A Deep Learning Method for Cultivated Land Parcels' (CLPs) Delineation From High-Resolution Remote Sensing Images With High-Generalization Capability

Yu Zhu, Yaozhong Pan, Dujuan Zhang, Hanyi Wu, *Graduate Student Member, IEEE*, and Chuanwu Zhao, *Graduate Student Member, IEEE*

***Abstract*— Accurate cultivated land parcels' (CLPs) information is essential for precision agriculture. Deep learning methods have shown great potential in CLPs' delineation but face challenges in detection accuracy, generalization capability, and parcel optimization quality. This study addresses these challenges by developing a high-generalization multitask detection network coupled with a specialized parcel optimization step. Our detection network integrates boundary and region tasks and designs distinct decoders for each task, employing performance-enhancing modules as well as more balanced training strategies to achieve both accurate semantic recognition and fine-grained boundary depiction. To improve the network's ability to train more generalized models, our study identifies the variations in image hue, landscape surroundings, and boundary granularity as the key factors contributing to generalization degradation and employs color space augmentation (CSA) and attention mechanisms on spatial and hierarchy to enhance the generalization. In addition, the parcel optimization step repairs long-distance boundary breaks and performs object-level fusion of delineated regions and boundaries, resulting in more independent and regular CLP results. Our method was trained and validated on GaoFen-1 images from four diverse regions in China, demonstrating high delineation accuracy. It also maintained stable spatiotemporal generalization across different times and regions. Comprehensive ablation and comparative experiments confirmed the rationale behind our model improvements and demonstrated our method's effectiveness over existing single-task models (SegNet, modified PSPNet (MPSPNet), DeeplabV3+, U-Net, ResU-Net, and R2U-Net) and recent multitask models (ResUNet-a, BSiNet, and SEANet). The implementation of our method is available at https://github.com/BNU-zhu/CLPs-delineation.**

***Index Terms*— Agricultural remote sensing, cultivated land parcels (CLPs), deep learning, GaoFen-1 image, semantic segmentation.**

Manuscript received 5 June 2024; accepted 5 July 2024. Date of publication 9 July 2024; date of current version 19 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42192580 and Grant 42192581 and in part by the National High Resolution Earth Observation System (the Civil Part) Technology Projects of China under Grant 20-Y30F10-9001-20/22. *(Corresponding author: Yaozhong Pan.)*

Yu Zhu, Yaozhong Pan, Hanyi Wu, and Chuanwu Zhao are with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China (e-mail: pyz@bnu.edu.cn).

Dujuan Zhang is with the National Supercomputing Center in Zhengzhou, Zhengzhou University, Zhengzhou 450001, China.

Digital Object Identifier 10.1109/TGRS.2024.3425673

I. Introduction

CULTIVATED land parcels (CLPs)—also known as crop plots or individual arable fields—are the essential unit of agricultural statistics that are demarcated by ridges, paths, and ditches, and are usually planted with one crop per growing season [1], [2]. CLPs' delineation is not only the identification of cultivated land areas but also the further subdivision of internal cropland units. Accurate delineation of CLPs is crucial for parcel-level assessments, providing valuable information on crop types, growth stages, and phenology for precision agriculture. In addition, these assessment outcomes offer precise data for agricultural insurance and policy development. In recent years, parcel-based analysis has expanded into various remote sensing applications, including crop-type classification [3], [4], [5], [6], [7], phenology estimation [8], and gap filling for time-series images [8]. These methods often outperform pixel-based analysis by incorporating parcel boundaries' constraints. Nonetheless, traditional CLP delineation methods typically rely on time-consuming and labor-intensive manual or on-site techniques. Fortunately, with advancements in image resolution and data accessibility, remote sensing-based delineation methods are increasingly becoming efficient and feasible alternatives, enabling the detailed delineation of large areas efficiently.

Current methods for delineating CLPs from remote sensing images can be divided into handcrafted feature-based and deep learning-based techniques based on the way of feature learning [9]. Among these, deep learning-based methods are generally preferred due to their superior feature description capabilities and more accurate parcel detection [1], [10], [11], [12], [13]. However, accurately delineating CLPs remains challenging, as it requires not only identifying cropland regions but also maintaining complete and accurate internal boundaries for subdivision. Although many methods exist for cropland region identification [14], [15], [16], they primarily focus on distinguishing cropland from noncropland using high-level features from multilayer convolutions. This process often results in significant loss of internal boundaries, making fine-grained delineation of internal parcels challenging. Therefore, most current methods target field boundaries to detect and

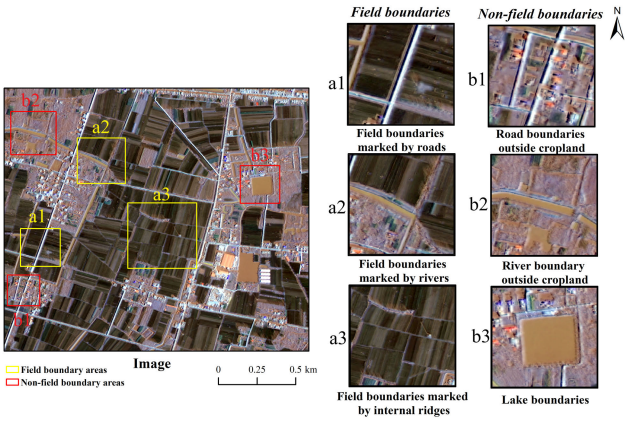


Fig. 1. Presentation of some types of field boundaries on localized images. (a1)–(a3) Yellow boxes show three different compositions of field boundaries. (b1)–(b3) Red boxes show three nonfield boundaries.

generate parcels [1], [12], [17], [18], [19], [20], [21]. However, field boundaries exhibit variable characteristics, as shown in Fig. 1, and they have complex compositions, which can be marked by roads, rivers, and fine ridges. When these features extend beyond croplands, they transform into nonfield boundaries, complicating model's learning processes and making direct semantic detection challenging. In response to this duality challenge, recent research has explored dual-branch network structures that concurrently predict cropland regions and boundaries [22], [23], [24], [25], [26], outperforming single-branch networks. While cropland region prediction emphasizes semantic discernment, boundary detection necessitates meticulous detail preservation. Current architectures struggle to optimize both aspects, and even some studies prioritize only region prediction [24], [25]. Yet, both are critical: the region predictions suppress noncropland boundary information, enhancing the semantic accuracy of boundary detection, while boundary predictions refine the granularity of region results for precise internal subdivision. Consequently, it is necessary to enhance the model's detection capabilities to maximize the quality of both predictions and fuse for more accurate CLPs.

Moreover, the model's generalization performance is also a significant concern. Robust generalization ability is essential for effective performance within the training dataset and also supports stable spatiotemporal generalization when applied to other nontraining regions. Some studies attempt to improve model spatial generalization by expanding the number of training areas [22]. Others achieve stable temporal generalization by utilizing multitemporal imagery for training and prediction [27], [28]. However, both multiregional and multitemporal high-resolution imagery and samples are scarce. This is primarily due to the narrow width of field boundaries, which imposes significant precision requirements on the samples. Obtaining high-quality samples necessitates manual delineation, which is both time-consuming and labor-intensive, especially in China where small-sized parcels are predominant. Therefore, it is necessary to enhance generalization from the perspective of network architecture, improving network's ability to train more generalizable CLP recognition features from limited samples.

Several studies have also extended the model's transfer capabilities through additional transfer learning. For instance, Wang et al. [29] retrained a pretrained model with a subset of samples from the target transfer area. This process updated the model to achieve improved performance in the transfer region. Similarly, Kerner et al. [27] enhanced transfer performance by freezing the shallow layers of the model and fine-tuning the deeper layers using samples from the transfer areas. In addition, Liu et al. [26] applied a fine-grained domain adaptation (FADA) module to align the features of the target domain with the source domain, enabling the trained model to be applicable in the transfer region. However, these methods of achieving model transfer rely on optimizing the model input images or updating the model weights for specific transfer scenarios. They do not alter the original network architecture and enhance the network's ability to train more generalized models. Moreover, their dependence on transfer area information for retraining also reduces their practical feasibility. Nevertheless, in specific transfer scenarios, these techniques remain an excellent way to further optimize the trained model's transfer performance and are complementary to the generalization-improved network architecture.

Another overlooked issue in existing research is the direct output of field boundary detection results. However, field boundaries in imagery often encounter challenges such as blurriness or occlusion. Despite the powerful detection performance of deep learning methods, they cannot guarantee that the predicted boundary results are entirely continuous. The integrity of boundaries is crucial because incomplete boundaries can lead to parcel undersegmentation, severely affecting subsequent parcel-based applications. Currently, few studies have designed additional optimization steps, and those that have attempted to address this issue typically employ morphological dilation operations to connect boundaries [26], [30], which can only rectify minor fractures. Therefore, a more effective optimization step for parcel results is necessary.

Based on the considerations mentioned above, we propose a novel CLPs' delineation method in this study. The method comprises a multitask detection network and a parcel optimization step, with improvements made to enhance the network's generalization capabilities. Overall, this work's contributions are threefold.

1) We analyze the duality of CLPs' detection and develop a multitask network that integrates boundary and region tasks. This network designs distinct decoder for each task to accommodate their different characteristics and employs some performance-enhancing modules as well as more balanced training strategies. It makes a balance between accurate semantic recognition and fine-grained boundary depiction, addressing the challenge of current research in predicting high-quality region and boundary results simultaneously.
2) We analyze the factors influencing model generalization. By employing hue transformation and attention mechanisms on spatial and hierarchy, we provide a foundational network architecture capable of training more generalized models across limited samples. Compared to existing networks, our trained models exhibit superior

spatiotemporal generalization performance, and they maintain this superior performance even after applying the same transfer learning techniques for optimization.

3) We propose a novel method to further optimize parcels, where boundary breaks are oriented repaired through breakpoint and extension direction detection, and where region and boundary results are fused through an object-based approach. Compared to existing studies, our optimization method can repair larger distance breaks and fuse to generate more regular parcels.

The remainder of this article is organized as follows. Section II presents some related work about the CLPs' delineation. Section III describes the proposed delineation method. Section IV introduces how to organize the experiment. Section V shows the performance of the method in experimental areas and the result of ablation and comparison experiments. Section VI provides some further insights and discussion on the CLPs' delineation. Finally, the conclusion and contributions are summarized in Section VII.

II. Related Work

This section provides a concise overview of the existing research on CLPs' delineation. In general, the delineation methods can be categorized into two primary categories based on the way of feature learning: handcrafted feature-based (Section II-A) and deep learning-based methods (Section II-B). Within each detection approach, three distinct detection objectives emerge: region-based methods, which target parcel regions; edge-based methods, which concentrate on parcel boundaries; and hybrid methods, which aim to extract both regions and boundaries. Furthermore, in Sections II-C and II-D, we provide a detailed introduction of existing parcel optimization methods and the research on further optimizing model using transfer learning techniques.

A. Handcrafted Feature-Based Methods

1) Region-Based Methods: Region-based methods utilize conventional image segmentation techniques, such as watershed, multiresolution, and mean-shift segmentation, which directly segment the images based on spectral homogeneity criteria, thereby yielding object-level CLPs' results. Nevertheless, the complexity of the imagery often leads to oversegmentation of resultant regions within high internal variation fields and undersegmentation between small adjacent fields [31]. In addition, the selection of segmentation parameters heavily relies on experience. Consequently, it is challenging to obtain accurate parcel objects directly through region-based methods.

2) Edge-Based Methods: Edge-based methods, on the other hand, focus on the detection of field boundaries by emphasizing the discontinuity (gradient) between pixels using edge operators. Operators, such as Sobel, Canny, and Scharr, have been employed for field boundary detection [32], [33], [34]. Nonetheless, due to the boundary detection operators solely focusing on gradient information without possessing the ability to semantically recognize field boundaries, the resulting boundaries may include other irrelevant boundaries. Furthermore, edge-based methods still encounter a considerable amount of boundary broken and noise information, making it difficult to transform nonclosed boundaries into independent and complete CLPs.

3) Hybrid Methods: There have also been studies attempting to combine the two methods, wherein region segmentation is performed on the detected gradient layer to obtain object-level results, thus avoiding the issue of nonclosed boundaries faced by singular boundary detection. Mueller et al. [33] initially applied this approach for field delineation, achieving improved results compared to direct segmentation by utilizing boundaries to guide the region segmentation. Watkins and van Niekerk [35] conducted comparative experiments on different combinations of boundary detection and region segmentation methods and found that the Canny combined with the watershed segmentation method (CEWS) demonstrated the best extraction performance, which has been applied in various parcel-based studies [3], [34], [36].

However, overall, the detection performance of edge operators and region segmentation methods is still limited as they rely solely on the shallowest features within the imagery. Real-world CLPs exhibit significant complexity in terms of crop types, terrain, cultivation practices, phenology, and imaging conditions. Therefore, this gap in feature description capability makes it challenging for handcrafted feature-based methods to accurately and extensively delineate CLPs.

B. Deep Learning-Based Methods

Due to the limited performance of handcrafted feature-based methods, more research has attempted deep learning-based methods to achieve accurate CLPs' identification by utilizing its powerful feature description capability.

1) Region-Based Methods: Early studies focused on utilizing deep learning models to detect the region of agricultural land distribution. For instance, Zhang et al. [14] employed a modified PSPNet (MPSPNet) model to map high-resolution croplands in four provinces in China. The added modules in the model significantly improved the accuracy of region identification, resulting in overall accuracies (OAs) exceeding 90% for all four provinces. In addition, architectures, such as DeeplabV3+ and UNet, have also been employed for cropland region recognition [16], [17].

However, the task of cropland identification primarily concerns distinguishing between cropland and noncropland categories, which corresponds to the high-level semantic features with low resolution in model. Nonetheless, the field boundaries often occupy only a few pixels in width. In the model's downsampling process, there is a substantial loss of boundary information. Consequently, these methods can solely determine the general region of croplands, failing to accurately represent the precise delineation of internal individual parcels.

2) Edge-Based Methods: Considering the significance of detailed boundary information in CLPs' delineation, an increasing number of studies have conducted delineation from the perspective of detecting field boundaries. These methods utilize deep learning models to learn the image features

of field boundaries for prediction. It is worth noting that the aforementioned edge detection operators can be viewed as models with only one convolutional layer. Therefore, CNN-based edge detection models with multiple convolutional layers, such as HED [37] and RCF [38], can provide more accurate boundary results. For instance, Marvaniya et al. [39] utilized the HED model to identify boundary information and obtained CLPs' results through a series of postprocessing steps. However, the field boundaries are semantic boundaries. Although these edge models possess excellent boundary awareness and can detect more potential field boundaries, they struggle to differentiate field boundaries from other types of boundaries (e.g., boundaries within residential areas) due to a lack of category judgment [11]. Incorrect boundary information can lead to the generation of incorrect parcels.

Semantic segmentation models, however, can treat field boundaries as a feature type and link each image pixel to the field boundary/nonfield boundary label, thus detecting them semantically. These models have been widely adopted for field boundary detection. FCN [10], [20], Segnet [1], U-Net [12], [19], [40], ResU-Net [17], ResUNet-a [22], [41], and R2U-Net [18] are extensively employed architectures. Among these, U-Net-based architectures exhibit the highest performance, as they incorporate high-resolution shallow features through skip connections after each decoding step, thereby mitigating the issue of boundary information loss in high-level features. Nevertheless, field boundaries are not inherent land cover types, and their complex composition poses challenges for models to learn their precise category features, consequently limiting the detection performance.

3) Hybrid Methods: The results of a single field boundary detection task can be mixed with other boundaries. To mask these irrelevant boundaries, some research has attempted to fuse boundary results with cropland region. Xia et al. [42] first detected boundary information based on the RCF model and then masked the irrelevant boundaries with the cropland region recognized by the U-Net model. Xu et al. [11], on the other hand, reversed this process by initially identifying the cropland region based on the U-Net model and then detecting the boundaries within the identified regions. Nevertheless, the two separate detection processes may result in inconsistent information correspondence.

Taravat et al. [17] have incorporated these two detection tasks into a multiclass network, which classified sample categories into three classes: cropland region, field boundary, and others, thus simultaneously predicting cropland region and boundary results, achieving good recognition results. However, such a multiclass network needs to consider both the "cropland recognition" and "field boundary detection" features at the same time, which limits the performance of the network.

Therefore, Waldner and Diakogiannis [22] have employed the multitask head model to predict the cropland region and boundary through two distinct detection heads within the same network. They proposed a multitask ResU-Net-a model, which utilized multitask heads to predict the cropland regions and boundaries within a modified U-Net model, achieving remarkable recognition accuracy. This method is regarded as a state-of-the-art technique for CLPs' delineation [26].

However, this multitask head model used the same model architecture for both tasks, which did not consider the differences between the cropland region and boundary during feature processing. In fact, image features of the region and boundary differ drastically. Also, in a model, there is a potential contradiction between the abstract semantic features required for cropland classification and the precise location information required for boundary detection, making ResU-Net-a model challenging to maintain strong performance in both tasks.

Recent studies have explored to design independent decoding components for multitasks, prioritizing the region task and utilizing the related signals from boundary task to provide boundary constraints, thereby facilitating parcel-level region identification. For instance, BSiNet employed separate decoding convolutions for boundary and region tasks, with the region task being responsible for delivering parcel-level outcomes [24]. The SEANet model advances this approach and provides the model with stronger boundary perception, thus enabling the region task to output higher quality parcel-level results [25].

However, while the region output benefits from boundary-related signals, the cropland region task, focusing on category differentiation, struggles to retain sufficient detail, resulting in region outcomes that lack the detail needed for internal delineation, especially neglecting fine field ridges within cropland. Boundary task results, which have higher predictive granularity, are necessary to supplement this. In addition, although these networks use independent decoding components for multitask output, their identical structure struggles to meet the differing semantic and detail requirements of the region and boundary tasks, making it difficult to balance both tasks effectively.

Therefore, in the study of parcel delineation, accurate cropland identification and detailed internal segmentation are equally important. This necessitates the equitable output of boundary and region results and their integration at the result level. However, existing networks have not yet fully exploited the potential of these two tasks, unable to simultaneously produce two high-quality outcomes: complete and accurate field boundaries, and well-shaped, categorically precise cropland regions. Consequently, there is a need to further enhance the network's detection performance, achieving both accurate semantic recognition and fine-grained boundary depiction.

C. Parcel Optimization

While deep convolutional networks can capture long-range dependencies to aid the comprehensive assessment of challenging boundaries, they cannot guarantee that the final predicted boundaries are entirely connected. However, the continuity of boundaries holds particular significance in CLPs. Therefore, the inclusion of additional parcel optimization procedures becomes indispensable to ensure boundary continuity and parcel independence.

Early parcel optimization methods have primarily focused on simplifying line segments. For instance, Turker and Kok [32] applied the Gestalt rule to group line segments and eliminate pseudo line segments. Similarly, Hong et al. [43]

employed the Suzuki85 algorithm for simplification. However, these approaches lack active connectivity for fragmented boundaries. Cheng et al. [30] proposed an improvement by employing morphological dilation operations to connect minor boundary discontinuities. Nonetheless, applying morphological dilation directly on a boundary may cause its width to increase and lead to attachment with adjacent field boundaries. Liu et al. [26] introduced the connecting boundaries and filling field (CB-FF) optimization method, where binary boundaries were first skeletonized and, then, dilation was applied to the one-pixel-wide skeleton lines to mitigate the aforementioned issues. However, due to the limited size of morphological structure operators, these methods only achieve minor repairs of boundary breaks.

D. Additional Transfer Learning

Several studies have employed some additional transfer learning strategies such as domain adaptation and model fine-tuning to adjust model's input images and weights, thereby ensuring good generalization in the transfer areas. These methods do not involve changes to the network architecture and can be considered additional transfer optimization strategies.

Wang et al. [29] utilized the FracTAL-ResUNet network architecture, obtained a pretrained model on the training dataset, and then directly retrained this model using the labeled dataset from the transfer area, updating the model weights to extend its performance to the transfer region. Kerner et al. [27], on the other hand, based on the pretrained model obtained from the spatiotemporal U-net (ST-U-net) architecture, froze the shallow layers and further trained the network's deeper layers using the labeled dataset from the transfer area. This fine-tuning approach leverages the original shallow features, updating only the deep layer weights related to the output, which accelerates the retraining efficiency and achieves similarly good performance transfer. This technique is not only applicable in the field of CLP detection but also widely used in other areas such as crop classification to help the model quickly update with the transfer area samples [44], [45].

Some studies have also tried using domain adaptation techniques to align the features of target domain with source domain. For instance, Liu et al. [26], based on a pretrained model trained from a parallel network of U-Net and DeeplabV3+, adopted the FADA technique for feature alignment. FADA simultaneously trains a feature extractor and a discriminator: the feature extractor extracts features from both the source and target domains, which are then evaluated by the discriminator. The training aims to make the features extracted by the feature extractor indistinguishable by the discriminator as originating from either the source or the target domain. Based on the trained feature extractor, the image features of the target domain can be adapted to the source domain, enabling the trained model to be applicable in the transfer region. This method avoids the need for samples from transfer areas. However, due to the lack of real samples, the gain for model transfer is limited, especially when the transfer region has high landscape heterogeneity compared to the source domain.

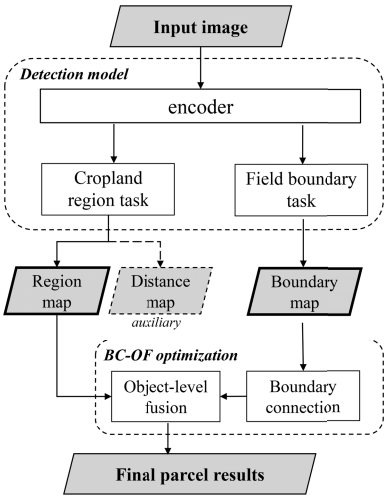


Fig. 2. Workflow of our method for delineating CLPs, which consists of a detection network and an optimization step.

In addition, for each area, retraining the feature extractor and discriminator for image adaptation is required, so the transfer efficiency is low.

Although these techniques have many limitations, they possess the potential to further enhance model transfer performance in specific transfer scenarios. Therefore, it is necessary to assess the need to apply these techniques additionally in different transfer scenarios.

III. Proposed Method

Fig. 2 illustrates the workflow of our method for delineating CLPs, which consists of a detection network and an optimization step. Initially, we establish a multitask detection network that encompasses cropland region and field boundary tasks. The region task predicts the coverage region of CLPs and additional distance maps to provide auxiliary constraints. The boundary task, on the other hand, offers a more refined prediction of the division of field boundaries. A distinct decoder was designed for each task to accommodate their different characteristics. Our network is designed on the image feature of CLPs, maximizing both detection and generalization performance. In addition, we implement an optimization step to enhance the boundary connectivity and integrate boundary and region results at the object level, resulting in individual and regular CLP final results.

A. Feature Analysis of Detection Model

1) Model Detection Performance: The complexity feature of field boundaries hinders direct detection. However, as shown in Fig. 3, we can conceptually define field boundaries as "boundary information within the cropland region." This decomposition divides the detection task into cropland identification and boundary detection tasks, avoiding the need for the model to learn the intricate features of field boundaries. These two tasks are mutually beneficial: cropland regions help constrain the presence of field boundaries, reducing irrelevant boundaries, while boundary information provides detailed

Fig. 3. Presentation of (a) cropland regions and (b) boundaries within the cropland regions, i.e., the field boundaries.

information for finer parcel delineation within cropland, compensating for the limitations of coarse-grained analysis inherent in region recognition. Therefore, integrating these two tasks is essential to enhance the feature interpretability of field boundaries and optimize the detection performance.

In contrast to typical semantic segmentation tasks such as cropland region recognition, the field boundary task demands a robust boundary sensing capability to identify all potential boundaries. Therefore, using the same detection architecture for both tasks is not practical. Field boundaries consist of boundaries with different granularity levels, which correspond to multilevel features in the network. High-level, low-resolution features predict the coarse-width overall parcel outline, such as road boundaries surrounding cropland. Low-level features describe finer boundaries within cropland, such as ridges. To detect all field boundaries effectively, it is necessary to employ a side architecture for the boundary task, enabling multiscale boundary prediction and appropriately weighting different granularity levels of boundaries.

In addition, in boundary task, the boundary information would be seriously lost in the low-resolution high-level features, and such inaccurate features cause errors in results. However, conventional supervised training methods lack sufficient motivation to correct them, so more powerful supervision needs to be imposed on boundary task. In addition, as a kind of large-span connectivity information, field boundary detection would benefit from the longer ranged context information. Expanding the feature's receptive field is also crucial for cropland region tasks. These long-range assistances can effectively identify land cover types, especially when detecting parcels with diverse shapes and sizes [25].

2) Model Generalization Performance: Spectral features of cropland vary across time periods, which can significantly restrict the temporal generalization ability of the model. However, the delineation of CLPs only necessitates the differentiation between cultivated and noncultivated areas, which relies more on the texture features, featuring the flat texture within the cropland, surrounded by boundaries with dramatic spectral transitions. The flat texture of cropland clearly distinguishes it from other land cover types with rough textures such as residential and forested regions, forming a distinct contrast. Moreover, the temporal fluctuations in the spectral are relatively subtle for cropland, and it consistently maintains discernible spectral distinctions from other flat-textured noncropland areas, such as factory roofs and water bodies. This

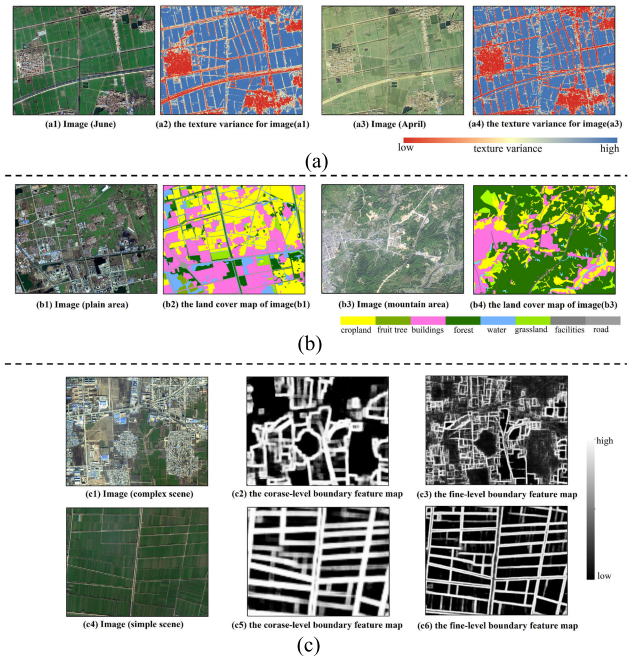


Fig. 4. Analysis of the factors affecting the model generalization performance. (a) Temporal generalization analysis. (a1) Image (June), (a2) texture variance for image (a1), (a3) image (April), and (a4) texture variance for image (a3). (b) Space generalization analysis (cropland region identification). (b1) Image (plain area), (b2) land cover map of image (b1), (b3) image (mountain area), and (b4) land cover map of image (b3). (c) Space generalization analysis (field boundary detection). (c1) Image (complex scene), (c2) and (c5) corase-level boundary feature map, (c3) and (c6) fine-level boundary feature map, and (c4) image (simple scene).

pattern remains constant across different temporal images, as demonstrated in Fig. 4(a), where the detection of two images with different phases yielded consistent texture results. Consequently, it is theoretically possible to improve the generalization ability by modeling spectral-insensitive features with little impact on cropland identification performance.

Moreover, recognizing the cropland region under different image scenes is complicated by significant variations in the surrounding landscapes, which can impede the space generalization of the model [Fig. 4(b)]. Therefore, to enhance the generalization of region task, it is crucial to minimize the interference of irrelevant information surrounding cropland during feature extraction and emphasize spatially salient features associated with cropland classification.

In the field boundary task, the final boundary result is a weighted aggregation of multilevel boundary prediction. The CNN-based boundary detection networks, such as HED and RCF, concatenated the multilevel prediction and passed them through a single 1×1 convolutional layer to achieve weight merging. However, the connection weights of each feature layer with the 1×1 convolutional layer are fixed within a single model and only represent weight allocation in the current training scene. It can be challenging for the model to produce desirable results when applied to different data scenes because the weight allocation for multilevel features differs across scenarios. As shown in Fig. 4(c), in areas where cropland distribution is sparse and mixed with other objects, higher level abstract features need to dominate, leading to a

tradeoff between sacrificing boundary accuracy and identifying accurate category information. Conversely, inner farmland scenarios require more low-level features to provide detailed information. Therefore, to enhance the model space generalization for field boundary detection, the model needs to be able to adaptively adjust the weight combinations of multilevel features in various scenarios.

B. Network Architecture

The architecture of our network is illustrated in Fig. 5. It comprises a multitask head network, constructed using an encoder–decoder structure. The encoding component employs a sequence of convolutional layers to aggregate contextual information and extract latent multilevel features pertinent to the CLPs' delineation. On the other hand, the decoding component, consisting of the region decoder and the boundary decoder, further decodes extracted features specific to the respective tasks and generates predictions, yielding corresponding results for the parcel regions and boundaries, respectively.

1) Image Input: We have analyzed that spectral variations across different temporal phases significantly impact the model's temporal generalization. Theoretically, texture features play a more crucial role in cropland recognition. Thus, the color space augmentation (CSA) was adopted in our method. By introducing random transformations within the hue-saturation-value (HSV) color space of the input samples, we effectively simulate the inherent heterogeneity of cropland's spectral characteristics arising from diverse imaging conditions, modeling spectral-insensitive semantic features and enhancing the model's temporal generalization. In addition, geometric transformations, encompassing horizontal, vertical, and diagonal flips, are also applied to augment the training dataset, thereby strengthening the model's robustness to rotation invariance [46].

2) Encoder: The encoder section employs the ResU-Net network's VGG-like encoding backbone, which is composed of a series of Conblock units, each stacked with a series of 3 × 3 convolutional layers. In addition, the network adds residual connections to alleviate the risk of gradient vanishing or explosion. Residual connections are preceded by 1 × 1 convolutional layers that adjust the channel dimensions of the original feature maps. Following each Conblock unit, squeeze-and-excitation (SE) operations are added to recalibrate the feature channels, thereby enhancing the focus on significant channels [47]. These Conblock units extract multilevel features from local to global scales.

3) Cropland Region Task: The cropland region task decodes the region-related features, which is designed based on the decoding component of the ResU-Net network [48]. This decoding approach facilitates the adaptation to the CLPs with diverse shapes and sizes through the fusion of multilevel features [49]. Specifically, the module progressively upsamples the feature maps through 3 × 3 convolutional units, and a skip layer connects the feature from the encoder after each upsample. This process yields a comprehensive feature representation of cropland regions.

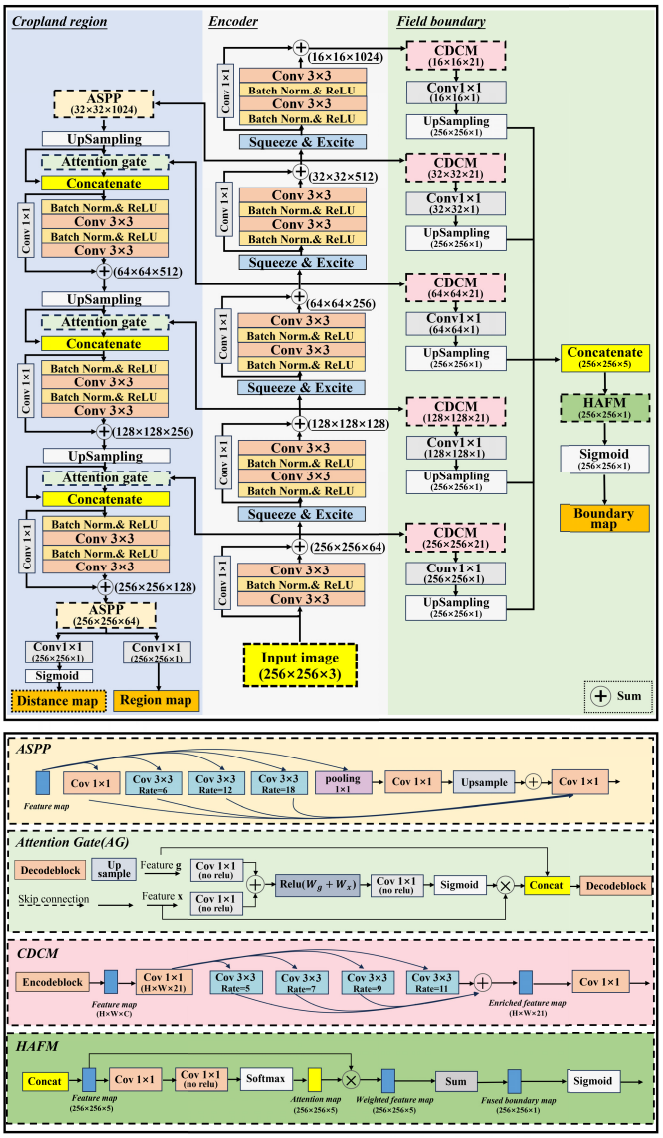


Fig. 5. Architecture of the proposed network, which has two decoders, corresponding cropland region and field boundary feature process. The numbers in parentheses refer to the height, width, and dimensionality of output channels.

In addition, we have analyzed that variations in the surrounding landscapes severely undermine the spatial generalization for cropland area recognition. Therefore, to suppress irrelevant regions and highlight the salient features in specific local areas, attention gate (AG) units [50] are introduced at each skip-layer connection. These AG units utilize dual 1 × 1 convolutions on the concatenated features to derive spatial attention coefficients, optimizing the corresponding layer's feature maps. Furthermore, the atrous spatial pyramid pooling (ASPP) [51], which consists of multiple parallel atrous convolutions with varying rates, was adopted to capture multiscale and larger range contextual information. Such information facilitates more comprehensive identification of land cover types and more effective detection of parcels with varying shapes and sizes. In this study, ASPP serves as a bridge between the encoder and decoder sections, and it is also applied after the final decoder block. Such strategic positioning

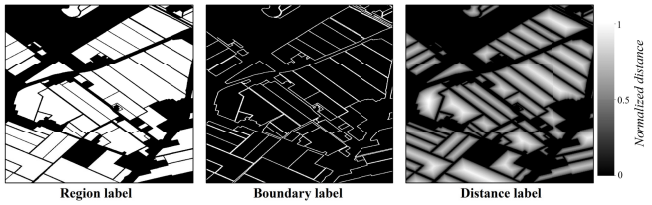


Fig. 6. Example of provided labeled samples of cropland regions, boundaries, and calculated distances.

has been proven to be more effective in capturing valuable multiscale information [50].

In addition, adding a distance (to cropland region mask) auxiliary prediction to the region task provides information for the topological connectivity of the segmentation regions, which can improve cropland segmentation in the geometric aspect and better avoids parcel undersegmentation [24], [48]. Here, we utilize the quasi-Euclidean distance transformation to compute the distance to the closest boundary for each pixel within the cropland region (Fig. 6). Compared to the traditional Euclidean distance that considers only diagonal pixels, the quasi-Euclidean distance, which accounts for horizontal, vertical, and diagonal pixels, offers a more accurate measure of distance [52].

4) Field Boundary Task: By drawing inspiration from the design of edge detection networks, such as HED [37] and RCF [53], we adopt the side structure as a decoder for field boundary task, which generates edge maps from each stage, and weighted integrates them into the interested field boundaries. The side structure can effectively capture multilevel boundary features, resulting in a more comprehensive hierarchical edge representation. Furthermore, the multilevel predictions of side structure can also facilitate applying deeper supervision to the boundary detection results.

To optimize feature maps at each stage, we developed a compact dilation convolution-based module (CDCM) to enrich the edge information before decoding the features. This module first reduces the dimensionality of the output multichannel features to 21 using a 1×1 convolutional kernel. This parameter value aligns with RCF's parameter setting for feature dimension reduction. Afterward, the dimension-reduced features undergo a series of dilated convolutional kernels to expand the receptive field of boundary features and enrich their representation. Due to the large span of field boundaries, the rich long-range assistance information provided by CDCM's stacked dilated convolution layers provides better preservation of the overall connectivity of the boundaries and more precise depiction of categorical features. In addition, this approach, without introducing downsampling, maintains fine-grained detailed information, which is vital for boundary objects with narrow widths.

The feature of each level was subjected to a 1×1 convolution for dimensionality reduction to a single channel. It is then upsampled to the original size for output, while the side loss is computed using the ground-truth map to provide deep supervision. Multilevel boundary predictions necessitate consolidation into a single-channel boundary result. Conventional boundary models employ a 1×1 convolution for this aggregation, featuring fixed weight allocation for multilevel predictions, which struggles to adapt to varying scenarios, thus limiting spatial generalizability of field boundary tasks. Therefore, we design a hierarchical attention fusion module (HAFM) to achieve adaptive weight fusion. This attention module utilizes dual 1×1 convolution kernels to simulate the nonlinear variations of attention, thereby learning the mechanism of weight fusion for different types of pixels.

5) Multitask Loss: In semantic segmentation tasks such as cropland region identification, the Dice loss function emerges as a better choice compared to the cross-entropy function as its faster training convergence and improved handling of class imbalances [54]. Thus, we defined the loss function L_{reg} for the cropland region task in our study as follows:

$$L_{\text{reg}} = 1 - \frac{2\sum_{i=1}^{N} P_i \times G_i + \varepsilon}{\sum_{i=1}^{N} P_i^2 + \sum_{i=1}^{N} G_i^2 + \varepsilon} \tag{1}$$

where N is the total pixel count, P_i represents the model's prediction for the ith pixel, G_i represents the label value for the ith pixel, and ε is a minute value introduced to prevent division by zero in the denominator.

The loss for the distance auxiliary task calculates the mean square error between the predicted distance y_D^i (the shortest distance of the pixel to the predicted cropland region) and the ground-truth distance $\hat{y}_D^i$ (the shortest distance to the ground-truth region)

$$L_{\text{dis}} = \frac{1}{N} \sum_{I=1}^{N} \left(y_D^i - \hat{y}_D^i \right)^2. \tag{2}$$

In addition, for the field boundary task, the classes "field boundary" and "nonfield boundary" are highly imbalanced. It is necessary to set a greater weight on the field boundary to help the model training to be more focused. The weight value is calculated from the ratio between the two classes [53]. The loss functions are defined as (3), and the ultimate boundary loss L_{bou} emerges as a composite value that sums up the losses of multilevel boundary predictions [see (4)]

$$L_{\text{bou}}^k = \frac{1}{N} \left(-\sum_{n=1}^{N} \left(\frac{|Y_-|}{|Y_+ + Y_-|} \hat{y}_+^k \log y_+^k + \frac{|Y_+|}{|Y_+ + Y_-|} \hat{y}_-^k \log y_-^k \right) \right) \tag{3}$$

$$L_{\text{bou}} = \sum_{k=1}^{5} l_{\text{bou}}^k + l_{\text{bou}} \tag{4}$$

where k is the level of the boundary predictions; y_+ and y_- are prediction positive pixels and negative pixels, respectively; $\hat{y}_+$ and $\hat{y}_-$ are the label category of them; and $|Y_+|$ and $|Y_-|$ are the number of their pixels.

Multitask loss functions are usually computed as a linear summation of individual task losses. Nevertheless, as task importance varies, manual weight tuning for each task proves labor-intensive. Thus, this study employs a homoscedastic uncertainty-based method to autonomously adjust task weights [55]. This technique models intertask uncertainty

(noise) to measure the relative confidence of tasks, thereby determining appropriate weights. Such an approach finds extensive utility in various multitask networks [25], [56], [57].

To provide specific details, the adaptive multitask loss function is formulated based on the maximization of homoscedastic uncertainty likelihood estimation. Let $f^w(x)$ represent the output of a task under weight w and input data x, and $P(y|f^w(x))$ signifies the model likelihood. For regression tasks, such as the distance task in our study, the probability distribution follows a Gaussian distribution, leading to Gaussian likelihood estimation: $P(y|f^w(x)) = \mathrm{N}(f^w(x), \sigma^2)$. For classification tasks, such as region task, employing the softmax function to normalize model outputs, the expression becomes $P(y|f^w(x)) = \mathrm{Softmax}(f^w(x))$. Here, the parameter σ^2 stands as a trainable model noise parameter.

For the losses of boundary, region, and distance tasks, their joint model likelihood is given by: $p(y_1, y_2, y_3|f^w(x)) = \mathrm{Softmax}(f^w(x), \sigma_1^2) \cdot (f^w(x), \sigma_2^2) \cdot \mathrm{N}(f^w(x))$. Through the logarithmic transformation of the joint likelihood, it is converted into a minimization objective function, serving as the loss $L(w, \sigma_1, \sigma_2, \sigma_3)$ for the three tasks, which can be further reduced to the form of the following equation:

$$\begin{aligned} &L(w, \sigma_1, \sigma_2, \sigma_3) \\ &= -\log \mathrm{Softmax}\left(f^w(x), \sigma_1^2\right) \cdot \left(f^w(x), \sigma_2^2\right) \cdot N(f^w(x)) \\ &= \frac{1}{2\sigma_1^2} L_{\mathrm{reg}}(W) + \log\sigma_1 + \frac{1}{2\sigma_2^2} L_{\mathrm{bou}}(W) + \log\sigma_2 + \frac{1}{\sigma_3^2} L_{\mathrm{dis}}(W) \\ &\quad + \log \frac{\sum_c' \exp\left(\frac{1}{\sigma_3^2} f_{c'}^w(x)\right)^{\frac{1}{\sigma_3^2}}}{\left(\sum_c' \exp\left(f_{c'}^w(x)\right)\right)} \\ &\approx \frac{1}{2\sigma_1^2} L_{\mathrm{reg}}(W) + \frac{1}{2\sigma_2^2} L_{\mathrm{bou}}(W) + \frac{1}{\sigma_3^2} L_{\mathrm{dis}}(W) + \log\sigma_1\sigma_2\sigma_3 \end{aligned} \tag{5}$$

where $L_{\mathrm{reg}}(W)$, $L_{\mathrm{bou}}(W)$, and $L_{\mathrm{dis}}(W)$ represent the loss for the region, boundary, and distance tasks, respectively. The parameters σ_1–σ_3 correspond to the noise parameters of the three tasks, reflecting their uncertainties. Higher noise values indicate increased uncertainty and lower task weights.

C. BC-OF Optimization

Owing to potential issues such as boundary occlusion and unclear display in the images, the delineated boundary results may exhibit discontinuities. This can lead to incomplete division between adjacent parcels. Furthermore, the boundary and region results need to be fused into more regularized parcels. In this study, a novel boundary connection and object-level fusion (BC-OF) approach is proposed for result optimization.

The oriented connection of broken boundaries is realized through breakpoint and junction detection, as illustrated in the implementation process depicted in Fig. 7. Initially, the method employs the Zhang–Suen algorithm to transform the detected boundaries into skeletal lines with a single-pixel width. This algorithm iteratively refines the width inward until it reaches a pixel-wide structure. Subsequently, leveraging the skeletonized boundaries, we directly conduct breakpoint and junction detection based on the eight-neighborhood information of boundary points. This is facilitated by the fact that the eight-neighbor window of a typical boundary pixel encompasses two boundary pixels, whereas breakpoints are singular, and junctions exceed two. The breakpoint information localizes the existence of broken boundaries, while the directions of connections between breakpoints and junctions indicate the extension direction of these broken boundaries. Finally, we establish connections between pairs of breakpoints that satisfy the conditions delineated in (6) and (7) in terms of distance and extension direction. After restoring the connected boundaries to their original width using morphological dilation, high-quality object-level boundaries can finally be obtained. The dilation's operator size is determined by the number of iterations in the skeletonization process, which signifies the original boundary width as each iteration only erodes a single pixel

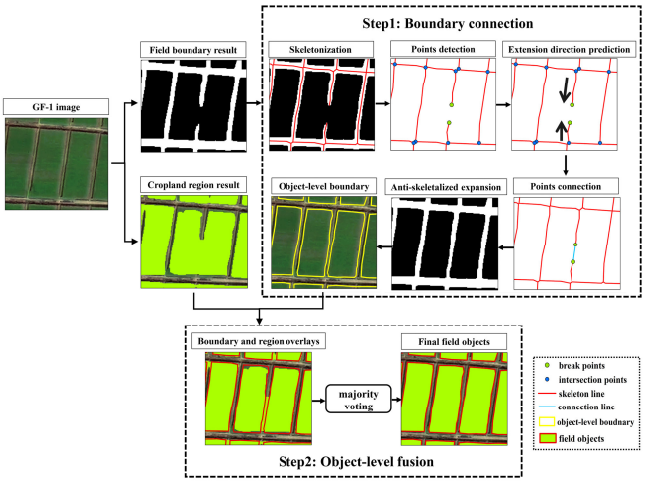


Fig. 7. Workflow of the proposed BC-OF optimization method.

$$|\mathrm{BP}_1, \mathrm{BP}_2| \le \frac{|\mathrm{IP}_1, \mathrm{IP}_2|}{3} \tag{6}$$

$$\pi - |\theta_{\mathrm{BP1}}, \theta_{\mathrm{BP2}}| \le \frac{\pi}{9} \tag{7}$$

where $|\mathrm{BP}_1, \mathrm{BP}_2|$ denotes the Euclidean distance between two breakpoints; IP_1 and IP_2 denote the intersection points corresponding to breakpoints BP_1 and BP_2, respectively; and $|\theta_{\mathrm{BP1}}, \theta_{\mathrm{BP2}}|$ denotes the angle of the predicted extension direction of the two breakpoints.

The optimized object-level boundaries are combined with the identified cropland regions using a pixel majority voting strategy [58]. Specifically, for each boundary-generated object G, we count the number of pixels predicted as cultivated land class (M) and the number of pixels classified as noncultivated land (N). If M is greater than N, the object is retained; otherwise, it is discarded. This voting approach is applied to all generated candidate objects, ultimately producing independent and accurate results for CLPs.

IV. Study Areas and Datasets

A. Study Sites and Image Data

This study selected four experimental regions to explore the space–time generalization performance of the model. These

TABLE I
SUMMARY OF THE STUDY AREAS AND IMAGES USED FOR EACH VALIDATION GOAL

Study site	Area and cropland share	Common crops	Topography	Field geometric characteristics	Image data	Image acquisition time
Source domain:						
Bincheng	1112km² (64%)	Wheat, maize, beans	Plain	large and regular	GaoFen-1 (2m, RGB)	June, 2017
Dong'e	729km² (55%)	Wheat, maize, beans	Plain	large and regular	GaoFen-1 (2m, RGB)	June, 2017
Luhe	986km² (5.7%)	Rice, wheat, fruit,	Mountain	small and irregular	GaoFen-1 (2m, RGB)	June, 2017
Funan	431km² (17%)	Rice, citrus, tea,	Hilly	small and regular	GaoFen-1 (2m, RGB)	January, 2021
Temporal generalization performance:						
Bincheng	1112km² (64%)	Wheat, maize, beans	Plain	large and regular	GaoFen-1 (2m, RGB)	April, 2017
Dong'e	729km² (55%)	Wheat, maize, beans	Plain	large and regular	GaoFen-1 (2m, RGB)	February, 2017
Spatial generalization performance:						
Netherlands	1495 km² (83%)	Rice, citrus, tea,	Plain	large and regular	Google earth (1m)	September, 2019
Netherlands	1495 km² (83%)	Rice, citrus, tea,	Plain	large and regular	Sentinel (10m)	August 2019

chosen areas cover typical cultivated land landscapes from north to south in China. Specifically, Bincheng and Dong'e, situated in Shandong Province, are characterized by plain terrain and a substantial proportion of cultivated land. Winter wheat and summer maize are the primary crops cultivated in these regions, with large and regular CLPs. In contrast, the Luhe experimental area, located in Guangdong Province in southern China, predominantly focuses on rice cultivation. This area is, indeed, characterized by a lower cropland proportion and more small and irregular parcels, as it is located in a mountain area, forming fragmented patterns. The Funan experimental area, situated at the crossroads between northern and southern China in Anhui Province, combines features of both regions, characterized by diverse crop types, and relatively small yet relatively regular land parcels. Overall, these study areas allow a comprehensive assessment of the delineation performance in different topographic features, crop types, field characteristics (Table I).

Given the prevalent small-sized CLPs in China, this study employed high-resolution fusion imagery from Gaofen-1 (RGB) with a spatial resolution of 2 m. The utilization of such high-resolution imagery facilitates a clear representation of the cultivated land demarcation. For model training, we selected a single image from each of the four counties as the source domain [Fig. 8(a)–(d)]. Moreover, in the Bincheng and Dong'e experimental areas, two temporally distinct images were selected to systematically evaluate the trained model's temporal generalization performance under various crop growth seasons (spring and winter).

In addition, using the model trained in the source domain, we also selected two distinct sensor images located in The Netherlands to further assess the model's spatial generalization performance and its effectiveness across different image resolutions acquired from various sensors. The selected transfer area is situated in a large-scale cultivated region of The Netherlands, encompassing approximately 3.6% of the country's total land area. The large size of the CLPs in this region ensures the feasibility of delineation under different image resolutions. The evaluation consisted of 10-m resolution Sentinel imagery and 1-m resolution Google Earth (GE) imagery, both acquired during the peak growing season.

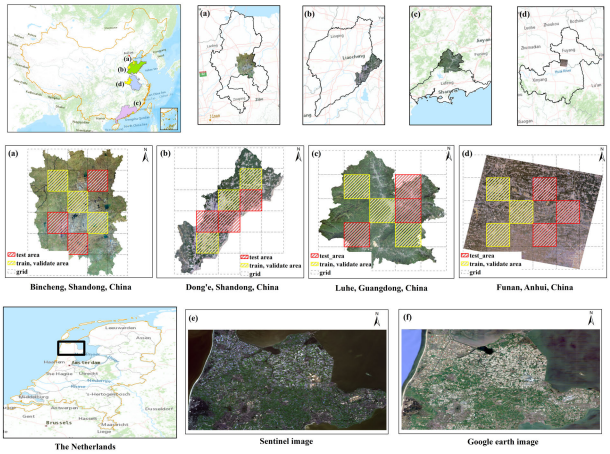


Fig. 8. Geographic location of the study sites and their images. (a) Bincheng County, Shandong; (b) Dong'e County, Shandong; (c) Luhe County, Guangdong; (d) Funan County, Anhui; and (e) and (f) Sentinel and GE images of the study area of the Netherlands, respectively.

B. Reference Data and Model Training

The ground-truth CLPs data for the GaoFen-1 dataset were manually delineated and cross-validated with situ identification data from the National Bureau of Statistics of China, exhibiting accuracy rates exceeding 95%, while the ground-truth data for The Netherlands datasets were obtained from The Netherlands' basic registration of crop plots (BRP), which are made available via the "Public Data on the Map" initiative of the Dutch Ministry of Economic Affairs and Climate via (https://www.pdok.nl/geo-services/-/article/basisregistratie-gewaspercelen-brp-#9464039d91ac261a857ee92a9f215250).

From each GaoFen-1 image, we divided it into grids (5 × 6 or 5 × 5) based on the bounding rectangle and selected three of these grids (the yellow boxes in Fig. 8) to produce the training and validation sets. The selection criterion was to cover the majority of the landscapes within that region. The selected areas were segmented into 256 × 256 sample sets and subjected to geometric transformations and CSA within an 8% fluctuation range. The augmented sample tiles were partitioned into training and validation sets using a 0.9:0.1 random split. For testing purposes, three grids located outside the training

areas (red boxes in Fig. 8) were randomly selected to verify the accuracy across each experimental region of the Gaofen-1 dataset. The accuracy validation for The Netherlands was conducted across the entire image domain.

The Adam algorithm [59] was used for gradient descent optimization with an initial learning rate of 0.001, a batch size = 8, and 80 epochs during the training process. The momentum was set to 0.9 to regularize learning. Using an Nvidia GTX 1080 Ti GPU, the training process took 58 h. Subsequent ablation and comparative experiments adhered to the same training samples, testing regions, and methodologies as our model.

C. Accuracy Assessment

1) Boundary Accuracy: The boundary accuracy is the correspondence degree between the detection and reference parcel boundaries. Considering the narrow width of the boundary, absolute correspondence can be difficult to achieve. Two-pixel accuracy tolerance offsets were set in this study.

The $F1$ score is used as a measure, which is the summed average of precision and recall [60]. The larger the $F1$ score, the closer to the optimal delineation. The equation is given as follows:

$$\text{Boundary Presicion(BP)} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

$$\text{Boundary Recall(BR)} = \frac{\text{TP}}{\text{FP}+\text{FN}} \tag{9}$$

$$\text{Boundary } F1 = 2 \times \frac{\text{BR} \times \text{BP}}{\text{BR}+\text{BP}}. \tag{10}$$

2) Geometry Accuracy: It can be divided into three parts: area accuracy, position accuracy, and shape accuracy.

a) Area accuracy: This accuracy analyzes the correctness and completeness of the delineated parcels. Consistent with the boundary accuracy, the $F1$ score was used as a metric for area accuracy.

b) Position accuracy: This accuracy represents the matching degree between the centroids of the delineated and reference parcels [41]. First, the Euclidean distance between the two centroids is calculated, and then, it is normalized by the diameter of equal area circle

$$P^i_{\text{centroid}} = 1 - \frac{d(C_T, C_E)}{D_{\text{cac}}} \tag{11}$$

$$D_{\text{cac}} = 2\sqrt{\frac{S_T + S_E}{\pi}} \tag{12}$$

$$P_{\text{centroid}} = \frac{P^i_{\text{centroid}}}{N} \tag{13}$$

where P^i_{centroid} denotes the position accuracy of the ith field, $d(C_T, C_E)$ is the Euclidean distance between the centroids, $S_T + S_E$ is the combined area of two parcels, and N is the number of all parcels.

c) Shape accuracy: This accuracy measures the shape similarity of the delineated and reference parcels. The normalized perimeter index (NPI) [61] is applied to define the shape factor and the accuracy was expressed by the ratio of the two

$$P^i_{\text{shape}} = \frac{\text{NPI}_{Ei}}{\text{NPI}_{Ti}} \tag{14}$$

$$\text{NPI} = \frac{P_{\text{eac}}}{P_{\text{object}}} \tag{15}$$

$$P_{\text{shape}} = \frac{P^i_{\text{shape}}}{N} \tag{16}$$

where P^i_{shape} is the shape accuracy of the ith field; P_{object} and P_{eac} are the perimeter of the object and its equal area circle, respectively; and NPI_{Ei} and NPI_{Ti} are the NPI of the extracted and reference parcels, respectively.

V. Result

A. Results of CLPs' Delineation

Fig. 9 and Table II present the method's delineation results and their accuracies in different regions, respectively. In the plain areas (Bincheng, Dong'e), the method achieves a remarkable ability to discriminate cropland from other land cover types, with area accuracies exceeding 0.95. The delineated CLPs are continuous in boundary and regular in shape, approaching visual interpretation standards. The boundary accuracy of both areas surpasses 0.93. Moving to the Funan experimental area, the method accurately identifies the cultivated land areas encompassing diverse crop types, with an area accuracy reaching 0.924. The method effectively captures small-sized parcels within this region. Extending the method to mountainous regions, as depicted in Fig. 9(d), it maintains a commendable identification accuracy, with boundary accuracy and area accuracy reaching 0.829 and 0.879, respectively. Notably, in the vicinity of residential areas at the foothills, the method accurately identifies cultivated land regions and refines the delineation of intricate parcels within. Moreover, the method exhibits a robust capability to distinguish isolated cropland dispersed within forests, effectively delineating independent and complete CLPs from spectrally similar forest areas.

Regarding temporal generalization performance, the model, benefiting from spectrally insensitive features, maintains good detection performance when transferred to images from different growing periods. Field boundaries, primarily constituted by fixed features such as roads and internal ridges, remain relatively stable across different growth stages. Consequently, the post-transfer accuracy is comparable to that of the source domain images. The slight decrease in accuracy is mainly due to the fact that the source domain's imageries acquired from the peak growth period can provide finer boundary results [Fig. 9(a) and (b)]. which can be attributed to three factors. First, during the peak growth period, the imagery accentuates the contrast between cultivated land crops and their boundaries, resulting in clearer boundary information. This facilitates the identification of more potential boundaries within cropland. Second, during this period, the imagery effectively highlights spectral disparities between adjacent parcels cultivated with different crops, enabling a more well-defined delineation. Third, the imagery during the peak growth period magnifies the distinctions between cultivated and noncultivated areas, consequently reducing misidentifications.

Even when transferring the model trained on source domain to The Netherlands' GE imagery, the delineation achieves remarkable precision ($F1$: 0.967), surpassing accuracy from

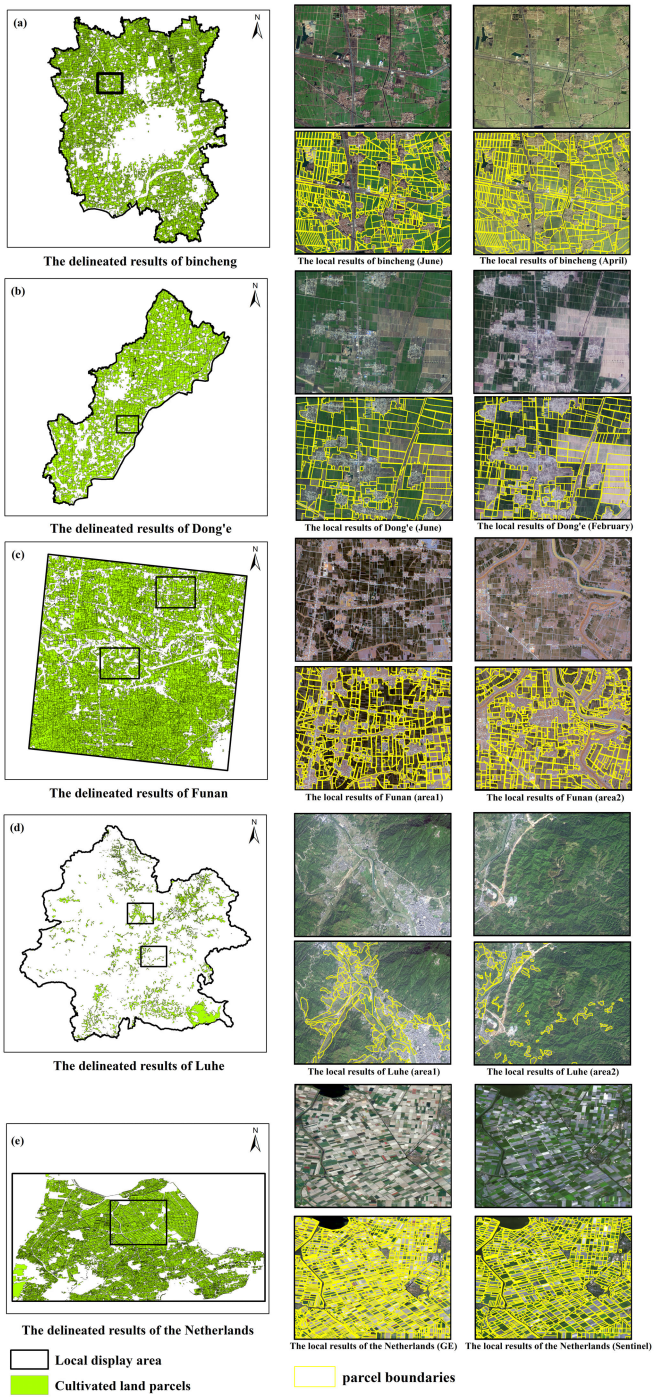


Fig. 9. Delineated field results in the five study areas. (a) Binchen, (b) Dong'e, (c) Funan, (d) Luhe, and (e) The Netherlands. (a) and (b) Delineation results on different phase images. (c) and (d) Results for different regions. (e) Results on different sensor images. The red circles show areas where images acquired from the peak growth period can provide finer boundary results.

any study area of source domain. This underscores the method's robust spatial generalization across regions with diverse sensors and image resolutions. Visually, in extensive agricultural landscapes such as The Netherlands, our approach effectively captures nearly all field boundaries, yielding continuous results. The delineated CLPs exhibit notable geometric regularity, akin to meticulous manual delineation. Upon application to 10-m resolution Sentinel imagery, the

TABLE II
ACCURACY OF THE DELINEATED RESULTS FOR THE STUDY AREAS

Area	Boundary accuracy			Geometric accuracy		
	F1	BP	BR	Area accuracy	Position accuracy	Shape accuracy
Binchen (June)	0.959	0.953	0.964	0.976	0.926	0.917
Dong'e (June)	0.941	0.957	0.926	0.963	0.941	0.924
Funan	0.891	0.869	0.915	0.924	0.902	0.886
Luhe	0.829	0.814	0.844	0.879	0.887	0.859
Binchen (Apr)	0.943	0.949	0.937	0.959	0.924	0.901
Dong'e (Feb)	0.927	0.943	0.911	0.948	0.933	0.915
Netherlands (GE)	0.967	0.972	0.962	0.973	0.932	0.943
Netherlands (Sentinel)	0.852	0.979	0.754	0.969	0.948	0.956

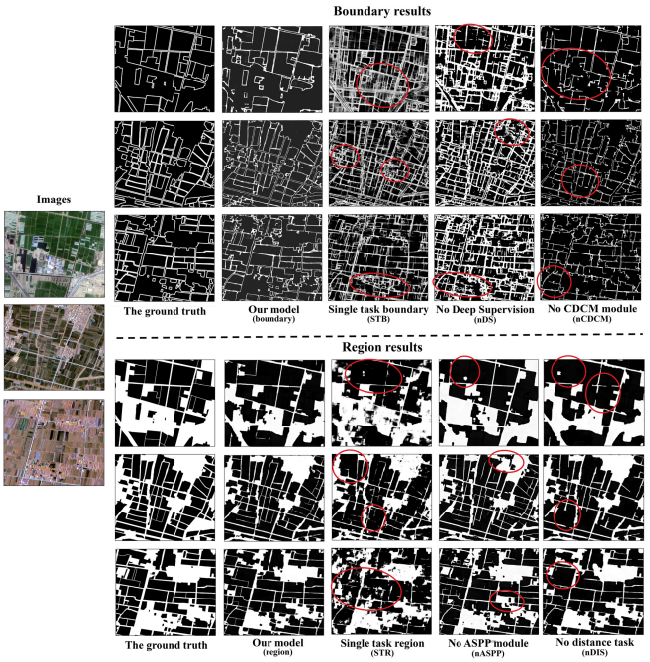


Fig. 10. Detection results of different networks. The upper part shows the results of the ablation experiment about the field boundaries detection, and the lower part shows the results of cropland regions, The red circles show areas where boundaries and regions are misidentified.

model similarly identifies clear, high-quality field boundaries, attaining a boundary precision (BP) of 0.979. In comparison to GE results, its delineations predominantly encompass large agricultural units, as finer internal boundaries are invisible on the 10-m resolution imagery. Considering that reference boundaries are generated from high-resolution imagery, with heightened granularity, the boundary recall (BR) result of Sentinel-based identifications is constrained, impacting the $F1$ score. Nevertheless, our model consistently showcases exceptional performance, effectively depicting comprehensive parcel divisions within the prevailing image resolution.

B. Ablation Experiments of Network

1) Ablation Experiments on Detection Performance: To enhance the network's detection performance, we have undertaken three improvements: 1) we transformed the network into a multitask network for regions and boundaries and individually designed decoder for each; 2) in the boundary detection

TABLE III
BOUNDARY AND AREA ACCURACY OF THE DELINEATED RESULTS BY DIFFERENT ABLATION EXPERIMENTS

Methods / Areas	Boundary accuracy (F1)				Area accuracy (F1)			
	Our model	STB	nDS	nCDCM	Our model	STR	nASPP	nDIS
Binchen (June)	**0.959**	0.861	0.813	0.921	**0.976**	0.939	0.960	0.948
Dong'e (June)	**0.941**	0.852	0.783	0.907	**0.963**	0.925	0.943	0.933
Funan	**0.891**	0.811	0.647	0.867	**0.924**	0.891	0.889	0.897
Luhe	**0.829**	0.708	0.619	0.792	**0.879**	0.857	0.832	0.849
Binchen (Apr)	**0.943**	0.848	0.792	0.898	**0.959**	0.927	0.951	0.952
Dong'e (Feb)	**0.927**	0.839	0.786	0.887	**0.948**	0.904	0.939	0.917

decoder, we introduced CDCM units to enrich the boundary information and applied deep supervision during training; and 3) in the region recognition decoder, we incorporated ASPP modules to capture extensive contextual information and added distance auxiliary task to avoid parcel undersegmentation. To validate the effectiveness of these strategies, we devised the following ablation experiments: 1) performing single-task detection for boundaries based on field boundary decoder (STB); 2) performing single-task detection for regions based on cropland region decoder (STR); 3) omitting the deep supervision strategy in boundary decoder (nDS); 4) removing the CDCM units from boundary decoder (nCDCM); 5) removing the ASPP module from region decoder (nASPP); and 6) removing the distance auxiliary task from region decoder (nDIS). The local results detected by different ablation networks are depicted in Fig. 10 and the accuracy results of Gaofen-1 dataset are presented in Table III.

a) Single task: Employing the same boundary/region training samples, we performed single-task detection using the corresponding field boundary/region task in our network. As depicted in Fig. 10, the single task for field boundary detection exhibited numerous erroneous boundaries extending beyond the cropland region, given the lack of region constraints. Essentially, this single task can be likened to directly employing a modified RCF edge detection network for field boundary detection. Such network struggles to achieve precise results due to the intricate semantic features of field boundaries. On the other hand, the single task for cropland region identification, equivalent to employing a modified U-Net network solely for region recognition, suffers from the loss of fine internal boundary details within the regions, resulting in the undersegmentation of adjacent parcels. The accuracy results of these two single-task networks fall significantly short of our multitask model's performance.

b) Deep supervision: Deep supervision was employed in the boundary task to compute corresponding losses for multilevel predictions and provide more powerful supervision. As illustrated in Fig. 10, this training strategy significantly impacts result accuracy. Without deep supervision, the network exhibits increased misidentifications in noncultivated land areas due to low-quality feature maps. In addition, the fine internal boundaries within cropland also suffer fragmentation.

c) Compact dilation convolution-based module: On the other hand, the added CDCM module, as depicted in Fig. 10, effectively enhances the connectivity of field boundary results, leading to improved boundary accuracy in each experimental area. The larger feature receptive field empowers field boundary predictions with comprehensive semantic understanding, mitigating the issue of boundary fragmentation arising from image blurriness and other objective conditions. Simultaneously, the enriched boundary features significantly reduce detection confusion, rendering the results with clearer and more explicit boundaries, devoid of noise information.

d) Atrous spatial pyramid pooling: The ASPP modules were integrated into region task to provide longer ranged context information. As illustrated in Fig. 10, the inclusion of this module effectively enhances the accuracy of recognition, enhancing area accuracy in each experimental area. Particularly in the areas of Funan and Luhe, characterized by complex land types, the accuracy gains are notable, reaching 0.035 and 0.047, respectively. The accuracy improvement primarily stems from reduced misidentifications outside the cropland regions and fewer internal holes.

e) Distance auxiliary task: The added distance auxiliary task, as depicted in Fig. 10, significantly addresses the issue of inadequate segmentation between parcels by providing topological connectivity information for the segmentation mask. This results in more individual and regular CLP results. Compared to the network without this auxiliary task, our network exhibits a noticeable improvement in area accuracies across all experimental areas.

2) Ablation Experiments on Generalization Performance: To enhance our network's generalization, we implemented three design strategies: 1) CSA for the input imagery to model weakly spectral sensitive features; 2) integration of AG units in region identification decoder to emphasize salient characteristics of specific local regions and suppress irrelevant areas; and 3) adoption of an HAFM in the boundary detection decoder for adaptive weight merging the multilevel predictions. A series of ablation experiments were conducted to analyze the effectiveness of these strategies.

a) Color space augmentation: Table IV demonstrates the significant negative impact of missing CSA on the model's generalization ability, leading to a noticeable decrease in accuracy when transferred to different time phases or regions. This is primarily due to the model without CSA relying on more spectral-related features, which are susceptible to variations in color across different study areas and temporal imageries. As a result, such a model has insufficient generalization performance, especially the temporal generalization. Moreover, the inherent color disparities in different sensor images also hinder model transfer across sensors. In contrast, our

TABLE IV
BOUNDARY AND REGION ACCURACY OF THE RESULTS BY EMPLOYING DIFFERENT COLOR FLUCTUATION RANGES FOR CSA

fluctuation / Areas	0% **(No CSA)**		5%		8% **(our setting)**		20%	
	Boundary (F1)	Area (F1)	Boundary (F1)	Area (F1)	Boundary (F1)	Area (F1)	Boundary (F1)	Area (F1)
Binchen (June)	0.919	0.906	0.943	0.952	**0.959**	**0.976**	0.849	0.841
Dong'e (June)	0.892	0.875	0.929	0.949	**0.941**	**0.963**	0.819	0.796
Funan	0.801	0.798	0.856	0.897	**0.891**	**0.924**	0.601	0.662
Luhe	0.752	0.714	0.804	0.839	**0.829**	**0.879**	0.584	0.628
Binchen (Apr)	0.887	0.894	0.936	0.967	**0.943**	**0.959**	0.836	0.834
Dong'e (Feb)	0.896	0.881	**0.931**	**0.954**	0.927	0.948	0.796	0.738
Netherlands (GE)	0.857	0.815	0.941	0.958	**0.967**	**0.973**	0.613	0.713
Netherlands (sentinel)	0.772	0.823	0.839	0.961	**0.852**	**0.969**	0.559	0.619

applied color augmentation on samples effectively simulates these spectral differences, capturing more robust features, thus enhancing the model's space–time generalization capacity and performance when applied to imagery from different times and regions.

In addition, when the color fluctuation range remains below 8%, the model consistently maintains robust extraction performance, with larger fluctuations leading to improved generalization. This can be attributed to the distinctive texture characteristics of cultivated land, featuring flat interiors and abrupt boundaries, which enables the model to differentiate it from spectral-similar forest areas, which have rough image texture due to the alternation of shadowed and bright spots. Moreover, other flat-textured land cover types in the imagery, such as factory roofs and water bodies, with significant color differences compared to cropland, exceeding the color range we set. Thus, such color transformations do not significantly weaken the model's ability to discriminate cultivated land. However, it should be noted that when the fluctuation range exceeds 20%, it still leads to a drastic decline in extraction performance.

b) Attention gate: In the region decoder, AGs were introduced at the skip-layer connections to overcome the heterogeneity of surrounding land types and maintain the network's focus on class-relevant areas. As shown in Table V, the model without this module exhibits a noticeable reduction in detection performance on the source domain's Gaofen-1 images, particularly in the complex land area of Funan, where the area accuracy decreased by 0.098. This impact becomes even more pronounced when the model is transferred to The Netherlands. On its GE and Sentinel images, incorporating the AG module results in area accuracy improvements of 0.114 and 0.158, respectively.

c) Hierarchical attention fusion module: In the boundary decoder, our network incorporates the HAFM module to achieve pixel-level adaptive fusion of multiscale boundary features. We compare this approach to the conventional 1 × 1 convolution-based fusion method. The latter assigns fusion weights in terms of the learning weights between features from different layers and 1 × 1 convolution kernel, which is fixed in a single model.

Fig. 11 counts the average of the boundary pixels' feature assignments in each region for the two methods. It can be

TABLE V
BOUNDARY AND AREA ACCURACY BY THE MODEL WITHOUT AG AND OUR MODEL

methods / Areas	No attention gate		Attention gate (our)	
	Boundary (F1)	Area (F1)	Boundary (F1)	Area (F1)
Binchen (June)	0.938	0.954	**0.959**	**0.976**
Dong'e (June)	0.914	0.939	**0.941**	**0.963**
Funan	0.806	0.832	**0.891**	**0.924**
Luhe	0.757	0.809	**0.829**	**0.879**
Binchen (Apr)	0.925	0.945	**0.943**	**0.959**
Dong'e (Feb)	0.920	0.937	**0.927**	**0.948**
Netherlands (GE)	0.877	0.859	**0.967**	**0.973**
Netherlands (sentinel)	0.786	0.811	**0.852**	**0.969**

found that the conv 1 × 1 method applies uniform weight allocation across various regions, and such inflexible distribution limits its generalization ability and restricts accuracy in some areas (Table VI). Only in the plain regions of Binchen and Dong'e, does the conv 1 × 1 method come close to the attention-based method in terms of assigning weights, achieving similar and competitive accuracy. However, in hilly regions such as Luhe, which demand more high-level features for semantic boundary recognition, the conv 1 × 1 method falls short, resulting in a BP of only 0.701, 0.128 lower than our attention-based approach. In contrast, our HAFM module computes the merging weights dynamically for each pixel, enabling it to dynamically balance semantic recognition and detail depiction across diverse scenes, yielding excellent delineation accuracy in all experimental areas of Gaofen-1 dataset. From a spatial transfer experiment perspective, the accuracy enhancement brought by the HAFM module to the boundary task is equally remarkable. When the model is transferred to The Netherlands, the HAFM module outperforms the conv 1 × 1 module in boundary $F1$ score improvements by 0.068 (GE image) and 0.063 (Sentinel image).

d) Feature separability of target domains: To further elucidate the model's spatial generalization performance, we assessed the feature separability of various ablation networks within the target domain (The Netherlands) through

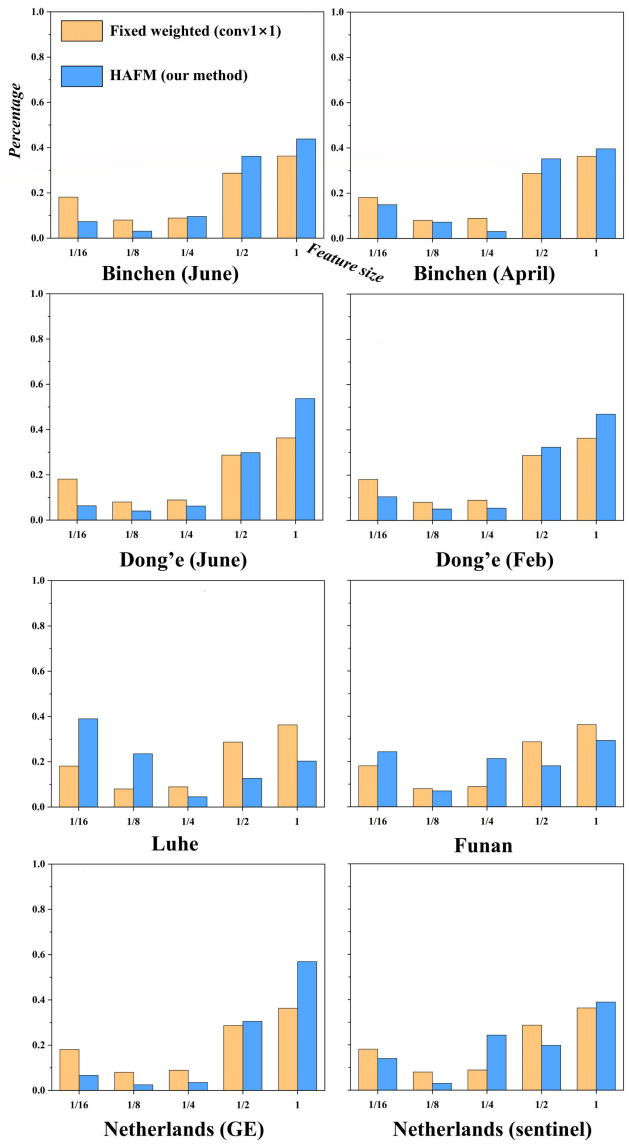


Fig. 11. Average of the boundary pixels' feature assignments in each region for the fixed weighted (1 × 1) and our hierarchical attention methods.

visualization techniques. The examined models included the model trained without CSA, the model with AGs removed, and the model with the HAFM module removed and only 1 × 1 convolution applied. The t-distributed stochastic neighbor embedding (t-SNE) algorithm [62] was employed for visualization. For the t-SNE visualization, we randomly selected 4000 field/nonfield boundary points and 4000 cropland/noncropland region points from the GE image of The Netherlands.

As shown in Fig. 12, despite being trained exclusively on the source domain's Gaofen-1 images, our model demonstrated remarkable class separability in the target domain, across both boundary delineation and regional classification tasks. In contrast, ablation models without CSA, AGs, and the HAFM module exhibited varying degrees of class overlap, highlighting the integral role each module plays in bolstering the model's generalization performance. The combined model (our model), integrating all three modules, showcases profound

TABLE VI

BOUNDARY AND AREA ACCURACY BY THE MODEL USING FIXED WEIGHTED (1 × 1 CONVOLUTION) FOR MULTISCALE BOUNDARY FUSION AND OUR HAFM

Areas \ methods	Fixed weighted		HAFM (ours)	
	Boundary (F1)	Area (F1)	Boundary (F1)	Area (F1)
Binchen (June)	0.941	0.970	**0.959**	**0.976**
Dong'e (June)	0.906	0.950	**0.941**	**0.963**
Funan	0.832	0.913	**0.891**	**0.924**
Luhe	0.701	0.862	**0.829**	**0.879**
Binchen (Apr)	0.919	0.951	**0.943**	**0.959**
Dong'e (Feb)	0.893	0.946	**0.927**	**0.948**
Netherlands (GE)	0.901	0.968	**0.967**	**0.973**
Netherlands (sentinel)	0.789	0.963	**0.852**	**0.969**

generalization capabilities, maintaining distinct differentiation between cropland areas and boundaries in our studied target domain.

3) Ablation Experiments on Network Settings:

a) RGB versus four-channel image: A further experiment on imagery from Funan County was conducted to evaluate the role and necessity of the near-infrared (NIR) band in the detection network.

We compared models trained on RGB imagery with those using RGB imagery supplemented with NIR (RGB + NIR). The experiments employed identical training settings, including sample areas and optimization methods. Surprisingly, the results from both RGB and RGB + NIR images were nearly indistinguishable in visual effect (Fig. 13) and accuracies in Funan County (Table VII).

This outcome can be attributed to the distinct spectral and textural differences between cultivated and noncultivated areas. The RGB band alone proved sufficient for effectively distinguishing these areas, resulting in remarkably high area accuracy. In addition, our research required delineating parcels with different growing periods and crops. The significant disparities in NIR responses between these parcels also underscored that the additional NIR band did not provide substantial benefits for cropland recognition. Furthermore, the limitations for boundary detection from RGB images mainly come from the unclear boundary presentation caused by thin widths or crop occlusion, However, introducing the NIR band did not alleviate this situation.

b) Multitask loss: In this study, we employ a homoscedastic uncertainty-based method to autonomously adjust the weights of multitask loss. To validate the effectiveness of this approach, two ablation experiments were conducted. In one experiment, we removed the homoscedastic uncertainty-based method, and the total loss was simplified to a straightforward summation of multitask losses (denoted as multitask summation). On the other hand, only the region and distance tasks were included within the homoscedastic uncertainty-based loss framework, with the boundary task being additionally summed. As shown in Table VIII,

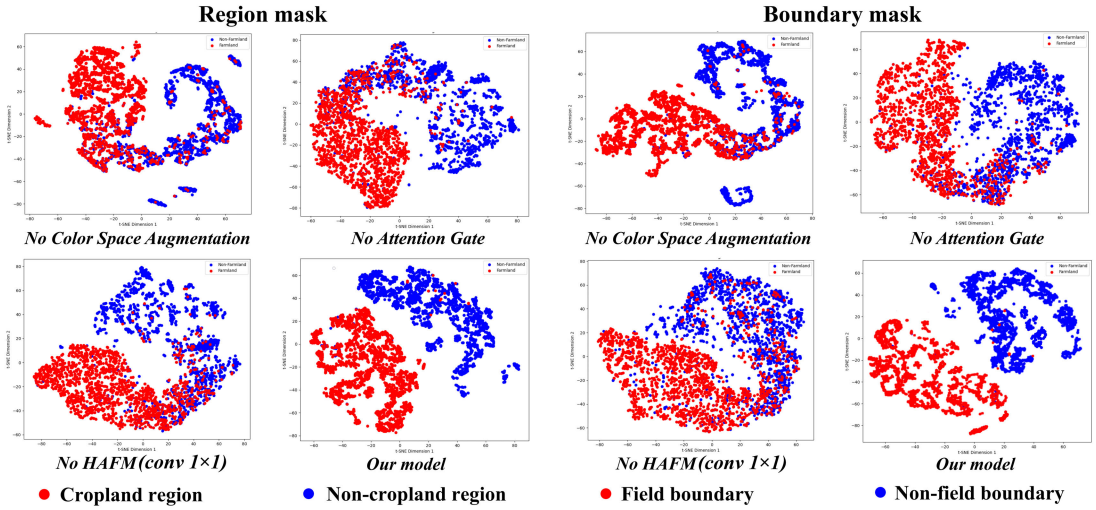


Fig. 12. Visualized feature separability of different ablation models in the target domain.

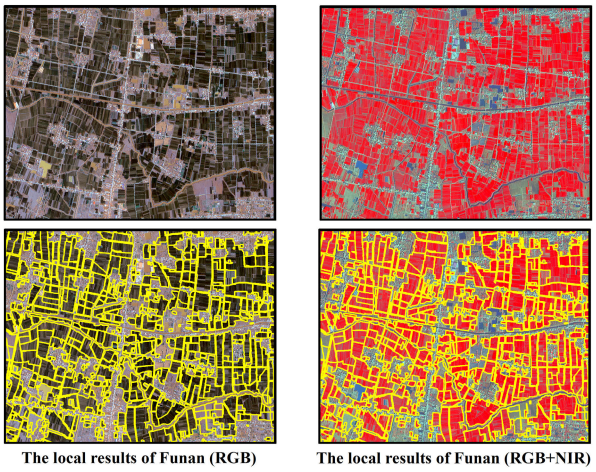


Fig. 13. Presentation of the delineation results from RGB images and four-band (RGB + NIR) images on a local region.

TABLE VII

BOUNDARY AND AREA ACCURACY OF THE RESULTS DELINEATED FROM RGB IMAGES AND FOUR-BAND (RGB + NIR) IMAGES

Input image	Boundary accuracy (F1)	Area accuracy (F1)
Funan (RGB)	0.891	**0.924**
Funan (RGB+NIR)	**0.899**	0.921

we calculated the average boundary and region accuracy for both methods across six study areas. The results show that the utilization of the homoscedastic uncertainty-based method facilitated the learning and improved accuracies for both tasks compared to the multitask summation method. Moreover, incorporating all tasks into this framework allowed for a better balance among them, improving the quality of boundary task results without compromising the area accuracy.

c) Network depth: Deeper network architectures facilitate the extraction of more abstract features, but they also introduce a greater computational burden to the model.

TABLE VIII

AVERAGE BOUNDARY AND REGION ACCURACY OF THE RESULTS USING THE DIFFERENT MULTITASK LOSS STRATEGIES

Multi-task loss strategies	Average boundary accuracy (F1)	Average area accuracy (F1)
Homoscedastic uncertainty-based	**0.915**	**0.942**
Homoscedastic uncertainty-based (no boundary)	0.904	0.940
Multi-task summation	0.901	0.929

TABLE IX

AVERAGE ACCURACIES AND COMPUTATIONAL EFFICIENCY OF THE RESULTS USING THE NETWORK WITH DIFFERENT DEPTHS (G: 10^9)

Network depth	Average boundary accuracy (F1)	Average area accuracy (F1)	FLOPs	Reasoning time per scene
4	0.832	0.876	118.4 G	1.6 min
5	0.915	0.942	157.3 G	2.5 min
6	**0.923**	**0.950**	196.1 G	3.9 min

To investigate this tradeoff, we conducted ablation experiments. As illustrated in Table IX, augmenting the network depth substantially increases both the floating-point operations per second (FLOPs)—a metric indicative of model complexity—and the average time required to predict each scene. While deeper networks enhance model precision, it was observed that when utilizing five resolution levels in the model, the accuracy gain from further increasing the network depth became very minimal (Table IX). Consequently, when balancing computational efficiency and performance, a five-depth network structure becomes the optimal configuration.

C. Comparison Experiments of Network

1) Comparison in Cropland Region Identification: To assess the performance of our network in identifying cropland regions, we conducted a comparative analysis with four commonly employed networks for cropland identification, which are described as follows.

TABLE X
AREA ACCURACY RESULTS FOR CROPLAND REGION IDENTIFICATION COMPARISON MODELS AND BOUNDARY ACCURACY FOR BOUNDARY COMPARISON MODELS

Networks	Area accuracy (F1)						Boundary accuracy (F1)					
	Our network (+optimization)	Our network (region)	MPSPNet	DeeplabV3+	U-Net	ResU-Net	Our network (+optimization)	Our network (boundary)	SegNet	U-Net	ResU-Net	R2U-Net
Binchen (June)	**0.976**	0.967	0.906	0.920	0.871	0.909	**0.959**	0.954	0.892	0.814	0.865	0.899
Dong'e (June)	**0.963**	0.958	0.912	0.913	0.849	0.904	**0.941**	0.935	0.858	0.799	0.842	0.886
Funan	**0.924**	0.916	0.874	0.890	0.827	0.886	**0.891**	0.879	0.806	0.757	0.816	0.841
Luhe	**0.879**	0.872	0.822	0.842	0.788	0.825	**0.829**	0.812	0.733	0.678	0.744	0.756
Binchen (Apr)	**0.959**	0.945	0.889	0.924	0.843	0.892	**0.943**	0.937	0.872	0.820	0.857	0.897
Dong'e (Feb)	**0.948**	0.939	0.901	0.889	0.851	0.899	**0.927**	0.921	0.861	0.809	0.850	0.872
Average	**0.942**	0.933	0.884	0.896	0.838	0.887	**0.915**	0.906	0.837	0.780	0.829	0.859

a) MPSPNet: It is a modified PSPNet network that has achieved more than 90% OA for cropland identification in four provinces of China [14].

b) DeeplabV3+: It is a widely used network that has been identified as the optimal method in a comprehensive model comparison study for recognizing cropland areas [26]. This network has also been applied for high-resolution cropland extraction [16].

c) U-Net: It is a network known for preserving fine details through shallow-level features and has been adopted in many works for cropland region recognition [17].

d) ResU-Net: It is an enhanced variant of U-Net network with improved cropland identification performance compared to the U-Net network [15].

These comparative networks were trained using the same cropland region labels as our model. Table X presents the area accuracy results of these networks on the GaoFen-1 dataset.

As indicated in Table X, our method exhibits superiority in area accuracy compared to all comparative approaches. Specifically, as shown in Fig. 14, both MPSPNet and DeepLabV3+ present clear region recognition results. However, due to insufficient high-resolution detail features, a considerable amount of internal boundary information is omitted in the identified results. This limitation hampers the direct applicability of these models for CLPs' delineation. In addition, the deficiency in boundary localization capability leads to irregular shapes of the recognized regions, negatively affecting the area accuracy. The UNet-based networks, benefiting from enriched high-resolution shallow-level information through skip-layer connections, are able to retain more detailed information. This enables them to depict partial internal boundaries and present more regular parcel shapes. Nonetheless, these internal boundaries remain fragmented. Furthermore, the limited type recognition capability of UNet-based networks introduces ambiguity in certain areas, resulting in low area accuracies. In contrast, our model, facilitated by the integration of boundary detection signals, produces continuous and comprehensive internal boundary segmentation. The clearer boundary constraints also yield remarkably regular parcel shapes. Moreover, our approach capitalizes on ASPP and AGs to significantly improve semantic recognition, thereby minimizing misidentifications. Consequently, our network achieves a notable average region accuracy of 0.942.

2) Comparison in Field Boundary Detection: To validate the performance of our network in field boundary detection, we conducted a comparative analysis with four commonly employed networks for field boundary detection, which are described as follows.

a) SegNet: It is a deep encoder–decoder network that has been used to detect the field boundaries from high-resolution WorldView-3 images [1].

b) U-Net: It is a U-shaped network with sufficient detail information for boundary preservation, which has been widely used in field boundary detection [12], [40].

c) ResU-Net: It is an enhanced variant of U-Net network with residual blocks and has been recently used to detect field boundaries from Sentinel-2 satellite images, with higher accuracy than U-Net [17].

d) R2U-Net: It is a U-shaped network with recurrent residual blocks, which improves feature representation by recurrently accumulating semantic features of multiscales and has been recently used to detect field boundaries from Sentinel-2 satellite images [18].

All comparative networks were trained using the same field boundary labels as our network. Table X presents the boundary accuracy results of these networks on the GaoFen-1 dataset.

As indicated in Table X, our method exhibits superiority in boundary accuracy compared to all comparative networks. Specifically, as shown in Fig. 14, The SegNet network produces fragmented boundary results, ignoring some internal boundaries within the fields. The U-Net architecture that has stronger descriptive capabilities can yield narrower boundary results. However, it exhibits substantial boundary omissions with an average boundary accuracy of only 0.780 due to its limited semantic recognition capability. The ResU-Net and R2U-Net networks enhance feature extraction by replacing more effective units and capture more abstract field boundary features. This greatly improved the boundary recognition rate, resulting in average boundary accuracies of 0.829 and 0.859, but they still suffered from some boundary breaks. Comparatively, our model identifies clear and continuous boundaries, achieving an average boundary accuracy of 0.915. This highlights that signals from the cropland region task can greatly improve the semantic recognition performance for boundary task. Furthermore, our employment of deep supervision and the CDCM module further ensures boundary quality.

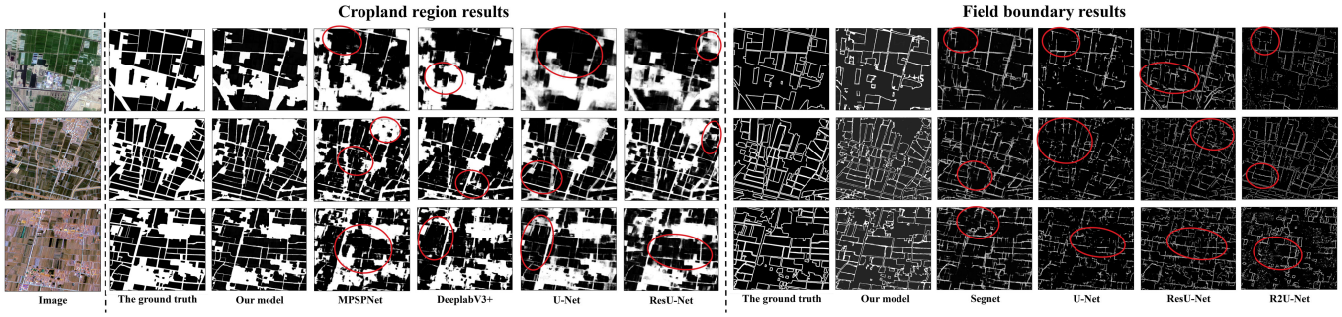


Fig. 14. Cropland region identification and field boundary detection results of different comparison networks. The red circles show areas where regions or boundaries are misidentified.

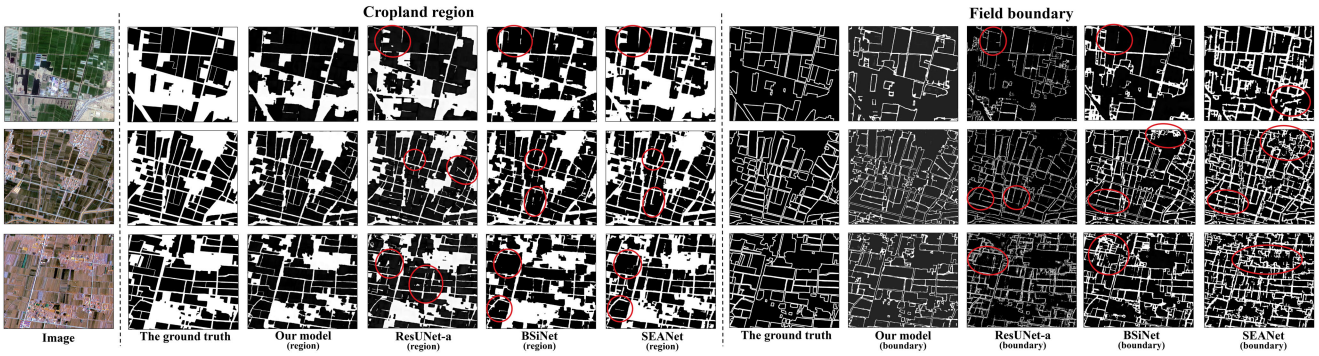


Fig. 15. Region and boundary results of our network and other recent multitask networks. The red circles show areas where boundaries and regions are misidentified.

3) Comparison Experiments With Recent Multitask Models: We further compare our method with three recent multitask networks: ResUNet-a, BSiNet, and SEANet, which are described as follows.

a) ResUNet-a: It is an augmented ResU-Net-based architecture for concurrently predicting field boundaries and regions, representing the first multitask model deployed for field boundary detection and achieving high detection accuracy. It is considered the state-of-the-art method by several studies.

b) BSiNet: It is a recent multitask model derived from PsiNet, which utilizes distinct decoding convolutions to generate region and boundary outputs. This network prioritizes region predictions as parcel outcomes.

c) SEANet: Similar to BSiNet in directly using region predictions as parcel outcomes, SEANet is renowned for its exceptional boundary awareness capability, thereby providing high-quality parcel-level region predictions that surpass BSiNet in terms of area accuracies.

All comparative networks were trained using the same cropland region and field boundary labels as our network and employed identical parcel optimization steps to generate the CLP results. Fig. 15 shows the predicted region and boundary results of different networks. Table XI presents the area and boundary accuracies ($F1$ score) of the final generated CLPs of these networks.

As shown in Fig. 15 and Table XI, ResUNet-a's region results ignore some internal boundaries and exhibit voids inside the cropland, leading to its area accuracy ($F1$) being 0.035 lower than our method. Similarly, its boundary detection produces slight discontinuities, resulting in a boundary accuracy gap of 0.036 compared to our method. This difference largely stems from ResUNet-a's unified decoder architecture for both boundary and region prediction tasks, ignoring their inherent heterogeneity. This renders the model relatively lacking in performance on individual tasks compared to our approach, struggling to achieve both accurate-type recognition and detailed boundary localization simultaneously.

Both BSiNet and SEANet demonstrated superior region accuracy but inferior boundary accuracy relative to ResUNet-a. Their identified regions appear more isolated with visibly wider separations; however, boundary inaccuracies were more frequent, aligning with their design philosophy that emphasizes region prediction as the core task. Despite this, their area accuracy remains below our network, missing some internal fine boundaries, with average area accuracy gaps of 0.024 and 0.009. This is primarily due to our network's design enhancements, such as AGs, SE blocks, and ASPP, which optimize semantic features from both spatial and channel dimensions and capture extended contextual information, enabling accurate identification across various cropland sizes and conditions. In boundary detection, BSiNet and SEANet significantly lagged behind our model, with average boundary accuracy gaps of 0.042 and 0.047, respectively. BSiNet only utilizes different decoding convolutional kernels to differentiate boundary and region tasks and essentially retains a U-Net-shaped architecture for boundary prediction, leading to insufficient boundary finesse. SEANet while enhancing boundary perception and identifying more potential boundaries fails to adequately balance region and boundary tasks,

TABLE XI
BOUNDARY AND REGION ACCURACY OF THE DELINEATED RESULTS USING OUR MODEL AND OTHER RECENT MULTITASK MODELS

Networks (+optimization)	Area accuracy (F1)				Boundary accuracy (F1)			
	Our network	ResUNet-a	BSiNet	SEANet	Our network	ResUNet-a	BSiNet	SEANet
Binchen (June)	**0.976**	0.927	0.946	0.963	**0.959**	0.917	0.910	0.902
Dong'e (June)	**0.963**	0.924	0.942	0.959	**0.941**	0.906	0.908	0.896
Funan	**0.924**	0.896	0.905	0.918	**0.891**	0.851	0.842	0.849
Luhe	**0.879**	0.856	0.848	0.867	**0.829**	0.796	0.783	0.789
Binchen (Apr)	**0.959**	0.910	0.931	0.948	**0.943**	0.898	0.902	0.883
Dong'e (Feb)	**0.948**	0.929	0.936	0.942	**0.927**	0.903	0.894	0.887
Netherlands (GE)	**0.973**	0.914	0.924	0.936	**0.967**	0.903	0.907	0.891
Netherlands (sentinel)	**0.969**	0.901	0.915	0.931	**0.852**	0.817	0.814	0.823

resulting in numerous recognition errors. In addition, its use of conventional 1×1 convolutions to aggregate multiscale boundary features and the fixed weight allocation result in overemphasized coarse-resolution features, leading to overly thick boundaries. In contrast, our network not only employs a side architecture for boundary tasks, leveraging its superior boundary perception capabilities, but also standardizes the boundary and region task losses at a unified scale and integrates them into an adaptive multitask loss function. This balances the learning of both tasks, achieving more accurate boundary results. Furthermore, our developed HAFM adaptively adjusts the weights of multiscale boundary features across various scenarios, predicting boundaries that more closely match actual widths. Our model also incorporates a CDCM module, enriching boundary features through stacked dilated convolution layers and capturing more long-distance dependencies, effectively aiding in producing more continuous boundary results.

It is important to note that solely outputting region predictions, as done by BSiNet and SEANet, exhibits limitations. Semantic tasks, such as cropland region prediction, do not offer the high predictive granularity of boundary tasks. Even our model, while excelling in region predictions, may omit boundaries that are fully presented in boundary predictions (illustrated by yellow circles in Fig. 15). Therefore, focusing solely on parcel-level region predictions limits the potential for further optimization. A more balanced approach involves simultaneously predicting high-quality cropland regions and boundaries, followed by result-level integration.

From a transfer experiment perspective, the lack of specialized design for transfer performance in ResUNet-a, BSiNet, and SEANet networks caused that when models trained on the source domain are transferred to the target domain, these comparative models exhibit significantly lower area and boundary accuracies compared to our model (Table XI).

Finally, in terms of network computational efficiency, our network demonstrates lower complexity and reasoning time per scene compared to SEANet, with a significantly reduced parameter count relative to the ResUNet-a network (Table XII).

TABLE XII
FLOPS, THE AVERAGE REASONING TIME FOR EACH SCENE, AND THE PARAMETERS OF THE THREE RECENT MULTITASK MODELS (G: 10^9 and M:10^6)

Networks	FLOPs	Reasoning time per scene	Parameters
Our network	157.3 G	2.5 min	49.1 M
ResUNet-a	73.9 G	2.3 min	134.7 M
BSiNet	13.4 G	1.4 min	7.9 M
SEANet	208.2 G	3.6 min	28.8 M

D. Comparison Experiments on Parcel Optimization

To validate the performance of the proposed BC-OF optimization method, we conducted a comparison with the CB-FF optimization method. The CB-FF method is known for its ability to better connect broken boundaries than other methods by performing morphological dilation on skeleton lines. This method then utilizes the repaired boundary to clip the identified region results and obtain the final CLPs.

However, as depicted in Fig. 16, our optimization method has noticeable progress in both boundary connection and result fusion over the CB-FF method. First, our method effectively addresses the boundary breaks with larger distances. Although the CB-FF approach can extend the boundary line through morphological dilation, it faces limitations in repairing long-distance breaks due to its restricted operator radius. In contrast, our boundary optimization approach identifies breakpoint and predicts the extension direction of the broken boundary, establishing directional connectivity between two breakpoints, thus achieving the connection of long-distance boundary breaks.

Second, our approach produces more regular parcels. The CB-FF method, which utilizes boundaries to clip region results, struggles to adapt to the situation where the identified cropland region is small, resulting in undersized and irregular parcels (black circles in Fig. 16). In fact, compared to the detailed boundary results, cropland region identification places more emphasis on semantic information, which may not accurately represent the transition between the cropland

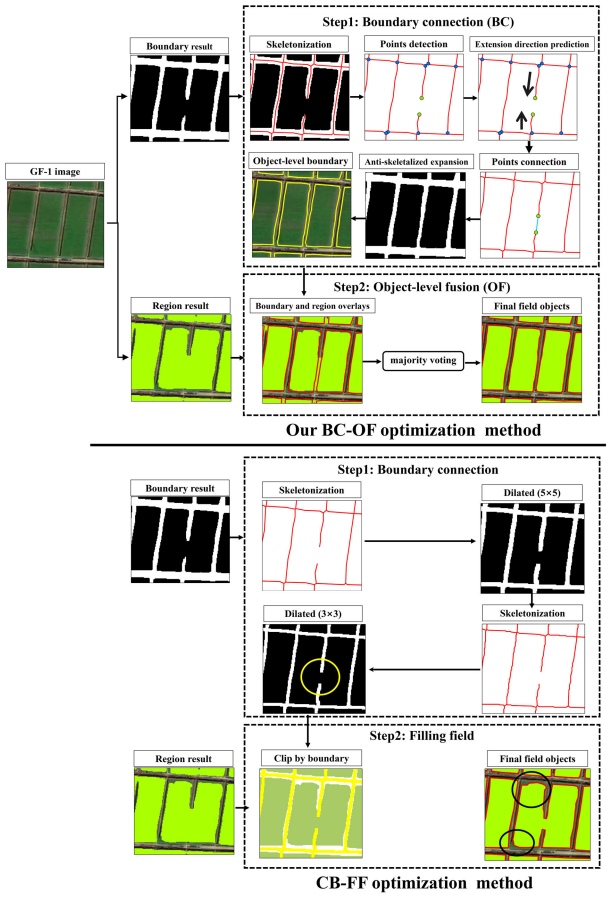


Fig. 16. Implementation process of our BC-OF optimization method and the CB-FF method used for comparison. Yellow circle indicates the limited distance extended by the CB-FF method and black circles indicate parcel irregularities resulting from inaccuracy region results.

and noncropland regions. In contrast, our method takes the boundary-generated objects as a reference, and the region results are only used to remove some noncultivated objects. The boundary results, with more detailed localization information, facilitate a more accurate and regular parcel delineation.

E. Effects of Addition Transfer Learning

Transfer learning can further optimize the transfer performance of the trained model in specific transfer regions by updating weights or adjusting input images, complementing our generalization-improved network architecture. To evaluate the effectiveness of different transfer learning techniques across various network architectures and compare the generalization performance of different network structures when applying these techniques, we conducted a cross-comparison experiment. The networks compared include our network, a variant without generalization-enhancement modules (Non-GEM), the parallel network of U-Net and DeeplabV3+ (Para-UNet-DeeplabV3+), and the recent SEANet network. The transfer learning strategies used include FADA [26] and fine-tuning techniques [27]. The transfer regions extended beyond The Netherlands to include China's Huaitai, Pingyuan, and Haifeng County (Fig. 17). Each transfer region independently trained a feature extractor based on FADA and updated

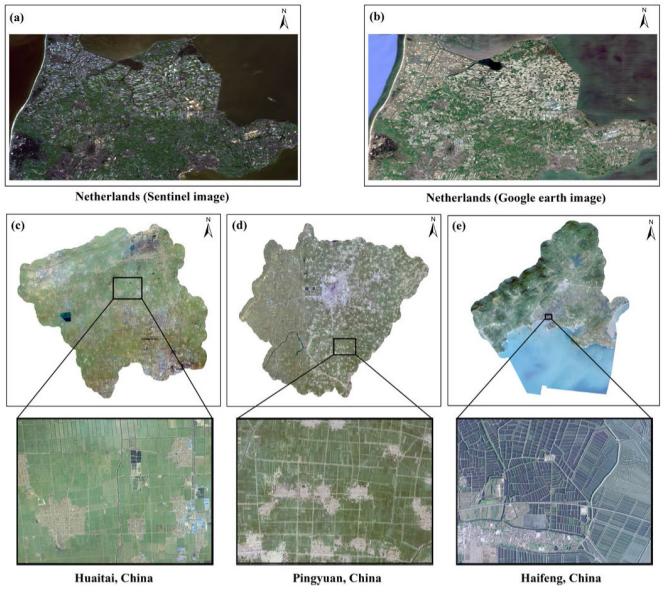


Fig. 17. Images of transfer regions that were used to validate additional transfer learning techniques. (a) and (b) Sentinel and GE images of the Nertherlands; (c) Huaitai County, China; (d) Pingyuan County, China; and (e) Haifeng County, China.

a new model based on fine-tuning with the target region's sample. Table XIII shows the delineated CLPs' area accuracy.

1) Domain Adaptation: As shown in Table XII, for our network, employing FADA to achieve image adaptation for transfer regions brought limited improvements in transfer accuracy. This indicates that our network architecture has already well generalized the recognition features, helping the model achieve stable performance on most untrained images during transfer. Therefore, in practical applications, the FADA module can serve as an optional plugin to optimize the model. However, the domain adaptation process involves the feature extractor and discriminator training, which is time-intensive. Thus, a balance must be struck between computational efficiency and performance gains in practical applications.

For the Non-GEM network, employing FADA for image adaptation brought significant accuracy gains. However, even so, the transfer performance of the Non-GEM + FADA was still far inferior to our network + FADA and it is even weaker than using our network alone without FADA for image adaptation. This demonstrates that the model trained by our network has stronger spatiotemporal generalization performance, which also maintains this enhanced performance when applying the same additional transfer learning optimizations. The cross-comparison, where our network-trained model outperforms the Non-GEM network-trained model optimized with FADA, highlights that our network's enhancements provide stronger transfer performance gains than FADA alone. This is primarily due to our improved network's ability to learn more generalizable CLP recognition features from limited samples, while FADA primarily adapted the features of a specific transfer region to the source domain. In addition, for SEANet and Para-UNet-DeeplabV3+, since there was no specialized enhancement of generalization performance in the network architecture, similar to Non-GEM, FADA also

TABLE XIII
AREA ACCURACIES OF TRANSFER AREAS USING FOUR NETWORK-TRAINED MODELS AND THE MODELS UPDATED WITH ADDITIONAL TRANSFER LEARNING

Models / Accuracy / Transfer areas	Networks (Pre-trained Models)				Networks (Pre-trained Model + **FADA**)				Networks (Pre-trained Model + **Fine-tune**)			
	Our network	Non-GEM	SEANet	Para-UNet-DeeplabV3+	Our network	Non-GEM	SEANet	Para-UNet-DeeplabV3+	Our network	Non-GEM	SEANet	Para-UNet-DeeplabV3+
Netherlands (GE)	**0.973**	0.907	0.936	0.895	**0.975**	0.929	0.954	0.951	**0.978**	0.941	0.958	0.949
Netherlands (sentinel)	**0.969**	0.896	0.931	0.873	**0.977**	0.931	0.951	0.942	**0.976**	0.943	0.954	0.943
Huaitai	**0.917**	0.795	0.829	0.801	**0.930**	0.862	0.919	0.889	**0.942**	0.906	0.927	0.926
Pingyuan	**0.909**	0.799	0.817	0.794	**0.928**	0.858	0.912	0.868	**0.940**	0.894	0.919	0.914
Haifeng	**0.763**	0.693	0.714	0.723	**0.793**	0.721	0.738	0.746	**0.901**	0.874	0.893	0.882

brought significant transfer accuracy improvements but still fell short compared to our network + FADA.

It is noteworthy that for transfer regions with agricultural landscapes that are very inconsistent with all regions in the training set, such as Haifeng County, which is dominated by paddy fields, the transfer performance of all models was poor, regardless of the addition of FADA. This highlights that both our generalization-enhancement strategies and domain adaptation technique have their limits.

2) Fine-Tuning: We further evaluated the effectiveness of fine-tuning the pretrained model based on labeled data from the transfer regions. Here, we adopted the same fine-tuning technique as Kerner et al. [27], which involves freezing the shallow layers and updating the weights of the deeper layers using labeled data. In this study, the labeled data wwere manually delineated and covered approximately 10% of each transfer image area.

As shown in Table XIII, fine-tuning the pretrained model by incorporating more accurate labeled data from the targeted transfer region significantly enhances the transfer performance, surpassing the FADA technique. This is particularly evident in areas such as Haifeng County, which exhibit significant landscape heterogeneity. For instance, in our network, after fine-tuning the model weights, the area accuracy improved from 0.763 to 0.901.

Overall, compared to existing networks, our network has ability to train more generalized models and consistently maintains superior performance when the same transfer learning techniques are applied. Based on our generalization-improved network architecture, some transfer learning strategies can be used to optimize the model performance for specific transfer scenarios. For instance, in cases without computational efficiency burdens, zero-sample domain adaptation techniques can be used to align the features of the target domain with the source domain. In situations where the transfer region possesses extremely heterogeneous landscapes, the pretrained model can be fine-tuned using manually delineated samples.

VI. DISCUSSION

In this article, we propose a high-precision, high-generalization method for delineating CLPs, validated on high-resolution imagery and medium-resolution imagery from different regions and time phases. Specifically, our method comprises a multitask detection network and a parcel optimization step. In the detection network, we constrain the concept of field boundaries as "boundaries within the cropland region," breaking down the complex detection into two conventional tasks: cropland region recognition and boundary detection. This enhances the feature interpretability of CLPs, significantly improving detection accuracy.

Combining these two detection tasks in the same network and detecting them simultaneously, i.e., multitask learning, better generalizes their respective tasks and improves accuracy through shared training signals and mutual constraints [63]. The results demonstrate that our multitask network significantly outperforms the accuracy of individual single-task networks, highlighting the mutual benefit of these tasks. Region signals help suppress boundary information outside cropland regions, while boundary signals assist in obtaining finer parcel results. In addition, we address conflicts between cropland region recognition and boundary detection tasks within a model. Our approach, distinct from the ResUnet-a network, employs separate decoders for two tasks. The boundary task uses a side architecture to capture multiscale edge semantics, improving the field boundary detection.

To further optimize the detection performance, we incorporate ASPP modules for richer spatial context in region task and add distance auxiliary prediction to constrain the region shape. In the boundary task, CDCMs enlarge the feature receptive field to better preserve the connectivity of large-span field boundaries. We opt for CDCM rather than the traditional approach of deepening the network to enlarge the receptive field because the latter tends to result in boundary details loss. In addition, an adaptive multitask loss function based on inter-task uncertainty was adopted to better coordinate the weights of different tasks and enhance result accuracy. In future research, strategies, such as replacing the ResNet backbone with more advanced transformer-based backbones, could further improve the detection performance. Introducing more auxiliary tasks could also provide more constraints for the model.

The scarcity of CLP samples places high demands on network's ability to train more generalized CLP recognition features from limited samples. To enhance it, we analyze factors affecting model generalization. We adopt CSA to simulate cultivated land's spectral heterogeneity under diverse imaging conditions (such as different times or sensors). We also employ attention mechanisms to enhance generalization. In the region task, we add AGs at skip-layer connections to highlight spatial salient features, ensuring generalization performance across different cropland landscapes. In the boundary task, we utilize

the HAFM to adaptively weight-fuse multilevel boundary predictions in various scenarios. Our experiments demonstrate that these modules effectively enhance the network's ability to train a more generalized model, which achieves higher identification accuracy in the source domain and maintains stable spatiotemporal generalization when transferred to untrained target domains. In addition, transfer learning can further optimize the performance in specific transfer scenarios. It is important to note that these transfer learning techniques do not contradict the generalization-enhancing modification of our network. Our network provides a foundational framework capable of training highly generalized models, focusing on achieving the highest possible generalization with a limited sample size. On this foundation, transfer learning can optimize the model performance for specific transfer scenarios. However, the application of these transfer learning techniques comes with certain conditions. For example, zero-sample domain adaptation techniques require a balance between limited accuracy improvement and significant computational burden, while fine-tuning techniques need labeled samples from the transfer area. Thus, practical application necessitates a careful tradeoff.

Compared to two recent multitask networks, BSiNet and SEANet, which prioritize the region task as the core and the boundary task as auxiliary to enhance the model's boundary perception, thereby outputting parcel-level region results, our method exhibits three advantages. First, the region task alone struggles to provide detailed internal delineation, necessitating the high granularity prediction offered by the boundary task to supplement its missing fine boundaries. Our strategy of simultaneously outputting regions and boundaries and integrating them at the result level is more rational. Second, to predict high-quality region and boundary results simultaneously, we design distinct decoders for each task to accommodate their different characteristics, adopting a side architecture for the boundary task to leverage its superior boundary perception capabilities. In addition, we employ an adaptive multitask loss to provide more balanced training, achieving both accurate semantic recognition and fine-grained boundary depiction. Third, based on the feature analysis of cropland regions and boundaries and improvements over existing models, our model maximizes detection performance through more effective feature encoding, introducing several performance-enhancement modules to enrich and optimize region and boundary features. It also employs CSA, distance auxiliary tasks, and other strategies to provide training constraints. These advancements help our network achieve accuracy superior to state-of-the-art methods. Furthermore, benefiting from the innovative enhancement of the network's ability to train more generalized models, our trained model can achieve a notable accuracy advantage over those trained on existing networks when transferred to the target domain.

It is essential to recognize that despite our detection model achieving sufficiently continuous boundary results, in challenging scenarios such as when boundaries are occluded or blurred, detection results may still exhibit discontinuities. Discontinuous boundaries lead to incomplete separation of neighboring parcels. Our study designed a method for further optimization, based on two observations concerning field boundaries. First, broken boundaries always possess two corresponding breakpoints, which can be easily identified based on neighborhood information of skeleton line. Second, due to the straightness of field boundaries, two breakpoints have opposite extension directions. Based on breakpoint positions and extension directions, oriented connections can be easily achieved. Compared to morphology-based optimization methods, our approach can repair longer distance breaks. In future research, the parcel optimization process could also be integrated into the network, achieving CLP delineation in an end-to-end model.

Numerous studies have proposed methods for delineating CLPs. However, a direct comparison with the reported accuracies of relevant studies remains challenging owing to the heterogeneous factors that must be considered. First, there were significant differences in the experimental setup among methods, such as image data (e.g., resolution, spectral bands, and sensors) and landscape complexity. Second, there is no standard method for assessing delineation accuracy, so the accuracies reported in previous studies cannot be directly compared. Existing studies have used numerous accuracy metrics, including mean absolute error [34], [35], [64], $F1$ score [10], [19], [65], [66], [67], OA [12], [41], [68], boundary displacement error [21], [69], and the Jaccard index [17], [70]. This highlights the need to develop a scientific accuracy validation system to provide comprehensive and fair comparisons. Furthermore, a common validation dataset would enable the systematic benchmarking of methods for future studies.

Regardless of the model's performance, deep learning-based delineation methods remain data-driven. Achieving both high accuracy and broad applicability necessitates abundant, high-quality parcel samples. However, this proves challenging within China's agricultural system, where small, frequently rotated parcels predominate. High-resolution remote sensing images can clarify land division in smaller parcels, but such images are scarce, especially in cloudy southern China. Frequent crop rotation causes field boundary changes, undermining sample consistency across imaging phases. Moreover, China has complex agricultural landscapes, and a comprehensive model requires diverse samples reflecting various topographies and agricultural patterns. Addressing sample scarcity could involve sample-free approaches such as graphical operators. For example, we can utilize traditional graphical operators (e.g., Canny and watershed) for parcel delineation, yielding a subset of usable training samples for model training. In addition, establishing a shared repository wherein individuals can contribute samples could also facilitate model development.

VII. Conclusion

This study developed a multitask and high-generalization detection network and an effective optimization method to delineate CLPs. Compared with the existing methods, notable improvements in detection accuracy, generalization performance, and optimization quality were achieved. The network enhances the feature interpretability of CLPs through multitask detection, and the network architecture accommodates the distinctions between boundary detection and region identification,

maximizing their individual performance through distinct decoders and some performance-enhancing modules. In addition, a homomorphic uncertainty-based multitask loss was adopted to coordinate intertask weights, balancing accurate semantic recognition and fine-grained boundary depiction. The developed network also improves the trained model's generalization by applying the CSA and attention mechanisms on spatial and hierarchy, providing a network architecture capable of training more generalized models. In addition, the designed optimization method can directionally repair larger distance boundary breaks, and object-level fuse boundary and region result in more regular and independent parcels. In summary, our method has robust performance and practical utility, which was validated across various temporal and geographical contexts, as well as diverse sensor imageries.

REFERENCES

[1] 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, Sep. 2019, Art. no. 111253, doi: 10.1016/j.rse.2019.111253.

[2] 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, doi: 10.1016/j.rse.2014.01.006.

[3] S. Xu et al., "A robust index to extract paddy fields in cloudy regions from SAR time series," *Remote Sens. Environ.*, vol. 285, Feb. 2023, Art. no. 113374, doi: 10.1016/j.rse.2022.113374.

[4] A. De Castro, J. Torres-Sánchez, J. Peña, F. Jiménez-Brenes, O. Csillik, and F. López-Granados, "An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery," *Remote Sens.*, vol. 10, no. 2, p. 285, Feb. 2018, doi: 10.3390/rs10020285.

[5] V. Lebourgeois, S. Dupuy, É. Vintrou, M. Ameline, S. Butler, and A. Bégué, "A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM)," *Remote Sens.*, vol. 9, no. 3, p. 259, Mar. 2017, doi: 10.3390/rs9030259.

[6] Q. Li, C. Wang, B. Zhang, and L. Lu, "Object-based crop classification with Landsat-MODIS enhanced time-series data," *Remote Sens.*, vol. 7, no. 12, pp. 16091–16107, Dec. 2015, doi: 10.3390/rs71215820.

[7] Y. Sun et al., "Geo-parcel-based crop classification in very-high-resolution images via hierarchical perception," *Int. J. Remote Sens.*, vol. 41, no. 4, pp. 1603–1624, Feb. 2020, doi: 10.1080/01431161.2019.1673916.

[8] Y. Wang, S. Fang, L. Zhao, X. Huang, and X. Jiang, "Parcel-based summer maize mapping and phenology estimation combined using Sentinel-2 and time series Sentinel-1 data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, Apr. 2022, Art. no. 102720, doi: 10.1016/j.jag.2022.102720.

[9] L. Xu et al., "Extraction of cropland field parcels with high resolution remote sensing using multi-task learning," *Eur. J. Remote Sens.*, vol. 56, no. 1, Dec. 2023, Art. no. 2181874, doi: 10.1080/22797254.2023.2181874.

[10] 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, p. 59, Dec. 2019, doi: 10.3390/rs12010059.

[11] L. Xu, D. Ming, T. Du, Y. Chen, D. Dong, and C. Zhou, "Delineation of cultivated land parcels based on deep convolutional networks and geographical thematic scene division of remotely sensed images," *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106611, doi: 10.1016/j.compag.2021.106611.

[12] B. Fetai, M. Rač ič, and A. Lisec, "Deep learning for detection of visible land boundaries from UAV imagery," *Remote Sens.*, vol. 13, no. 11, p. 2077, May 2021, doi: 10.3390/rs13112077.

[13] W. Liu et al., "Farmland parcel mapping in mountain areas using time-series SAR data and VHR optical images," *Remote Sens.*, vol. 12, no. 22, p. 3733, Nov. 2020, doi: 10.3390/rs12223733.

[14] D. Zhang et al., "A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111912, doi: 10.1016/j.rse.2020.111912.

[15] A. O. Onojeghuo, Y. Miao, and G. A. Blackburn, "Deep ResU-Net convolutional neural networks segmentation for smallholder paddy Rice mapping using Sentinel 1 SAR and Sentinel 2 optical imagery," *Remote Sens.*, vol. 15, no. 6, p. 1517, Mar. 2023, doi: 10.3390/rs15061517.

[16] W. Zhang et al., "A novel knowledge-driven automated solution for high-resolution cropland extraction by cross-scale sample transfer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4406816, doi: 10.1109/TGRS.2023.3299956.

[17] A. Taravat, M. P. Wagner, R. Bonifacio, and D. Petit, "Advanced fully convolutional networks for agricultural field boundary detection," *Remote Sens.*, vol. 13, no. 4, p. 722, Feb. 2021, doi: 10.3390/rs13040722.

[18] H. Zhang et al., "Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102557, doi: 10.1016/j.jag.2021.102557.

[19] R. Yang, Z. U. Ahmed, U. C. Schulthess, M. Kamal, and R. Rai, "Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: A case study from Bangladesh," *Remote Sens. Appl., Soc. Environ.*, vol. 20, Nov. 2020, Art. no. 100413, doi: 10.1016/j.rsase.2020.100413.

[20] X. Xia, C. Persello, and M. Koeva, "Deep fully convolutional networks for cadastral boundary detection from UAV images," *Remote Sens.*, vol. 11, no. 14, p. 1725, Jul. 2019, doi: 10.3390/rs11141725.

[21] A. García-Pedrero, M. Lillo-Saavedra, D. Rodríguez-Esparragón, and C. Gonzalo-Martín, "Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system," *IEEE Access*, vol. 7, pp. 158223–158236, 2019, doi: 10.1109/ACCESS.2019.2950371.

[22] 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, Aug. 2020, Art. no. 111741, doi: 10.1016/j.rse.2020.111741.

[23] Y. Zhu, Y. Pan, T. Hu, D. Zhang, C. Zhao, and Y. Gao, "A generalized framework for agricultural field delineation from high-resolution satellite imageries," *Int. J. Digit. Earth*, vol. 17, no. 1, pp. 1–31, Dec. 2024, doi: 10.1080/17538947.2023.2297947.

[24] 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, doi: 10.1016/j.jag.2022.102871.

[25] M. Li, J. Long, A. Stein, and X. Wang, "Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 200, pp. 24–40, Jun. 2023, doi: 10.1016/j.isprsjprs.2023.04.019.

[26] S. Liu, L. Liu, F. Xu, J. Chen, Y. Yuan, and X. Chen, "A deep learning method for individual arable field (IAF) extraction with cross-domain adversarial capability," *Comput. Electron. Agricult.*, vol. 203, Dec. 2022, Art. no. 107473, doi: 10.1016/j.compag.2022.107473.

[27] H. Kerner, S. Sundar, and M. Satish, "Multi-region transfer learning for segmentation of crop field boundaries in satellite images with limited labels," 2024, *arXiv:2404.00179*.

[28] Z. Cai et al., "Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by integrating multimodal satellite data," *ISPRS J. Photogramm. Remote Sens.*, vol. 205, pp. 34–49, Nov. 2023, doi: 10.1016/j.isprsjprs.2023.09.021.

[29] S. Wang, F. Waldner, and D. B. Lobell, "Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision," *Remote Sens.*, vol. 14, no. 22, p. 5738, Nov. 2022, doi: 10.3390/rs14225738.

[30] 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. Agricult.*, vol. 178, Nov. 2020, Art. no. 105787, doi: 10.1016/j.compag.2020.105787.

[31] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sens. Environ.*, vol. 204, pp. 509–523, Jan. 2018, doi: 10.1016/j.rse.2017.10.005.

[32] M. Turker and E. H. Kok, "Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 106–121, May 2013, doi: 10.1016/j.isprsjprs.2013.02.009.

[33] M. Mueller, K. Segl, and H. Kaufmann, "Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery," *Pattern Recognit.*, vol. 37, no. 8, pp. 1619–1628, Aug. 2004, doi: 10.1016/j.patcog.2004.03.001.

[34] B. Watkins and A. Van Niekerk, "Automating field boundary delineation with multi-temporal Sentinel-2 imagery," *Comput. Electron. Agricult.*, vol. 167, Dec. 2019, Art. no. 105078, doi: 10.1016/j.compag.2019.105078.

[35] B. Watkins and A. van Niekerk, "A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery," *Comput. Electron. Agricult.*, vol. 158, pp. 294–302, Mar. 2019, doi: 10.1016/j.compag.2019.02.009.

[36] P. Zhang, S. Hu, W. Li, and C. Zhang, "Parcel-level mapping of crops in a smallholder agricultural area: A case of central China using single-temporal VHSR imagery," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105581, doi: 10.1016/j.compag.2020.105581.

[37] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403, doi: 10.1109/ICCV.2015.164.

[38] 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, doi: 10.1109/TPAMI.2018.2878849.

[39] S. Marvaniya, U. Devi, J. Hazra, S. Mujumdar, and N. Gupta, "Small, sparse, but substantial: Techniques for segmenting small agricultural fields using sparse ground data," *Int. J. Remote Sens.*, vol. 42, no. 4, pp. 1512–1534, Feb. 2021, doi: 10.1080/01431161.2020.1834166.

[40] D. K. Gopidas et al., "Integrated deep learning based segmentation and classification method for boundary delineation of agricultural fields in multitemporal satellite images," *Int. J. Mod. Agric.*, vol. 10, no. 2, pp. 1804–1822, 2021.

[41] F. Waldner et al., "Detect, consolidate, delineate: Scalable mapping of field boundaries using satellite images," *Remote Sens.*, vol. 13, no. 11, p. 2197, Jun. 2021, doi: 10.3390/rs13112197.

[42] 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-Geoinf. (Agro-Geoinf.)*, Aug. 2018, pp. 1–5, doi: 10.1109/Agro-Geoinformatics.2018.8476002.

[43] R. Hong, J. Park, S. Jang, H. Shin, H. Kim, and I. Song, "Development of a parcel-level land boundary extraction algorithm for aerial imagery of regularly arranged agricultural areas," *Remote Sens.*, vol. 13, no. 6, p. 1167, Mar. 2021, doi: 10.3390/rs13061167.

[44] P. Hao, L. Di, C. Zhang, and L. Guo, "Transfer learning for crop classification with cropland data layer data (CDL) as training samples," *Sci. Total Environ.*, vol. 733, Sep. 2020, Art. no. 138869, doi: 10.1016/j.scitotenv.2020.138869.

[45] L. Xun, J. Zhang, F. Yao, and D. Cao, "Improved identification of cotton cultivated areas by applying instance-based transfer learning on the time series of MODIS NDVI," *CATENA*, vol. 213, Jun. 2022, Art. no. 106130, doi: 10.1016/j.catena.2022.106130.

[46] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, and D. Warde-Farley, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 1–9.

[47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. CVPR*, Jun. 2018, pp. 7132–7141. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html

[48] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[50] D. Jha et al., "A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2029–2040, Jun. 2021.

[51] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[52] V. Singh, V. Devgan, and I. Anand, "Determining image similarity with quasi Euclidean metric," 2018, *arXiv:2006.14644*.

[53] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5872–5881, doi: 10.1109/CVPR.2017.622.

[54] A. A. Novikov, D. Lenis, D. Major, J. Hladuvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1865–1876, Aug. 2018, doi: 10.1109/TMI.2018.2806086.

[55] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7482–7491.

[56] H. Wang, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "MaX-DeepLab: End-to-end panoptic segmentation with mask transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5459–5470, doi: 10.1109/CVPR46437.2021.00542.

[57] B. Cheng et al., "Panoptic-DeepLab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12472–12482, doi: 10.1109/CVPR42600.2020.01249.

[58] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sens.*, vol. 11, no. 21, p. 2586, Nov. 2019, doi: 10.3390/rs11212586.

[59] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.

[60] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[61] I. Lizarazo, "Accuracy assessment of object-based image classification: Another STEP," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6135–6156, Aug. 2014, doi: 10.1080/01431161.2014.943328.

[62] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[63] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.

[64] H. P. Meyer and A. Van Niekerk, "Assessing edge and area metrics for image segmentation parameter tuning and evaluation," in *Proc. 6th Int. Conf. Geographic Object-Based Image Anal.*, 2016, pp. 1–4, doi: 10.3990/2.440.

[65] M. P. Wagner and N. Oppelt, "Deep learning and adaptive graph-based growing contours for agricultural field extraction," *Remote Sens.*, vol. 12, no. 12, p. 1990, Jun. 2020, doi: 10.3390/rs12121990.

[66] J. Graesser and N. Ramankutty, "Detection of cropland field parcels from Landsat imagery," *Remote Sens. Environ.*, vol. 201, pp. 165–180, Nov. 2017, doi: 10.1016/j.rse.2017.08.027.

[67] S. Crommelinck, M. Koeva, M. Y. Yang, and G. Vosselman, "Application of deep learning for delineation of visible cadastral boundaries from remote sensing imagery," *Remote Sens.*, vol. 11, no. 21, p. 2505, Oct. 2019, doi: 10.3390/rs11212505.

[68] O. Vlachopoulos, B. Leblon, J. Wang, A. Haddadi, A. LaRocque, and G. Patterson, "Delineation of crop field areas and boundaries from UAS imagery using PBIA and GEOBIA with random forest classification," *Remote Sens.*, vol. 12, no. 16, p. 2640, Aug. 2020, doi: 10.3390/rs12162640.

[69] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," in *Computer Vision—ECCV* (Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2352. Berlin, Germany: Springer, 2002, pp. 408–422, doi: 10.1007/3-540-47977-5_27.

[70] G. O. Tetteh, A. Gocht, and C. Conrad, "Optimal parameters for delineating agricultural parcels from satellite images based on supervised Bayesian optimization," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105696, doi: 10.1016/j.compag.2020.105696.

Yu Zhu received the B.S. degree in geography from China University of Geosciences, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree in remote sensing science and technology with the State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing.

His research interests include deep learning-based image processing and agricultural intelligent remote sensing.

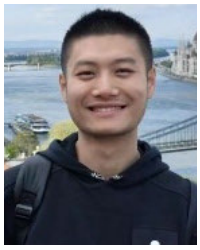
Yaozhong Pan received the B.Sc. degree in remote sensing technology from Zhejiang University, Hangzhou, China, in 1988, and the M.S. and Ph.D. degrees in physical geography from Beijing Normal University, Beijing, China, in 1994 and 1997, respectively.

He is currently a Professor with Beijing Normal University, where he is currently the Dean of the Institute of Remote Sensing Science and Engineering. He has authored or co-authored more than 150 peer-reviewed international journal articles. His research interests include agricultural intelligent remote sensing and real-time remote sensing of surface anomalies.

Dr. Pan has won several research awards, including the Precursor of the Science and Technology Production Transformation in Beijing, awarded by the Beijing Economy Committee, Beijing Science and Technology Committee, and Beijing Education Committee in 2002; the First Prize of the National Natural Science Award superscripted by the Ministry of Education in 2004; and the 8th National Youth Geographic Science and Technology Award issued by the Ministry of Education in 2005.

Dujuan Zhang received the B.S. degree in geography from Wuhan University, Wuhan, China, in 2016, and the Ph.D. degree in geography from Beijing Normal University, Beijing, China, in 2022.

She is currently a Lecturer with the National Supercomputing Center in Zhengzhou, Zhengzhou University, Zhengzhou, China. Her research interests include remote sensing big data and machine learning, with a focus on deep learning and its applications in remote sensing.



Hanyi Wu (Graduate Student Member, IEEE) received the B.S. degree in remote sensing science and technology from Nanjing University of Information Science and Technology, Nanjing, China, in 2022. He is currently pursuing the M.S. degree in cartography and geographic information system with the State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, China.

His research interests include remote sensing image processing, urban remote sensing, health risk assessment, and agricultural intelligent remote sensing.

Chuanwu Zhao (Graduate Student Member, IEEE) was born in Xinyang, Henan, China, in 1995. He is currently pursuing the Ph.D. degree in cartography and geographic information system with the State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, China.

His research interests include remote sensing image processing, urban remote sensing, and vegetation anomaly detection.