ISPRS Journal of Photogrammetry and Remote Sensing 221 (2025) 1–19

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A large-scale VHR parcel dataset and a novel hierarchical semantic boundary-guided network for agricultural parcel delineation

Hang Zhao [a,b], Bingfang Wu [a,b,*], Miao Zhang [a,**], Jiang Long [c], Fuyou Tian [a], Yan Xie [a,b], Hongwei Zeng [a,b], Zhaoju Zheng [a], Zonghan Ma [a], Mingxing Wang [a,d], Junbin Li [a,b]

[a] Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101 China
[b] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049 China
[c] Key Lab of Spatial Data Mining & Information Sharing of Education, Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350108 China
[d] College of Resources and Environment, Hubei University, Wuhan 430062 China

ARTICLE INFO

Keywords:
Agricultural parcel delineation
Very high-resolution remote sensing images
Multitask neural networks
Boundary-guided
FHAPD

ABSTRACT

Current agricultural parcels (AP) extraction faces two main limitations: (1) existing AP delineation methods fail to fully utilize low-level information (e.g., parcel boundary information), leading to unsatisfactory performance under certain circumstances; (2) the lack of large-scale, high-resolution AP benchmark datasets in China hinders comprehensive model evaluation and improvement. To address the first limitation, we develop a hierarchical semantic boundary-guided network (HBGNet) to fully leverage boundary semantics, thereby improving AP delineation. It integrates two branches, a core branch of AP feature extraction and an auxiliary branch related to boundary feature mining. Specifically, the boundary extract branch employs a module based on Laplace convolution operator to enhance the model's awareness of parcel boundary. For AP feature extraction, a local–global context aggregation module is designed to enhance the semantic representation of AP, improving the adaptability across different AP scenarios. Meanwhile, a boundary-guided module is developed to enhance boundary details of high-level AP semantic information. Ultimately, a multi-grained feature fusion module is designed to enhance the capacity of HBGNet to extract APs with various sizes and shapes. Regarding the second limitation, we construct the first large-scale very high-resolution (VHR) agricultural parcel dataset (FHAPD) across seven different areas, covering more than 10,000 km^2, using data from GaoFen-1 (2-meter) and GaoFen-2 (1-meter). Detailed experiments are conducted on the FHAPD, a publicly European dataset (i.e., Al4boundaries), and medium-resolution Sentinel-2 images from the Netherlands and HBGNet is compared with other eight AP delineation methods. Results show that HBGNet outperforms the other eight methods in attribute and geometry accuracy. The Intersection over Union (IOU), F1-score of the boundary (F_{bdy}), and global total-classification (GTC) exceed other methods by 0.61 %-7.52 %, 0.8 %-36.3 %, and 1.7 %-31.8 %, respectively. It also effectively transfers to unseen regions. We conclude that the proposed HBGNet is an effective, advanced, and transferable method for diverse agricultural scenarios and remote sensing images.

1. Introduction

Agricultural parcels (APs) are fundamental units for agricultural activities, the accurate delineation of which is essential for efficient resource management and critical agricultural applications, including crop mapping, yield estimation, and sustainable agricultural planning (Belgiu and Csillik 2018; Waldner and Diakogiannis 2020). Currently, there are several AP datasets at regional- and national scales (Boryan et al. 2011; Teluguntla et al. 2018). However, traditional AP delineation methods rely heavily on time-consuming and laborious field surveys or manual interpretation of very-high-resolution (VHR) remote sensing images, which can hinder timely updating and maintaining affordability in accurate cropland monitoring (Hasituya et al. 2017; Liu et al. 2020; Rydberg and Borgefors 2001). Fortunately, the increasing availability of satellite imagery across diverse spatial, temporal, and spectral resolutions has unprecedentedly promoted the development of automatic AP

* Corresponding author at: Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China.
** Co corresponding author.
E-mail addresses: zhaohang201@mails.ucas.ac.cn (H. Zhao), wubf@radi.ac.cn (B. Wu).

https://doi.org/10.1016/j.isprsjprs.2025.01.034
Received 5 July 2024; Received in revised form 16 December 2024; Accepted 27 January 2025
Available online 3 February 2025
0924-2716/© 2025 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

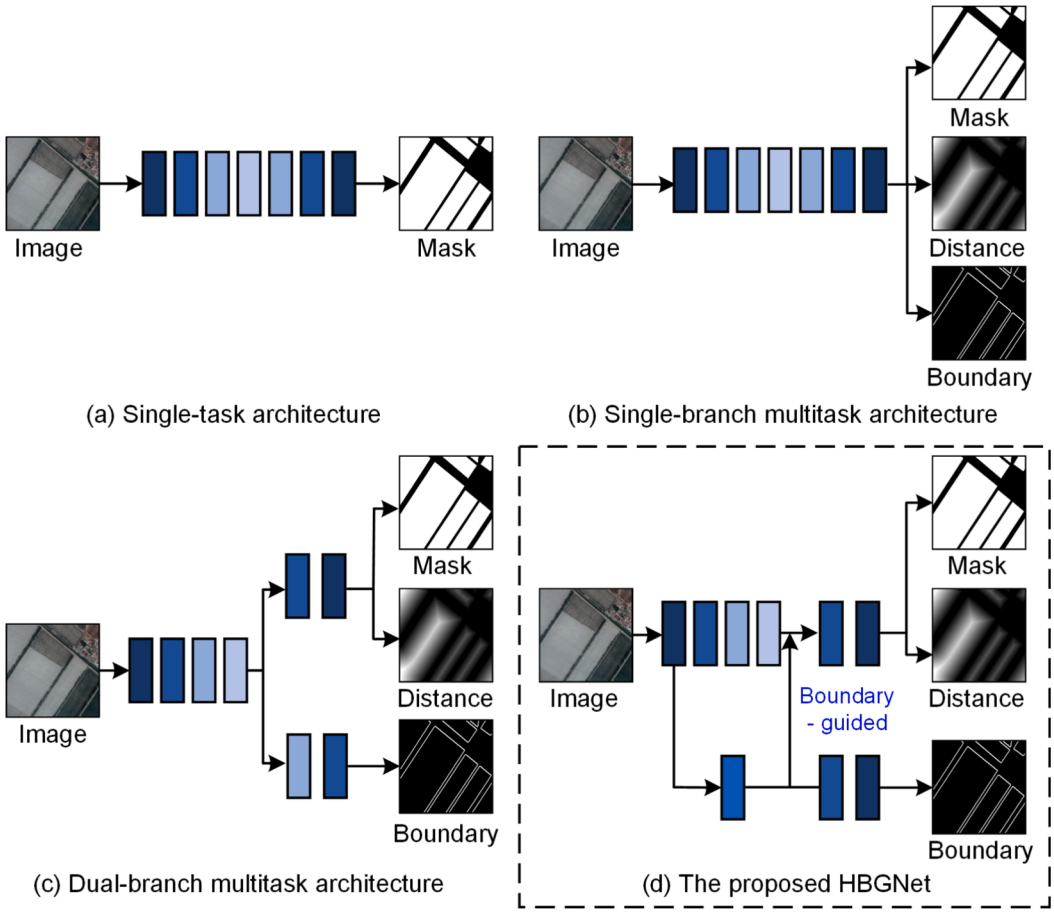


Fig. 1. Three AP delineation types according to the model structure: (a) single-task structure, (b) single-branch architecture of multitask method, (c) dual-branch architecture of multitask method, and (d) our structure, hierarchical semantic boundary-guided network (HBGNet).

extraction techniques (Jong et al. 2022; Persello et al. 2019; Tang et al. 2020).

Traditional AP delineation methods are mainly divided into three categories: edge-based, region-based, and hybrid methods. Edge-based methods focus on AP boundary detection by using different operators like Sobel, Canny, and Laplace filters edge operators to emphasize discontinuity (gradient) between pixels (Robb et al. 2020; Singh and Singh 2015). Nonetheless, these methods limit the utilization of high-level edge features, resulting in high sensitivity to high-frequency noise and incorrect parcel boundaries. To solve this problem, region-based methods utilize self-defined homogeneity criteria to adaptively split and merge similar pixels into larger regions (Nock and Nielsen 2004). They yield spectral intensity, textural, geometrical, and spatial properties of object-level APs based on conventional image segmentation techniques, such as watershed, multiresolution, and mean-shift segmentation (Cai et al. 2022; Garcia-Pedrero et al. 2017; Lebourgeois et al. 2017). However, region-based approaches often neglect surroundings and inter-object heterogeneity in generating optimal image objects and the segmentation parameters depend heavily on experience (Hay et al. 2003; Hossain and Chen 2019). Some researchers that attempt to integrate edge-based and region-based methods to improve AP delineation performance (Cheng et al. 2020; Rydberg and Borgefors 2001; Watkins and Van Niekerk 2019). Mueller et al. (2004) implement region segmentation under the guidance of the detected gradient layer to obtain object-level parcel results. Rydberg and Borgefors (2001) use a gradient detector with the magnitudes and direction of edges, and combine the clustering algorithm to delineate APs.

Currently, deep neural networks have been widely used in extracting thematic information due to their powerful feature extraction capabilities (Li et al. 2023; Long et al. 2022; Lu et al. 2022; Waldner and Diakogiannis 2020). Fig. 1 shows two common architectures of deep learning models for AP extraction, i.e., single task and multitask models. Single-task networks typically use an encoder-decoder architecture, where the encoder captures multi-scale semantic features, and the decoder subsequently refines these gathered features (Taravat et al. 2021; Zhang et al. 2020a; Zhang et al. 2021). Although these approaches produce satisfactory results, they tend to overlook the benefits of intrinsically related tasks that share optimal hypothesis classes, which can enhance generalization and reduce overfitting risk (Ruder 2017). Therefore, researchers have increasingly adopted multitask architectures that allow for joint representation learning across multiple related tasks (Li et al. 2023; Long et al. 2022; Waldner and Diakogiannis 2020). Multitask architectures are usually divided into single-branch (Fig. 1(b)) and dual-branch models (Fig. 1(c)). Single-branch architecture directly feeds the images into the model to simultaneously extract AP and AP boundary (or distance) (Lin et al. 2024; Taravat et al. 2021; Zhang et al. 2020a; Zhang et al. 2021). For example, Long et al. (2022) introduced a lightweight multitask network (BsiNet) based on a singular encoder-decoder architecture using the single-branch structure, which improves the geometric accuracy of APs in hilly terrains. Nevertheless, such method fails to fully consider the multi-level boundary semantics, resulting in the incomplete and unclosed parcel boundaries. To deal with these deficiencies, the dual-branch architecture perform the tasks of AP boundary detection and AP delineation separately, using a shared encoder but distinct decoders, which improves the representation of boundary features, thus enhancing the accuracy of AP delineation in diverse areas (Li et al. 2023; Liu et al. 2022; Lu et al. 2024; Shunying et al. 2023). For example, Li et al. (2023) developed a semantic edge-

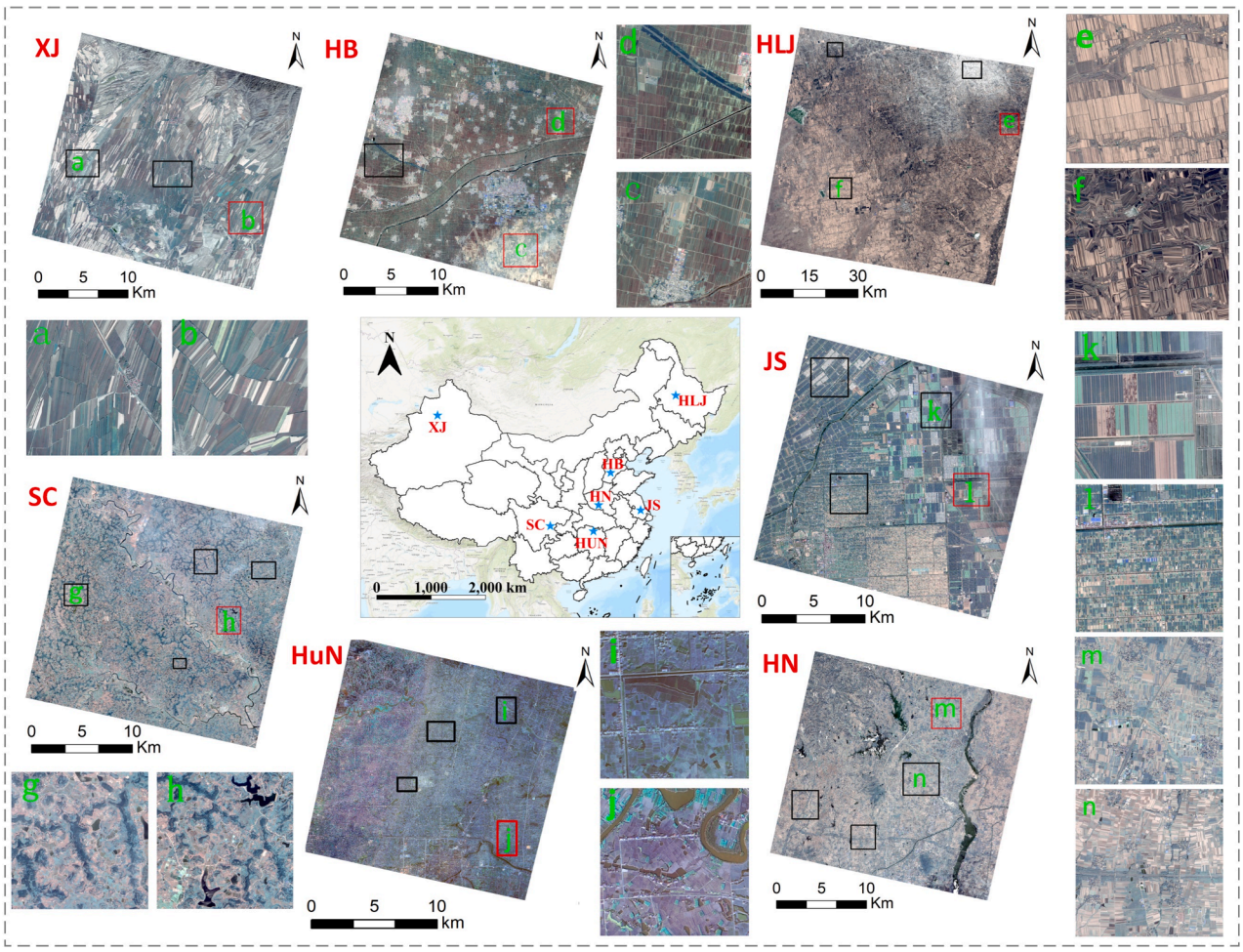


Fig. 2. Overview of the FHAPD including seven areas (a-n refers to AP examples and the black box represents the training and validation areas, and the red box represents the testing area).

aware multitask network (SEANet) that comprises a dense edge detection branch and a parcel mask extraction branch to capture multi-level edge semantics, evidently improving the extraction of parcels in varied agricultural landscapes. However, such approaches have drawbacks of increased computation loads due to overemphasis on high-level features in boundary branch.

Instance segmentation methods based on multitask learning technologies for AP delineation are also effective (Pan et al. 2023; Xu et al. 2023; Zhong et al. 2023), which assigns semantic labels and instance identifiers to all pixels (De Brabandere et al., 2017; Fathi et al. 2017). AP delineation based on instance segmentation is mainly divided into two categories, i.e., top-down and bottom-top. For the top-down architecture, the location of each AP instance box is located by object detection, and then the AP instance is obtained by semantic segmentation inside the box, such as Mask R-CNN (He et al. 2017), DAFPN (Cao et al. 2023), E2EVAP (Pan et al. 2023), etc. However, AP segmentation accuracy is limited by the accuracy of box detection, and AP delineation results may weaken in complex landscapes. For the bottom-top architecture of instance segmentation, the pixel classification is first performed by semantic segmentation, and then AP instances are obtained by clustering (Brabandere et al., 2017; Li et al. 2017). These methods usually require cumbersome post-processing steps and filter out some small and narrow AP instances.

Despite the progress made in AP delineation, there are still several significant limitations. First, existing multitask methods mainly use convolutional neural networks (CNNs) as the model backbone, which struggles to model the long-distance contextual dependencies among objects due to the locality of convolution operations and leads to low extraction accuracy in complex scenarios. Second, existing methods commonly ignore the role of low-level semantics (e.g., boundary) in refining the spatial details of high-level semantics, leading to irregular parcel boundaries. Third, existing model architectures with skip connections may fail to retain vital multi-scale contextual information, which reduces the ability to accurately delineate APs across diverse scenarios. Finally, accurately delineate APs with irregular shapes, fuzzy boundaries, or small sizes remains a significant challenge for existing methods.

Moreover, another problem lies in the availability of datasets with high-resolution images. To our knowledge, no large-scale VHR AP dataset is currently available within China. Although there are multiple available land cover products including the AP category with medium spatial resolution, they are not sufficiently detailed and have limited coverage. Current algorithms and datasets, particularly VHR AP datasets, are not readily accessible as open-source resources, leading to a scarcity of fundamental data. This data scarcity hampers the comparative experiments across various methods and stifles innovation in related techniques. Therefore, an urgent need exists for a large-scale AP dataset with high spatial resolution.

This study aims to develop an efficient multitask semantic segmentation network for delineating APs with fine boundary details from VHR remote sensing images. It presents a hierarchical semantic boundary-guided network (i.e., HBGNet) to focus on utilizing spatially explicit low-level boundary information to constrain the detection of high-level parcel features. Initially, an effective global and local context extraction

Table 1
Details of the study areas on the FHAPD.

Region	Satellites	Dates	AP number	Average areas	Areas (km^2)	Image sizes
HLJ	GF1	20,201,113	8238	15136 m^2	6735.68	42266 × 39841
HB	GF2	20,200,602	13,923	4361 m^2	760.21	27924 × 27224
HN	GF2	20,200,919	19,165	1748 m^2	810.01	29200 × 27740
JS	GF2	20,200,507	19,656	3331 m^2	771.98	28076 × 27496
SC	GF2	20,200,505	14,396	732 m^2	741.43	27536 × 26926
HuN	GF2	20,181,006	19,526	613 m^2	758.95	29659 × 26522
XJ	GF2	20,200,921	9757	32316 m^2	865.80	30163 × 28704

module is employed to improve the model's adaptability under diverse AP scenarios. Subsequently, a boundary information extraction branch is constructed to fully explore the spatial explicit AP boundary information. Further, the AP delineation capability is improved by constraining high-level AP semantic information with boundary features. Finally, by integrating multi-grained information related to APs, HBGNet can be adapted to detect APs with various sizes and shapes. The main contributions of this study include:

It develops a novel parcel delineation architecture that jointly learns multiple related task features. A Local-Global Contextual Aggregation (LGCA) module is developed to comprehensively utilize local and global spatial information, improving the perception ability of APs, particularly in complex terrain scenarios like mountain and hilly.

It proposes a Boundary Guided Module (BGM) that fosters semantic interaction between boundary features and high-level AP features, thus alleviating boundary fuzzy of extracted parcels.

It introduces a Multi-Grained feature Fusion Module (MGFM) to augment the semantic representation of discrepant semantic information, bolstering the model's capability to extract APs with small sizes effectively.

It constructs the first large-scale high-resolution AP dataset (i.e., FHAPD) covering seven different agricultural scenarios of China, which was collected from VHR remote sensing images for model comparison. The code and datasets is available at https://github.com/NanNanmei/HBGNet.

2. Datasets and study areas

2.1. FHAPD: The first large-scale high-resolution remote sensing AP dataset in China

The indispensable role of a substantial dataset for training has been thoroughly substantiated within the domain of deep learning (Götz et al. 2022; Sambasivan et al. 2021). However, there is a lack of publicly available large-scale parcel datasets, particularly those containing VHR remote sensing imagery with fine spatial and texture information. To fill the gap in VHR APs dataset, this study constructs the first large-scale VHR remote sensing AP dataset in China, named FHAPD. FHAPD includes seven agricultural scenarios across different provinces in China, collected by GaoFen-1 (GF1) GaoFen-2 (GF2) satellites (Fig. 2). Detailed information of FHAPD is given in Table 1. We manually delineated the ground-truth APs for the FHAPD dataset, which includes over 100,000 individual APs. Additionally, subset areas were selected for training and testing our model, as indicated by the black and red boxes in Fig. 2. The landscape shape index, fractal dimension and patch density were selected to characterize landscape pattern using FRAGSTATS version 4.2 with eight-neighbor rule (McGarigal and Marks 1995). Please refer to Table S1 and Table S2 in the supplementary materials.

The Gram-Schmidt pan-sharpening process is implemented to integrate the multispectral imagery from GF1 and GF2 satellites with their panchromatic bands respectively. We then resampled the pan-sharpened images from GF1 to a resolution of 2 m and from GF2 to 1 m. The dataset covers an area of more than 10,000 km^2, featuring a diverse range of

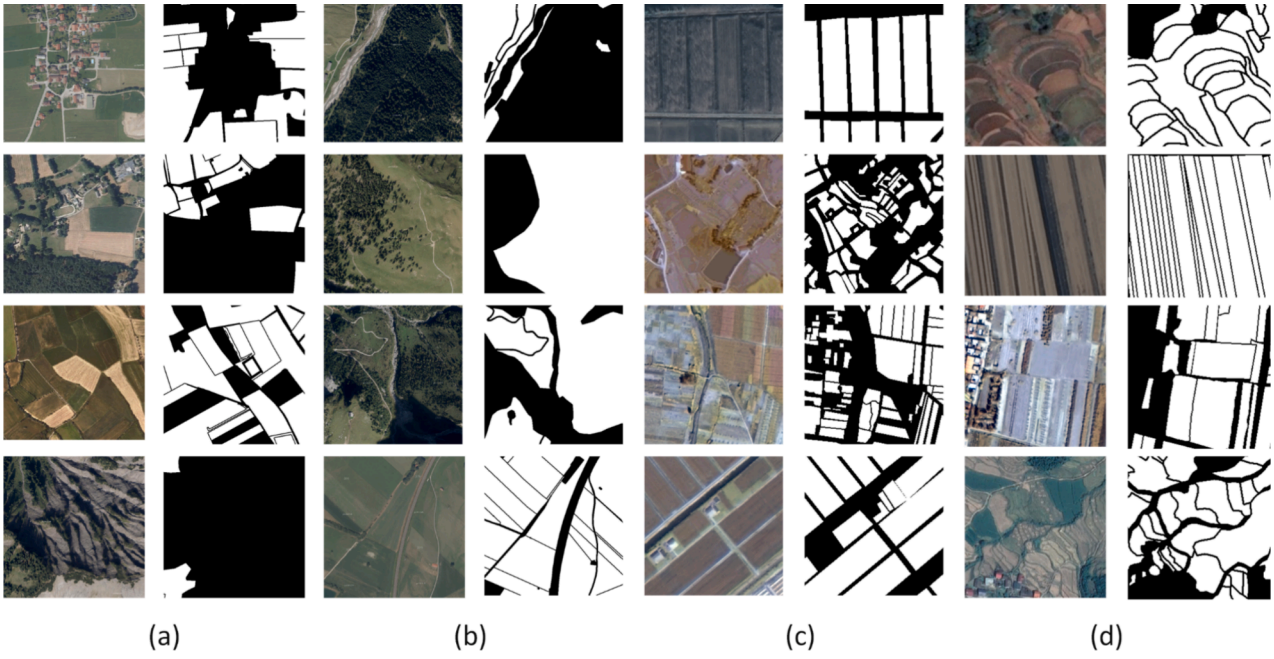


Fig. 3. Parcel examples on the FHAPD and AI4Boundaries datasets (a and b are the parcel examples from the AI4Boundaries dataset, and c and d are the parcel examples from the FHAPD).

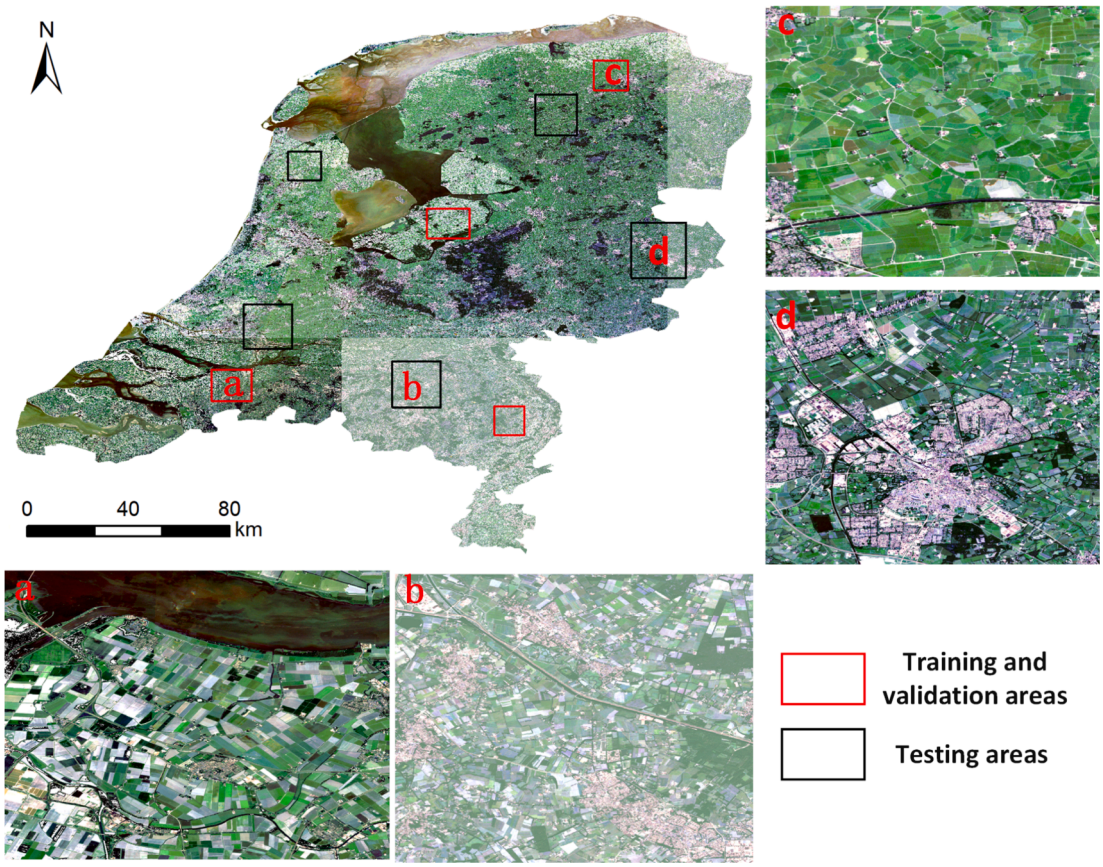


Fig. 4. Overview of Netherlands Sentinel-2 dataset. (a-d refers to the examples of typical agricultural parcels).

agricultural landscapes—including plains, hills, and mountains—a diversity of crop types across different climatic regions.

FHAPD covers seven different districts (Fig. 2), i.e., Heilongjiang (HLJ), Hebei (HB), Henan (HN), Jiangsu (JS), Sichuan (SC), Hunan (HuN), and Xinjiang (XJ). The Xinjiang area is in northwest China, where APs are large, flat, and have clear boundaries. The image here is acquired during the cotton harvest season. The Hebei area is in the North China Plain, where APs are small but precisely delineated, and the images were taken during the ripening of winter wheat. The Henan area, also in the North China Plain, was selected to capture a diverse landscape, includes plains, mountains, and hills. Cropland in Henan is predominantly used for planting corn and rapeseed, resulting in small, irregularly shaped parcels. The Heilongjiang area features a mix of plain and hilly landscapes, with large, flat APs that have clear boundaries. The Jiangsu area displays small, regularly shaped APs, where diverse crops are cultivated here. The Sichuan and Hunan areas are both typical mountainous agricultural areas, featuring small, fragmented, and irregularly shaped APs. Effectively identifying and mapping these smallholder parcels from satellite imagery is widely recognized as a challenging task (Long et al. 2022).

2.2. Public AI4Boundaries dataset

AI4Boundaries is an open access AP dataset with 1-m orthophoto images of 512 by 512 pixels (d'Andrimont et al. 2023). The labels are derived from public parcel data of Austria, Catalonia, France, Luxembourg, Netherlands, Slovenia, and Sweden, provided by the Geospatial Aid Application. In AI4Boundaries, APs are homogeneous and large, especially in the plains (Fig. 3(a) and (b)). In contrast, the APs in FHAPD outline boundaries across plains, hills, and mountains with greater precision. Furthermore, FHAPD includes a variety of cropland types such as drylands, terraces, and paddy fields.

2.3. Netherlands Sentinel-2 dataset

The dataset from the Netherlands encompasses five cloud-free 10-m Sentinel-2 Level-2A surface reflectance products with true-color composites, which are also selected for model comparison. The original imagery is downloaded from Google Earth Engine (GEE) and covers almost the entire country (i.e., 41,528 km^2) with a size of 42,845 × 31,580 pixels. Each cloud-free image is mosaicked from images taken between May 1 to October 1, 2020 (Li et al., 2023). These images cover an area of more than 64,800 km^2, providing sufficient material to evaluate our method in diverse agricultural scenarios. The subset areas (red and black boxes in Fig. 4) are used to train and test our method, and the training and validation subsets are randomly partitioned from designated areas in an 8:2 ratio.

3. Methods

3.1. HBGNet architecture

Fig. 5 illustrates the architecture of the proposed method (i.e., HBGNet), which focuses on three tasks: a core task of AP mask extraction, and two auxiliary tasks involving the parcel distance map estimation and boundary extraction. The distance map is generated from the mask label using Euclidean distance transformation, representing the distance from the internal pixels of APs to the nearest boundary. This improves the geometric accuracy of the model by providing the topological connectivity of APs and alleviating isolated segmentation errors (Diakogiannis et al. 2020; Long et al. 2022). Specifically, HBGNet consists of two branches: the agricultural parcel feature extraction (APFE) branch and the parcel boundary feature extraction (PBFE) branch. The PBFE branch is designed to capture precise boundary information by leveraging fine-grained spatial details from low-level semantics, while

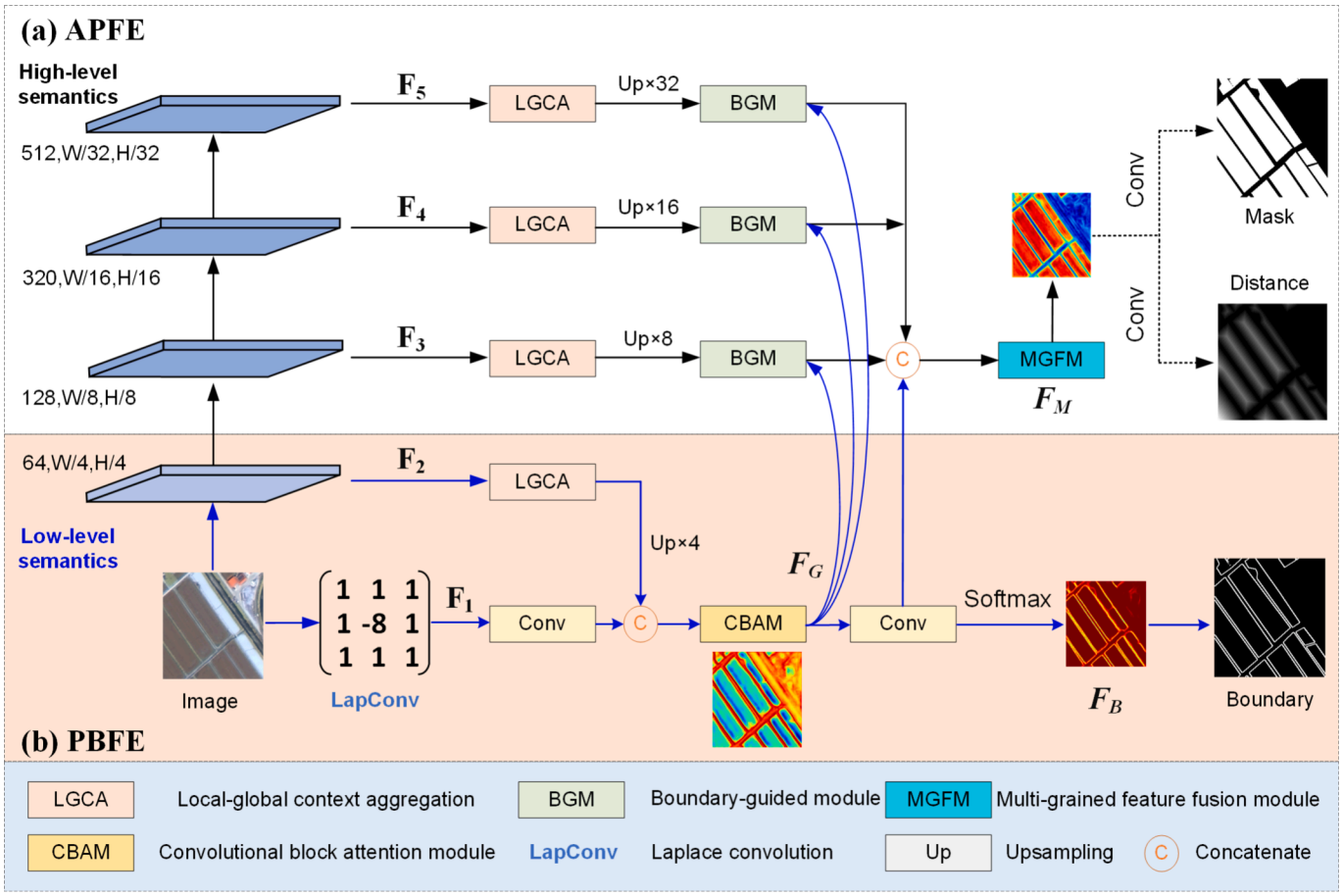


Fig. 5. Architecture of the proposed AP delineation using HBGNet from VHR images. It consists of two branches, corresponding to APFE and PBFE.

the APFE branch focuses on extracting high-level semantic features of APs. Our method efficiently utilizes boundary features from PBFE to guide and constrain multi-level features from APFE.

The Pyramid Vision Transformer (PVT-V2) pre-trained on the ImageNet dataset is employed as the backbone of HBGNet (Wang et al. 2022b). PVT-V2 adopts a pyramidal architecture with a spatially reduced attention module, which enables the effective capture of features at different levels in an image while maintaining low computational loads. For an input image $I \in R^{H\times W\times C}$, (where H, W and C denote heights, widths and the number of channels, respectively). Laplace convolution is firstly used to extract AP boundary features (i.e., F_1). And then, multi-level features $F_i \in R^{(H/2^i)\times(W/2^i)\times C_i}$ are derived from PVT-V2 (Fig. 5) using the overlapping patch embedding and Transformer encoder, where $C_i \in \{64, 128, 320, 512\}$ and $i \in \{2, 3, 4, 5\}$. F_1 and F_2 contain fine-grained AP spatial features (Fig. 5b), while F_3 to F_5 provide high-level semantic features (Fig. 5a) to assist in recognizing

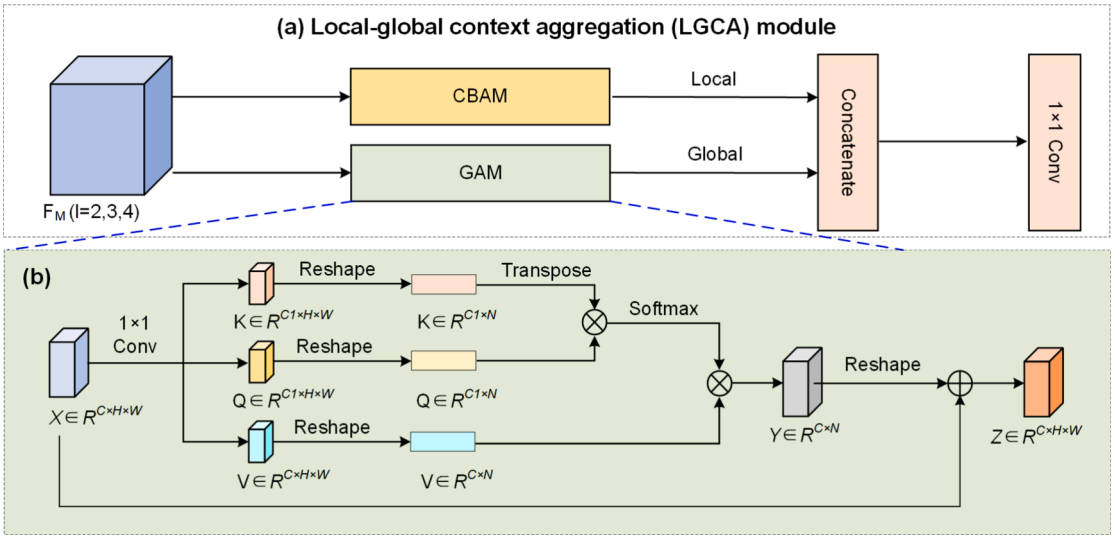


Fig. 6. The structure of LGCA module for enhancing the representation capability of multi-scale contextual information.

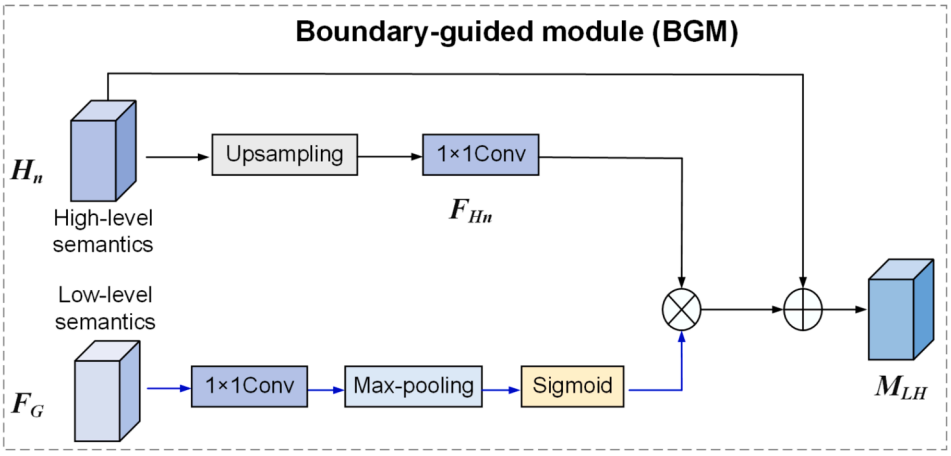


Fig. 7. Architecture of the proposed BGM for modeling the interaction of low-level and high-level semantics features.

abstract APs. These multi-level semantic features derived from PVT-V2 (i.e., F_2, F_3, F_4 and F_5) are further improved using the Local-global context aggregation (LGCA) module, enhancing the representation capability of multi-scale contextual information. For fine-grained features F_1, spatial details are enhanced by a convolutional block attention module (CBAM) (Woo et al. 2018). In addition, the boundary-guided module (BGM) leverages fine-grained semantics derived from low-level (e.g., PBFE) to guide and constrain high-level semantics (e.g., APFE), compensating for the loss of boundary information in high-level semantics. Finally, HBGNet aggregates all multi-level features and introduces the multi-grained feature fusion module (MGFM) to integrate discrepant semantic features from multi-levels, improving the extraction of AP with various shapes and sizes.

3.1.1. Local and global semantic information integration

Due to variability in spectra and seasons, APs often exhibit significant intra-class variation and inter-class similarity, posing challenges for accurate delineation. Utilizing multi-scale contextual information facilitates the identification of ground objects with various appearances, particularly those exhibiting similar spectral and textural characteristics (Long et al. 2024a). Therefore, the LGCA module is proposed to fuse global and local contextual features. The LGCA module consists of global attention module (GAM) and CBAM (Woo et al. 2018):

$$F_G^C = f^{1\times1}(Cat(I_{GAM}(F_M), I_{CBAM}(F_M))) \tag{1}$$

where $F_M(M = 2, 3, 4, 5)$ indicates semantic features of different levels and F_G^C denotes the fused features. $f^{1\times1}$ and *Cat* indicate a 1 × 1 convolutional and concatenation operation, respectively. I_{GAM} and I_{CBAM} mean the GAM and CBAM operations, respectively.

Conventional convolution operations struggle to effectively extract global contextual information of the target object at spatial scale due to its local characteristics. Inspired by the non-local network, the GAM is proposed to leverage long-range contextual dependencies of target objects based on a self-attention mechanism (Fu et al. 2019; Shaw et al. 2018) to make up for the inadequate expression of global information of CBAM at the spatial level. Specifically, as shown in Fig. 6, given a feature map $X \in R^{C\times H\times W}$, where *C*, *H*, and *W* represent the number of channels, heights, and widths, respectively, and a $f^{1\times1}$ is applied firstly to independently produce two feature maps *K* and *Q*, respectively, where $\{K, Q\} \in R^{C\times H\times W}$, then reshape them to $R^{C\times N}$, where $N = H \times W$. Subsequently, a matrix multiplication is conducted on the transpose of *K* and *Q*. Following this, a *SoftMax* layer is applied to derive the spatial attention matrix $S^{N\times N}$, which Eq.(2) indicates the result of each pixel.

$$S_{ji} = \frac{\exp(K_i \cdot Q_j)}{\sum_{i=1}^{N} \exp(K_i \cdot Q_j)} \tag{2}$$

where S_{ji} measures the i^{th} position's impact on j^{th} position.

Meanwhile, we use $f^{1\times1}$ and reshape *X* to $V^{C\times N}$. After that, a matrix multiplication is performed between *V* and *S*. Let reshaped *Y* add *X* to obtain $Z \in R^{C\times H\times W}$, and enhance representation by encoding context information across the horizontal, vertical, left diagonal, and right diagonal directions (Fig. 6b). The global context I_{GAM} for the input features is defined as follows:

$$I_{GAM} = \left((V^R) \times \left(Softmax\left(\left((K^R)^T\right) \times (Q^R)\right)\right)\right)^R + X \tag{3}$$

where *R* and *T* indicate reshape and transpose, respectively.

3.1.2. Parcel boundary feature extraction and enhancement

The PBFE branch is propose to further incorporate fine-grained boundary information into HBGNet. Specifically, an eight-neighborhood Laplace convolution kernel (LapConv) is used to highlight boundary information and extract features F_1. Simultaneously, LGCA-enhanced F_2 is concatenated with F_1. After applying CBAM operation, the boundary-guided feature, F_G, is obtained. F_G is used as the input data to BGM module for guiding and constraining the APFE semantic information, resulting in multilevel AP mask features. Finally, the boundary feature F_B is obtained as follows:

$$F_B = SoftMax\left(f^{1\times1}(I_{CBAM}(LGCA(F_2) + I_{Res}(F_1)))\right) \tag{4}$$

where I_{CBAM} and I_{Res} mean the CBAM and residual operation, respectively. $f^{1\times1}$ and *SoftMax* indicate a 1 × 1 convolutional and *SoftMax* operation, respectively.

3.1.3. Parcel boundary refinement using BGM

High-level semantic features contain more representative and discriminative global information but lack spatial detail. Moreover, traditional CNNs like ResUNet (Zhang et al. 2018) employ skip connections to aggregate both low- and high-level semantics, which struggle to effectively utilize reliable boundary information, thereby reducing the accuracy of AP boundaries. Therefore, the BGM is constructed to enable low-level boundary information to constrain high-level semantic information, enhancing the interactivity of multi-level information.

As shown in Fig. 7, the low-level semantic features (i.e., boundary-guided feature), named F_G. H_n, represent high-level semantic features from the n_{th} layer. For H_n, it is unsampled to the original size firstly and then a $f^{1\times1}$ is applied to obtain the refined features, i.e., F_{Hn}. For F_G, a $f^{1\times1}$ is used to extract the feature, followed by max pooling and a

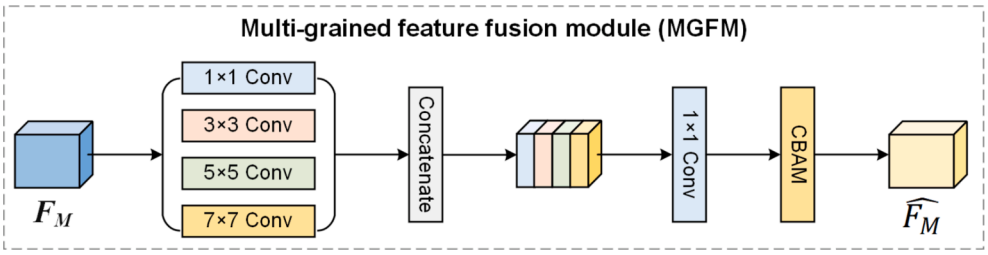


Fig. 8. Structure of the proposed MGFM module for the fusion of multi-grained features to improve the extraction capability of APs with various shapes and sizes.

sigmoid function. The features of each channel are then enhanced by multiplying with F_{Hn}. Lastly, H_n is added to obtain M_{LH}. Here, the BGM is formulated as follows:

$$M_{LH} = \left(\delta\left(MaxPool\left(f^{1\times1}(L_n)\right)\right) \times F_{Hn}\right) + H_n \tag{5}$$

where δ indicates the sigmoid function and *MaxPool* is used to highlight the low-level boundary features of APs.

3.1.4. Multi-grained information enhancement using MGFM module

APs vary in size and shape, making accurate identification challenging. Existing research indicates that extending the model's receptive field helps capture multi-grained information of the target objects (Nam et al. 2017; Zhang et al. 2020b). Inspired by this, the MGFM module is designed to aggregate disparate semantic features from multiple receptive fields (Fig. 8). Specifically, multi-level features derived from low-level and high-level semantics are incorporated firstly (Fig. 5) to obtain concatenated features F_M. Next, four convolutional layers are introduced with different kernel sizes are introduced to create a multi-scale receptive field, capturing fine-grained local features and coarse-grained global features. These features are then fused and fed to a convolution layer to obtain more representative features. Furthermore, the spatial information associated with APs is enhanced, resulting in $\widehat{F_M}$, using the CBAM. Here, MGFM is formulated as follows:

$$\widehat{F_M} = I_{CBAM}\left(f^{1\times1}\left(Cat\left(f^{1\times1}, f^{3\times3}, f^{5\times5}, f^{7\times7}\right)\right)\right) \tag{6}$$

3.2. Multitask loss construction

AP mask and boundary predictions are regarded as major classification tasks. In this study, the binary cross-entropy (BCE) l_{bce} is used to calculate the mask loss. Specifically, let $\widehat{y_m}$ and y_m be the prediction and label values, respectively. Then l_{bce} is defined as:

$$l_{bce} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_m \log \widehat{y_m} + (1-y_m)\log(1-\widehat{y_m})\right) \tag{7}$$

where i is the *ith* pixel and N is the sum of image pixels. y_m and $\widehat{y_m}$ represent the label value and prediction result after a sigmoid function in the *ith* pixel, respectively.

For the classification task of boundary prediction, the Negative Log-Likelihood Loss (NLL) l_{nll} is used to improve training stability (Long et al. 2022; Murugesan et al. 2019). Let x be the pixel position in image space $x_E^i (i = 1, \ldots, N)$. The loss function of edge prediction is defined as:

$$l_{nll} = -\frac{1}{N}\sum_{i=1}^{N} \log p_e(x; y_e(x)) \tag{8}$$

where $p_e\left(x; y_e(x)\right)$ indicates the predicted probability of the edge label value $y_e(x)$ after *SoftMax* function.

Distance prediction provides constraints about the topology and shape of the APs, which represents the distance from the pixels inside the APs to the parcel boundary (which ranges from 0 to 1). Therefore, we regard it as a regression task. To calculate this loss, the mean square error (MSE) l_{mse} is used to compute their bias of the predicted $\widehat{y_d}$ and ground truth distance y_d.

$$l_{mse} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y_d} - y_d\right)^2 \tag{9}$$

Finally, the total loss l_{total} is formulated by integrating the loss values of three subtasks and setting equal weights:

$$l_{total} = w_1 \cdot l_{bce} + w_2 \cdot l_{nll} + w_3 \cdot l_{mse} \tag{10}$$

where $w_1 = w_2 = w_3 = 1$. referring to Murugesan et al. (2019), Waldner and Diakogiannis (2020), Long et al. (2022), and Xu et al. (2024).

3.3. Comparison with existing methods for AP delineation

To further evaluate the performance of the proposed HBGNet, it is compared with six existing open-source models, corresponding to a single-task network, i.e., ResUNet (Taravat et al. 2021; Zhang et al. 2018), three single-branch multitask networks, i.e., ResUNet-a (Waldner and Diakogiannis 2020), BsiNet (Long et al. 2022), REAUNet (Lu et al. 2024) and two dual-branch multitask networks, i.e., SEANet (Li et al. 2023), BFINet (Zhao et al. 2024). Two popular instance segmentation models (i.e., Mask-RNN (He et al. 2017) and E2EVAP (Pan et al. 2023)) are also tested and the result are shown in the supplementary materials.

ResUNet is a U-shaped semantic segmentation network based on residual connectivity. It achieves better segmentation performance with fewer parameters than UNet, while better balancing feature richness and detail trade-offs (Zhang et al. 2018).

REAUNet is a single-branch multitask network based on UNet backbone integrating edge detection block, deep supervision, the refine module and a post-processing workflow to delineate APs from high-resolution images (Lu et al. 2024).

ResUNet-a is a neural network that uses a UNet encoder/decoder backbone combined with residual connectivity, atrous convolutions, pyramid scene parsing pooling and multitask inference. It has been successfully used to extract APs from Sentinel-2 satellite images (Waldner and Diakogiannis 2020).

BsiNet is a lightweight multitask network based on Psi-Net. It used a single encoder-decoder to simultaneously predict parcel extent, boundary, and distance. It has shown advantages in small AP extraction of complex terrain (Long et al. 2022).

SEANet is a multitask network using a dense boundary prediction branch that considers multi-level boundary semantics to improve the identification of parcel boundaries (Li et al. 2023).

BFINet is a dual-branch multitask network mainly for the delineation of APs with blurred boundaries, irregular shapes and small sizes in mountains and hills using a semantic interaction module and a dense multibranch fusion module (Zhao et al. 2024).

3.4. Performance assessment

In this study, we evaluate the resulting accuracy of the model through typical pixel-based evaluation metrics, i.e. Precision (P), Recall (R), F1-score (F1), Overall accuracy (OA), and Intersection over Union (IOU). These measures are defined as follows:

$$IOU = \frac{TP}{TP + FP + FN} \tag{11}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{14}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{15}$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

The geometric metrics (both boundary-based and object-based) are employed to evaluate the geometric accuracy of APs. Specifically, three boundary-based assessment metrics, i.e., *Completeness* (*Com*), *Correctness* (*Corr*), and F1-score of the boundary (F_{bdy}) (Marshall et al. 2019; Persello et al. 2019; Xu et al. 2022) are used. *Com* measures the ratio of accurately detected parcel boundaries, while *Corr* assesses the error of incorrectly detected boundaries. F_{bdy} is the reconciled mean of completeness and correctness, by setting extracted boundary and reference boundary be the B_{pred} and B_{truth}, respectively. The buffer distance of boundary is set to 5-m, refer to Cai et al. (2023). These metrics are expressed as Eq.(16)~(18).

$$Com = \frac{B_{pred} \cap B_{truth}}{B_{truth}} \tag{16}$$

$$Corr = \frac{B_{truth} \cap B_{pred}}{B_{pred}} \tag{17}$$

$$F_{bdy} = \frac{2 \times Com \times Corr}{Com + Corr} \tag{18}$$

An object-based evaluation metric, based on the errors of under-segmentation and over-segmentation (Persello and Bruzzone 2009),is further used to assess the geometric accuracy of identified parcels. Let R_i represent the area where reference parcel object overlaps most with the predicted object P_i, $i = 1, 2, \ldots, n$, where n is the parcel number. Both spatially represent the same parcel object: i.e., area(R_i) and area(P_i) represent the area of R_i and P_i, respectively. That is, area $(R_i \cap P_i)$ represents their overlapping area. We then calculate the over-segmentation error $S_O(i)$ (Eq. (19)) and the under-segmentation error $S_U(i)$ (Eq. (20)). The total-segmentation error $S_T(i)$ is obtained based on $S_O(i)$ and $S_U(i)$. (Eq. (21)).

$$S_O(i) = 1 - \frac{\text{area}(R_i \cap P_i)}{\text{area}(R_i)} \tag{19}$$

$$S_U(i) = 1 - \frac{\text{area}(R_i \cap P_i)}{\text{area}(P_i)} \tag{20}$$

$$S_T(i) = \sqrt{\frac{S_O(i)^2 + S_U(i)^2}{2}} \tag{21}$$

Meanwhile, three global error measures (i.e., *GOC*, *GUC,* and *GTC*) are derived based upon $S_O(i)$, $S_U(i)$ and $S_T(i)$ to evaluate different segmentation errors comprehensively (Li et al. 2015).

$$\mathrm{GOC} = \sum_{i=1}^{n} \left(S_O(i) \times \frac{\text{area}(P_i)}{\sum_{i=1}^{n} \text{area}(P_i)} \right) \tag{22}$$

$$\mathrm{GUC} = \sum_{i=1}^{n} \left(S_U(i) \times \frac{\text{area}(P_i)}{\sum_{i=1}^{n} \text{area}(P_i)} \right) \tag{23}$$

Table 2
AP accuracy evaluation of different models on FHAPD_HLJ.

Methods	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
HBGNet	**91.75**	**85.11**	**0.764**	0.112	**0.152**	**0.159**
SEANet	90.56	83.06	0.727	0.114	0.238	0.215
BsiNet	88.89	81.44	0.686	0.137	0.211	0.208
BFINet	91.44	84.57	0.759	0.099	0.206	0.186
REAUNet	90.42	82.88	0.721	**0.063**	0.355	0.246
ResUNet-a	88.97	81.23	0.401	0.087	0.640	0.477
ResUNet	90.31	81.13	0.720	0.089	0.308	0.245

Table 3
AP accuracy evaluation of different models on FHAPD_JS.

Methods	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
HBGNet	**95.12**	**90.78**	**0.842**	0.053	**0.086**	**0.085**
SEANet	94.78	90.16	0.831	0.054	0.137	0.120
BsiNet	93.66	88.23	0.798	0.065	0.147	0.131
BFINet	93.86	88.91	0.803	0.087	0.112	0.107
REAUNet	93.52	87.93	0.768	0.061	0.171	0.200
ResUNet-a	93.68	87.67	0.680	0.096	0.254	0.213
ResUNet	93.13	88.17	0.834	**0.052**	0.153	0.129

$$\mathrm{GTC} = \sum_{i=1}^{n} \left(S_T(i) \times \frac{\text{area}(P_i)}{\sum_{i=1}^{n} \text{area}(P_i)} \right) \tag{24}$$

where *GOC*, *GUC,* and *GTC* represent global over-classification, global under-classification and global total-classification.

3.5. Implementation details

For the experiments, images of the study area and the corresponding ground truth are cropped to dimensions of 256 × 256 pixels. The overlap rate is set to 50 % based on empirical observations. To ensure model comparability and robust generalization, the data augmentation operations are applied including horizontal and vertical flips, diagonal mirroring, the addition of Gaussian noise, contrast enhancement, zoom, and mix-up. We finally obtain (10220, 2555, 1600), (15272, 3818, 1120), (14604, 3651, 1600), (20228, 5057, 2160), (17468, 4367, 3640), (18864, 4716, 4224) and (15468, 3867, 2000) samples in HLJ, HB, HN, JS, SC, HuN and XJ for training, validation and testing, respectively. Meanwhile, AI4Boundaries (256 × 256 pixels) and Netherlands Sentinel-2 datasets (256 × 256 pixels) are used to obtain (21632, 5408, 1200) and (23200, 5800, 1350) samples for training, validation, and testing, respectively. The Adam optimizer with an initial learning rate of 10^{-4} is employed with an NVIDIA GeForce RTX 4090 GPU to train HBGNet. The batch sizes and epochs are set to 8 and 100, respectively, and same parameters are used to train other models.

4. Experimental result

4.1. Comparing AP delineation using different methods on the FHAPD_HLJ and FHAPD_JS datasets

To demonstrate the superior performance of HBGNet with other models, a comparison based on FHAPD_HLJ and FHAPD_JS datasets (Table 2 and Table 3) was conducted. The reasons for choosing these two regions were: (1): APs in FHAPD_JS were more complicated and imaging time is May 7, when the crop types were abundant. (2): APs in FHAPD_HLJ were regular and distributed in the plain region. (3): The sensor and spatial resolution were different, FHAPD_JS was from GF2 (1-m) and FHAPD_HLJ was from GF1 (2-m). The results indicated that in FHAPD_HLJ (GF-1), HBGNet obtained the highest accuracy in F1 and IOU, which were 2.05 % and 3.67 % higher than SEANet and BsiNet in IOU, respectively. HBGNet also achieved the highest F_{bdy} score and the lowest GTC error on FHAPD_HLJ. Similarly, HBGNet achieved the

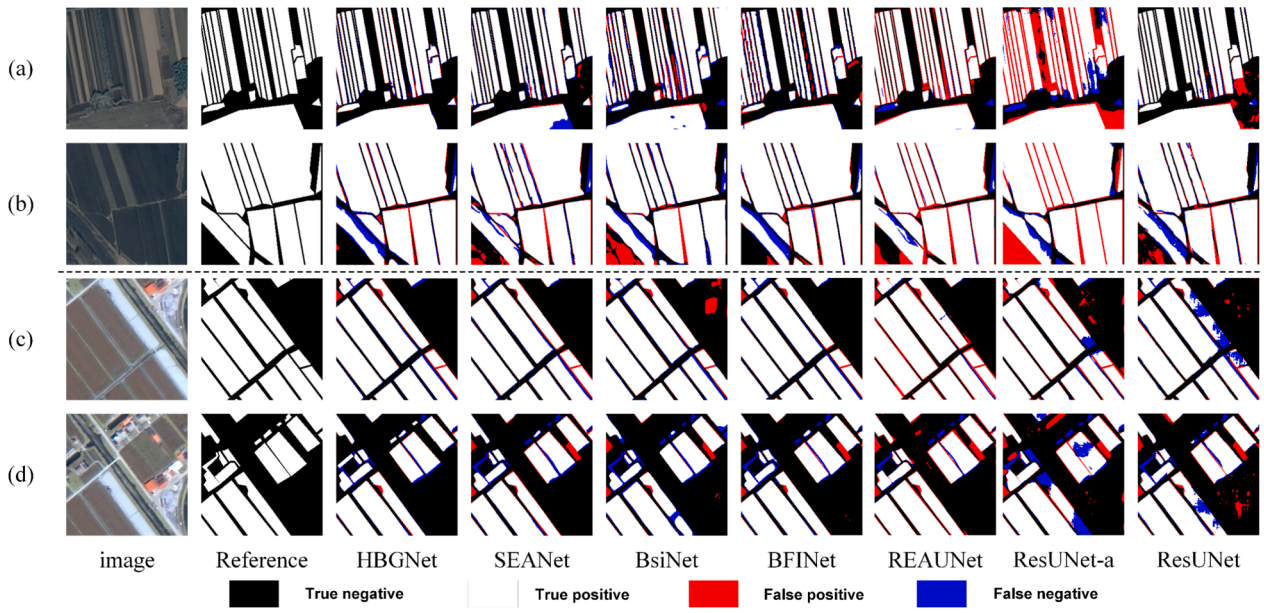


Fig. 9. Visual comparisons of different models for AP delineation. (a) – (b) and (c) – (d) exhibited extracted results on the FHAPD_HLJ and FHAPD_JS, respectively.

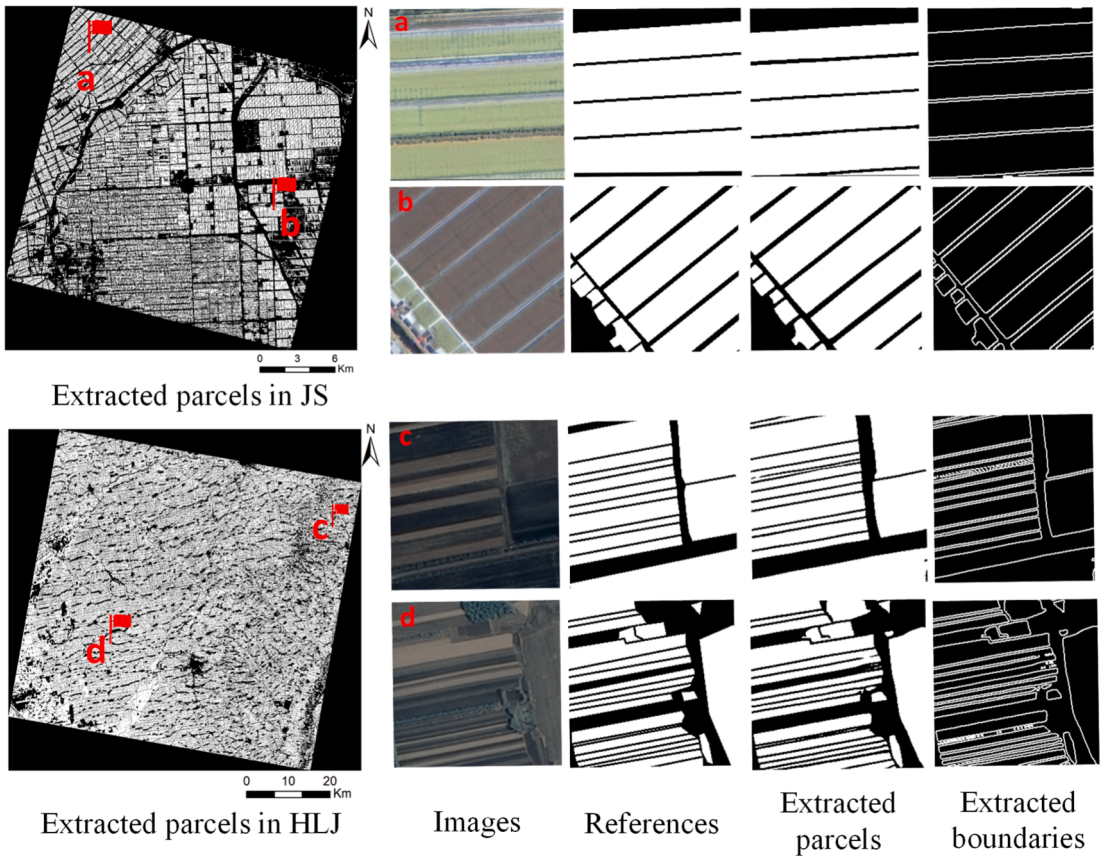


Fig. 10. Extracted APs by HBGNet on Jiangsu (JS) GF-2 image and Heilongjiang (HLJ) GF-1 image.

highest F1, IOU, and F_{bdy} scores and had the lowest GUC and GTC error on FHAPD_JS. These results demonstrated that HBGNet was superior in capturing robust AP boundary information across different sensors (GF-1, GF-2).

The recognition errors of the different models were visualized in Fig. 9. HBGNet achieved the lowest segmentation error compared to the other models and had better APS extraction results than the recently proposed SEANet, which was proved to achieve a state-of-the-art performance. Fig. 10 showed the inference results for the whole GF-1 and GF-2 images, demonstrating satisfactory AP boundary detection details, indicating that utilizing boundary-guided information improved AP delineation.

Table 4
AP accuracy evaluation of different models on the public AI4Boundaries dataset.

Methods	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
HBGNet	**91.96**	**86.06**	**0.623**	**0.087**	0.346	**0.280**
SEANet	89.84	83.05	0.573	0.120	**0.344**	0.297
BsiNet	88.94	81.28	0.508	0.119	0.429	0.350
BFINet	90.69	84.36	0.574	0.103	0.431	0.328
REAUNet	87.98	80.55	0.476	0.099	0.493	0.385
ResUNet-a	88.54	80.23	0.466	0.316	0.561	0.468
ResUNet	88.23	80.63	0.491	0.110	0.459	0.369

Table 5
AP accuracy evaluation of different models on Netherlands Sentinel-2 image datasets.

Methods	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
HBGNet	**90.67**	**83.22**	**0.793**	0.328	**0.142**	**0.299**
SEANet	87.88	78.83	0.749	0.314	0.234	0.334
BsiNet	87.25	77.76	0.717	**0.257**	0.453	0.436
BFINet	89.35	80.94	0.775	0.343	0.155	0.309
REAUNet	90.55	82.38	0.779	0.310	0.227	0.331
ResUNet-a	86.11	76.56	0.751	0.301	0.162	0.351
ResUNet	85.97	75.70	0.722	0.310	0.304	0.369

4.2. Comparing AP delineation using different methods on the AI4Boundaries and the Netherlands datasets

The AP delineation was also conducted using the AI4Boundaries orthophoto dataset (1-m) and Netherlands Sentinel-2 images (10-m) (Table 4 and Table 5). These tables indicated that the proposed method exhibited the highest extraction accuracy across all measures. In IOU, it surpassed SEANet and BsiNet by 3.01 % and 4.78 % on the AI4Boundaries, 4.39 % and 5.46 % on Netherlands datasets, respectively. Additionally, HBGNet outperformed ResUNet-a, specifically designed for Sentinel-2 images by achieving 4.56 % and 6.66 % higher F1 and IOU than ResUNet-a on the Netherlands dataset. HBGNet also exhibited satisfactory performance in geometric accuracy, with GTC decreased by 1.7 % to 8.9 % on the AI4Boundaries and 3.5 % to 7 % on the Netherlands datasets, respectively. These results demonstrated HBGNet's effectiveness in extracting features related to APs from images with varying resolutions, regions, and sensors, and in handling spectral heterogeneity.

A similar pattern was observed in the results shown in Fig. 11. Compared to other methods, HBGNet performed better in identifying varied parcels with different landscapes, producing the fewest omission and commission on the AI4Boundaries and NL datasets. Moreover, Fig. 12 showed the inference results of Sentinel-2 imagery using HBGNet for the entire Netherlands, demonstrating excellent AP boundary detection. Our model accurately identified fine agricultural parcels in 10-m resolution Sentinel-2 imagery, effectively excluding irrelevant features such as buildings.

The results compared with two popular instance segmentation models were shown in Table S3, Tabel S4, and Fig. S1 of the supplementary materials. HBGNet outperformed them in both attribute and geometric accuracy, and produced the fewest FN and FP errors.

4.3. AP delineation for other FHAPD areas

To assess the adaptability of the proposed method in varied agricultural environments, AP delineation was performed on GF-2 images in the other five FHAPD regions. Despite variations in color and fragmentation levels among parcels in these regions, HBGNet demonstrated satisfactory results, indicating strong generalization performance (Fig. 13). It was evident from the pixel-based visualization that the delineation results on the XJ, HB, and HN areas outperformed those in the HuN and SC areas. This was primarily because HuN and SC are in the mountainous regions of China, where the APs were more fragmented, making the segmentation tasks more challenging. The segmentation performance (Table 6) of these five regions was quantified and all metrics indicated satisfactory results generated by HBGNet.

4.4. Transfer testing across different areas

To validate the transferability of HBGNet on VHR images in different regions, the HBGNet trained on FHAPD_XJ GF-2 images was applied to predict AP results directly on JS GF-2 images (i.e., from XJ to JS). The two regions exhibited similar parcel characteristics with neat and regular alignment of APs, but distinct crop planting patterns. Similarly, the model trained on HuN GF-2 images was transferred to directly predict the APs on GF-2 images from SC. The APs in the two areas were characterized by mountainous agriculture and fragmentation in both regions. Meanwhile, the transfer experiment was also conducted in agricultural regions with different parcel characteristics (i.e., from JS to SC). Fig. 14 displayed the transferred results, revealing that the APs were identified accurately, while other types (e.g., road, building) not associated with APs were discarded, even for images with evident spectral differences.

Table 7 listed the accuracies of attribute and geometric measures

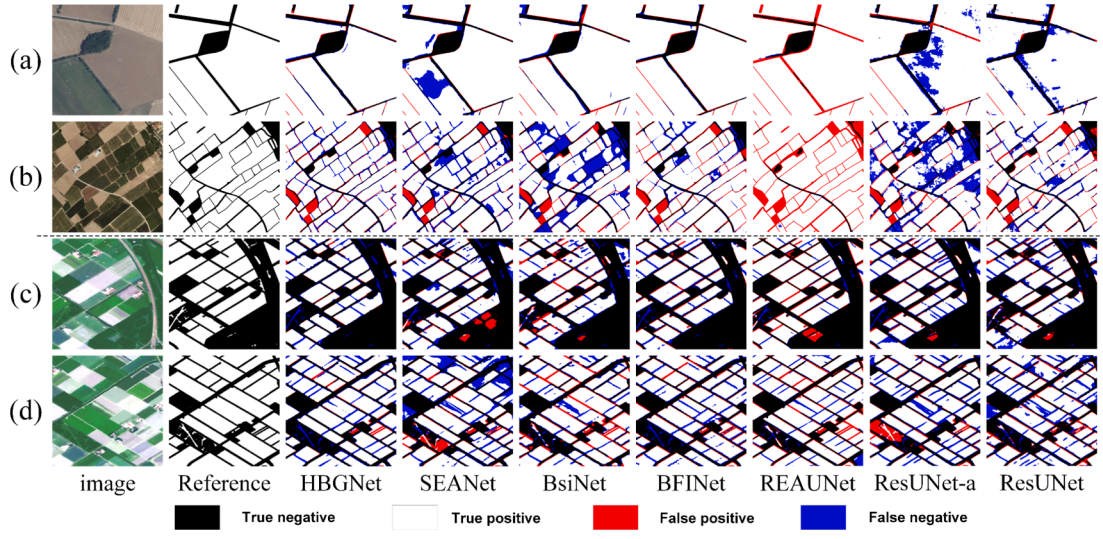


Fig. 11. Visual AP results of different models. (a) – (b) and (c) – (d) are the extracted results on the AI4Boundaries and Netherlands Sentinel-2 dataset, respectively.

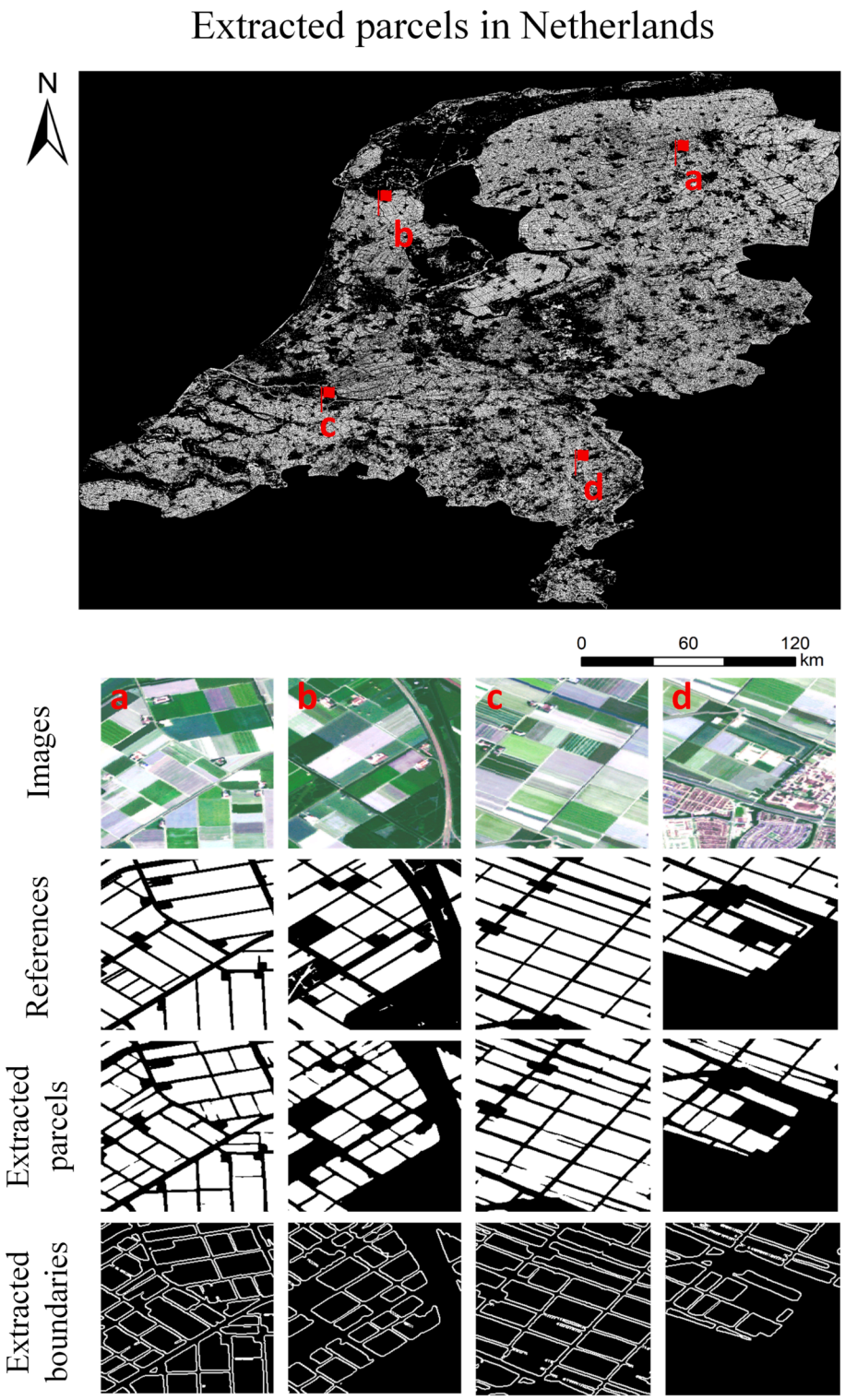


Fig. 12. Extracted parcels and boundaries by HBGNet on the Netherlands Sentinel-2 images.

after transferring. The accuracy of XJ-to-JS was close to that directly trained on the JS areas. Additionally, in the experiments of HuN-to-SC, the two datasets had a large difference in spectra and landscapes, posing a significant challenge for AP delineation. However, HBGNet still achieved high accuracy similar to that directly trained on SC areas, which further confirmed that HBGNet remained a promising approach in these regions with evident fragmentation parcels. As for the experiments of JS-to-SC, although the accuracy metrics have a small gap compared with the results directly trained on SC areas, the results are acceptable in combination with Fig. 14. These results verify the high transferability of HBGNet in regions with different AP features, even across different agricultural landscapes and image acquisition times.

5. Discussion

In this study, a novel and efficient AP delineation method is constructed based on multitask learning, i.e., HBGNet. It includes a boundary extraction branch based on Laplace convolution (i.e., PBFE) to

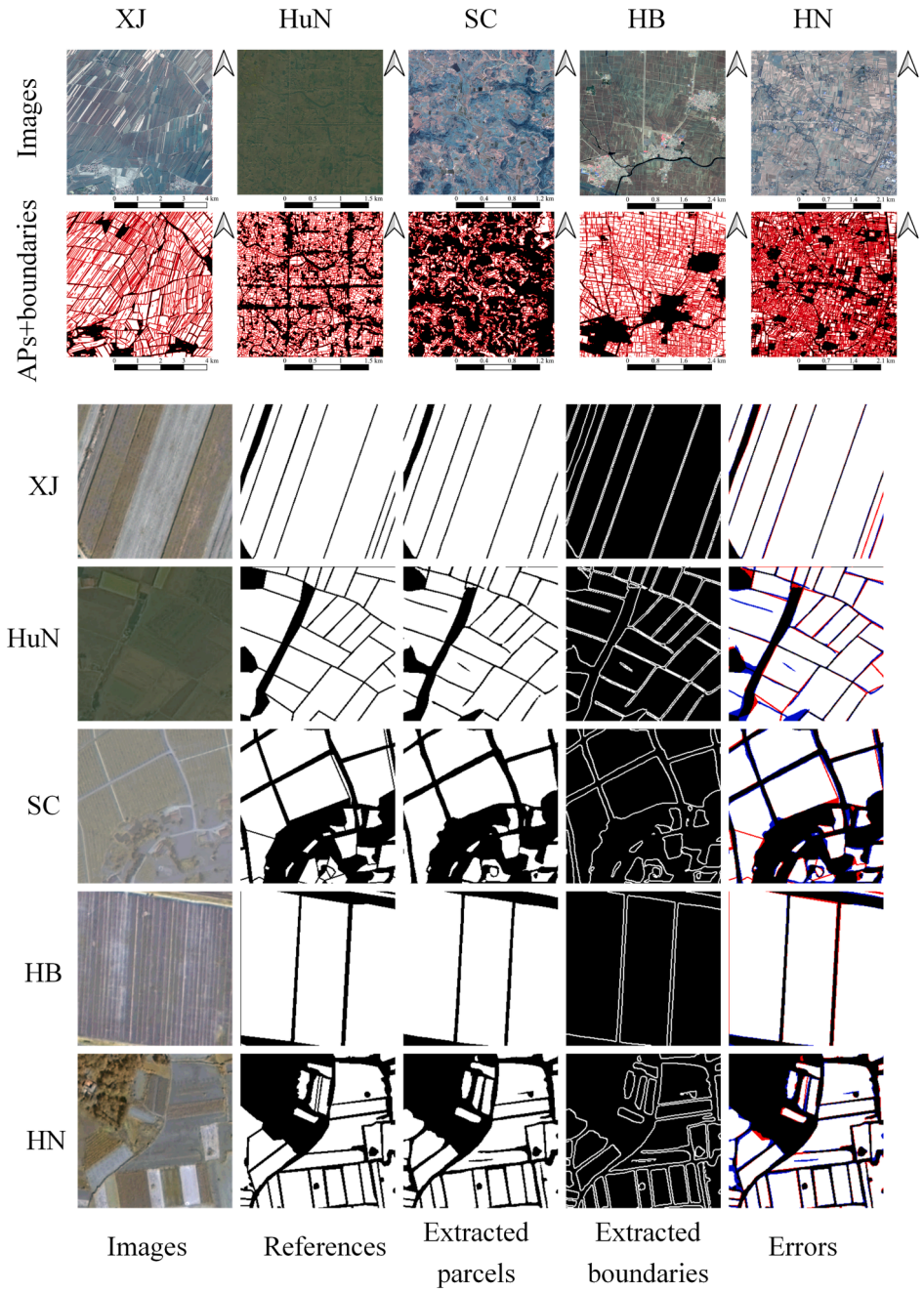


Fig. 13. AP delineation results and detail display for other regions in FHAPD (APs boundaries are red vector and shown overlaid with mask).

Table 6
AP accuracy evaluation of HBGNet on other FHAPD.

Dataset	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
FHAPD_XJ	82.37	91.52	0.785	0.123	0.108	0.186
FHAPD_HuN	84.23	82.58	0.792	0.102	0.206	0.188
FHAPD_SC	84.16	84.31	0.791	0.190	0.351	0.318
FHAPD_HB	92.23	85.98	0.763	0.093	0.186	0.172
FHAPD_HN	90.65	88.52	0.827	0.119	0.156	0.163

obtain low-level fine-grained boundary semantics to refine high-level AP features. This information interaction and weight sharing strategy is different from existing single-branch and dual-branch multitask models that focus on capturing high-level semantics (Diakogiannis et al. 2020; Li et al. 2023; Long et al. 2022; Murugesan et al. 2019). Moreover, we construct the first large-scale VHR parcel dataset (i.e., FHAPD) collected from GF-1 and GF-2 images covering seven different agricultural areas of China. The extensive experiments are conducted on three datasets, including two high-resolution benchmark datasets, i.e., the FHAPD, a public AI4Boundaries dataset, and a medium-resolution Netherlands Sentinel-2 dataset. Our results show that HBGNet obtains satisfactory results in all datasets with varying sensors, spatial resolution, and terrains, and evidently outperforms eight existing methods, i.e., SEANet, BsiNet, REAUet, BFINet, ResUNet-a, ResUNet, Mask R-CNN, and E2EVAP, in terms of attribute and geometric measures. Furthermore, HBGNet has great potential to transfer to different areas across various dates, landscapes, and images.

5.1. *The first large-scale VHR parcel dataset (FHAPD)*

This study presents the first-of-its-kind agricultural parcel dataset in

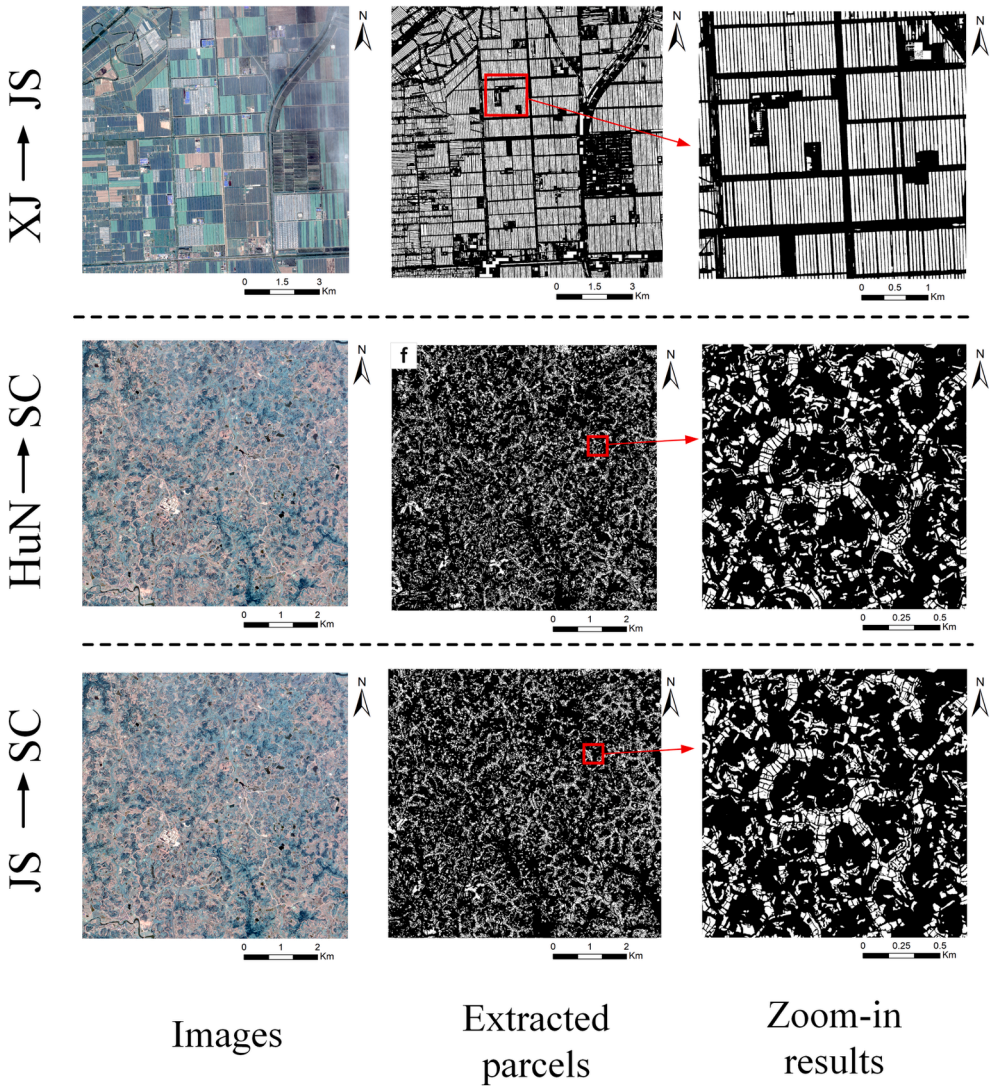


Fig. 14. Extracting APs results of different transfer experiments.

Table 7
Transfer accuracy evaluation of HBGNet across different areas.

Methods	F1 (%)	IOU (%)	F_{bdy}	GOC	GUC	GTC
JS	95.12	90.78	0.842	0.053	0.086	0.085
XJ → JS	94.21	88.92	0.811	0.152	0.168	0.186
SC	84.16	84.31	0.791	0.190	0.351	0.318
HuN → SC	80.46	81.91	0.706	0.105	0.401	0.482
JS → SC	77.56	77.03	0.633	0.239	0.583	0.611

China with high accuracy and spatial resolution. To our knowledge, there are no publicly available large-scale AP VHR datasets in China. Additionally, due to various reasons, relevant methodology papers have not made their datasets publicly available. This limitation hinders benchmarking, performance improvement of data-driven deep learning models, and comparisons between different models. The constructed FHAPD dataset features various parcel types and terrains, different phenological stages, and rich spectral information. It can be regarded as a benchmark AP dataset for model comparison and advancement. Future efforts will focus on collecting more parcel samples in various scenarios to enrich the FHAPD.

Table 8
Ablation experiments of HBGNet, $HBGNet_{noLap}$, $HBGNet_{noPBFE}$, and $HBGNet_{noBT}$ on FHAPD_JS.

Methods	F1 (%)	mIOU (%)	F_{bdy}	GOC	GUC	GTC
HBGNet	**95.12**	**90.78**	**0.842**	**0.051**	**0.086**	**0.085**
$HBGNet_{noLap}$	90.93	86.79	0.781	0.128	0.147	0.129
$HBGNet_{noPBFE}$	90.14	84.61	0.743	0.145	0.151	0.163
$HBGNet_{noBT}$	94.08	88.63	0.826	0.094	0.113	0.104

5.2. *Effectiveness of PBFE branch in extracting APs*

Many studies have demonstrated that the boundary task is a crucial component in existing multitask network models of AP delineation (Cai et al. 2023; Li et al. 2023; Long et al. 2022). However, these methods usually use a down-sampling block to capture multi-level features, which leads to the loss of significant spatial details that retain finer resolution and valuable information for predicting pixel-level locations (Chen et al. 2014; He et al. 2016). Significant changes in pixel values at AP boundaries make predictions even more challenging.

Three experiments were conducted to evaluate the effectiveness of the PBFE branch. First, the Laplacian convolution in the PBFE branch was replaced with traditional convolution to verify its validity, referred to as $HBGNet_{noLap}$. Second, the PBFE branch (except for the first level of PVT-V2) was removed to verify the validity of the PBFE, called $HBGNet_{noPBFE}$. Third, eliminate the auxiliary task, but keep the PBFE

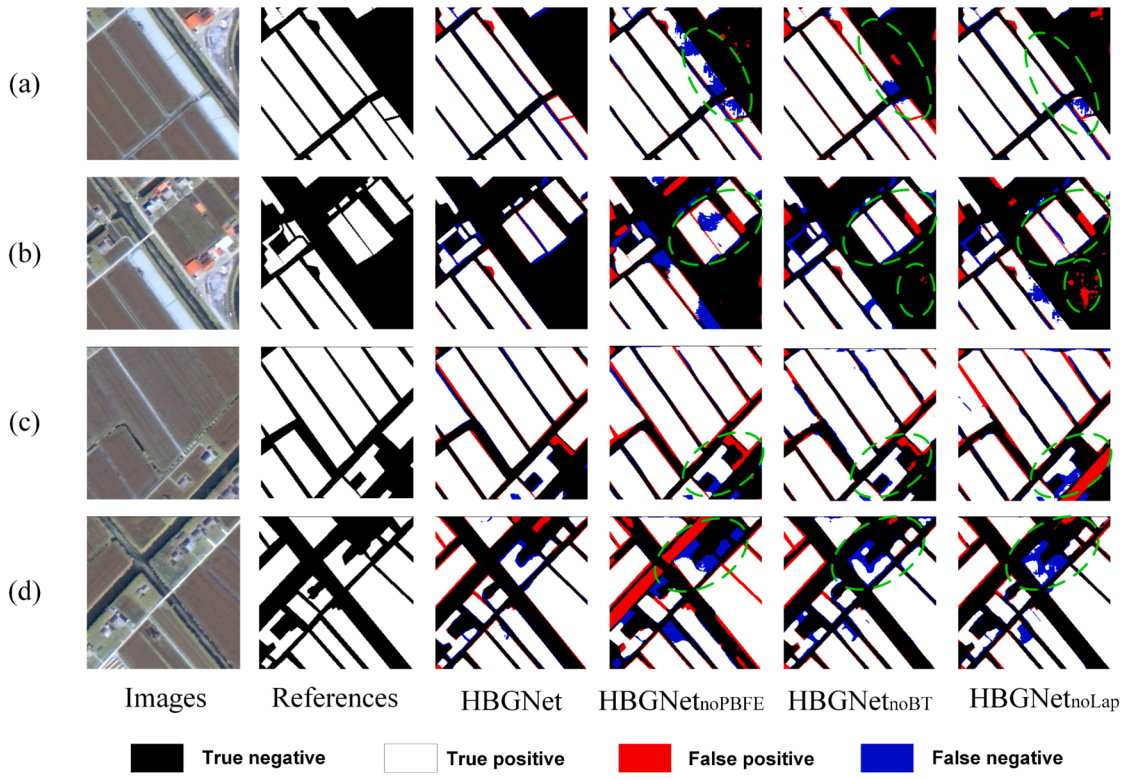


Fig. 15. Visual comparisons of HBGNet, HBGNet$_{noLap}$, HBGNet$_{noPBFE}$, and HBGNet$_{noBT}$ on FHAPD_JS.

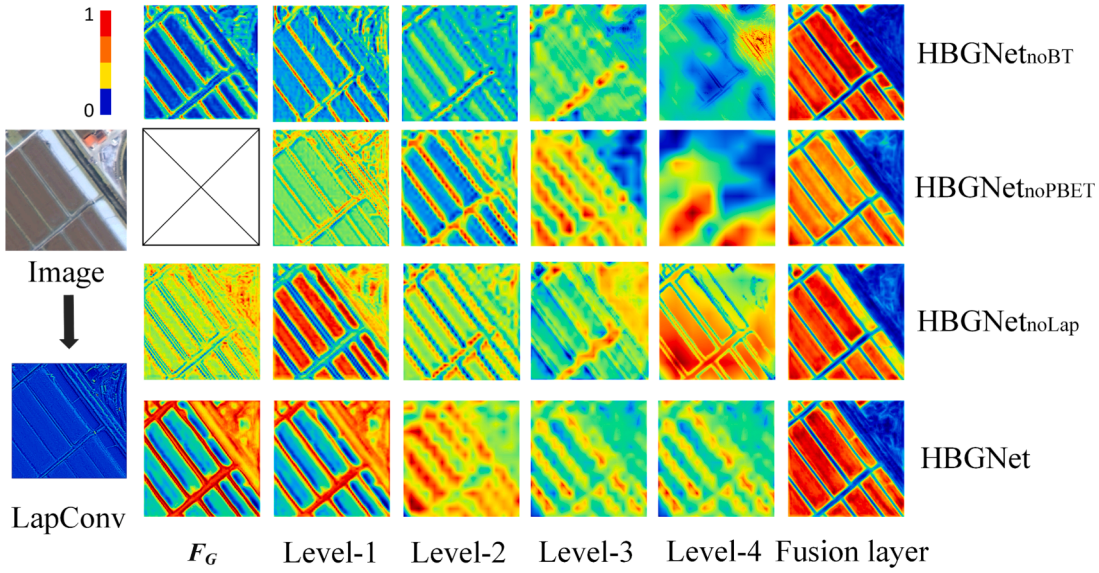


Fig. 16. Comparison of feature maps for the three tasks at different levels on the FHAPD_JS dataset. F_G is the boundary feature input to BGM. LapConv denotes the feature by Laplace convolution. All image resolution is 256 × 256 pixels.

branch, called HBGNet$_{noBT}$. The results (Table 8 and Fig. 15) indicated that HBGNet$_{noPBFE}$ performed the poorest in terms of both attribute and geometric accuracy, with a clear increase in false positive (FP) and false negative (FN). In contrast, HBGNet$_{noLap}$ showed improvements but did not produce distinct AP boundaries. The results of HBGNet$_{noBT,}$ which significantly reduced FP and FN pixels, were only surpassed by those of HBGNet.

Meanwhile, the visualization of feature maps was presented in Fig. 16. Clearly visible boundary features (e.g., red pixels) are observed in HBGNet. In contrast, AP structures are also observed in F_G of HBGNet$_{noBT}$, but with limited boundary activation. F_G of HBGNet$_{noLap}$ barely activated boundary information, demonstrating the effectiveness of Laplace convolution in delineating AP boundaries. Therefore, HBGNet effectively maintains boundary semantics across various levels, utilizing them to achieve superior AP mask delineation.

The PBFE addresses this issue by utilizing Laplace convolution, which combines the properties of first-order and second-order derivatives. First-order derivatives capture general edge information and highlight rough edges, facilitating the extraction of regular AP boundaries. Second-order derivatives, in contrast, are more sensitive to image

Table 9
Ablation experiments of proposed modules and pixel-based AP accuracy evaluation on FHAPD_JS.

Methods	OA (%)	precision (%)	recall (%)	F1 (%)	IOU (%)	IOU↑ (%)
Baseline	93.58	95.02	93.76	94.37	89.45	
Baseline + LGCA	93.71	95.99	94.05	94.50	89.67	0.22
Baseline + LGCA + BGM	93.84	94.99	94.31	94.63	**90.39**	**0.72**
Baseline + LGCA + BGM + MGFM	**94.38**	**95.32**	**94.96**	**95.12**	**90.78**	**0.39**

Table 10
Geometric accuracy evaluation of ablation experiments on FHAPD_JS.

Methods	Com	Corr	F_{bdy}	GOC	GUC	GTC	GTC↓ (%)
Baseline	0.826	0.824	0.819	0.057	0.160	0.136	
Baseline + LGCA	0.829	0.829	0.825	0.054	0.150	0.128	0.8 %
Baseline + LGCA + BGM	0.830	0.822	**0.838**	0.057	0.143	**0.105**	**2.3 %**
Baseline + LGCA + BGM + MGFM	0.836	0.847	0.842	0.051	0.086	0.085	2.0 %

details such as fine lines, isolated points, and noise, thus enhancing the accuracy of extracting irregular AP boundaries (see the green boxes of Fig. 15, Fig. 16). Consequently, the boundary features extracted by this method are more representative and distinguishable. Exploring effective combinations of different edge detection operators with deep learning-based methods is a promising area of research. Furthermore, incorporating design logic inspired by traditional image segmentation methods into deep learning-based semantic segmentation models has significant potential for improving multitask learning. Adaptive adjustment of loss weights between different subtasks and selecting optimal weights will enhance the interpretability of deep learning models. Investigating more efficient post-processing methods will further improve AP delineation performance.

5.3. Effectiveness of the proposed modules

5.3.1. Ablation experiment of the proposed modules

To assess the effectiveness of the proposed modules (i.e., LGCA, BGM, and MGFM), ablation experiments were conducted based on the FHAPD_JS dataset. The evaluation results of attribute and geometric accuracy were presented in Table 9 and Table 10, respectively. The results showed that the incremental addition of these modules enhances the accuracy of AP delineation. Specifically, removing all proposed modules led to a decrease of 1.33 % in IOU and 2.3 % in F_{bdy}, while *GTC* increased by 5.1 %. Notably, incorporating the BGM module significantly improves accuracy, with a 0.72 % and 1.3 % increase in IOU and F_{bdy}, 2.3 % decrease in *GTC* errors.

Fig. 17 illustrates that adding the LGCA module enhances the identification of isolated APs, while incorporating the BGM module further eliminates these APs, resulting in smoother AP boundaries, as shown in the green boxes in Fig. 17 (b). This suggested that AP mask predictions could be effectively optimized using low-level boundary information. In contrast, the introduction of MGFM improved the accuracy of the model in recognizing APs of different shapes (see the green boxes in Fig. 17 (a) and (d)). Additionally, the results demonstrated that LGCA, BGM, and MGFM modules possess strong capabilities to capture discriminate semantic cues, thereby enhancing the spatial details of AP boundaries and integrates semantic representations at different receptive fields.

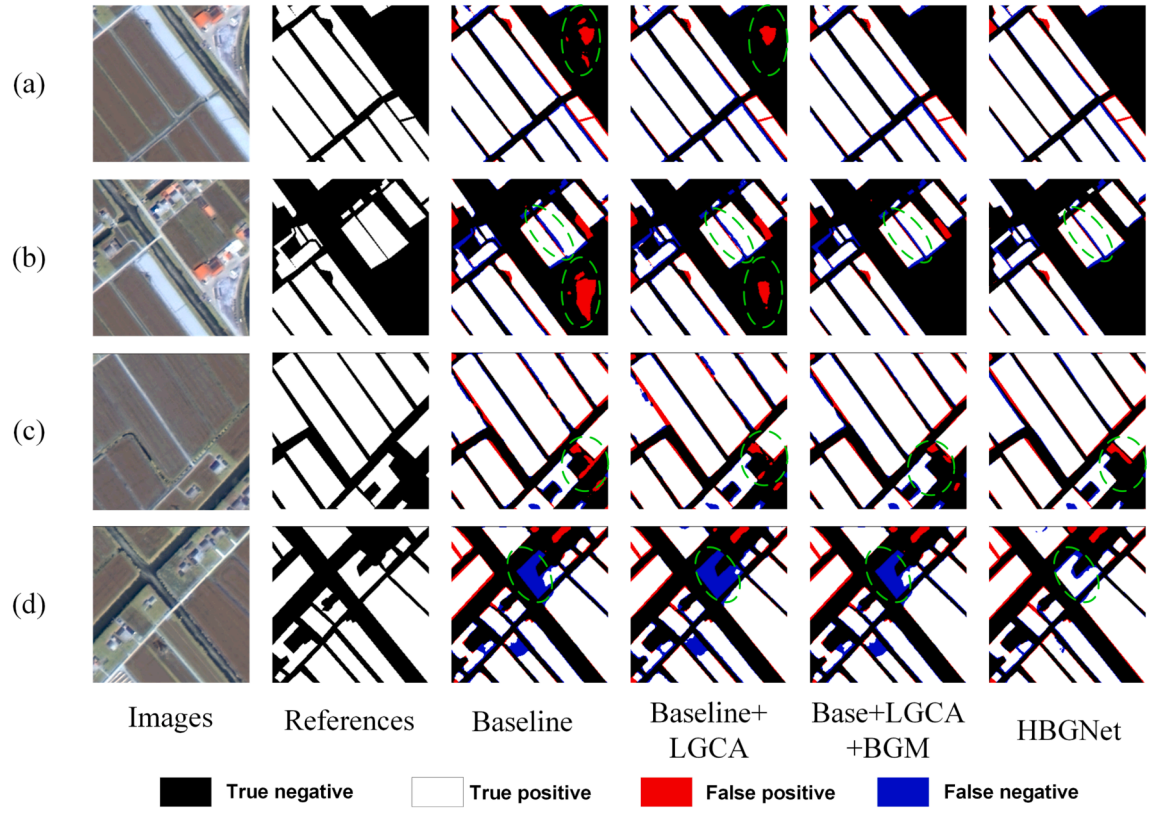


Fig. 17. Visual comparisons of ablation experiment results.

Table 11
Floating point of operations (FLOPs) and parameters evaluation of different methods. 1GFLOPs = 10^9 FLOPs and 1 M = 10^6.

Method	Image size	FLOPs	Parameters
HBGNet	256 × 256	58.94G	30.45 M
SEANet	256 × 256	208.14G	28.75 M
BsiNet	256 × 256	13.34G	7.85 M
BFINet	256 × 256	14.75G	27.38 M
REAUNet	256 × 256	254.42G	142.41 M
ResUNet-a	256 × 256	70.90G	131.47 M
ResUNet	256 × 256	81.02G	13.04 M

5.3.2. BGM improves the boundary accuracy of AP delineation

Existing multitask learning models are typically categorized into single-branch (Cai et al. 2023; Long et al. 2022; Waldner and Diakogiannis 2020; Waldner et al. 2021) and dual-branch (Li et al. 2023; Liu et al. 2022; Shunying et al. 2023). Single-branch architectures attempt to simultaneously extract parcels and boundaries, but do not fully utilize boundary information. In contrast, dual-branch architecture uses independent branches to capture local–global boundary information, thereby improving parcel delineation accuracy. However, such methods fail to effectively constrain the AP mask and increase model complexity. To effectively utilize the boundary information, we combine single-branch and dual-branch approaches through the BGM module. Fig. 17 illustrates that the BGM module reduces the FP errors and optimizes AP boundary, while simultaneously yielding clear and well-defined AP boundary features at different layers (Fig. 16). Ultimately, the number of HBGNet parameters is comparable to dual-branch networks like SEANet, but it significantly reduces the model complexity (Table 11) while enhancing performance (4.1, 4.2). This finding aligns with Yuan et al. (2020), which shows that adding boundary orientation and intensity information can further constrain object boundaries.

5.3.3. LGCA module improves the identification of complex scenarios

Recent studies have shown that incorporating multi-scale contextual information aids in recognizing objects with various shapes and spectra, especially those with similar spectral and textural characteristics (Jin et al. 2024). However, effectively addressing the jagged edge remains challenging when relying solely on local and global contexts (Wang et al. 2022a). To improve the AP identification in complex scenarios, we propose the LGCA module to enhance the representation capability of multi-scale contextual information (Table 9). As shown in the green boxes of Fig. 17, the LGCA module significantly mitigates the impact of small objects. This finding further demonstrates that the LGCA module helps compensate for the lack of global contextual feature extraction capability at different layers of traditional CNNs. Future research can explore more efficient fusion modules for local and global context information, which is an interesting direction for further study.

5.4. Analysis of model adaptability and transferability

This study tests the effectiveness of HBGNet on two high-resolution datasets (i.e., FHAPD and AI4Boundaries) and medium-resolution Netherlands dataset, which cover almost all agricultural landscapes. Overall, the accuracy of APs extracted from flat and regular agricultural landscapes (e.g., XJ, JS, and the Netherlands) is higher than that in irregularly shaped mountainous areas with predominantly small farmers (e.g., SC and HuN). Our results demonstrate the robustness of HBGNet to imagery from different sensors and resolutions (4.3, 4.4). In addition, our method successfully transfers to different agricultural landscapes, and effectively avoids the disturbance of other objects even with limited training samples (Fig. 14). However, current delineation results are often confounded by abandoned land and grassland, an issue that could be addressed by incorporating medium-resolution time-series data, particularly for areas with complex crop cultivation. Meanwhile, a large-scale AP delineation based on deep learning is still a major challenge. Combining visual foundation models, e.g., segment anything model, to produce considerable high-quality pseudo labels to achieve small sample based fine-tuning tasks could reduce the labor-intensive task of manual outlining, is promising for improving the robustness and transferability of AP delineation models (Kirillov et al. 2023; Long et al. 2024b).

6. Conclusions

This paper develops a hierarchical semantic boundary-guided network, namely HBGNet, to delineate APs using remote sensing imagery of multiple resolutions and sensors. It includes two branches: (1) an agricultural parcel feature extraction (APFE) branch to extract multi-scale contextual information related to APs based on the LGCA module, and (2) a parcel boundary feature extraction (PBFE) branch to mine spatial explicit AP boundary features using Laplace convolution. These two branches are then combined by a BGM module, which utilizes boundary features from PBFE to guide and constrain high-level semantic features derived from APFE. In addition, HBGNet captures multi-grained AP features by extending the receptive field using MGFM module. Furthermore, this study establishes the first large-scale high-resolution remote sensing AP dataset in China (i.e., FHAPD) that covers various agricultural landscapes to promote the comparison and innovation of AP delineation models.

Experiments were conducted on three datasets: FHAPD, Netherlands Sentinel-2, and AI4Boundaries. First, ablation experiments show that FP and FN error pixels are significantly reduced under the guidance and constraint of boundary branches (Table 8 and Fig. 15). BGM refines high-level AP boundary features, evidently improving the geometric accuracy of delineation results (Table 10 and Fig. 17). Second, comparative results show that HBGNet outperformed the other eight methods in both attribute and geometry. The IOU, F_{bdy}, and GTC exceeded other methods by 0.61 %-7.52 %, 0.8 %-36.3 %, and 1.7 %-31.8 %, respectively. Finally, to facilitate the AP delineation in unknown region, transfer testing show HBGNet can transfer to various areas, even across different dates, landscapes, and crop types (Table 7 and Fig. 14).

Overall, the developed HBGNet model exhibits strong performance and transferability, substantially enhancing the monitoring and understanding of agricultural industrialization and intensification in regions predominantly characterized by smallholder farming systems.

CRediT authorship contribution statement

Hang Zhao: Writing – review & editing, Writing – original draft, Validation, Software, Methodology. **Bingfang Wu:** Writing – review & editing, Methodology. **Miao Zhang:** Writing – review & editing. **Jiang Long:** Writing – review & editing, Software, Methodology. **Fuyou Tian:** Writing – review & editing, Methodology. **Yan Xie:** Formal analysis, Data curation. **Hongwei Zeng:** Writing – review & editing. **Zhaoju Zheng:** Writing – review & editing. **Zonghan Ma:** Validation. **Mingxing Wang:** Visualization, Software. **Junbin Li:** Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the third Xinjiang Scientific Expedition (No. 2022xjkk030106), the National Key Research and Development Program of China (No. 2022YFD1900802, No. 2023YFD2000105).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2025.01.034.

References

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26, 341–358.

Brabandere, B.D., Neven, D., & Gool, L.V. (2017). Semantic Instance Segmentation with a Discriminative Loss Function. ArXiv, abs/1708.02551.

Cai, Z., Hu, Q., Zhang, X., Yang, J., Wei, H., He, Z., Song, Q., Wang, C., Yin, G., Xu, B., 2022. An adaptive image segmentation method with automatic selection of optimal scale for extracting cropland parcels in smallholder farming systems. Remote Sens. (Basel) 14, 3067.

Cai, Z., Hu, Q., Zhang, X., Yang, J., Wei, H., Wang, J., Zeng, Y., Yin, G., Li, W., You, L., 2023. Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by integrating multimodal satellite data. ISPRS J. Photogramm. Remote Sens. 205, 34–49.

Cao, Y., Zhao, Z., Huang, Y., Lin, X., Luo, S., Xiang, B., Yang, H., 2023. Case instance segmentation of small farmland based on Mask R-CNN of feature pyramid network with double attention mechanism in high resolution satellite images. Comput. Electron. Agric. 212, 108073.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

Cheng, T., Ji, X., Yang, G., Zheng, H., Ma, J., Yao, X., Zhu, Y., Cao, W., 2020. 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. Agric. 178, 105787.

d'Andrimont, R., Claverie, M., Kempeneers, P., Muraro, D., Yordanov, M., Peressutti, D., Batič, M., Waldner, F., 2023. AI4Boundaries: an open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. Earth Syst. Sci. Data 15, 317–329.

De Brabandere, B., Neven, D., & Van Gool, L. (2017). Semantic instance segmentation with a discriminative loss function. arXiv preprint arXiv:1708.02551.

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote Sens. 162, 94–114.

Fathi, A., Wojna, Z., Rathod, V., Wang, P., Song, H.O., Guadarrama, S., & Murphy, K.P. (2017). Semantic instance segmentation via deep metric learning. arXiv preprint arXiv:1703.10277.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: In, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.

Garcia-Pedrero, A., Gonzalo-Martin, C., Lillo-Saavedra, M., 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. Int. J. Remote Sens. 38, 1809–1819.

Götz, T.I., Göb, S., Sawant, S., Erick, X., Wittenberg, T., Schmidkonz, C., Tomé, A., Lang, E., Ramming, A., 2022. Number of necessary training examples for Neural Networks with different number of trainable parameters. J. Pathol. Inform. 13, 100114.

Hasituya, C., Z., Li, F., & Hongmei, 2017. Mapping plastic-mulched farmland with C-band full polarization SAR remote sensing data. Remote Sens. (Basel) 9, 1264.

Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. ISPRS J. Photogramm. Remote Sens. 57, 327–345.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: In, 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hossain, M.D., Chen, D., 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. ISPRS J. Photogramm. Remote Sens. 150, 115–134.

Jin, Y., Liu, J., Zhou, Y., Chen, R., Chen, H., Duan, W., Chen, Y., Zhang, X.-L., 2024. CRDet: A circle representation detector for lung granulomas based on multi-scale attention features with center point calibration. Comput. Med. Imaging Graph. 102354.

Jong, M., Guan, K., Wang, S., Huang, Y., Peng, B., 2022. Improving field boundary delineation in ResUNets via adversarial deep learning. Int. J. Appl. Earth Obs. Geoinf. 112, 102877.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., 2023. Segment anything. In: In, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026.

Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 2017. 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. (Basel) 9, 259.

Li, M., Bijker, W., Stein, A., 2015. Use of binary partition tree and energy minimization for object-based classification of urban land cover. ISPRS J. Photogramm. Remote Sens. 102, 48–61.

Li, M., Long, J., Stein, A., Wang, X., 2023. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. ISPRS J. Photogramm. Remote Sens. 200, 24–40.

Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y., 2017. Fully Convolutional Instance-Aware Semantic Segmentation. In: In, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4438–4446.

Lin, R., Zhang, Y., Zhu, X., Chen, X., 2024. Local-Global Feature Capture and Boundary Information Refinement Swin Transformer Segmentor for Remote Sensing Images. IEEE Access.

Liu, S., Liu, L., Xu, F., Chen, J., Yuan, Y., Chen, X., 2022. A deep learning method for individual arable field (IAF) extraction with cross-domain adversarial capability. Comput. Electron. Agric. 203, 107473.

Liu, W., Wang, J., Luo, J., Wu, Z., Chen, J., Zhou, Y., Sun, Y., Shen, Z., Xu, N., Yang, Y., 2020. Farmland parcel mapping in mountain areas using time-series SAR data and VHR optical images. Remote Sens. (Basel) 12, 3733.

Long, J., Li, M., Wang, X., Stein, A., 2022. Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images. Int. J. Appl. Earth Obs. Geoinf. 112, 102871.

Long, J., Li, M., Wang, X., Stein, A., 2024a. Semantic change detection using a hierarchical semantic graph interaction network from high-resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 211, 318–335.

Long, J., Zhao, H., Li, M., Wang, X., Lu, C., 2024b. Integrating Segment Anything Model Derived Boundary Prior and High-Level Semantics for Cropland Extraction From High-Resolution Remote Sensing Images. IEEE Geosci. Remote Sens. Lett. 21, 1–5.

Lu, R., Wang, N., Zhang, Y., Lin, Y., Wu, W., Shi, Z., 2022. Extraction of agricultural fields via dasfnet with dual attention mechanism and multi-scale feature fusion in south xinjiang, china. Remote Sens. (Basel) 14, 2253.

Lu, R., Zhang, Y., Huang, Q., Zeng, P., Shi, Z., Ye, S., 2024. A refined edge-aware convolutional neural networks for agricultural parcel delineation. Int. J. Appl. Earth Obs. Geoinf. 133, 104084.

Marshall, M., Crommelinck, S., Kohli, D., Perger, C., Yang, M.Y., Ghosh, A., Fritz, S., Bie, K.d., & Nelson, A, 2019. Crowd-driven and automated mapping of field boundaries in highly fragmented agricultural landscapes of Ethiopia with very high spatial resolution imagery. Remote Sens. (Basel) 11, 2082.

McGarigal, K., Marks, B.J., 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. In: USDA Forest Service General Technical Report PNW-GTR-351.

Mueller, M., Segl, K., Kaufmann, H., 2004. Edge- and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. Pattern Recogn. 37, 1619–1628.

Murugesan, B., Sarveswaran, K., Shankaranarayana, S.M., Ram, K., Joseph, J., & Sivaprakasam, M. (2019). Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation. In, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 7223-7226): IEEE.

Nam, H., Ha, J.-W., Kim, J., 2017. Dual attention networks for multimodal reasoning and matching. In: In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 299–307.

Nock, R., Nielsen, F., 2004. Statistical region merging. IEEE Trans. Pattern Anal. Mach. Intell. 26, 1452–1458.

Pan, Y., Wang, X., Zhang, L., Zhong, Y., 2023. E2EVAP: End-to-end vectorization of smallholder agricultural parcel boundaries from high-resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 203, 246–264.

Persello, C., Bruzzone, L., 2009. A novel protocol for accuracy assessment in classification of very high resolution images. IEEE Trans. Geosci. Remote Sens. 48, 1232–1244.

Persello, C., Tolpekin, V., Bergado, J.R., De By, R., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. Remote Sens. Environ. 231, 111253.

Robb, C., Hardy, A., Doonan, J.H., Brook, J., 2020. Semi-automated field plot segmentation from UAS imagery for experimental agriculture. Front. Plant Sci. 11, 591886.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098.

Rydberg, A., Borgefors, G., 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. IEEE Trans. Geosci. Remote Sens. 39, 2514–2520.

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., Aroyo, L.M., 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In: In, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–15.

Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

Shunying, W., Ya'nan, Z., Xianzeng, Y., Li, F., Tianjun, W., Jiancheng, L., 2023. BSNet: Boundary-semantic-fusion network for farmland parcel mapping in high-resolution satellite images. Comput. Electron. Agric. 206, 107683.

Singh, S., & Singh, R. (2015). Comparison of various edge detection techniques. In, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 393-396).

Tang, Z., Li, M., Wang, X., 2020. Mapping tea plantations from VHR images using OBIA and convolutional neural networks. Remote Sens. (Basel) 12, 2935.

Taravat, A., Wagner, M.P., Bonifacio, R., Petit, D., 2021. Advanced fully convolutional networks for agricultural field boundary detection. Remote Sens. (Basel) 13, 722.

Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J. Photogramm. Remote Sens. 144, 325–340.

Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote Sens. Environ. 245, 111741.

Waldner, F., Diakogiannis, F.I., Batchelor, K., Ciccotosto-Camp, M., Cooper-Williams, E., Herrmann, C., Mata, G., Toovey, A., 2021. Detect, consolidate, delineate: Scalable mapping of field boundaries using satellite images. Remote Sens. (Basel) 13, 2197.

Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., Atkinson, P.M., 2022a. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS J. Photogramm. Remote Sens. 190, 196–214.

Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L., 2022b. Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media 8, 415–424.

Watkins, B., Van Niekerk, A., 2019. Automating field boundary delineation with multi-temporal Sentinel-2 imagery. Comput. Electron. Agric. 167, 105078.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module. In: In, Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.

Xu, L., Ming, D., Du, T., Chen, Y., Dong, D., Zhou, C., 2022. Delineation of cultivated land parcels based on deep convolutional networks and geographical thematic scene division of remotely sensed images. Comput. Electron. Agric. 192, 106611.

Xu, W., Wang, J., Wang, C., Li, Z., Zhang, J., Su, H., Wu, S., 2024. Utilizing Dual-Stream Encoding and Transformer for Boundary-Aware Agricultural Parcel Extraction in Remote Sensing Images. Remote Sens. (Basel) 16, 2637.

Xu, Y., Zhu, Z., Guo, M., Huang, Y., 2023. Multiscale edge-guided network for accurate cultivated land parcel boundary extraction from remote sensing images. IEEE Trans. Geosci. Remote Sens.

Yuan, Y., Xie, J., Chen, X., & Wang, J. (2020). Segfix: Model-agnostic boundary refinement for segmentation. In, Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16 (pp. 489-506): Springer.

Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., Chen, Q., 2020a. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. Remote Sens. Environ. 247, 111912.

Zhang, H., Liu, M., Wang, Y., Shang, J., Liu, X., Li, B., Song, A., Li, Q., 2021. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. Int. J. Appl. Earth Obs. Geoinf. 105, 102557.

Zhang, J., Lin, S., Ding, L., Bruzzone, L., 2020b. Multi-scale context aggregation for semantic segmentation of remote sensing images. Remote Sens. (Basel) 12, 701.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. 15, 749–753.

Zhao, H., Long, J., Zhang, M., Wu, B., Xu, C., Tian, F., Ma, Z., 2024. Irregular agricultural field delineation using a dual-branch architecture from high-resolution remote sensing images. IEEE Geosci. Remote Sens. Lett.

Zhong, B., Wei, T., Luo, X., Du, B., Hu, L., Ao, K., Yang, A., Wu, J., 2023. Multi-swin mask transformer for instance segmentation of agricultural field extraction. Remote Sens. (Basel) 15, 549.