Capsule-Based Networks for Road Marking Extraction and Classification From Mobile LiDAR Point Clouds

Lingfei Ma, *Student Member, IEEE*, Ying Li, *Graduate Student Member, IEEE*, Jonathan Li, *Senior Member, IEEE*, Yongtao Yu, *Senior Member, IEEE*, José Marcato Junior, *Member, IEEE*, Wesley Nunes Gonçalves, *Member, IEEE*, and Michael A. Chapman

***Abstract*— Accurate road marking extraction and classification play a significant role in the development of autonomous vehicles (AVs) and high-definition (HD) maps. Due to point density and intensity variations from mobile laser scanning (MLS) systems, most of the existing thresholding-based extraction methods and rule-based classification methods cannot deliver high efficiency and remarkable robustness. To address this, we propose a capsule-based deep learning framework for road marking extraction and classification from massive and unordered MLS point clouds. This framework mainly contains three modules. Module I is first implemented to segment road surfaces from 3D MLS point clouds, followed by an inverse distance weighting (IDW) interpolation method for 2D georeferenced image generation. Then, in Module II, a U-shaped capsule-based network is constructed to extract road markings based on the convolutional and deconvolutional capsule operations. Finally, a hybrid capsule-based network is developed to classify different types of road markings by using a revised dynamic routing algorithm and large-margin Softmax loss function. A road marking dataset containing both 3D point clouds and manually labeled reference data is built from three types of road scenes, including urban roads, highways, and underground garages. The proposed networks were accordingly evaluated by estimating robustness and efficiency using this dataset. Quantitative evaluations indicate the proposed extraction method can deliver 94.11% in precision, 90.52% in recall, and 92.43% in F_1-score, respectively, while the classification network achieves an average of 3.42% misclassification rate in different road scenes.**

Manuscript received December 4, 2019; revised February 19, 2020 and April 10, 2020; accepted April 21, 2020. This work was supported in part by the Natural Science Foundation of China under Grant 41871380. The Associate Editor for this article was C. Wen. *(Corresponding author: Jonathan Li.)*

Lingfei Ma and Ying Li are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: l53ma@uwaterloo.ca; y2424li@uwaterloo.ca).

Jonathan Li is with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China (e-mail: junli@uwaterloo.ca).

Yongtao Yu is with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China (e-mail: allennessy.yu@gmail.com).

José Marcato Junior and Wesley Nunes Gonçalves are with the Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil (e-mail: jose.marcato@ufms.br; wesley.goncalves@ufms.br).

Michael A. Chapman is with the Department of Civil Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail:mchapman@ryerson.ca).

Digital Object Identifier 10.1109/TITS.2020.2990120

***Index Terms*— Point cloud, LiDAR, capsule network, road surface, road marking, extraction, classification, dynamic routing.**

I. Introduction

NOWADAYS, many leading digital mapping corporations (e.g., Here, TomTom, Google Maps, and Bing Maps) and multinational courier services companies (e.g., UPS, FedEx, and SF Express), are investing increasingly and dedicating themselves to produce high-definition (HD) maps [1]. Such HD maps are capable of providing sub-lane level road information and highly detailed road inventories, including traffic signs, pole lights, roadside trees, lanes, boundaries, curbs, and all other essential road assets required for the development of autonomous vehicles (AVs) and intelligent service robotics (ISRs) [2]. As a critical element in HD maps, road markings play a significant role in guiding, regulating, and forbidding all road participants [3]. For instance, lane lines regulate driving zones, painted texts indicate traffic rules, and arrows show allowable driving directions. Therefore, accurately extracting and classifying road markings have a significant impact on transportation-related policymaking, driving behavior regulation, and traffic collision reduction.

A series of research has been conducted for road marking segmentation and classification using 2D images obtained from vehicle-borne optical imaging systems [4], [5]. In [6], a group of geometric feature functions in a probabilistic Random Under Sampling Boost (RUSBoost) and Conditional Random Field (CRF) classification framework was employed to automatically learn the rules embodied in road markings from stereo images. A trainable multi-task model was developed in [7] for pavement marking recognition and segmentation from images acquired under complex road topotaxy and varying traffic conditions. Moreover, line markings were extracted in [8] by creating a novel line proposal unit embedded in a fully convolutional network (FCN) for valid feature encodings, which achieved the promising performance on MIKKIand TuSimple image datasets. However, such image-related methods are highly susceptible to weather and illumination variations [9].

Mobile laser scanning (MLS) systems comprising a combined Global Navigation Satellite System and Initial Measurement Unit (GNSS/IMU) subsystem, a Light Detection

and Ranging (LiDAR) subsystem, a Radio Detection and Ranging (RADAR) subsystem, CCD cameras, and a central computing subsystem, can collect highly dense and accurate point clouds with intensity or reflectance information in large-scale urban environments and highways [10]. Compared to vehicle-mounted cameras, LiDAR sensors are less sensitive to ambient lighting conditions [11]. The point density collected by MLS systems can achieve over 10,000 pts/m^2 with cm-level resolutions, while it is quite challenging for both terrestrial and airborne laser scanning (TLS/ALS) platforms to deliver such precision and flexibility [12].

Therefore, many studies focusing on the road marking extraction and classification have been addressed by using MLS point clouds [13]. However, massive and unevenly distributed 3D point clouds make the intelligent point cloud processing challenging. Occlusions and distortions, intensity variations, density variations, noisy points, and incomplete pavements during MLS data acquisition also result in considerable difficulties. Since thresholding-based methods at a global scale cannot effectively extract road markings from georeferenced images with various point distributions, multi-threshold methods are accordingly proposed by partitioning road surface point clouds into a set of data blocks and determining an adaptive threshold within each data block [14]. Nevertheless, such methods highly rely on suitable data block sizes. Meanwhile, a normalized intensity-based approach was performed in [15] to minimize the impacts of different intensity values due to varying distances from the onboard laser scanners to scanning objects. However, the normalization parameters defined in such methods are different from scene to scene.

Generally, there exist three main challenges for road marking extraction and classification from mobile LiDAR data: (1) the contrast between pavements and road markings is relatively low. Road damage is inevitable regarding poor maintenance, which leads to the unevenly distributed intensity, thereby resulting in intensity-related methods ineffective. (2) The intensity values and point densities are varying. Point clouds are generally acquired by vehicle-based MLS systems that are driven through changing lanes at varying driving speeds. Depending on the profiling scanning mechanism of MLS systems, the incident angle of laser beams grows larger with an increased scanning range. Consequently, road markings have higher intensity values and point densities if they are closer to the trajectory of MLS systems. It is challenging for thresholding-based extraction methods to effectively extract road markings by assuming that intensity and point density are uniformly distributed. (3) Some road markings are incomplete. The damage of road surfaces resulting from on-road overloaded trucks and severe weather conditions, such as acid-alkali erosion, could create worn and decaying road markings. Moreover, occlusions from all road participants (e.g., vehicles and cyclists) also bring in dilemmas and uncertainties for the accurate extraction and classification of road markings. Accordingly, manual editing and post-refinement are required to improve the completeness and accuracy of extracted road markings. However, it is time-consuming and labor-intensive.

To deal with these challenges, we investigate the feasibility of combing capsule networks with hierarchical feature encodings of georeferenced feature images. Compared to the conventional convolutional neural networks (CNNs), capsule networks have achieved superior performance in image segmentation and classification tasks, which captures more intrinsic features in pose and spatial relationships of different objects in images [16], [17]. In this paper, we develop two capsule-based neural network architectures for road marking extraction and classification by using MLS point clouds. To this end, a pixel-wise U-shaped road marking extraction network is proposed to segment road markings from input images. At first, the road surface is partitioned into a collection of image patches. Then, the Intersection-over-Union (IoU) loss is employed, rather than cross-entropy, to guide weight updates in the U-shaped segmentation architecture. Finally, road markings are extracted based on binary classification. Moreover, combined with the fully connected (FC) capsule layers, a capsule-based network is constructed to classify road markings. First of all, the extracted road markings are resized to 28×28 pixels for computational complexity reduction. Then, two sibling classification networks (i.e., a capsule-based network and a fully-connected capsule network) are trained to encode both low-level and high-level features for different road marking classes, followed by a revised dynamic routing algorithm. Meanwhile, a large-margin Softmax (L-Softmax) loss function is adopted in the capsule-based classification model to guide training, instead of a standard Softmax loss. Finally, road markings are effectively categorized into several groups.

The whole road marking extraction and classification framework provides a promising solution for preloaded HD map creation, which further produces an essential road inventory dataset for road marking updates to support the development of AVs. The significant contributions of this paper are as follows. (1) A novel U-shaped convolution-deconvolution capsule network is constructed to road markings. The impacts of low-intensity contrast between road markings and their surrounding pavements, as well as varying point densities, are remarkably decreased through encoding the image patches at various locations. (2) A hybrid capsule network is proposed to categorize road markings with the assistance of a revised dynamic routing algorithm. The sibling framework of the capsule model and the fully connected capsule model achieves more effective performance for road marking classification. (3) To our knowledge, it is the first use of capsule-based neural networks for road marking extraction and classification in literature. And (4) a road marking dataset containing both 3D point clouds and manually labeled reference data in three types of road scenes (i.e., urban roads, highways, and underground garages) is constructed, which will be publicly accessible to motivate relevant research.

II. Related Work

A. Road Marking Extraction

Road markings are decorated on asphalt concrete pavements with highly light-reflective coatings, the intensities backscattered from road markings are considerably higher than surrounding pavements [18]. Accordingly, thresholding-related approaches have been widely applied to achieve road marking

extraction [19], [20]. To overcome the uneven distribution of intensities and point densities, a multi-thresholding approach was developed in [21] by first segmenting raw point clouds into data blocks with trajectory support. Next, each block was divided into different profiles with a certain width. Finally, road markings were extracted based on the peak values of intensity in each profile, followed by the spatial density filtering (SDF) algorithm for noise removal. In [22], combined with the multi-thresholding method, geometric feature filtering was employed to segment lane markings.

Furthermore, by converting 3D point clouds into 2D georeferenced intensity images, a multiscale tensor voting (MSTV) algorithm was proposed in [24] for discrete pixel elimination and road marking preservation. A weighted neighboring difference histogram (WNDH) algorithm was first performed to compute the intensity histogram of raw point clouds and determine adaptive thresholds. Subsequently, the MSTV and upward region-growing approaches were applied to ascertain candidate road marking pixels, accompanied by a morphological nearest algorithm for road marking extraction. However, it is still very challenging for such methods to effectively extract road markings from unorganized and high-density point clouds, especially with distinctive concavo-convex features.

Deep learning is taking off in object segmentation and object classification communities [23]. In [24], a lane marking extraction method was proposed based on the CNNs from MLS point clouds. A CNN framework designed for learning hierarchical features from upsampling-downsampling modules was first introduced for effectively and accurately detect lane-shaped markings. Then, both the length and spatial information related filters were utilized to optimize the extracted road markings. Moreover, in [25], an improved U-Net encoder-decoder framework was proposed by learning inherent features of road markings embedded in different data patches, which achieved promising flexibility and performance on point clouds with low-intensity contrast ratios. Nevertheless, these methods mainly concentrate on regular-shaped road markings (e.g., dashed lines and zebra crossings), it remains a challenge to deliver satisfactory results for complicated road markings (e.g., texts). Although it dramatically reduces the computational complexity by converting 3D point clouds into 2D rasterized images, these neural networks cannot capture pose or spatial information that is quite significant for road marking extraction in fluctuant terrain environments.

B. Road Marking Classification

Following the extraction process, many classification methods were developed to classify road markings into various categories for specific applications [26], [27]. In [28], an Euclidean distance-based clustering approach was implemented, followed by a voxel-based normalized segmentation algorithm for clustering unorganized road marking point clouds into large-size and small-size clusters. Afterward, large-size road markings were classified with the assistance of trajectory data and curb-lines. Then, a jointly trained Deep Boltzmann Machine (DBM) neural network followed by a multi-layer classifier was developed to recognize and categorize small-size markings effectively. Additionally, based on the geometric parameters (e.g., perimeter, area, and calculated width), the extracted road markings were classified in [29] by constructing a manually defined decision tree. However, it is challenging for this rule-based method to effectively classify complex road markings, such as words and arrows. Due to discrete noise, faded markings, and varied road environments, it is also difficult for these methods to accurately classify the incomplete road markings.

To achieve the superior road marking classification performance, a hierarchical classification framework was first designed in [20] by employing a multi-layer neural network to recognize arrows and pedestrian crossings. Then, the Structural Similarity Index (SSIM) algorithm was carried out to classify different types of arrows. Furthermore, a two-stage CNN-based hierarchical classification framework was introduced in [25]. At first, a multi-scale Euclidean clustering algorithm was implemented to classify large-size road markings (e.g., zebra crossing). Then, the remaining small-size road markings (e.g., texts and diamonds) were successfully classified into different groups by using a four-layer convolution network, followed by an optimized conditional generative adversarial network (c-GAN) to enhance the completeness of the extracted road markings. Although their experimental results indicated a highly promising solution in road marking classification, it is still challenging to eliminate the influences of small incompletions and deliver an end-to-end deep learning framework.

III. METHODS

This section details the theoretical and mathematical implementations of the developed capsule-based network architectures for road marking extraction and classification by using mobile LiDAR point clouds. This framework mainly contains three modules: data-preprocessing, road marking extraction, and road marking classification.

A. Module I: Data-Preprocessing

Since we mainly concentrate on road markings in this paper, the off-ground point clouds (e.g., trees, traffic lights, fences, and buildings) are first filtered out to strengthen the computational efficiency and reduce GPU memory consumption in the following processes. In our previous research [21], a revised curb-based road surface extraction method was introduced. Given the fact that urban roads are constructed with concrete curbs as road separation zones, road surface point clouds can be accurately and effectively segmented from the input point clouds depending on the fitted curb-lines. Detailed descriptions of this method and relevant parameter refinement are depicted in [21]. Herein, we employ this curb-based extraction method to segment road surface point clouds.

Moreover, existing studies have demonstrated that the height information of road point clouds conduces little to road marking segmentation [30], [31]. Thus, in this paper, road surface point clouds are projected to a 2D xy-plane and transformed into georeferenced intensity raster images. To this end, we employed an inverse distance weighting (IDW)

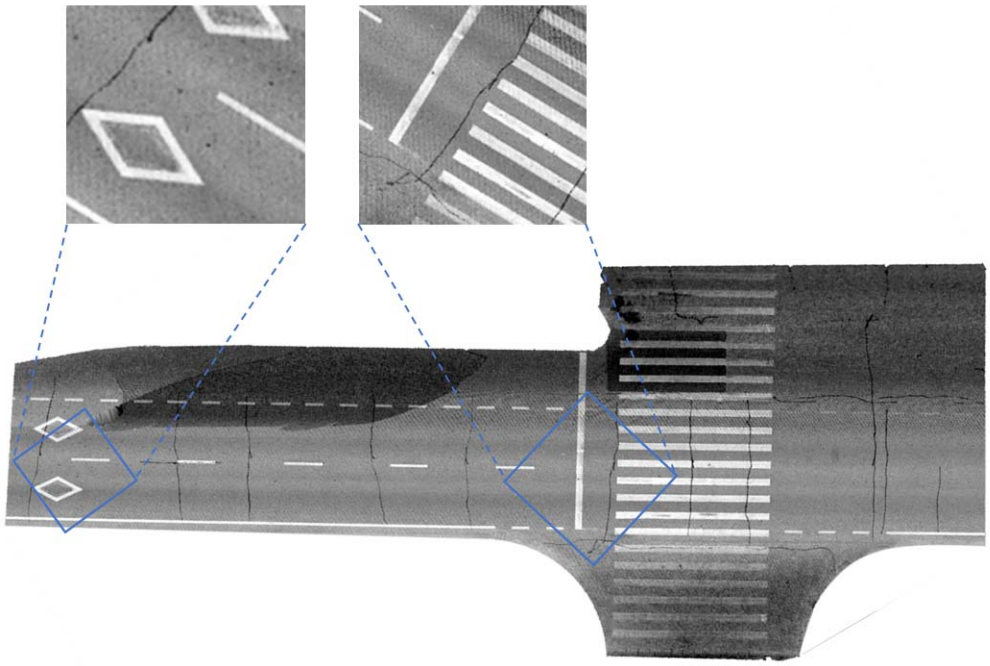

Fig. 1. The intensity image generated by using IDW interpolation.

interpolation algorithm to produce 2D intensity images by calculating the grey-scale value of a specific cell from its surrounding neighbors. Two rules are designed to determine the weight associated with each point: (1) a point with a larger intensity value has a higher weight, and (2) a point closer to the trajectory has a higher weight. The grid cell size should adequately preserve the details of different road markings and dramatically decreases the number of data that should be handled. Theoretically, A larger grid cell size is selected when performed on point clouds with lower density. Based on our previous experiments [29], we tested several grid cell sizes from 2.5 cm to 10 cm. The generated intensity images became blurred, and the computational cost was reduced. Moreover, with a grid cell size of 4 cm, the thinnest road markings (i.e., lane lines with a width of 15 cm) are well preserved in the generated intensity images, and the gaps among 3D point clouds are accordingly interpolated.

Furthermore, a high-pass filtering operation with a suitable kernel size is performed on the generated intensity images to minimize the influence of varying intensity values. The kernel size of this high-pass filter is ascertained based on prior knowledge and multiple experiments. Specifically, this kernel size should be not only large enough to comprise both road marking and road surface pixels but small enough to reduce the influence of the spatial variance and uneven distribution of the intensity. In this paper, the raster grid size of the generated intensity image and the kernel size of the high-pass filter are defined as 4 cm and 25 $\times$ 25, respectively. Fig. 1 indicates an example of the generated intensity image after implementing IDW interpolation and high-pass enhancement.

B. Module II: U-Shaped Capsule Network

Since a rasterized cell either denotes some road marking pixels or pavement pixels in the intensity images, we could regard the road marking extraction process as a basic binary classification task. Meanwhile, although capsule networks introduced by in [32] have achieved remarkable success for a broad range of computer vision problems particularly for digit recognition and small image classification (e.g., CT scans [33]), no studies yet exist in literature that employs capsule networks for road marking extraction from MLS point clouds. Comparing with conventional CNNs, capsule networks utilize vectorial neurons rather than scalar neurons to encode entity features. The instantiation parameters of different capsules indicate varying types of entities, while different lengths of capsules encode the probabilities of the existence of these entities, and different directions indicate their pose information [34]. Therefore, to demonstrate the effective performance of capsule networks in extracting road markings, we design a U-shaped capsule-based network using the 2D georeferenced intensity images.

In the training process, the generated intensity images are first manually labeled into two groups: positive training samples containing road marking pixels and negative training samples containing road surface pixels. Subsequently, a collection of local image patches with the size of 512 $\times$ 512 pixels are derived from the generated intensity image based on a sliding window mechanism. To ensure complete and extensive coverage of the training image, two adjacent image patches are generated with an overlapping size of p_s pixels. Moreover, such patches are fed into the multi-layer capsule networks for intrinsic feature extraction.

Fig. 2 shows a U-shaped convolution-deconvolution capsule network, which can learn not only intensity variance from massive labeled image patches but the shape and position information of road markings. This U-shaped capsule network consists of traditional convolutional layers, primary capsule layers, convolutional capsule layers, and deconvolutional

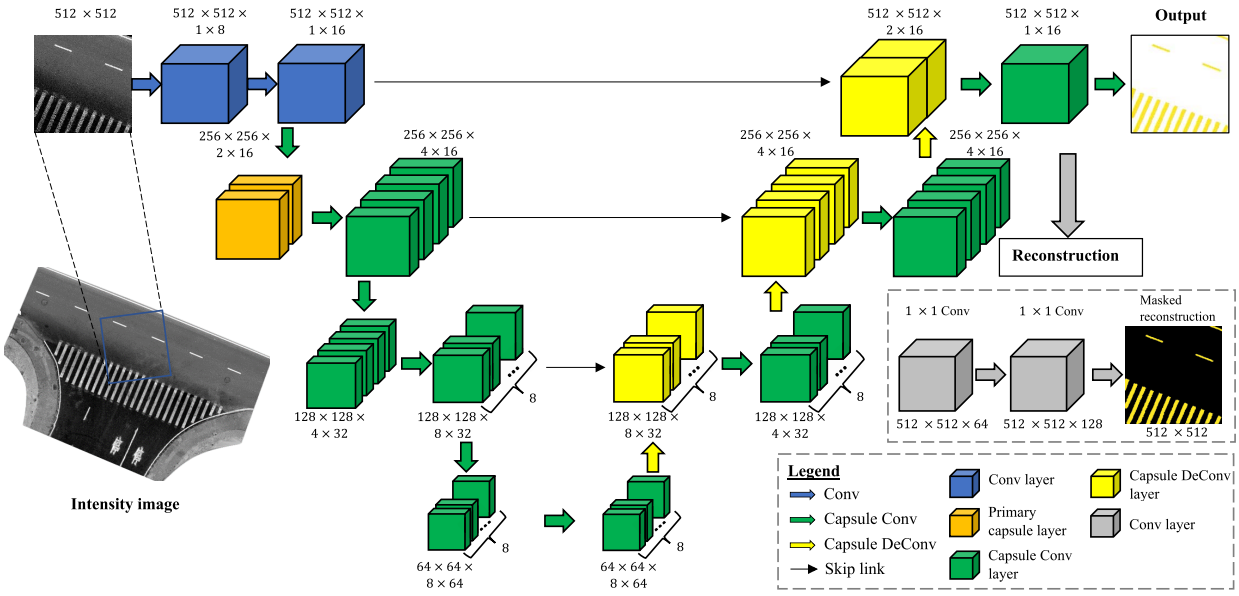

Fig. 2. Architecture of the proposed U-shaped capsule network.

capsule layers. The traditional convolution layers are designed to encode locally shallow features (e.g., edges and shapes) from the input local image patches via convolutional encodings. Such low-level features are afterward fed into high order capsules to learn in-depth features. Herein, we adopt the commonly employed rectified linear unit (ReLU) as the activation function.

In primary capsule layers, the shallow scalar feature encodings are transformed into high-order vectorized capsule representations. Assuming F_m and D_c are the number of feature maps and the dimension of capsules, respectively. Then, the kernels with the size of $F_m \times D_c$ are implemented in the following convolutional layer. Finally, the output feature maps are categorized as F_m groups, each of which consists of D_c feature maps.

The following convolutional capsule networks focus on encoding high-level capsule features and orientations by using capsule convolution operations. In general, the whole inputs to a capsule j are a weighted sum over all outputs from the capsules in the convolutional kernel in the previous layer:

$$C_j = \sum_i h_{ij} \cdot \widehat{V}_{j|i} \tag{1}$$

where C_j denotes the whole input to the capsule j, h_{ij} is the coupling coefficient showing the level of significance that capsule i in the previous layer activates capsule j, and $\widehat{V}_{j|i}$ indicates the predictions between capsule i and capsule j, which is calculated by:

$$\hat{V}_{j|i} = W_{ij} V_i \tag{2}$$

where W_{ij} denotes the weight matrix and V_i is the outcomes of capsule i. The sum of the weighting coefficients between capsule i and all its linked capsules in the previous layer is equal to 1, which is ascertained through a dynamic routing mechanism [32]. Moreover, a nonlinear "Squashing" activation function is employed to guarantee that different lengths of vectors in capsules are shrunk in the range of [0, 1] and results in the different probabilistic predictions. This squashing function is calculated by:

$$S_j = \frac{\|C_j\|^2}{1 + \|C_j\|^2} \cdot \frac{c_j}{\|C_j\|} \tag{3}$$

where C_j and S_j is the total input and output vector of capsule j, respectively. Furthermore, three deconvolutional capsule layers are designed to construct a diverse set of capsule types and propagate learned features from down-sampled images to the original images, thereby allowing the capsules to capture the most critical and intrinsic parameters of the input images. Based on the IDW interpolation algorithm, the feature propagation process is performed by interpolating feature values f of p_n pixels at coordinates of the p_{n-1} pixels. Then, such interpolated features are locally-constrained and concatenated with skip linked pixel features from the convolution capsule layers. Since we only focus on the distributions of the positive input class (i.e., road marking pixels) and regard the remaining pixels as background, we mask out all capsules except whose class labels match to the input image patch. This reconstruction process is conducted by employing three 1×1 convolutional layers.

Compared to the standard capsule network, we iteratively update the parameters by using the IoU as the cost function for model performance refinement, rather than binary cross-entropy. The IoU loss function, namely L_{IoU}, is calculated as follows:

$$L_{IoU} = -\frac{\sum_{i=1}^{N} \left(p_{pred}^i \cap p_{gt}^i \right)}{\sum_{i=1}^{N} \left(p_{pred}^i \cup p_{gt}^i \right)} \tag{4}$$

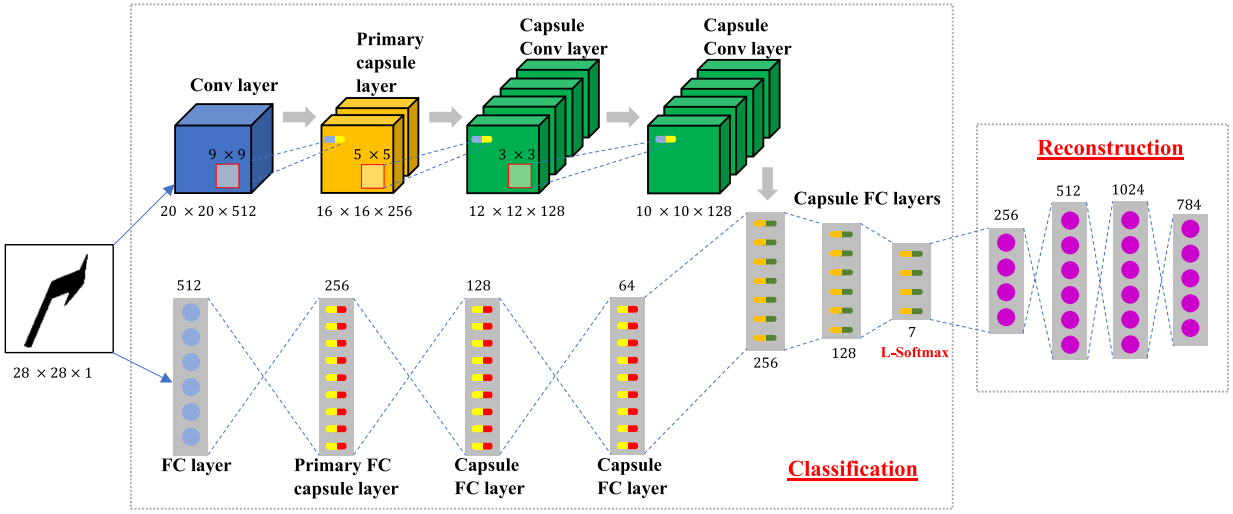

Fig. 3. Architecture of the proposed hybrid capsule network.

Algorithm 1 Revised Dynamic Routing Algorithm

1: procedure DYNAMIC ROUTING ($\hat{\mathbf{v}}_{\mathbf{j|i}}$, $\mathbf{r}$, $\mathbf{k}_h$, $\mathbf{k}_w$)
2: for all capsule $\mathbf{i}$ within a $\mathbf{k}_h * \mathbf{k}_w$ kernel in layer $\mathbf{l}$ and all capsule $\mathbf{j}$ in layer $(\mathbf{l}+\mathbf{1})$: $\mathbf{b_{ij}} \leftarrow \mathbf{0}$
3: for $\mathbf{r}$ iterations **do**
4: for all capsule $\mathbf{i}$ in layer $\mathbf{l}$: $\boldsymbol{l_i} \leftarrow \boldsymbol{L}_{\text{Softmax}}(\boldsymbol{b_i})$ $\Rightarrow$ L-Softmax computes Eq.5
5: for all capsule $\mathbf{j}$ in layer $(\mathbf{l}+\mathbf{1})$: $\boldsymbol{c}_j \leftarrow \sum_i \boldsymbol{h}_{ij} \cdot \widehat{\boldsymbol{V}}_{j|i}$
6: for all capsule $\mathbf{j}$ in layer $(\mathbf{l}+\mathbf{1})$: $\boldsymbol{s}_j \leftarrow$ squash $(\boldsymbol{c}_i)$ $\Rightarrow$ Squash computes Eq.3
7: for all capsule $\mathbf{i}$ in layer $\mathbf{l}$ and all capsule $\mathbf{j}$ in layer $(\mathbf{l}+\mathbf{1})$: $\boldsymbol{b}_{ij} \leftarrow \boldsymbol{b}_{ij} + \widehat{\boldsymbol{V}}_{j|i} \cdot \boldsymbol{s}_j$
8: **return** $\boldsymbol{s}_j$

where p^i_{pred} and p^i_{gt} is the i-th predicted road marking pixel and corresponding ground truth pixel, respectively. To minimize the L_{IoU}, the proposed U-shaped capsule network can extract more accurate and complete road markings in image patches. Moreover, images with various intensity are utilized as training data to decrease the impacts of intensity variation.

C. Module III: Hybrid Capsule Network

After road markings are segmented from the georeferenced intensity images, a hybrid capsule-based network is further proposed to categorize these road markings into several classes. Fig. 3 shows the workflow of the hybrid capsule framework, which mainly consists of two hierarchical networks (a convolutional capsule network and an FC capsule network) for, respectively, encoding high-level and low-level features from input images. As indicated in Fig.3, the convolutional capsule network comprises a standard convolutional layer and a primary capsule layer, followed by two convolutional capsule layers. By taking advantage of convolutional operation, the first convolutional layer functions to encode locally low-level features from the input images. Such extracted features are then fed into high order vectorial capsules for further feature encodings.

To enhance the weight update efficiency and improve inter-class separability, we propose a revised dynamic routing algorithm in this paper. Different from the dynamic routing method conducted in [32], the revised dynamic routing algorithm only route the child capsules within the user-specified kernel to the parent, rather than routing all child capsules to all parent capsules. The revised dynamic routing algorithm is described in Algorithm 1. Moreover, a large-margin Softmax loss [35] is adopted to emphasize intra-class compactness and overcome inter-class imbalance, which usually poses challenges by using standard Softmax loss. The large-margin Softmax loss is calculated as follows:

$$L_{Softmax} = -\log\left(\frac{e^{\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i})}}{e^{\|W_{y_i}\|\|x_i\|\psi(\theta_{y_i})} + \sum_{j\neq y_i} e^{\|W_j\|\|x_i\|\cos(\theta_j)}}\right) \quad (5)$$

where W_{y_i} indicates the y_i-th column of a FC-capsule layer W, θ_j $(0 < \theta_j < \pi)$ represents the angle between the vector W_j and x_i. Generally, we define $\varphi(\theta) = \cos(m\theta)$, $0 \le \theta \le \frac{\pi}{m}$ and $\varphi(\theta) = \mathcal{F}(\theta)$, $\frac{\pi}{m} \le \theta \le \pi$. m presents an integer that is closely correlated to the classification margin. With smaller m, the classification margin becomes smaller, and the learning objective becomes easier accordingly. Based on prior knowledge, m is defined as 5 in this paper. Furthermore, $\mathcal{F}(\theta)$ is a function that monotonically decreases in the range of $[0, \pi]$, while $\mathcal{F}\left(\frac{\pi}{m}\right)$ equals to $\cos\left(\frac{\pi}{m}\right)$. By such an introduction, the large-margin Softmax loss not only gains the main

strengths from Softmax loss but encodes inherent features at a large angular margin between different classes.

As perceived in Fig.3, the FC-capsule network contains a standard FC layer, a primary FC-capsule layer, and two FC-capsule layers. Intuitively, the FC layer is employed to encode shallow global features from the raw images. Then, such extracted global features are fed into high order capsules. Likewise, according to traditional fully-connected operations, the primary FC-capsule layer is generated. The corresponding units are equally divided into categories to construct a group of capsules. Meanwhile, two FC-capsule layers focusing on extracting inherent capsule features on a global scale are employed.

The two hierarchical networks encoding both local and global capsule features are combined through flattening and concatenation operations and further input into three FC-capsule layers for the classification task. Finally, four FC layers are employed to rebuild the input images, thus enable capsules to learn the most intrinsic and critical instantiation parameters of the raw images. Accordingly, we first mask out all classification capsules and keep the remaining capsules whose class labels correspond to the raw image. The instantiation parameters of these capsules are input to the reconstruction module for reconstruction.

The training parameters are effectively fine-tuned at the error back-propagation stage. Accordingly, the following multi-task loss function is adopted to refine the whole framework:

$$L = \sum_{i=1}^{M} \sum_{c=1}^{N} L_1^c + \delta \sum_{i=1}^{M} L_2^i \tag{6}$$

where L_1^c and L_2^i denote the classification loss and reconstruction loss, respectively. M and N are, respectively, the number of input training images and class-related capsules within the L-Softmax layer. δ indicates a regularization term. Accordingly, the classification loss for the specified class c is calculated as follows:

$$L_1^c = T_c \cdot \max\left(0, m^+ - \|S_c\|\right)^2 + \lambda\left(1 - T_c\right) \cdot \max\left(0, \|S_c\| - m^-\right)^2 \tag{7}$$

where $T_c = 1$, if the training image corresponds to class c. Otherwise, $T_c = 0$. m^+ and m^- are thresholds, respectively, that determine if a training image belongs to class c or not. In this paper, $m^+ = 0.93$ and $m^- = 0.07$ are predefined based on a set of experiments. λ represents a regularization term. Additionally, a smooth-L loss proposed by Girshick [36] is adopted to determine the reconstruction loss.

In the classification process, a group of road marking training data is manually labeled. Then, such training samples are augmented through rotation, scaling, crop, and illumination changes. Thus, we generate 7,000 training samples in total, with 1,000 samples for each road marking type. Moreover, to reduce the computational cost, all training samples are resized to 28×28 pixels before feeding into the classification network. Finally, the class label of a road marking image is ascertained using the following equation:

$$K^* = \text{argmax}_c \, \|S_c\| \tag{8}$$

TABLE I

THE CATEGORY AND QUANTITY OF LABELED ROAD MARKINGS

Category	Quantity of labeled road markings	Quantity of labeled road markings after augmentation
Lane line	366	1,000
Dashed line	423	1,000
Zebra crossing	219	1,000
Straight arrow	330	1,000
Turn arrow	298	1,000
Diamond	305	1,000
Text	318	1,000

where $\|S_c\|$ denotes the output of the classification network. This class label is therefore assigned to the image as its predicted road marking type.

IV. DATASET AND IMPLEMENTATION DETAILS

In this paper, the mobile LiDAR point clouds were captured using a RIEGL-VMX 450 MLS system in both urban and highway road sections with free-flowing traffic. The RIEGL-VMX 450 MLS platform contains two full-circle RIEGL VQ-450 laser heads, which could achieve a 400 lines/sec scan frequency in a "Butterfly" configuration pattern. The average point density of these data is over 4,500 pts/m^2, and the absolute measurement accuracy can reach 8 mm. Additionally, we collected some pavement point clouds in underground garage environments using our self-assembled backpack laser scanning (BLS) system. The configuration specifications of this BLS system are detailed in [37]. This BLS system is equipped with two Velodyne VLP-16 laser heads, which can achieve a scanning distance of 100 m with 1,700 pts/m^2 point density and 3 cm measurement accuracy. Compared to point clouds obtained by MLS systems, the point clouds captured by the BLS system are of low quality due to low-end LiDAR sensors and poor illumination conditions in underground garage environments.

Moreover, we manually annotated the data to build a road marking dataset. To end this, all road markings were first labeled pixel-by-pixel on the generated intensity images based on visual interpretation. Then, all these road markings were segmented as separate training samples by employing a clustering method. Thus, each labeled image only contains one type of road marking. Finally, the 2D coordinates and class type of each road marking pixel were recorded. Since the number of some road markings in certain classes (e.g., different Chinese words) is limited, such classes were merged and accordingly augmented through rotation, translation, and scaling operations. As listed in Table I, a total of seven categories of road markings were generated in this paper.

Since the road marking types in these three road scenes (e.g., urban roads, highways, and underground garages) are similar, 7,000 samples were utilized to extract and classify road markings. The whole dataset was split into 60%, 20%, and 20% subsets for training, validation, and testing, respectively. According to prior knowledge and multiple experiments, different hyperparameters, i.e., the batch size, initial

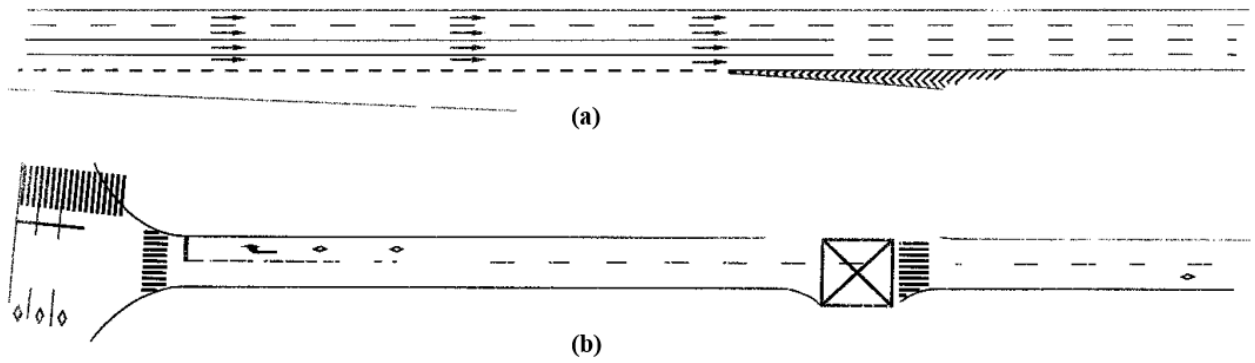

Fig. 4. Road marking extraction results using the proposed U-shaped capsule network. (a) Highway scene, and (b) urban road scene.

learning rate, and dropout rate, were fine-turned in the training process for the optimal combination. Accordingly, the batch size, initial learning rate, dropout rate, and epochs were [8, 0.0001, 0.80, 400] for the U-shaped road marking extraction model, respectively, and [32, 0.0005, 0.80, 300] for the hybrid capsule-based road marking classification model. We tested the proposed models using TensorFlow 2.0 on Intel® i7-8700K CPU @3.70 GHz, Nvidia® 1080-Ti GPU, and 32 GB RAM.

V. Results and Discussion

A. Road Marking Extraction Results

In this paper, according to the manually labeled reference data, the following three evaluation metrics, i.e., precision, recall, and F_1-score [38], were adopted to conduct the quantitative performance evaluation of road marking extraction:

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{9}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{10}$$

$$F_1\text{-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

where T_P, F_P, and F_N present true positive, false positive, and false negative segmentation outputs, respectively. Specifically, the precision shows the percentage that the extracted road markings are valid, while the recall represents the completeness of the extracted road markings. Moreover, F_1-score is a weighted average score of by analyzing both precision and recall.

The U-shaped convolution-deconvolution capsule architecture was proposed for road marking extraction on the generated intensity image patches with 4 cm resolutions. Fig. 2 presents the fine-tuned network configurations based on multiple experiments, which details the number of capsules and the sizes of feature maps after standard or capsule-based convolutional operations. The training samples captured in urban roads, highways, and underground garages were used to evaluate the extraction performance. A series of experiments were performed to determine the optimal parameters, such as the overlapping size p_s between two adjacent image patches. In fact, an increasing overlapping size can produce not only better classification performance but also more image patches resulting in slow training speed. Therefore, to balance the classification performance and computational burden, p_s was defined as 128 in this paper.

Fig. 4(a) indicates the road marking extraction results in a highway scene. These highways are newly built and well maintained with clear road markings (i.e., lane lines and dashed lines), which enables the vectorial capsules to effectively learn inherent road marking features (e.g., varying intensities) at the training stage. Although we conducted binary classification to minimize the influence of low contrast between road surfaces and road markings, some pixels belonging to the lane lines were misclassified into road surfaces. Compared to highway road scenes, road markings painted on urban road surfaces are usually worn, which leads to dilemmas and uncertainties for high-accuracy road marking extraction. As shown in Fig. 4(b), most road markings were successfully extracted, which demonstrates our proposed network can effectively extract road markings even in complex urban road scenes. However, due to heavy traffic flow and pavement corrosion, some road markings are heavily worn and incomplete, which brings in enormous difficulties for road marking extraction. Besides, some road markings are covered by thick dust due to late maintenance, thus leading to varying intensities and low contrasts with the surrounding pavements. Although these conditions inevitably occurred in urban road scenes rather than highways, our proposed U-shaped capsule network could achieve reliable performance and deliver promising results for road marking extraction under various road conditions.

Additionally, Fig. 5 illustrates the road marking extraction results by utilizing the proposed networks in a 50 $\times$ 40 m^2 underground garage scene. Although the point density of these underground garage data is lower than point clouds obtained from the MLS systems, it can be solved by converting 3D point clouds into 2D images. The intensity value of a certain pixel is calculated based on its surrounding neighbors. Moreover, capsule convolutions encode not only intensity contrast but pose (e.g., shape and position) information. Consequently, the majority of road markings were accurately detected and extracted, while only a few pixels belonging to arrows and zebra crossings were misclassified into road surfaces due to inevitable occlusions and poor illuminations.

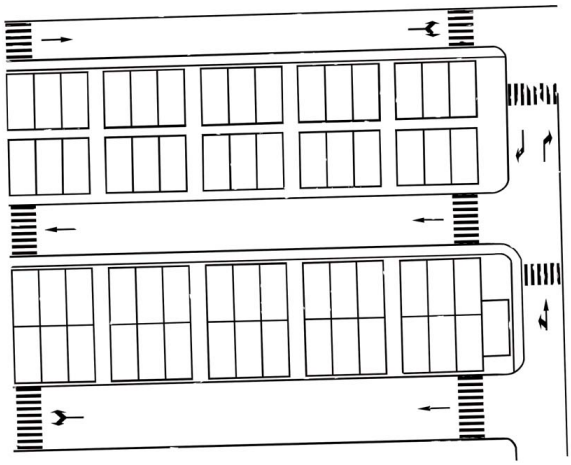

Fig. 5. Road marking extraction results in an underground garage scene.

TABLE II
ACCURACY ASSESSMENT OF ROAD MARKING EXTRACTION IN URBAN ROADS, HIGHWAYS, AND UNDERGROUND GARAGES

Test dataset	Precision(%)	Recall(%)	F_1-score(%)
Urban scene			
1	94.97	90.16	92.50
2	94.76	89.31	91.95
3	95.03	91.28	93.12
Average	94.92	90.25	92.52
Highway scene			
1	97.08	92.03	94.49
2	95.46	91.62	93.50
3	95.89	89.74	92.71
Average	96.14	91.13	93.57
Underground garage			
1	91.87	89.93	90.89
2	92.06	90.02	91.03
3	92.84	90.55	91.68
Average	91.26	90.17	91.20

Tables II indicates the quantitative accuracy assessment of the road marking extraction results in three different road scenes. Consequently, the propose U-shaped capsule network delivered an average precision, recall, and F_1-score of 94.92%, 90.25%, 92.52% in urban road scenes, and 96.14%, 91.13%, and 93.57% in highways, and 91.26%, 90.17%, and 91.20% in underground garages, respectively. Because of the high point density, few occlusions, and good illumination conditions in highways, our proposed U-shaped capsule network achieved more superior performance for road marking extraction in highway scenes than both urban roads and underground garages. The issue of intensity variation is considerably solved by learning the patches at different locations. The unavoidable errors occurring at the stage of manually annotated label generation could bring in challenges for robust and effective road marking extraction. Furthermore, some road markings are worn and incomplete, resulting in the sizes of such road markings smaller than the manually labeled reference data. Therefore, the road marking extraction performance of the developed model is underestimated in the experimental results. Our experiments demonstrated that our proposed U-shaped capsule-based model is able to learn inherent features (e.g., intensity and shape) for road marking extraction by using different kinds of point clouds. Such data are obtained from complex road scenes, with low point densities, in poor illumination conditions, and with uneven intensity distributions.

B. Comparative Study for Road Marking Extraction

A comparative study was conducted to evaluate the road marking extraction performance by using the developed models and existing algorithms, including Cheng *et al.*[29], Ma *et al.* [21], and Wen *et al.* [25]. The test datasets were collected from urban roads, highways, and underground garages with low-intensity contrast and incomplete point clouds, which contain many categories of road markings (e.g., lines, arrows, and texts). We re-implemented these three methods on our testing datasets. Fig. 6 shows the road markings on the large-scale roadways extracted using four methods.

Accordingly, mobile LiDAR point clouds were transformed from 3D point clouds into 2D georeferenced images at the stage of road marking extraction in both Cheng's [29] and Wen's [25] methods. Cheng's [28] method employed Otsu's thresholding approach [39] for road marking segmentation according to the discriminant analysis, which requires the generated intensity image should be bimodal with uniform illumination conditions. Thus, it is difficult to completelyy extract road markings from low-intensity contrast point clouds, especially from underground garage data with poor illuminations. Meanwhile, by projecting 3D point clouds onto a horizontal xy-plane, Wen's method [25] performed a revised U-Net neural network to segment different types of road markings. However, the Softmax activation function used at the stage of road marking extraction cannot capture intra-class compactness, thereby resulting in limitations in road marking extraction from low-intensity contrast point clouds. Additionally, Ma's [21] method mainly concentrated on determining adaptive intensity thresholds on local scales for road marking extraction. Nevertheless, it is quite difficult to define suitable threshold values in different road scenes.

Table III shows the overall performance of different methods for road marking extraction by using precision, recall, and F_1-score evaluation metrics. Cheng's method [29], Ma's method [21], and Wen's method [25] achieved an average of precision, recall, and F_1-score of 27.48%, 32.32%, and 29.69%, 67.16%, 59.38%, and 63.01%, and 93.36%, 88.97%, and 91.09%, respectively; while our proposed method delivered an average of precision, recall, and F_1-score of 94.11%, 90.52%, and 92.43%, respectively. As can be seen, the road markings were extracted incompletely or even lost information by using such three comparative methods. In contrast, the proposed U-shaped capsule network is capable of achieving better performance with higher accuracy and less noise in all road scenes.

Moreover, the capsule-based convolutional operations in our proposed model can not only capture the salient features embedded in intensity values but also the shape and

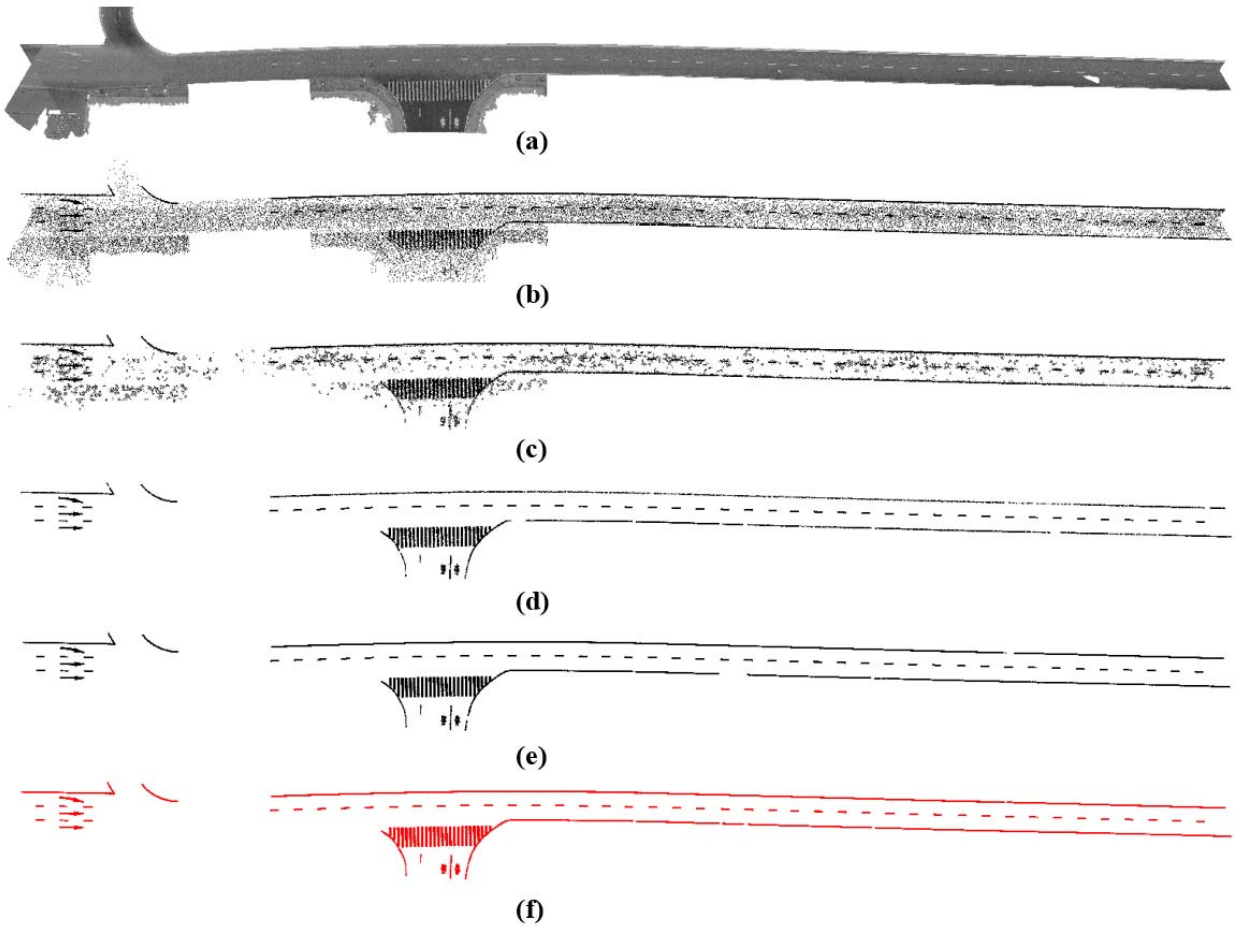

Fig. 6. Road marking extraction results using different methods. (a) Raw road surface, (b) Cheng *et al.* [29], (c) Ma *et al.* [21], (d) Wen *et al.* [25], (e) ours, and (f) manually labeled reference data.

TABLE III
ROAD MARKING EXTRACTION RESULTS BY USING DIFFERENT METHODS

Method	**Road scene**	**Precision**(%)	**Recall**(%)	**F_1-score**(%)
Cheng et al. [29]	Urban	27.35	33.82	30.24
	Highway	30.57	34.10	32.24
	Underground garage	24.52	29.03	26.59
	Average	27.48	32.32	29.69
Ma et al. [21]	Urban	62.63	53.19	57.53
	Highway	70.13	65.54	67.76
	Underground garage	68.73	59.42	63.74
	Average	67.16	59.38	63.01
Wen et al. [25]	Urban	92.15	89.33	90.72
	Highway	95.97	87.52	91.55
	Underground garage	91.95	90.07	91.00
	Average	93.36	88.97	91.09
Ours	Urban	94.92	90.25	92.52
	Highway	96.14	91.13	93.57
	Underground garage	91.26	90.17	91.20
	Average	94.11	90.52	92.43

position information of the road markings, which makes our proposed model outperformed than other methods in terms of correctness and completeness. However, some road markings were not correctly segmented from the generated intensity image patches because of the occlusions of other road users (e.g., vehicles and cyclists) during the data acquisition of MLS systems. Additionally, uneven intensity distribution and varying illumination conditions from different road scenes also make effective and accurate road marking extraction challenging. On the whole, the proposed U-shaped capsule network

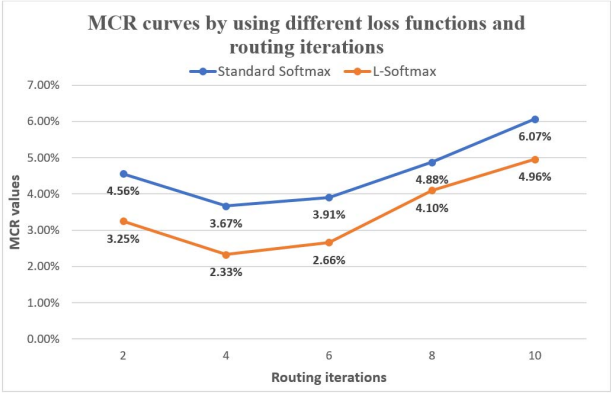

Fig. 7. Road marking classification results by using Softmax loss and L-Softmax loss with different routing iterations.

designs a promising solution for road marking extraction from massive and unordered 3D MLS point clouds.

C. Road Marking Classification Results

The experimental results of the hybrid capsule-based road marking classification neural network were evaluated based on the misclassification rate (MCR), which is calculated by:

$$\mathrm{MCR} = \frac{\sum_{i=1}^{N} T_i}{N} \tag{12}$$

where N is the total number of road marking pixels. Specifically, $T_i = 0$, if the road marking is correctly classified. Otherwise, $T_i = 1$. The proposed road marking classification method was evaluated in urban roads, highways, and underground garages.

Accurate and robust road marking classification is essential for fully autonomous driving to design efficient navigation paths and avoid accidents in changing road conditions. Based on the proposed hybrid capsule-based road marking classification method, the extracted road markings were further classified into seven categories, i.e., lane line, dashed line, zebra crossing, straight arrow, turn arrow, diamond, and text. Fig. 3 details the optimal network configurations through computational complexity and classification accuracy analysis. In the revised dynamic routing algorithm, the number of routing iterations, namely r, plays a significant role to balance between classification performance and computational complexity. Accordingly, multiple experiments were carried out to verify the robustness and convergence of the revised dynamic routing algorithm with different iterations. In fact, more routing iterations usually strength the classification performance but results in overfitting problems.

Additionally, to emphasize intra-class compactness and overcome inter-class imbalance, the L-Softmax was adopted in this paper to guide weight updates in the hybrid capsule-based classification architecture. By calculating different MCRs after 800 epochs from an urban road scene, Fig. 7 illustrates the classification performance of different loss functions (i.e., standard Softmax loss and L-Softmax loss) with varying routing iterations. Intuitively, the standard Softmax loss and L-Softmax loss deliver minimal MCRs of 3.67% and 2.33%, respectively, by defining the routing iterations r as 4. Consequently, the L-Softmax loss function can achieve a lower misclassification rate compared with the standard Softmax loss, which demonstrates the L-Softmax loss function can significantly boost the capsule-based classification network performance by using mobile LiDAR point clouds.

TABLE IV

THE MISCLASSIFICATION RATES OF THE PROPOSED CLASSIFICATION NETWORK IN DIFFERENT ROAD SCENES

Evaluation metric	Road scene			Average
	Urban	Highway	Underground	
MCR	2.16%	4.87%	3.23%	3.42%

Table IV shows that the proposed hybrid capsule-based network can achieve an average of 3.42% MCR, which demonstrates that most road markings were correctly classified in three road scenes. Fig. 8 shows the road marking classification results from a complex urban road environment. This scenario is a typical urban road that consists of zebra crossings, lane lines, diamonds, and texts, etc. Various colors denote different road marking types, while the misclassified markings are identified with black boxes. As can be seen, most road markings were correctly classified with a 2.16% MCR. Some lane lines are broken due to moving overloaded trucks and late road maintenance, resulting in broken lane lines similar to dashed lines. Therefore, such lane lines were inaccurately grouped into dashed lines. Additionally, some zebra crossings and straight arrows were misclassified as lane lines due to the erroneous results obtained in the process of road marking extraction. Similarly, Fig.9 indicates the classification results from highway point cloud data with a 4.87% MCR. Intuitively, the proposed model is capable of correctly classifying most road markings. However, some lane lines were incorrectly identified as dashed lines, resulting from the incomplete extraction of road markings. Additionally, the shapes of some broken lane lines are very similar to straight arrows, which also leads to false classification results.

Moreover, we evaluated the road marking classification performance of the proposed hybrid capsule-based network in underground garage scenes. As shown in Fig. 10, our proposed method can deliver satisfactory classification results from low-intensity contrast point clouds in poor illumination and GNSS-signal denied environments. However, some lane lines were misclassified into dashed lines, because these lane lines were not thoroughly extracted at the stage of road marking extraction. The hybrid capsule road marking classification model regarded them as independent dashed lines and incorrectly trained. Besides, a turn arrow was incorrectly identified as straight arrows since the shape of this turn arrow looks much like straight arrows. The MCR of road marking classification in underground garage environments is 3.23%.

To further demonstrate the effectiveness and robustness of our proposed models, we also evaluated the road marking extraction and classification performance by using low-quality

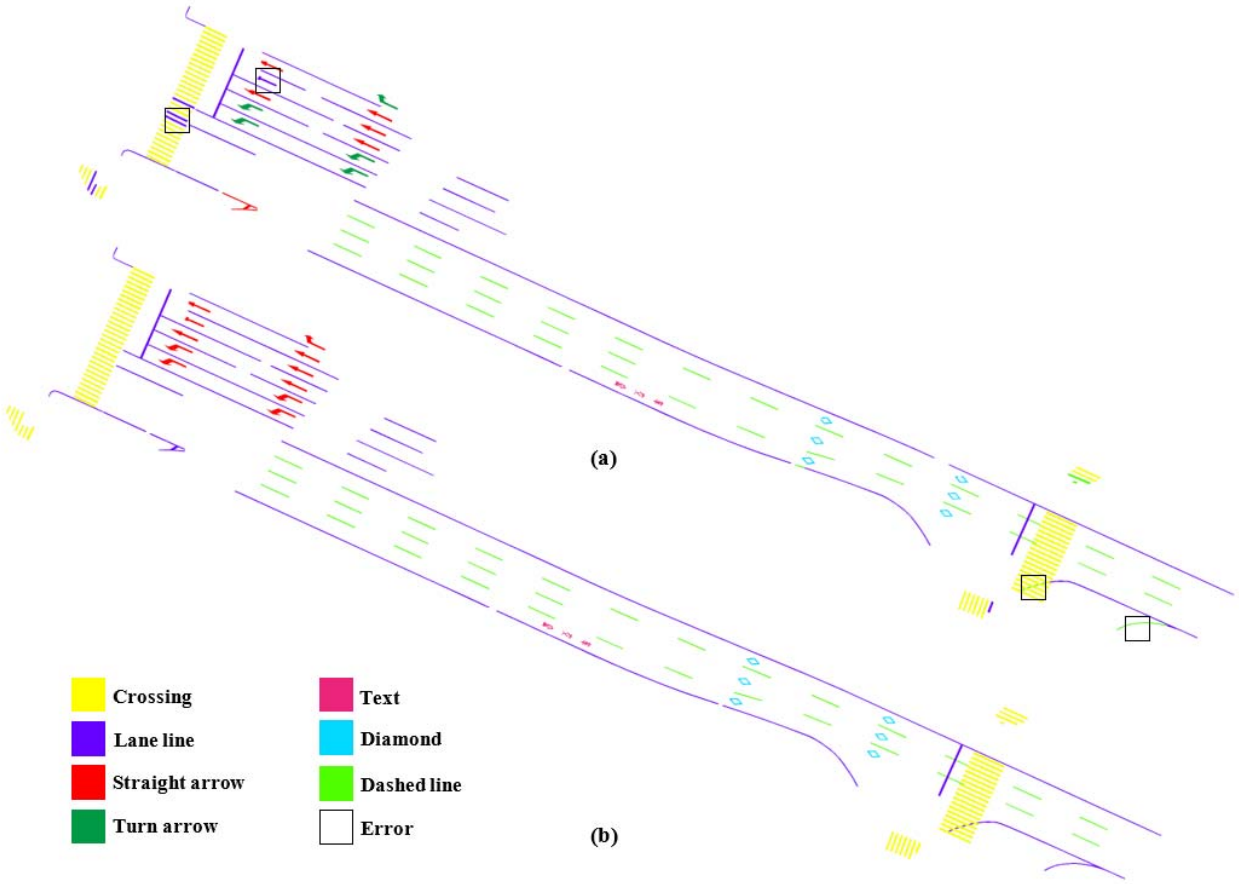

Fig. 8. Road marking classification results from an urban road scene. (a) Classification results, and (b) manually labeled reference data.

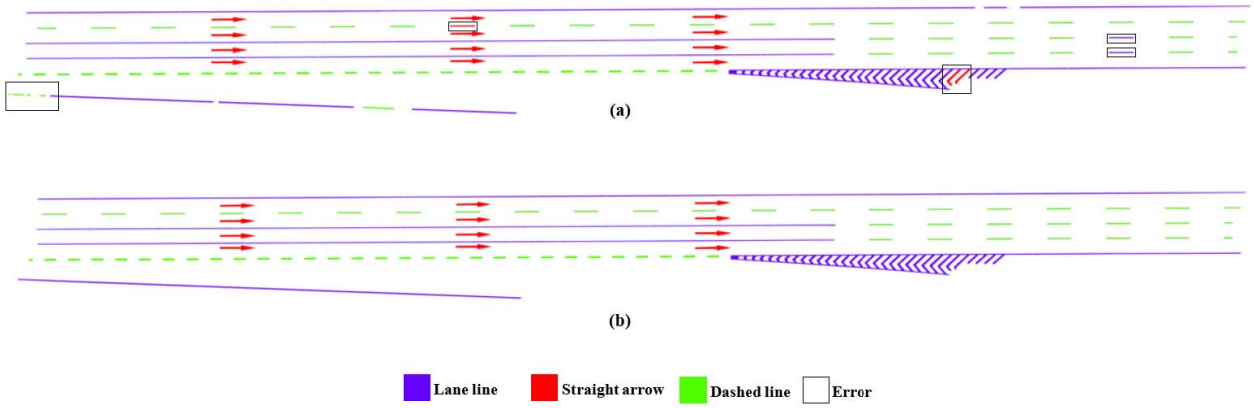

Fig. 9. Road marking classification results from a highway road scene. (a) Classification results, and (b) manually labeled reference data.

point cloud data. Accordingly, Figs. 11(a)-(c) presents the road surface with low-intensity contrast between road markings and their surrounding environments, the generated intensity image with diverse point densities, and the road surface with worn and incomplete road markings, respectively. Figs. 11(d)-(f) indicates the corresponding road marking extraction and classification results, respectively. As can be perceived, the proposed capsule-based deep learning networks are capable of effectively extracting and classifying road markings from low-quality input data. On the whole, the proposed capsule-based networks can deliver accurate road marking extraction and classification results on complex road environments, which provides a promising solution in fully autonomous driving and HD map creation.

D. Computational Efficiency Evaluation

In this paper, the proposed framework mainly contains three modules: data-preprocessing, U-shaped capsule network for road marking extraction, and hybrid capsule network for road marking classification. Table V lists the computational cost

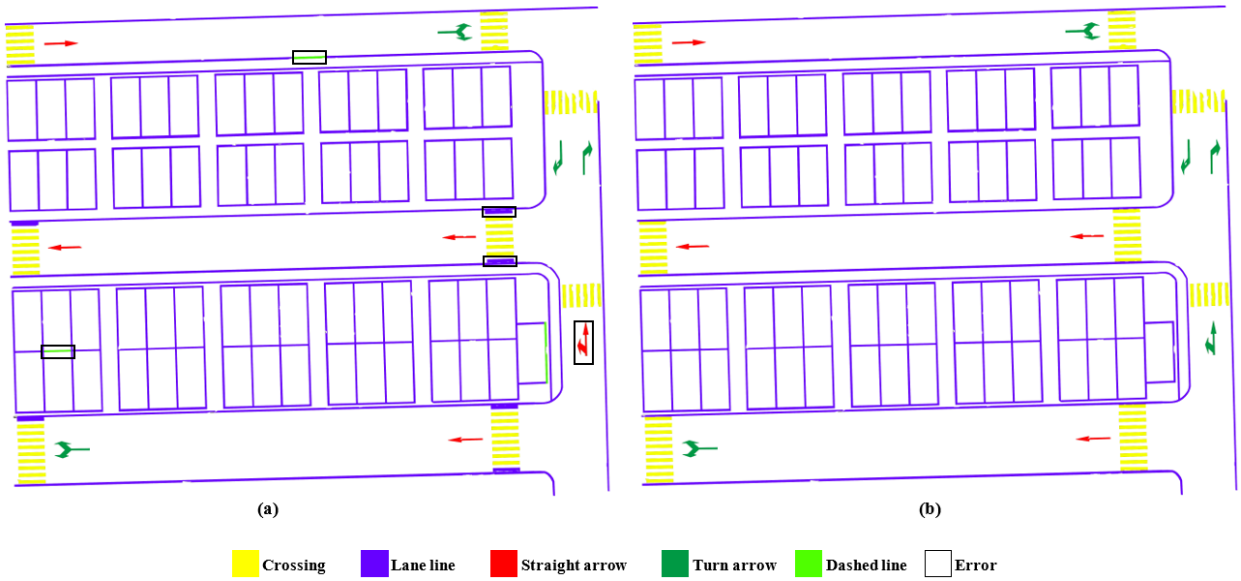

Fig. 10. Road marking classification results from an underground garage scene. (a) Classification results, and (b) manually labeled reference data.

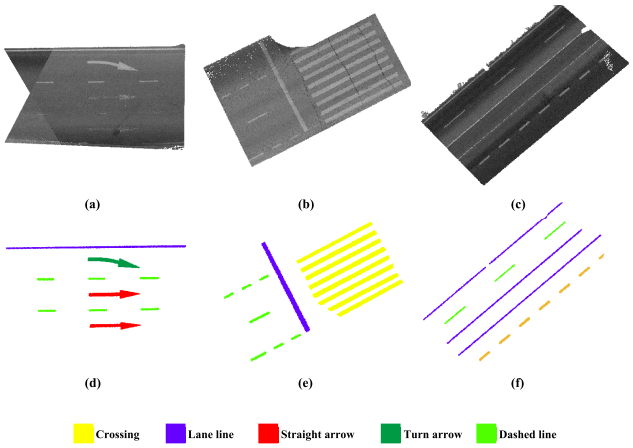

Fig. 11. Road marking extraction and classification results on low-quality data. (a) road surface with low-intensity contrast between road markings and their surrounding environments, (b) generated intensity image with varied point densities, (c) road surface with worn and incomplete road markings, and (d)-(f) are corresponding road marking extraction and classification results.

TABLE V
THE COMPUTATIONAL EFFICIENCY OF THE PROPOSED METHODS IN DIFFERENT ROAD SCENES

Processing module	Road scene			Average
	Urban	Highway	Underground	
Data processing (s)	40.25	31.36	37.80	36.47
Road marking extraction (s)	3.37	2.74	3.11	3.07
Road marking classification (s)	2.77	2.35	2.63	2.58

for each module, as well as the average time complexity across all over three road scenes. The average processing time of data-preprocessing, road marking extraction, and road marking classification are 36.47s, 3.07s, 2.58s, respectively. In fact, most of the processing time is spent in the data-preprocessing phase. Accordingly, multiple threads and GPU parallel computing techniques can be further applied to not only boost the computational efficiency in the process of 3D point cloud projection but dramatically accelerate capsule-based networks.

VI. CONCLUSION

This paper handles the dilemmas related to thresholding-based methods for road marking extraction and classification. Such dilemmas result in robustness reduction and computational complexity when dealing with 3D unordered and high-density point clouds captured by MLS systems, most remarkably due to its varying point density and intensity, as well as low-intensity contrast between road markings and their neighboring pavements. In this paper, we have designed two novel capsule-based network architectures for road marking extraction and classification, respectively, from highly dense MLS point clouds with an irregular data format. Moreover, a road marking dataset containing both 3D point clouds collected by both MLS and BLS systems and manually labeled reference data is created from three types of road environments, including urban roads, highways, and underground garages, while the proposed models were accordingly evaluated by estimating robustness and efficiency using this self-built dataset.

In the extraction process, we designed a U-shaped capsule-based network to extract road markings using 2D georeferenced intensity images. The experimental results demonstrated that our proposed model is capable of effectively encoding high-level features (e.g., changing intensity and pose information) with significantly enhanced road marking extraction performance. The comparative study indicated that our developed method achieved better performance than other thresholding-based and U-Net based methods, while delivered

an average of precision, recall, and F_1-score of 94.11%, 90.52%, and 92.43%, respectively, in three different road scenes.

In the classification process, a hybrid capsule-based network was proposed to classify seven types of road markings. Compared to those manually defined rule-based classification methods, our proposed method can automatically learn more salient features embedded in intensity values, as well as the shape information of the road markings by using a revised dynamic routing algorithm and powerful L-Softmax loss function. The quantitative evaluation indicated that the hybrid capsule-based network achieved an average of 3.42% MCR in changing road environments.

In conclusion, our experimental results have demonstrated that capsule-based networks are capable of effectively extracting inherent features from massive MLS point clouds and achieving superior performance in road marking extraction and classification tasks. For further research, we are dedicated to developing a point-wise end-to-end deep learning framework for robust and effective road marking extraction and classification.

REFERENCES

[1] R. C. Luo and M. Chiou, "Hierarchical semantic mapping using convolutional neural networks for intelligent service robotics," *IEEE Access*, vol. 6, pp. 61287–61294, 2018.

[2] H. Chu, L. Guo, B. Gao, H. Chen, N. Bian, and J. Zhou, "Predictive cruise control using high-definition map and real vehicle implementation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11377–11389, Dec. 2018.

[3] D. Betaille and R. Toledo-Moreo, "Creating enhanced maps for lane-level vehicle navigation," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 786–798, Dec. 2010.

[4] H. Gi Jung, Y. Hee Lee, and J. Kim, "Uniform user interface for semiautomatic parking slot marking recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 616–626, Feb. 2010.

[5] M. Feng, B. Fang, and W. Yang, "Fast reducing perspective distortion for road marking recognition," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[6] B. Mathibela, P. Newman, and I. Posner, "Reading the road: Road marking classification and interpretation," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2072–2081, Aug. 2015.

[7] S. Lee *et al.*, "VPGNet: Vanishing point guided network for lane and road marking detection and recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1947–1955.

[8] 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.

[9] G. Mattyus, 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. (CVPR)*, Jun. 2016, pp. 3611–3619.

[10] L. Ma, Y. Li, J. Li, C. Wang, R. Wang, and M. Chapman, "Mobile laser scanned point-clouds for road object detection and extraction: A review," *Remote Sens.*, vol. 10, no. 10, p. 1531, 2018.

[11] B. Yang, Y. Liu, Z. Dong, F. Liang, B. Li, and X. Peng, "3D local feature BKD to extract road information from mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 329–343, Aug. 2017.

[12] Y. Chen *et al.*, "Rapid urban roadside tree inventory using a mobile laser scanning system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3690–3700, Sep. 2019.

[13] R. Wan, Y. Huang, R. Xie, and P. Ma, "Combined lane mapping using a mobile mapping system," *Remote Sens.*, vol. 11, no. 3, p. 305, 2019.

[14] H. Guan, J. Li, Y. Yu, C. Wang, M. Chapman, and B. Yang, "Using mobile laser scanning data for automated extraction of road markings," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 93–107, Jan. 2014.

[15] J. Jung, E. Che, M. J. Olsen, and C. Parrish, "Efficient and robust lane marking extraction from mobile lidar point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 1–18, Jan. 2019.

[16] A. Jaiswal, W. AbdAlmageed, Y. Wu, and P. Natarajan, "Capsule-gan: Generative adversarial capsule network," in *Proc. ECCV*, 2018, pp. 1–10.

[17] K. Duarte, Y. Rawat, and M. Shah, "CapsuleVOS: Semi-supervised video object segmentation using capsule routing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8480–8489.

[18] D. Zai *et al.*, "3-D road boundary extraction from mobile laser scanning data via supervoxels and graph cuts," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 802–813, Mar. 2018.

[19] S. Soilán, P.-C. Río-Barral, and R. Arias, "Review of laser scanning technologies and their applications for road and railway infrastructure monitoring," *Infrastructures*, vol. 4, no. 4, p. 58, 2019.

[20] M. Soilán, B. Riveiro, J. Martínez-Sánchez, and P. Arias, "Segmentation and classification of road markings using MLS data," *ISPRS J. Photogramm. Remote Sens.*, vol. 123, pp. 94–103, Jan. 2017.

[21] L. Ma, Y. Li, J. Li, Z. Zhong, and M. A. Chapman, "Generation of horizontally curved driving lines in HD maps using mobile laser scanning point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1572–1586, May 2019.

[22] C. Ye, J. Li, H. Jiang, H. Zhao, L. Ma, and M. Chapman, "Semi-automated generation of road transition lines using mobile laser scanning data," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 9, 2019, doi: 10.1109/TITS.2019.2904735.

[23] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[24] B. He, R. Ai, Y. Yan, and X. Lang, "Lane marking detection based on convolution neural network from point clouds," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 2475–2480.

[25] C. Wen, X. Sun, J. Li, C. Wang, Y. Guo, and A. Habib, "A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 178–192, Jan. 2019.

[26] E. Che, J. Jung, and M. Olsen, "Object recognition, segmentation, and classification of mobile laser scanning point clouds: A state of the art review," *Sensors*, vol. 19, no. 4, p. 810, 2019.

[27] H. Guan, J. Li, S. Cao, and Y. Yu, "Use of mobile LiDAR in road information inventory: A review," *Int. J. Image Data Fusion*, vol. 7, no. 3, pp. 219–242, Jul. 2016.

[28] Y. Yu, J. Li, H. Guan, F. Jia, and C. Wang, "Learning hierarchical features for automated extraction of road markings from 3-D mobile LiDAR point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 709–726, Feb. 2015.

[29] M. Cheng, H. Zhang, C. Wang, and J. Li, "Extraction and classification of road markings using mobile laser scanning point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 1182–1196, Mar. 2017.

[30] B. Yang, L. Fang, Q. Li, and J. Li, "Automated extraction of road markings from mobile lidar point clouds," *Photogramm. Eng. Remote Sens.*, vol. 78, no. 4, pp. 331–338, Apr. 2012.

[31] B. Yang, L. Fang, and J. Li, "Semi-automated extraction and delineation of 3D roads of street scene from mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 80–93, May 2013.

[32] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.

[33] R. LaLonde and U. Bagci, "Capsules for object segmentation," 2018, *arXiv:1804.04241*. [Online]. Available: http://arxiv.org/abs/1804.04241

[34] Y. Yu *et al.*, "A hybrid capsule network for land cover classification using multispectral LiDAR data," *IEEE Geosci. Remote Sens. Lett.*, early access, Sep. 23, 2019, doi: 10.1109/LGRS.2019.2940505.

[35] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. ICML*, vol. 2, no. 3, 2016, p. 7.

[36] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[37] C. Wang *et al.*, "Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud," *ISPRS J. Photogramm. Remote Sens.*, vol. 143, pp. 150–166, Sep. 2018.

[38] D. M. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *J. Mach. Learn. Tech.*, vol. 2, no. 1, pp. 37–63, 2011.

[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

Lingfei Ma (Student Member, IEEE) received the B.Sc. degree in GIScience from the China University of Geoscience, Beijing, China, in 2015, and the B.Sc. and M.Sc. degrees in geomatics specializing in remote sensing from the University of Waterloo, Canada, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree in photogrammetry and remote sensing.

His research interests include autonomous driving, high-definition mapping, mobile laser scanning, intelligent processing of point clouds, 3D scene modeling, and deep learning.

Ying Li (Graduate Student Member, IEEE) received the B.Eng. degree in geomatics engineering from the Hefei University of Technology, China, in 2014, and the M.Sc. degree in remote sensing from Wuhan University, China, in 2017. She is currently pursuing the Ph.D. degree with the Mobile Sensing and Geodata Science Laboratory, Department of Geography and Environmental Management, University of Waterloo, ON, Canada.

Her research interests include autonomous driving, mobile laser scanning, intelligent processing of point clouds, geometric and semantic modeling, and augmented reality.

Jonathan Li (Senior Member, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa, in 2000.

He is currently a Professor and the Head of the Mobile Sensing and Geodata Science Group, Department of Geography and Environmental Management, cross-appointed with the Department of Systems Design Engineering, University of Waterloo, Canada. He is also a Founding Member of the Waterloo Artificial Intelligence Institute. His research interests include AI-based information extraction from mobile LiDAR point clouds and Earth observation images. He has coauthored more than 400 publications, over 200 of which were published in refereed journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, ISPRS-JPRS, and RSE. He is the Chair of the ISPRS WG I/2 on LiDAR, Air- and Space-borne Optical Sensing from 2016 to 2020 and the ICA Commission on Sensor-driven Mapping from 2015 to 2023. He is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and *Canadian Journal of Remote Sensing*. He was a recipient of the Outstanding Achievement in Mobile Mapping Technology Award in 2019, for his pioneering contributions in developing and promoting mobile mapping technology, and the ISPRS Samuel Gamble Award in 2020, for his significant contributions to the development, organization or professional activities of the photogrammetry, remote sensing, and spatial information sciences at national or international level.

Yongtao Yu (Senior Member, IEEE) received the Ph.D. degree in computer science and technology from Xiamen University, Xiamen, China, in 2015.

He is currently an Assistant Professor at the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, China. He has authored or coauthored over 30 research articles published in refereed journals and proceedings, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and ISPRS-JPRS. His research interests include intelligent transportation systems, computer vision, deep learning, intelligent interpretation of 3-D point clouds, and remotely sensed imagery.

José Marcato Junior (Member, IEEE) received the Ph.D. degree in cartographic science from Sao Paulo State University, Brazil. He is currently a Professor with the Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil. His current research interests include UAV photogrammetry and deep neural networks for object detection, classification, and segmentation. He has published more than 30 articles in refereed journals and over 70 papers in conferences, including *ISPRS Journal of Photogrammetry and Remote Sensing* and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Wesley Nunes Gonçalves (Member, IEEE) received the Ph.D. degree in computational physics from the University of Sao Paulo, Brazil. He is currently a Professor with the Faculty of Computer Science and Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil. His current research interests include computer vision, machine learning, and deep neural networks for object detection, classification, and segmentation. He has published more than 30 articles in refereed journals and over 60 papers in conferences, including *Pattern Recognition*, *Pattern Recognition Letters*, and *Neurocomputing*.

Michael A. Chapman received the Ph.D. degree in photogrammetry from Laval University, Qubec City, QC, Canada.

He was a Professor with the Department of Geomatics Engineering, University of Calgary, Canada, for 18 years prior joining Ryerson University in 1999. He is currently a Professor at the Department of Civil Engineering, Ryerson University, Toronto, Canada. He has authored or coauthored over 200 technical articles, including those in top remote sensing journals such as *ISPRS Journal of Photogrammetry and Remote Sensing* and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include algorithms and processing methodologies for airborne sensors using GNSS/IMU, geometric processing of digital imagery in industrial environments, terrestrial imaging systems for transportation infrastructure mapping, mobile laser scanning, and algorithms and processing strategies for biometrology applications.