RBFNet: A Region-Aware and Boundary-Enhanced Fusion Network for Road Extraction From High-Resolution Remote Sensing Data

Weiming Li, *Member, IEEE*, Tian Lan, Shuaishuai Fan, *Member, IEEE*, and Yonghua Jiang, *Member, IEEE*

***Abstract*—Most existing road extraction methods prioritize region accuracy at the expense of ignoring road boundaries and connectivity quality. The occluding objects such as buildings, trees, and vehicles in remote sensing data usually cause discontinuous mask outputs, and consequently affect road extraction accuracy. In this article, a road extraction fusion network perceiving region and boundary features is proposed. The combination of a location-aware transformer and convolutional neural network is responsible for focusing regional semantic information through adaptive weight filtering. Combining spatial and channel information, the graph convolutional network is improved by constructing an integrated adjacency matrix to consider the relationships between nodes at different scales, which allows for better capture of multi-scale contexts. Boundary details are used to complement regional features, thereby enhancing the connectivity of masks. Comprehensive quantitative and qualitative experiments demonstrate that our method significantly outperforms state-of-the-art methods on two public benchmarks, which can improve road extraction by handling interruptions related to shadows and occlusions, producing high-resolution masks.**

***Index Terms*—Boundary enhancement, integrated GCN, parameter filtering, region-aware, road extraction.**

I. Introduction

ROAD coverage, structure, pavement conditions, and the surrounding environment have a significant impact on the convenience, smoothness, and safety of traffic-related activities. Accurate and comprehensive road mask information is essential for various applications such as road planning, traffic capacity analysis, functional zoning, map navigation, providing services that are crucial in these domains. Satellite remote sensing images provide a relatively inexpensive, objective, and repeatedly covered way for obtaining surface information over large areas, making them a primary data source for road network extraction. However, the factors such as shapes, widths, lane numbers, intersection types, as well as occlusions from trees, buildings, and cloud shadows may obstruct roads in images with a certain resolution, resulting in blurred and discontinuous boundaries, which pose challenges in accurately extracting road masks from remote sensing images.

Recently, significant progress has been made by convolutional neural network (CNN)-based variants [1], [2], [3] in the application of road extraction. Design concepts such as multiscale fusion [4], [5], [6], attention [7], [8], lightweight [9], [10], few-shot [11], and transfer learning [12] have improved their ability to handle the diversity and complexity of roads. Simultaneously, optimizing CNNs to enhance the generalization ability is another important trend. The NLinkNet proposed by Wang et al. [13] used nonlocal blocks that can capture the relationships between global features as a way to improve network generalization. Zhang et al. [14] introduced a ResUnet, which combines the strengths of U-Net and residual connection, simplifying the parameter count and enhancing accuracy. However, the disadvantage of these variants lies in limited local receptive field caused by the fixed-size convolutional kernels, which makes them more sensitive to the road occlusions caused by other objects or complex backgrounds. Consequently, the above framework restricts the effective capture of road structures at different scales and resolutions, and hinders the modeling of long-range dependencies across different regions.

Unlike CNNs, transformer [15], [16], [17] captures global relationships among sequences with its self-attention mechanism, enabling better understanding of the overall road topologies, which also exhibits a certain robustness to scale variations. Luo et al. [18] proposed a bidirectional transformer that enhances the extraction of global and local information, capturing contextual road information across features of various scales. Similarly, Tao et al. [19] proposed a Seg-Road model, which leverages the long-distance dependencies of transformers to improve road connectivity and alleviate road mask fragmentation. Another example is RoadFormer, proposed by Jiang et al. [16], which focuses on semantically relevant features in a sparse global manner, effectively enhancing the quality and robustness of road feature representation. Currently, the trend is to combine CNNs and transformers to further enhance the performance of such applications, capitalizing on CNNs' efficiency in spatial

Manuscript received 27 April 2024; revised 11 June 2024; accepted 21 July 2024. Date of publication 25 July 2024; date of current version 25 September 2024. This work was supported by the National Natural Science Foundation of China (NSFC), under Grant Grant U2031138. *(Corresponding author: Shuaishuai Fan.)*

Weiming Li, Tian Lan, and Shuaishuai Fan are with the School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264000, China (e-mail: liweiming_xd@sdtbu.edu.cn; lt399124289@163.com; fanshuaishuai@sdtbu.edu.cn).

Yonghua Jiang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China (e-mail: jiangyh@whu.edu.cn).

The code is available at https://github.com/ENDYC/RBFNet.

Digital Object Identifier 10.1109/JSTARS.2024.3433552

feature extraction and transformers' ability to handle global context dependencies. However, the aforementioned methods still struggle to effectively address road occlusions by objects such as buildings, trees, etc., as well as poor connectivity in road inspection due to different geomorphological backgrounds.

Graph neural networks (GNNs) [20], [21] leverage topological relationships between nodes to aggregate features of neighbors, and can effectively distinguish roads from other features in road extraction tasks. Yan et al. [22] introduced a regularized GNN to process preconstructed road maps derived from accessible road centerline data. Zhou et al. [23] developed a SGCN with an adjacency matrix constructed via the Sobel gradient operator, capturing global contexts in both channels and spatial features. In addition, Cui et al. [24] proposed a GDCNet combining GCN and CNN frameworks to generate complementary spatial-spectral features at the superpixel and pixel levels, respectively, thereby effectively handling incomplete and discontinuous roads. Although the aforementioned methods have demonstrated promising performance in road extraction applications, there are still some unresolved issues:

1) Insufficient capability to extract road details. Extracting details such as road boundaries, signs, intersections, and other fine-grained information is crucial at a determined spatial resolution, which often require focused processing to ensure accurate mask outputs.
2) Limited ability to handle mask connectivity. Roads in remote sensing images are often occluded by buildings, trees, and other objects, making it difficult to extract continuous features and resulting in fragmented road masks.
3) Room for improvement in road extraction accuracy. The complexity of the background terrain increases the difficulty of extracting road features within a region, while how to make comprehensive use of multiple features to improve the accuracy and robustness of mask extraction is the difficulty of method design.

The motivation in this work is to enhance extraction accuracy by incorporating boundary details alongside region features, and improve mask connectivity with extending the feature representation dimensions. To address all the above problems concurrently, we propose a fusion network for road extraction from high-resolution remote sensing data, named RBFNet, which simultaneously focuses on local regions and road boundaries. The proposed RBFNet enhances the perception capability of the region-aware branch for road regions by leveraging the boundary-enhanced branch, which provides additional image details and edge features. While the region-aware branch delivers low-dimensional features to the boundary-enhanced branch to supplement the context information, thus helping it to better analyze the environment around the road boundaries. The main contributions of this article are summarized as follows:

1) We propose a region-aware and boundary-enhanced fusion network for road extraction application from high-resolution remote sensing data. Stacking transformers with weight-adaptive filtering is employed in region-aware branch to extract global semantic information, and an integrated GCN variant is adopted in boundary-enhanced branch to collect edge relationships. Under the premise of ensuring that global information and collecting details, extraction performance and road connectivity are further improved.
2) A region awareness module (RAM) composed of a multiscale feature extraction module (MFEM) and a location-aware Transformer (LaFormer) based on positioning information is proposed to effectively extract road regional features through multiscale information capture and global semantic modeling, taking into account both the details and topologies. LaFormer is a parallel structure that utilizes dynamic parameter filtering and bias to simultaneously focus on global and local information.
3) A boundary enhancement module (BEM) is introduced to extract details from multiple directions and generate edge features, thereby enhancing the perception of road boundaries. An integrated spatial-channel perception GCN (ISCP-GCN) is proposed for boundary-enhanced branch to consider the relationship between nodes at different scales by constructing an integrated adjacency matrix, which captures multiscale contexts with the features of two-branch fusion as input, further enhancing the feature representation of road boundaries.

The rest sections of this article are organized as follows. Section II introduces our RBFNet and the design details of each module, while Section III reports extensive experimental analyses. Finally, Section IV summarizes the research results, shortcomings, and future work prospects.

II. Method

A. Architecture

To refine road boundary and enhance segmentation accuracy, and considering the strengths of CNNs, transformers, and GCNs, we propose a dual-branch fusion network for road extraction, namely RBFNet, as illustrated in Fig. 1. RBFNet takes high-resolution remote sensing images $\boldsymbol{X} \in \mathbb{R}^{3\times H\times W}$ with height H and width W as input, and outputs well-defined road masks $\boldsymbol{Y} \in \mathbb{R}^{3\times H\times W}$. The architecture consists of two branches, one dedicated to region awareness and the other focused on boundary details. The region-aware branch comprises multiple cascaded RAMs, which is composed of a MFEM with a downsampling operation and LaFormer in series to model global semantic information. Furthermore, the boundary-enhanced branch consists of BEM, MFEM, and ISCP-GCN, where BEM strengthens road boundaries and feeds it into MFEM to supplement the fine-grained details. ISCP-GCN constructs an integrated adjacency matrix by integrating spatial and channel information, expanding the information propagation between nodes to reduce overfitting, which allows the model to learn more general and robust feature representations. The fusion of these two branches enables RBFNet to collect road region representations more comprehensively while retaining more detailed and global semantic information, thereby enhancing the extraction accuracy and mask connectivity. The decoding part corresponds to the encoding part, and the last layer of the decoding part uses a $sigmoid$ activation function to obtain the prediction masks.

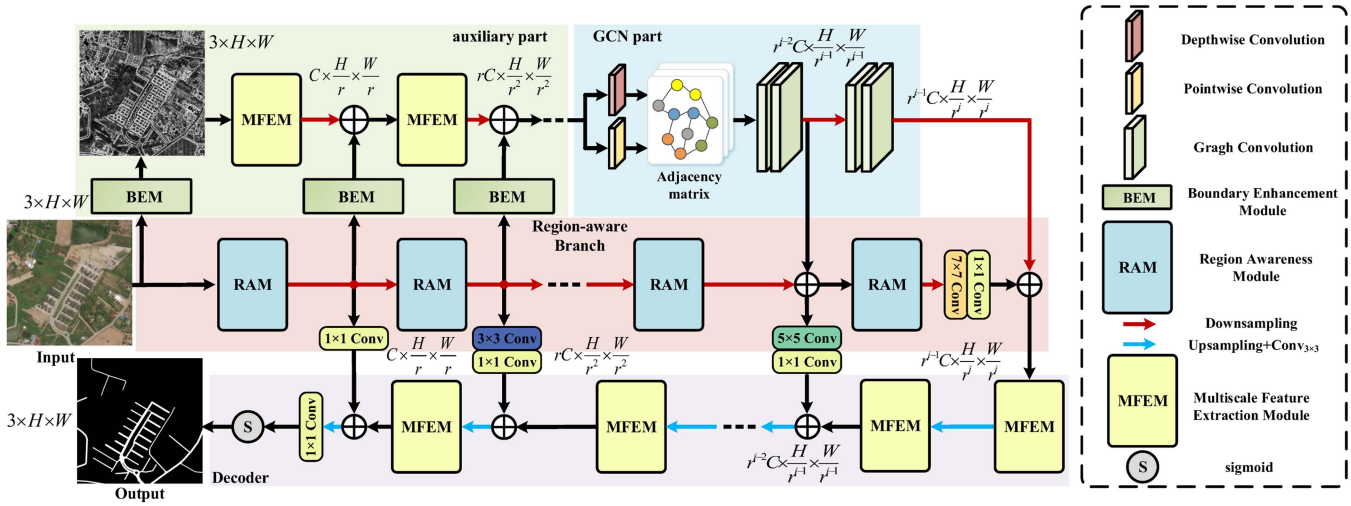

Fig. 1. Overall architecture of the proposed RBFNet.

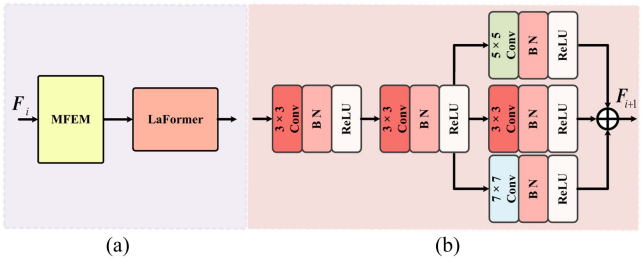

Fig. 2. Internal structure of RAM and MFEM. (a) RAM. (b) MFEM.

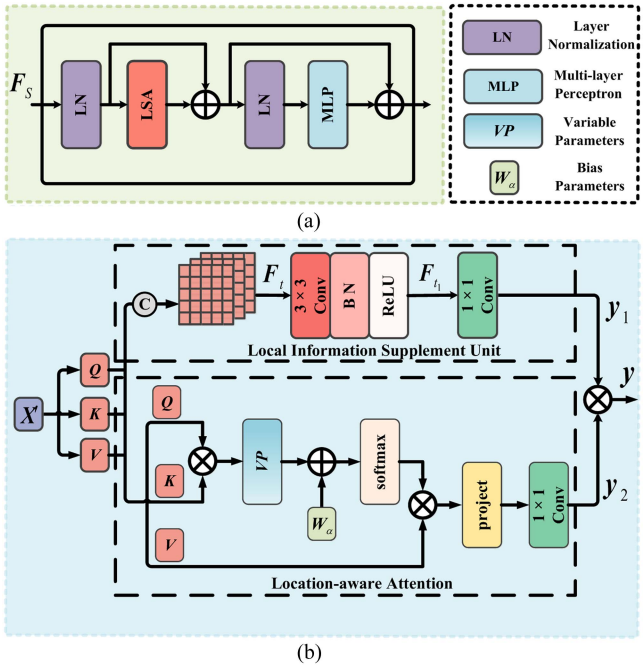

Fig. 3. Internal structure of LaFormer and LSA. (a) LaFormer. (b) LSA.

B. Region Awareness Module

The region-aware branch is designed to gather multiscale information and model the global semantics of roads, and the feature encoding part consists of multiple concatenated RAMs. Ulteriorly, RAM consists of MFEMs, downsampling operations, and LaFormers, and the structure is shown in Fig. 2. MFEM first applies two 3×3 convolutions for feature extraction, and then employs three parallel convolutions using 3×3, 5×5, and 7×7 filters, aiming to increase the diversity of semantic features.

Most traditional transformers and their variants treat the importance of all information equally when extracting global semantic information [17], [19], [25]. In contrast, LaFormer consists of layer normalization (LN), location-aware self-attention (LSA), and multilayer perceptron (MLP), as illustrated in Fig. 3(a), and it focuses more on the regions of interest through dynamic parameter filtering and bias, thereby better capturing local details while maintaining global information. Fig. 3(b) shows that LSA in LaFormer is composed of a local information supplement unit and a location-aware attention in parallel. The first part supplements local information by convolutions to address the issue of insufficient detail information processing in transformers. Meanwhile, the latter captures positional information using variable parameters ($\boldsymbol{VP}$), allowing LaFormer to prioritize regions with road distribution, and then utilizes bias parameters to reduce the influence of features resembling roads on the outputs.

LaFormer performs a linear transformation on the output $\boldsymbol{F}_S \in \mathbb{R}^{r^{i-1}C \times \frac{H}{r^i} \times \frac{W}{r^i}}$ with $i \in (1, 2, \ldots, i)$ of MFEM to obtain sequence $\boldsymbol{X}' \in \mathbb{R}^{c \times (h \times w)}$ $(c = r^{i-1}C, h = \frac{H}{r^i}, w = \frac{W}{r^i})$, and then the components of $Query$ ($\boldsymbol{Q} \in \mathbb{R}^{c \times (h \times w)}$), Key ($\boldsymbol{K} \in \mathbb{R}^{c \times (h \times w)}$), and $Value$ ($\boldsymbol{V} \in \mathbb{R}^{c \times (h \times w)}$) in LSA can be formalized as

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{X}' \cdot \boldsymbol{W}_Q \\ \boldsymbol{K} = \boldsymbol{X}' \cdot \boldsymbol{W}_K \\ \boldsymbol{V} = \boldsymbol{X}' \cdot \boldsymbol{W}_V \end{cases} \tag{1}$$

where $\boldsymbol{W}_Q \in \mathbb{R}^{c\times c}$, $\boldsymbol{W}_K \in \mathbb{R}^{c\times c}$, and $\boldsymbol{W}_V \in \mathbb{R}^{c\times c}$ represent learnable weights, respectively. c denotes the channel dimension, and h, w are the height and width of features.

To ensure that the features extracted by RBFNet contain more details, in the local information supplement unit, $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are concatenated along the channel dimension to obtain the feature $\boldsymbol{F}_t \in \mathbb{R}^{3c\times h\times w}$

$$\boldsymbol{F}_t = \text{concat}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \tag{2}$$

where concat$(*)$ is the channel concatenation operation.

Furthermore, local detail extraction and dimension adjustment are carried out on $\boldsymbol{F}_t$ after a 3×3 convolution operation $\text{Conv}_{3\times3}(*)$ and a 1×1 convolution $\text{Conv}_{1\times1}(*)$ successively. The output $\boldsymbol{y}_1 \in \mathbb{R}^{c\times h\times w}$ of this part can be calculated as

$$\boldsymbol{y}_1 = \text{Conv}_{1\times1}(\sigma(\text{Conv}_{3\times3}(\boldsymbol{F}_t))) \tag{3}$$

where $\sigma(*)$ represents the activation function.

The attention $\boldsymbol{An} \in \mathbb{R}^{c\times c}$ of the output from the location-aware attention can be represented as

$$\boldsymbol{An} = \text{softmax}\left(\boldsymbol{QK}^{\mathrm{T}}/\sqrt{d}\cdot \boldsymbol{VP} + \boldsymbol{W}_\alpha\right)\boldsymbol{V} \tag{4}$$

where $\boldsymbol{W}_\alpha \in \mathbb{R}^{c\times c}$ denotes the bias parameters for adjusting the weight distribution, and $\boldsymbol{VP} \in \mathbb{R}^{c\times c}$ represents the VPs used to effectively capture location information, which is adaptively generated based on the road positions in each patch, allocating greater weights to these key regions. Both of the above are obtained through network learning.

Assuming $\boldsymbol{W}_s \in \mathbb{R}^{c\times c}$ represents the weight parameter matrix for the learned road region features, we perform a filtering operation on it by using the adaptive threshold k_s to obtain w_{t_i}, which can be expressed as

$$w_{t_i} = \begin{cases} 0, & \text{if } w_s < k_s \\ w_s, & \text{otherwise} \end{cases} \tag{5}$$

where $w_s \subset \boldsymbol{W}_s$, $w_{t_i} \subset \boldsymbol{W}_{t_i}$, and $\boldsymbol{W}_{t_i} \in \mathbb{R}^{c\times c}$ is the filtered parameter matrix.

Afterward, $\boldsymbol{VP}$ undergoes self-updating using $\boldsymbol{W}_{t_i}$ to achieve adaptive weight allocation, thereby allowing the RBFNet to pay more attention to road regions and enhancing it sensitivity to road features

$$\boldsymbol{VP} = \boldsymbol{W}_{t_i} \odot \boldsymbol{VP} \tag{6}$$

where symbol $\odot$ denotes the dot product operation.

However, we have observed that some patches may contain features that are similar to road features but not relevant. To prevent the proposed RBFNet from overly emphasizing these irrelevant features, the matrix $\boldsymbol{W}_\alpha$ is employed to fine-tune the weight distribution. Similar to the generation of $\boldsymbol{VP}$, $\boldsymbol{W}_\alpha$ is learned based on the features extracted by MFEM in the boundary-enhanced branch, specifically targeting the road distribution in each patch.

Similarly, we define the weight parameter matrix from the learned road boundary features as $\boldsymbol{W}_b \in \mathbb{R}^{c\times c}$, and each element w_b is filtered based on the adaptively learned threshold k_b to obtain the filtered matrix $\boldsymbol{W}_{t_j} \in \mathbb{R}^{c\times c}$. That is

$$w_{t_j} = \begin{cases} w_b, & \text{if } w_b < k_b \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $w_{t_j} \subset \boldsymbol{W}_{t_j}$.

Then, $\boldsymbol{W}_\alpha$ is obtained via multiplying $\boldsymbol{W}_{t_j}$ by a small scalar ε, allowing for slight adjustments to the attention regions, which flexibility is crucial for adapting to different road scenarios and environmental changes, ensuring that our model performs well under various conditions.

After mapping $\boldsymbol{An}$, we pass it through a 1×1 convolution to obtain the final output $\boldsymbol{y}_2 \in \mathbb{R}^{c\times h\times w}$ of the location-aware attention

$$\boldsymbol{y}_2 = \text{Conv}_{1\times1}(\text{project}(\boldsymbol{An})) \tag{8}$$

where project$(*)$ represents the mapping operation.

The feature $\boldsymbol{y} \in \mathbb{R}^{c\times h\times w}$ of LSA is obtained by fusing the two above outputs, and the pseudocode procedure of LaFormer is presented in Algorithm 1

$$\boldsymbol{y} = \boldsymbol{y}_1 \times \boldsymbol{y}_2 \tag{9}$$

Algorithm 1: Feature Extraction Process of LSA in LaFormer.

Input: The features $\boldsymbol{F}_S \in \mathbb{R}^{c\times h\times w}$, adaptive parameters threshold k_s, k_b, boundary features $\boldsymbol{F}_B \in \mathbb{R}^{c\times h\times w}$.
Output: results $\boldsymbol{y} \in \mathbb{R}^{c\times h\times w}$ are the output of LSA
$\boldsymbol{F}_S$ is converted to $\boldsymbol{X}'$ by linear layers
$\boldsymbol{X}'$ is converted to $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ through the matrices $\boldsymbol{W}_Q, \boldsymbol{W}_K$, and $\boldsymbol{W}_V$
Concat $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ together in dimensions to get $\boldsymbol{F}_t$
$\boldsymbol{y}_1 = \text{Conv}_{1\times1}(\sigma(\text{Conv}_{3\times3}(\boldsymbol{F}_t)))$
for each feather patch **do**
    $\boldsymbol{W}_s = S(\boldsymbol{F}_S), \quad \boldsymbol{W}_{t_i} \overset{k_s}{\leftarrow} \boldsymbol{W}_s$
    Multiply $\boldsymbol{W}_{t_i}$ by the adaptive matrix $\boldsymbol{VP}$
    $\boldsymbol{W}_b = S(\boldsymbol{F}_B), \quad \boldsymbol{W}_{t_j} \overset{k_b}{\leftarrow} \boldsymbol{W}_b$
    Multiply $\boldsymbol{W}_{t_j}$ by ε as $\boldsymbol{W}_\alpha$
    $\boldsymbol{An} = \text{softmax}(\boldsymbol{QK}^{\mathrm{T}}/\sqrt{d}\cdot \boldsymbol{VP} + \boldsymbol{W}_\alpha)\boldsymbol{V}$
end for
$\boldsymbol{y}_2 = \text{Conv}_{1\times1}(\text{project}(\boldsymbol{An}))$
$\boldsymbol{y} = \boldsymbol{y}_1 \times \boldsymbol{y}_2$

C. Boundary Enhancement Module

The boundary-enhanced branch is designed to strengthen the road boundary and acquire global contextual information, so as to supplement the lost details in the region feature extraction process, and improve the segmentation accuracy and road connectivity. The BEM is utilized to generate edge features and perceive road boundaries, thereby reducing confusion between roads and the surrounding backgrounds, which helps to improve the accuracy of road boundary recognition, and its structure is shown in Fig. 4.

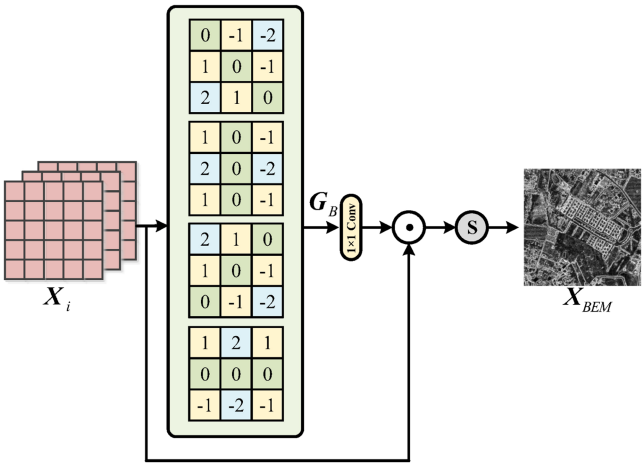

Fig. 4. Structure of BEM.

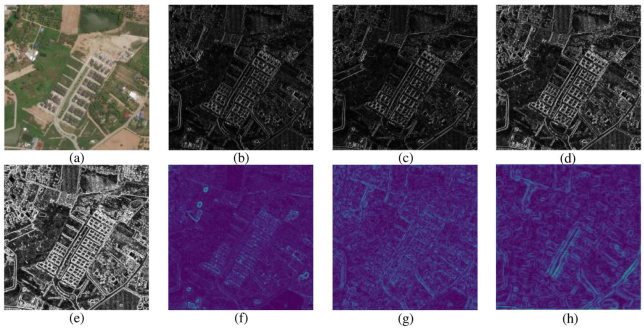

Fig. 5. Boundary features generated by operators at different angles and the outputs from different stages of BEM. (a) is the original image, (b) is the result of horizontal filtering, (c) is the result of vertical filtering, (d) is the fused output in the horizontal and vertical directions, (e) is the output with our method, (f)–(h) are the outputs of BEM at each stage.

Most traditional edge detection methods, such as Roberts [26], Wallis [27], morphological operators [28], and wavelet [29], primarily filter images by convolutional templates to extract edge information. However, these algorithms have limitations in considering the directional features of object boundaries, leading to some loss of details. The Sobel operator [30] extracts edge information in images using first-order gradients in the horizontal and vertical directions. However, for the application of road network extraction, the edge details extracted from these two directions alone are insufficient, as seen in the results in Fig. 5(b)–(d). To address this, we draw inspiration from the Sobel operator and design BEM for edge detection, which is designed with a span of 45° and incorporates four directional operators, then fused to capture more detailed features from different orientations. The specific kernel parameters for the filtering operators are given as

$$\boldsymbol{K}_{0^\circ} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \boldsymbol{K}_{45^\circ} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix}$$

$$\boldsymbol{K}_{90^\circ} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \boldsymbol{K}_{135^\circ} = \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} \tag{10}$$

where $\boldsymbol{K}_{0^\circ}$, $\boldsymbol{K}_{45^\circ}$, $\boldsymbol{K}_{90^\circ}$, and $\boldsymbol{K}_{135^\circ}$ represent edge filting operators with angles of 0°, 45°, 90°, and 135°, respectively.

By applying these operators at different angles to the input image $\boldsymbol{X}_i$, the gradient $\boldsymbol{G}_B$ can be expressed as

$$\boldsymbol{G}_B = \sum_j \boldsymbol{X}_i \odot \boldsymbol{K}_j, j = 0^\circ, 45^\circ, 90^\circ, 135^\circ. \tag{11}$$

After smoothing features and dimensional adjustments with a 1×1 convolution, $\boldsymbol{G}_B \in \mathbb{R}^{1\times h\times w}$ is fused with $\boldsymbol{X}_i \in \mathbb{R}^{c\times h\times w}$, and finally normalized with the function sigmoid$(*)$ to obtain edge enhancement feature $\boldsymbol{X}_{BEM} \in \mathbb{R}^{c\times h\times w}$, that is

$$\boldsymbol{X}_{\text{BEM}} = \text{sigmoid}(\boldsymbol{X}_i \odot \text{Conv}_{1\times 1}(\boldsymbol{G}_B)). \tag{12}$$

To verify the effectiveness of BEM in segmenting edges of low-dimensional features, we visualize the outputs of BEM at each stage, as exhibited in Fig. 5(f)–(h). Observation of the visualization results reveals that BEM can effectively segment low-dimensional road features from the background. Comparing the third stage to the second stage, the road features become more prominent, indicating that as BEM progressively extracts and enhances features, it can effectively focus on the road regions, further improving the accuracy of RBFNet in detecting road boundaries.

D. Integrated Spatial-Channel Perception GCN

GCN is adopted to process high-dimensional features in the boundary-enhanced branch, better capturing the global relationships among pixels, thus improving the understanding of the semantic structure of the entire image. However, the traditional GCNs only consider the information propagation between nodes and their neighbors, which cannot effectively process multiscale information [31], [32]. This limited local perception capability leads to decreased extraction performance, especially for large-scale road remote sensing images [33]. To address the aforementioned issues, ISCP-GCN is constructed by integrating spatial and channel dimensions to enhance feature representation, thus improving the model's awareness of global contexts in images and reducing overfitting. Furthermore, edge detail features are fused into the constructed adjacency matrix to enhance the representation of road boundaries. The construction of the channel feature matrix allows the network to learn the correlations and interactions between channels, which helps strengthen the focus on specific channels in road extraction tasks and improves the network's perception of key information. The construction of the spatial feature matrices enable the model to learn the relative positions of spatial pixels, thereby better understanding the semantic structure and contextual relationships of road regions.

The spatial and channel features $\boldsymbol{X}_S \in \mathbb{R}^{N'\times C'}$, $\boldsymbol{X}_C \in \mathbb{R}^{N'\times C'}$ are obtained by depth separable convolutions and flattening operations for the high-dimensional feature $\boldsymbol{X}^* \in \mathbb{R}^{C'\times H'\times W'}$ generated by MFEM in boundary-enhanced branch, where N' denotes the number of nodes, with its value

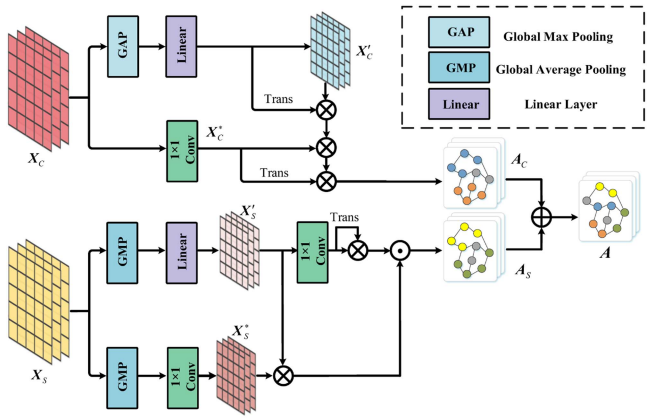

Fig. 6. Construction process of the integrated adjacency matrix.

being the product of H' and W' $\left(C' = r^{i-3}C,\ H' = \frac{H}{r^{i-2}},\ \text{and}\ W' = \frac{W}{r^{i-2}}\right)$.

$$\boldsymbol{X}_S = \text{Flatten}(\text{Conv}_{Dw}(\boldsymbol{X}^*)) \tag{13}$$

$$\boldsymbol{X}_C = \text{Flatten}(\text{Conv}_{Pw}(\boldsymbol{X}^*)) \tag{14}$$

where $\text{Conv}_{Dw}(*)$ and $\text{Conv}_{Pw}(*)$ represent the depthwise and pointwise convolution operations, respectively. $\text{Flatten}(*)$ denotes the flattening operation.

To construct the integrated adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N' \times N'}$ for ISCP-GCN to better aggregate node features, $\boldsymbol{X}_S$ and $\boldsymbol{X}_C$ are treated as undirected graph $\boldsymbol{\mathcal{G}} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$, $\boldsymbol{\mathcal{V}}$ is the set of nodes, and $\boldsymbol{\mathcal{E}}$ is the set of edges. The construction process is shown in Fig. 6, the spatial matrices $\boldsymbol{X}^*_S \in \mathbb{R}^{N' \times C'}$, $\boldsymbol{X}'_S \in \mathbb{R}^{C' \times N'}$, and the channel matrix $\boldsymbol{X}'_C \in \mathbb{R}^{C' \times C'}$ are obtained through the linear processing $\text{Linear}(*)$, global average pooling (GAP), and global max pooling (GMP) operations, which can be formalized as

$$\boldsymbol{X}^*_S = \text{Conv}_{1\times1}(\text{Pool}_{GM}(\boldsymbol{X}_S)) \tag{15}$$

$$\boldsymbol{X}'_S = \text{Linear}(\text{Pool}_{GM}(\boldsymbol{X}_S)) \tag{16}$$

$$\boldsymbol{X}'_C = \text{Linear}(\text{Pool}_{GA}(\boldsymbol{X}_C)) \tag{17}$$

where $\text{Pool}_{GM}(*)$ and $\text{Pool}_{GA}(*)$ represent global max and average pooling operations, severally.

The calculation of similarity and weights between different channels helps the model to better understand the relationships and dependencies between multiple channels, and improve the model's perception of channel features. Therefore, the channel feature matrix $\boldsymbol{A}_C \in \mathbb{R}^{N' \times N'}$ can be further calculated as

$$\boldsymbol{A}_C = \boldsymbol{X}^*_C \cdot \text{Trans}P(\boldsymbol{X}'_C) \cdot \boldsymbol{X}^{*\text{T}}_C \tag{18}$$

where $\boldsymbol{X}^*_C \in \mathbb{R}^{N' \times C'} = \text{Conv}_{1\times1}(\boldsymbol{X}_C)$, $\boldsymbol{X}^{*\text{T}}_C \in \mathbb{R}^{C' \times N'}$ is the transpose matrix of $\boldsymbol{X}^*_C$, and $\text{Trans}P(*)$ represents the transposition product operation.

Similarly, the spatial feature matrix $\boldsymbol{A}_S \in \mathbb{R}^{N' \times N'}$ is obtained by calculating the similarity and weights of different spatial positions to enhance the model's understanding and perception of the dependencies between various regions in the images, that is

$$\boldsymbol{A}_S = \text{Conv}_{1\times1}(\boldsymbol{X}'_S)^{\text{T}} \cdot \text{Conv}_{1\times1}(\boldsymbol{X}'_S) \odot \boldsymbol{X}'_S \cdot \boldsymbol{X}^*_S. \tag{19}$$

The final integrated adjacency matrix $\widetilde{\boldsymbol{A}} \in \mathbb{R}^{N' \times N'}$ can be represented as

$$\widetilde{\boldsymbol{A}} = \boldsymbol{A}_S + \boldsymbol{A}_C + \boldsymbol{I}. \tag{20}$$

Then, the graph convolution operations are performed by utilizing $\widetilde{\boldsymbol{A}}$, which is formulized as

$$\boldsymbol{H}^{l+1} = \sigma\left(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^l\boldsymbol{W}^l\right) \tag{21}$$

where $\widetilde{\boldsymbol{D}} \in \mathbb{R}^{N' \times N'}$ is a diagonal matrix and its elements $\widetilde{d}_{ii}$ on the diagonal represent the degrees of the corresponding nodes, such that $\widetilde{d}_{ii} = \sum_j \widetilde{a}_{ij}, \widetilde{a}_{ij} \subset \widetilde{\boldsymbol{A}}$. $\boldsymbol{H}^l \in \mathbb{R}^{N' \times C'}$ is the output of the previous layer l, and $\boldsymbol{W}^l \in \mathbb{R}^{C' \times C'}$ is the trainable weight parameters for the lth layer.

E. Loss Function

RBFNet employs Dice as the loss function, which calculates the similarity between the predicted results and the actual labels. Specifically, the Dice coefficient computes the ratio of the overlap between two sets to their total size. The relationship between Dice loss $\mathcal{L}_{\text{Dice}}$ and Dice coefficient Cf_{Dice} is as follows:

$$\mathcal{L}_{\text{Dice}} = 1 - Cf_{\text{Dice}} \tag{22}$$

$$Cf_{\text{Dice}} = \frac{2 \times TP}{TP + FN + TP + FP} \tag{23}$$

where TP represents the number of correctly classified pixels, FP is the number of misclassified positive sample pixels, TN and FN are the number of correctly and misclassified classified background pixels, respectively.

III. Experimental Analysis

To demonstrate the effectiveness of our RBFNet, performance comparison tests on two publicly available datasets are conducted with various state-of-the-art methods, and some ablation experiments are further performed to compare the contributions and applicability of each proposed functional module to the overall performance. Mean precision (*mP*), mean recall (*mR*), mean intersection over union (*mIoU*), road intersection over union (*Road IoU*), and mean F1-score (*mF1*) are used as evaluation metrics. The specific computation formulas can be found in [34], and the recorded values of each metric in the experiments are the average values obtained after multiple trials.

The experimental evaluations in this article are performed on the following two publicly available datasets, namely, DeepGlobe Road Extraction (DGRE) [35] and Massachusetts Roads (MR) datasets [36], which contain samples of different land cover categories, including urban, rural, and tropical rainforest areas, etc. All the samples are randomly split into training, validation, and testing sets, and the configuration parameters are listed in Table I.

All methods are implemented using PyTorch 1.10.1 and tested on an NVIDIA RTX3060ti GPU with a consistent software

TABLE I
PARAMETERS OF TWO DATASETS USED IN THIS ARTICLE

Dataset	Total	Size	Training	Validation	Testing
MR	6226	512× 512	4980	623	623
DGRE	1171	512× 512	1108	14	49

environment. During the training process, $LeakyReLU$ is used as the activation function, and $Adam$ algorithm is optimizer, COS is employed for adjusting the learning rate. It is worth emphasizing that the runtime environments of all methods are consistent, and the internal parameters of the comparative methods are default and consistent with their original papers. The number of stages in the encoding and decoding parts of the region-aware branch of RBFNet is set to 4. In addition, we fine tune the parameters through multiple experiments based on existing theories and experiences, setting the learning rate from 1e−6 to 1e−4, training batch size to 8, ε to -0.01, iterating 200 epochs, and most networks converged within 125 epochs of training.

A. Comparison Experiments

Fifteen state-of-the-art methods recently applied to road extraction tasks, including NLinkNet [13], DLinkNet [37], Resunet [14], DeepLabV3+ [38], CoANet [39], PSPNet [40], Seg-Road [19], DSMSA-Net [41], RoadFormer [16], Segformer-b4 [42], SwinUnet [43], DGCN [44], TGDAUNet [45], GD-CNet [24], and SGCN [23], are compared with the proposed RBFNet on the aforementioned datasets to validate their effectiveness. The experimental results on the testing set are detailed in Table II, the best results are highlighted in bold, while the suboptimal values are underlined. The corresponding visual results are shown in Figs. 7, 8, 9, and 10.

RBFNet outperforms the second-best method (i.e., SGCN) by 0.86% and 1.86% in terms of *mF1* and *mIoU* on the MR dataset, respectively. Similarly, our main evaluation metrics, *mF1* and *mIoU*, are respectively 0.77% and 1.04% higher than the suboptimal SGCN on the DGRE dataset, which further demonstrates that RBFNet exhibits robustness and generalization capabilities across different datasets. The above performance advantages can be attributed to the ability of region-aware branch to collect multiscale information and conduct global modeling, and the ability of boundary-enhanced branch to enhance edge information and obtain global context information, thereby improving segmentation accuracy and road connectivity.

Fig. 7 depicts the performance of the above methods outlined in Table II on the MR dataset, and four representative sample instances are selected for comparison and explanation. PSPNet performs poorly on this dataset, identifying only the rough outline of fewer roads. CoANet and DSMSA-Net exhibit more missed and incorrect detections, as observed from Fig. 7(b), owing to their failure to adequately address features similar to roads. In addition, DeepLabV3+, Segformer, and SwinUnet face challenges in maintaining good connectivity and exhibit a higher number of missing regions. Although NLinkNet, Resunet, and Seg-Road perform better overall, they are less effective at extracting complex roads and small objects, as highlighted boxes in Fig. 7. In contrast, our proposed RBFNet demonstrates excellent performance on this dataset, which is attributed to the important consideration of global semantic information for road segmentation within our model. Furthermore, as shown in Fig. 7(a), the results of the four GCN-based methods compared with our RBFNet indicate that RBFNet demonstrates superior road connectivity with little difference in evaluation metrics, because ISP-GCN enables the model to better understand the topology of the road network.

The magnified view exhibited in Fig. 8 also confirms the above conclusions, it is evident that our RBFNet maintains better road connectivity when dealing with road features. Among other comparison networks, NLinkNet exhibits slightly better extraction performance than the transformer-based approaches, indicating that it leverages the collection and fusion of local information at different scales to enhance overall performance as a variant architecture of CNNs. However, as seen in the instance of Fig. 8(c), NLinkNet has some limitations in extracting road features in complex environments, due to its insufficient ability to perceive global information and contextual topology. Seg-Road performs best among transformer-based approaches, as it effectively understands long-range road features and ensures road connectivity to a certain extent, thereby addressing the issue of fragmented roads. However, it faces difficulties in extracting fine-grained road features, as evident from Fig. 8(d). SGCN stands out as the best-performing model in GCN-based networks, utilizing different features to collect global context information and capturing road feature relationships using graph structures, thereby improving road extraction accuracy. However, as depicted in Fig. 8(b), this model evidently ignores the treatment of similar features.

Sample instances are selected from rural, urban, city, and high vegetation coverage images to further validate the algorithm's performance. The visual comparison results of all the mentioned comparison methods on the DGRE dataset are presented in Fig. 9, while Resunet, PSPNet, and SwinUnet exhibit the poorest extraction performance among them. Fig. 9(b) shows that most models experience missed and incorrect regions. Although our model also had some false positives, it demonstrates better overall performance compared to others. The case shown in Fig. 9(c) shows that the masks extracted by RBFNet, SGCN, and TGDAUNet are relatively complete, especially in terms of fine-grained road features. However, there are some errors detected by TGDAUNet, while RBFNet achieves a higher level of completeness in feature extraction than SGCN. The same trend in Fig. 9(a) further visually supports the aforementioned observations. The underwhelming performance of Resunet on the DGRE dataset indicates that the model is affected differently by different data samples. However, our model performs optimally on both datasets, which indicates its decent generalization ability to some extent.

Similarly, Fig. 10 is a highlighted magnification of Fig. 9, from which it clearly shows that our RBFNet surpasses others in terms of segmentation accuracy. This is attributed to the

TABLE II
PERFORMANCE METRICS OF VARIOUS METHODS ON THE MR AND DGRE DATASETS

Type	Method	MR					DGRE				
		Road IoU(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑	*Road IoU*(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑
CNNs	NLinkNet	53.88	75.39	88.30	80.88	84.43	50.57	73.72	86.53	79.62	82.93
	DLinkNet	46.28	71.79	89.73	75.63	82.08	47.30	72.16	83.17	79.70	81.40
	Resunet	54.72	75.02	87.99	81.25	84.49	45.77	71.02	89.57	74.62	81.41
	DeepLabV3+	45.50	70.42	80.13	79.62	79.87	51.84	74.24	84.60	81.83	83.20
	CoANet	44.85	69.83	78.02	80.88	79.42	50.36	73.75	85.12	80.66	82.83
	PSPNet	08.36	51.54	62.86	54.67	58.48	20.49	61.06	61.37	64.32	62.81
Transformers	Seg-Road	51.64	74.27	86.57	80.49	83.42	60.87	78.73	87.87	85.77	86.81
	DSMSA-Net	41.25	68.45	80.41	75.89	78.08	48.19	72.89	86.17	78.67	82.25
	RoadFormer	48.17	72.50	88.71	75.63	81.65	56.51	76.97	89.01	82.45	85.60
	Segformer-b4	46.32	70.89	78.92	81.94	80.40	55.29	76.51	86.70	83.53	85.09
	SwinUnet	41.76	68.36	79.02	76.86	77.93	48.73	72.65	82.68	80.93	81.80
GCNs	DGCN	45.97	70.72	86.77	75.55	80.77	53.70	75.40	86.73	81.89	84.24
	TGDAUNet	55.71	75.92	86.47	82.95	84.67	60.81	79.25	89.54	85.12	87.27
	GDCNet	50.23	73.60	**92.08**	76.82	83.76	58.63	77.88	88.21	84.27	86.19
	SGCN	58.49	77.68	89.73	83.67	86.59	63.17	80.66	**91.03**	85.84	88.36
	Ours	**61.96**	**79.54**	87.77	**87.13**	**87.45**	**65.48**	**81.70**	90.85	**87.48**	**89.13**

TABLE III
ABLATION RESULTS ON THE MR DATASET UNDER DIFFERENT MODULE COMBINATIONS

Case	BEM	LaFormer	auxiliary part	ISCP-GCN	*Road IoU*(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑
(I)	**w/o**	w/	w/	w/	59.84	78.25	85.59	86.44	86.01
(II)	w/	w/	**w/o**	w/	58.39	77.63	86.57	83.74	85.13
(III)	w/	**w/o**	w/	w/	57.04	75.96	83.11	84.32	83.71
(IV)	w/	w/	w/	**w/o**	58.77	76.92	84.26	84.55	84.40
(V)	w/	w/	w/	w/	**61.96**	**79.54**	**87.77**	**87.13**	**87.45**

w/ and w/o indicate that the module is adopted or not in the test.

TABLE IV
ABLATION RESULTS ON THE DGRE DATASET UNDER DIFFERENT MODULE COMBINATIONS

Case	BEM	LaFormer	auxiliary part	ISCP-GCN	*Road IoU*(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑
(I)	**w/o**	w/	w/	w/	63.82	80.79	90.68	86.15	88.44
(II)	w/	w/	**w/o**	w/	61.47	79.65	89.62	84.14	86.79
(III)	w/	**w/o**	w/	w/	58.14	77.88	88.21	84.24	86.18
(IV)	w/	w/	w/	**w/o**	59.91	79.18	89.19	83.43	86.21
(V)	w/	w/	w/	w/	**65.48**	**81.70**	**90.85**	**87.48**	**89.13**

w/ and w/o indicate that the module is adopted or not in the test.

utilization of CNN to extract multiscale features, combined with transformer's detail information extraction module and location information attention mechanism, so as to more effectively process local details while maintaining global information integrity. The incorporation of GCN further enhances the model's ability to perceive road features, endowing it with powerful modeling capabilities. In Fig. 10(b), the segmentation effect of most models is not ideal due to occlusions caused by buildings and shadows. However, our model still exhibits good segmentation performance, indicating its ability to capture road topology more effectively. In addition, Fig. 10(c) illustrates that our RBFNet collects more detailed information compared to other models in situations such as curves and intersections, thus improving the segmentation accuracy. The instance shown in Fig. 10(d) also demonstrates the superiority of our RBFNet in complex environments.

B. Ablation Experiments

1) Single Module Ablation Experiment: The ablation experiments are organized on the two datasets to test the contribution of multiple functional modules, including BEM, LaFormer, the auxiliary part (the pipeline of this component is depicted with the green range in Fig. 1), and ISP-GCN. Tables III and IV and Fig. 11 record the test results for different module combinations. In case (I), the removal of BEM reduces the ability of the model to extract the boundaries of the sample instances for both datasets, as can be seen from the gap in the table relative to case (V). The main evaluation metrics, *mF1* and *mIoU*, decreased by 1.44%, 1.29%, and 0.69%, 0.91%, respectively, compared to case (V). Case (II) displays a decrease in *mF1* and *mIoU* by 2.32%, 1.91%, and 2.34%, 2.05%, respectively, compared to case (V), due to the excisions of the auxiliary part, leading to a weakening of detail extraction.

In the absence of LaFormer, case (III) is significantly detrimental to make model to perform global modeling and handle long-range dependencies, and has a decrease in *mF1* and *mIoU* by 3.74%, 3.58%, and 2.95%, 3.82%, respectively, compared to case (V). Case (IV), after ISCP-GCN is removed, leads to the model's inability to collect sufficient global contextual information, resulting in a reduction of *mF1* and *mIoU* by 3.05%, 2.62%, and 2.92%, 2.52%, respectively, compared to case (V).

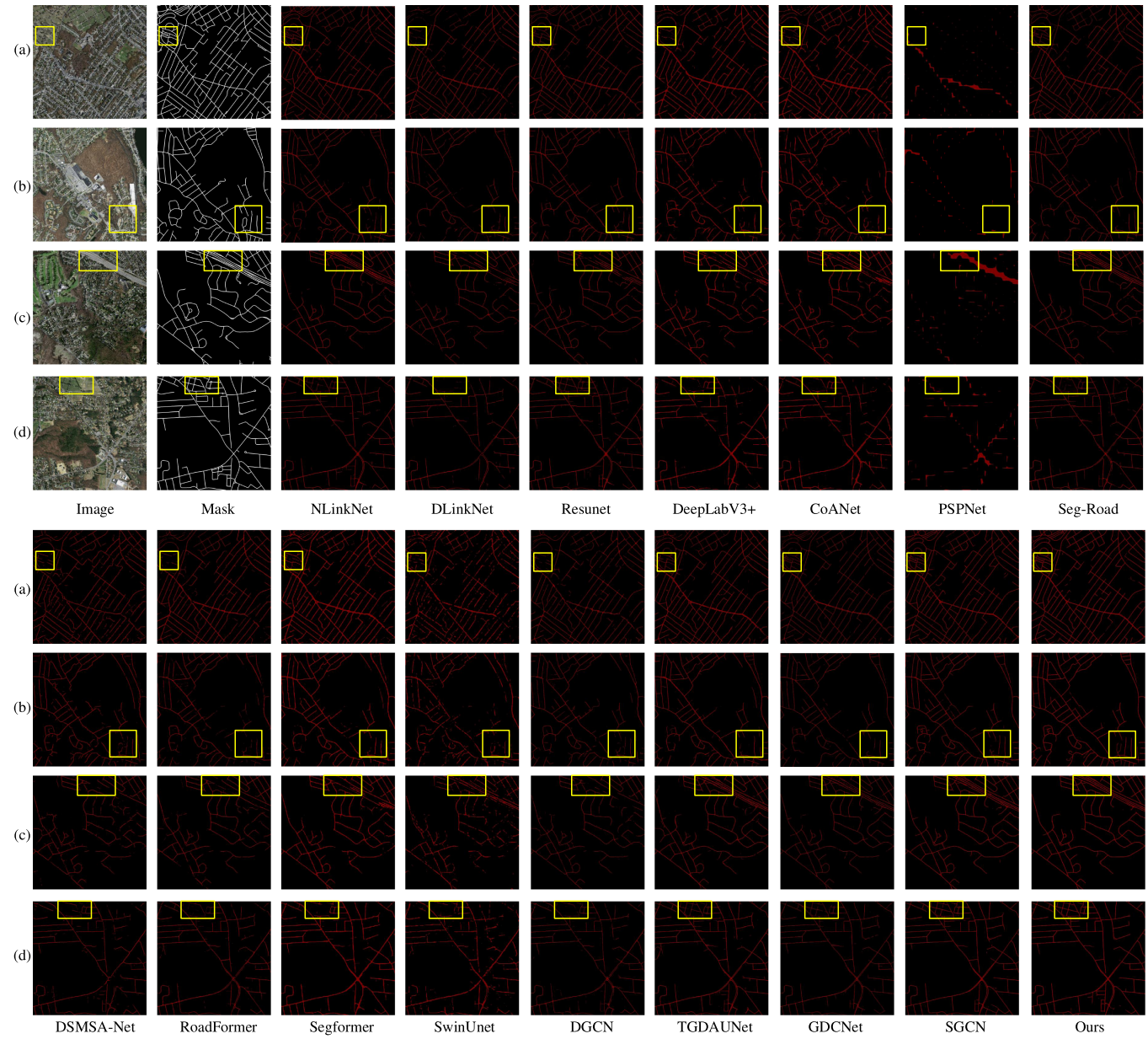

Fig. 7. Visual comparison results of various methods on the MR dataset.

Furthermore, Fig. 11 visually reflects the impact of each module on our model. Obviously, removing LaFormer would have the most significant impact on the RBFNet. By modeling global semantic information, LaFormer helps models better understand the overall structure and semantic information of images. As a result, when LaFormer is removed, the model's ability to understand global information decreases, resulting in a significant drop in performance. In contrast, removing the auxiliary part or BEM has relatively little effect on the model because their effects can be partially compensated by other parts. This does not mean that their role is not important, but indicates that the model has some robustness and can adapt to these changes to a certain extent. The above results show that each module in the proposed RBFNet contributes positively to the improvement of performance.

Figs. 12 and 13 display the qualitative test results of the sample instances corresponding to Tables III and IV, respectively. The completeness is highlighted with yellow, while the correctness is highlighted with pink. The BEM is beneficial to extract edge information from the images, thereby enhancing the recognition of fine-grained features. However, when the BEM is removed, RBFNet struggles to accurately judge the features of small objects during the feature extraction process, consequently affecting the extraction accuracy. Fig. 12 and Table III validate the aforementioned observations. The design motivation behind LaFormer is to enhance the network's focus on regions of interest by introducing dynamic and bias parameters, thus increasing its sensitivity to road regions. Therefore, removing LaFormer would decrease the network's attention to regions of interest, resulting in a significant decline in extraction effect. From Fig. 13,

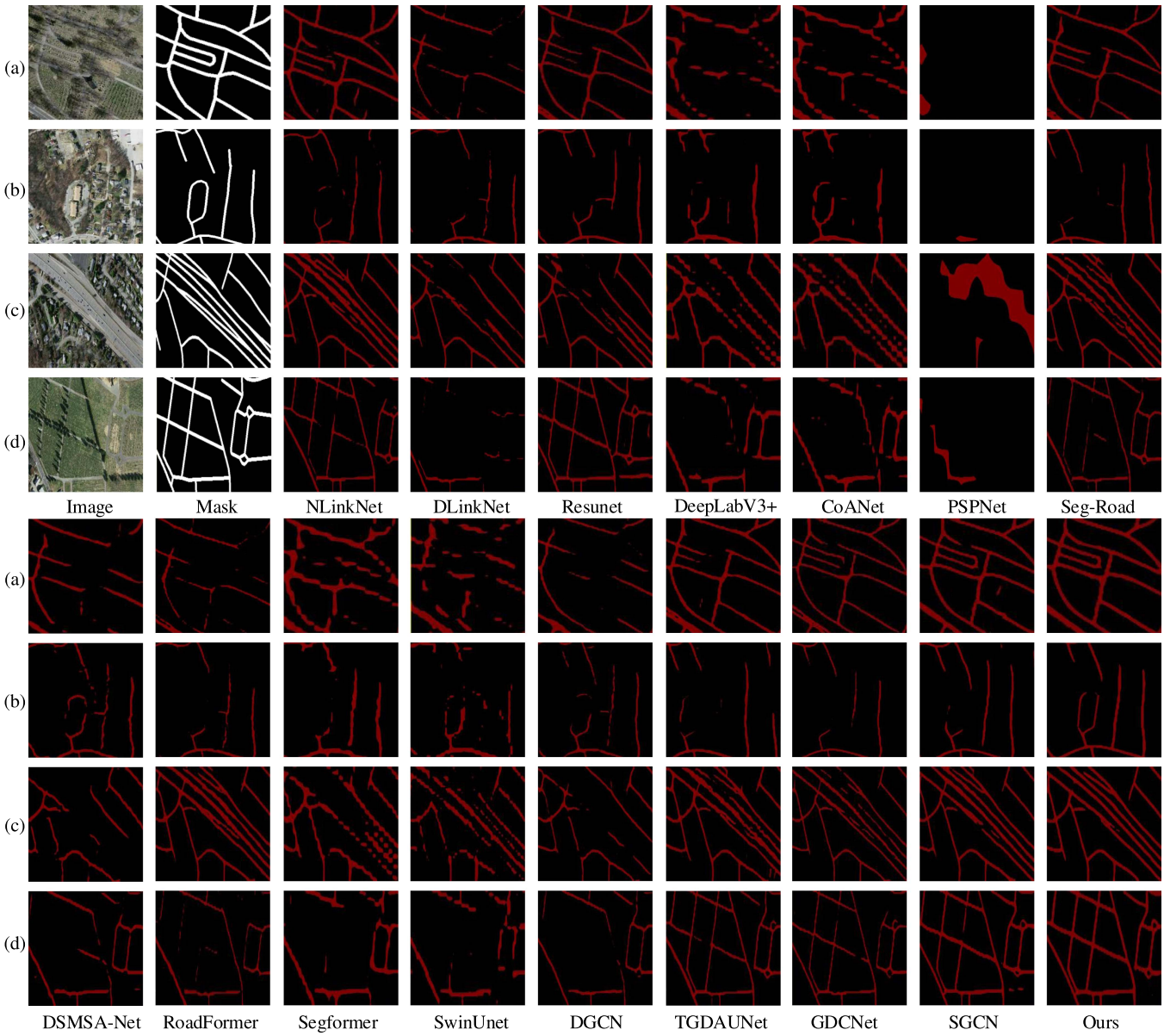

Fig. 8. Comparison of magnified highlighted regions on the MR dataset.

we can observe that extensive missing occurs once LaFormer is removed. Further examination of Fig. 12 reveals the presence of some false detections in the visualization results, which is primarily because the absence of LaFormer leads to a decrease in the model's ability to resist interference from similar features. In the boundary-enhanced branch, the auxiliary part processes low-dimensional edge features to compensate for the lack of detailed information, which plays an active role in improving extraction. Removing it weakens the network's ability to handle fine-grained details, thereby affecting the accurate capture of road edges and details. The difficulty of detecting small roads in Fig. 12 confirms the functionality of the auxiliary part. In addition, the removal of the auxiliary part prevents the transmission of edge information to LaFormer, which results in the inability to exclude the interference of similar features, resulting in some localized error detection cases.

The ISCP-GCN is significant for enhancing the capture of multiscale information and considering the relationships between nodes. Removing ISCP-GCN results in limited local perception, leading to a decline in detection performance, especially when dealing with large-scale road networks. The road connectivity degradation in Fig. 13 validates the above description. In conclusion, the experimental results affirm the beneficial impact of the aforementioned key modules on the overall performance of RBFNet. Their organic combination enables the network to comprehensively understand road images and enhances the extraction performance of small objects, regions of interest, edges, and multiscale information.

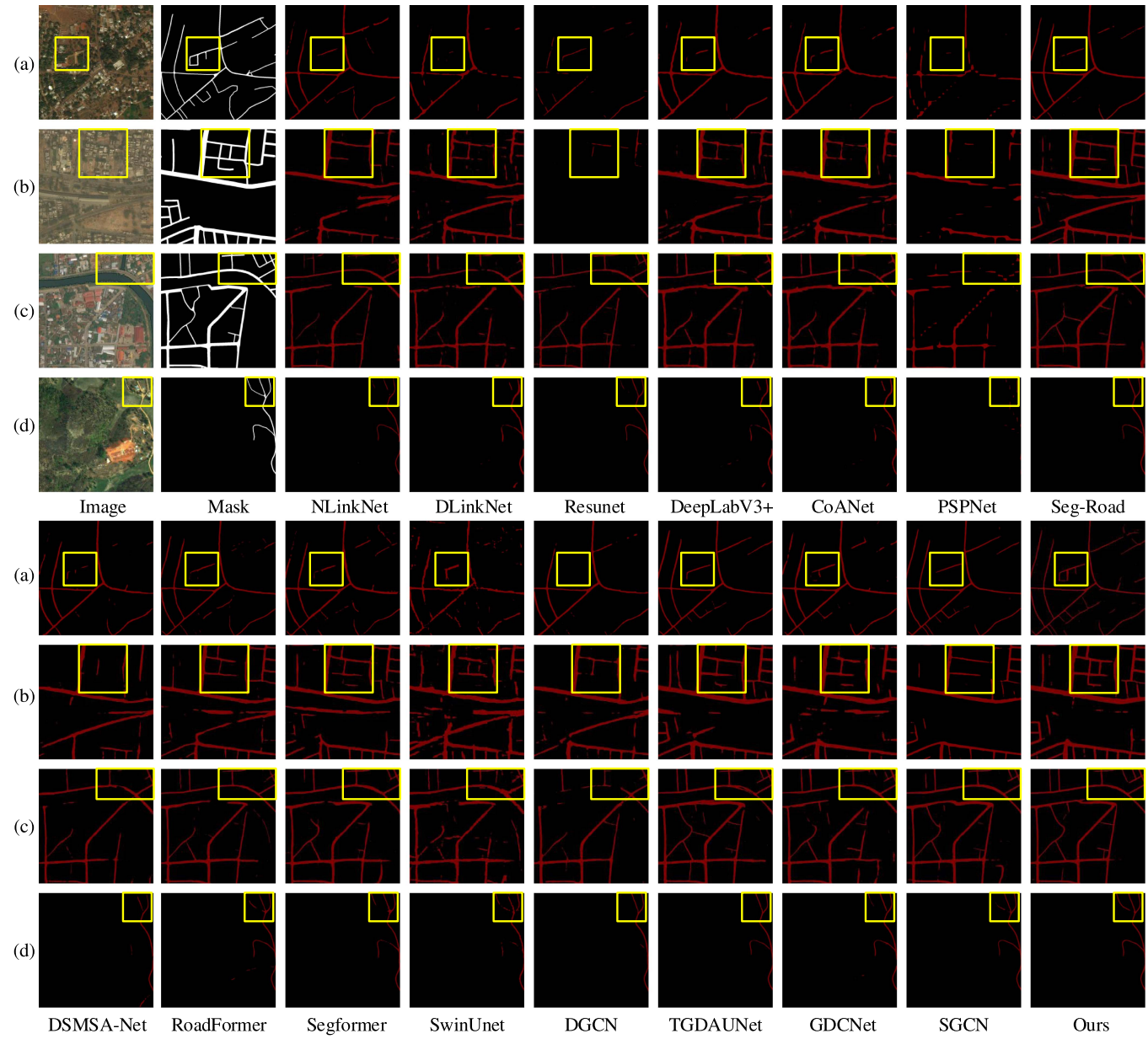

Fig. 9. Visual comparison results of various methods on the DRGE dataset.

2) Multimodule Ablation Experiment: To investigate the effectiveness of combining any two modules, ablation experiments are still conducted on the both datasets, which are recorded in Tables V and VI. From case (I), it is evident that after removing the auxiliary part and ISCP-GCN, the model's ability to extract boundary detail information is significantly decreased. Compared to case (VII), *mF1* and *mIoU* are dropped by 3.59%, 2.78%, and 5.28%, 4.08%, respectively. The removal of LaFormer and ISCP-GCN significantly weakens the RBFNet to eliminate the interference of similar features owing to feebly capturing global context information and multiscale features, thus significantly reducing the performance. In terms of *mF1* and *mIoU*, there are 6.3%, 6.48% and 6.77%, 6.29% decreases compared with case (VII), respectively. When both LaFormer and the auxiliary part are removed, the global modeling capability is weakened, along with its ability to extract local boundary information, leading to *mF1* and *mIoU* in case (III) decreasing by 4.67%, 4.53%, and 6%, 4.25%, respectively, compared to case (VII). Similarly, cases (IV)–(VI) are mainly the removal of BEM combined with several other modules, all of which reduce the boundary detail enhancement ability of the whole model and the semantic sensitivity to varying degrees.

The above phenomenon shows that the combination of different modules can significantly improve the performance of the model. Each module plays a different role in obtaining global context information, extracting multiscale features, enhancing boundary detail processing, and improving road boundary perception. When any two modules are removed, the performance of the model decreases to varying degrees, further verifying the indispensability of each module in the overall model. Figs. 14

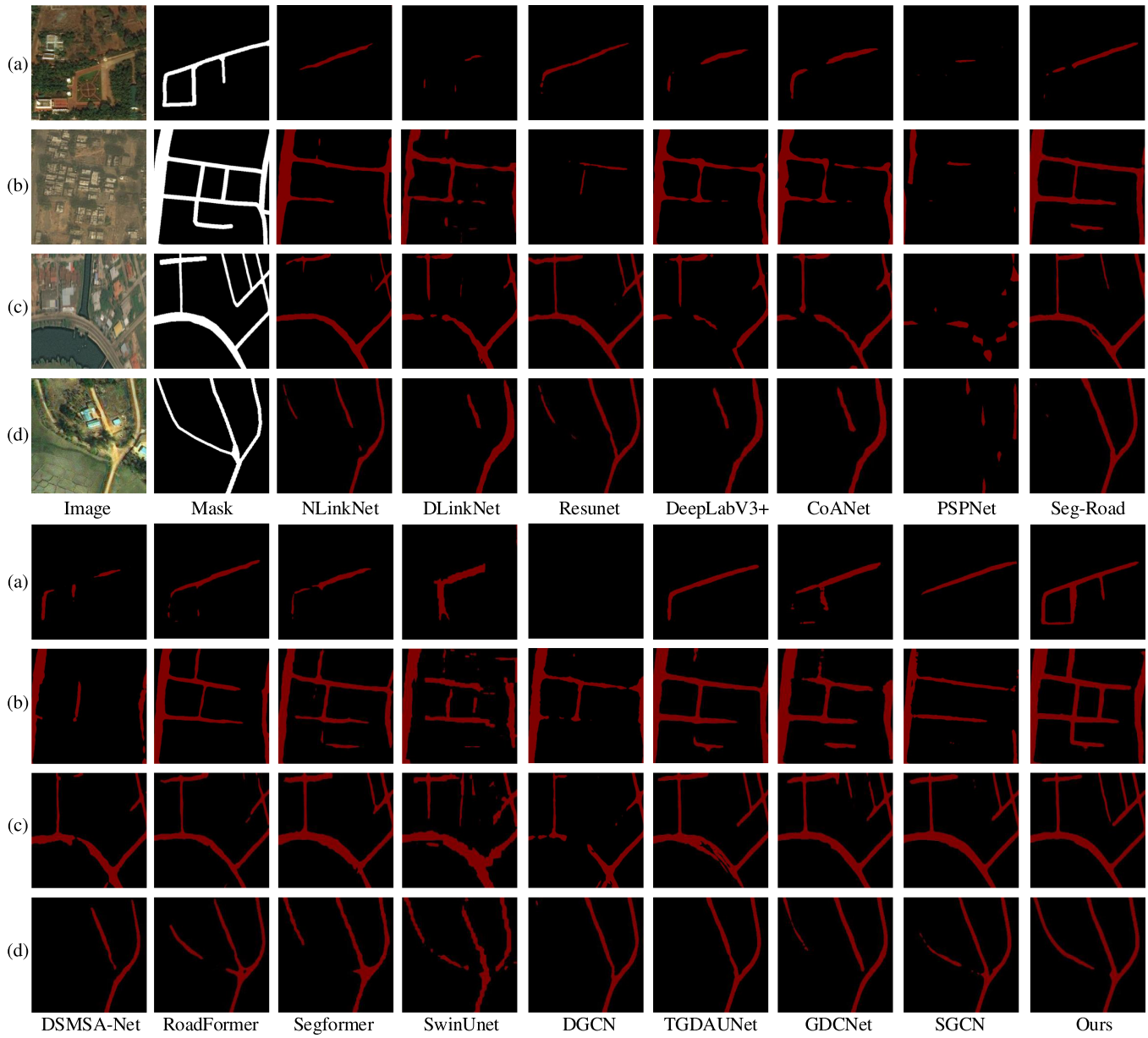

Fig. 10. Comparison of magnified highlighted regions on the DGRE dataset.

TABLE V
ABLATION RESULTS ON THE MR DATASET UNDER ANY TWO MODULES COMBINATIONS

Case	BEM	LaFormer	auxiliary part	ISCP-GCN	*Road IoU*(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑
(I)	w/	w/	**w/o**	**w/o**	55.27	75.95	85.01	84.33	84.67
(II)	w/	**w/o**	w/	**w/o**	50.74	73.24	81.16	80.78	80.97
(III)	w/	**w/o**	**w/o**	w/	55.38	74.87	82.67	83.18	82.92
(IV)	**w/o**	w/	w/	**w/o**	55.21	76.19	84.02	84.08	84.04
(V)	**w/o**	w/	**w/o**	w/	57.83	77.02	84.35	85.40	84.87
(VI)	**w/o**	**w/o**	w/	w/	54.17	75.04	82.37	84.35	83.35
(VII)	w/	w/	w/	w/	**61.96**	**79.54**	**87.77**	**87.13**	**87.45**

w/ and w/o indicate that the module is adopted or not in the test.

TABLE VI
ABLATION RESULTS ON THE DGRE DATASET UNDER ANY TWO MODULES COMBINATIONS

Case	BEM	LaFormer	auxiliary part	ISCP-GCN	*Road IoU*(%)↑	*mIoU*(%)↑	*mP*(%)↑	*mR*(%)↑	*mF1*(%)↑
(I)	w/	w/	**w/o**	**w/o**	55.29	76.42	87.25	82.95	85.05
(II)	w/	**w/o**	w/	**w/o**	50.84	74.93	86.85	79.19	82.84
(III)	w/	**w/o**	**w/o**	w/	54.28	75.70	87.70	82.23	84.88
(IV)	**w/o**	w/	w/	**w/o**	55.41	76.54	88.72	80.97	84.67
(V)	**w/o**	w/	**w/o**	w/	58.63	77.31	89.39	81.78	85.42
(VI)	**w/o**	**w/o**	w/	w/	54.38	76.03	86.83	80.39	83.49
(VII)	w/	w/	w/	w/	**65.48**	**81.70**	**90.85**	**87.48**	**89.13**

w/ and w/o indicate that the module is adopted or not in the test.

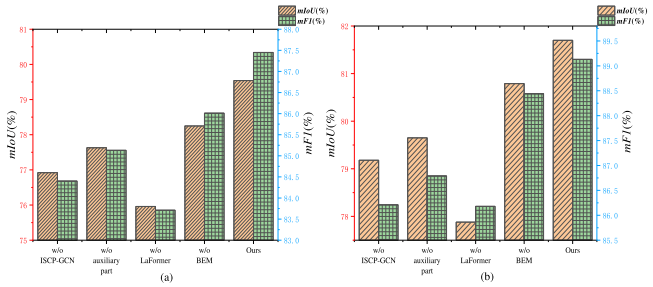

Fig. 11. Ablation experimental results of different modules on the two datasets. (a) MR. (b) DGRE.

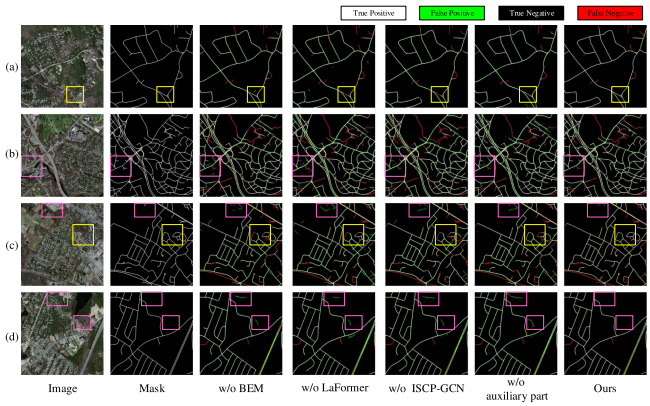

Fig. 12. Visualization results of ablation experiments on the MR dataset for different modules.

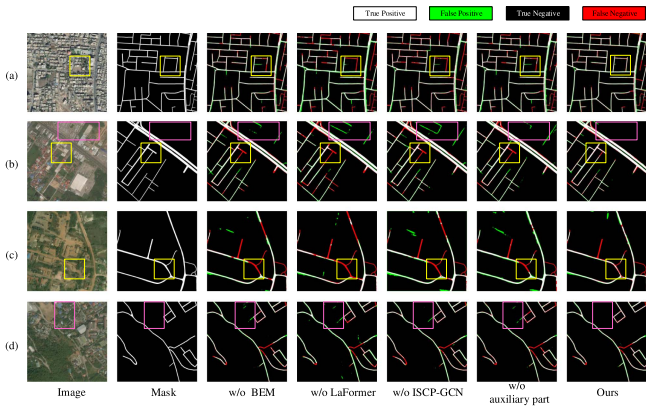

Fig. 13. Visualization results of ablation experiments on the DGRE dataset for different modules.

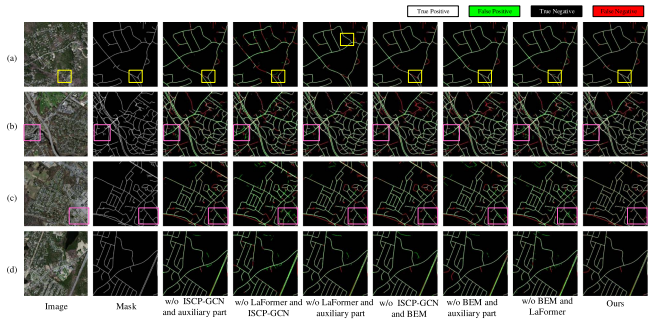

Fig. 14. Visualization results of ablation experiments on the MR dataset for any two modules.

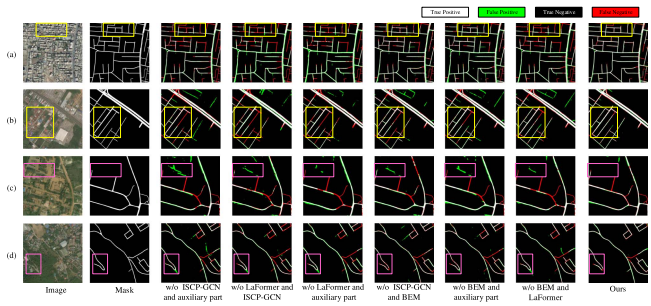

Fig. 15. Visualization results of ablation experiments on the DGRE dataset for any two modules.

and 15 show the qualitative test results of sample instances corresponding to Tables V and VI, respectively, with completeness highlighted in yellow and correctness highlighted in pink. From Fig. 15(b), it can be seen that the combination of BEM and ISCP-GCN plays a crucial role in handling boundary texture details, and the combination of BEM and the auxiliary part is vital for enhancing road boundary perception and ensuring road continuity. Fig. 15(a) exhibits the reduced boundary continuity caused by the above two modules removal. The combination of LaFormer and ISCP-GCN is critical to improving the model's ability to process multiscale features and eliminate interference from similar features, which is particularly highlighted in the result of case (II) in Tables V and VI and Fig. 14(c). In summary, each module plays a unique role in capturing and processing specific information, and their collaborative operation significantly improves the model's performance in complex scenarios.

TABLE VII
PERFORMANCE ANALYSIS OF VARIOUS TRANSFORMER VARIANTS ON THE TWO DATASETS

Variant	MR					DGRE				
	Road IoU (%)↑	*mIoU* (%)↑	*mP* (%)↑	*mR* (%)↑	*mF1* (%)↑	*Road IoU* (%)↑	*mIoU* (%)↑	*mP* (%)↑	*mR* (%)↑	*mF1* (%)↑
MaxVit	57.68	77.30	85.83	86.34	86.08	60.83	79.17	88.76	83.79	86.67
Edter	58.37	78.29	86.71	86.11	86.41	62.18	80.28	89.02	85.36	87.15
Restormer	59.44	78.76	87.63	86.08	86.85	62.71	80.13	88.89	86.04	87.44
DAT	60.34	79.01	87.19	86.91	87.05	63.29	80.64	89.81	86.43	88.09
LaFormer	**61.96**	**79.54**	**87.77**	**87.13**	**87.45**	**65.48**	**81.70**	**90.85**	**87.48**	**89.13**

The bold values indicate the optimal value and underscores indicate the meaning of the suboptimal value.

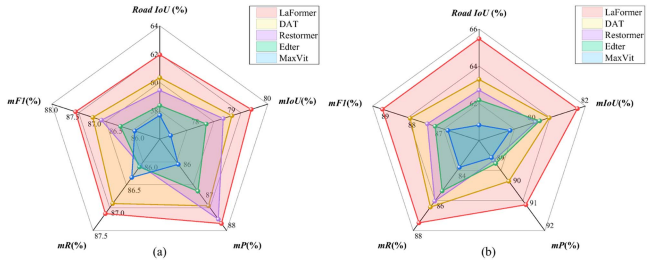

Fig. 16. Comprehensive performance comparison of various transformer variants on the two datasets. (a) MR. (b) DGRE.

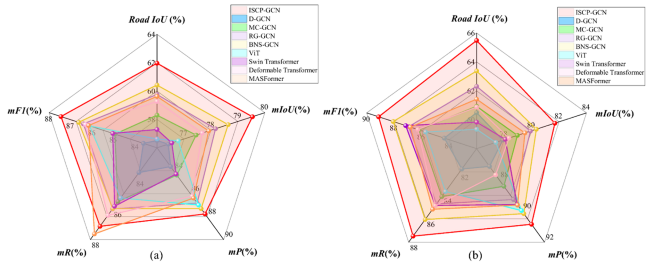

Fig. 17. Comprehensive performance comparison of various GCN and transformer variants on the two datasets. (a) MR. (b) DGRE.

C. LaFormer Analysis

Comparison tests are organized to analyze the performance differences between LaFormer and other transformer variants (such as MaxVit [46], Edter [47], Restormer [48], and DAT [49]), and the results are recorded in Table VII, the corresponding visual results are shown in Fig. 16. DAT combines high-dimensional with low-dimensional features to effectively integrate global and local information, which can enhance the model's performance and generalization ability. However, the collection of local information by the model is relatively limited, failing to provide sufficient detailed information for road segmentation tasks. LaFormer demonstrates excellent performance in road extraction, achieving the best performance on both datasets. As observed from Table VII, LaFormer outperforms the second-best model (i.e., DAT) by 0.54% and 1.62% in the term of *mIoU* and *Road IoU* on the MR dataset. This indicates that the design of LaFormer considers the specific characteristics of road extraction tasks, allowing the model to better handle road structures under different scenarios and environmental conditions. Other models, on the other hand, fail to fully consider the integration of global and local information such as LaFormer, resulting in a more comprehensive understanding of road structures. Similarly, on the DGRE dataset, *mIoU* of LaFormer is 1.06% higher than DAT, while *Road IoU* is 2.19% higher, indicating that the local detail feature enhancement and location adaptive mechanism enable LaFormer to achieve significant improvement. The learning strategy of the model is adjusted adaptively, so that RBFNet can maintain high performance in different scenarios. Fig. 16 also shows that LaFormer offers greater performance and versatility compared to other transformer variants.

D. ISCP-GCN Analysis

To analyze the performance advantages and contributions of ISCP-GCN in road extraction tasks, comparative analyses are conducted with other four GCNs and four transformer variants, including D-GCN [50], MC-GCN [51], RG-GCN [52], BNS-GCN [53], ViT [54], swin transformer [55], deformable transformer [56], and MASFormer [57]. The experimental results are presented in Table VIII and Fig. 17. ISCP-GCN achieves the best results on the MR dataset, outperforming the second-best model (i.e., BNS-GCN) by 0.9% on *mIoU*. This is attributed to its ability to generate graph topologies using different features, resulting in more informative feature representations and excellent analytical capabilities for different-scale road structure patterns, Fig. 17 also fully demonstrates the above situation. The second-best method, BNS-GCN, divides the entire graph by minimizing the number of boundary nodes to reduce communication and memory costs. However, the random sampling of boundary nodes may lead to the loss of important information, thus affecting the accuracy of road detection. Similarly, on the DGRE dataset, ISCP-GCN achieves an *mIoU* 1.35% higher than BNS-GCN, which indicates that integrating boundary features helps improve the perception of road structures. ISCP-GCN comprehensively synthesizes different types of attention matrices, avoiding excessive reliance on specific information from a particular aspect, and can learn more general and robust feature representations.

Both quantitative and qualitative results show that the GCN-based methods outperform the transformer variants in overall performance due to differences in mechanism and scenario application solution requirements. One of the best performing transformer variant is MASFormer, which has a mixed attention

TABLE VIII
PERFORMANCE ANALYSIS OF VARIOUS GCN AND TRANSFORMER VARIANTS ON THE TWO DATASETS

Variant	MR					DGRE				
	Road IoU (%)↑	*mIoU* (%)↑	*mP* (%)↑	*mR* (%)↑	*mF1* (%)↑	*Road IoU* (%)↑	*mIoU* (%)↑	*mP* (%)↑	*mR* (%)↑	*mF1* (%)↑
D-GCN	56.51	76.42	83.64	83.62	83.63	60.49	77.91	87.14	81.74	84.35
MC-GCN	58.30	77.45	84.39	85.51	84.95	60.93	78.97	88.41	84.10	86.20
RG-GCN	59.76	78.17	86.82	86.01	86.41	62.31	79.92	89.26	84.73	86.94
BNS-GCN	60.42	78.64	87.30	85.98	86.63	63.37	80.35	90.17	86.06	88.07
ViT	56.62	76.79	86.99	85.28	86.13	59.34	77.40	89.97	83.67	85.71
Swin Transformer	57.29	76.55	84.29	85.82	85.05	59.83	78.09	89.58	84.36	86.16
Deformable Transformer	59.27	77.68	86.22	86.42	86.32	60.88	78.36	87.68	84.88	86.28
MASFormer	59.62	77.92	86.38	**87.63**	86.17	61.42	79.52	89.62	85.18	86.82
ISCP-GCN	**61.96**	**79.54**	**87.77**	87.13	**87.45**	**65.48**	**81.70**	**90.85**	**87.48**	**89.13**

The bold values indicate the optimal value and underscores indicate the meaning of the suboptimal value.

span and can achieve full focus on long-range dependencies and sparse focus on short-range dependencies. Although it can compensate for a certain degree of local information loss, its performance is still not comparable to other GCNs. The reason ISP-GCN performs well in road extraction tasks can be attributed to its ability to handle detailed textures. The structure of GCN allows it to efficiently capture spatial relationships between pixels and interactions between channels by constructing adjacency matrices, which compels the model to focus more on the details and textures of the road boundaries. In contrast, although transformer variants have certain advantages in global modeling, they often result in blurred boundaries and an inability to capture detail and texture effectively, which is detrimental to tasks requiring precise road extraction.

IV. CONCLUSION

This article proposes a dual-branch fusion model RBFNet for road extraction task in high-resolution remote sensing images. The region-aware branch is based on LaFormer and pays more attention to the regions of interest, thus increasing sensitivity. The boundary-enhanced branch is employed to extract edge details, enhance the model's ability to capture fine-grained features, and supplement the multiscale global semantic features. The introduction of ISP-GCN enables the model to better capture multiscale contextual information and topological structures, and effectively improve the connectivity of road masks. Extensive comparison and ablation experiments demonstrate that RBFNet achieves an average *mIoU* of 79.54% and an *mF1* of 87.45% on the MR dataset, while for the DGRE dataset, the average *mIoU* and *mF1* are 81.70% and 89.13%, respectively. RBFNet performs exceptionally well in extracting road masks from high-resolution remote sensing images, particularly in capturing coverage and small roads. However, the computational complexity of RBFNet is high, which will affect its performance in real-time applications. Meanwhile, the model's quantification and understanding of uncertainty are not thorough enough, which may lead to performance degradation or instability when encountering invisible complex scenes. Future work will focus on addressing the aforementioned potential limitations. Integrated uncertainty modeling techniques are considered to solve the forecast uncertainty, while improving the robustness and reliability of the model under complex scenarios.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed and constructive comments and suggestions. They would like to thank Tian Lan and Associate Professor Shuaishuai Fan of SDTBU for the numerical calculations in this article, and thank Associate Professor Yonghua Jiang of WHU for his suggestion and RAW data.

REFERENCES

[1] Y. Lin, F. Jin, D. Wang, S. Wang, and X. Liu, "Dual-task network for road extraction from high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 66–78, 2023.

[2] R. Hang, S. Xu, P. Yuan, and Q. Liu, "AANet: An ambiguity-aware network for remote-sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5612911.

[3] J. Yang, Z. Gu, T. Wu, and Y. A. E. Ahmed, "RUW-Net: A dual codec network for road extraction from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1550–1564, 2023.

[4] R. Hang, P. Yang, F. Zhou, and Q. Liu, "Multiscale progressive segmentation network for high-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5412012.

[5] W. Li, Q. Liu, S. Fan, H. Bai, and M. Xin, "Multi-stage superpixel-guided hyperspectral image classification with sparse graph attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5519718.

[6] R. Hang, G. Li, M. Xue, C. Dong, and J. Wei, "Identifying oceanic eddy with an edge-enhanced multiscale convolutional network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9198–9207, 2022.

[7] A. Akhtarmanesh, D. Abbasi-Moghadam, A. Sharifi, M. H. Yadkouri, A. Tariq, and L. Lu, "Road extraction from satellite images using attention-assisted UNet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1126–1136, 2023.

[8] Z.-X. Yang, Z.-H. You, S.-B. Chen, J. Tang, and B. Luo, "Semi-supervised edge-aware road extraction via cross teaching between CNN and transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8353–8362, 2023.

[9] Y. Fu, W. Li, S. Fan, Y. Jiang, and H. Bai, "CAL-Net: Conditional attention lightweight network for in-orbit landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4408515.

[10] L. Zhao, D. Guo, and Q. Xu, "Road network information extraction from high-resolution remote sensing images based on improved U-Net," in *Proc. 3rd Int. Conf. Artif. Intell. Comput. Eng.*, 2023, pp. 476–480.

[11] K. Zhang, Y. Han, J. Chen, S. Wang, and Z. Zhang, "Few-shot semantic segmentation for building detection and road extraction based on remote sensing imagery using model-agnostic meta-learning," in *Proc. Int. Conf. Guid. Navigation Control*, Tianjin, China, 2022, pp. 1973–1983.

[12] Y. Zhang, Q. Zhu, Y. Zhong, Q. Guan, L. Zhang, and D. Li, "A modified D-linknet with transfer learning for road extraction from high-resolution remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1817–1820.

[13] Y. Wang, J. Seo, and T. Jeon, "NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 3000105.

[14] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
[15] X. Zhu, X. Huang, W. Cao, X. Yang, Y. Zhou, and S. Wang, "Road extraction from remote sensing imagery with spatial attention based on swin transformer," *Remote Sens.*, vol. 16, no. 7, 2024, Art. no. 1183.
[16] X. Jiang et al., "RoadFormer: Pyramidal deformable vision transformers for road network extraction with remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, 2022, Art. no. 102987.
[17] L. Gao et al., "Road extraction using a dual attention dilated-linknet based on satellite images and floating vehicle trajectory data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10428–10438, 2021.
[18] L. Luo, J.-X. Wang, S.-B. Chen, J. Tang, and B. Luo, "BDTNet: Road extraction by bi-direction transformer from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2505605.
[19] J. Tao et al., "SEG-Road: A segmentation network for road extraction based on transformer and CNN with connectivity structures," *Remote Sens.*, vol. 15, no. 6, 2023, Art. no. 1602.
[20] S. Huang, L. Liu, J. Dong, X. Fu, and F. Huang, "SPGCN: Ground filtering method based on superpoint graph convolution neural network for vehicle LiDAR," *J. Appl. Remote Sens.*, vol. 16, no. 1, pp. 016512–016512, 2022.
[21] W. Li, Y. Fu, S. Fan, M. Xin, and H. Bai, "DCI-PGCN: Dual channel interaction portable graph convolutional network for landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4403616.
[22] J. Yan, S. Ji, and Y. Wei, "A combination of convolutional and graph neural networks for regularized road surface extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4409113.
[23] G. Zhou, W. Chen, Q. Gui, X. Li, and L. Wang, "Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5614115.
[24] F. Cui, Y. Shi, R. Feng, L. Wang, and T. Zeng, "A graph-based dual convolutional network for automatic road extraction from high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 3015–3018.
[25] X. Liu et al., "RoadFormer: Road extraction using a swin transformer combined with a spatial and channel separable convolution," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 1049.
[26] Z. Z. Abidin, S. Asmai, Z. A. Abas, N. Zakaria, and S. Ibrahim, "Development of edge detection for image segmentation," in *IOP Conference Series: Materials Science and Engineering*. Bristol, U.K.: IOP Publishing, 2020.
[27] Y.-J. Zhang and Y.-J. Zhang, "Edge detection," *Handbook Image Eng.*, pp. 859–899, 2021.
[28] M. Hytch and P. W. Hawkes, *Morphological Image Operators*. Cambridge, MA, USA: Academic Press, 2020.
[29] B. Cui and H. Jiang, "An image edge detection method based on haar wavelet transform," in *Proc. Int. Conf. Artif. Intell. Comput. Eng.*, 2020, pp. 250–254.
[30] O. R. Vincent et al., "A descriptive algorithm for Sobel image edge detection," in *Proc. Inf. Sci. IT Educ. Conf.*, 2009, pp. 97–107.
[31] X. Liang, Z. Hu, H. Zhang, L. Lin, and E. P. Xing, "Symbolic graph reasoning meets convolutions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018.
[32] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, and H. Liu, "Spatial pyramid based graph reasoning for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. pattern Recognit.*, 2020, pp. 8950–8959.
[33] L. Dang, Y. Nie, C. Long, Q. Zhang, and G. Li, "MSR-GCN: Multi-scale residual graph convolution networks for human motion prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 11467–11476.
[34] M. Xin, Y. Fu, W. Li, H. Ma, and H. Bai, "Delformer: Detail-enhanced lightweight transformer for road segmentation," *J. Appl. Remote Sens.*, vol. 17, no. 4, pp. 046507–046507, 2023.
[35] I. Demir et al., "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–181.
[36] V. Mnih, "Machine learning for aerial image labeling," Univ. Toronto, Toronto, ON, Canada, 2013.
[37] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 182–186.
[38] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
[39] J. Mei, R.-J. Li, W. Gao, and M.-M. Cheng, "CoANet: Connectivity attention network for road extraction from satellite imagery," *IEEE Trans. Image Process.*, vol. 30, pp. 8540–8552, 2021.
[40] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.
[41] S. D. Khan, L. Alarabi, and S. Basalamah, "DSMSA-Net: Deep spatial and multi-scale attention network for road extraction in high spatial resolution satellite images," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 1907–1920, 2023.
[42] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 12077–12090.
[43] H. Cao et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 205–218.
[44] H. Liang, J. Lv, Z. Wang, and X. Xu, "Medical image mis-segmentation region refinement framework based on dynamic graph convolution," *Biomed. Signal Process. Control*, vol. 86, 2023, Art. no. 105064.
[45] P. Song, J. Li, H. Fan, and L. Fan, "TGDAUNet: Transformer and GCNN based dual-branch attention UNet for medical image segmentation," *Comput. Biol. Med.*, vol. 167, 2023, Art. no. 107583.
[46] Z. Tu et al., "Maxvit: Multi-axis vision transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 459–479.
[47] M. Pu, Y. Huang, Y. Liu, Q. Guan, and H. Ling, "Edter: Edge detection with transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1402–1412.
[48] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5728–5739.
[49] J. Park, M. Son, S. Lee, and C. Kim, "Dat: Domain adaptive transformer for domain adaptive semantic segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2022, pp. 4183–4187.
[50] F. Ye, S. Pu, Q. Zhong, C. Li, D. Xie, and H. Tang, "Dynamic GCN: Context-enriched topology learning for skeleton-based action recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 55–63.
[51] X. Shi, X. Chai, J. Xie, and T. Sun, "MC-GCN: A multi-scale contrastive graph convolutional network for unconstrained face recognition with image sets," *IEEE Trans. Image Process.*, vol. 31, pp. 3046–3055, 2022.
[52] W. Ding et al., "RG-GCN: Improved graph convolution neural network algorithm based on rough graph," *IEEE Access*, vol. 10, pp. 85582–85594, 2022.
[53] C. Wan, Y. Li, A. Li, N. S. Kim, and Y. Lin, "BNS-GCN: Efficient full-graph training of graph convolutional networks with partition-parallelism and random boundary node sampling," in *Proc. Mach. Learn. Syst.*, vol. 4, pp. 673–693, 2022.
[54] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[55] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.
[56] P.-C. Hu, S.-B. Chen, L.-L. Huang, G.-Z. Wang, J. Tang, and B. Luo, "Road extraction by multi-scale deformable transformer from remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2503905.
[57] Q. Zhang, D. Ram, C. Hawkins, S. Zha, and T. Zhao, "Efficient long-range transformers: You need to attend more, but not necessarily at every layer," 2023, *arXiv:2310.12442*.

Weiming Li (Member, IEEE) received the B.S. and Ph.D. degrees in weapon science and technology from the School of Energy and Power Engineering, Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2009 and 2014, respectively.

From 2015 to 2017, he was a postdoctoral research fellow with the China Academy of Space Technology, Beijing, China. He is currently an Associate Professor with the Shandong Technology and Business University, Yantai, China. His fields of interest include remote sensing image processing, machine learning, and pattern recognition.

Tian Lan received the B.S. degree in network engineering from the Liaocheng University, Liaocheng, China, in 2022. He is currently working toward the M.S. degree in electronic information with the School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, China.

His fields of interest include computer vision and remote sensing image processing.

Shuaishuai Fan (Member, IEEE) received the M.S. degree in applied mathematics from the Beijing Institute of Technology, Beijing, China, in 2013.

She is currently an Associate Professor with the Shandong Technology and Business University, Yantai, China. Her research interests include the remote sensing image processing, pattern recognition, and deep learning.

Yonghua Jiang (Member, IEEE) received the B.S. and Ph.D. degrees in remote sensing science and technique from the School of Remote Sensing and Information Engineering, Wuhan University (WHU), Wuhan, China, in 2010 and 2015, respectively.

He is currently an Associate Professor with the WHU. His research interests include geometry processing of spaceborne optical imagery.