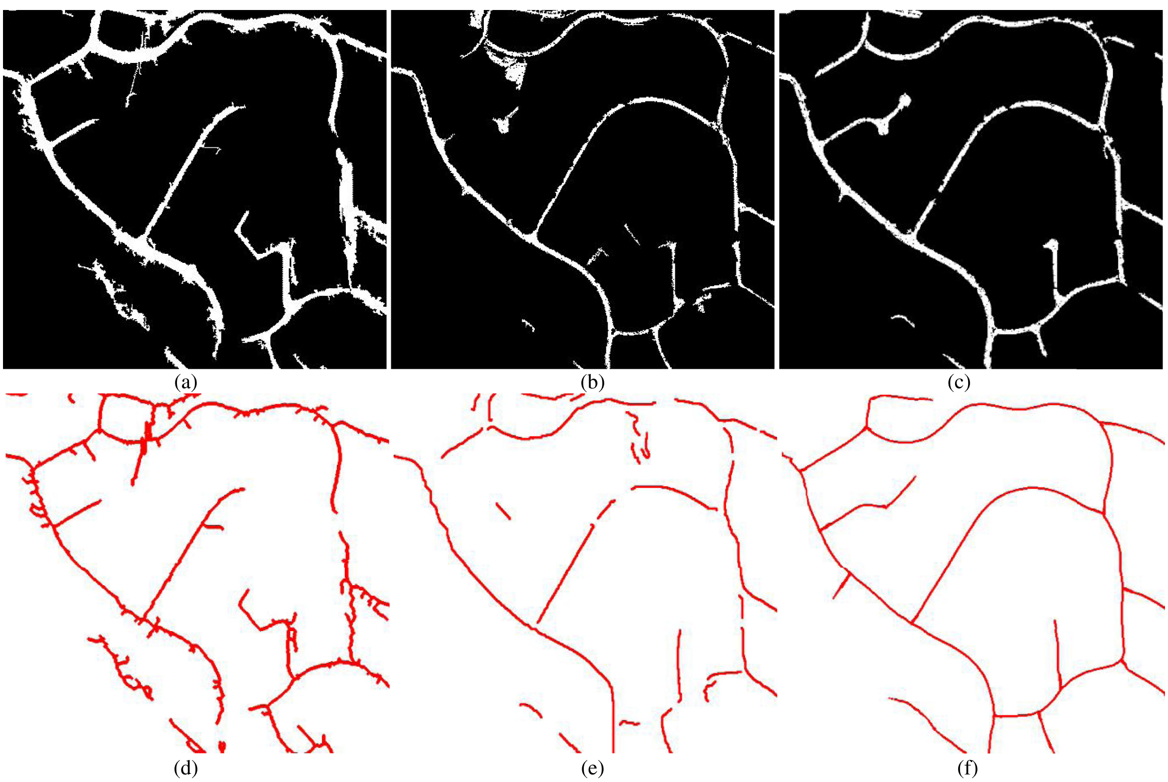


Fig. 5. Result comparison with supervised road extraction method in experiment 2. (a) Road binary image by Huang method. (b) Road binary image by Shi method. (c) Road binary image by our SSLF method. (d) Road centerline by Huang method. (e) Road centerline by Shi method. (f) Road centerline by our SSLF method.

TABLE II
COMPARISON RESULT WITH THE SUPERVISED METHOD (%)

Method	Completeness	Correctness	Detection quality
Our SSLF	**92.05**	**96.14**	**88.24**
Shi	82.23	88.68	82.89
Huang	84.40	64.54	68.47

because the Huang method appears more "spur" resulting in the lower accuracy.

From the above mentioned analysis, the proposed SSLF method gets optimal results both in vision and in quantitative evaluation compared with the traditional supervised road extraction algorithm, which verifies the effectiveness of the proposed algorithm.

C. Comparing With Positive Sample One-Class Classification Algorithm

In order to verify the robustness of the proposed SSLF algorithm, a classical one-class classification algorithm—support

Fig. 6. Beijing-2 image in experiment 3.

TABLE III
COMPARISON RESULT WITH ONE-CLASS METHOD (%)

Method	Completeness	Correctness	Detectionquality
Our PULRF	**72.23**	**90.68**	**76.69**
SVDD	58.41	88.59	60.57

TABLE IV
COMPARISON WITH DEEP LEARNING METHOD USING BENCHMARK DATASET (%)

Method		Completeness	Correctness	Detection quality
CVPR dataset	Our SSLF	36.21	54.00	25.61
	D-LinkNet	**53.90**	**57.08**	**37.00**
Massachusetts dataset	Our SSLF	41.96	**34.55**	22.55
	D-LinkNet	**42.68**	33.88	**22.64**

vector data description (SVDD) [38] —is selected for comparison. The SVDD classifier has achieved satisfactory experimental results in the medium resolution remote sensing image change detection [39]. Since, the SVDD is also suitable for the small sample biased data as our proposed PULRF classifier, the SVDD has high comparability. Beijing-2 image of Tongcheng County, Hubei Province is used and the spatial resolution is 1 m after the multispectral and panchromatic image fusion. The spectral composition is RGB band and the image size is 2693×2409 as shown in Fig. 6.

The SVDD one-class classification algorithm is implemented as LIBSVM open source package [40]. The radial basis function kernel is selected. Parameters C and g are set by the method proposed in [41]. The centerline extraction also uses the tensor voting algorithm. For our SSLF method, the parameters are set as $T1 = 0.90$, $R = 1.0$, $T2 = 0.18$, $P1 = 0.7$. The road centerline parameter is 20. The final experimental results are shown in Fig. 7.

From Fig. 7, it can be seen that both the road binary image and the centerline extracted by our PULRF classifier are superior to that of the SVDD classifier. Although, both SVDD and PULRF have extracted the main road network, the SVDD still produces leak extraction for some small roads, because these small roads are shadowed by trees in long distance. For some complex road intersections, there are some deviations in the centerlines extracted by both SVDD and PULRF. This is because the tensor voting algorithm is easy to over fit for multiple adjacent road areas. The experimental area is located in the complex urban area. There are many vehicles on the road, and the spectral characteristics of different road networks are different. However, our algorithm still extracts most main roads. Therefore, the good robustness of our PULRF classifier is verified. About the speed, SVDD is 1560 s and PULRF is 150 s, which further validates the high efficiency of our PULRF.

The quantitative evaluation of two methods is shown in Table III. The PULRF classifier is superior to the SVDD in all the completeness, correctness, and detection quality. For the correctness, both methods have good results. In the aspect of the completeness, the SVDD method has more leak extraction than the PULRF, because only the label-classification results are output. However, our PULRF classifier is based on the posterior probability.

D. Comparing With Deep Learning Algorithm

In order to compare with the current state-of-the-art baseline method, the deep learning D-LinkNet algorithm [19] in literature is chosen as a contrast. The D-LinkNet is a semantic segmentation neural network, which adopts the encoder–decoder structure, dilated convolution and pretrained encoder for road extraction. To valid the general ability of the proposed method, two benchmark road datasets are considered in this experiment, including the Massachusetts dataset [10] and the DeepGlobe CVPR 2018 dataset [19]. We trained the D-LinkNet on the CVPR dataset and test images in both CVPR dataset and Massachusetts dataset. Our proposed method also tests on both datasets. The optimal parameters of our SSLF are designed as $T1 = 0.85$, $R = 1.0$, $T2 = 0.2$, $P1 = 0.6$, which are the same as that in experiment 2. For the deep learning D-LinkNet method, the initial learning ratio is 0.001, the number of iteration is 60, the input batch size is 8, and the loss function is the cross entropy. In this experiment, the centerline extraction parameter is 25. The representative road extraction results of CVPR and Massachusetts datasets are shown in Figs. 8 and 9, respectively. The accuracy evaluation average value is given in Table IV. The value of the CVPR dataset is the average of 5709 images, and the value of the Massachusetts dataset is the average of 38 images.

From Figs. 8 to 9, in most cases, the results of D-LinkNet are more coherent and comprehensive than that of our SSLF, and vice versa in some cases such as in the case of vegetation shadow occlusion and interference from cars on the road. In another cases, the results of both methods are similar. Especially, the D-LinkNet deep learning method performs very well in the complex urban area. For the accuracy evaluation, the deep learning method has higher values than that of the SSLF method in most cases. The deep learning method is based on the training results of a large number of samples, and in our SSLF, there is no need for the labeled samples (training samples).

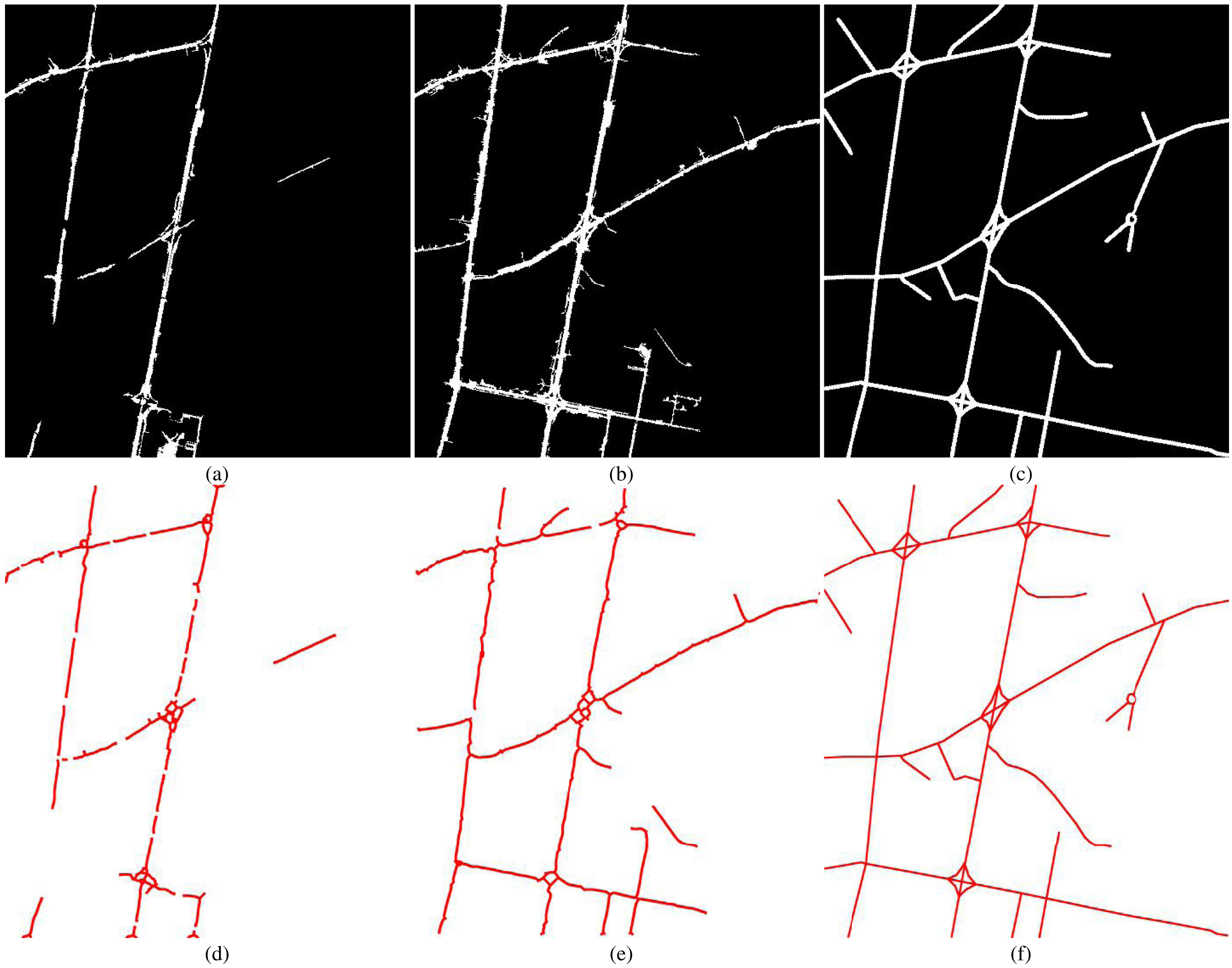


Fig. 7. Result comparison with one-class classification road extraction method in experiment 3. (a) Road binary image by SVDD. (b) Road binary image by our PULRF. (c) Reference road binary image. (d) Road centerline by SVDD. (e) Road centerline by our PULRF. (f) Reference road centerline.

E. Parameter Sensitivity Analysis

In order to verify the influence of experimental parameters on the accuracy of our method, the accuracy effects of parameters T1, T2, P1, and σ are evaluated with experiment 2 as an example. Fig. 10 shows the parameter sensitivity evaluation results.

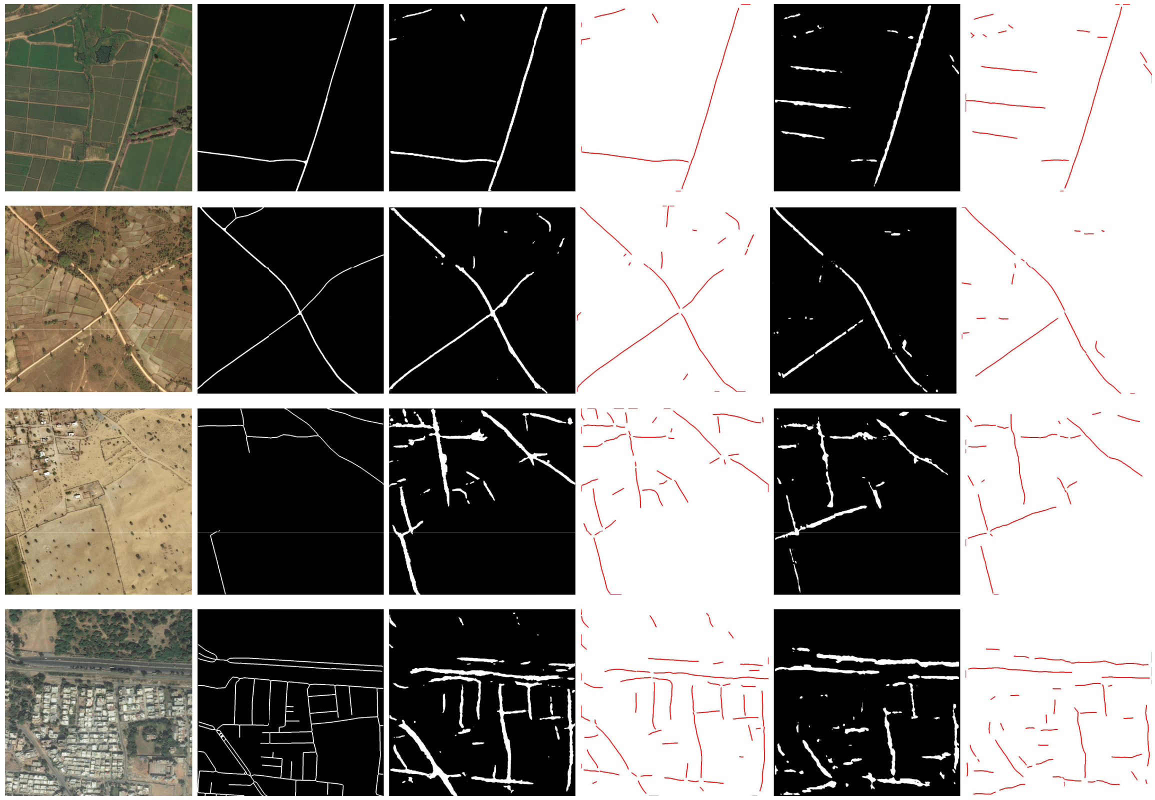
In Fig. 10(a), when the parameter T1 increases higher than the value of 0.85, the completeness, correctness, and the detection quality indicators remained relatively stable trend. When T1 is below 0.85, all the three indicators show a decreasing trend. This is because when the parameter T1 is set larger, more pure and effective road samples are obtained, and when T1 is too low, other ground objects are easily misidentified as road samples, which results in the accuracy reduction. With the increasing of T1, the number of samples is decreasing, but the accuracy is not changed obviously, which indicates that our PULRF classifier achieves good results even in the case of small samples.

For Fig. 10(b), when the parameter T2 is lower than the value of 0.2, the completeness value is basically the same, while the correctness value is decreasing. When T2 is higher than 0.2, the completeness shows a downward trend, but the correctness shows an upward trend. This is because when T2 is low, the road completeness is retained well, but it is easy to introduce other nonroad areas, which leads to the decreased correctness. Meanwhile, when T2 is set too large, other nonroad areas are effectively reduced, but it is easy to produce some fractures resulting in the low completeness.

As can be seen from Fig. 10(c), the effect of parameter P1 on the accuracy is similar to that of parameter T2 and the reason causing this phenomenon is also similar. About Fig. 10(d), when the parameter σ is low, the accuracy decreases. This is because low σ produces other unnecessary "spur" phenomenon. However, too high σcauses the over-fitting deviation especially for the ring road area.

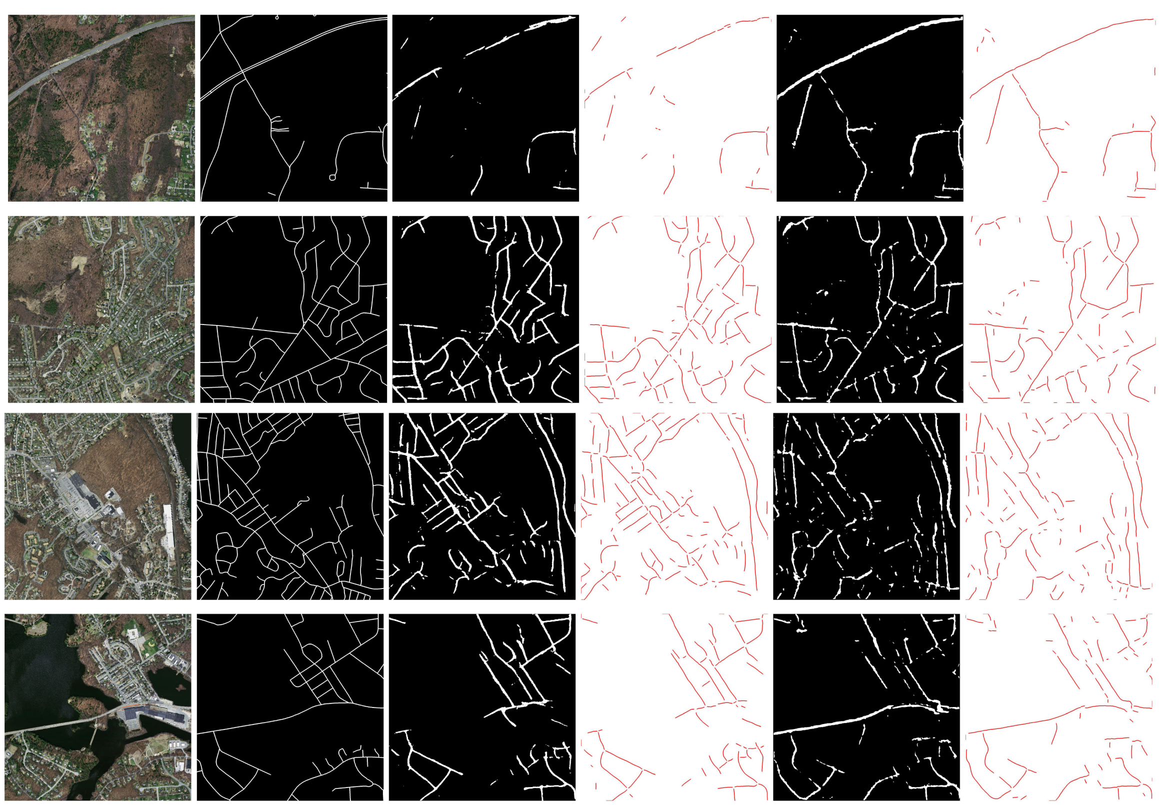
Since samples are almost automatically obtained in our algorithm, in order to get more efficient samples, the sample selection parameter T1 is set higher. For the final road network extraction, the road acquisition parameter T2 is set lower to extract more road pixels.

In summary, from the experiments comparing with the unsupervised and the supervised road extraction methods, our SSLF method achieves high accurate results both in the qualitative and quantitative evaluation, which verifies the effectiveness of the proposed algorithm. From the parameter analysis, our SSLF



(a) (b) (c) (d) (e) (f)

Fig. 8. Result comparison with deep learning road extraction algorithm using the CVPR dataset in experiment 4. (a) Images in CVPR dataset. (b) Reference centerline. (c) D-LinkNet binary. (d) D-LinkNet centerline. (e) Our SSLF binary. (f) Our SSLF centerline.



(a) (b) (c) (d) (e) (f)

Fig. 9. Result comparison with deep learning road extraction algorithm using the Massachusetts dataset in experiment 4. (a) Images in Massachusetts dataset. (b) Reference centerline. (c) D-LinkNet binary. (d) D-LinkNet centerline. (e) SSLF binary. (f) SSLF centerline.

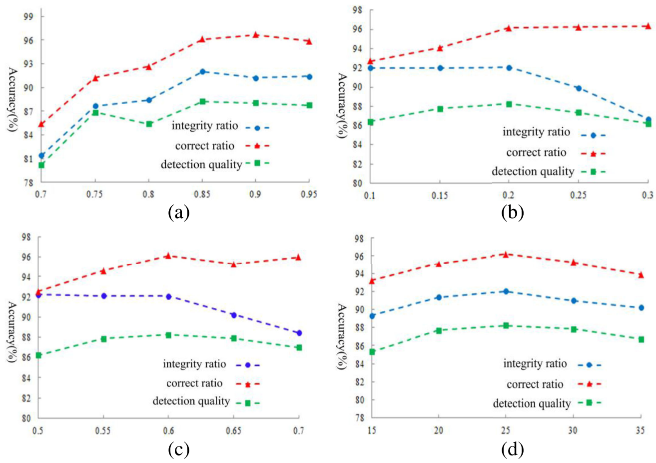


Fig. 10. Accuracy effects of parameters. (a) Accuracy effect of parameter T1. (b) Accuracy effect of parameter T2. (c) Accuracy effect of parameter P1. (d) Accuracy effect of parameter σ.

algorithm requires some adjustment of the experimental parameters. But, from the perspective of algorithm implementation, the automation degree of our method is higher than that of the supervised road extraction algorithm. Moreover, the noise resistance of SSLF is superior to that of the unsupervised algorithm. Furthermore, a classical one-class classifier SVDD is compared with our PULRF classifier, which shows the PULRF has good robustness and high efficiency. Compared with the deep learning method, although the SSLF does not get high accuracy, the interpretable SSLF method does not require high hardware support, long training time, and a huge number of samples. It is believed that our SSLF algorithm can give a revelation to other information extraction algorithm.

IV. Conclusion

In this article, a road centerline extraction algorithm is proposed under the self-supervised learning framework. Aiming at the problem that the existed supervised algorithm needs to extract the training samples manually, an automatic positive sample acquisition is proposed to decrease the manual intervention from two aspects of the spectral feature and the shape feature. Differing from the two-class or multiclass classifiers widely used in the existing road extraction algorithms, the one-class classifier is introduced. A new positive and unlabeled learning of RF one-class classifier is constructed and the effectiveness of this classifier is proved by experiments. Moreover, the final road area is directly extracted from the combination of the posterior probability and the shape feature. The automation degree of our algorithm is improved much without further step-by-step optimization for filtering the nonroad areas. The proposed one-class self-supervised learning framework is probably a new promising starting point in the information extraction field.

References

[1] W. Z. Shi, C. Q. Zhu, and Y. Wang, "Road feature extraction from remotely sensed image: Review and prospects," *Acta Geodaetica Et Cartographica Sinica*, vol. 30, no. 3, pp. 257–262, 2001.

[2] J. Mena, "State of the art on automatic road extraction for GIS update: a novel classification," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3037–3058, 2003.

[3] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1538–1544, 2012.

[4] C. Unsalan and B. Sirmacek, "Road network detection using probabilistic and graph theoretical methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4441–4453, Oct. 2012.

[5] I. Coulibaly, N. Spiric, M. O. Sghaier, W. Manzo-Vargas, R. Lepage, and M. St-Jacques, "Road extraction from high resolution remote sensing image using multiresolution in case of major disaster," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2014, pp. 2712–2715.

[6] Z. Shen, J. Luo, and L. Gao, "Road extraction from high-resolution remotely sensed panchromatic image in different research scales," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 453–456.

[7] M. O. Sghaier and R. Lepage, "Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1946–1958, May 2016.

[8] J. Liu, Q. Qin, J. Li, and Y. Li, "Rural road extraction from high-resolution remote sensing images based on geometric feature inference," *ISPRS Int. J. Geo Inf.*, vol. 6, no. 10, pp. 314–337, 2017.

[9] S. Das, T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3906–3931, Oct. 2011.

[10] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: preliminary results, "in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1591–1594.

[11] 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.

[12] X. Lu *et al.*, "Multi-scale and multi-task deep learning framework for automatic road extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9362–9377, Nov. 2019.

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

[14] G. Fu, H. Zhao, C. Li, and L. Shi, "A method by improved circular projection matching of tracking twisty road from remote sensing imagery," *Acta Geodaetica Et Cartographica Sinica*, vol. 43, no.7, pp. 724–730, 2014.

[15] H. Mayer, I. Laptev, and A. Baumgartner, "Multi-scale and snakes for automatic road extraction," in *Proc.Eur. Conf. Comput. Vis.*, 1998, pp. 720–733.

[16] X. Huang and L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines," *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 1977–1987, 2009.

[17] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3359–3372, Jun. 2014.

[18] Z. Hong, D. Ming, K. Zhou, Y. Guo, and T. Lu, "Road extraction from a high spatial resolution remote sensing image based on richer convolutional features," *IEEE Access*, vol. 6, pp. 46988–47000, 2018.

[19] 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.

[20] H. He, D. Yang, S. Wang, S. Wang, and Y. Li, "Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss," *Remote Sens.*,vol. 11, 2019, Art. no. 1015.

[21] V. Morard, P. Dokladal, and E. Decenciere, "Parsimonious path openings and closings," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1543–1555, Apr. 2014.

[22] Z. Miao, W. Shi, H. Zhang, and X. Wang, "Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 583–587, May 2013.

[23] R. Liu, J. Song, Q. Miao, P. Xu, and Q. Xue, "Road centerlines extraction from high resolution images based on an improved directional segmentation and road probability," *Neurocomputing*, vol. 212, no. C, pp. 88–95, 2016.

[24] Cao Y., Z. Wang, L. Shen, X. Xiao, and L. Yang, "Fusion of pixel-based and object-based features for road centerline extraction from high-resolution

satellite imagery," *Acta Geodaetica Et Cartographica Sinica*., vol. 45, no. 10, pp. 1231–1240, 2016.
[25] P. Soni, N. Rajpal, and R. Mehta, "Semiautomatic road extraction framework based on shape features and LS-SVM from high-resolution images," *J. Indian Soc. Remote Sens.*, vol. 48, no. 3, pp. 513–524, 2020.
[26] C. Senaras and F., Vural, "A self-supervised decision fusion framework for building detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1780–1791, May 2016.
[27] J. A. Hartigan and M. A.Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. Roy. Statist. Soc., Ser. C (Appl. Statist.)*, vol. 28, no. 1, pp. 100–108, 1979.
[28] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[29] Z. Miao, W. Shi, P. Gamba, and Z. Li, "An object-based method for road network extraction in VHR satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4853–4862, Oct. 2016.
[30] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, vol. 58, no. 1, pp. 213–220.
[31] X. Chen, D. Yin, J. Chen, and X. Cao, "Effect of training strategy for positive and unlabelled learning classification: Test on Landsat imagery," *Remote Sens. Lett.*, vol. 7, no. 11, pp. 1063–1072, 2016.
[32] M. Belgiu, and L. Dragut, "Random forest in remote sensing: A review of applications and future directions." *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.
[33] A. Bosch, A. Zisserman, X. Munoz, "Image classification using random forests and ferns," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[34] M. Fauvel, J. Chanussot, and J. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.
[35] G. Li and Y. Wang, "Synthesis classification of remote sensing image based on improved pixel-Level and object-level methods," *Acta Geodaetica Et Cartographica Sinica*, vol. 41, no. 6, pp. 891–897, 2012.
[36] G. Medioni, C. K. Tang, and M. S. Lee, "Tensor voting: Theory and applications" in *Proc. RFIA*, Paris, France, 2000, pp. 1482–1495.
[37] C. Heipke, H. Mayer, C. Wiedemann, and O. Jamet, "Evaluation of automatic road extraction," *Int. Archives Photogrammetry Remote Sens.*, vol. 32, no. 3, pp. 151–160, 1997.
[38] J. Tax, E. Ypma, and R. Duin, "Support Vector Data Description applied to machine vibration analysis," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 1999.
[39] Z. Xu, G. Mountrakis, and L. J. Quackenbush, "Impervious surface extraction in imbalanced datasets: Integrating partial results and multi-temporal information in an iterative one-class classifier," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 43–63, 2017.
[40] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[41] M. Wu and J. Ye, "A small sphere and large margin approach for novelty detection using training data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2088–2092, Nov. 2009.

Qing Guo (Member, IEEE) received the M.Sc. and Ph.D. degrees in optics from the Harbin Institute of Technology, Harbin, China, in 2006 and 2010, respectively.

She joined the Chinese Academy of Sciences in 2010. She is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. From 2007 to 2009, she was an exchange Ph.D. student with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB, Canada. From 2014 to 2015, she is a Visiting Scholar with the Institute for Geoinformatics and Remote Sensing, University of Osnabrück, Osnabrück, Germany. Her research interests include the remote sensing information extraction and processing, including the road extraction, the image registration and fusion.

Zhipan Wang received the M.Sc. degree in photogrammetry and remote sensing from the Southwest Jiaotong University, Chengdu, China, in 2017.

He was a Visiting Master Student with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. He is currently with the School of Urban-Rural Development, Shaoyang University, Shaoyang, China. His research interests include road extraction, computer vision, and remote sensing image processing.