Road Extraction Methods in High-Resolution Remote Sensing Images: A Comprehensive Review

Renbao Lian, Weixing Wang, Nadir Mustafa, and Liqin Huang

***Abstract*—Road extraction from high-resolution remote sensing images is a challenging but hot research topic in the past decades. A large number of methods are invented to deal with this problem. This article provides a comprehensive review of these existing approaches. We classified the methods into heuristic and data-driven. The heuristic methods are the mainstream in the early years, and the data-driven methods based on deep learning have been quickly developed recently. With regard to the heuristic methods, the road feature model is first introduced, then, the classic extraction methods are reviewed in two subcategories: semiautomatic and automatic. The principles, inspirations, advantages, and disadvantages of these methods are described. In terms of the data-driven methods, the road extraction methods based on deep neural network, particularly those based on patched convolutional neural network, fully convolutional network, and generative adversarial network are reviewed. We perform subjective comparisons between the methods inner each type. Furthermore, the quantity performances achieved on the same dataset are compared between the heuristic and data-driven methods to show the strengthening of the data-driven methods. Finally, the conclusion and prospects are summarized.**

***Index Terms*—Data-driven, heuristic, high resolution, remote sensing image, road extraction.**

I. Background

REMOTE sensing images acquired by airborne or spaceborne sensors are the main resource for the earth surface observation, environment monitoring, objects identification, etc. [1]. In particular, the high-resolution remote sensing images (HRSI) become very important for geographic information system (GIS) application, ecological research, land management, and disaster monitoring [2]. The key issue is how to recognize the objects of interest in HRSI quickly and accurately [3], [4]. The basic processing of HRSI is to extract information for objects classification and recognition according to not only the spectral features but also the shape characters and spatial relations [5]. The bottleneck of these problems is the recognition of the small terrestrial objects like roads, junctions, buildings, etc. [6].

Man-made objects are most important elements in spatial geographic information databases, including buildings, bridges, roads, farmlands, etc. Buildings and roads are unquestionably the two most important object classes. A large number of articles have been published discussing the detection of buildings and roads during the last decades [7]. For example, we will get more than 2.6 million related papers if we search Google Scholar with the keyword "road extraction," and more than 129 000 articles remain if the publish date is restricted to post 2016. Faced with such a large number of works of literature, a systematic review of road extraction algorithms is valuable for beginners. Although we have reviewed the traditional methods before [8], we stick to present a more comprehensive review from a new perspective given the rapid development of deep learning and the numerous articles discussing automatic road extraction employing deep learning models [9]. We divide the existing methods into two categories, namely, heuristic and data-driven. The heuristic methods extract road features using the prior of road regions, whereas the data-driven methods depend on big data. For heuristic methods, we supplemented some methods missing in the previous version and made a more concrete description of the basic principles of each method, which will be more friendly to beginners. In order to review the heuristic and data-driven methods systematically, we collected and sieved more 200 papers discussing road extraction from HRSI published in the past two decades. We only kept the pieces of literature that extract roads directly from optical images captured by satellites or aircraft instead of from the data generated by light detection and ranging or synthetic aperture radar. We restricted the articles to peer-reviewed ones and preferred those with high citations. To the best of our knowledge, this article is the first comprehensive review that surveys the road extraction methods based on traditional algorithms and deep learning technology.

II. Introduction

In 1972, the first Earth Resources Observation Technology Satellite, later renamed Landsat, was launched by the United States. The interpretation technologies of remote

Manuscript received June 17, 2019; revised December 23, 2019, April 7, 2020, and July 28, 2020; accepted September 1, 2020. Date of publication September 11, 2020; date of current version September 25, 2020. This work was supported in part by the Natural Science Foundation of China under Grant 61170147, in part by the Education and Scientific Research Project for Middle-Aged and Young Teachers in Fujian Province under Grant JT180595, in part by the Key Program of Fostering Young Scientific Research Talents in Fujian Jiangxia University under Grant JXZ2018002 and Grant JXZ2016001, and in part by the Fuzhou Science and Technology Project under Grant 2020-G-066. *(Corresponding authors: Weixing Wang; Liqin Huang.)*

Renbao Lian is with the College of Physics and Information Engineering, Fuzhou University, Fuzhou 350008, China, with the Digital Fujian, Internet-of-Things Key Lab of Information Collection and Processing in Smart Home, Fuzhou 350108, China, and also with the College of Electronics and Information Science, Fujian Jiangxia University, Fuzhou 350108, China (e-mail: luoshao@163.com).

Weixing Wang is with the KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: znn525d@qq.com).

Nadir Mustafa and Liqin Huang are with the College of Physics and Information Engineering, Fuzhou University, Fuzhou 350008, China (e-mail: nadirmu17@hotmail.com; hlq@fzu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3023549

sensing imagery have been developed rapidly, including image compression, transmission, segmentation, fusion, understanding, etc. With the use of modern sensors (e.g., IKONOS, QuickBird, and GeoEye), the spatial, spectral, and time resolutions of the RS images have gradually increased. HRSI provides a new way to obtain detailed geographic information, which consequently motivates further development of the processing technology of HRSI [10].

At present, Image recognition technologies have been successfully applied to many specific fields, such as fingerprint identification [11], face recognition [12], scene description [13], etc. However, there still have many problems with understanding RS (including aerial) images [8]. As man-made objects, roads are the important information processed in the GIS. Road information can be used in many aspects of social life, such as vehicle navigation, traffic management, map updating, and geological disaster emergency [14]. With the further application of GIS, the manual extraction methods cannot fulfill the quick updating requirements. Therefore, extracting road information from RS images by (or aided by) machine is in demand.

We can easily collect a large number of articles about road extraction method from Web of Science or Google Scholar, but the classification of them is very difficult due to the various applied technologies. In accordance with distinct classification criteria, road extraction methods can be classified into various types. For example, based on the different algorithms, these methods can be categorized into clustering, classification, morphology, dynamic programming, active contour models, etc. [8]. According to the objects to be handled in the algorithms, these methods can be pixel based, region based, and knowledge based [5]. On the basis of the different outputs of these methods, they can be divided into road segmentation and road centerline extraction [15]. In this work, we divide the road extraction methods into two types: the heuristic methods and the data-driven methods. The heuristic methods are further divided into semiautomatic methods and automatic methods according to the degree of the manual intervention. The data-driven methods are subdivided into several types based on the architecture of neural network (see Fig. 1).

The rest of this article is organized as follows. Section III introduces the road features and road model. Section IV describes the classic algorithms that belong to the heuristic strategy. Section V shows the data-driven methods according to different DCNN model. In Section VI, we list the metrics used to evaluate road segmentation or road network extraction. Comprehensive comparisons are listed in Section VII. The conclusion and prospects are summarized in Section VIII.

III. Road Features and Model

According to the theories of human visual systems [16], there are three computation levels starting from the primal sketch extraction to the 3-D model construction. Similarly, road extraction also undergoes three processing levels, which are, extracting edges, lines, and texture in the low level, filtering and grouping low-level features according to the road models in the middle level and construct road network using road prior in the high level. The definition of road features and model is important to road extraction.

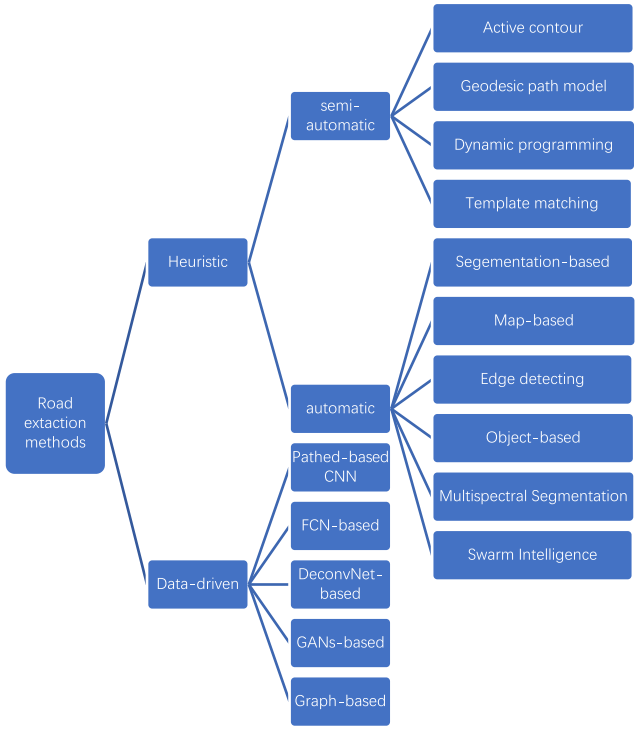


Fig. 1. Classification of road extraction methods.

A. Road Features

The road network is difficult to extraction from HRSI because the road features can be affected by different sensor type, distinct spectral and spatial resolution, volatile weather conditions, diverse road material, complex backgrounds, etc. It is important to analyze the road features and road model in the normal situation without considering noise interference. Overall, roads in HRSI present ribbons with slow intensity change [17]. We summarize road features as follows.

1) Geometric Features: Roads are stripe-shaped objects whose width does not change dramatically. Its length is much longer than its width. The intersections of roads can be visually described as the shape of "+," "Y," or "T." Besides, when the spatial resolution of RS images reduced, road ribbons may be degraded as linear geometric objects.

2) Spectral Features: Roads are characterized by the evident parallel edges. Gradient magnitudes are high on road edges but low inside roads. Strong intensity contrast exists between roads and the surroundings, but the contrast may be completely opposite due to the different road materials [18]. For example, cement roads generally present brighter than bitumen roads in panchromatic satellite images.

3) Topological Features: Road networks typically have intersections, and roads are not suddenly interrupted. In images, roads generally span across the entire scenes, unless the roads are dead ends. The road network can be viewed as a graph composed by vertices and edges.

4) Texture Features: The texture of images is a strong local characteristic, which is local homogeneity reflected in the vision. It is not related to color and intensity. Generally, the texture is smoother on the road surface than on the background.

In practice, many road extraction methods used multiple road features rather than only one feature. However, subject to the influence of illumination, shadow, and occlusion, roads in images do not have all the above-mentioned features, which complicates the extraction of roads from RS images [19].

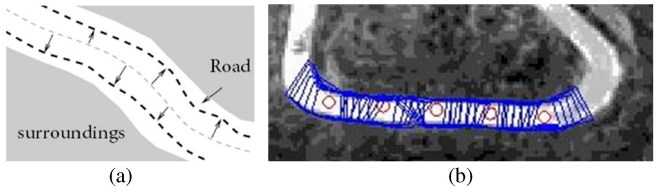


Fig. 2. Road extraction based on Snake model. (a) Deformed from initial lines (ribbon snakes). (b) Deformed from seed points (balloon snakes). (a) Ribbon snakes. (b) Balloon snakes.

B. Road Model

Establishing a road model can help us extract road more effectively. Baumgartner *et al.* [20] proposed a classic road model based on the composition of roads. The road model is divided into three layers, i.e., realistic road network layer, geographic geometric feature layer, and image feature layer. The model shows how the different of road materials and geometric shapes in the real world presented in the images. The model also demonstrates the road features from the perspective of high and low resolutions. More precise information can be extracted from RS image with higher resolution, such as road lanes and zebra crossings. However, higher resolution may introduce more interferences, which will disturb the extraction of global road networks. In the coarse scale, most interferences on road surfaces are eliminated, and prominent road edges are preserved to identify road networks. However, the extracted roads are typically broken and imprecise given the lack of resolution. On this basis, the road extraction methods based on multiscale segmentation are extensively researched [21].

IV. Heuristic Road Extraction Methods

We subdivide the heuristic road extraction methods into two types according to the degree of interaction: semiautomatic and automatic. In semiautomatic methods, some seeds need to be placed in advance or during the extraction, or request users to decide whether the extracted results are correct or not. In automatic extraction algorithms, the parameters usually need to be preset according to different images. Then, the whole workflow can be executed without supervision.

A. Semiautomatic Methods

Tracking roads manually in the HRSI is the most time-consuming part of geo-databases updating [22]. Researchers eagerly hope that the computer can automatically extract the road network and update the map in real-time. However, automatic extraction methods cannot achieve satisfying results in consideration of the complexity of RS images. Therefore, researchers have to adopt a compromised solution, that is, semiautomatic extraction, because human can perform the task of road identification flawlessly and almost effortlessly. Some semiautomatic methods are successfully applied to the existing commercial software, such as Feature Analyst, Definiens eCognition, and GeoEye Road Tracker [23]. In semiautomatic road extraction methods, seeds (typically the center points or contours of a road segment) are manually provided, and road directions are also requested sometimes. Then, other road centers are searched according to certain road features and linked to form road segments. In semiauto mode, the algorithms can request intervention when they fail to find the other road centers due to the complexity of RS images, or rollback the previous results when the algorithms extract wrong roads, which increases the performance and robustness of these methods [24]. Due to the various technologies applied, it is difficult to describe all the existing methods. Some classic methods of existing methods are described in the following sections.

1) Methods Based on Active Contour Model: The active contour model, also called Snake, was proposed by Kass *et al.* [25]. In this model, an energy function is defined first on the contour drawn by the user; then, the minimum value of the energy function is iteratively computed by adjusting the contour until the contour finally converges. The algorithm has been successfully applied to medical image segmentation and has been extended to other domains. In road extraction, these algorithms are called ribbon snakes or balloon snakes [26]–[32]. The deformation of contours can start from initial lines or points. In the case of initial lines, the initial lines are manually drawn, and ribbon contours are initialized around the lines. The algorithm calculates the minimum energy by deforming the ribbon contours until the contours are no longer changed. An illustrative diagram of a ribbon snake is depicted in Fig. 2(a), in which the gray dashed line is an initial road line. For starting from the initial points, the road region is regarded as a pipeline, and seed points are manually placed inside the pipeline. Then, circular or rectangular contours are initialized centered at these seeds. The contours deform like expansive or shrunk balloons by calculating the minimum value of the energy function iteratively. Ideally, the contours are exactly on the edges of the road when the energy is minimized. The road regions can be easily obtained by merging the final contours. The principle of the balloon snake is exhibited in Fig. 2(b), where the red circles are the initial circular contours centered at the seeds, and the blue contours are the deformed contours (for clarity, the contours are drawn once in every three iterations). The methods are practicable by combining the prior knowledge of human and the accurate calculation of computer. However, the energy optimization is easy to fall into the local optimal and extracts the wrong road regions owing to the complexity of RS images. The level set algorithm is an evolutionary Snake algorithm, which also requires an initial contour curve, and then performs contour evolution according to the functional energy optimization [33]. Niu [34] studied a method that integrated the boundary gradient

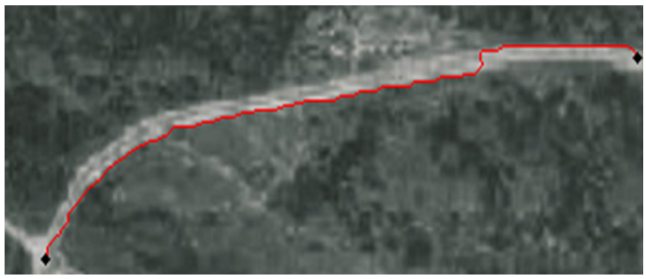
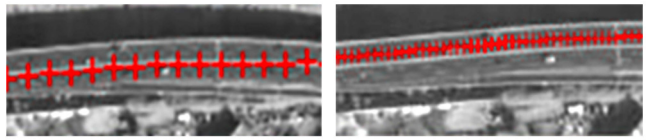


Fig. 3. Extraction of road edge by FMM.



(a) (b)

Fig. 4. Template matching for road centerline extraction. (a) Correct road centerline. (b) Deviated road centerline.

with the area information to construct a model to minimize an objective function that connects the optimization problem with the propagation of regular curves. It is worth noting that, these methods can be fully automatic if the initial lines or seed points can be automatically detected.

2) Methods Based on Geodesic Path: In road linking, a road that connects two seed points x_s and x_e can be approximately defined as a smooth curve that has a constant gray value c [35]. Let W denote the potential (or saliency) map of an image, the length of the smooth curve connecting x_s and x_e in image γ can be defined by a weighted length, as shown in (1). Function (1) can be regarded as an energy function, and the geodesic path $\gamma*$ can be solved by minimizing the energy function, as defined in (2). The fast marching method (FMM) is extensively used in minimizing energy function, which is proposed by Cohen [36]. In FMM, an energy map is computed by accumulating the gradients in potential map from x_s (start point) to x_e (end point), and then, the geodesic path can be tracked starting from x_e to x_s conversely following the steepest descent in energy map [37]. The essence of this method is to find the path with the greatest gradient between the two seeds. As such, a continuous, single-pixel path is found. On this basis, only the edges of the road can be obtained, as displayed in Fig. 3. Meanwhile, this method is inefficient because it calculates the accumulated energy map for a huge number of pixels. To enhance the efficiency, Yang *et al.* [38] improved this method by using multipoints fast marching simultaneously. Moreover, to obtain the road centerline, Miao *et al.* [35] extracted the road edges between seeds by a fast marching run. Then, they produced a road probability map using the extracted road edges. After that, the image was classified into road and nonroad according to the road probability map. Finally, a kernel density estimation (KDE) map was generated using the road class pixels, and the road centerline was extracted by applying fast marching again on the KDE map. Later, they proposed a similar algorithm [39], which applied the geodesic method on the hue–saturation–value color space instead of red–green–blue color space, and achieved better performance

$$L\ (\gamma) = \int_0^1 W\left(\gamma\left(t\right)\right) \left\|\gamma'\left(t\right)\right\| dt,\ \gamma\ (0) = x_s\ ,\ \gamma\ (1) = x_e \tag{1}$$

$$\gamma^* = \underset{\gamma \in (x_s, x_e)}{\operatorname{argmin}} L\left(r\right). \tag{2}$$

3) Methods Based on Dynamic Programming: Dynamic programming is a technique for solving optimization problems when not all variables in the evaluation function are interrelated simultaneously [40]. In general, a parameter model of the road should be given and expressed as a metric function; then, dynamic programming is taken as a computational tool to determine the optimal path among the seed points. The basic process can be listed as follows. First, a series of seed points are manually placed on the road regions as the initial vertices. Connecting these seed points, an initial polygon is formed. Second, dynamic programming is used to find the maximum value of the merit function by adjusting these seeds in a local window. Third, new equidistance vertices are inserted by a linear interpolation between every two adjacent vertices, and the dynamic programming is applied on this new polygon. Finally, the third step is repeated until convergence is reached, and the path represented by the final polygon is regarded as the road centerline. Gruen and Li [41] proposed a semiautomatic road extraction scheme that combined the wavelet decomposition for road sharpening and a model-driven linear feature extraction algorithm based on dynamic programming. Inspired by [41], Poz and Do Vale [42] introduced road width into the metric function, which can be used for road extraction in HRSI. Furthermore, Poz *et al.* [43] improved the dynamic programming optimization in object space to overcome a certain degree of noise. Moreover, the follow-up works [44]–[46] autoestimate seeds using maximum road likelihoods, and connect these seeds by minimizing cost paths and construct road networks by mixed integer programming, and finally overcome the inefficiency of artificial intervention.

4) Methods Based on Template Matching: The template matching methods are applicable because the road features are stable and the road surface shows high similarity in a certain range. In template matching methods, a specific matching window (e.g., rectangular window [46], [47], circular window [48], or T-shaped window [49], etc.) is constructed at the starting point. Then, the initial window is pushed forward at a certain step iteratively to find a series of matching points by the best similarity criterion with respect to some characteristic or statistic features, such as color, intensity, profile, and texture. Finally, the matching points are connected to form road segments, as presented in Fig. 4(a). However, the matching window may encounter occlusion (e.g., shadow, trees, and buildings) that violates the road features. Thus, it may fail to find the next road point. The manual intervention is required to reselect a new seed and resume the matching process. In some situations, the position of the seed significantly influences the extraction result. If the seed is not exactly on the road center, then the extracted lines may also deviate from the road center [46], as

illustrated in Fig. 4(b). The new template matching methods typically included a preprocessing of the automatic adjustment of the seed points.

B. Automatic Methods

The human intervention required by semiautomatic methods reduces the working efficiency significantly. To improve the automation, researchers have studied automatic road extraction methods for decades. Despite the lack of a fully automatic algorithm to produce satisfactory results for all types of RS images, some successful works have been achieved in certain special scenes.

1) Methods Based on Segmentation: There are many algorithms applied to image segmentation, which can also be used in the road extraction issue. In fact, many segmentation techniques are already used in road extraction to obtain road regions [50]. The following algorithms are commonly used: artificial neural network (ANN), support vector machine (SVM), Bayesian classifier, watershed algorithm, mean shift (MS), *K*-means, Gaussian mixture models (GMM), superpixel segmentation, conditional random field (CRF), graph-based segmentation, etc.

The ANN is inspired by the biological neural system, and developed rapidly after the introduction of the back propagation algorithm. A back-propagation neural network for road extraction was proposed by Mokhtarzade and Zoej [51]. They searched the best network structure by designing the network with different hidden layer sizes and training the network with different epochs. Neural network was applied for road detection in the paper [52]. At first, a binary image was generated by neural network using only spectral information. Then, different texture features are computed for each pixel using a gray level co-occurrence matrix from the source image, which were used to produce a segmented image. Finally, the road map was optimized by fusing the binary image and the segmented image. Although neural networks are widely used for image segmentation, they are more likely to get into the local minima and become overfitting.

The SVM is a generalized linear classifier for binary classification trained by supervised learning. SVM uses a hinge loss function to calculate empirical risk and adds regularization terms to the solution system to optimize structural risk. SVM can perform nonlinear classification by the kernel function. Yager and Sowmya [53] exploited the SVM classifier for solving the problem of road extraction from remotely sensed images using edge-based features, but the correctness is relatively low as many researchers reported. In paper [54], the SVM was employed merely to classify the image into two groups of categories: road and nonroad. Abdollahi *et al.* [55] presented an automatic method for road extraction by integrating the SVM and level set methods. The estimated probability of classification by SVM was used as input in the level set method, which achieved high performance of road extraction from Google Earth images. The road and building detection using a multiclass SVM method was proposed by Simler [56], in which the both spatial and spectral information were used at the object level. The SVM methods can minimize the structural risk and the have good generalization ability, but have the difficulties in the selection of kernel function and sensitive to the training samples, etc.

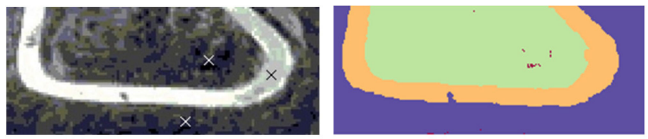


Fig. 5. Segment by mark-based watershed algorithm. (a) Origin image with basin makers. (b) Segment of watershed algorithm.

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. The principle of the classifier is to calculate the posterior probability of an object using its prior probability and the Bayesian theorem, that is, the object belongs to a certain class with the largest posterior probability. In the paper [57], the buildings were detected by Bayesian decision theory in regard to Laplacian probability density function followed by the discerning of roads by a special intensity threshold. Storvik *et al.* [58] described a Bayesian framework for classification based on the multiscale features, which was realized by the iterative conditional modes algorithm. The naive Bayes classifier shows better robustness unlike SVMs, which are easily affected by noise samples.

Watershed is a classic image segmentation algorithm with the defect of over-segmentation. Two means are commonly used to solve this problem: region merging and marker based. A road extraction algorithm based on region merging was proposed in the work [59]. The region merging process was constrained by the threshold of minimum region area and the maximum intensity variance between regions (named hydronephrosis basin). Then, the initial road segments were identified by combining the shape features of the regions and the road's prior. However, the threshold minimum region area is difficult to be preset. The marker-based watershed algorithm, which was proposed by Meyer [60], requires a marker-map provided by the operator. Fig. 5(a) depicts an RS image with three initial basin markers indicating that the image will be divided into three regions. Fig. 5(b) demonstrates the result of Meyer's algorithm. The outputs of the marker-based algorithm are remarkable, but the manual intervention reduces the degree of automation of the watershed algorithm. In some application field, researchers tried to automatically determined the initial basin markers to improve the automation degree [61]–[63]. Wu *et al.* [64] proposed a regional adaptive segmentation to automatically find the initial markers. Inspired by [64], Li and Zhang [65] presented a fast automatic road extraction algorithm. The algorithm initially collects the road regions using the technology proposed in [64], then it extracts lines using linear features [66]. Finally, it identifies road objects automatically by fusing region features, line features and road's prior. Watershed-based segmentation algorithm also can be used to find the road seed points for fully automatic road extraction method [67].

MS method is a clustering technique used to classify data into different categories and does not require information about

(a) (b) (c)

Fig. 6. Road extraction by *K*-means. (a) Original RS image. (b) *K*-means clustering. (c) Extracted roads.

(a) (b) (c)

Fig. 7. Road region recognition based on GMM. (a) Origin image. (b) Merged superpixel. (c) Clustering by GMM.

the specific object [68]. An analysis framework based on MS was proposed to extract spectral and spatial features from HRSI data [69]. In this approach, the MS was used to obtain an object-oriented representation of hyperspectral imagery. A road extraction method-based MS was suggested in [68], in which the MS procedure was obtained by successive computation of an MS vector and translation of the kernel by the MS vector. A simple yet effective method for road network extraction was proposed in [70], which achieved the detection of potential road center points using MS based on homogenous property and ribbon-like shape of road. By the way, the MS algorithm also can be used to adjust the manual seeds in semiautomatic road extraction based on the road KDE [35], [71].

K-means clustering is an unsupervised learning algorithm. It uses the similarity indices as the distance to classify the data into different categories based on the features of the data. The algorithm does not need to manually set the cluster seeds, but randomly initializes K centroids. Then, it calculates the Euclidean distances between data and K centroids. It classifies the data into the cluster with the smallest distance. When all the data are classified, the mean feature vector of each cluster will be calculated again and be treated as the new centroid of the cluster. The process is repeated until each cluster centroid converges. In some HRSI, an evident intensity difference is observed between the roads and the background. The *K*-means clustering algorithm can be used to segment an image into different clusters [72], [73], as displayed in Fig. 6. Fig. 6(a) presents an IKONOS RS image generated by fusing four bands (blue, green, red, and near-infrared); Fig. 6(b) demonstrates the result of *K*-means clustering, where $K = 3$; and Fig. 6(c) presents the result of recognition using road's prior. The method is suitable for situations where the road and the surrounding background are significantly different (e.g., suburban and rural roads). In the case of urban roads, the spectral features of buildings are similar to those of roads. Thus, roads are difficult to be classified effectively. For example, the clustering error is illustrated in the northeast corner of Fig. 6(c). In addition, the hyperparameter of K in these algorithms is not constant for different scenes, thereby reducing automation.

GMM is a parametric probability density function that is extensively used in computer vision and pattern recognition. It assumes that the distribution of data can be modeled by a mixture of K different Gaussian distributions. Expectation-maximization algorithm can be used to solve the parameters optimization of GMM. Li *et al.* [74] presented a road region extraction method

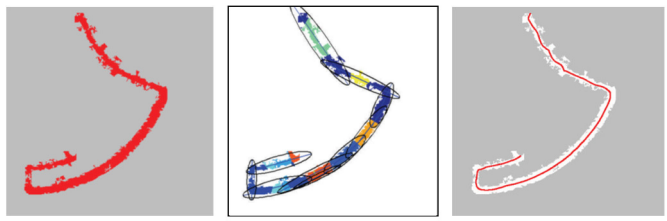


Fig. 8. Road centerline extraction based on GMM. (a) Road region. (b) Long axes of GMM. (c) Axes adjusted by SCMS.

based on GMM using object features. The algorithm used the superpixel segmentation algorithm [75] to divide the image into a series of superpixels initially. Then, it merged these superpixels into larger objects following graph theory and then extracted the 13-D features of these objects, including gradient, color (YUV space), and texture. Finally, the GMM was applied to cluster these objects into different categories using the 13-D features. The algorithm framework is depicted in Fig. 7. Fig. 7(a) exhibits the original image, Fig. 7(b) displays the merged superpixels, and Fig. 7(c) presents the GMM clustering result, where the hyperparameter K equals 4. The pixels in the segmented road regions can be represented as 2-D discrete joint random variables. On this basis, Miao *et al.* [76] presented a novel application of GMM for extracting the road centerlines from the segmented road regions. The road regions were cut into multiple segments by taking road pixels as observations. Then, different road segments were fitted with various Gaussian models to obtain the long axis of the Gaussian ellipse, which corresponds to each road segment. Meanwhile, each long axis is regarded as the initial centerline of each road segment. Finally, the SCMS [77] algorithm is used to adjust these initial centerlines to the exact positions. The process diagram is illustrated in Fig. 8.

Superpixel segmentation technology was first proposed by Ren and Malik [78], which combined adjacent pixels with similar features into pixel blocks and replaced massive pixels with a minimal quantity of irregular blocks. The texture interference was reduced, and the efficiency of the algorithm was improved after superpixel segmentation. After that, several new superpixel segmentation algorithms were introduced including Graph-Cut [79], Quick-Shift [80], GCa10 and GCb10 [81], Turbopixel [82], SLIC [83], ERS [75], PBO [84], etc. In [45], the image was first segmented into superpixels. These superpixels were treated as the smallest units (entities) to be labeled. Then, a feature vector per superpixel was extracted and fed into a binary

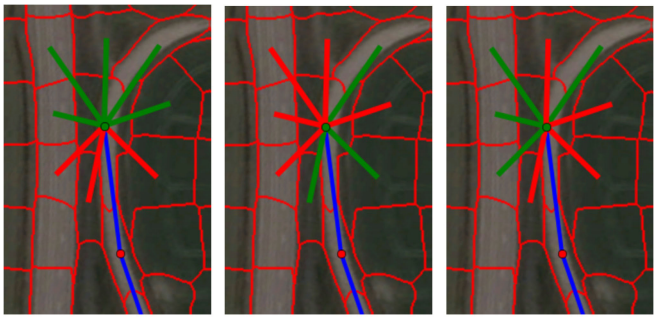


Fig. 9. Cost functions, from left to right: angle cost, boundary cost, region cost. Red: impossible to extend the road. Green: may extend the road. Blue: the tracked road information.

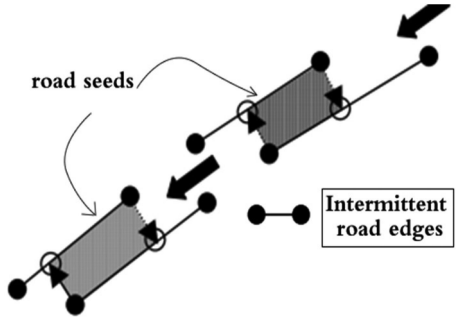


Fig. 10. Road seeds composed of edge fragments.

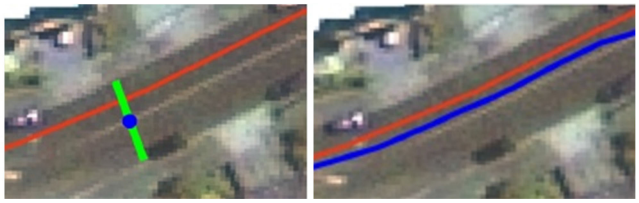
random forest classifier, which assigned each superpixel a unary road likelihood. Next, superpixels with high road likelihoods were sampled randomly as seed nodes and linked by minimum cost paths. Seppke *et al.* [85] proposed a superpixel-based road extraction algorithm. The algorithm cut the image into superpixels using SLIC and constructs a region shell (RS) graph, which held not only the geometry of the superpixels but also the boundaries (shells) between them. Then, the algorithm tracked roads starting from the user-provided seeds and directions in accordance with the cost functions [24] including an angular cost function, a boundary cost function and a regional cost function. The angular cost function punished the situation where the road changes drastically. The boundary cost function was used to judge the adjacency of two superpixels based on the RS graph. The region cost function was used to determine which neighboring superpixels are similar to the starting superpixel. The tracking algorithm judged the most likely next block by combining the three cost functions, as depicted in Fig. 9. Wegner *et al.* [86] developed a higher-order CRF formulation for road extraction based on superpixel segmentation, in which the prior is represented by long-range cliques with robust P^N-Potts potentials. To overcome the drawback of too many irrelevant cliques, minimum-cost paths are embedded in a higher-order CRF framework in order to construct an explicit prior about the shape of roads.

Hierarchical graph-based image segmentation for road extraction was described in [87], which consisted of a graph representation of initial segmentation and hierarchical merging and splitting of these segments based on color and shape features. Cem and Beri [88] proposed a road extraction method based on probability and graph theory, which consisted of three modules: road center detection, road shape extraction, and graph theory-based road network formation. In the third module, the detected road center pixels were represented in a parametric curve form, which allowed the refining of some road segments due to their shape and neighborhood conditions. An unsupervised framework to identify buildings and roads from VHR was proposed in [7]. First, the buildings, shadow, vegetation, and others were classified by a four-label graph optimization. Then, the regions that might belong to the road were extracted. Finally, a graph optimization was used again to characterize the regions belonging to buildings and roads.

Recently, in order to improve classification accuracy, the combination of different classifiers has become a hot research topic. The essence is to train different classification models by the same samples. The classification results are achieved by combining the outputs of these classifiers [55], [89].

2) Edge Analysis Methods: According to Marr's vision theory, the road recognition process can be divided into three levels: road edges detecting, pairing and connecting in the low level, road feature information processing in the middle level and road object recognition in the high level. Inspired by this idea, the edge-based methods detect (generally by Canny) the road edges initially and then remove the false positives in accordance with the predefined road model. Then, the parallel edges, also called road primitives, are paired into road seeds; finally, these road seeds are connected to road segments according to prior [90], as depicted in Fig. 10. Baumgartner *et al.* [67] detect roads based on the extraction of edges in a high-resolution image and the extraction of lines in the image with reduced resolution. Roadsides are generated using both resolution levels and explicit knowledge about roads. Then, roadsides are used to construct road parts and intersections. In [91], the Canny edge detection algorithm was applied on R, G, B, and NIR band separately and the union of the edges on these four edge maps was obtained. Then, parallel-line pairs were found. These methods are more robust than the other methods because they neglect the radiation characteristics of different road materials and only depend on the gradients. However, given the excessive dependence on the road edges, the performance of these methods relies on the outputs of the road edge detection and is limited by the complexity of RS images. Moreover, automatically setting the parameters of the edge detection algorithm is difficult. Therefore, these methods are only suitable for extracting simple roads, such as main roads.

3) Map-Based Methods: The coordinates of most roads in the world are already available in several map databases. The famous free map called OpenStreetMap (OSM) is used by most researchers, which has covered more than 88% of the urban roads globally [92]. Over 2 million registered users in the OSM project collect data using the manual survey. However, OSM roads are not very accurate because they are contributed by volunteers without quality control, and some of them are computed from GPS trajectories. The misalignment of road coordinates motivated the study of map-based road centerline extraction (or refinement) methods [23], [93], [94]. These methods



(a) (b)

Fig. 11. Road centerline extraction based on map matching. (a) Calculated road width and road center point. (b) Adjusted road centerline.

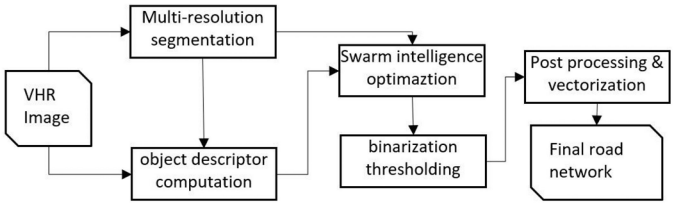


Fig. 12. Overview of the road extraction based on swarm intelligence optimization.

generally have two objectives: one is to refine the road coordinates combining the road map and the local texture, and the other is to extract the road profile. These methods locate the rough road positions using the road map initially and then adjust the coordinates to the road center points or extract the road shape exploiting the texture of the local region. Fig. 11 demonstrates the road centerline extraction process [92]. In Fig. 11(a), the red line is a road curve generated according to the road reference. The green stripe is the actual road width estimated based on the local intensity, and the blue dot is the exact center point of the road obtained using the algorithm. Fig. 11(b) exhibits the adjusted road centerline, which is marked by the blue line.

4) Swarm Intelligence-Based Methods: Bio-inspired swarm intelligence is an evolutionary computing technique that has attracted considerable interest over the past two decades [95]. Various biomimetic swarm intelligence technologies including particle swarm optimization, ant colony optimization (ACO), artificial bee colony and firefly algorithm, etc., were invented to solve the problems of optimization. These algorithms aim to solve certain problems by simulating the behavior of real biological populations. In realistic intelligent groups, a simple intelligent individual can solve complex problems beyond itself. Swarm intelligence algorithms perform well in solving discrete and network-based problems; in addition, the parallel processing characteristics of swarm intelligence pave the way for efficient target detection [96]. Maboudi *et al.* [97] introduced an algorithm for extracting road regions using ACO technology. First, multispectral data were used to segment the image for generating image objects; meanwhile, the feature vector of each object was calculated. Second, a directed graph was constructed regarding these objects as nodes and the migration costs (bidirectional difference) between two nodes as edges. Finally, the ants moved in the graph and aim to connect the nodes for generating a road map. The algorithm process is demonstrated in Fig. 12. Moreover, the author has further improved the abovementioned algorithm in the work [22].

5) Object-Based Methods: In RS imagery, the road probability of pixels is difficult to evaluate given the lack of context information. By contrast, if multiple adjacent pixels with similar features are merged into objects, then the spectrum, shape, texture, and other features of each object can be extracted more stably. In addition, more accurate road probabilities can be achieved by integrating these features [98]. In comparison with different object-oriented road extraction methods, the main differences between them are the methods of object segmenting, feature depicting, and the decision making to classify object types. Ding *et al.* [99] discovered that road regions typically have consistent local directions. The pixels were merged into the objects with similar main directions to calculate geometric measurements, such as linear feature indices and the area of these objects. Afterward, a segment-linking algorithm was used to extract road objects amongst them. In [22], an image with a very high resolution was smoothed to reduce the heterogeneity of a road surface. Then, the multiresolution segmentation algorithm was used to segment the image into many objects. Afterward, the road membership descriptor of each object was defined by a fuzzy inference system, which will be used in the ACO algorithm to distinguish the road objects in the following steps. In [74], the image was segmented into superpixels using ERS [75]; the grayscale, color (YUV space), and texture features of each superpixel were extracted, and the pixels with similar features were assembled into large objects through a graph-based algorithm based on the abovementioned features. Finally, the GMM algorithm was used to classify these large objects into several different types, as presented in Fig. 7. In order to detect the objects in different granularity, multiscale segmentation and clustering algorithms are often employed to extract road objects [98]. Huang and Zhang [100] proposed a road extraction method based on multiscale structural features of the objects. It classified different terrain objects using SVM, and recognized the road objects using majority voting machine. Finally, the road centerlines were extracted using connected area analysis. Article [101] constructed a point process, which was able to simulate and detect thin networks using an object-based approach. First, the road network was approximated by a configuration of connected segments. Then, the global minimum of the energy function was found by a simulated annealing technique based on a Monte Carlo dynamics (RJMCMC) for finite point processes. An object-based method was proposed in [102], which started from an initial segmentation obtained by fractal net evolution approach [103]. The road region of interest (RoadROI) was obtained after vegetation and shadows are removed. The RoadROI was used to construct a binary partition tree for representing the road-like regions.

6) Multispectral Segmentation Methods: Different from the road extraction methods based on the segmentation method described in Section IV-B1, the methods based on multispectral segmentation require multispectral images and even hyperspectral images support. A multispectral image is composed of pixels with multiband features of radiation. Generally, various types of object have different reflection intensities, which can be used to classify terrestrial objects into distinct catalogs [104]. Maboudi *et al.* [97] and Xue-Wen and Han-Qiu [105] analyzed the spectral

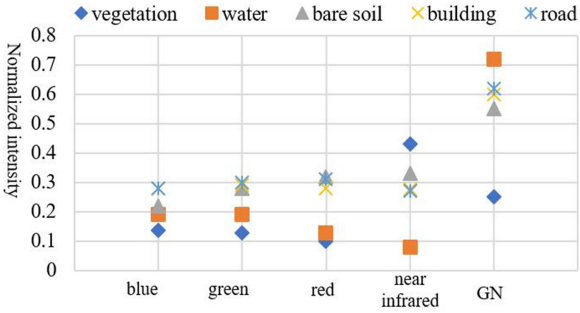


Fig. 13. Multispectral features of different objects in IKONOS images.

characteristics of water, vegetation, bare soil, buildings, cement roads, and asphalt roads in IKONOS satellite imagery and find that roads can be well distinguished from vegetation, water, and bare soil using GN feature (i.e., normalized the intensity difference between the green and the near-infrared bands). However, distinguishing between roads and buildings is difficult, as displayed in Fig. 13. Usually, the object shape features and the Gabor wavelet transform are used after multispectral segmentation to distinguish between roads and buildings further. Thus, using the multispectral features, the unconcerned objects can be excluded efficiently. However, these methods are not applicable in panchromatic or RGB images. Moreover, most of the aerial images are RGB images. In addition, images captured by different satellite sensors have different numbers of bands, and the spectral features of the same objects in different satellite images may be different, which further decreases the versatility of this method.

V. Data-Driven Road Extraction Methods

The data-driven methods mentioned in this work mainly refer to learning-based methods. These methods have been used for road detection in the early years [106], [107]. Although some promising results have been achieved by these approaches, they still failed to scale up to large challenging datasets that suffered from the small size of the neural network and the absence of big data [108]. In 2006, Hinton published a famous paper [109] in *Science*, in which he verified the following points: first, multihidden layer neural network has an excellent ability of feature representation; second, the difficulty of training deep neural network can be effectively overcome by "layer-by-layer initialization." Since then, the neural network with multiple hidden layers are called deep neural network, and the machine learning based on deep neural network is called deep learning.

To the best of our knowledge, the first attempt of extract road from HRSI using deep neural network was conducted by Minh and Hinton [108]. They built a patch-based convolutional neural network with millions of learnable parameters with large receptive field to predict the class of pixels centered at the patch and achieved significant improvement of precision and recall. Experiments showed that deep neural network exhibits significant ability of feature representation. In this section, we describe several attempts for road extraction based on the deep neural network in recent years.

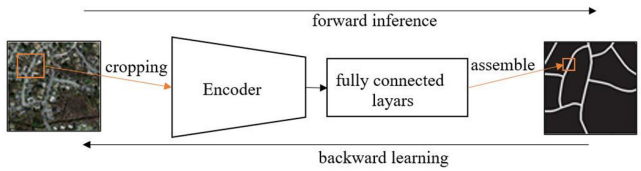


Fig. 14. Framework of methods based on patch-based DCNN.

A. Patch-Based DCNN Methods

It is infeasible to predict the class of a pixel directly because the features of a single pixel are insufficient for distinguishing its type. Therefore, the contextual information of a pixel is important to determine its membership. Mnih [110] first attempted to build a GPU-based, deep convolutional neural network, which is capable of exploiting a large image context as well as learning discriminative features. The DCNN is used to predict a small patch of labels from the same context. To utilize the correlation between neighbor pixels, the road probabilities of a small block of adjacent pixels were simultaneously predicted by the DCNN taking a larger patch as input. Thus, the classification accuracy is improved and the computation cost is reduced. Mnih [110] formulated the core problem as a probability distribution estimation, as expressed in (3), where $n(S, i, w_s)$ denotes a $w_s x w_s$ patch of the image S centered at the pixel i, S is the input image and $\tilde{M}$ is the predicted map of the model. Generally, w_m is less than w_s, especially when $w_m = 1$ means predicting the label of a pixel at a time. However, predicting the labels of a small block in one pass is more efficient, as defined in (4), and is more effective in handling noisy training data. The model predicted the road probability by sliding window with a specific stride (i.e., patches may overlap with each other). Finally, it assembled all the labeled patches to generate the global map of the entire image. The generic architecture of patched CNN models is illustrated in Fig. 14, where the encoder is a deep convolutional neural network based on any kind of backbone, such as VGG, ResNet, Inception, and MobileNet. The fully connected layers act as a linear discriminator. In the encoder, only the first layer is equipped with a max-pooling layer to prevent the loss of spatial information. In general, postprocessing is conducted for further reasoning pixels' classes, which enhances the smoothness of the predict results. Similarly, Saito *et al.* [6] designed a single patch-based DCNN for extracting roads and buildings from HRSI simultaneously followed by postprocessing to improve the performance

$$P\left(n\left(\tilde{M}, i, w_m\right) | n\left(S, i, w_s\right)\right) \tag{3}$$

$$P\ (\tilde{m}|s) = \prod_{i=1}^{w_m^2} P\left(\tilde{m}_i|s\right). \tag{4}$$

In the patch-based DCNN models, an image patch around a pixel is used as the input for model training and reasoning. These methods have several disadvantages: first, the prediction is time-consuming because the test image needs to be cut into numerous overlaid patches, which are inferred one by one followed by the reassembling of these patches to generate

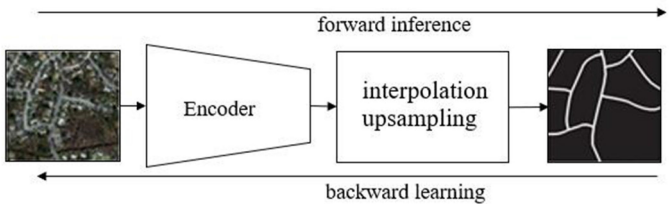


Fig. 15. Framework of methods based on FCN.

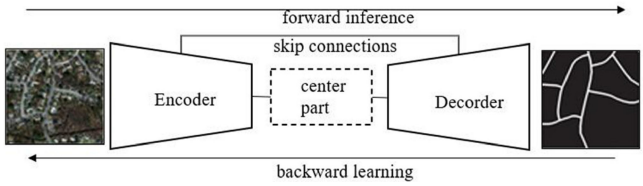


Fig. 16. Framework of methods based on DeconvNets.

the global segmentation. Second, it is inefficient in computing. The adjacent pixels are usually duplicated, and the patches are computed separately and repeatedly. Third, the small size of the patch limits the perspective field, but large size leads to poor efficiency of computation.

B. Fully Convolutional Network (FCN)-Based Methods

FCN predicts images end to end unlike the patch-based DCNN. They are implemented based on the CNN by replacing the last fully connected layers with convolutional layers. The final convolutional layer then outputs a map of the labels. In CNN models, the invariance of shifting, scaling, and distortion achieved by max pooling, weights sharing, and spatial subsampling guarantees the excellent performance in pattern recognition, e.g., the accuracy of document recognition of LeNet [111] exceeds 99%. Although CNN performs well in image recognition, the identification of a specific part of the image was still a worldwide problem until 2015 when the FCNs were proposed by Long [112]. FCN classifies the image at the pixel level and solves the problem of semantic segmentation. In contrast to the classic CNNs, which accept images with fixed size and output a fixed-length feature vector for classification using fully connected layers (full-connection layer + softmax layer), FCN accepts image of any size and up samples the output of the encoder (the front part of the FCN) to restore feature map to the same size as input image using the interpolation layers and the map presents the membership of each pixel whilst preserving its spatial information. The schematic network is shown in Fig. 15. In road extraction, road segmentation can be treated as semantic segmentation problem where roads and nonroads are labeled by 1 and 0, respectively.

Zhong *et al.* [113] applied FCN to extract roads and buildings in areal images. The paper showed that it is inappropriate to directly use the existing FCN pretraining model to extract roads or buildings, and found that the accuracy is significantly improved by adding the outputs of pooling layers into the final score layer, which suggests the direction of further research. The experiments also revealed that excessive pooling will hurt the performance of segmentation. After testing and adjusting the hyperparameters of the proposed network, the final road segmentation precision, recall, and intersection of union (IoU) on the Massachusetts road dataset (Mass. Roads) [110] are 0.71, 0.66, and 0.52, respectively.

C. DeconvNet-Based Methods

Since the advent of the FCN, the technology of semantic segmentation based on deep neural networks has become an independent research field. However, FCNs resort to *ad hoc* methods to up sample features that result in noisy predictions and also restrict the number of pooling layers in order to avoid too much upsampling and thus reduces spatial context [114]. DeconvNets are variant of FCN replacing the interpolation layers as deconvolutional layers (called decorder), such as SegNet [114], DeepLab [115], and U-Net [116]. The decoder helps map low-resolution feature maps at the output of the encoder stack to full input image size feature maps [114]. The generic architecture of DeconvNets is shown in Fig. 16, where the center part is an important component to capture multiscale features in many articles. In this section, we describe several attempts based on these models.

Wei *et al.* [9] built a road structure refined CNN (RSRCNN) for road extraction based on VGG encoding road structure in the cross-entropy loss. RSRCNN achieved 0.662 $F1$-score on Massachusetts road dataset. Panboonyuen *et al.* [117] presented an enhanced deep convolutional encoder-decoder network incorporation of exponential linear unit for road segmentation in aerial images taking SegNet as backbone, which obtained 0.857 $F1$-score. Mattyus *et al.* [118] developed a variant of FCN using ResNet as an encoder with a fully deconvolutional decoder to estimate road topology from aerial images directly. Mosinska *et al.* [119] introduced an iterative refinement method to extract topologies based on U-Net and proposed a novel topological loss term considering that pixel-wise losses alone are unsuitable to reflect the topological impact of prediction mistakes. Gao *et al.* [120] proposed an end-to-end framework called the multiple feature pyramid network similar to the RSRCNN. They exploited the multilevel semantic features of HRSI and designed a novel loss function to focus on the problem of unbalanced categories. Inspired by ResNet [121] and U-Net [116], Zhang *et al.* [15] designed a deep residual U-Net for road extraction, which combined the strengths of residual learning and U-Net. The network consisted many residual units, and the architecture was similar to U-Net with rich skip connections. The residual units make the deep network easy to be trained, and the skip connections facilitated the information propagation and reduced the number of trainable parameters. The input images were cropped from the original RS images with an overlap of 14 pixels. The global map was reassembled by stitching all the submaps, and the values in the overlap parts were averaged. The experiments showed that the breakeven point of precision and recall curve on Mass. Roads was 0.9187, which surpassed the state-of-the-art methods, although the parameters are only one-fourth of U-Net. Inspired by DenseNet [122] and U-Net, Xu *et al.* [123] proposed a GL-Dense-U-Net for extracting roads from aerial imagery. The authors introduced the local and global

information to focus on the local texture and morphological structure of roads, respectively. The network was designed based on DenseNet and U-Net because DenseNet has fewer trainable parameters, thereby making it easy to be trained. Meanwhile, U-Net has an elegant architecture with convenient information propagation. Thus, the GL-Dense-U-Net consists of two parts, namely, contracting and expanding, similar to encoder and decoder. These parts were connected by the local attention units to maximize the features in different stages. The global attention units in the expansive part extracted the road information when recovering the map from deep-level features. In this work, edge enhancement was conducted to reduce the noise and computation complexity. This method achieved remarkable performance (i.e., 0.9572 $F1$-score) on a special, relatively small dataset [124].

In the public road extraction competition, methods based on deconvolutional network are most commonly used. D-LinkNet [125], the best solution in Deepglobe-2018 [126], was a UNet-like network combined with a dilation part in the center. The dilation part contains dilated convolutions both in cascade mode and parallel mode. The receptive field of each path is different. So, the network can capture multiscale features. A modified UNet was used in the second-best solution [127]. It used ResNet-34 pretrained on ImageNet as its encoder and decoder adapted from vanilla U-Net. Its loss function was composed by binary cross entropy and IoU. The other solutions in the leaderboard of Deepglobe-2018 almost used the encoder-decoder architecture with skip connections, like UNet. The power of the deconvolutional network can also be confirmed by the top 3 solutions of SpaceNetChallange-2018 [128]. They all used the UNet-like network structure to predict the road regions in the first stage.

Inspired by the success of D-LinkNet, the encoder-decoder architecture with dilated convolutions is further applied in later researches [129]–[133]. A road extraction method using weakly-supervised learning was proposed in [132] for road extraction at global scale. The authors found that the trained model is highly specific to particular regions and produce worse in those regions, which did not adequately represented in the training set. So, they use the data from OSM to train more generalizable models. The dataset covers more than 700 000 square miles consisting of about 1.8 million tiles spanning six continents, which is 1000 times the area covered by the DeepGlobe dataset. The model was designed based on D-LinkNet. The batch normalization layers were replaced by group normalization layers, if not, the training fails to converge. Finally, they found that more regions in the training set results in better performance in prediction. Inspired by [134], Wu *et al.* [133] presented a road extraction method using weakly labeled data. The scheme combined graph cut theory and deep learning technique in the loss function. The loss function consisted of two parts: a partial loss used to reflect the feature of the pixels labeled road or nonroad, and a regularized loss used to reflect the similarity between these pixels.

D. Generative Adversarial Nets (GANs)-Based Methods

GANs are initially designed by Goodfellow *et al.* [135], which have received considerable attention in recent years. GANs consist of two parts, namely, a generator and a discriminator. Throughout the training process, the generator strives to make the generated images realistic, and the discriminator aims to distinguish the forged images from the real ones. In the end, the two parts finally achieve a dynamic equilibrium, and the generator can be used to segment images. Fig. 17 shows the architecture of GANs used in road extraction, where the generator produces the road segmentation map, the discriminator takes the segmentation map or ground truth map as the input, combined with the RGB image, to decide the probability that the label map comes from a true map or a generated map.

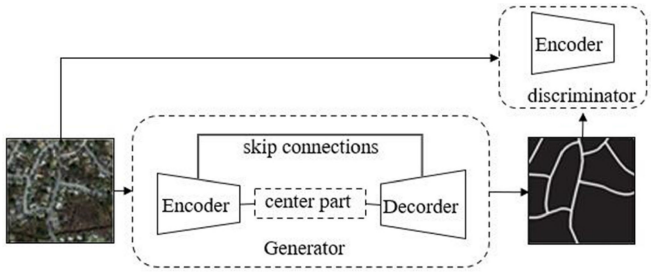


Fig. 17. Framework of methods based on GANs.

Inspired by [136], which confirmed that the GANs can enforce the continuity of spatial labels and generate accurate and smooth results versus nonadversarial training, Shi *et al.* [137] used a novel end-to-end conditional GAN for road region extraction by optimizing a structure loss. In this work, SegNet was adopted to build the generator part to output smooth segmentation maps containing higher-order spatial consistency and detailed boundary information. The model presented high performance on a small dataset including 550 images, which were cut out of Google Maps and manually labeled by the author. Costea *et al.* [138] argued that the road map must be recognized by high-level information, such as road graphs. To achieve this goal, they assembled a dual-hop GAN with two conditional GANs cascaded with each other. The first GAN detected the road region map according to the RGB image. The second GAN extracted the intersections using the road region map and the source image. After the road regions and interactions were extracted, an optimal road graph was identified by applying a smooth-based graph optimization [139] in the postprocess. The network was evaluated on a large dataset of European roads and outperforms the state-of-the-art methods.

E. Graph-Based Methods

Most of the above-mentioned methods are based on pixel-level segmentation. These segmentation-based methods may be suffered from the noisy CNN outputs, which are difficult to be corrected in postprocess stage [140]. Moreover, the pixelwise classification supervision leads to road networks with fragmented road segments and poor connectivity [141]. On the other hand, the vectorized representation of road maps (road graph) is more convenient in real-world applications, such as road mapping and driving navigation. Thus, graphed-based methods are getting more attention. Currently, there are two strategies to

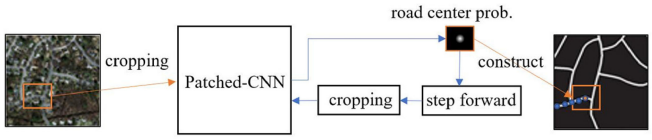


Fig. 18. Framework of road tracking methods (taking iterative road center detection strategy as an example).

generate road graphs, i.e., iterative road tracking and polygon detection.

The road tracking methods usually iteratively construct the road graphs according to the output (the next road point or marching direction) of a patched CNN, as shown in Fig. 18. Ventura *et al.* [142] designed a patched CNN to predict the local connectivity amongst the central pixel of an input patch and its border points. By iterating this local connectivity, they swept the whole image and inferred the global topology of the road network. Bastani *et al.* [140] proposed a new method called RoadTracer to extract accurate road network directly from aerial images using an iterative search process guided by a patch-based CNN decision function. Inspired by [142] and [140], Lian and Huang [143] presented a road network tracking algorithm. The key component is a road center points estimation DCNN model. The DCNN model predicts the road center points iteratively during the tracking process. VecRoad, a point-based iterative graph exploration scheme with segmentation-cues guidance and flexible step, was proposed in [144]. They used the framework of RoadTracer and proposed several schemes to improve the performance of road graph construction. The location of the next moving was represented as a "point" that unified the representation of multiple constraints such as direction and step size in each moving step. Moreover, the informative cues such as road segmentation and road junctions were jointly detected and utilized to guide the next moving, which results in better alignment of roads.

Some researchers argued that a road segment can be regarded as a polygon in a local image patch. Therefore, some methods extracted roads using polygon detection, such as active contour models and topology tracking. Most of the existing algorithms run in iterative mode. Directly predicting polygons from image is a relatively new research direction. Recent approaches such as PolygonRNN [145] and PolygonRNN++ [146] addressed the problem of polygon prediction directly using deep neural network. The approaches consisted of an RNN coupled with a CNN. The CNN encoder produces image features that are used to predict the first vertex, which is fed to the recurrent decoder. The RNN exploits the visual attention at each time step to predict other polygon vertices and generates polygons. Inspired by PolygonRNN, Li *et al.* [147] proposed the PolyMapper to construct roads in a vector representation directly from aerial images. The approach interpreted the topology of roads as a simply connected maze and followed the principle of a maze solving algorithm to reformulate the topology of roads. The PolyMapper also consisted of two parts: the CNN part and the RNN part. The CNN part was a VGG-16 to extract a set of skip features based on an image patch, and the RNN part predicted the potential position $P(y_{t+1}|y_t, y_{t-1}, y_0)$ of y_{t+1} polygon vertex according to the output of the patched CNN part. Unlike PolygonRNN, the PolyMapper is a fully automated, end-to-end learnable approach. Belli and Kipf [148] presented a generative graph transformer (GGT), a deep autoregressive model based on self-attention. The GGT model was designed for the recurrent, conditional generation of graphs and consists of an encoder–decoder architecture. A gray-scale image was passed through the encoder, which produces a conditioning vector using a context attention mechanism on the previously generated node. The self-attentive decoder used the conditioning vector and the previously generated nodes to predict the next node in the graph. This sequential process incrementally generates the graph.

VI. Metrics

The most common used metrics for evaluating the road extraction methods are completeness, correctness, quality, precision, recall, $F1$-score [68], [73], [108], [149], and IoU [94], [113], [150], [151]. The completeness is the percentage of the reference network that lies within the buffer around the extracted road centerlines, which is defined as (5), where true positive (TP) indicates the road segments detected correctly, false positive (FP) represents the wrong parts, and false negative (FN) denotes the unextracted fragments. The correctness represents the percentage of the extracted road centerlines, which lie within the buffer around the reference network, as defined by (6). The quality is a measure of the "goodness" of the final result. It takes into account the completeness of the extracted data as well as its correctness, which is defined as (7). The $F1$-score is the harmonic mean of completeness and correctness, which is defined as (8). Recall and precision are usually used to evaluate the performance of segmentation methods. In the field of road extraction, they are regarded as relax completeness and correctness by some researchers [110], [152]. Some researchers also use them to measure the performance of road segmentation (a subtask of road extraction) in remote sensing images [124], just as they are used to evaluate the performance of semantic segmentation methods. They are also defined by (5) and (6), respectively, but TP means the number of the correctly segmented pixels, and FP and FN are the number of the wrongly classified pixels. The IoU used in road extraction describes the accuracy of road area segmentation, which is defined as (9), where the target is the road reference and the prediction is the predicted road segmentation. By the way, the mean IoU (mIoU) is infeasible to evaluate the road segmentation. Because roads in the HRSI are small objects, even if the algorithm predicts the entire image as the background, mIoU can be close to 50% while IoU is 0.

Because $F1$-score and IoU fail to adequately incentivize the creation of connected road networks, more evaluation metrics have been proposed. The TOPO metric simulates a car driving a certain distance from several seed locations, and compares the destinations that can be reached in the ground truth graph G with those that can be reached in the estimated graph G' in terms of precision and recall [140]. The SP metric can be found in [86], which randomly samples two points lying both on the true and the estimated road network, and checks whether the

TABLE I
COMPARISON OF DIFFERENT SEMIAUTOMATIC METHODS

Algorithm	Road feature	Preprocessing	Seed requirement	Time consumption	Disadvantages
Active contour	Intensity gradient	smoothen , Euclidean distance transform, morphologic transformation, centerline detection	1 seed in a certain distance or 1 line for a segment	Medium	Require more seeds, varied cost function and the local minimum
Dynamic programming	Intensity gradient, road probability	Sharpening, road probability estimation,	1 seed in a certain distance	Medium	Require more seeds, varied cost function
Geodesic path models	Intensity gradient, road probability	Edge enhancement, road probability estimation	2 seeds per segment	High	Extract road edges only
Template matching methods	Texture, Intensity gradient	Smoothen and enhancement, texture extraction	1-2 seeds and march direction (optional)	Low	Unsuitable for twisty roads

shortest path between the two points has the same length in both networks. The procedure repeats with different random points and records the percentages of correct, too short, too long, and infeasible paths, until these percentages have converged. The SP metric is hardly reproducible, so the connected road ratio (CRR) metric is proposed in [118] to capture the ratio of road segments, which are estimated without discontinuities. Inspired by CRR, The CONN metric is defined in [142] to measure the similarity of connectivity between G and G'. A novel metric [average path length similarity (*APLS*)] is introduced in the third SpaceNet competition [128] to measure the similarity between G and G'. The *APLS* is defined as (10), where N is the number of unique paths, whereas $L(a, b)$ is the length of the path (a, b). The sum is taken over all possible source (a) and target (b) nodes in the ground truth graph. The node a' denotes the node in the proposal graph closest to the location of ground truth node a. The APLS metric scales from 0 (poor) to 1 (perfect). StreetMover Distance was introduced in [148], which was computed as the optimal cost of moving the predicted proposal point cloud to the ground-truth target point cloud. The main benefits of the StreetMover distance are its interpretability, scalability, and invariance with respect to node permutation graph translations and rotations

$$\text{Completeness} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{5}$$

$$\text{Correctness} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{6}$$

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

$$F1-\text{score} = \frac{2*\text{Completeness}*\text{Correctness}}{\text{Completeness}+\text{Correctness}} \tag{8}$$

$$\text{IoU} = \frac{\text{target}\cap\text{prediction}}{\text{prediction}\cup\text{prediction}} \tag{9}$$

$$\text{APLS} = 1-\frac{1}{N}\sum\min\left\{1,\frac{|L(a,b)-L(a',b')|}{L(a,b)}\right\}. \tag{10}$$

VII. COMPARISONS

A. Comparison of Semiautomatic Methods

Table I displays the subjective comparison of different semiautomatic road extraction methods. The performance comparison of these algorithms is not presented because the performances usually can be improved by increasing the density and precision of the seeds.

Table I also indicates that most of the algorithms exploit similar features (e.g., gray, texture, and edge features) and some extracted features as auxiliary information. However, different algorithms require different degrees of manual intervention. The template matching methods request fewer seeds and achieve real-time performance. Therefore, these methods are extensively used. Shortest path methods are rarely used in practice because it is highly time-consuming, although they require the least seeds. Dynamic programming and snakes require additional seeds and have to define complex cost function according to specific scenarios, which leads to less application.

B. Comparison of Automatic Methods

Table II displays the subjective comparison of the different automatic road extraction methods. The performances of the algorithms are also not listed in Table II considering the lack of benchmark dataset. In Table II, each algorithm has its own advantages and disadvantages. It is difficult to obtain satisfactory results from large scale RS images using only one algorithm. Thus, research on integrating different methods is promising. In addition, high resolution presents detailed features but is accompanied by additional noise, such as cars, shadows, and occlusions. On this basis, many researchers have emphasized object features and structures rather than local intensity.

C. Comparison of Data-Driven Methods

Thanks to the open-source benchmark dataset, we can perform quantitative comparisons of the data-driven road extraction algorithms. Several benchmarks are available for evaluating road extraction methods. Mass. Roads dataset[1] is the most popular one, in which each pixel in an image is annotated as foreground or background. The spatial resolution of the images in the Mass. Roads dataset is 1 meter per pixel, and each image consists of 1500 × 1500 pixels. Specifically, the road dataset contains 1172 aerial images that cover more than 2600 km^2 in total [110]. Another popular benchmark is provided by the DeepGlobe Road

[1][Online]. Available: http://www.cs.toronto.edu/~vmnih/data/

TABLE II
COMPARISON OF DIFFERENT AUTOMATIC METHODS

No	Algorithm	Road features	Advantages	Disadvantages
1	Parallel-pairs	Edge, direct of line, width	Intuition	Sensitive to noise, only suitable for the main load, discontinuity
2	Map-based	Texture, gradient	Automation, precision	Require GPS support
3	Watershed	Intensity, gradient	Continuity, efficient	Over-segmentation
4	K-means clustering	Intensity	Simple, efficiency,	Difficult K presetting, sensitive to initial centroids, hard-clustering
5	Gaussian mixture model	Intensity, object features	Universality, soft-clustering	Difficult K presetting, relay on the object segmentation
6	Swarm intelligence	Intensity, spectral, structure	Robust, parallel	Relay on the object segmentation, local optimum
7	Object-based	Intensity, object features	Robust, efficient	Relay on the object segmentation, difficult feature selection

TABLE III
COMPARISON OF DIFFERENT DATA-DRIVEN METHODS ON MASSACHUSETTS ROAD DATASET

Algorithm	Advantages	Disadvantages	Methods	Precision	Recall	Fl
Patch-based DCNN	weights sharing, less parameter	Inefficiency, large-scale training samples requirement	Mnih et al. [110]	0.901	0.901	0.901
			Saito et al. [6]	0.905	0.905	0.905
			Ventura et al. [142]	0.835	0.808	0.816
			Alshehhi et al. [154]	0.917	0.917	0.917
FCN-based	Arbitrary image size, end to end training	Low fineness, low position accuracy, spatial, lack of spatial consistency	Zhong et al. [113]	0.710	0.660	0.684
DeconvNet-based	Arbitrary image size, end to end training, better fineness	High cost of computing and storage	Wei et al. [9]	0.606	0.729	0.662
			Panbooyuen et al. [117]	0.854	0.861	0.857
			Panbooyuen et al. [155]	0.858	0.894	0.876
			Zhang et al. [15]	0.919	0.919	0.919
			Gao et al. [120]	0.851	0.742	0.793
GAN-based	More consistent	non-convergence, gradient vanishing, and model collapse	Costea et al. [138][4]	0.841	0.841	0.841
			Shi et al. [137][5]	0.883	0.910	0.896
Graph-based	Higher connectivity	Complex graph reconstruction & optimization	Ventura et al. [142]	83.5	80.8	81.6
			Lian et al. [143]	82.3	82.7	82.5
			Batra et al. [141][6]	-	-	73.35
			Tan et al. [144][6]	-	-	73.69

Extraction Challenge,[2] which consists of 8570 images and spans a total land area of 2220 km^2. A total of 6626 images from this dataset are selected for training, 1243 images for validation, and 1101 images for testing. SpaceNet road dataset[3] is specifically built for the SpaceNet road detection and routing challenge, which consists of 8000 km of road centerlines with associated attributes, such as road type, surface type, and the number of lanes. This dataset first introduces the data-agnostic APLS metric to evaluate road network proposals. The dataset is divided into three parts as follows: 60% of the dataset is distributed for training, 20% for testing, and 20% for validating.

In this article, we use the Mass. Roads benchmark for performance comparison because it has the most performance reports, thereby allowing fair and comprehensive comparisons amongst various methods. It should be noted that we do not find the performance reports on the Mass. Roads dataset achieved by GAN-based methods. Thus, we present the performance tested on the other dataset. The evaluation metric is $F1$-score, which is a tradeoff between precision and recall. We present the scores reported in the original papers. For an approach with multiple variations in model architecture, we only show the score of the optimal model for the sake of brevity. To be consistent with the expression in this work, we classify these models as patched CNN, FCN, DeconvNet, GAN, and graph-based. The comparison of different neural network models is summarized in Table III, where the tags 4, 5, 6 denote Evaluated on the European Road Dataset, a private dataset cut out from Google Earth, and RoadTracer dataset, respectively. The patch-based CNN models use a large local context to infer the class of a pixel, thereby achieving the high performance (relax version). However, it is inefficient to reason each pixel ignoring the correlation between adjacent pixels. The FCNs achieve pixel-to-pixel reasoning, but the accuracy is low due to the simple interpolation used in the upsampling stage. The deconvolutional networks significantly improve spatial accuracy

[2][Online]. Available: http://deepglobe.org/
[3][Online]. Available: https://medium.com/the-downlinq

TABLE IV
PERFORMANCES ACHIEVED IN THE SAME DATASET BY THE HEURISTIC AND DATA-DRIVEN ALGORITHMS

Dataset	Description	Method type	Completeness	Correctness	Quality
Google Earth[7]	224 HR images with 600×600 pixels, with a spatial resolution of 1.2 m/pixel.	Heuristic	0.8509	0.6837	0.6471[100]
			0.8113	0.6735	0.6218[156]
			0.8486	0.8417	0.7463[5]
			0.8769	0.8394	0.7631[157]
		Data Driven	0.9487	0.9275	0.8963[124]
Google Earth	550 HR images with 360×480 pixels, with unfixed spatial resolution.	Heuristic	0.8312	0.7885	0.7642[100]
		Data Driven	0.9101	0.8831	0.8732[137]
Shaoshan City	2255 HR images with 1000×1000 pixels cut from Pleiades-1, with spatial resolution of 0.5m/pixel.	Heuristic	0.7848	0.8778	0.7075[18]
		Data Driven	0.8397	0.9568	0.8090[158]
Vaihingen[8]	33 HR patches of aerial image of less than 0.5 m/pixel provided by International Society for Photogrammetry and Remote Sensing (ISPRS)	Heuristic	0.621	0.495	0.379[159]
			0.694	0.750	0.556[86]
			0.662	0.833	0.583[74]
		Data Driven	0.774	0.832	0.647[4]

by adopting deconvolutional layers and utilizing the low-level convolutional output (called skip connections). Although the GANs can produce consistent output, the problems of model collapse, non-convergence, and gradient vanishing make the network training more complicated.

D. Comparison Between Heuristic and Data-Driven Methods

We compare the performances of heuristic algorithms and data-driven algorithms in the same dataset, see Table IV, where the tags 7 and 8 denote Cheng-TGRS-2018 and Vaihingen data. It can be found that data-driven methods achieve better performances, and the performances have been improved by more than 10% basically. Therefore, the data-driven methods are the mainstream in this field now. However, the integration of heuristic algorithms and deep learning is getting attention. A coarse-to-fine road extraction method was presented in [153]. They argued that intensity distribution and structure feature information are both crucial for road extraction task. In [153], a local Dirichlet mixture model (LDMM) was introduced to presegment images into roads and nonroad using intensity distribution at the coarse level. Then, based on the results of LDMM, a multiscale-high-order deep learning strategy was employed to further remove FP at the fine scale. The ablation experiments showed that the performance combining heuristic algorithms and deep learning is better than using any one alone.

VIII. CONCLUSION AND PROSPECTS

This article summarizes the methods extracting roads from HRSI in the past two decades, which are classified into heuristic and data-driven. For heuristic algorithms, we further divide the algorithms into two types (i.e., automatic and semiautomatic) and introduce the principles and inspirations of them. For data-driven methods, we focus on the methods based on deep learning, and survey how the different deep learning technologies and frameworks are applied in road extraction methods. Comprehensive comparisons are described inner and inter different type of methods. The hand-crafted feature engineering required in the heuristic algorithms decreases the algorithms' generalization, makes them difficult to apply to large-scale datasets. Comparisons of performances achieved by heuristic methods and data-driven methods in the same dataset show the advance of the data-driven method. This article cannot elaborate on all kinds of heuristic algorithms due to the diversity of them and fails to compare their performances objectively for the lack of benchmark dataset used by them.

In general, the existing road extraction algorithms are not smart enough to fulfill practical applications. Interventions, such as adjusting some parameters, specifying the road type or manually placing the seeds, etc., are required by most of the algorithms. Even by manual intervention, satisfactory results are sometimes difficult to obtain. In addition, different extraction methods have their own advantages and disadvantages. Therefore, the intelligence of road extraction for RS imagery still requires further study. Future research may continue from the following aspects.

1) *Methods based on large-scale neural network*: Experiments showed that deep neural networks are ideal solutions for pattern recognition. However, DNN models trained by one scenario may be inapplicable to the other scenes. If the neural network is designed reasonably and the training data are sufficient and sampled from all over the world, then the network should be able to perform well in most common scenes.
2) *Methods based on high-resolution DNN*: In the end-to-end process models, the backbone networks (such as ResNet and VGGNet) mainly acquire low resolution of representation, and the spatial discrimination of the restored high-resolution representation is insufficiently strong. We hold that it is necessary to study high-resolution deep neural network that can learn abundant high-resolution representations by maintaining high resolution throughout the whole process.
3) *Integration of heuristic algorithms and deep learning*: We believe that if a method can make full use of statistical features and learning features, it can achieve higher performance. There are two directions that can be developed.

One is weakly or semisupervised learning, which uses heuristic algorithms to generate initial labels, and the other uses heuristic algorithms to preprocess images and postprocess outputs of DCNN.

4) *Methods based on graph neural networks (GNNs)*: GNNs are deep learning based methods that operate on graph domain, which have been widely applied in recent years due to their convincing performance and high interpretability. The road network is essentially a graph composed of intersections (vertices) and road segments (edges). We believe that road extraction based on graph learning is a promising research topic.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, "Semi automatic road extraction from digital images," *Egypt. J. Remote Sens. Space Sci.*, vol. 20, no. 1, pp. 117–123, 2017.

[2] G. Panteras and G. Cervone, "Enhancing the temporal resolution of satellite-based flood extent generation using crowdsourced data for disaster monitoring," *Int. J. Remote Sens.*, vol. 39, no. 5, pp. 1459–1474, 2018.

[3] Z. Zhu, S. Yang, G. Xu, X. Lin, and D. Shi, "Fast road classification and orientation estimation using omni-view images and neural networks," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 1182–1197, Aug. 1998.

[4] J. Wang, J. Song, M. Chen, and Y. Zhi, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *Int. J. Remote Sens.*, vol. 36, no. 12, pp. 3144–3169, 2015.

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

[6] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks," *J. Imag. Sci. Technol.*, vol. 60, no. 1, pp. 10402–10409, 2016.

[7] A. O. Ok, "Automated extraction of buildings and roads in a graph partitioning framework," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 79–84, 2013.

[8] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *J. Traffic Transp. Eng. (Engl. Ed.)*, vol. 3, no. 3, pp. 271–282, 2016.

[9] Y. Wei, Z. Wang, and M. Xu, "Road structure refined CNN for road extraction in aerial image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 709–713, May 2017.

[10] T. Perciano, F. Tupin, R. Hirata Jr., and R. M. Cesar Jr., "A two-level Markov random field for road network extraction and its application with optical, SAR, and multitemporal data," *Int. J. Remote Sens.*, vol. 37, no. 16, pp. 3584–3610, 2016.

[11] R. P. Krish, J. Fierrez, D. Ramos, F. Alonso-Fernandez, and J. Bigun, "Improving automated latent fingerprint identification using extended minutia types," *Inf. Fusion*, vol. 50, pp. 9–19, 2019.

[12] H. S. Dadi and G. K. M. Pillutla, "Improved face recognition rate using HOG features and SVM classifier," *IOSR J. Electron. Commun. Eng.*, vol. 11, no. 4, pp. 34–44, 2016.

[13] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4651–4659.

[14] 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.

[15] 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.

[16] D. Marr, "Vision: A computational investigation into the human representation and processing of visual information," *Quart. Rev. Biol.*, vol. 8, pp. 9–37, 1982.

[17] Y. Trinder, C. John, and Wang, "Knowledge-based road interpretation in aerial images," *Int. Arch. Photogramm. Remote Sens.*, vol. 32, pp. 635–640, 1998.

[18] Y. Zang, C. Wang, L. Cao, Y. Yu, and J. Li, "Road network extraction via aperiodic directional structure measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3322–3335, Jun. 2016.

[19] H. Pan, Y. Jia, and Z. Lv, "An adaptive multifeature method for semi-automatic road extraction from high-resolution stereo mapping satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 201–205, Feb. 2019.

[20] A. Baumgartner, C. Steger, H. Mayer, W. Eckstein, and H. Ebner, "Automatic road extraction based on multi-scale, grouping, and context," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 7, pp. 777–785, 1999.

[21] S. Hinz and A. Baumgartner, "Automatic extraction of urban road networks from multiview aerial imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, pp. 83–98, 2003.

[22] M. Maboudi, J. Amini, S. Malihi, and M. Hahn, "Integrating fuzzy object based image analysis and ant colony optimization for road extraction from remotely sensed images," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 151–163, 2018.

[23] P. Manandhar, P. R. Marpu, and Z. Aung, "Segmentation based traversing-agent approach for road width extraction from satellite images using volunteered geographic information," *Appl. Comput. Inform.*, 2018, doi: 10.1016/j.aci.2018.07.004.

[24] V. Shukla, R. Chandrakanth, and R. Ramachandran, "Semi-automatic road extraction algorithm for high resolution images using path following approach," in *Proc. 3rd Indian Conf. Comput. Vis., Graph. Image Process.*, Ahmadabad, India, Dec. 2002, pp. 231–236.

[25] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[26] 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.*, 2010, pp. 263–266.

[27] Z. Zheng, W. Liang, and Y. Hu, "An improved balloon snake method for road contour extraction," in *Proc. Int. Conf. Comput. Appl. Syst. Model.*, 2010, pp. 227–231.

[28] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes," *Mach. Vis. Appl.*, vol. 12, no. 1, pp. 23–31, 2000.

[29] 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.

[30] I. Laptev, "Road extraction based on snakes and sophisticated line extraction," M.S. thesis, Dept. Numer. Anal. Comput. Sci., Roy. Inst. Technol., Stockholm, Sweden, 1997.

[31] Y. Hu and K. J. Zu, "Road extraction from remote sensing imagery based on road tracking and ribbon snake," in *Proc. Pac.–Asia Conf. Knowl. Eng. Softw. Eng.*, 2010, pp. 201–204.

[32] A. Maarir and B. Bouikhalene, "Roads detection from satellite images based on active contour model and distance transform," in *Proc. 13th Int. Conf. Comput. Graph., Imag. Vis.*, 2016, pp. 94–98.

[33] W. Wang, H. Li, K. Wang, C. He, and M. Bai, "Pavement crack detection on geodesic shadow removal with local oriented filter on LOF and improved level set," *Construction Building Mater.*, vol. 237, no. 375, 2020, Art. no. 117750, doi: 10.1016/j.conbuildmat.2019.117750.

[34] X. Niu, "A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model," *ISPRS J. Photogramm. Remote Sens.*, vol. 61, no. 3/4, pp. 170–186, 2006.

[35] Z. Miao, B. Wang, W. Shi, and H. Zhang, "A semi-automatic method for road centerline extraction from VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1856–1860, Nov. 2014.

[36] L. D. Cohen, "Multiple contour finding and perceptual grouping using minimum paths," *J. Math. Imag. Vis.*, vol. 14, no. 3, pp. 225–236, 2001.

[37] X. G. Sun, M. C. Li, Y. X. Liu, W. Liu, and L. Tan, "A semi-automation road extraction approach based on fast marching method and mean shift algorithm," in *Proc. Global Congr. Intell. Syst.*, 2009, pp. 355–359.

[38] K. Yang, M. Li, Y. Liu, and C. Jiang, "Multi-points fast marching: A novel method for road extraction," in *Proc. 18th Int. Conf. Geoinform.*, 2010, pp. 1–5.

[39] Z. Miao *et al.*, "Use of colour transformation and the geodesic method for road centreline extraction from VHR satellite images," *Int. J. Remote Sens.*, vol. 40, no. 10, pp. 4043–4058, 2019.

[40] D. H. Ballard and C. M. Brown, *Computer Vis.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1982.

[41] A. Gruen and H. Li, "Road extraction from aerial and satellite images by dynamic programming," *ISPRS J. Photogramm. Remote Sens.*, vol. 50, no. 4, pp. 11–20, 1995.

[42] A. P. D. Poz and G. M. Do Vale, "Dynamic programming approach for semi-automated road extraction from medium-and high-resolution images," *ISPRS Arch.*, vol. 34, no. 3, pp. 87–92, 2003.

[43] A. P. D. Poz, R. A. B. Gallis, and J. F. C. da Silva, "Three-dimensional semiautomatic road extraction from a high-resolution aerial image by dynamic-programming optimization in the object space," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 796–800, Oct. 2010.

[44] E. Türetken, F. Benmansour, B. Andres, H. Pfister, and P. Fua, "Reconstructing loopy curvilinear structures using integer programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1822–1829.

[45] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "Road networks as collections of minimum cost paths," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 128–137, 2015.

[46] S.-R. Park and T. Kim, "Semi-auto road extraction algorithm from IKONOS images using template matching," in *Proc. 22nd Asia Conf. Remote Sens.*, Singapore, 2001, pp. 1209–1213.

[47] X. Lin, J. Shen, and Y. Liang, "Semi-automatic road tracking using parallel angular texture signature," *Intell. Autom. Soft Comput.*, vol. 18, no. 8, pp. 1009–1030, 2012.

[48] G. Fu, H. Zhao, C. Li, and L. Shi, "Road detection from optical remote sensing imagery using circular projection matching and tracking strategy," *J. Indian Soc. Remote Sens.*, vol. 41, no. 4, pp. 819–831, 2013.

[49] X. Lin, J. Zhang, Z. Liu, and J. Shen, "Semi-automatic extraction of ribbon roads from high resolution remotely sensed imagery by T-shaped template matching," in *Proc. Geoinform. Joint Conf. GIS Built Environ., Classification Remote Sens. Images*, 2008, vol. 7147, pp. 168–175, doi: 10.1117/12.813220.

[50] J. B. 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.

[51] M. Mokhtarzade and M. J. V. Zoej, "Road detection from high-resolution satellite images using artificial neural networks," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 1, pp. 32–40, 2007.

[52] A. Kirthika and A. Mookambiga, "Automated road network extraction using artificial neural network," in *Proc. Int. Conf. Recent Trends Inf. Technol.*, 2011, pp. 1061–1065.

[53] N. Yager and A. Sowmya, "Support vector machines for road extraction from remotely sensed images," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2003, pp. 285–292.

[54] M. Song and D. Civco, "Road extraction using SVM and image segmentation," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 12, pp. 1365–1371, 2004.

[55] A. Abdollahi, H. R. R. Bakhtiari, and M. P. Nejad, "Investigation of SVM and level set interactive methods for road extraction from Google Earth images," *J. Indian Soc. Remote Sens.*, vol. 46, no. 3, pp. 423–430, 2018.

[56] C. Simler, "An improved road and building detector on VHR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 507–510.

[57] B. Yousefi, S. M. Mirhassani, and H. Marvi, "Classification of remote sensing images from urban areas using Laplacian image and Bayesian theory," *Proc. SPIE*, vol. 6718, 2007, Art. no. 67180F.

[58] G. Storvik, R. Fjortoft, and A. H. S. Solberg, "A Bayesian approach to classification of multiresolution remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 539–547, Mar. 2005.

[59] H. Cai and G. Yao, "Optimized method for road extraction from high resolution remote sensing image based on watershed algorithm," *Remote Sens. Land Resour.*, vol. 25, no. 3, pp. 25–29, 2013.

[60] F. Meyer, "Color image segmentation," in *Proc. Int. Conf. Image Process. ITS Appl.*, 1992, pp. 303–306.

[61] S. Xu, H. Liu, and E. Song, "Marker-controlled watershed for lesion segmentation in mammograms," *J. Digit. Imag.*, vol. 24, no. 5, pp. 754–763, Oct. 2011, doi: 10.1007/s10278-011-9365-2.

[62] K. Parvati, B. S. Prakasa Rao, and M. Mariya Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation," *Discr. Dyn. Nature Soc.*, vol. 2008, 2008, Art. no. 384346, doi: 10.1155/2008/384346.

[63] R. Gaetano, G. Masi, G. Poggi, L. Verdoliva, and G. Scarpa, "Marker-controlled watershed-based segmentation of multiresolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2987–3004, Jun. 2015.

[64] Z. Wu, Z. Hu, and Q. Ouyang, "A regional adaptive segmentation algorithm for remote sensing image," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 36, no. 3, pp. 293–296, 2011.

[65] L. Li and X. Zhang, "A quickly automatic road extraction method for high-resolution remote sensing images," *Geomatics Sci. Technol.*, vol. 3, pp. 27–33, 2015.

[66] Y. Shao, B. Guo, X. Hu, and L. Di, "Application of a fast linear feature detector to road extraction from remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 626–631, Sep. 2011.

[67] A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein, "Multi-resolution, semantic objects, and context for road extraction," in *Proc. Semantic Model. Acquisition Topograph. Inf. Images Maps*, 1997, pp. 140–156.

[68] M. Rajeswari, K. S. Gurumurthy, L. P. Reddy, S. N. 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.

[69] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[70] R. Li and F. Cao, "Road network extraction from high-resolution remote sensing image using homogenous property and shape feature," *J. Indian Soc. Remote Sens.*, vol. 46, no. 1, pp. 51–58, 2018.

[71] Z. Lv, Y. Jia, Q. Zhang, and Y. Chen, "An adaptive multifeature sparsity-based model for semiautomatic road extraction from high-resolution satellite images in urban areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1238–1242, Aug. 2017.

[72] J. Zhang, L. Chen, Z. Li, W. Geng, and C. Wang, "Multiple saliency features based automatic road extraction from high-resolution multispectral satellite images," *Chin. J. Electron.*, vol. 27, no. 1, pp. 133–139, 2018.

[73] R. Maurya, P. R. Gupta, and A. S. Shukla, "Road extraction using K-means clustering and morphological operations," in *Proc. Int. Conf. Image Inf. Process.*, 2011, pp. 1–6.

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

[75] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.

[76] Z. Miao, B. Wang, W. Shi, H. Wu, and Y. Wan, "Use of GMM and SCMS for accurate road centerline extraction from the classified image," *J. Sensors*, vol. 2015, 2015, Art. no. 784504, doi: 10.1155/2015/784504.

[77] U. Ozertem and D. Erdogmus, "Locally defined principal curves and surfaces," *J. Mach. Learn. Res.*, vol. 12, no. 4, pp. 1249–1286, 2008.

[78] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, vol. 1, pp. 10–17.

[79] A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[80] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 705–718, doi: 10.1007/978-3-540-88693-8_52.

[81] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 211–224.

[82] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

[83] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[84] Y. Zhang, R. Hartley, J. Mashford, and S. Burn, "Superpixels via pseudo-Boolean optimization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2012, pp. 1387–1394.

[85] B. Seppke, L. Dreschler-Fischer, and C. Wilms, "Robust superpixel based segmentation for urbanization monitoring by means of spectral remote sensing images," in *Proc. DLT DGPF Symp.*, 2016.

[86] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "A higher-order CRF model for road network extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1698–1705.

[87] R. Alshehhi and P. R. Marpu, "Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 126, pp. 245–260, 2017.

[88] 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.

[89] U. Çinar, E. Karaman, E. Gedik, Y. Yardımcı, and U. Halıcı, "A new approach to automatic road extraction from satellite images using boosted classifiers," *Proc. SPIE*, vol. 8537, 2012, Art. no. 85370O.

[90] A. P. Dal-Poz, G. M. Do Vale, and R. B. Zanin, "Automatic extraction of road seeds from high-resolution aerial images," *An. Acad. Bras. Cienc.*, vol. 77, no. 3, pp. 509–520, 2005.

[91] E. Karaman, U. Cinar, E. Gedik, Y. Yardemci, and U. Halici, "A new algorithm for automatic road network extraction in multispectral satellite images," in *Proc. 4th GEOBIA*, 2012, pp. 455–459.

[92] C. Cao and Y. Sun, "Automatic road centerline extraction from imagery using road GPS data," *Remote Sens.*, vol. 6, no. 9, pp. 9014–9033, 2014, doi: 10.3390/rs6099014.

[93] J. B. Mena and J. A. Malpica, "An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery," *Pattern Recognit. Lett.*, vol. 26, no. 9, pp. 1201–1220, 2005.

[94] G. Mattyus, S. Wang, S. Fidler, and R. Urtasun, "Enhancing road maps by parsing aerial images around the world," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1689–1697.

[95] X. S. Yang, Z. Cui, R. Xiao, A. H. Gandomi, and M. Karamanoglu, *Swarm Intelligence and Bio-Inspired Computation*. New York, NY, USA: Elsevier, 2013.

[96] A. Delévacq, P. Delisle, M. Gravel, and M. Krajecki, "Parallel ant colony optimization on graphics processing units," *J. Parallel Distrib. Comput.*, vol. 73, no. 1, pp. 52–61, 2013.

[97] M. Maboudi, J. Amini, M. Hahn, and M. Saati, "Object-based road extraction from satellite images using ant colony optimization," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 179–198, 2017.

[98] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, "A comparison of three image-object methods for the multiscale analysis of landscape structure," *ISPRS J. Photogramm. Remote Sens.*, vol. 57, no. 5, pp. 327–345, 2003.

[99] L. Ding, Q. Yang, J. Lu, J. Xu, and J. Yu, "Road extraction based on direction consistency segmentation," in *Proc. Chin. Conf. Pattern Recognit.*, 2016, pp. 131–144.

[100] 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, pp. 1977–1987, 2009.

[101] R. Stoica, X. Descombes, and J. Zerubia, "A Gibbs point process for road extraction from remotely sensed images," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 121–136, 2004.

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

[103] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3/4, pp. 239–258, 2004.

[104] P. Mitra, B. Uma Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, Jul. 2004, doi: 10.1016/j.patrec.2004.03.004.

[105] W. U. Xue-Wen and X. U. Han-Qiu, "Level set method major roads information extract from high-resolution remote-sensing imagery," *J. Astronaut.*, vol. 31, no. 5, pp. 1495–1502, 2010.

[106] U. Bhattacharya and S. K. Parui, "An improved backpropagation neural network for detection of road-like features in satellite imagery," *Int. J. Remote Sens.*, vol. 18, no. 16, pp. 3379–3394, 1997.

[107] J. E. Boggess, "Identification of roads in satellite imagery using artificial neural networks: A contextual approach," Comput. Sci. Dept., Mississippi State Univ., Starkville, MS, USA, 1993.

[108] 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.

[109] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi: 10.1126/science.1127647.

[110] V. Mnih, "Machine learning for aerial image labeling," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.

[111] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, Nov. 1998.

[112] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

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

[114] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[115] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *Comput. Sci.*, vol. 4, pp. 357–361, 2014.

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

[117] T. Panboonyuen, P. Vateekul, K. Jitkajornwanich, and S. Lawawirojwong, "An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery," in *Proc. Int. Conf. Comput. Inf. Technol.*, 2017, pp. 191–201.

[118] G. Mattyus, W. Luo, and R. Urtasun, "DeepRoadMapper: Extracting road topology from aerial images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3458–3466.

[119] A. Mosinska, P. Marquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3136–3145.

[120] X. Gao *et al.*, "An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network," *IEEE Access*, vol. 6, pp. 39401–39414, 2018.

[121] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 770–778.

[122] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 11–19.

[123] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, pp. 1461–1476, 2018, doi: 10.3390/rs10091461.

[124] 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.

[125] 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 Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 192–196.

[126] "DeepGlobe," 2018. [Online]. Available: http://deepglobe.org/

[127] A. Buslaev, S. Seferbekov, V. Iglovikov, and A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2018, pp. 197–200.

[128] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "SpaceNet challenge: Road extraction and routing." [Online]. Available: https://spacenetchallenge.github.io/Challenges/Challenge-3.html, Accessed: Sep. 20, 2020.

[129] 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, no. 9, pp. 1015–1026, 2019.

[130] L. Gao, W. Song, J. Dai, and Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network," *Remote Sens.*, vol. 11, no. 5, pp. 552–567, 2019, doi: 10.3390/rs11050552.

[131] A. Wulamu, Z. Shi, D. Zhang, and Z. He, "Multiscale road extraction in remote sensing images," *Comput. Intell. Neurosci.*, vol. 2019, 2019, Art. no. 2373798, doi: 10.1155/2019/2373798.

[132] D. Bonafilia, J. Gill, S. Basu, and D. Yang, "Building high resolution maps for humanitarian aid and development with weakly-and semi-supervised learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1–9.

[133] S. Wu, C. Du, H. Chen, Y. Xu, N. Guo, and N. Jing, "Road extraction from very high resolution images using weakly labeled OpenStreetMap centerline," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 11, pp. 478–496, 2019.

[134] M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov, "On regularized losses for weakly-supervised CNN segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 507–522, doi: 10.1007/978-3-030-01270-0_31.
[135] I. J. Goodfellow *et al.*, "Generative adversarial Nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[136] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," in *Proc. NIPS Workshop Adversarial Training*, 2016.
[137] Q. Shi, X. Liu, and X. Li, "Road detection from remote sensing images by generative adversarial networks," *IEEE Access*, vol. 6, pp. 25484–52494, 2017.
[138] D. Costea, A. Marcu, M. Leordeanu, and E. Slusanschi, "Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2017, pp. 2100–2109.
[139] M. Leordeanu and M. Hebert, "Smoothing-based optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[140] F. Bastani, S. He, S. Abbar, M. Alizadeh, and H. Balakrishnan, "RoadTracer: Automatic extraction of road networks from aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4720–4728.
[141] A. Batra, S. Singh, G. Pang, S. Basu, C. V Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10385–10393.
[142] C. Ventura, J. Pont-Tuset, S. Caelles, K.-K. Maninis, and L. Van Gool, "Iterative deep learning for road topology extraction," in *Proc. Brit. Mach. Vis. Conf.*, 2018.
[143] R. Lian and L. Huang, "DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1905–1916, 2020.
[144] Y. Tan, S. Gao, X. Li, and B. Ren, "VecRoad: Point-based iterative graph exploration for road graphs extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, vol. 1, pp. 8910–8918.
[145] L. Castrejon, K. Kundu, R. Urtasun, and S. Fidler, "Annotating object instances with a polygon-RNN," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5230–5238.
[146] D. Acuna, H. Ling, A. Kar, and S. Fidler, "Efficient interactive annotation of segmentation datasets with polygon-RNN++," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 859–868.
[147] Z. Li, J. D. Wegner, and A. Lucchi, "Topological map extraction from overhead images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 1715–1724.
[148] D. Belli and T. Kipf, "Image-conditioned graph generation for road network extraction," in *Proc. Graph Represent. Learn. Workshop*, 2019.
[149] C. Sujatha and D. Selvathi, "Connected component-based technique for automatic extraction of road centerline in high resolution satellite images," *EURASIP J. Image Video Process.*, vol. 2015, 2015, Art. no. 8.
[150] G. Máttyus, S. Wang, S. Fidler, and R. Urtasun, "HD Maps: Fine-grained road segmentation by parsing ground and aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3611–3619.
[151] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "RoadNet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2043–2056, Apr. 2019.
[152] C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet, "Empirical evaluation of automatically extracted road axes," in *Proc. Empirical Eval. Techn. Comput. Vis.*, 1998, vol. 12, pp. 172–187.
[153] Z. Chen, W. Fan, B. Zhong, J. Li, J. Du, and C. Wang, "Corse-to-fine road extraction based on local Dirichlet mixture models and multiscale-high-order deep learning," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–11, 2019, doi: 10.1109/tits.2019.2939536.
[154] R. Alshehhi, P. R. Marpu, L. W. Wei, and M. D. Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, 2017.
[155] 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. 680–698, 2017.
[156] 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.
[157] G. Cheng, F. Zhu, S. Xiang, Y. Wang, and C. Pan, "Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting," *Neurocomputing*, vol. 205, pp. 407–420, 2016.
[158] Y. Zang, C. Wang, J. Li, M. Cheng, L. Luo, and Y. Yu, "Road network extraction via deep learning and line integral convolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1599–1602.
[159] X. Hu and C. V Tao, "A reliable and fast ribbon road detector using profile analysis and model-based verification," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 887–902, 2005, doi: 10.1080/0143116042000298243.

Renbao Lian received the B.S. degree in computer science and technology from the University of Jinan, Jinan, China, in 2002. He is currently working toward the Ph.D. degree in communication and information system with the College of Physics and Information Engineering, Fuzhou University, Fuzhou, China.

He is currently an Associate Professor with the Collage of Electronics and Information Science, Fujian Jiangxia University, Fuzhou, China. His research interests include remote sensing image processing, computer vision, and information processing.

Weixing Wang received the Ph.D. degree in information engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 1997.

He is currently a Professor in Information Engineering. Since 2001, he has been a Ph.D. Supervisor with the KTH Royal Institute of Technology, and is currently a Visiting Professor with Chang'an University, Xi'an, China. His research interests include information engineering, image processing and analysis, pattern recognition, and computer vision.

Nadir Mustafa received the B.S. degree in electronic engineering from Neelain University, Khartoum, Sudan, in 2012. He is currently working toward the Ph.D. degree in communication and information system with the College of Physics and Information Engineering, Fuzhou University, Fuzhou, China.

His research interests include multimedia signal processing, machine learning, and computer vision.

Liqin Huang received the Ph.D. degree in communication and information system from Fuzhou University, Fuzhou, China, in 2009.

He is currently a Full Professor and a Ph.D. supervisor with the College of Physics and Information Engineering, Fuzhou University. His research interests include image processing, computer vision, artificial intelligence, and computer network communication.