IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 18, 2018, accepted August 22, 2018, date of publication August 27, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2867210

Road Extraction From a High Spatial Resolution Remote Sensing Image Based on Richer Convolutional Features

ZHAOLI HONG, DONGPING MING, KEQI ZHOU, YA GUO, AND TINGTING LU
School of Information Engineering, China University of Geosciences, Beijing 10083, China

Corresponding author: Dongping Ming (mingdp@cugb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41671369, in part by the Fundamental Research Funds for the Central Universities, and in part by the National Key Research and Development Program under Grant 2017YFB0503600.

ABSTRACT The extraction and vectorization of roads from high spatial resolution remote sensing (HSRRS) images are of great significance to city planning and development. However, significant as they are, it is usually an arduous task to put them into practice because the HSRRS images are often filled with complex ground information. Furthermore, extracted roads may suffer from netsplit or brokenness. This paper thus proposes Richer convolutional features (RCFs)-based road extraction (Road-RCF) as a method which targets these issues. A modified roads sample set and RCF network are applied to generate road probabilities in order to extract initial road information. After the road centerlines extraction by the refinement algorithm, vectorized roads are ultimately extracted. The compared experiment results show that the Road-RCF method produce better road extraction results than the other four state-of-the-art methods, in both quantitative road extraction accuracy metrics and the qualitative visual evaluation. The benefits of this model are threefold. First, the image-to-image network structure of side-output realizes multi-scale and multi-level road feature fusion in order to make a full use of the information from a low level to a high level. Second, according to the deep supervision of the side-output, it guides the learning of the correct road information. Third, after the detection of the road, the road centerlines are vectorized to facilitate the attribute information management and electronic map production. In a word, the proposed Road-RCF method is both practical and meaningful toward updating the geo-information system database.

INDEX TERMS High spatial resolution remote sensing images, Richer convolutional features, road detection, road centerlines extraction and vectorization.

I. INTRODUCTION

As one of the major components of a city, urban roads play an important role in the formation and development of that city. The vectorized road can integrate some non-spatial attribute information into the spatial database, which can effectively model the traffic information and further aid traffic management. Traditional electronic maps and urban roads traffic networks are mainly produced based on vectorized human-informed paper maps and updated based on the manual digitization of high spatial resolution remote sensing (HSRRS) images, which involves large workloads and has long production cycles. Therefore, it is difficult to effectively maintain the accuracy and real-time capability of traffic maps. The rapid development of remote sensing technology makes it possible to use the road information extracted from HSRRS images in the urban geographic databases updating, urban transportation planning [1], citizen tourism planning, urban mapping techniques [2], etc. However, there are a large number of mixed pixels in HSRRS image, which makes the boundaries between roads and other objects in the image data unclear. In the image data, urban roads with rich spectral information can easily produce incorrect border information. The complex features presented in HSRRS images (such as green belts, shadows and road making line.) are liable to cause the poor geometric accuracy of road information extraction results. Consequently, it is always a hard task to accurately extract roads from HSRRS images.

Road extraction from remote sensing images is a hot topic with a rich research history. The methodological system related to road extraction was summarized in Fig. 1. As shown

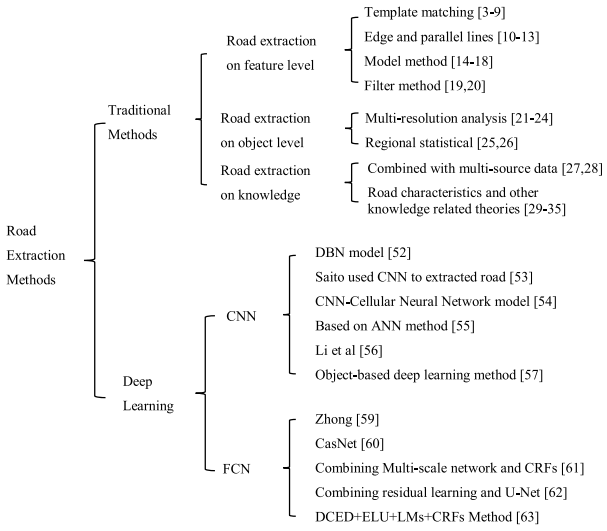

FIGURE 1. **Main methods of road extraction.**

in Fig. 1, traditional works on this topic can be generally summarized into three groups, feature level, object level and knowledge level.

(1) Road extraction on the feature level, includes the template matching [3]–[9], the edge and parallel lines [10]–[13], the model method [14]–[18] and the filter method [19], [20]. The template matching method extracts the roads from the image according to extracted seed pixels or a specific template to form an initial road network. The edge and parallel lines use the trait that the road edges are usually parallel lines. For example, Unsalan and Sirmacek [12] extracted the initial edge of the road and then used the Binary Ballon algorithm and graph theory to extract roads. The model method extracted the road network based on the snake model [14], [15], the Markov model [16], and the like. The filter method obtains the road network based on enhanced road pixels using a specific filter. However, these methods lead to the ''salt and pepper'' phenomenon and poor extraction accuracy, especially in complex scenes.

(2) Road extraction on the object level, includes multi-resolution analysis [21]–[24] and regional statistical analysis [25], [26]. The multi-resolution analysis method improves the precision of road extraction by combining different resolutions of remote sensing images or single image at different scales. Regional statistical methods, such as Yi *et al*. [25], initially segmented the image into objects, and then built a ''word-theme'' model to extract the road network. Region-based road extraction algorithms have achieved some progresses in complex urban environments, but the initial segmentation or clustering of images leads to the phenomenon of ''adhesion''.

(3) Road extraction on the knowledge level, such as the multi-source data fusion [27], [28], the road characteristics and the other knowledge related theories [29]–[35]. The multi-source data fusion method guides or assists the extraction of road networks based on existing road databases or other data, such as vector maps and documents. The road characteristics and other knowledge related theories extract roads based on their own characteristics, such as spectrum and context features. These methods also have made advances in complex scenes [36], but the design of such algorithms is more complicated and the operating efficiency is not ideal.

In 2006, Hinton *et al*. [37] proposed the concept of deep learning (DL), which provides the basis for the revival of DL. After that, deep learning techniques are also widely used in the field of remote sensing [38], like scene classification [39]–[42], object detection [43]–[46] and image retrieval [47], [48]. Recently, researchers have attempted to adopt DL to extract roads from HSRRS images. Two kinds of networks have been involved in this topic, the Convolutional Neural Network (CNN) and the Fully Convolutional Neural (FCN) [49], and Fig. 1 summarizes methods for road extraction based on the CNN and the FCN.

Unlike traditional algorithms that just use low-level information for road extraction, the CNN can reduce false detections by embedding many high-level [50], [51] and multi-scale information. Starting with Mnih and Hinton [52] (2010), the Deep Belief Network (DBN) model has been used to detect roads in airborne remote sensing images. Saito *et al*. [53] employed the CNN to extract buildings and roads directly from raw remote sensing images. Sarhan *et al*. [54] used cellular networks to extract roads from images, and proposed a framework called the ''CNN-Cellular Neural Network'', which makes full use of the spectral and geometric characteristics of roads in remote sensing images. Based on the adaptive artificial neural network (ANN) method, a self-learning supervised learning neural network was trained and applied to road extraction from the WorldView-II satellite's panchromatic images by Wijesingha [55]. Li *et al*. [56] used the CNN to judge whether a pixel belongs to a road, then smoothed it through post-processing and obtained the centerline of a road. Zhao *et al*. [57] proposed an object-based deep learning method to accurately extract roads from HSRRS images. Although the CNN has achieved certain results in road extraction, the local processing strategy still yields many errors in the extraction results. For example, roads extracted from images of complex features still have patches and incomplete fragments.

Compared with CNN, the features extracted from FCN network are high-level features that contain more abstract semantic information [58]. Zhong *et al.* [59] used the FCN to extract roads and buildings from the Massachusetts dataset and achieved acceptable results. Cheng *et al*. [60] proposed a cascaded end-to-end convolutional neural network (CasNet) which contains two convolutional neural networks: a road detection network and a centerline extraction network. Fu *et al*. [61] designed a multi-scale network based on the FCN and adopted Conditional Random Fields (CRFs) to refine output class maps, which could get better road classification results on GF-2 and IKONOS images. In [62],

a semantic segmentation neural network combining residual learning and U-Net strength is proposed for road area extraction, which can ease the training of deep networks and maintain networks with fewer parameters. Panboonyuen *et al*. [63] replaced the activation function with an exponential linear unit (ELU) in a deep convolutional encoder-decoder network (DCED). Then, they added landscape metrics (LMs) and CRFs to extract roads more accurately. These network structures operate on the entire image to facilitate the acquisition of high-level information. However, only the information output from the last convolutional layer can be used. As a result, the road information between each layer cannot be fully acquired and used.

The networks mentioned above are only for end-to-end and pixel-to-pixel training, which can't fully acquire the abstract information on each convolution layer. Liu et al. [64] firstly proposed the Richer Convolutional Features (RCF) network which developed from the Holistically-Nested Edge Detection (HED) network [65]. The RCF model changes the locally handled way of the traditional network. Based on the training and prediction of the entire image, high-level semantic information can be obtained. It is worth mentioning that deep supervision helps side-output layers to produce multi-scale density predictions and a fusion output, which can make full use of the complementary information between different convolution layers. Up to present, the RCF has been trained and tested on the Berkeley Segmentation Dataset and Benchmark (BSDS500), the NUY Depth Dataset (NYUD), as well as the Multicue Datasets, and all achieves good accuracy.

As is known to all, although natural scene belongs to different categories, they may resemble each other in many respects [66]. However, HSRRS images include various types of objects with different sizes, colors, rotations, and locations in a single scene, which results in high complexity of remote sensing images and high difficulty of road extraction. In this paper, we not only want to get a smooth and complete road network map with the consideration of high-level semantic information, but also want the road network map to contain more details to describe the boundary information. Thus, we develop a side-output image-to-image fusion and deep supervision road detection system by using Richer Convolutional Features based road extraction (Road-RCF) method. Richer multi-scale road features are learned in the side-output layer, which combines the local information of the road with the high-level semantic information to reduce the influence of occlusion and shadow. Manually labeling road samples makes the deep supervision of each layer of network in reality, thus achieving optimal fitting of road information at different scales and enhancing the saliency-guided road feature learning. The paper is organized as follows: Section 2 describes the proposed methodology. Section 3 presents the processed datasets and experimental results. Discussions are shown in Section 4. The conclusions are demonstrated in Section 5.

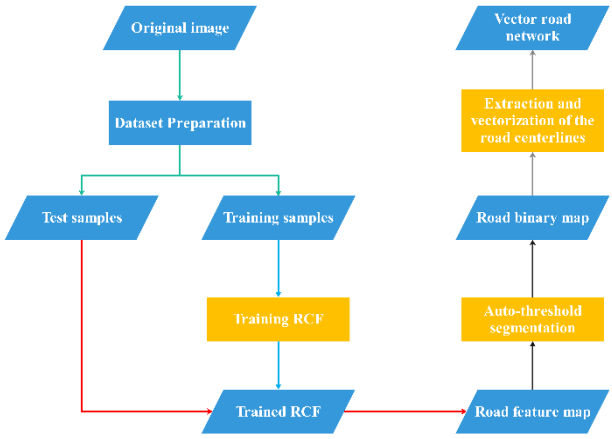

FIGURE 2. Framework of Road-RCF model.

II. METHODOLOGY

This paper proposes a Road-RCF model for road information extraction from HSRRS images. As shown in Fig. 2, this model is mainly composed of the following four parts: (i) dataset preprocessing is performed on the original image to generate training samples and testing samples; (ii) training samples are used to train the RCF network, and then the trained RCF network is performed on input test samples to generate a rough road feature map; (iii) non-road information is eliminated with a low grayscale value in the feature map by auto-threshold segmentation, which generates a road binary map; (iv) extraction and vectorization of the road centerlines are performed to obtain a complete vector road network.

A. DATASET PREPARATION

The production and processing of datasets are essential for the training of the network and the final prediction.

Unlike natural pictures, HSRRS images are not only large, but also occupy a large storage space, which requires high performance hardware environment. To ease processing, it is preferable to separate a HSRRS image into several suitable sized sub-images.

Manual road labeling is then performed to use software such as ArcGIS or some related algorithms. In the process of labeling, it should be noted that the size of the groundtruth must be the same as the size of the original image, and the position of the road in the groundtruth is consistent with that of the road in the original image. Every sample data is composed of two types of image: the image (Hight*Width*Channel) and the labeled groundtruth raster image (Hight*Width*{0,1}). The sample data is divided into the training set and testing set in accordance with the ratio of 20:1.

Deep learning usually requires a lot of data, and it is difficult to make larger datasets by manual labeling. Data augmentation has been proved to be a crucial technique in deep network. Data augmentation can effectively avoid

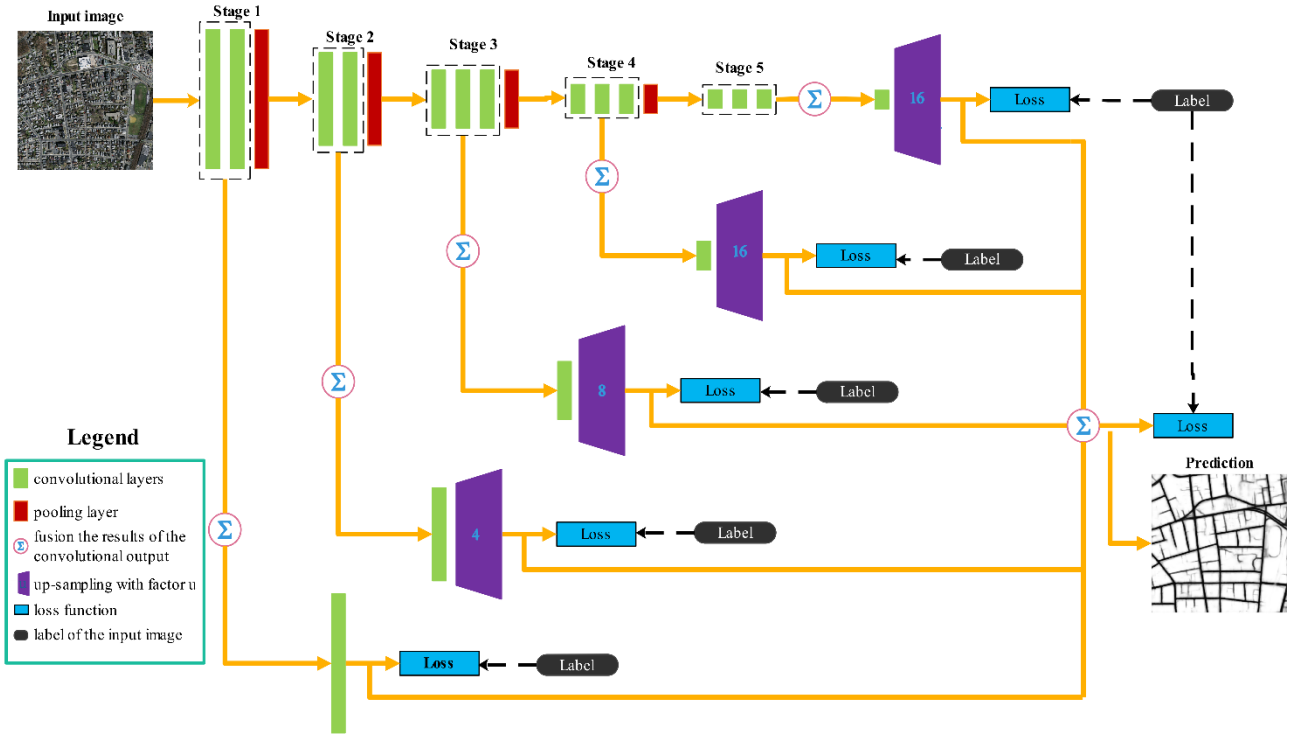

FIGURE 3. Road-RCF network architecture.

over-fitting problems and improve the accuracy of road extraction. In fact, there are many ways for data augmentation, such as rotation, random scale, color jittering and deformation. This paper utilizes the method of rotation at different angles for data augmentation.

B. Road-RCF NETWORK

The proposed Road-RCF model of road information extraction benefits from the recent success of the RCF network. The design of the RCF network is based on the modification of the VGG16 [67] network. An advantage of the RCF-based road detection method is that it uses the side-output image-to-image and deep supervision network architecture. In this architecture, the side-output layer is added to incorporate the feature responses from different levels of the primary network stream. Also, deep supervision uses the label (groundtruth) to guide the correct side-output of the road information. Our Road-RCF network architecture is illustrated in Fig. 3.

1) NETWORK STRUCTURE OF VGG16

The structure of the VGG16 network is really simple. Nevertheless, it is actually an ideal structure for image processing, such as image classification and target positioning. As demonstrated in Fig. 4, VGG 16 contains 5 stages, 3 fully connected layers and other basic components (activation and dropout layers).

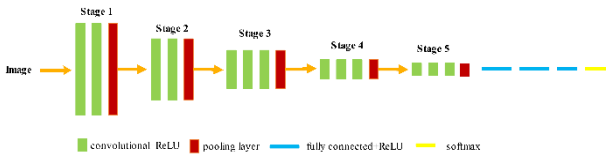

FIGURE 4. Illustration of the architecture of VGG 16.

Each stage has 2 or 3 convolutional layers and is connected with a maximum pooling layer. The entire network has the same size convolution kernel (3×3) and pool kernel (2×2).

Oftentimes, multiple identical 3×3 convolution layers which are stacked together make more non-linear transformations available which can enhance the network's ability to learn the features. As the receptive field size increases, useful information captured by the convolutional layers becomes increasingly rough. The receptive field is an area where the region mappings on the original image according to the pixels on the feature map output by each layer of the convolutional neural network. The detailed receptive field sizes of different layers can be seen in Table 1 in detail.

2) SIDE-OUTPUT IMAGE-TO-IMAGE

The Road-RCF network discards the fully connected layer, and uses the deconvolutional layer for upsampling to restore the original image size. This strategy enables input images to

TABLE 1. Detailed receptive field sizes of different layers.

	c1_1	c1_2	p1	c2_1	c2_2	p2	c3_1	c3_2	c3_3
rf	3	5	6	10	14	16	24	32	40
st	1	1	2	2	2	4	4	4	4

	p3	c4_1	c4_2	c4_3	p4	c5_1	c5_2	c5_3	p5
rf	44	60	76	92	100	132	164	196	212
st	8	8	8	8	16	16	16	16	32

rf: receptive field size; st: stride; c1_1: the first convolution layer in the stage 1, and p1: pool layer of stage 1.

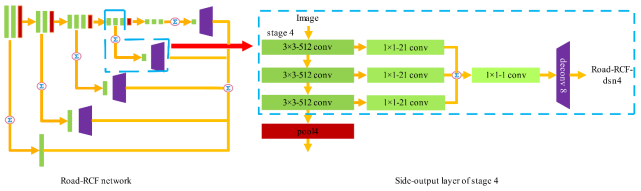

FIGURE 5. Illustration of side-output image-to-image for side-output feature map generation.

be flexible in size, and finally maintains the corresponding size of the classification image.

The CNN autonomously learns features of multiple levels through the convolutional layer and the pooling layer. The hidden network with smaller receptive field can learn some local information of objects in the image. However, as the number of layers increases, the receptive field also becomes larger, which leads to a higher level of information output. Unlike the CNN, the Road-RCF takes full advantage of the complementary information between different convolutional layers to obtain more accurate results for road extraction. The skip-layer connection provides the important ability to the Road-RCF with the important ability to use features at different layers for extracting information. After each convolution of conv1_2, conv2_2, conv3_3, conv4_3, and conv5_3, the VGG16 performs a pooling operation respectively. As shown in Fig. 5, the Road-RCF network removes the pooling layer 5 and the fully connected layer, and then connects each convolutional layer with an output layer which includes the convolution kernel 1×1-21. The features learned by each layer are preserved. Finally, the resulting layers in each stage are accumulated by an eltwise layer to attain the hybrid features.

In the proposed model, 5 stages are involved in side-output image-to-image. In stage 1, a 1×1-1 side-output layer follows each eltwise layer, and then the low-level road feature map is acquired. In stage 2, the original image is reduced to 1/2 after a pooling operation, thus the Road-RCF-dsn2 is derived from the second convolution and output of the image. At this point, the Road-RCF-dsn2 is half the size of the original image. Then, the feature map is upsampled by using a deconvolution layer based on bilinear interpolation with a kernel_size of 4 and a stride of 2. According to different convolution and pooling layers, the parameters of kernel_size and stride are

TABLE 2. The parameters kenel_size and stride of the deconvolution layer in the Road-RCF.

	stage 1	stage 2	stage 3	stage 4	stage 5
kenel_size		4	8	16	16
stride		2	4	8	8

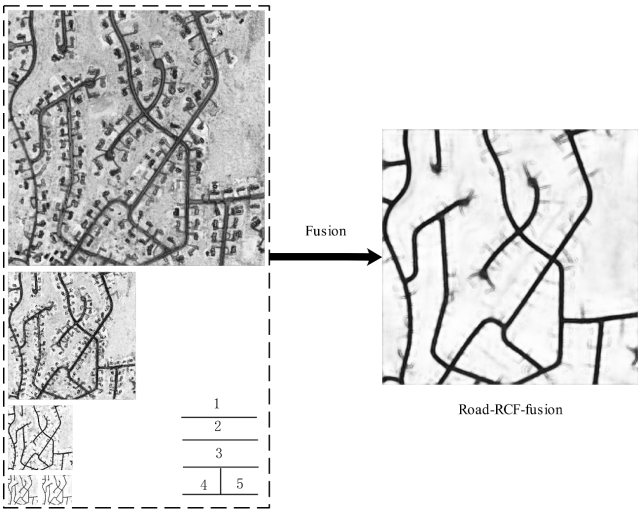

FIGURE 6. Side-outputs of multi-scale road feature in every stage of Road-RCF and fusion.

set for the deconvolution layer of stage 3, stage 4 and stage 5 respectively, as listed in Table 2.

Fig. 6 illustrates the exemplary results of side-output and the multi-scale fusion. As shown in Fig. 6, the shallow layers (Road-RCF-dsn1, Road-RCF-dsn2, and Road-RCF-dsn3) capture more spatial details, while they lack sufficient semantic information. In contrast, the deeper layers (Road-RCF-dsn4 and Road-RCF-dsn5) encode richer semantic information, but spatial details are missing. Finally, the concat layer fuse the side-output road feature map to generate the prediction.

3) DEEP SUPERVISION

The Road-RCF network inserts a side-output layer into each convolutional layer to obtain multi-scale information. In the meantime, the Road-RCF network joins the deep supervision that in embedded in each side-output layer. The traditional FCN performs weighted-fusion supervision. For example, the [49] proposed two skip strategies for object segmentation: FCN-16s and FCN-8s. They mapped the groundtruth on the final fusion layer and adjusted network parameters by calculating loss values of the fused image. The deep supervision showed in Fig. 7 is different from the weighted-fusion supervision. The deep supervision is used to guide image output of side-output layer which obtains corresponding road feature at different scales. Fig. 8 shows comparison of road feature extracted in stage 1 where two situations are tested: deep supervision and without deep supervision. We observe that

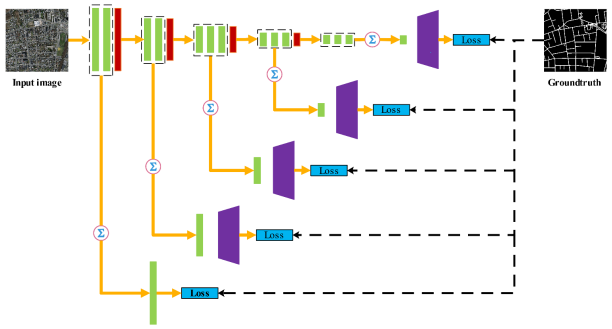

FIGURE 7. Illustration of deep supervision for guiding the side-outputs towards road predictions.

FIGURE 8. Comparison of road feature extracted in stage 1 with or without deep supervision. (a) Groundtruth. (b) Saliency road features are obtained through deep supervision. (c) Without deep supervision.

with deep supervision, the road feature is more obvious in stage1.

C. AUTOMATIC THRESHOLD SEGMENTATION

Automatic threshold segmentation can automatically convert a grayscale image into a binary image according to the distribution of grayscale values. Through the network, a road feature map which is (grayscale map) full of ''shadows'' is predicted based on the datasets. The reason for the occurrence of shadows is that the spatial resolution of the image features gradually decreases as the pooling operation and receptive field of the network increase. Thus, the linear classifier in the fully convolution structure produces similar responses in adjacent pixels, which generates of fuzzy edges. Furthermore, when the images are finally merged and output, they are affected by the characteristics of low-level network information, and results in some fuzzy ''shadows''. We use the largest inter-class variance method to enhance the road region with a large gray value while eliminating the ''shadows'' with a smaller gray value. The pixels in the graph are divided into two types: w0 and w1, by a threshold T. Pixels with gray values in $[0, T-1]$ marked with w_0, and pixels with gray value in $[T, L-1]$ are marked with w_1. The average gray values of areas w_0 and w_1 are

$$\mu_0 = \frac{1}{P}\sum_{i=0}^{T-1} ip_i = \frac{\mu(T)}{P_0} \tag{1}$$

Algorithm 1 Calculation ***Z0*** (***P1***)

```
nCount = 0
if (P2==0 && P3==1) nCount ++;
if ( P3==0 && P4==1) nCount ++;
if (P4==0 &&  P5==1) nCount ++;
if (P5==0 && P6==1) nCount ++;
if (P6==0 && P7==1) nCount ++;
if (P7==0 &&  P8==1) nCount ++;
if (P8==0 && P9==1) nCount ++;
if (P9==0 && P2==1) nCount ++;
Z0 (P1) = nCount
```

$$\mu_1 = \frac{1}{P}\sum_{i=T}^{L-1} ip_i = \frac{\mu - \mu(T)}{1 - P_0} \tag{2}$$

where, μ_0 and μ_1 are respectively the average gray values of areas w_0 and w_1, P_0 and P_1 respectively represent the probabilities of regions w_0 and w_1, and P_i represents the probability of occurrence of gray values i.

Then, the average gray level of the entire image μ is computed as

$$\mu = \sum_{i=0}^{T-1} iP_i + \sum_{i=T}^{L-1} iP_i = P_0\mu_0 + P_1\mu_1 \tag{3}$$

The total variance of the two regions is

$$\sigma_B^2 = P_0(\mu_0 - \mu)^2 + P_1(\mu_1 - \mu)^2 = P_0P_1(\mu_0 - \mu_1)^2 \tag{4}$$

Let T be in the range of $[0, L-1]$ in order so that the maximum T value of σ_B^2 is the optimal segmentation threshold.

D. ROAD CENTERLINE EXTRACTION AND VECTORIZATION

In general, centerline extraction of roads highlights the roads' shape characteristics and topology that can reduce the amount of redundant information. In large-scale navigation maps, spatial and attribute information of roads can be clearly displayed. This refinement algorithm has two advantages: there is no glitch on the extracted centerline, and the centerline of the extracted road is relatively smooth.

In the process of extracting the centerline of the road. As shown in Fig. 9, 9 points in 3×3 the area are marked as *P1*, *P2*, ... *P9*. Here, it is specified that 1 represents black and 0 represents white. In the case of the center point $P1 = 1$, rules can be formulated as follows.

(1) $2 \le NZ\ (P1) \le 6$, where $NZ(P)$ represents the number of 1s in the 8 fields of P points.

(2) $Z0\ (P1) = 1$.

(3) $P2 \times P4 \times P8 = 0$ or $Z0(P1) \ne 1$.

(4) $P2 \times P4 \times P6 = 0$ or $Z0(P4) \ne 1$.

If the above four rules are satisfied, delete *P1* (that is, let *P1*=0). Repeat steps for each point in the image until no further points can be deleted.

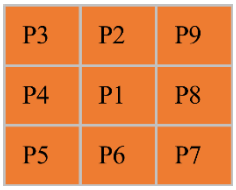

FIGURE 9. 3×3 neighborhoods.

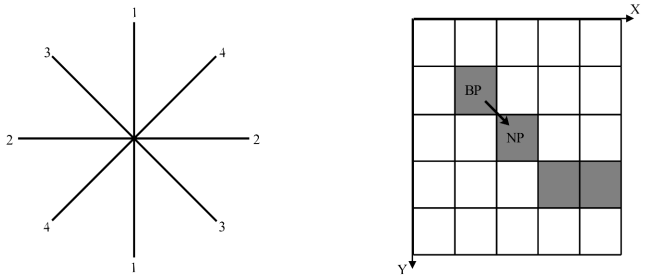

FIGURE 10. The process of the auto-tracking-vectorization algorithm. (a) The current feature point search direction. (b) The vectorization in the BP-NP direction, where BP represents the previous feature pixel and NP represents the current feature point.

The vectorization of refined roads in grid format realizes through an auto-tracking-vectorization algorithm [68]. As shown in Fig. 10, all pixels except for the intersection pixels that are marked as current feature pixel (NP) are set as 0. The basic steps of the algorithm are as following.

(1) Search for endpoint pixels. If the last pixel of the image has been found, perform step (4). Otherwise, it is marked as the NP. Then, the line number increases by 1, and the pixel is converted into a vector point. Eventually, the point number is stored in the current line.

(2) In eight-neighbor field (as shown in Fig. 10(a)) whose center is the NP, reach next refined pixel and set it as the new NP. Meanwhile, the center pixel which is the previous NP is set as previous feature pixel (BP) (as shown in Fig. 10(b)).

(3) If the pixel being searched is a non-endpoint feature pixel, it is converted into a vector point and the point number shall be stored in the current line. Then, step (2) is repeated. The pixel shall be directly converted into a vector point and point number should be saved in the current line before step (1) is performed. If the pixel is an intersection point, it should be treated as an ordinary feature point, and the intersection point should be stored in the intersection table. The search for the next pixel starting from the direction of the BP is performed and point to the intersection point to ensure that the intersection point is not broken.

(4) Based on the intersection table, the intersection point is transferred as the starting point. The tracking method is the same as above, and once the tracking is completed, a reverse tracking shall be operated from the same intersection point. After all the intersections are tracked, the vector data file of the road will be output.

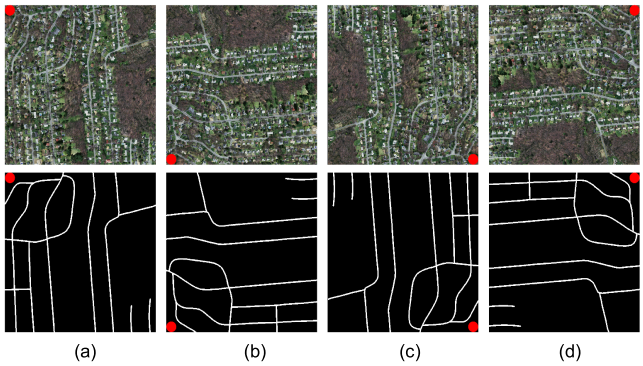

FIGURE 11. Samples of the dataset and data augmentation through anticlockwise rotation. (a) Cropped image and groundtruth; (b), (c) and (d) respectively indicate that the image and groundtruth are rotated 90 degrees, 180 degrees and 270 degrees.

III. EXPERIMENT

A. EXPERIMENTAL DATASETS

The Massachusetts roads dataset (publicly available and provided by [69]) consists of 1171 images of Massachusetts. The dataset covers cities, suburbs and rural areas, with a total area of more than 2600 square kilometers. It also contains a wealth of information, including roads, rivers, oceans, various buildings, vegetation, schools, bridges, ports and vehicles. Each image is 1500 × 1500 pixels in size. Considering the computer memory constraint, 865 images of good quality and complete information were selected from the Massachusetts roads dataset and divided each image into four 750 × 750 pixels parts suitable for the network learning environment. The data is randomly split into a training set of 3300 images and testing set of 160 images.

In deep network training, the model need to ensure that enough data is entered to avoid overfitting. Therefore, data augmentation has been proved to be a key technology in deep networks. In this study, data augmentation was achieved through rotating the segmented training image in three different angles. Samples of this dataset and data augmentation are shown in Fig. 11.

B. EXPERIMENTAL ENVIRONMENT OF Road-RCF

As one of the top ten frameworks for deep learning, Caffe has been a leader in the computer vision. We implemented our network architecture based on the Caffe platform. We used the pre-training parameters of the first 13 convolutional layers of VGG as the initial parameters for the fine-tuning of our networks. The network was trained for 40000 iterations, the learning rate was set to 10^{-7} and it was gradually decreased by 0.1 every 10000 iterations. Due to the same size of the input image, we applied a non-standardized missing learning rate. On a single NVIDIA GTX1080 GPU, we trained the Road-RCF model for approximately 31 hours.

C. COMPARISON EXPERIMENT

To test the validity of the proposed model for road detection, by using four different typical images, this paper

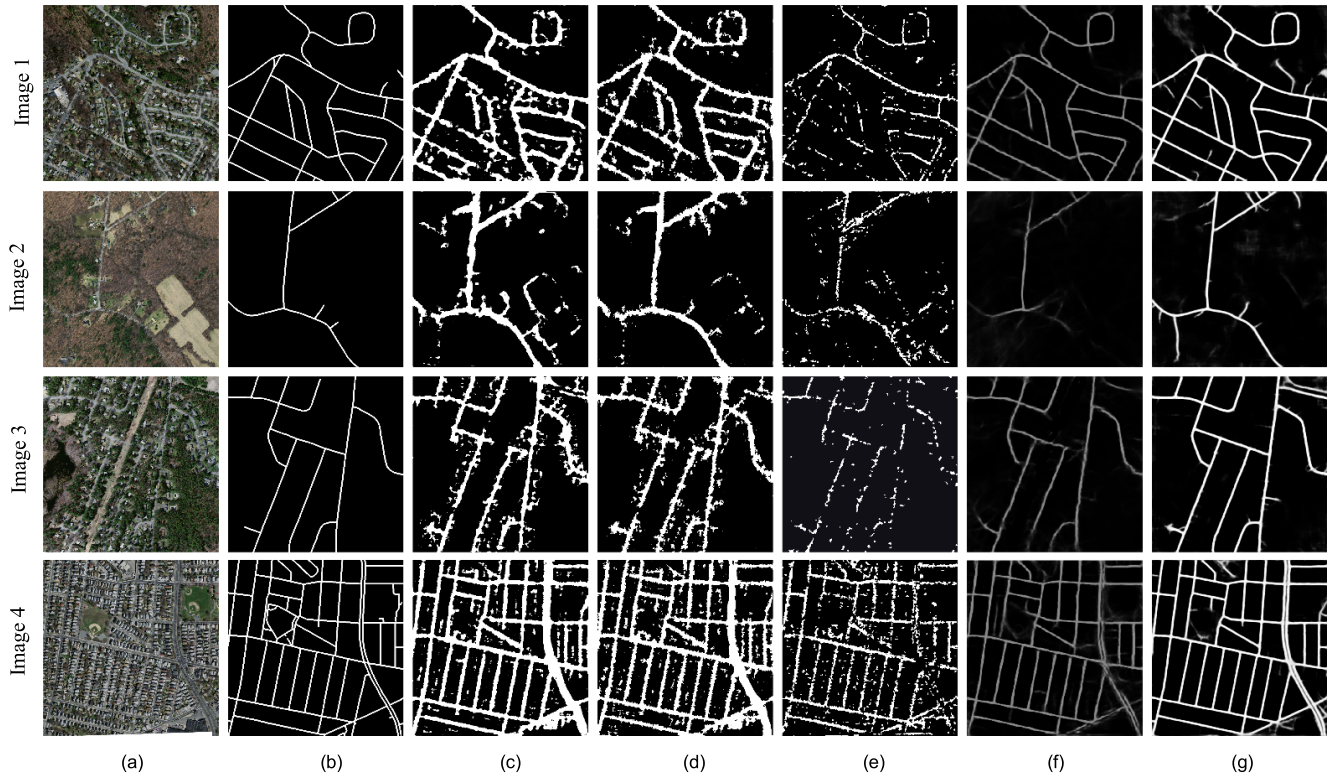

FIGURE 12. Visual comparisons of road extraction results with different comparing algorithms. (a) Input image. (b) Groundtruth. (c)-(g) The extracted road probability results in the Massachusetts roads dataset by Pixel-CNN, SLIC-CNN, SEEDS-MCNN, U-Net and Road-RCF.

compared the performance of road extraction using Road-RCF with the Pixel-based Convolutional Neural Network (Pixel-CNN) [70] (pixel-wise classification), the super-pixel classification based on the Simple Linear Iterative Clustering Convolutional Neural Network (SLIC-CNN) [72], the super-pixel classification based on the Multi-Scale Convolutional Neural Network (SEEDS-MCNN) [73], and the semantic segmentation methods based on the CNN, such as U-Net [74]. The lately proposed U-Net tries to deal with the problem of the pixel-wise classification, which used a hierarchy of decoders to gradually repair the details and spatial dimensions of the segmentation map. It is worth noting that the U-Net model has achieved good results in the scene feature detection on satellite images as reported on the Kaggle website.[1]

D. ROAD EXTRACTION RESULTS

Fig. 12 illustrates the four road extraction results of the Pixel-CNN, SLIC-CNN, SEEDS-MCNN, U-Net and the proposed Road-RCF. Without the influence of light, shelter, shadows, etc. (as shown in image 1), the roads exhibit long-length characteristics and gray-scale uniformity characteristics in remote sensing images, which objectively guarantees the accuracy and completeness of road extraction. Despite this, Road-RCF shows higher extraction accuracy on the boundary and could obtain smoother results. It is worth noting that, compared with image 1, the original images of image 2 and image 3 are more shaded with trees and shadows. Meanwhile, the shape and grayscale information are also seriously disturbed. The road boundary information extracted by the Pixel-CNN and SLIC-CNN methods were partly missed, although sufficient samples have been selected. There are serious "salt and pepper" phenomena and discontinuity due to the SEEDS-MCNN based road extraction results. The U-Net method takes into account the semantic information of the road, but the phenomenon of leakage is still more serious due to the influence of trees and shadows. In this case, roads extracted by Road-RCF are more continuous and complete than those produced by the other four methods. Meanwhile, the spectral and textural information of roads, buildings, parking lots, etc. are similar in the original image of image 4, resulting in plenty of small and wrong patch information in Pixel-CNN, SLIC-CNN and SEEDS-MCNN extracting much wrong small patch information. By comparison as shown in Fig. 12, it can be concluded that the proposed Road-RCF method is superior to Pixel-CNN, SLIC-CNN, SEEDS-MCNN and U-Net methods.

Based on the road probability results, the road centerlines were extracted after automatic threshold segmentation and vectorization. Fig. 13 shows the results of road centerline extraction and vectorization.

[1] https://deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentation/

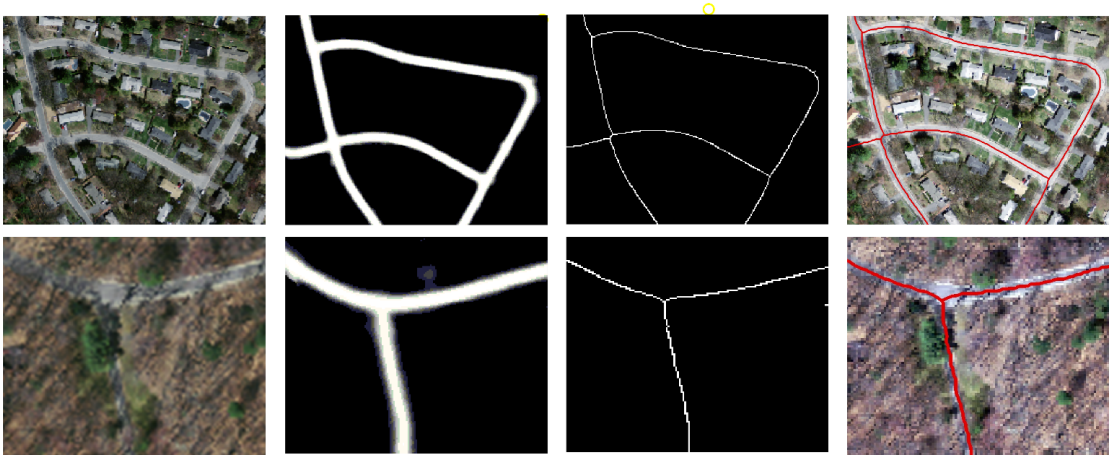

FIGURE 13. The results of road centerline extraction and vectorization. (a) Original image. (b) The extracted road probability results by Road-RCF. (c) Result of road centerline extraction. (d) Vectorized road superimposed with the original image effect.

Fig. 13 presents the visual effect of road centerline extraction and vectorization. It can be seen that the road in this image are concurrently affected by trees, buildings, vehicles, and branches, which is very common in real-world applications. The results show that the centerline extracted from the road is very smooth and well connected without interruption. Therefore, a complete road network can be obtained (as shown in Fig.13(c)). Fig. 13 (d) shows the result of the superposition of the original image and the vector road, from which it can be seen that the vectorized road is in good agreement with the original road.

E. EVALUATION METRICS

We extract roads' information from remote sensing images with complex features, which can be viewed as binary classifications. That is, road pixels are positive and non-road pixels are negative. According to this situation, we divided the predicted image pixels into four types according to the combination of the real category and the predicted category: true positive (*TP*), false positive (*FP*), true negative (*TN*), and false negative (*FN*). *Precision* is the percentage of correctly classified road pixels among all predicted pixels by the classifier. *Recall* is used to evaluate the percentage of road pixels that are correctly predicted as the actual road pixels. The *F1-score* is the harmonic average of *Precision* and *Recall*. In particular, we use the *Accuracy* index to perform an overall assessment of the classifier. Thus, the following four evaluation metrics are defined for assessing the performance of the proposed road extraction method.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

TABLE 3. Performance comparison of different methods.

	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Pixel-CNN	57.0	96.2	71.2	86.5
SLIC-CNN	58.1	96.2	72.1	89.2
SEEDS-MCNN	78.0	80.4	79.0	92.2
U-Net	89.3	88.5	88.9	93.4
Road-RCF	**85.8**	**98.5**	**91.5**	**96.3**

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}$$

Because of the difficulty in correctly labeling all the road pixels, we used the relaxed precision and recall scores [75] to evaluate road extraction. In this experiment, the parameter ρ is set to 3, which is consistent with previous studies [52], [62].

To better explore the contrasting effects between Road-RCF, Pixel-CNN, SLIC-CNN, SEEDS-MCNN and U-Net, we provided the quantitative comparisons of different road extraction methods in Table 3.

Table 3 shows the evaluation results of the five methods. In this paper, the four indicators including *Precision*, *Recall*, *F1-score*, and *Accuracy* were used. It is noteworthy that our method achieved a high *Accuracy* rate (96.3%) in these five methods. Simultaneously, the proposed approach achieved a *Precision* rate of 85.8% and a *Recall* rate of 98.5%, which are improvements of at least 7.8% and 2.3% relative to Pixel-CNN, SLIC-CNN and SEEDS-MCNN, respectively. Compared to Road-RCF, the U-Net model has a higher *Precision* rate (89.3%), but its *Recall* rate (88.5%) is much lower than 98.5%. Particularly, amongst the five methods, the Road-RCF achieved the highest *F1-score* (91.5%). By analyzing the *Precision*, *Recall*, *F1-score*, and *Accuracy* values from Table 3, it is clear that the proposed Road-RCF model yields more reliable and acceptable results and improves the general

classification performance, which is more suitable for the extraction of road information.

IV. DISCUSSION

Being different from other road extraction methods, the Road-RCF method fully combines the low-level features (e.g., spectral and texture) from the bottom level and the high-level features from the top level to perform precise extraction of road information. Meanwhile, the completeness and validity of road information influence the result of road vectorization directly. Normally, shadows of objects on and beside roads and the impact of the phenomenon that different object with same spectral characteristics would lead to incomplete road information extraction and failures. However, the issue cannot be solved by classical road extraction methods nor by the CNN road extraction. For quite some time, the road extraction from HSRRS images required manual work or the help of traditional techniques. However, there will be challenges when the workload gets greater. The network of Road-RCF which uses the training data manually labeled can overcome the problem that road information can't be obtained completely. The validity of extraction is promoted by deep supervision guidance and side-outputs multi-scale fusion, which meets the visual cognition pattern. Road centerlines are obtained and vectorized based on the information extracted by Road-RCF network. The final vectorized road centerlines possess accurate location information, better completeness and shapes as well as better visual explanation.

Natural images possess salient features, less information and a single structure. Unlike natural images, HSRRS images contain complex object information [76]. Road extractions in the image are greatly influenced by object spectrums, textures and positions. Meanwhile, geometric features of roads such as changes of curvature and width also bring challenges to road extraction.

According to experiment results, the proposed method has advantages in extracting road information from HSRRS images. Roads extracted by the Pixel-CNN has a serious problem of ''salt and pepper'' phenomena and fuzzy boundaries because the Pixel-CNN only counts on spectral and texture features and the mixed pixels in road boundaries lead to undesirable error classifications. Methods based on superpixels such as SLIC-CNN and SEEDS-MCNN effectively restrain the influence of mixed pixels by segmentation process and thus improve classification performance to a certain extent. Nevertheless, it's hard to find a proper segmentation scale among existing methods to fit road information with complex features in HSRRS images. U-Net owns an encoder-decoder structure. The encoder obtains sematic information of roads by constantly pooling while the decoder gradually renovates details of road information. Compared to the Pixel-CNN, SLIC-CNN and SEEDS-MCNN, the U-Net shows better accuracy in boundary extraction and has better performance in pixel classification.

In comparison to four excellent methods mentioned above, the Road-RCF shows better classification performance. It outputs road features from each convolutional layer and integrates them on multi scales. Multi-scale fusion of roads not only uses low-level information to maintain detail information features, but also utilizes high-level sematic information to obtain accurate boundaries and avoid impacts of shadows, curvatures and width changes. The optimal fits in different scales are obtained via constant learning of road features under the guidance of deep supervision. Many shallow and hidden road information are learned by multi-scale information fusion and under the guidance of manual labels.

The main limitation of our approach is that the road width can't be accurately obtained because the continuous pooling operation and the receptive field become large, causing the linear classifier in the full convolution structure to produce similar responses at the beginning of adjacent pixels, thus making the original image details cannot be restored by simple upsampling. In addition, the emergence of problems such as mixed pixels are also inevitable. Furthermore, there is a structural limitation of CNNs in conducting fine-grained detection. If we wish to keep a low number of learnable parameters, the ability to learn long-range contextual features comes at the cost of losing spatial accuracy. That is, there is a tradeoff between detection and localization. This is a well-known issue and still a scientific challenge [77].

V. CONCLUSIONS

This paper presents a Road-RCF model for road information extraction from HSRRS images. The proposed method applies an RCF network with deep supervision and side-output image-to-image fusion to obtain road feature maps. Deep supervision is imposed at each side-output layer to guide the side-outputs towards road predictions with the characteristics that we desire. The use of high-level semantic information that is acquired by side-output layer can reduce the impact of interference information and is helpful for effective road detection. The centerline extraction using the refinement algorithm shows a good smoothness and continuity, and the road centerlines are vectorized to facilitate the attribute information management and electronic map production. Compared with the Pixel-CNN, SLIC-CNN, SEEDS-MCNN and U-Net methods, the proposed method is proven to be a promising approach for road extraction.

Highlights of this paper are listed as follows:

(1) For the first time, the RCF model is used to extract road information from complex HSRRS images. Multi-scale fusion of road information and guidance of deep supervision are utilized to obtain better road extraction results, which performs best compared to other methods.

(2) About construction of RCF-Road network. Based on Massachusetts roads dataset, representative images are chosen after deleting false labels, which greatly reduces the interference of false information. Meanwhile, the dataset was enhanced by rotation based on data augmentation to avoid over-fitting phenomenon during the construction of RCF-Road network.

(3) From the view of practical use, other than edge detection by traditional RCF model, this work provides a new practicable idea for vectorized road extraction. Since it can produce integrated and complete road network map which can be used as the basic map of road vectorization. The vectorized road centerline can store additional non-spatial attribute information, which facilitates the updates and management of the electronic map.

However, our proposed method could not get accurate information about road width because of the problems of the network structure itself and the problem of mixed pixels. Thus, future research may focus on extracting the accurate location of the road network.

REFERENCES

[1] M. Kumar, R. Singh, P. Raju, and Y. Krishnamurthy, "Road network extraction from high resolution multispectral satellite imagery based on object oriented techniques," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 8, pp. 107–110, 2014.

[2] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy," *Int. J. Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, 2009.

[3] Y. Cao, Z. Wang, and Y. Lei, "Advances in method on road extraction from high resolution remote sensing images," (in Chinese), *Remote Sens. Technol. Appl.*, vol. 32, no. 1, pp. 20–26, 2017.

[4] J. Zhang, X. Lin, Z. Liu, and J. Shen, "Semi-automatic road tracking by template matching and distance transformation in urban areas," *Int. J. Remote Sens.*, vol. 32, no. 23, pp. 8331–8347, 2011.

[5] X. Lin, J. Zhang, and Z. Liu, "Semi-automatic extraction of ribbon roads from high resolution remotely sensed imagery by improved profile matching algorithm," (in Chinese), *J. Sci. Surv. Mapping*, vol. 34, no. 4, pp. 64–66, 2009.

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

[7] R. Li, Y. Si, S. Zhu, and B. Zhu, "Searching road centerline from remote sensing images using the angular texture signature," (in Chinese), *J. Geomatics Sci. Technol.*, vol. 31, no. 4, pp. 393–398, 2014.

[8] 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. (IGARSS)*, Jul. 2014, pp. 2712–2715.

[9] X. Lin, J. Zhang, H. Li, and J. Yang, "Semi-automatic extraction of ribbon roads from high resolution remotely sensed imagery by a T-shaped template matching," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 34, no. 3, pp. 293–296, 2009.

[10] R. Gaetano, J. Zerubia, G. Scarpa, and G. Poggi, "Morphological road segmentation in urban areas from high resolution satellite images," in *Proc. 17th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2011, pp. 1–8.

[11] S. Zhou and Y. Xu, "To extract roads with no clear and continuous boundaries in RS images," *Acta Geodaetica Cartograph. Sinica*, vol. 37, no. 3, pp. 301–307, 2008.

[12] 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, Nov. 2012.

[13] M. Airouche, M. Zelmat, and M. Kidouche, "Statistical edge detectors applied to SAR images," *Int. J. Comput. Commun. Control*, vol. 3, no. 2, pp. 144–149, 2008.

[14] S. Leninisha and K. Vani, "Water flow based geometric active deformable model for road network," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 140–147, Apr. 2015.

[15] P. N. Anil and S. Natarajan, "A novel approach using active contour model for semi-automatic road extraction from high resolution satellite imagery," in *Proc. 2nd Int. Conf. Mach. Learn. Comput. (ICMLC)*, Feb. 2010, pp. 263–266.

[16] M. Wang, J. Luo, C. Zhou, D. Ming, Q. Chen, and Z. Shen, "Extraction of road network from high resolution remote sensed imagery with the combination of Gaussian Markov random field texture model and support vector machine," *J. Remote Sens.*, vol. 37, no. 3, pp. 271–276, 2005.

[17] M. Rajeswari, K. S. Gurumurthy, L. P. Reddy, S. Omkar, and J. Senthilnath, "Automatic road extraction based on level set, normalized cuts and mean shift methods," *Int. J. Comput. Sci. Issues*, vol. 8, no. 3, pp. 250–257, 2011.

[18] X.-W. Wu and H.-Q. Xu, "Level set method major roads informaiton extract from high-resolution remote-sensing imagery," *J. Astronaut.*, vol. 31, no. 5, pp. 1495–1502, 2010.

[19] 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, Oct. 2012.

[20] C. Cheng and T. Ma, "Automatic recognition of landscape linear features from high-resolution satellite images," *J. Remote Sens.*, vol. 7, no. 1, pp. 26–30, 2003.

[21] T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia, "An extended phase field higher-order active contour model for networks and its application to road network extraction from VHR satellite images," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 509–520.

[22] H. Mayer, I. Laptev, A. Baumgartner, and C. Steger, "Automatic road extraction based on multi-scale modeling, context, and snakes," *Int. Arch. Photogramm. Remote Sens.*, vol. 32, no. 3, pp. 106–113, 1997.

[23] 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. (IGARSS)*, vol. 4, Jul. 2010, pp. 453–456.

[24] 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.

[25] W. Yi, Y. Chen, H. Tang, and L. Deng, "Experimental research on urban road extraction from high-resolution RS images using probabilistic topic models," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2010, pp. 445–448.

[26] K. Hedman, S. Hinz, and U. Stilla, "Road extraction from SAR multi-aspect data supported by a statistical context-based fusion," in *Proc. Urban Remote Sens. Joint Event*, Apr. 2007, pp. 1–6.

[27] Z. Qiaoping and I. Couloigner, "Automatic road change detection and GIS updating from high spatial remotely-sensed imagery," *Geo-Spatial Inf. Sci.*, vol. 7, no. 2, pp. 89–95, 2004.

[28] H. Chen, L. Yin, and L. Ma, "Research on road information extraction from high resolution imagery based on global precedence," in *Proc. 3rd Int. Workshop Earth Observ. Remote Sens. Appl. (EORSA)*, Jun. 2014, pp. 151–155.

[29] S. Movaghati, A. Moghaddamjoo, and A. Tavakoli, "Road extraction from satellite images using particle filtering and extended Kalman filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2807–2817, Jul. 2010.

[30] C. Poullis, "Tensor-Cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 95, pp. 93–108, Sep. 2014.

[31] J. Qian, J. Wang, R. Ma, and Y. Deng, "Feature semantics information extraction of urban road based on quickbird imagery," (in Chinese), *Remote Sens. Technol. Appl.*, vol. 29, no. 4, pp. 653–659, 2014.

[32] 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.

[33] Z. Sun, H. Fang, M. Deng, A. Chen, P. Yue, and L. Di, "Regular shape similarity index: A novel index for accurate extraction of regular objects from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3737–3748, Jul. 2015.

[34] L. Wang, X. Lin, and Y. Liang, "Automatic extraction of main roads from high resolution remote sensing imagery based on perceptual organization," (in Chinese), *Sci. Surv. Mapping*, vol. 42, no. 7, pp. 127–131, 2017.

[35] 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.

[36] Z. Hui, Y. Hu, S. Jin, and Y. Z. Yevenyo, "Road centerline extraction from airborne LiDAR point cloud based on hierarchical fusion and optimization," *ISPRS J. Photogramm. Remote Sens.*, vol. 118, pp. 22–36, Aug. 2016.

[37] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[38] X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
[39] Y. Zhong, F. Fei, Y. Liu, B. Zhao, H. Jiao, and L. Zhang, "SatCNN: Satellite image dataset classification using agile convolutional neural networks," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 136–145, 2017.
[40] F. P. S. Luus, B. P. Salmon, F. van den Bergh, and B. T. J. Maharaj, "Multi-view deep learning for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2448–2452, Dec. 2015.
[41] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, Jan. 2017.
[42] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning Earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.
[43] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.
[44] P. Zhou, G. Cheng, Z. Liu, S. Bu, and X. Hu, "Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 925–944, 2016.
[45] L. Zhang, Z. Shi, and J. Wu, "A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4895–4909, Oct. 2015.
[46] I. Ševo and A. Avramović, "Convolutional neural network based automatic object detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 740–744, May 2016.
[47] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval," *Remote Sens.*, vol. 9, no. 5, pp. 489–509, 2017.
[48] T. Jiang, G.-S. Xia, and Q. Lu, "Sketch-based aerial image retrieval," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3690–3694.
[49] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[50] M.-M. Cheng, Q.-B. Hou, S.-H. Zhang, and P. L. Rosin, "Intelligent visual media processing: When graphics meets vision," *J. Comput. Sci. Technol.*, vol. 32, no. 1, pp. 110–121, 2017.
[51] M.-M. Cheng *et al.*, "HFS: Hierarchical feature selection for efficient image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 867–882.
[52] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 210–223.
[53] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks," *Electron. Imag.*, vol. 60, no. 1, pp. 1–9, 2016.
[54] E. Sarhan, E. Khalifa, and A. M. Nabil, "Road extraction framework by using cellular neural network from remote sensing images," in *Proc. Int. Conf. Image Inf. Process. (ICIIP)*, Nov. 2011, pp. 1–5.
[55] J. S. J. Wijesingha, "Automatic road feature extraction from high resolution satellite images using LVQ neural networks," *Asian J. Geoinform.*, vol. 13, no. 1, pp. 30–36, 2013.
[56] P. Li *et al.*, "Road network extraction via deep learning and line integral convolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1599–1602.
[57] W. Zhao, S. Du, and W. J. Emery, "Object-based convolutional neural network for high-resolution imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3386–3396, Jul. 2017.
[58] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.
[59] 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. (IGARSS)*, Jul. 2016, pp. 1591–1594.
[60] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017.
[61] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, pp. 498–519, 2017.
[62] 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.
[63] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields," *Remote Sens.*, vol. 9, no. 7, pp. 680–699, 2017.
[64] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5872–5881.
[65] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.
[66] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[67] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[68] R. Liu and Y. Li, "Study of auto-vectorization based on scan-thinning algorithm," *Acta Geodaetica Cartograph. Sinica*, vol. 41, no. 2, pp. 309–314, 2012.
[69] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.
[70] W. Shao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Oct. 2016.
[71] A. Darwish, K. Leukert, and W. Reinhardt, "Image segmentation for the purpose of object-based classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 3, Jul. 2003, pp. 2039–2041.
[72] W. Zhao *et al.*, "Superpixel-based multiple local CNN for panchromatic and multispectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4141–4156, Jul. 2017.
[73] X. Lv, D. Ming, Y. Chen, and M. Wang, "Very high resolution remote sensing image classification with SEEDS-CNN and scale effect analysis for superpixel CNN classification," *Int. J. Remote Sens.*, doi: 10.1080/01431161.2018.1513666.
[74] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[75] M. Ehrig and J. Euzenat, "Relaxed precision and recall for ontology matching," in *Proc. K-Cap Workshop Integr. Ontol.*, 2005, pp. 25–32.
[76] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.
[77] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez, "Recurrent neural networks to correct satellite image classification maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4962–4971, Sep. 2017.

ZHAOLI HONG is currently a Graduate Student with the School of Information Engineering, China University of Geosciences, Beijing. His research interests include deep learning and information extraction from high spatial resolution remote sensing image.

DONGPING MING received the B.E. degree in land administration and cadastral surveying from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1999, the M.E. degree in cartography and geographic information engineering from Wuhan University, Wuhan, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2006. She is currently an Associate Professor with the School of Information Engineering, China University of Geosciences, Beijing. Her research interests include remote-sensing image processing and analysis, and information extraction from high spatial resolution satellite remote sensing image.

KEQI ZHOU is currently pursuing the master's degree with the School of Information Engineering, China University of Geosciences, Beijing. His research interests include high spatial resolution remote sensing imagery analysis, deep learning, and environment protection.

YA GUO is currently pursuing the master's degree with the School of Information Engineering, China University of Geosciences, Beijing. His research interests include high spatial resolution remote sensing classification based on machine learning and deep learning.

TINGTING LU is currently pursuing the master's degree from the School of Information Engineering, China University of Geosciences, Beijing. Her main research direction is to detect building edge from high resolution imagery based on CNNs.