Lane-Level Road Network Construction Based on Street-View Images

Jinlin Shi , *Member, IEEE*, Guannan Li, Liangchen Zhou, and Guonian Lü

Abstract—With the advancement of autonomous driving technologies, road network data have attracted a lot of attention as a virtual source of information. Traditional node–arc road networks are no longer able to match the demands of high-precision location awareness. Thus, lane-level road networks with more information have become a research hotspot. Furthermore, street-view pictures are a popular data source for building a lane-level road network because they provide a significant quantity of road information. The current method of constructing lane-level road networks based on street-view images performs feature extraction in image space and then projects it into geographic space. Hence, due to perspective and other rules, there are conflicts and overlaps in the exact locations after projecting the results of the street-view pictures from different viewpoints into geographic space. Because, to the best of the author's knowledge, there is no process for optimizing the overall geographic space results, the current study does not meet the demand for the accurate and comprehensive acquisition of lane-level road network in complete areas. However, this study proposed a lane-level road network construction method based on street-view image data, focusing on aggregating and optimizing the picture space extraction results in geographic space to improve the accuracy while aligning the results more consistent with the vector data requirements of geographic information systems. The experimental results show that by using street-view picture data, this technology can establish a submeter lane-level road network, which can be used for low-cost road data collection and updating.

Index Terms—Data collection, geographic optimization, lane-level road network, street-view images.

I. INTRODUCTION

GRAPH theory-based node–arc road network data [1] have enabled many critical applications, including shortest path planning [2], resource allocation [3], and navigation [4], and they are an important data source for geographic information systems (GISs), as well as intelligent transportation systems [5]. Recently, there has been an increase in demand for road information due to the development of autonomous driving, digital twins, and virtual scenarios. Using conventional node–arc network data, which involve abstract expressions of details in the road and ignore elements such as ground markings, it is challenging to meet the requirements of the aforementioned technology for highly accurate and detailed road information. Furthermore, lane-level

road network data add extra road information, such as lane lines and ground symbols. Thus, lane-level road network data, which provide more detailed information about roads, have become the focus of recent studies and are crucial for autonomous driving and other domains [6]. Street-view photos are rich in semantic information and offer the advantages of rich data sources and low acquisition cost, making them ideal for quickly creating and updating lane-level road networks at a low cost. The recent methods [7], [8] that are based on street-view photos used image segmentation to detect relevant elements in the image space. Although the outputs are projected into geographic space, there is still a significant gap to becoming a usable lane-level road network for the following reasons.

1) Previous research did not thoroughly classify the constituent pieces that make up the lane-level road network; hence, certain elements were left out.
2) The findings were challenging to satisfy the typical vector format of lane-level road networks because elements were retrieved using segmentation in the picture space.
3) Because the overall results have not been aggregated or optimized in geographic space, these methods are incapable of meeting the accuracy, uniqueness, and completeness requirements for regional data collection.

Thus, the contributions of this research are the following.

1) The elements that make up the lane-level road network concerning Chinese national standards were classified, as well as the proposed corresponding detection and subsequent processing procedures for each category.
2) The lane line detention algorithm with the result as the key point and the object detection algorithm as the image space element extractor were selected.
3) A set of approaches for optimizing the results in geospatial space, making the results more accurate and more in line with the requirements of GIS for data collection were presented, which can be directly applied to downstream tasks.

The experimental section of this research shows that the produced results achieve submeter accuracy and meet the cost and accuracy requirements for lane-level road network data by comparing them with manually acquired results in a calibrated high-resolution orthophoto.

II. RELATED WORK

The reconstruction of ground marks and symbols is a crucial part of the road network creation process. In this

research, according to [9], the ground markings and symbols are divided into longitudinal and transverse markings. Longitudinal markings run parallel to the road traffic direction, whereas transverse markings run perpendicular to the road traffic direction. The symbols and text in this article are included in the unified transverse markings. Many studies on road network data have focused on detecting and rebuilding longitudinal markings while ignoring transverse markings. This study, however, gives a detailed review of deep learning-based algorithms for detecting transverse markings in pictures.

A. Road Network Generation

In this study, road network generation methods were categorized according to the data source types as follows:

1) trajectory data-based methods;
2) remote sensing image-based methods;
3) point cloud data-based methods;
4) street-view image-based methods.

1) Trajectory Data-Based Methods: These methods use mainly trajectory data of various types of vehicles to generate a road network. The advantage of these methods is that the trajectory data contain many road geometric features, and the road network structure can be recovered quickly [10]–[15]. Although Uduwaragoda *et al.* [16] and Zhang *et al.* [17] attempted to establish a lane-level road network by extracting lane centerlines using trajectory data, a large amount of semantic information, such as lane type, could not be recovered efficiently using this method.

2) Remote Sensing Image-Based Methods: These methods primarily use feature extraction algorithms extracting ground makers from remote sensing photos and constructing road networks [18]–[21]. With the development of deep learning, an increasing number of scholars are trying to use deep learning for landmark extraction. Henry *et al.* [22] wanted to use a convolutional neural network to extract and rebuild a road network. Hong *et al.* [23] presented a general multimodal deep learning framework for the fusion of different deep learning classification outcomes, which captured more spatial information for better accuracy. Zhang *et al.* [24] extracted and improved a road network using generative adversarial neural networks. More complex methods have recently been used in road extraction. Thus, Hong *et al.* [25] proposed miniGCNs in road extraction, which fuses GCN and CNN, and the extraction accuracy exceeded single-CNN or -GCN models. Furthermore, the structure of the transformer was utilized for road extraction to mine and represent the sequence attributes of spectral [26]. Remote sensing image-based methods have a wide range of application areas and great extraction accuracy. However, to extract lane-level road networks, these approaches require more data sources. Weather conditions frequently disrupt remote sensing images, thus resulting in the unavailability of crucial ground-marking attributes. In cities with dense tree cover, such as the central city of Nanjing, severe tree shading leads to poor results when these methods are employed.

3) Point Cloud-Data-Based Methods: This approach is based on point cloud data, and it has two basic processing methods: 1) After the intensity is projected to a 2-D plane, feature engineering [27] or deep learning [28] was used to extract the landmarks and then build the road networks. 2) Lane extraction was performed in 3-D point cloud space based on geometric features of lane markings and ground symbols [29]. The benefits of point cloud-data-based methods are their high geometric accuracy, but laser point cloud equipment is expensive. Moreover, light detection and ranging is very sensitive to noise and occlusion, so it is not accessible in data acquisition.

4) Street-View Image-Based Methods: 1) Feature engineering methods: These methods extract relevant aspects of ground lane lines and symbols for identification and reconstruction using various computer vision operators [30]–[33]. 2) Deep learning-based methods: Recently, these methods have achieved higher accuracy in lane detection. The main methods are ① semantic segmentation, ② row classification, and ③ anchor-based detection. ① The semantic segmentation method classifies all the pixel points in an image and clusters and extracts the contours of the pixels that correspond to road markings to obtain the marking outlines. The benefit of this method is that it realizes higher accuracy than feature engineering methods [7], [8]. Nonetheless, many operations must be conducted due to the enormous amount of operational data, and the time efficiency is lowered because of the necessity to categorize the kind of each pixel point in the image. The landmark pixel classification results require additional processing, such as removing outlier points and generating pixel outlines. ② The geometric qualities of the marking lines are used to classify the rows. According to the row direction, the image is divided into equal distances; in each row, the columns that correspond to the marking lines are identified; marking line units are formed, and the coordinates of the marking line image after postprocessing are acquired [34], [35]. The advantage of this type of method is its high speed, but it is relatively weak in terms of accuracy. ③ Inspired by region proposal networks (RPNs) [36], Li *et al.* [37] suggested an anchor point-based lane detection method. The method creates the lane lines as rays whose starting points are at the three boundaries (left, bottom, and right) offers 15 directions at the starting point and has numerous horizontal deviation points (offsets) on each candidate line for fitting the actual lane lines. The procedure achieves state-of-the-art accuracy performance on public datasets. Tabellini enhanced the model with an attention module and proposed LaneATT to give it a stronger global information perception capability, which can overcome marking lines occlusion and absence [38].

B. Object Detection

Because the development of deep neural network technology has improved accuracy recently, this study presents a literature review on object detection in deep learning. Deep neural network-based object detection methods are divided into two main categories: 1) anchor-based and 2) anchor-free approaches.

1) Anchor-Based Approaches: Anchor-based approaches are subdivided into two categories: 1) two-stage and 2) one-stage approaches. The two-stage approaches use a backbone network and then extract feature maps and subsequently provide region

proposals for target objects. Anchors with different aspect ratios are used afterward for object boundary regression. Finally, these methods use a classification network to determine the proposal's class [36], [39], [40]. The methods have a high accuracy rate but a vast number of generated proposals leads to many computational parameters and high time consumption. One-stage approaches: These approaches gather an object's boundary and category results directly through a single neural network [41]–[43]. They are fast but at the expense of accuracy.

2) Anchor-Free Approaches: Although the anchor-based approaches provide significant improvements in accuracy, a huge number of anchors lead to high memory and time consumption, as well as poor generalization ability, and the aspect ratios of anchors usually must be designed for different tasks. Recently, many researchers have conducted target detection by detecting critical locations [44]–[46]. This method does not require a large number of anchor calculations; the speed of operation is much enhanced, but the accuracy is not significantly better than that of anchor-based detection methods.

C. Summary of Study Status

This study concludes that deep learning-based detection methods may effectively detect transverse and longitudinal markings in picture space after synthesizing the aforesaid method. Considering that this study focuses on data collection for lane-level road networks, it is also important to perform optimization operations such as aggregation and filtering of image space results from various perspectives in geospatial space to ensure the integrity and correctness of the lane data for the entire road.

III. METHODOLOGY

This study uses street-view images as the main data, which contain many elements for constructing lane-level road networks. It is important to classify these elements before introducing the research methods. This study refers to [9], which divides markings into two categories: 1) longitudinal markings, that is, markings that are set along the direction of road traffic; and 2) transverse markings, that is, markings that are set across the direction of road traffic. Showing arrows, text, and so forth, are considered transverse markings in this research.

Fig. 1 is a flowchart of the methodology used in this study. This study includes three main sections.

1) *Element extraction in the image space:* Lane-level road network elements, which are primarily longitudinal and transverse markings, are extracted from street-view photos, including geometric and semantic information.

2) *Projection from the image space to geographic space:* The detected elements in the photo space are projected into geospatial space.

3) *Optimization in geographic space:* Street-view photos from different viewpoints in the image space are seen with a large quantity of information, which is projected directly into the geographic space with conflicts, overlaps, and so forth.

This research tends to create an accurate, full regional vector lane-level road network; as a result, this section proposes a series
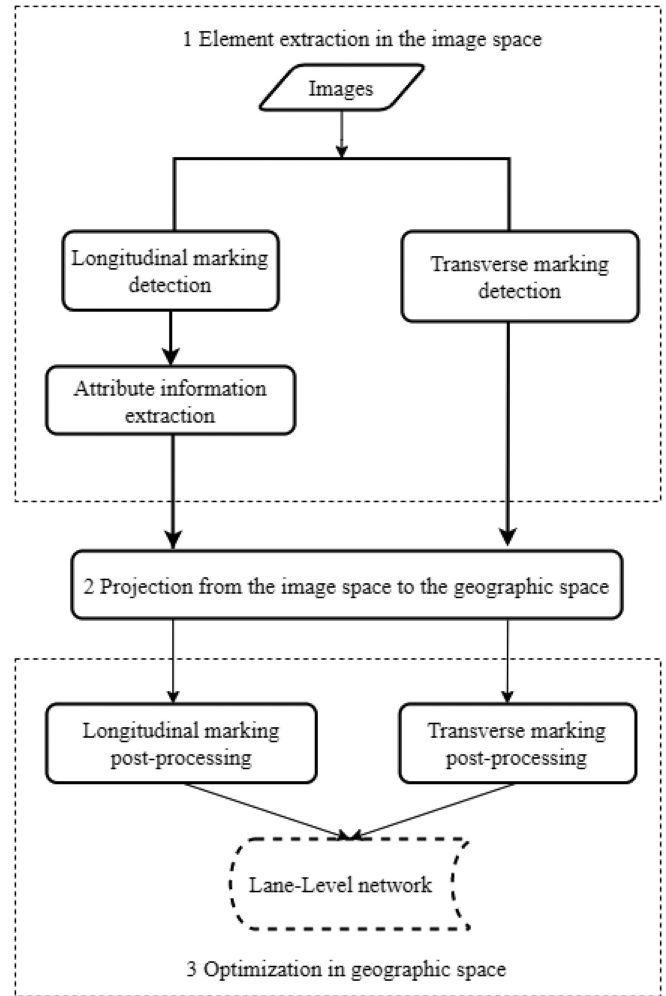

Fig. 1.   Methodology flowchart.

of optimization solutions to address the above problems while enhancing the data correctness and ensuring the data integrity of the whole street. This study thus focuses on geospatial optimization methods.

A. Element Extraction in the Image Space

1) Longitudinal Marking Extraction: Traditional feature engineering methods are less resilient and cannot recognize lane lines well in the face of complex driving scenarios when choosing detection methods. Thus, this study selected the deep learning-based longitudinal lane detection algorithm LaneAtt [38] to extract geometric information in the image space. The algorithm's output is shown in Fig. 2, in which different colors are used to differentiate between the lanes.

The result is a list of *KeyPoint* coordinates that identify the various markings, presented as points in Fig. 2. The resulting output is then universally defined for subsequent work. The line segments of the main points adjacent to the same marking line are called sections, and the whole lane line is called the *image-level lane line*.

Fig. 2. Visualization of output results (left); schematic diagram of output results (right).
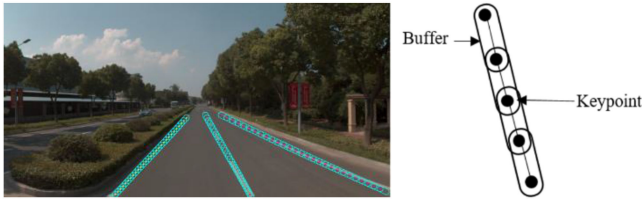

Fig. 3. Buffer visualization (left); schematic diagram of the buffer (right).

TABLE I
CLASSIFICATION TYPES

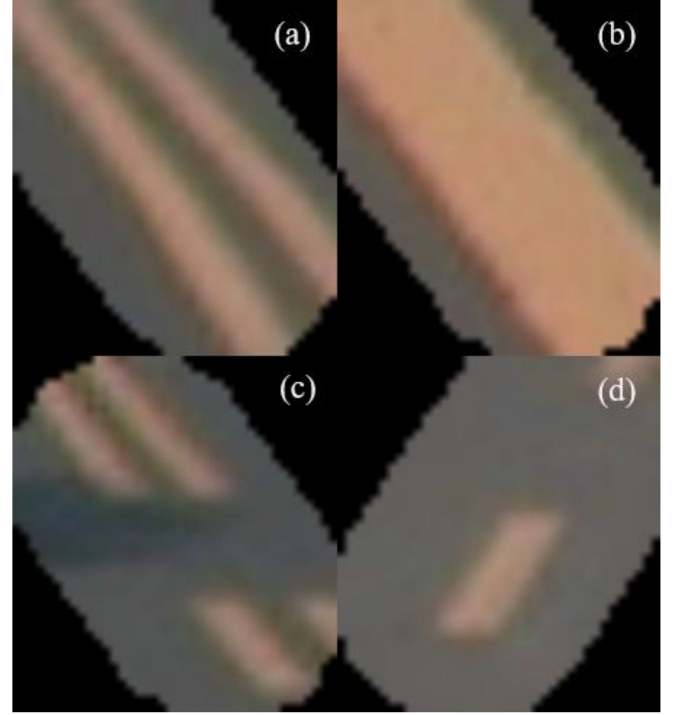| Type Index	Type Name
0	*Background (Bg)*
1	*White-real*
2	*White-broken*
3	*White-real-double*
4	*White-broken-double*
5	*Yellow-real*
6	*Yellow-broken*
7	*Yellow-real-double*
8	*Yellow-broken-double*
9	*Other*

Fig. 4. Images in the buffer. (a) Yellow-real-double. (b) Yellow-real. (c) Yellow-broken-double. (d) Yellow-broken.

Fig. 5. Visualization of classification results.

To construct a more detailed lane-level road network, after obtaining the geometric information, we wanted to further obtain the semantic attribute information of the longitudinal markings, including the color and type of each marking. However, the detection result is a line made up of many key points without width, which makes it challenging to obtain the above information. Therefore, in this study, a buffer is built with the section as the smallest unit, and the visual result is as Fig. 3.

The picture in each buffer range is classified to obtain the attribute information of the whole line. This research classifies the picture in the section buffer to get the line types, and the classification types are listed in Table I.

Example images are presented in Fig. 4.

In this study, the ResNeXt model [47] was utilized as a classifier to improve higher classification accuracy. Because of the residual structure of the ResNet [48] model, the ResNeXt model offers a more powerful feature extraction capacity. The addition

of cardinality enhances the model accuracy under approximately the same number of operations. Fig. 5 depicts a visualization of the section classification results, where the *while-real* class is rendered in white and the *background* class is rendered in black.

2) Transverse Marking Extraction: Transverse markings are subdivided into the following:

1) line elements, such as crosswalk lines, vehicle stop lines, and deceleration markings;
2) symbol elements, such as arrows, deceleration mound markings, crosswalk preview lines, and deceleration yield lines;
3) text elements, such as road speed limit characters, special lane characters, and no U-turn messages.

The above elements are clearly defined in terms of shape, size, and orientation in [9], and each element of a transverse marking is treated as a separate type in target detection, and there are a total of 45 classes. The location and type of each object of
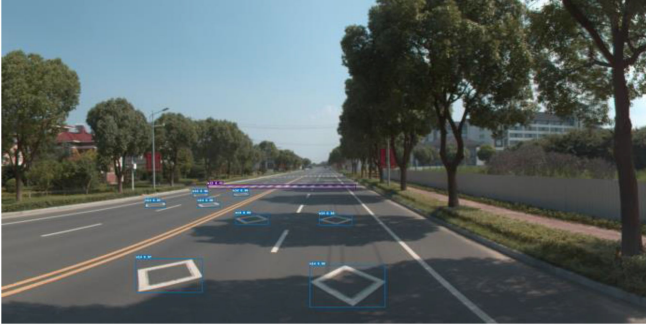
Fig. 6. Visualization of the transverse marking object detection results.

Fig. 7. Depth of the view diagram.

interest in the image space can be properly obtained using the detection algorithm. It is possible to obtain the geometric and semantic properties of the transverse markings.

This study used the object detection algorithm cascade R-CNN [49] as the transverse marking detector for the following reasons: The quality of the proposals generated by the RPN is weak, and the positive and negative judgment frequently relies on proposals and a single threshold (e.g., 0.5) for the intersection-over-union (IoU) ratio of the ground truth; thus, the subsequent output tends to produce noisy bounding boxes. To overcome this problem, cascade R-CNN uses a multistage proposal resampling at each stage, which increases the IoU values of the obtained proposals stage, lowering the noise discussed above, and finally, further improving the object detection accuracy. Cascade R-CNN achieves the highest accuracy on the public dataset. Because of the need for a high accuracy object detector when creating the road network in this research, this model was used. The image coordinates of the bounding box for each transverse marking are retrieved following detection by the cascade R-CNN target detector. The results of transverse marking object detection are presented in Fig. 6.

B. Projection From the Image Space to the Geographic Space

The preceding work primarily explains the acquisition of geometric and semantic information about longitudinal and transverse markings in the image space. The construction of road networks must take place in a specific geographic location. Thus, this section focuses on projecting the above results from the image space to the geographic location.

A panoramic image with a width W and height H was captured. The IMU+GNSS combination was used to record the pose and position information when this image is captured. Any point in the picture can be noted as $P(\text{col}, \text{row})$, where col is the column number and row is the row number. The object t denoted by point P is the observation object (Object, Obj). The camera is fixed to form a strong body above the vehicle in this study. The ground on which the car travels is presumed to be horizontal, and the phase plane is perpendicular to the ground. It can be noted abstractly, as shown in Fig. 7.

The camera optical center is represented as O, the height of O from the ground is H_o, the angle between the vertical ground line of O and the observation object is θ, and the distance from
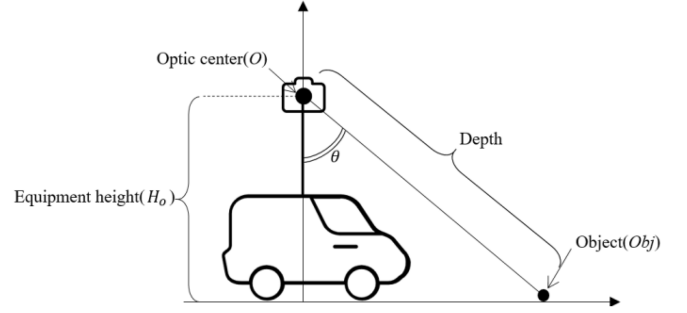
the observation object to the optical center is denoted to be the depth.

The following formula calculates the depth D_p at point P:

$$D_p = D_{(\text{col},\text{row})} = \frac{H_o}{\cos(\theta)} = \frac{H_o}{\cos\left(\frac{\pi}{2} + \left(\frac{H}{2} - \text{row}\right) \cdot \frac{180}{H}\right)}. \tag{1}$$

A camera coordinate system was built for the panorama, which is a right-handed system Fig. 8 shows the diagram.
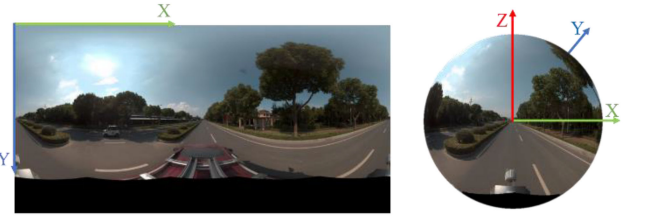
Fig. 8. Establishment of the camera coordinate system.

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = R_{(\text{col},\text{row})} \cdot \begin{bmatrix} 0 \\ D_{(\text{col},\text{row})} \\ 0 \end{bmatrix}. \tag{2}$$

The following formulas are used to calculate p_{row} and h_{col}:

$$p_{\text{row}} = \frac{\frac{H}{2} - \text{row}}{H} \cdot 180 \tag{3}$$

$$h'_{\text{col}} = \frac{\left(\text{col} - \frac{W}{2}\right)}{W} \cdot 360 \tag{4}$$

$$h_{\text{col}} = \begin{cases} h'_{\text{col}} + 360 \ h'_{\text{col}} < 0 \\ h'_{\text{col}} - 360 \ h'_{\text{col}} > 360 \end{cases} \tag{5}$$

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R_{(\text{roll, heading,pitch})} \cdot R_{(\text{col},\text{row})} \cdot \begin{bmatrix} 0 \\ D_{(\text{col},\text{row})} \\ 0 \end{bmatrix} + T. \tag{6}$$

Inspired by Tsai and Chang [50], this article uses (6) to project any point P in the image into the geographic space. Where $R_{(\text{roll, heading,pitch})}$ is the external reference matrix, which is composed of the positional data obtained from the IMU; $R_{(\text{col},\text{row})}$ is the rotation matrix from any point P to the camera optical center; and T represents the world coordinates of the camera, which are obtained by the global satellite navigation system (GNSS).
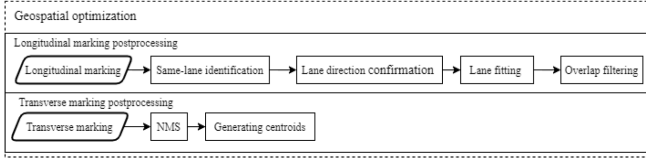
Fig. 9.   Geospatial optimization flowchart.

Fig. 10.   Projection results of longitudinal markings for three consecutive street-view images.

Fig. 11.   Longitudinal marking buffer results.

C. Optimizations in the Geographic Space

The acquisition of longitudinal and transverse marks, as well as the projection of significant points from the image space to the geographic space, were covered in previous sections. Due to the error of coordinates and poses obtained from the GNSS and IMU devices and the influence of camera perspective, there are back-and-forth conflicts and repetitions at the exact locations; hence, the geospatial results must be optimized. Fig. 9 thus shows a flowchart of the optimization process.

1) Longitudinal Marking Postprocessing: Due to the limited effective area of the image, the longitudinal markers that are extracted from the continuous image are projected to the geographic space to express polyline segments, which cannot yet become lane data in the road network; see Fig. 10 for details.

This is because of the impact of imaging perspective rules and chance mistakes, such as GNSS signal lag and misalignment with IMU time. To produce the most basic unit of the lane-level road network, namely the road-level longitudinal marking, numerous images, that is, multiple polylines, need to be fitted into a smooth lane line to produce.

a) Same-lane identification: First, it is necessary to determine which polylines are longitudinal markings of the same road level. Following the projection in Section III-B, the image-level markings corresponding to the same road-level longitudinal markings are close, whereas the projection results of different markings are far apart. Therefore, for each image-level-marking projection result, a smaller buffer is created first as shown in Fig. 11.

If there is an intersection, buffers of different image-level longitudinal markings are considered to be the same road-level longitudinal marking and are given the same ID.

b) Lane direction confirmation: It is necessary to examine the primary direction of the fitting, that is, whether the *X*-coordinate is fitted to the *Y*-coordinate or the *Y*-coordinate is fitted to the *X*-coordinate. In this study, three adjacent pictures are used as the minimum unit for direction identification, and the significant directions of these line segments are used for the fitting. The formula for calculating the significant directions of these lines is

$$\text{Direction} = \begin{cases} X \text{ if } |X_{\text{start}} - X_{\text{end}}| \geq |Y_{\text{start}} - Y_{\text{end}}| \\ Y \text{ if } |X_{\text{start}} - X_{\text{end}}| < |Y_{\text{start}} - Y_{\text{end}}| \end{cases} \quad (7)$$

c) Lane fitting: After the same line confirmation, we installed the longitudinal markers corresponding to the same road. For cubic curve lane fitting, this research employed the least-squares method, with control points being the section endpoints in the three adjacent pictures, assigned to the same ID in the previous section:

$$\begin{cases} X = R_c \cdot Y \text{ if Main Direction} = X \\ Y = R_c \cdot X \text{ if Main Direction} = Y \end{cases} \quad (8)$$

where R_c represent the least squares-derived cubic function coefficient matrix. Fig. 12 shows the fitting results.

d) Overlap Filtering: Due to repetitive portions of the same longitudinal markers being shown in adjacent street-view pictures, different image-level lane lines were fitted together in the aforementioned section; hence, there is a superposition of markings at the same location, as detailed in Fig. 13(a).

The result that is closer to the camera has a higher confidence level, according to the camera imaging standards. Therefore, when overlap occurs, we choose to keep the result that is closest to the camera's optical center. The ultimate result is depicted in Fig. 14.

2) Transverse Marking Postprocessing: The projection method in part *B* was used to project outsourced rectangles generated from transverse marking detection projected into the geographic space. There were instances where the results of
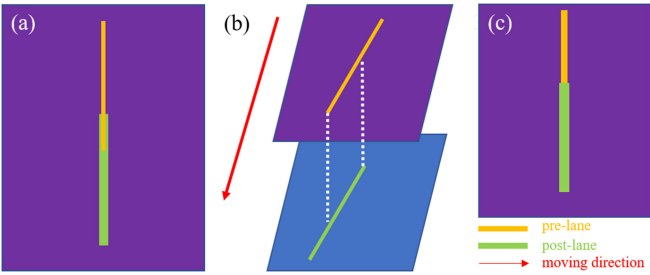
Fig. 12. Fitting results.

Fig. 13. Filtering process. (a) Lane overlap. (b) Layer-by-layer exhibition. (c) Filtered result.

Fig. 14. Final result.

numerous street-view photos appear to overlap, as shown in Fig. 15.

The projection method is a more accurate optical center of the camera. However, because of the high robustness, transverse markings cut by the imaging range can also be identified. Thus, Fig. 16 depicts the detection results of two adjacent street-view images. The image is on the left, whereas the following image is on the right.

The image width cuts the left turn arrow in half in the right panel of Fig. 16, and only a portion of it is still recognizable.
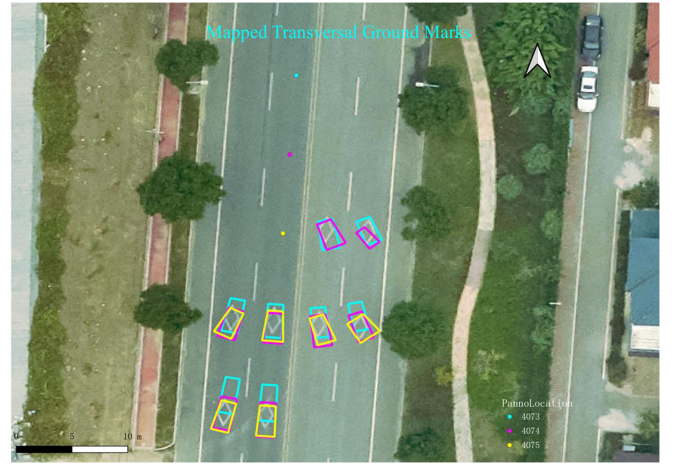

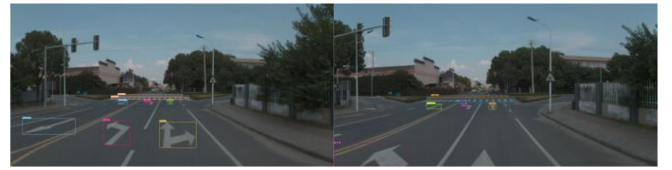Fig. 15. Projection results of transverse markings.

Fig. 16. Transverse marking detection results for two adjacent street-view images.

We used nonmaximum suppression (NMS) [51] filtering on the above findings in the geographic space to maintain just the detection results with the highest confidence. The confidence of a cut object is often lower than that of the complete object, so this operation can effectively solve the above problem. The national standard fully defines the shapes and widths, and as transverse marking line is abstracted into a point-like element, for postprocessing.

IV. EXPERIMENTS AND ANALYSIS

A. Experimental Area Introduction

The experimental area of this research was in Yincun, Changzhou City, Jiangsu Province, which is located at the core of the Yangtze River Delta region. The location was a vocational education base in Jiangsu Province, a complete road infrastructure construction, demonstrating the method's efficacy. The experimental area covered a total of 2.619 km^2, and the total length of the road was 19.1 km. The red range in Fig. 17 covers the experimental area, and the blue lines represent the vehicle driving track.

B. Equipment Introduction

The primary equipment used in this experiment is as follows:
1) a Ladybug5 optical camera, which is mainly used to capture street-view images;
2) a Novatel 718D GNSS, which is utilized in making records of optical camera position information;
3) a Honeywell 4930 inertial sensor (IMU), primarily to record the optical camera position information.
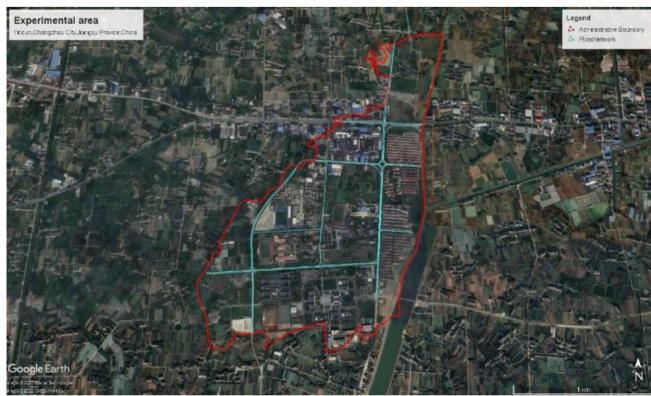
Fig. 18.   Experimental equipment.

Fig. 17.   Yincun location in Google Maps.

All of the equipment listed above is attached to the vehicle to form a hard body. Calibration of the camera, GNSS, and IMU equipment is done in real-time. Fig. 18. shows our equipment.

C. Experimental Method

1) Experiment Preparation: The experiments in this study use a vehicle with the aforementioned equipment to drive through the experimental region, following the blue line in Fig. 17. The camera position and pose information are recorded at the time of capture for a total of 2868 street-view photos.

a) Specific parameters: As discussed in Section III-A, if the buffer is too large, the algorithm's efficiency will be impaired while adding extra background information to influence classification accuracy, whereas if the buffer is too small, it will not be able to cover the marker lines closer to the camera optical center. Following repeated experiments, we select a buffer size of 15 pix. As discussed in Section III-A, the vertical distance of the camera optical center from the ground (H_o) is 2.65 m after calibration. We optimized the markings within 15 nm of the camera in Section III-C. For same-lane identification, we employ a buffer radius of 0.4 m in Section III-C. Section III-C uses the NMS algorithm, where the IoU threshold is set to 0.2.
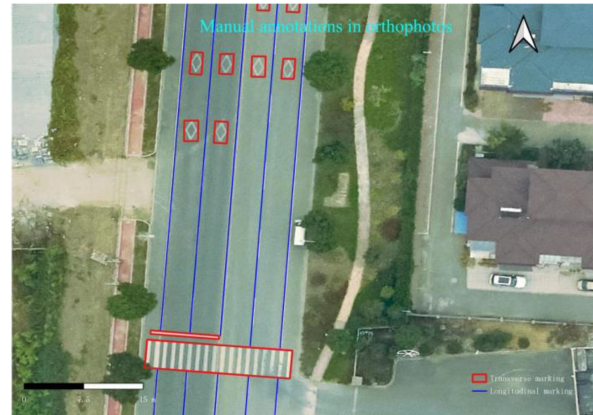

Fig. 19.   Manual annotations in orthophotos.

b) Deep learning models: For longitudinal marking detection, we used the LaneAtt algorithm of BackBone for ResNet122, and the training dataset was the CULane [52]. F1 on the test set was 77.02%. Pytorch was the algorithm's implementation platform used.

We employed the ResNeXt152_vd [47] as the classifier for longitudinal marking attribute categorization. We also used the ImageNet pretraining model to obtain the pretraining parameters. Training data were collected from Nanjing, Jiangsu Province, China. The training set contained a total of 74 267 images; the test set contained 8250 images. As the loss function, we employed CEloss. On the test set, the model achieved Top1 accuracy of 95% and Top5 accuracy of 99.952%. PaddlePaddle was the algorithm's implementation platform used.

The transverse marking detection algorithm was Cascade-RCNN [49], the backbone is CBResNet with a depth of 200 layers [53], and deformable convolution (DCN V2) was added to the model [54] to improve the detection accuracy. The training dataset was gathered in Nanjing and Suzhou, both located in Jiangsu Province, China. The training set has a total of 13 532 photos, whereas the test set has 1502 images. At the IoU 0.5 level, we used the model with an AP of 74.2% and an AR of 72.3%. The implementation platform of the algorithm used was PaddlePaddle

c) Validation data introduction: This study employed a 0.1 m resolution orthophoto that was captured by a UAV, and these data were matched with a ground control point to achieve centimeter-level accuracy. Manual annotation was used to gather geometric and semantic attribute information of longitudinal and transverse markings from orthophotos, where the line type was collected for longitudinal markings, and the type and boundary boxes were collected for transverse markings. Fig. 19 shows several examples of data.

2) Experiment Details:

a) Geometric accuracy estimation of longitudinal markings: The manual orthophoto acquisition results were used as the benchmark in this work. A vertical line was created at 1 cm intervals for each hand-gathered longitudinal marker. The error was the distance between the intersection of the vertical line and the findings generated in this article and the vertical
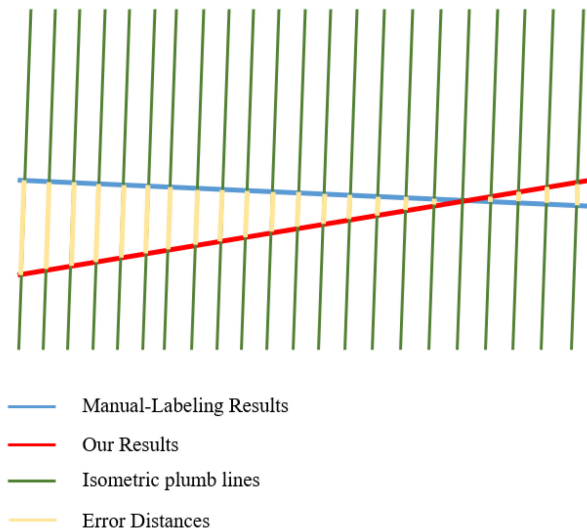
Fig. 20.   Error calculation schematic diagram.

Fig. 21.   Geometric and semantic error example diagram.

foot. A schematic diagram is presented in Fig. 20. The practical conditions are shown in Fig. 21.

Using the method above, a total of 2 099 842 vertical lines were obtained in the experiment, and the average error distance was 0.2082 m.

b) Semantic attribute accuracy estimation of longitudinal markings: In Section IV-C1, the accuracy of the longitudinal marking semantic attribute classifier is discussed. Because of the projection, fitting, and filtering steps, the final accuracy needs to be evaluated in the geospatial space. The procedure is identical to that used to assess the geometric accuracy of longitudinal markers. To determine if the attribute information of the crossing point and the vertical foot is the same, the original form for computing the error distance is updated. If they are the same, the attribute is correct, and if they are different, the attribute is erroneous. The accuracy was computed to be 94.18%.

c) Transverse marking geometry and semantic accuracy estimation: The geometric results of the transverse markings are evaluated in this work using the results of manual acquisitions in the orthophoto as actual values. The orthophoto transverse

marking acquisition results are the smallest outsourcing rectangles and are turned into the corresponding center points. We also convert our results into the center points. The gap between two center points was considered to be a geometric error. In the experimental area, there are a total of 169 transverse markers with an average error calculated to be 0.61 m. To test if the kinds of two centroids are consistent, a semantic accuracy evaluation is performed: They are right if they are consistent; otherwise, they are erroneous. The semantic accuracy was calculated to be 100%.

D. Analysis

1) Accuracy Discussion: First, we discuss the geometric accuracy: 1) There was a discrepancy between the assumed conditions and the actual situation: The projection method used in this study is based on the assumption that the road surface is always horizontal and the phase plane is always perpendicular to the ground, but the above conditions are unlikely to be satisfied during the actual vehicle driving process because the road surface has high and low undulations, and the vehicle also has bumps and other conditions. 2) The camera shutter time has not been synchronized with the IMU and GNSS times. Even if the Kalman filter interpolation operation is employed in the preprocessing section, this study may not correctly capture the camera location or attitude data due to the various periods of IMU and GNSS. Next, we discuss the semantic accuracy: 1) The effect of light changes and shading on optical camera imaging: There are many trees and buildings in the experimental area, and the shadows have a significant impact on the classification of the marking attributes. Simultaneously, the region is under construction. The road surface has become dustier, and some of the markings have become obscured, affecting the categorization of the marking kinds on both sides of the road. 2) Geometric deviations generate the following effects: Although the image space is correct, the projection into geographic space deviates from the real place and does not match the orthophoto findings.

2) Parameter Analysis: Because the optical center height determines the simulated depth of field (H_o), various parameters such as the optimization range, the buffer size of the longitudinal marking coincidence confirmation, and the threshold of the transverse marking filtering algorithm must be selected with the optical center height in mind. For example, in cases where the optical center is higher than in this study, the size of the longitudinal marking same line identification buffer may be lowered.

3) Noise and Variability Effects: When projecting image space findings into geographic space, we assume that the optical center is always perpendicular to the ground. In real situations, there are noise and variation disturbances, such as sudden ground bumps, emergency avoidance, and so forth, which affect the accuracy of our proposed method. Hong *et al.* proposed a linear mixing model to express the above noise and variation effects and achieved state-of-the-art accuracy performance [55]. In our practical application, we refer to the aforementioned solution.

4) Method Extensibility: This method works with any device that can record street-view images, such as a car recorder or smartphone. Extending the data source to various sorts of devices
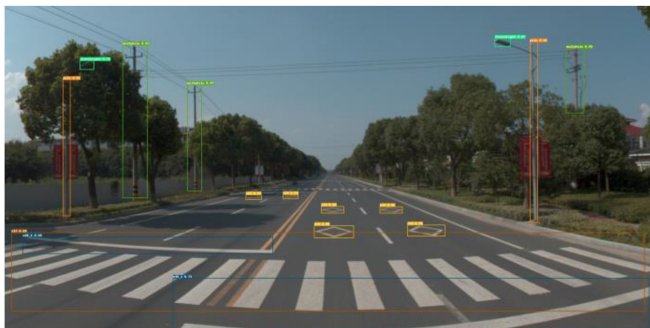
Fig. 22.   Facility detection results in street-view images.

can help the road network be updated faster. This technique can collect ground markings, text, and symbols, but roadside entities such as utility poles, guardrails, and signage lights can also be managed by this method. As shown in Fig. 22, it can efficiently serve urban statistical infrastructure construction and is routinely checked.

V. Conclusion

This study proposed a lane-level road network generation method that uses street-view images as the main data source to detect longitudinal and transverse markings in the image space according to the national standard classification system and project the image space results into the geographic area. This study focused on aggregating, fitting, and filtering image space results from different perspectives in the geographic space, ensuring this method's completeness accuracy in regional lane-level road network data collection. In the experimental part of a 19.1 km road, a longitudinal marking geometric error of 0.2082 m with 94.18% accuracy on semantic attributes and a transverse marking geometric accuracy error of 0.61 m with or no inaccuracies in semantic features were obtained. The results demonstrate that this method can acquire submeter lane data at a lower cost and can be widely used in road data acquisition, change detection, and updating. Our subsequent research will focus on the improvement and enhancement of the longitudinal marker detection algorithm in image space to further improve the efficiency by integrating the current detection and classification algorithms into the same algorithm.

References

[1] M. F. Goodchild, "Geographic information systems and disaggregate transportation modeling," *Geograph. Syst.*, vol. 5, pp. 9–17, 1998.

[2] W. Shuxi and L. Anyu, "Multi-adjacent-vertexes and multi-shortest-paths problem of Dijkstra algorithm," *Comput. Sci.*, vol. 41, no. 6, pp. 217–224, Sep. 2014.

[3] Y. He *et al.*, "K-shortest-path-based evacuation routing with police resource allocation in city transportation networks," *PLos One*, vol. 10, no. 7, Jul. 2015, Art. no. e0131962, doi: 10.1371/journal.pone.0131962.

[4] Q. Guo *et al.*, "Graphic simplification and intelligent adjustment methods of road networks for navigation with reduced precision," *Int. J. Geo-Inf.*, vol. 9, no. 8, Aug. 2020, Art. no. 490, doi: 10.3390/ijgi9080490.

[5] E. Agyemang, "A cost-effective Geographic Information Systems for Transportation (GIS-T) application for traffic congestion analyses in the developing world," *Ghana J. Geograph.*, vol. 5, no. 1, pp. 51–72, 2014.

[6] R. Liu *et al.*, "High definition map for automated driving: Overview and analysis," *J. Navig.*, vol. 73, no. 2, pp. 324–341, Mar. 2020, doi: 10.1017/S0373463319000638.

[7] T. Qin *et al.*, "RoadMap: A light-weight semantic map for visual localization towards autonomous driving," in *Proc. IEEE Intl. Conf. Robot. Autom.*, 2021, pp. 11248–11254, doi: 10.1109/ICRA48506.2021.9561663.

[8] M. Herb, T. Weiherer, N. Navab, and F. Tombari, "Lightweight semantic mesh mapping for autonomous vehicles," in *Proc. IEEE Intl. Conf. Robot. Autom.*, 2021, pp. 6732–6738.

[9] *Road Traffic Signs and Markings-Part 3: Road Traffic Markings*, GB 5968.3-2009, 2009.

[10] W. Fang *et al.*, "A novel road network change detection algorithm based on floating car tracking data," *Telecommun. Syst.*, vol. 75, no. 2, pp. 161–167, Oct. 2020.

[11] Y. Guo *et al.*, "A scalable method to construct compact road networks from GPS trajectories," *Int. J. Geograph. Inf. Sci.*, vol. 35, no. 7, pp. 1309–1345, Jul. 2021.

[12] L. Manière *et al.*, "Building a model to reconstruct the Hellenistic and Roman road networks of the Eastern Desert of Egypt, a semi-empirical approach based on modern travelers' itineraries," *J. Comput. Appl. Archaeol.*, vol. 4, no. 1, pp. 20–46, Mar. 2021.

[13] R. S. de Sousa *et al.*, "On the prediction of large-scale road-network constrained trajectories," *Comput. Netw.*, vol. 206, 2022, Art. no. 108337.

[14] Y. Zhang *et al.*, "A hybrid method to incrementally extract road networks using spatio-temporal trajectory data," *Int. J. Geo-Inf.*, vol. 9, no. 4, Mar. 2020, Art. no. 186.

[15] C. Zhang *et al.*, "Generating road networks for old downtown areas based on crowd-sourced vehicle trajectories," *Sensors*, vol. 21, no. 1, Jan. 2021, Art. no. 235, doi: 10.3390/s21010235.

[16] C. M. Uduwaragoda, A. S. Perera, and S. A. D. Dias, "Generating lane level road data from vehicle trajectories using kernel density estimation," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst.*, 2013, pp. 384–391, doi: 10.1109/ITSC.2013.6728262.

[17] Y. Zhang *et al.*, "Generation of lane-level road networks based on a trajectory-similarity-join pruning strategy," *Int. J. Geo-Inf.*, vol. 8, no. 9, Sep. 2019, Art. no. 416.

[18] P. Gamba, F. Dell'Acqua, and G. Lisini, "Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 387–391, Jul. 2006, doi: 10.1109/LGRS.2006.873875.

[19] W. He *et al.*, "A multiscale method for road network extraction from high-resolution SAR images based on directional decomposition and regional quality evaluation," *Remote Sens.*, vol. 13, no. 8, Apr. 2021, Art. no. 1476.

[20] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 6, pp. 764–773, Nov. 1988, doi: 10.1109/36.7708.

[21] F. Tupin, H. Maitre, J. Mangin, J. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Application to road network extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 434–453, Mar. 1998, doi: 10.1109/36.662728.

[22] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018, doi: 10.1109/LGRS.2018.2864342.

[23] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021, doi: 10.1109/TGRS.2020.3016820.

[24] Y. Zhang *et al.*, "Road topology refinement via a multi-conditional generative adversarial network," *Sensors*, vol. 19, no. 5, Mar. 2019, Art. no. 1162.

[25] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021, doi: 10.1109/TGRS.2020.3015157.

[26] D. Hong *et al.*, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615, doi: 10.1109/TGRS.2021.3130716.

[27] H. Ma *et al.*, "Automatic extraction of road markings from mobile laser scanning data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. W7, pp. 825–830, Sep. 2017, doi: 10.5194/isprs-archives-XLII-2-W7-825-2017.

[28] C. Wen *et al.*, "A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds," *ISPRS J. Photogramm.*, vol. 147, pp. 178–192, Jan. 2019, doi: 10.1016/j.isprsjprs.2018.10.007.

[29] Y. Pan *et al.*, "Automatic road markings extraction, classification and vectorization from mobile laser scanning data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. W13, pp. 1089–1096, Jun. 2019, doi: 10.5194/isprs-archives-XLII-2-W13-1089-2019.

[30] W. Jingyu and D. Jianmin, "Lane detection algorithm using vanishing point," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2013, pp. 735–740, doi: 10.1109/ICMLC.2013.6890384.

[31] H. Jung, J. Min, and J. Kim, "An efficient lane detection algorithm for lane departure detection," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 976–981, doi: 10.1109/IVS.2013.6629593.

[32] P. S. Rahmdel, D. Shi, and R. Comley, "Lane detection using Fourier-based line detector," in *Proc. IEEE 56th Int. Midwest Symp. Circuits Syst.*, 2013, pp. 1282–1285, doi: 10.1109/MWSCAS.2013.6674889.

[33] S. Yun, Z. Guo-Ying, and Y. Yong, "A road detection algorithm by boosting using feature combination," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 364–368, doi: 10.1109/IVS.2007.4290141.

[34] Z. Qin *et al.*, "Ultra fast structure-aware deep lane detection," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 276–291, doi: 10.1007/978-3-030-58586-0_17.

[35] S. Yoo *et al.*, "End-to-end lane marker detection via row-wise classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 4335–4343, doi: 10.1109/CVPRW50498.2020.00511.

[36] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[37] X. Li, J. Li, X. Hu, and J. Yang, "Line-CNN: End-to-end traffic line detection with line proposal unit," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 248–258, Jan. 2020, doi: 10.1109/TITS.2019.2890870.

[38] L. Tabelini *et al.*, "Keep your eyes on the lane: Real-time attention-guided lane detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 294–302.

[39] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.

[41] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[42] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.

[43] X. Hu *et al.*, "Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network," *Comput. Electron. Agriculture*, vol. 185, 2021, Art. no. 106135, doi: 10.1016/j.compag.2021.106135.

[44] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2019, pp. 734–750.

[45] X. Zhou *et al.*, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6569–6578.

[46] C. Zhu *et al.*, "Feature selective anchor-free module for single-shot object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 840–849.

[47] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5987–5995.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[49] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162, doi: 10.1109/CVPR.2018.00644.

[50] V. J. D. Tsai and C. T. Chang, "Three-dimensional positioning from Google street view panoramas," *IET Image Process*, vol. 7, no. 3, pp. 229–239, Apr. 2013.

[51] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. 18th Int. Conf. Pattern Recognit.*, 2006, pp. 850–855, doi: 10.1109/ICPR.2006.479.

[52] X. Pan *et al.*, "Spatial as deep: Spatial CNN for traffic scene understanding," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.

[53] Y. Liu *et al.*, "CBNet: A novel composite backbone network architecture for object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11653–11660.

[54] X. Zhu *et al.*, "Deformable convnets v2: More deformable, better results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9308–9316.

[55] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2018, doi: 10.1109/TIP.2018.2878958.

Jinlin Shi (Member, IEEE) received the B.S. degree from the Suzhou University of Science and Technology, Suzhou, China and the M.S. degree from Nanjing Normal University, Nanjing, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree with the School of Geography, Nanjing Normal University, all in geography information science.

His research interests include GeoAI, deep learning, data mining, crowd counting, and computer vision.

Guannan Li received the B.S. degree from Nantong University, Nantong, China, in 2017. He is currently working toward the Ph.D. degree with the School of Geography, Nanjing Normal University, Nanjing, China, both are in geography information science.

His research interests include 3-D modeling and data mining.

Liangchen Zhou received the B.S. and Ph.D. degrees in geography information science from Nanjing Normal University, Nanjing, China, in 2001 and 2009, respectively.

He is currently a Professor at the School of Geography, Nanjing Normal University, Nanjing, China. His research interests include 3-D modeling, virtual reality, 3-D reconstruction, and computer graphics.

Guonian Lü received the B.S. degree and the M.S. degree in geography from Nanjing University in 1984 and 1987, and the Ph.D. degree in geography information science from Institute of Geographic Science and Resource, CAS, in 1990.

He is currently the managing director at Key Laboratory of Virtual Geographic Environment, Ministry of Education of PRC, Nanjing Normal University, China. His areas of interest include VGE, Urban GIS, Sharing and Integration of Geographic model, Socialization application, and service of GIS.