A Fuzzy-Boundary Enhanced Trident Network for Parcel Extraction in the Urban–Rural Area

Jiajun Zhu , Zhuoqun Chai, Zhaopu Song, Zhanpeng Chen, and Qian Shi , *Senior Member, IEEE*

Abstract—As the basic unit of farmland, parcel is crucial for remote sensing tasks, such as urban management. Previous studies of farmland parcels extraction are based on boundary detection and instance segmentation methods. However, these methods perform poorly in the parcels with complex shape and fuzzy boundary due to the insufficient feature extraction capability. Moreover, for the lack of multiscale features extraction and fusion, they are difficult to extract different scale farmland parcels accurately. Based on these issues, we propose a fuzzy-boundary enhanced trident network, named FBETNet, to enhance the feature of fuzzy boundary and generate multiscale parcels. First, a semantic-guided multitask strategy is introduced in order to enhance the feature of fuzzy boundary. Second, we design a multiscale trident module to further improve the performance of multiscale feature extraction. Finally, an adversarial data augmentation strategy is employed in the training phase to strengthen the robustness and stability of our proposed method. Experiments show that our proposed method improves significantly in both accuracy and visualization, especially for the parcels with fuzzy boundary and complex shape.

Index Terms—Adversarial data augmentation (ADA), deep learning, fuzzy boundary, multiscale feature, parcel extraction, urban—rural area.

I. Introduction

RAPID urbanization has brought drastic changes to the regional land use structure. Due to the impact of urbanization, the croplands in urban–rural area are usually maliciously occupied for other usages [1]. With the rapid expansion of urban construction land, the sharp decrease of cropland resources, and the serious destruction of the urban ecological environment, we need to seek a sustainable and growing way of development more urgently. Nowadays, in some cities [2], [3], [4], the cropland information of land use and land cover (LULC) data has been used to simulate and analyze the process of urban expansion in order to obtain the accurate urban management measures. However, due to the coarse resolution and detail, LULC is not sufficient to carry out precise urban expansion simulation. Therefore, we require more refined cropland data to assist in urban decision making. As the basic unit of farmland, parcel is crucial for cropland information extraction [5]. Thanks to its clear boundaries and distinct geometry properties, cropland parcel gradually substitutes LULC in urban management. Nevertheless, the cropland boundaries in urban–rural area are often fuzzy, making parcels extraction be a tough task. As a consequence, how to enhance the capability of fuzzy-boundary extraction has become an international research hotspot in remote sensing.

In recent years, based on the rapid development of imaging sensors and operating platforms, the increased availability of very high resolution (VHR) remote sensing imagery provides clearer texture and spatial information [6], [7], thus providing richer details for parcels. However, for the lack of available training data and the limited ability of feature extraction, parcels extraction from VHR imagery still remains a challenge. The existing cropland parcel datasets [8] are mainly concentrated in Europe cities, which lack the parcels with complex shape and fuzzy boundary, failing to meet the demands of urban management. In addition to the limitation of training data, the distribution of cropland is restricted by topography properties and complex planting structure with different shapes and textures, making it harder to obtain precise parcels mapping by a single model [9]. Thereby, how to consider more comprehensive parcels information in single model to generate the parcel-level data is of great significance for the development of urban management.

The cropland parcels extraction methods are usually based on edge detection, image segmentation, and deep learning. The traditional edge detection methods obtain the contours of parcels by using specific operators, such as Sobel [10], Laplacian [11], and Canny [12], and then utilize the contours information to form the segmented area in the image. In contrast, methods based on image segmentation aggregate similar pixels by calculating homogeneous information in order to obtain more compact and closed parcels' boundaries [13]. However, these traditional methods will lead to rough and chaotic cropland boundaries delineation due to the poor performance of feature extraction, resulting in misclassification and omission, which cannot meet the demand of urban management. In recent years, the deep

Jiajun Zhu, Zhuoqun Chai, and Qian Shi are with the Guangdong Provincial Key Laboratory for Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: zhujj26@mail2.sysu.edu.cn; chaizhq@mail2.sysu.edu.cn; shixi5@mail.sysu.edu.cn).

Zhaopu Song and Zhanpeng Chen are with the Lands and Resource Department of Guangdong Province, Surveying and Mapping Institute, Guangzhou 510500, China, also with the Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical Area of South China, Ministry of Natural Resources, Guangzhou 510631, China, and also with the Guangdong Natural Resources Science and Technology Collaborative Innovation Center, Guangzhou 510663, China (e-mail: 446724136@qq.com; 543283684@qq.com).

learning method has been successfully applied to remote sensing due to its success in image processing [14], [15]. Convolutional neural network (CNN) can implicitly learn advanced deep features from training data automatically through local connection, weight sharing and sampling, reducing the shortcomings of artificial designed features, and error accumulation [16], [17]. As a result, it is potential to extract parcels' boundaries through CNN-based methods. Nevertheless, parcels' boundaries present different features in VHR images, which is the major obstacle to parcels extraction [18]. But the existing methods pay less attention to the fuzzy boundaries and, thus, making it difficult to extract the fuzzy boundaries of parcels. Therefore, how to comprehensively consider sufficient boundary and semantic features in CNN-based methods to obtain more refined results is crucial for urban management.

At present, boundary extraction is also a hot topic in deep learning, so there are some advanced methods suitable for detecting boundary features [19], [20]. As the deep learning boundary extraction methods gradually mature in remote sensing, they are directly applied to parcels extraction. Garcia-Pedrero et al. [21] used a CNN method to automatically extract the boundaries of large cropland parcels with heterogeneous landscapes based on the open data of the land parcel identification system of the Navarra concession community in Spain. Masoud et al. [22] used a fully convolutional network (FCN) to detect cropland boundaries from Sentinel-2 images at a resolution of 10 m and then employ FCN structure to increase the spatial resolution of the output from 10 to 5 m. Although the existing methods succeed in the parcels extraction, they still perform poorly in the parcels with complex shape and fuzzy boundary due to the insufficient feature extraction capability, making the boundaries result disconnected. Moreover, there is a lack of methods to extract parcels on VHR images, which cannot meet the resolution demand of current urban management. Finally, in order to obtain the detailed mapping result of parcels, we need to consider the multiscale features of parcels. However, current methods hardly take this problem into account. Due to the diversity and complexity of parcels' features, it is difficult for single-scale models to accurately express the appearance features of complex parcels, resulting in some missing detection. The lack of multiscale consideration will affect the performance of parcels extraction and make it hard to obtain more refined results. These issues will cause some negative effects on the practical application. Therefore, it is very important to propose a fine-grained cropland parcel extraction method suitable for VHR images.

In view of the above problems, we propose a fuzzy-boundary enhanced trident network, named FBETNet, to enhance the feature of fuzzy boundary and generate multiscale parcels. First, we introduce a semantic-guided multitask strategy to learn the semantic features and boundary features of parcels simultaneously. At the same time, on the basis of the simple multitask learning methods, a guidance module is added to enhance the semantic information of the fuzzy boundaries by using the feature in the semantic branch to guide the boundaries detection and to supplement the missing parcels' boundaries, thereby improving the detection performance of parcels. Second, due to the complexity of the parcels, it is vital to consider the multiscale

features. Therefore, a multiscale trident (MST) module is designed to obtain multiscale outputs, which can ensure the feature consistency and reduce computational cost. Finally, in order to further enhance the robustness of the proposed method and make it more generalized in different scenarios, an adversarial data augmentation (ADA) strategy is employed in the training phase, which simulates feature learning process of both the clear and fuzzy boundaries. We validate our method on a high-quality parcels' annotations google Earth images dataset.

The main contributions of this article are summarized as follows.

1) We propose a semantic-guided multitask strategy to enhance the extraction ability of fuzzy boundaries.
2) We propose an MST module, which fully considers the features of parcels at different scales, making the results more accurate.
3) We propose an ADA strategy to make the model more robust and generalized to various scenarios.

II. RELATED WORK

Current parcel boundary extraction approaches rely on image segmentation-based methods, such as thresholding segmentation, edge detection, and region segmentation algorithms. However, with the expansion of remote sensing image data sources and the popularity of deep learning, CNN-based neural networks, attention mechanism, multitask learning, multiscale learning, and other deep learning methods have gained more and more concentration and become the mainstream methods of parcels extraction.

The traditional parcel extraction methods primarily exploit the limited information of local regions, which are computationally efficient, overcoming the repetitive and time-consuming work of manually labeling process. Torre and Radeva [13] combined region growing algorithm and deformable models to design a semiautomatic framework for cropland segmentation. Yan and Roy [23] used multitemporal Landsat data and watershed algorithm to decompose the multiple parcels into isolated parcel and connected the correlated circular parcels with geometry-based algorithmic detection. Cheng et al. [24] fused the spatial features of VHR images with temporal features of multitemporal images to extract cropland boundaries from submeter images (2–3 m). However, because of the lack of automation and generalization ability of the traditional methods, they cannot meet the needs of current applications.

In contrast, deep learning approaches, such as instance segmentation, attention mechanisms, adversarial generative learning, edge detection, and multitask learning, are gradually maturing and have been extensively applied. Accurately detecting the location and category of objects in images is the core issue focused by deep learning and remote sensing task, and the instance segmentation methods can effectively determine the location of parcels and closed objects. Rieke [25] used FCN to conduct the instance segmentation task and obtained cropland parcels in Sentinel 2 images without any preprocessing and postprocessing process. Potlapally et al. [26] performed instance segmentation on rural India using mask R-CNN and achieved better farmland extraction results. Mei et al. [27] used mask

R-CNN and WorldView-3 satellite images to better delineate the cropland boundaries in northeast India. However, these methods suffer from problems, such as inaccurate extraction of shared edges and omission detection of fuzzy boundaries, failing to meet accuracy requirements in urban management.

The attention mechanism is widely exploited in the deep learning model to further highlight the parcel characteristics. Attention mechanism simulates the attention pattern of human eyes and can focus on the area of interest according to the attention weight. Therefore, it is really suitable in parcels extraction. Li et al. [28] introduce the spatial and channel attention modules between the basic convolution layers and the max pooling layers, which can adaptively enhance the feature to improve the capability of the network for small cropland targets in complex scenarios. Xu et al. [29] combine LSTM structure and attention mechanism to achieve dynamic mapping of corn and soybean based on multispectral and multitemporal images.

To produce further refined parcels data, it is necessary to incorporate the multiscale features of the parcels, and hence, the multiscale learning strategy is gradually used in the remote sensing application. Multiscale learning can consider the features of the target under different scales so as to provide more accurate information for model learning. In deep learning, there are a variety of multiscale models, mainly in the input side [30] and feature fusion [31]. In order to match the scale variations of different objects in remote sensing images, Deng et al. [32] proposed a multiscale target detection model for remote sensing images by combining a multiscale object proposal network and the accurate object detection network. Garnot and Landrieu [8] used time-sequence coding network to extract rich and adaptive multiscale spatiotemporal features of crops from multitemporal images. However, the existing methods for parcels extraction lack multiscale features learning. Therefore, we design an MST module to capture parcel features at different scales in this article.

In the process of parcels extraction, different boundary clarity will make the model difficult to distinguish parcels' boundaries, so adversarial learning will occur in this process. Adversarial learning strategies have been demonstrated to be effective for semantic segmentation [33] as well as object detection [34], including building extraction [35], cloud removal [36], road detection [37], and ship identification [38]. For example, Jong et al. [39] attempt to use Res-UNET structure as a generative adversarial network generator to improve cropland boundary prediction through an adversarial training strategy. In this article, instead of designing the model structure, we incorporates adversarial learning into the data augmentation strategy to enhance the robustness of model.

Currently, edge detection is still the mainstream method of parcels extraction, which is more potential as it can directly identify the exact location of parcels' boundaries. Xia et al. [18] integrated UNET model and RCF model to extract hard and soft parcels edges, respectively. Masoud et al. [22] designed a multiple dilation full convolutional network and incorporated the semantic contour information extracted by the super-resolution network to obtain parcels' boundaries from Sentinel-2 images. Persello et al. [40] also modified the FCN to learn complex spatial contextual features for better detection of sparse cropland

contours, followed by a grouping algorithm to obtain cropland parcels from the hierarchical segmentation features. However, these single parcels' boundaries detection methods usually fail to accurately identify ambiguous fuzzy boundaries, leading to omission detection in complex scenarios.

Considering the limitations of single boundary detection methods, multitask learning is used in some methods in order to obtain more adequate parcels' features [41]. Multitask learning strategy focuses on the connection between different tasks and learns the joint representation features to improve the performance of the model. Long et al. [42] proposed a new multitask neural network Bsi-Net to construct three parallel decoders to learn the core task of cropland identification and two auxiliary tasks for cropland boundary prediction and distance estimation, respectively. Waldner and Diakogiannis [43] designed a multitask semantic segmentation model using ResNet as the backbone to jointly learn the extent, boundary, and distance information of farmland, and obtained closed boundaries after postprocessing. Sharifi et al. [44] constructed a multitask UNET using residual modules and skip connections to obtain precise cropland parcels on Sentinel-2 and Landsat-8 images. Although these methods leverage the potential of multitask learning in parcel extraction, they still fail to exploit the direct correlation between different tasks and lack an interaction in the model, so the enhancement of parcels extraction through multitask learning strategy is limited and the ability of fuzzy boundaries extraction is still insufficient. Therefore, in this article, we further explore the interaction guidance process in the multitask learning strategy to obtain better fuzzy boundaries extraction results.

III. METHOD

A. Overview

The cropland parcel types within the urban–rural area are various and complicated to extract precisely, leading to errors in urban management and expansion simulation. The parcel extraction inaccuracies in these area are primarily concentrated in the fuzzy boundaries, which fail to meet the demand of urban management. In order to solve these problems, in this research, we propose a fuzzy-boundary enhanced trident network, named FBETNet, to enhance the feature of fuzzy boundary and generate multiscale parcels in urban–rural area. The overall network structure is shown in Fig. 1, which contains three main sections:

1) guidance multitask (GMT) strategy;
2) MST module;
3) ADA strategy.

We employed the D-LinkNet [45] structure as the backbone with pretraining weights from resnet34 in the encoder and added the designed modules and strategies to the whole model for experiments.

First, to address the issue of difficulty in extraction of fuzzy boundaries, we introduce a GMT learning strategy with two branches for semantic segmentation and boundary detection, respectively. We guide the boundary detection branch with the semantic information obtained from the segmentation branch to enrich the features of fuzzy parcels' boundaries. Second, due to the multiscale features of parcels in urban–rural area, specific-scale method cannot obtain fine-grained parcels results.
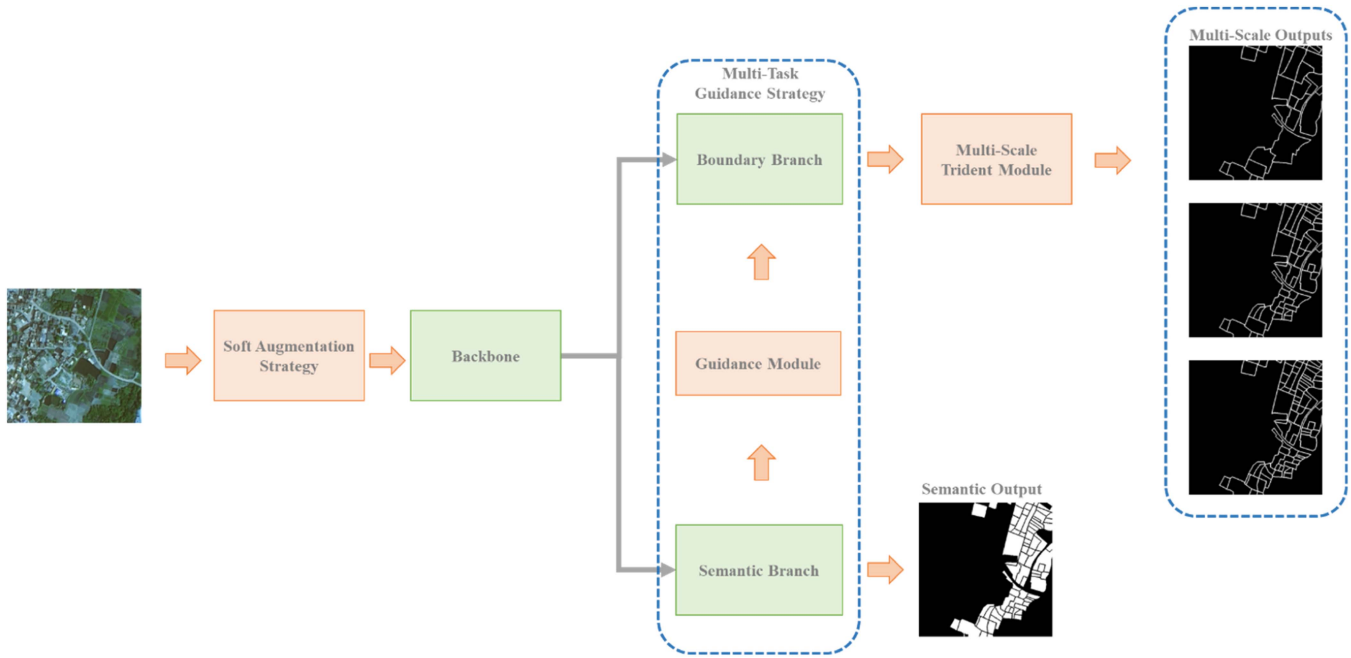
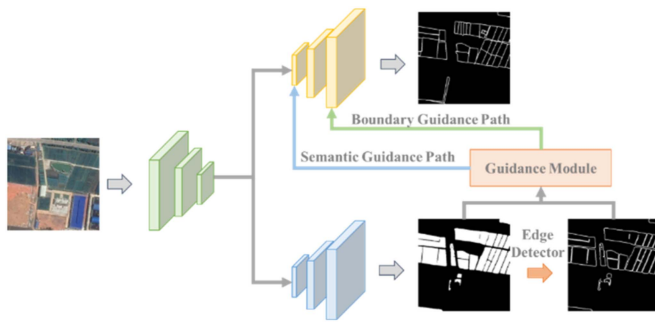Fig. 1. Overview architecture of the proposed network.

Fig. 2. GMT learning strategy.

Therefore, we propose an MST module to capture different scales' features for refined parcels. In order to further improve the robustness of the model so that it can achieve accurate extraction for complex fuzzy boundaries in numerous circumstances, we adopt an ADA strategy in the training phase learn clear and fuzzy features simultaneously.

B. GMT Strategy

In order to improve the extraction capability on fuzzy and complex boundaries, we devised a GMT strategy, as illustrated in Fig. 2.

The structure differs from common simple multitask learning in which it has a close relationship between different task branches, and two interaction routes, semantic-guided and boundary-guided, are designed. Since boundary information is identified by elaborate low-level features, detecting the boundary directly is difficult if the boundary is vague or complex. Different from simple multitask learning strategy, GMT designs

an interaction between the two branches. It contains two interaction paths, semantic guidance path and boundary guidance path. Since boundary information is identified by low-level features, directly to detect the fuzzy boundaries is difficult. However, semantic information can be captured through high-level features without the assistance of high-resolution detailed features, which can identify the coarse location of the parcels more accurately so that it can incorporate the semantic features into boundary detection branch to make the model pay more attention to the cropland region. Therefore, the semantic guidance path can enhance the feature extraction ability of fuzzy boundaries, and thus enabling more refined and ample boundaries detected by the boundary branch. The boundary guidance path enhances the fuzzy boundaries extraction results more directly by using the boundary information detected by edge operator, such as Sobel, and incorporating it into the boundary detector to further supplement the omissive fuzzy-boundary features. Through these two paths, we can enhance the performance of multitask learning strategy and can capture more complete parcel boundary features in the boundary branch, including the fuzzy-boundary features, which are difficult to learn.

The details of these two paths are shown in Fig. 3. In the semantic guidance path, we downsample the segmentation result in semantic branch and directly incorporate it into the high-level feature of the boundary branch to provide richer and more abundant semantic information. As shown in (1), the semantic features $S_{m \times n}$ are downsampled to the same size as the high-level feature by average pooling layer to obtain $S_{\text{down}}(m_d \times n_d)$, which is operated by © with the high-level features $F_h^B(m_d \times n_d)$, of the boundary branch. In boundary guidance path, we obtain the boundary information directly from the segmentation result by edge detection operator and incorporate it into the low-level feature of boundary branch after
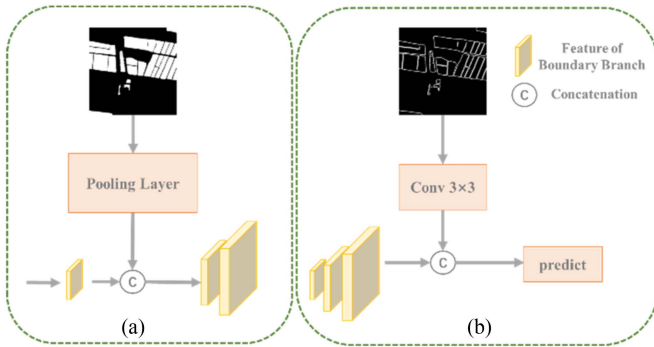
Fig. 3. Architecture of two guidance paths. (a) Semantic guidance path. (b) Boundary guidance path.

convolution operation. As shown in (2), the boundary information from semantic branch is superimposed onto the low-level features F_l^B in the boundary branch after the convolution operation to supplement missing fuzzy boundary directly. By using such an interactive multitasking strategy, it can greatly enhance the features of fuzzy boundaries and extract more refined and complete parcels

$$\widetilde{F_h^B} = \text{Avg}(S) © F_h^B \qquad (1)$$

$$\widetilde{F_l^B} = \text{Conv}_{m \times m}(B) © F_l^B \qquad (2)$$

where F_h^B and F_l^B are the high-level features and low-level features of the decoder in the boundary branch, respectively. $\widetilde{F_h^B}$ and $\widetilde{F_l^B}$ are the high-level features and low-level features of the after optimization by the guidance module. S and B denote the segmentation results and their corresponding edge detection results from the semantic branch. © is the concatenation operation, Avg is the average pooling operation, and $\text{Conv}_{m \times m}$ represents the convolution operation of $m \times m$ size.

C. MST Module

The croplands located at urban–rural area have multiscale features. In order to obtain refined parcels data to assist the urban management, it is essential to consider the multiscale parcels' features learning. We adopt a module similar to the article presented in [46] and modified it for multiscale parcels extraction. As mentioned in [46], we have summarized four main structures of multiscale module, as illustrated in Fig. 4, which are image pyramid, simple feature pyramid, top-down pyramid structure, and trident pyramid. The image pyramid [see Fig. 4(a)] represents that in the inference stage, the input images are constructed into different resolution and then obtain the results of different scales. Although this module obtain better performvvance, it is rarely used in practical applications on account of huge calculational cost. The simple feature pyramid [see Fig. 4(b)] refers to the direct predictions of the feature of different scales and layers. This module usually performs poorly due to information lost in the downsampling operation. The top-down feature pyramid [see Fig. 4(c)] designs skip connections and top-down upsampling in the simple feature pyramid to reduce the information lost. However, this module

has a flaw that makes it an unsatisfactory alternative for image pyramid. The feature pyramid generates multilevel features, thus sacrificing the feature consistency across different scales. This leads to a decrease in effective training data and a higher risk of overfitting for each scale. The MST pyramid [see Fig. 4(d)] is a further improvement on the basis of the previous modules, which can not only reduce the computational cost but also ensure the feature consistency. Our proposed multiscale module is introduced based on this structure.

The MST module used in this article is shown in Fig. 5. We combined the top-down feature pyramid and trident pyramid to extract the multiscale features with the backbone of D-LinkNet. The purpose is to take advantage of the two multiscale structures by maintaining the integrity of the semantic information through the top-down structure in the front and then ensure the feature consistency of the multiscale results through the trident structure.

The details of the trident module can be shown in Fig. 6. It can be seen that the three paths are weight sharing, making it a better performance similar to image pyramid. Therefore, after the atrous convolution operation with different dilated rates, we can obtain parcels at different scales. After obtaining the multiscale results, we directly fusion them through addition operation. Then, the results are optimized by mathematical morphology processing, such as erode and dilate algorithm. Finally, we extract the skeleton lines of the results and obtain the refine parcels results.

D. Adversarial Data Augmentation

Due to the limited training data, it is difficult for the model to be stable in a variety of scenarios. Therefore, in order to promote the robustness of model so that it can achieve accurate and refined extraction results in different scenarios with complex fuzzy boundaries, we employ an ADA strategy in the training phase. During the model training, we use both edge sharpening and edge blurring methods to augment the training data. Edge blurring will weaken the information of the image so that blur the image boundary. After edge sharpening process, the details of the parcels' boundaries will be richer and clearer. While after edge blurring process, the parcels' boundaries will tend to be smooth and blurred. Edge sharpening can simulate the process of learning clear boundaries in model training. However, the edge blurring is tend to simulate the process of fuzzy boundaries extraction. Therefore, the model will balance the clear and fuzzy boundaries' features through an adversarial manner after convergence in order to perform more robust in parcels extraction. Edge sharpening will strengthen the edge features of the image so as to highlight the details of the image texture. Sobel operator and median filter are used for edge sharpening and blurring, respectively, and the results are shown in Fig. 7. We can see that the image boundary information is richer in detail after sharpening(clear boundary) and relatively vaguer after blurring (fuzzy boundary).

In addition to these two adversarial augmentation methods, we also perform the random color transformation for the robust performance in multiple scenarios. Fig. 8 shows the overall ADA
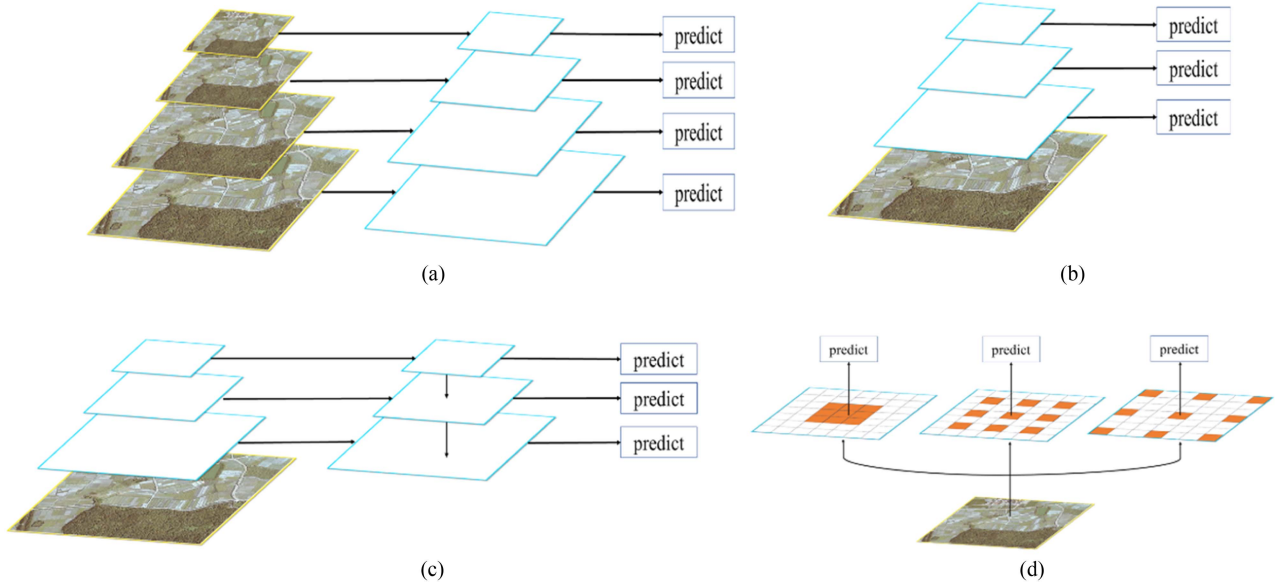
Fig. 4. Four structures of pyramid module. (a) Image pyramid. (b) Simple feature pyramid. (c) Top-down feature pyramid. (d) Triden pyramid.
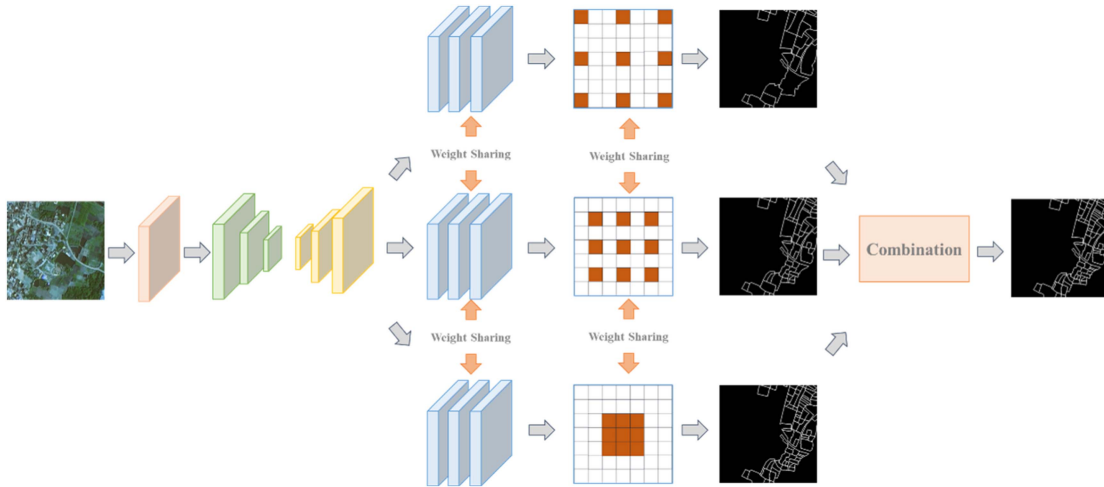

Fig. 5. Architecture of our MST module.

strategy. After sharpening and blurring, we obtain three different images. And then the three images are input as a batch for model training after random color transformation and random flip operation. After converging in the train phase, the model can perform well in both clear and fuzzy boundaries and can have robust results in different scenarios.

E. Loss Function

To better supervise our proposed model and reach better performance, we used cross-entropy loss and Dice loss [47] simultaneously. The cross-entropy loss, commonly used in semantic segmentation, examines each pixel separately and compares the class prediction with the target. The equation for the cross-entropy loss is given as follows:

$$L_{CE} = \frac{1}{N} \sum_i - [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3)$$

where L_{CE} is the overall cross-entropy loss, y_i is the label of i, and p_i is the corresponding predicted probability.

Since boundary detection is a category-imbalanced task, dice loss is selected as the auxiliary loss to help boost the performance of parcels extraction. The dice loss is an ensemble similarity measurement used to calculate the similarity of two samples. Dice loss is calculated as follows:

$$L_{dice} = 1 - \frac{2 \left| \hat{Y} \right| \cap |Y|}{\left| \hat{Y} \right| + |Y|} \quad (4)$$

where L_{dice} is the overall dice loss, Y is the label, and $\hat{Y}$ is the corresponding prediction result.

Considering that fuzzy boundaries extraction is a hard sample in model learning, we adopt the strategy proposed in [64], called online hard example mining (OHEM) loss. OHEM loss is a
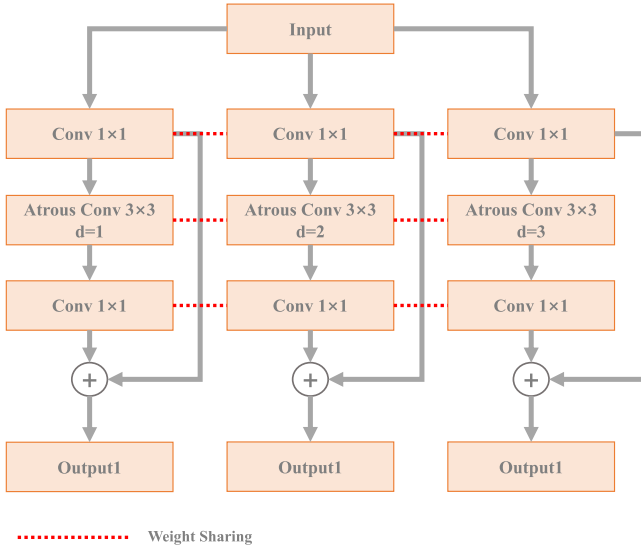
Fig. 6. Details of trident pyramid.

training strategy that focuses on hard samples during the training process and applies higher weights to them. This strategy can enable the model to concentrate more on hard samples, which can be applied to further optimize the two loss functions in this article. Ultimately, our combining loss function for boundary detection branch can be shown as follows:

$$L_B = \lambda \text{OHEM}(L_{\text{CE}}) + (1 - \lambda)\text{OHEM}(L_{\text{Dice}}) \quad (5)$$

where λ is the weight to control the importance between CE loss and Dice loss, and OHEM() represents the OHEM strategy.

For semantic segmentation branch, CE loss is directly employed for network training, and OHEM strategy is also added in the supervision phase to enhance the semantic recognition ability of complex parcels, as shown in (6). Finally, the whole loss for model training is obtained by weighted sum operation, as shown in (7). Since the semantic segmentation branch only plays an auxiliary role in this article, in order to improve the capability of fuzzy boundaries extraction, we reduce the weight of the loss in the semantic segmentation branch

$$L_S = \text{OHEM}(L_{\text{CE}}) \quad (6)$$

$$L = \beta L_S + (1 - \beta) L_B \quad (7)$$

where β is the weight to control the importance between semantic segmentation branch and boundary detection branch.

IV. EXPERIMENTS

A. Implementation Details

We download the google images of Guangdong and Chongqing in urban–rural area with a resolution of 0.5 m and crop them into 512×512 size with a 50% overlap rate to obtain 20 000 patches. And then we manually annotate both clear and fuzzy cropland parcels' boundaries precisely. We divided the training set, validation set, and test set at a ratio of 6:2:2, and carried out subsequent experiments on this dataset.

Our proposed network is implemented by PyTorch. We select the Adam optimizer, and set the initial learning rate to 0.0002.

During the training process, the learning rate is reduced by 5% after 20 epochs. Considering that our own ADA strategy will increase the batch size, we set the input batch size to 4. All experiments were run on an NVIDIA GTX 3080ti GPU. In the following comparison and ablation experiments, the value of β and λ in the loss function is set to 0.3 and 0.6, respectively.

B. Experimental Settings

To evaluate the performance of our proposed method, we compare it with several superior semantic segmentation and boundary detection methods. Then, to demonstrate the effectiveness of our proposed modules, we perform the corresponding ablation experiments to validate the performance of each module and strategy. In order to verify the advantage of our proposed GMT strategy, we need to compare it with several multitask learning structure. Therefore, we use four kinds of structure for this experiment: simple multitask learning, multitask learning with semantic guidance path, multitask learning with boundary guidance path, and the proposed GMT strategy. In the subsequent experiments, four metrics are used to verify the performance of our model. The precision rate and recall rate can reflect the misclassification and omission of the results, respectively, while IoU and F1 score can comprehensively verify the similarity between the predicted results and the ground truth labels. The formulae for these metrics are given as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$IoU = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (11)$$

C. Comparison Experiment

To verify the advantage of our model, we compare it with four state-of-the-art semantic segmentation and boundary extraction methods as follows.

1) D-LinkNet [45] is a boundary extraction method based on UNNE-like architecture for roads detection by using skipping connection, residual block, and dilated convolution.
2) HRNet [48] is a multibranch fusion network, which implements information interaction among multiple resolutions simultaneously through parallel multiresolution subnetworks.
3) BiSeNet [49] is a bidirectional segmentation method, which integrates a spatial path module and a context path module to extract spatial shallow and deep features, respectively.
4) RCF [20] is an edge detection method using rich convolution features, which utilize multilevel representation of objects for accurate edge prediction.

According to the selected model, comparison experiments are conducted on the labeled dataset, and the results are shown in Table I. Our proposed method outperforms the comparison approach in all metrics and has a significant improvement. In

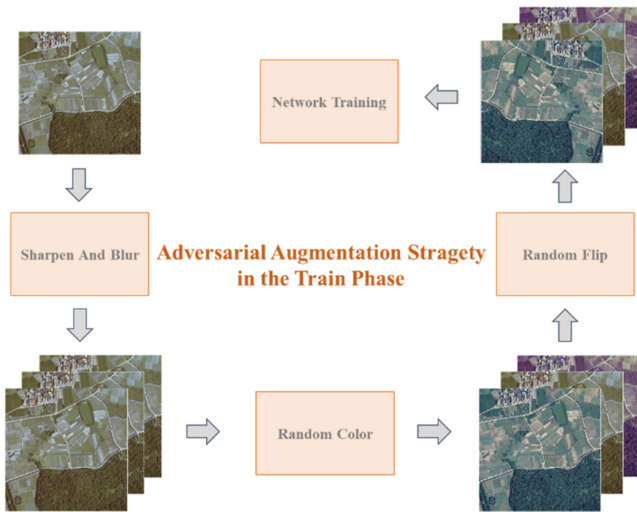Fig. 7.   Visualization of sharpening and blurring.

Fig. 8.   ADA strategy.

TABLE I
QUANTITATIVE RESULTS OF DIFFERENT METHODS

Method	Recall	Precision	F1	IoU
D-linkNet	0.8219	0.8271	0.8245	0.7014
BiSeNet	0.7438	0.8565	0.7962	0.6614
HRNet	0.8498	0.8496	0.8497	0.7388
RCF	0.8631	0.8675	0.8653	0.7626
Proposed	**0.8752**	**0.8798**	**0.8775**	**0.7817**

The bolded values represent the highest values of the evaluation metrics among the methods.

comparison to D-LinkNet and RCF, we achieve an 8% and 2% improvement in IOU, respectively. And it can be concluded that deeper and larger networks can perform better, such as HRNet and RCF. The result in the experiment further illustrates the suitability of our proposed method for cropland parcels extraction.

The visualization results are shown in Fig. 9. It can be seen that the boundary details obtained from our proposed method significantly improve. BiseNet obtain the worst result and cannot

TABLE II
QUANTITATIVE RESULTS OF DIFFERENT MULTITASK STRUCTURES

Method	Multi-task	Semantic Guidance	Boundary Guidance	IoU
Baseline				0.7014
	√			0.6955
	√	√		0.7328
	√		√	0.7213
	√	√	√	**0.7411**

The bolded values represent the highest values of the evaluation metrics among the methods.

extract the fuzzy boundary at all. Although the result from HRNet has a better detail due to its strong ability to capture detailed features, there is still a certain gap with the real labels. The two boundary extraction networks, D-LinkNet and RCF, are better at extracting significant boundary information but have weak ability to extract internal fuzzy-boundary texture. Especially for D-Linknet, the results obtained by this method still remain a great deal of omissions. In contrast, our method is able to capture more fuzzy boundaries' features and produce the best result, which is closer to the ground truth labels. It proves that our proposed method is very suitable for complex parcels boundary extraction.

D. Ablation Study

Ablation experiments are conducted to verify the effectiveness of the proposed method. In the ablation experiments, we name the GMT strategy, MST module, and ADA strategy as GMT, MST, and ADA, respectively. We employ D-LinkNet as our baseline (BL) method and conduct some experiments on it to show the advantages of our proposed method. As a representative method in the field of boundary extraction, comparing with D-LinkNet can fully reflect the effectiveness of our method.

The effectiveness of the GMT is first verified, and the results are shown in Table II. It can be seen that the proposed GMT performs the best, whereas the simple multitask method has the worst performance. It can be explained that, due to the difference
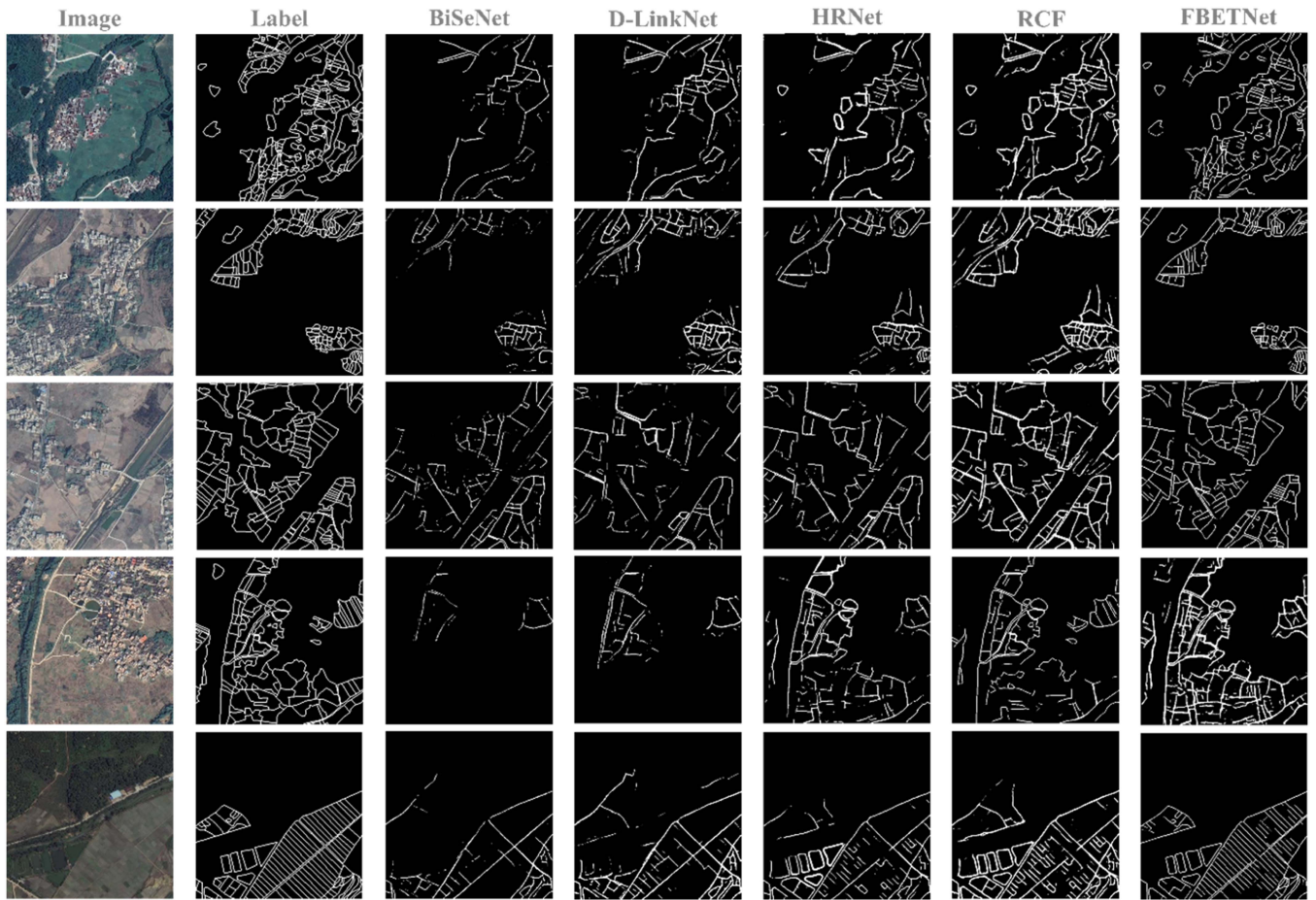
Fig. 9.   Visualization of the results obtained from different methods.

between the semantics and boundary features and the lack of interactive operation between the two branches, the model fails to learn these features effectively. However, after adding the guidance path as the interaction between the two tasks, it can obviously improve the performance of network.

We visualized the result of GMT and the simple multitask learning method, as shown in Fig. 10. It can be seen that our GMT is suitable for the boundaries extraction and can complement some fuzzy boundaries. In contrast, the extracted boundary information is vaguer and the results are even worse than BL when using the simple multitasking strategy. As a result, we conduct the subsequent ablation experiments with the proposed GMT strategy.

The ablation study results of each module and strategy are presented in Table III. As can be seen, our proposed module and strategy in this article results in an improvement in accuracy. GMT obtained the most obvious improvement in accuracy, which indicates that the interaction between semantic and boundary features is very effective in parcels extraction. The MST module and ADA strategy also have a very significant accuracy improvement, verifying that multiscale features extraction and ADA can further help refine the extraction of parcels. Ultimately, the proposed method, which integrates the three modules, obtains a substantial accuracy improvement, further

TABLE III
QUANTITATIVE RESULTS OF ABLATION STUDY

Method	Recall	Precision	F1	IoU
Baseline	0.8219	0.8271	0.8245	0.7014
Baseline+GMT	0.8439	0.8588	0.8513	0.7411
Baseline+MST	0.8392	0.8472	0.8432	0.7289
Baseline+ADA	0.8321	0.8438	0.8379	0.7210
Proposed	**0.8752**	**0.8798**	**0.8775**	**0.7817**

The bolded values represent the highest values of the evaluation metrics among the methods.

proving that these modules can complement each other and facilitate the network to capture more detailed boundary features and gain more refined results.

Fig. 11 shows the results after gradually adding the proposed module. It can be seen that the performance of fuzzy boundaries extraction has a significant improvement by using MTG in BL. With the fusing of MST module, the results are better due to its enhanced ability of multiscale features extraction, which is able to provide more detailed information of parcels. It is obvious that the ADA strategy can obtain a more robust performance, especially for the fuzzy boundaries, making the extraction results
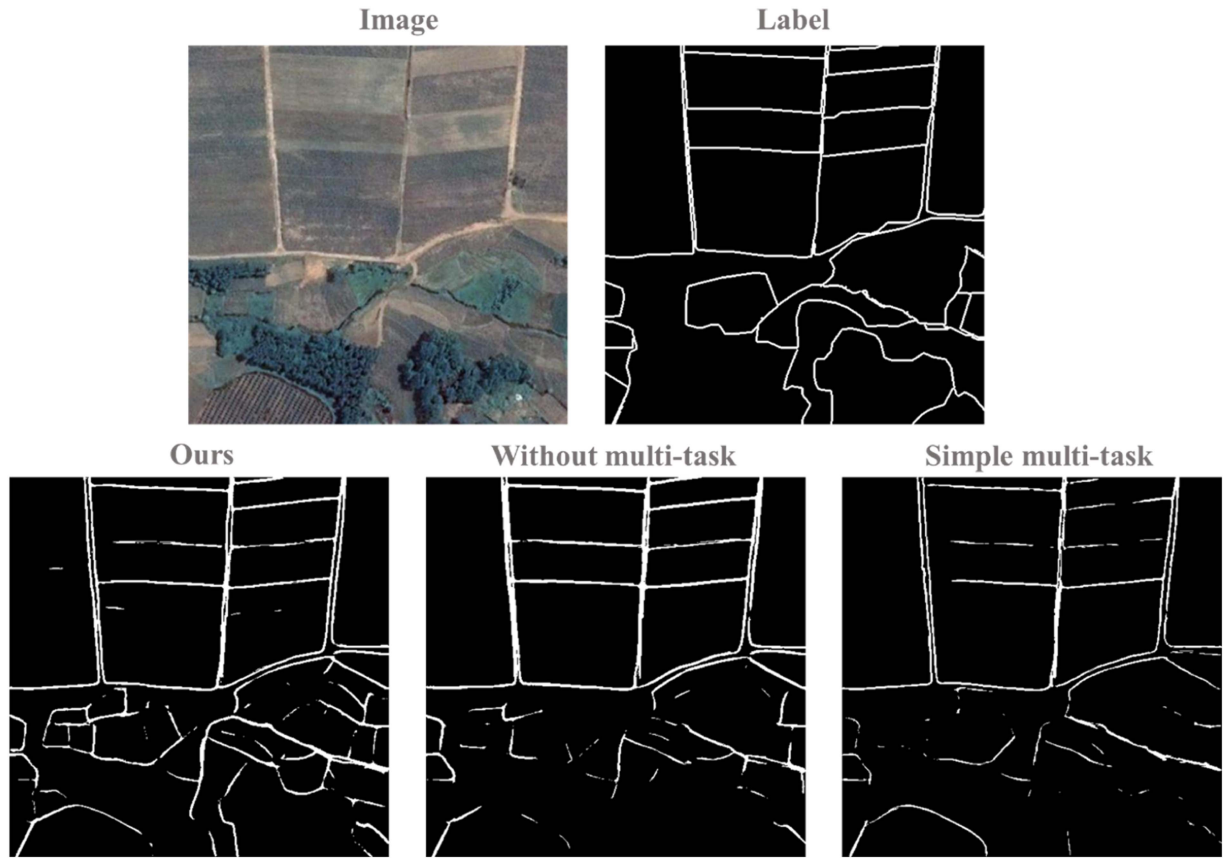
Fig. 10. Visualization of the results with different multitask learning strategy.
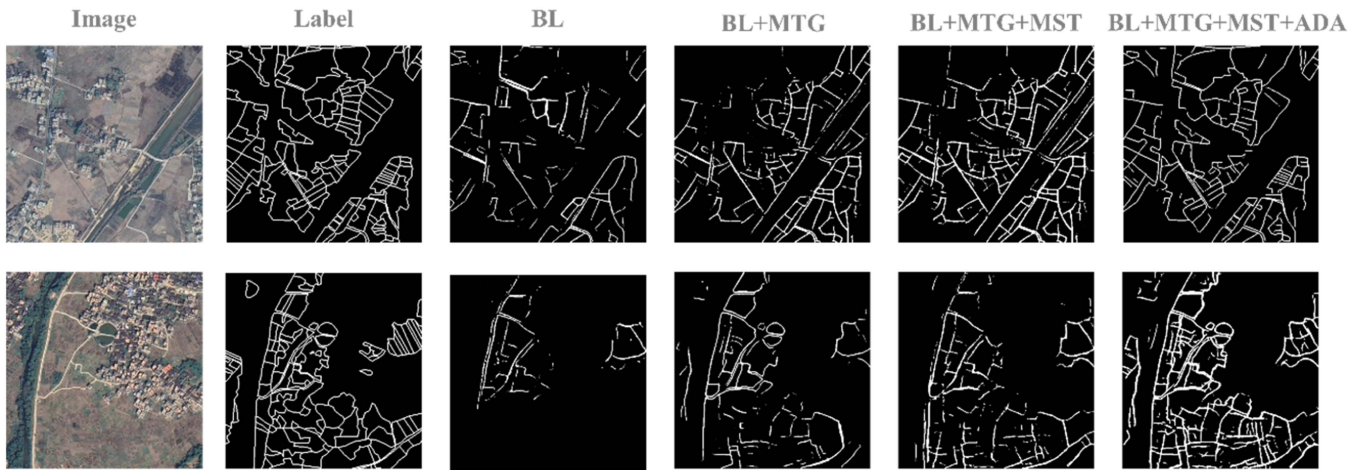

Fig. 11. Visualization of the results with ablation study.

closer to the labels, further verifying the effectiveness of our proposed modules.

V. DISCUSSION

In the discussion part, we discuss the detailed model design of our proposed method and conduct some experiments to analyze the performance difference of different structures.

For the proposed guidance module in the multitask learning strategy, we introduce two guidance paths and conduct the module design experiments on different pooling methods and feature combination methods. We name the semantic guidance path and the boundary guidance path as SGMT and BGMT, respectively.

First, for the pooling layer in the semantic guidance path, we design four models that contain average pooling, max pooling, and pooling after convolution, respectively. The accuracy results
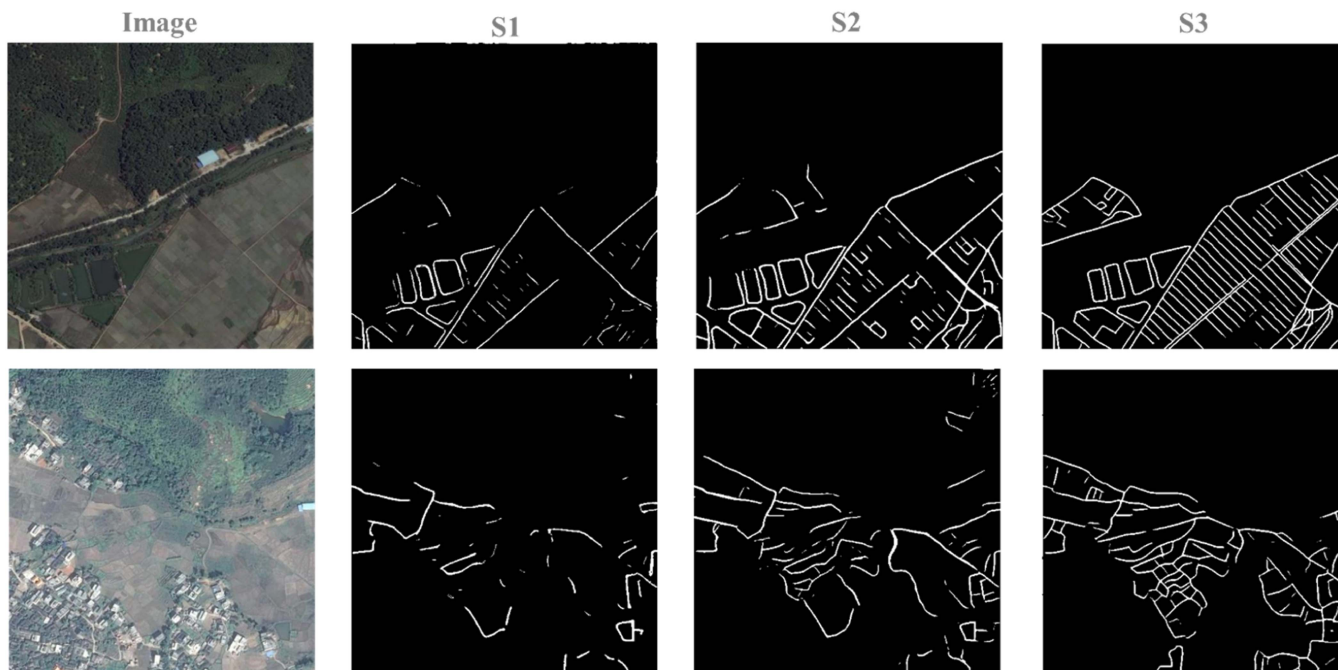
Fig. 12. Visualization of the multiscale results: S1, S2, and S3 represent the coarse, intermediate, and refine results, respectively.

TABLE IV
QUANTITATIVE RESULTS OF DIFFERENT SGMT STRUCTURES

Method	IoU
SGMT + MaxPooling	0.7792
SGMT + AvgPooling	**0.7817**
SGMT +Conv + MaxPooling	0.7707
SGMT +Conv + AvgPooling	0.7723

The bolded value represent the highest values of the evaluation metrics among the methods.

TABLE V
QUANTITATIVE RESULTS OF DIFFERENT BGMT STRUCTURES

Method	IoU
BGMT	0.7808
BGMT + Conv	**0.7817**

The bolded value represent the highest values of the evaluation metrics among the methods.

TABLE VI
QUANTITATIVE RESULTS OF DIFFERENT PYRAMID MODULES

Method	IoU
Baseline	0.7014
Image Pyramid	**0.7301**
Simple Feature Pyramid	0.7119
Top-down Feature Pyramid	0.7152
Trident Pyramid	0.7193
Proposed	<u>0.7289</u>

The bolded value represent the highest value of the evaluation metrics among the methods and the second best value is underlined.

are summarized in Table IV. It can be seen that the accuracy is worse after adding the convolution operation. The result with average pooling obtains better than max pooling, and it can be explained that the max pooling operation will reduce more semantic information while the average pooling can consider the nearby local features that can maintain more useful information.

Second, we verify the effectiveness of convolution operation in the BGMT, and the results are presented in Table V. We can see that the methods after convolution obtain better performance, and it can be explained that the boundary of the semantic segmentation result is blurred and will misguide the boundary detection. Therefore, performing a convolution operation to capture

an auxiliary feature can provide more accurately information of fuzzy parcels' boundaries.

For the multiscale extraction requirements pointed out in this article, we compare the performance of different multiscale modules, as shown in Table VI. The four multiscale strategies compared are illustrated in Fig. 4. The results show that the highest accuracy is obtained by the image pyramid, and our proposed module ranks second. There is only a slight difference in accuracy between the two methods. However, in terms of computational efficiency, our model is obviously faster than the image pyramid (at about 2.4 times faster than image pyramid). Therefore, on the whole, our model is still very suitable for multiscale parcels extraction. In Fig. 12, we visualize the multiscale outputs of parcels. We discover that the outputs at different scales represent different boundary details. With the multiscale outputs, it is feasible to fuse them into a more refined result.

For the ADA strategy, we conduct different design experiments, and the results are shown in Table VII. It can be seen that the ADA obtain the best performance. When using image blurring only, the network performs even worse. However, by
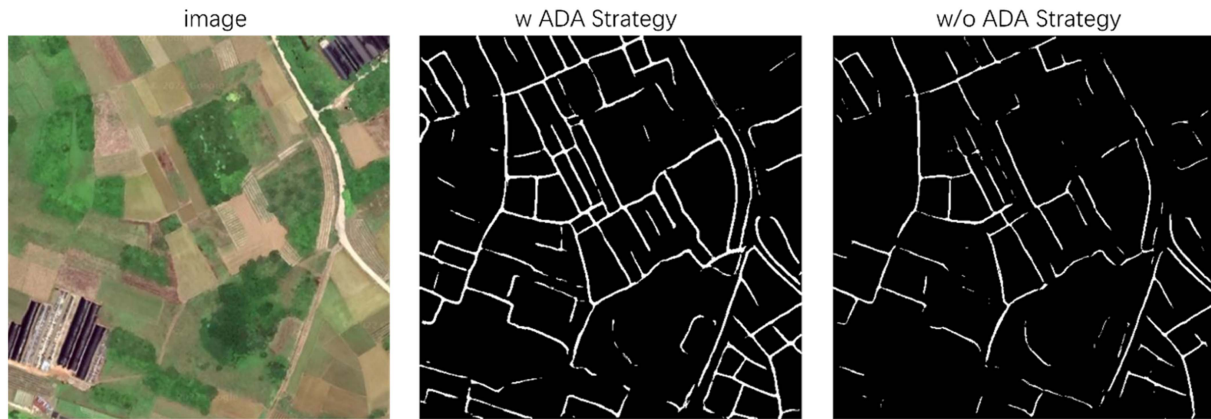
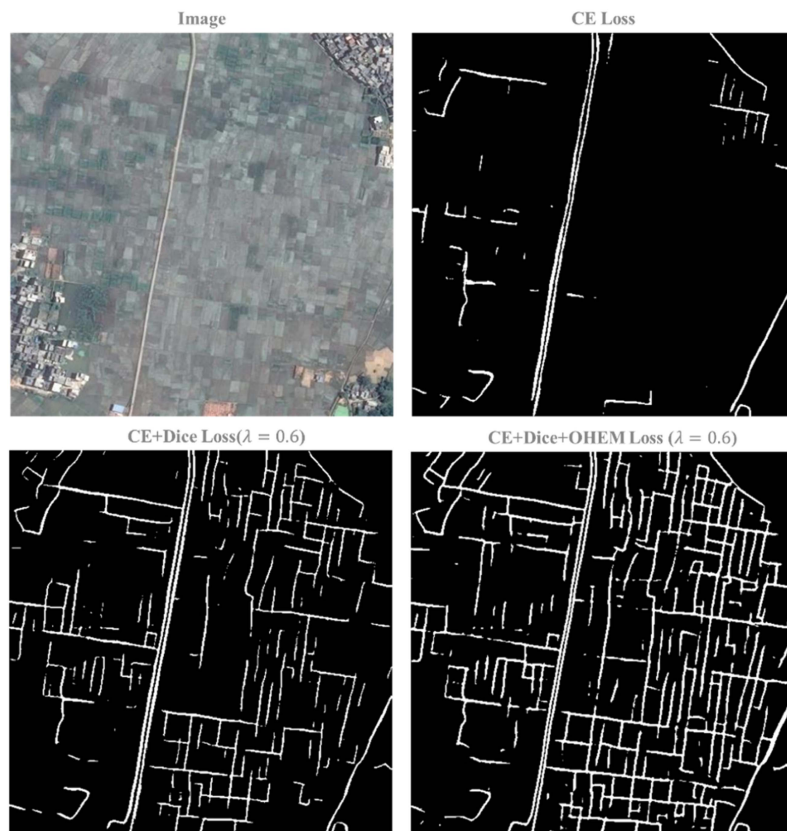Fig. 13. Visualization of the results of ADA strategy.

Fig. 14. Visualization of the results with different loss functions.

combining blurring and sharpening, it can obtain a significant improvement. Therefore, it is further demonstrated that our designed ADA strategy can simultaneously strengthen the recognition of fuzzy boundaries as well as capture the features of clear boundaries, enhancing the robustness of the model. The accuracy is also improved after adding random color transformation and random flip, which illustrates that these data augmentation methods can also help strengthen the performance of network. We visualize the results before and after using our ADA strategy in Fig. 13. It is obvious that the fuzzy boundaries extraction result is much more precise, after using our ADA strategy.

In order to verify the effectiveness of our designed loss function, comparison experiments with different combinations and weights are designed, as shown in Table VIII. It can be seen that incorporating the dice loss can improve significantly, and the optimal weight between dice loss and CE loss is 4:6. The accuracy is further increased after using OHEM strategy, indicating that the hard sample learning of OHEM is very successful in extracting cropland parcels.

TABLE VII
QUANTITATIVE RESULTS OF DIFFERENT AUGMENTATION STRATEGIES

Method	Blur	Sharpen	Random color transform	Random flip	IoU
Baseline					0.7014
	√				0.6907
		√			0.7088
	√	√			0.7165
	√	√	√		0.7197
	√	√	√	√	**0.7210**

The bolded value represent the highest values of the evaluation metrics among the methods.

TABLE VIII
QUANTITATIVE RESULTS OF DIFFERENT LOSS FUNCTIONS

Method	IoU
CE	0.6954
CE+Dice($\lambda = 0.8$)	0.7626
CE+Dice($\lambda = 0.7$)	0.7669
CE+Dice($\lambda = 0.6$)	<u>0.7701</u>
CE+Dice($\lambda = 0.5$)	0.7683
CE+Dice($\lambda = 0.4$)	0.7613
CE+Dice+OHEM($\lambda = 0.6$)	**0.7817**

The bolded value represent the highest value of the evaluation metrics among the methods and the second best value is underlined.

We visualized the results of different loss combinations in Fig. 14. It shows that the network trained by CE loss only fail to detect parcels' boundaries. However, when combining the dice loss and CE loss, it obtain a significant improvement in boundaries extraction. After adopting OHEM strategy, the network can mine more information of fuzzy boundaries and generate more complete and refined parcels' boundaries.

VI. CONCLUSION

In this article, a fuzzy-boundary enhanced trident network, named FBETNet, is proposed to enhance the feature of fuzzy boundary and generate multiscale parcels. We first introduce a GMT learning strategy to fully use the semantic information of parcels' boundaries and then utilize these features to supplement fuzzy boundaries extraction result in the boundary branch. Second, an MST module is designed to generate parcels of different scales by considering the complexity of parcels appearance in urban–rural area. Third, in order to further improve the robustness of model and make it generalized to various scenarios, we employ an ADA strategy in the training phase, which can simulate feature learning process of both the clear and fuzzy boundaries. In experiments, we use our own well-annotated dataset to verify the effectiveness and advantage of our proposed modules. It further proves that our FBETNet can outperform other state-of-the-art methods. Moreover, through the ablation study and model design experiments, it can demonstrate the effectiveness of our proposed strategy and obtaining the best design of our network. The proposed method can not only capture the clear boundaries' features precisely but also enhance the performance of fuzzy boundaries extraction, making it possible

to meet the demand of urban precise management. Although our proposed method employs semantic information and multiscale features to help extract more precise parcels' boundaries, it is still limited by the integrity of semantic information. For those parcels with complex semantic information, our method still hard to detect. Therefore, the subsequent work will continue to explore the weak semantic information for parcels and further enhance the performance of parcels extraction so as to identify more accurate and refined parcels and make better practical applications.

REFERENCES

[1] D. Domingo, G. Palka, and A. M. Hersperger, "Effect of zoning plans on urban land-use change: A multi-scenario simulation for supporting sustainable urban growth," *Sustain. Cities Soc.*, vol. 69, 2021, Art. no. 102833.

[2] L. Lin, H. Fan, and Y. Jin, "Multi-scale and multi-model simulation of land use/land cover change in the mountainous county: A case study of Mengla County in Yunnan Province, China," *Mountain Res.*, vol. 38, no. 4, pp. 630–642, 2020.

[3] H. Liu, R. Homma, Q. Liu, and C. Fang, "Multi-scenario prediction of intra-urban land use change using a cellular automata-random forest model," *ISPRS Int. J. Geo-Inf.*, vol. 10, 2021, Art. no. 503.

[4] H. E. Haishan, Y. Zhao, and W. U. Jiansheng, "Simulation of urban landscape pattern under the influence of low carbon: A case study of Shenzhen," *Acta Ecol. Sinica*, vol. 41, pp. 8352–8363, 2021.

[5] W. Liu et al., "Farmland parcel mapping in mountain areas using time-series SAR data and VHR optical images," *Remote Sens.*, vol. 12, 2020, Art. no. 3733.

[6] H. Zhang, Y. Liao, H. Yang, G. Yang, and L. Zhang, "A local-global dual-stream network for building extraction from very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1269–1283, Mar. 2022.

[7] P. Liu et al., "Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network," *Remote Sens.*, vol. 11, 2019, Art. no. 830.

[8] V. S. F. Garnot and L. Landrieu, "Panoptic segmentation of satellite image time series with convolutional temporal attention networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4852–4861.

[9] A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2514–2520, Nov. 2001.

[10] Z. Jin-Yu, C. Yan, and H. Xian-Xiang, "Edge detection of images based on improved Sobel operator and genetic algorithms," in *Proc. Int. Conf. Image Anal. Signal Process.*, 2009, pp. 31–35.

[11] P. A. Mlsna and J. J. Rodríguez, "Gradient and Laplacian edge detection," in *Handbook of Image and Video Processing*, 2nd ed., B. Al, Ed. Burlington, ON, Canada: Academic Press, 2005, pp. 535–553.

[12] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[13] M. Torre and P. Radeva, "Agricultural field extraction from aerial images using a region competition algorithm," *Int. Arch. Photogramm. Remote Sens.*, vol. 33, pp. 889–896, Jan. 2000.

[14] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, 2018.

[15] F. Xu, C. Hu, J. Li, A. Plaza, and M. Datcu, "Special focus on deep learning in remote sensing image processing," *Sci. China Inf. Sci.*, vol. 63, 2020, Art. no. 140300.

[16] H. Liu et al., "Contour-oriented cropland extraction from high resolution remote sensing imagery using richer convolution features network," in *Proc. 8th Int. Conf. Agro-Geoinform.*, 2019, pp. 1–6.

[17] Z. Xiao, Y. Long, D. Li, C. Wei, G. Tang, and J. Liu, "High-resolution remote sensing image retrieval based on CNNs from a dimensional perspective," *Remote Sens.*, vol. 9, 2017, Art. no. 725.

[18] L. Xia, J. Luo, Y. Sun, and H. Yang, "Deep extraction of cropland parcels from very high-resolution remotely sensed imagery," in *Proc. 7th Int. Conf. Agro-Geoinform.*, 2018, pp. 1–5.

[19] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, vol. 125, pp. 3–18.

[20] Y. Liu et al., "Richer convolutional features for edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1939–1946, Aug. 2019.

[21] A. Garcia-Pedrero, M. Lillo-Saavedra, D. Rodriguez-Esparragon, and C. Gonzalo-Martin, "Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system," *IEEE Access*, vol. 7, pp. 158223–158236, 2019.

[22] K. M. Masoud, C. Persello, and V. A. Tolpekin, "Delineation of agricultural field boundaries from Sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 59.

[23] L. Yan and D. P. Roy, "Automated crop field extraction from multi-temporal web enabled Landsat data," *Remote Sens. Environ.*, vol. 144, pp. 42–64, Mar. 2014.

[24] T. Cheng et al., "DESTIN: A new method for delineating the boundaries of crop fields by fusing spatial and temporal information from WorldView and Planet satellite imagery," *Comput. Electron. Agriculture*, vol. 178, 2020, Art. no. 105787.

[25] C. Rieke, "Deep learning for instance segmentation of agricultural fields," 2017. [Online]. Available: https://github.com/chrieke/InstanceSegmentation_Sentinel2

[26] A. Potlapally, P. S. R. Chowdary, S. S. Raja Shekhar, S. N. Mishra, C. S. V. D. Madhuri, and A. V. V. Prasad, "Instance segmentation in remote sensing imagery using deep convolutional neural networks," in *Proc. Int. Conf. Contemporary Comput. Inform.*, 2019, pp. 117–120.

[27] W. Mei et al., "Using deep learning and very-high-resolution imagery to map smallholder field boundaries," *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 3046.

[28] C. Li et al., "Attention enhanced U-Net for building extraction from farmland based on Google and WorldView-2 remote sensing images," *Remote Sens.*, vol. 13, 2021, Art. no. 4411.

[29] J. Xu et al., "DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111946.

[30] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[31] 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. 6230–6239.

[32] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 3–22, 2018.

[33] U. Michieli, M. Biasetton, G. Agresti, and P. Zanuttigh, "Adversarial learning and self-teaching techniques for domain adaptation in semantic segmentation," *IEEE Trans. Intell. Veh.*, vol. 5, no. 3, pp. 508–518, Sep. 2020.

[34] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 50–65, Mar. 2021.

[35] X. Li, X. Yao, and Y. Fang, "Building-A-Nets: Robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3680–3687, Oct. 2018.

[36] J. Li et al., "Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 373–389, 2020.

[37] W. Chen, G. Zhou, Z. Liu, X. Li, X. Zheng, and L. Wang, "NIGAN: A framework for mountain road extraction integrating remote sensing road-scene neighborhood probability enhancements and improved conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Jan. 2022.

[38] M. Ju and Q. Hu, "Vision-inspired filtering algorithm for SAR ship detection based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Jan. 2022.

[39] M. Jong, K. Guan, S. Wang, Y. Huang, and B. Peng, "Improving field boundary delineation in ResUNets via adversarial deep learning," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102877.

[40] C. Persello, V. A. Tolpekin, J. R. Bergado, and R. A. de By, "Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111253.

[41] Y. Zhang and Q. Yang, "An overview of multi-task learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 30–43, 2018.

[42] J. Long, M. Li, X. Wang, and A. Stein, "Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102871.

[43] F. Waldner and F. I. Diakogiannis, "Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network," *Remote Sens. Environ.*, vol. 245, 2020, Art. no. 111741.

[44] A. Sharifi, H. Mahdipour, E. Moradi, and A. Tariq, "Agricultural field extraction with deep learning algorithm and satellite imagery," *J. Indian Soc. Remote Sens.*, vol. 50, pp. 417–423, Feb. 2022.

[45] 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/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 192–1924.

[46] Y. Li, Y. Chen, N. Wang, and Z.-X. Zhang, "Scale-aware trident networks for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6053–6062.

[47] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[48] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5686–5696.

[49] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet V2: Bilateral network with guided aggregation for real-time semantic segmentation," *Int. J. Comput. Vis.*, vol. 129, pp. 3051–3068, 2021.

Jiajun Zhu received the B.S. degree in geographic information science in 2021 from the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China, where he is currently working toward the M.S. degree in cartography and geographic information system.

His research interests include remote sensing image processing, deep learning, and boundary detection.

Zhuoqun Chai is currently working toward the B.S. degree in geographic information science with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China.

His research interests include urban spatial analysis and image processing with deep learning.

Zhaopu Song received the B.S. degree in cartography and geographic information systems from South China Agricultural University, Guangzhou, China, in 2012.

He is working with Lands and Resource Department of Guangdong Province, Surveying and Mapping Institute, Guangzhou, China. His research interests include interpretation of remote sensing images and surface classification.

Zhanpeng Chen received the B.S. degree in Information engineering from Jinan University in 2011. He is currently working with Lands and Resource Department of Guangdong Province, Surveying and Mapping Institute, Guangzhou, China. His research interests include interpretation of remote sensing images and geographic information system.

Qian Shi (Senior Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2015.

She is currently an Associate Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote sensing image classification, including deep learning, active learning, and transfer learning.