IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 August 2023, accepted 11 September 2023, date of publication 22 September 2023,
date of current version 11 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3318000

RESEARCH ARTICLE

Multi-Scale Bilateral Spatial Direction-Aware Network for Cropland Extraction Based on Remote Sensing Images

WEIMIN HOU, YANXIA WANG, JIA SU, YANLI HOU, MING ZHANG, AND YAN SHANG
School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

Corresponding author: Yan Shang (shangyan@hebust.edu.cn)

This work was supported in part by the Science and Technology Program of Hebei Province under Grant 20355901D and Grant 21355901D.

ABSTRACT The information of cropland is obtained efficiently and accurately as the basis for achieving precision agriculture (PA). As the boundary between cropland and other types of land in remote sensing images with different resolutions is fuzzy, the characteristics of cropland are easily confused when extracting cropland, resulting in inaccurate identification and extraction of cropland under large and complex backgrounds and rough localization of marginal areas. We proposed a two-path multiscale attention self-supervised network with the perception of four directions in pixel space, called the multiscale bilateral spatial direction-aware network (MBSDANet), to solve these problems and improve the model's ability to extract cropland in small samples. One path extracts attentional feature maps by spatial directions to preserve detailed direction-aware information and generate high-resolution features; the other path obtains local-to-global information through pyramid pooling and attention awareness to capture dense multiscale cropland features to separate targets in complex contexts. The features of the two branches at different levels are fused by weighting the multi-aware information. Triangular self-supervised and boundary-aware losses are used to achieve fine segmentation and extraction of cropland in small samples. We tested the extraction method on cropland in Denmark and the Hebei Province of China, demonstrating its effectiveness and generalization. Compared to other neural network models, MBSDANet achieves better accuracy with a precision of 0.9481, an IoU of 0.8937, and an F1 score of 0.9438.

INDEX TERMS Cropland extraction, remote sensing images, multi-aware information, spatial directions-aware, two-path branch.

I. INTRODUCTION

Cropland is closely related to the development of human society and economy in terms of quantity, quality, and spatial distribution [1], [2]. At present, remote sensing technology is widely used in digital precision agriculture (PA), such as crop area statistics, crop pest monitoring, crop yield prediction, etc. [3], [4], [5], [6]. The identification and extraction of cropland are crucial for the subsequent acquisition of cropland change and cropland utilization information, which is helpful for crop growth monitoring and yield prediction [7], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

Manual mapping is an early method of cropland extraction. When monitoring scattered and large-scale farmland, manual mapping is not only time-consuming and laborious but also affects the accuracy of data. At present, methods based on deep learning show performance that is difficult for traditional methods to achieve in some scenarios. However, due to the large differences in scale, shape, and other aspects of different cropland regions, cropland extraction is still faced with some difficulties at this stage: First, the resolution of remote sensing images is limited, and errors often occur when extracting cropland in low-resolution images, thus affecting the accuracy of subsequent studies; Second, the spectral characteristics of the same type of cropland may be different

or similar to those of other types of land, so there may be omissions or errors in the extraction of cropland; Third, the positioning of the edge region is very rough, and the boundary will be blurred.

Cropland extraction methods based on remote sensing images are mainly in the three most common types of learning: unsupervised learning [9], semi-supervised learning [10], and supervised learning [11]. Xue et al. [12] applied the simulated submerged watershed method to merge the cropland division units. Their method is designed to incorporate objects that are pre-segmented well from high-resolution images. In [13] and [14], some object-based methods are proposed to pick out cropland in high-resolution remote sensing images. To improve discontinuity and maintain the smoothness of the divided fields, normal and uniform kernels are used to filter internal fields and boundary areas. Hong et al. [15] designed a set of mathematical methods for extracting land-level boundaries from regularly arranged agricultural areas. Pen et al. [16] Although multi-temporal spectra, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) have different effects on multi-temporal classification, the expression of spatial shape features is ignored, which cannot effectively eliminate the influence of spectral-spatial heterogeneity, resulting in poor mapping accuracy. The cropland boundaries extracted by these methods are quite detailed. However, it is usually necessary to artificially design features or parameters in combination with prior knowledge of the size, shape, and texture of cropland in multi-source remote sensing images. For large-area farmland extraction, it is a difficult task to accurately classify the objects in the farmland using unsupervised methods.

Graesser and Ramankutty [17] proposed a semi-supervised edge detection method to detect a single cropland plot from Landsat images with time series, but the image resolution limited the accuracy of cropland boundary and area. Teluguntla et al. [18] used Random Forest (RF) method to extract and classify the cropland area of different agroecological zones (AEZs) in Australia and China on Google Earth Engine (GEE) platform. Waldner et al. [19] added feature interpolation to the decision tree algorithm to generate farmland maps, thereby improving the stability of unsupervised classifiers. These methods have the problem of low accuracy of cultivated land extraction and are prone to the wrong and missing cropland extraction. However, semi-supervised learning can reduce the cost of manpower, time, and resources, and is of great significance to the research of cropland extraction.

With the rise of deep learning, new opportunities have been brought to natural image applications, and some mature algorithms have been applied in remote sensing [20], [21], [22], [23]. Shelhamer et al. [24] proposed a fully convolutional network (FCN) on the basis of convolutional neural network (CNN), which realizes pixel-level classification of images by replacing the last fully connected layer of the network with the upper sampling layer. Since then, many segmentation algorithms have extended FCN, from the initial convolutional neural network U-Net [25], [26] to deep neural network models such as pyramid scene parsing network (PSPNet) [27], [28] and Deeplabv3+ [29], [30].

Most of the existing cropland extraction methods based on deep learning are supervised learning. Garcia-Pedrero et al. [31] used global boundary probability to obtain the boundaries of farmland plots. Masoud et al. [32] redesigned a new super-resolution semantic contour detection network on the basis of the FCN, and realized the division of cultivated land boundary by using spatial background information, thus improving the spatial resolution of cultivated land boundary output. Lu et al. [33] proposed a dual attention and scale fusion network (DASFNet) to extract cultivated land from GaoFen-2 images of Aral City, southern Xinjiang, China. However, the low-resolution feature map of this network may lose some key features during the up-sampling process, resulting in blurred boundaries. Yang et al. [34] performed multi-source fusion of RGB images and multi-spectral images collected by drones, and combined with a deep learning algorithm, accurately extracted soybean planting areas on the scale of farmland. However, this method often aimed at a single crop and needed to perform spectral feature statistics on a large number of crops when extracting multi-crop farmland. Zhang et al. [35] improved PSPNet and combined depth distance features with shadow local features to provide more detailed predictions. Sun et al. [36] proposed successive pooling attention network (SPANet), using continuous pooling operators to connect intermediate pooling features of different scales to extract deeper semantic features. Huan et al. [37] proposed Multiple attention encoder-decoder network (MAENet), embedding dual-pooling efficient channel attention (DPECA) module into the trunk. A dual-feature attention (DFA) module is designed to extract the context information of advanced features. Cao et al. [38] combined semantic segmentation network U-Net and feature extraction residual network (ResNet) into an improved Res-UNet network for extracting spatial and spectral features of remote sensing images. Xu et al. [39] proposed a high-resolution context extraction network (HRCNet) based on a high-resolution network (HRNet), which uses HRNet structure to retain spatial information and significantly improves boundary and segmentation performance, with an overall accuracy of 92.0% and 92.3%, respectively.

The above methods provide convincing segmentation results but have not been evaluated on multi-resolution remote sensing images and complex mountain cropland. We propose the multi-scale bilateral spatial direction awareness network (MBSDANet) based on the direction-aware spatial background (DSC) [40], which trains pre-models with a small number of labeled farmland samples. The cropland can be accurately extracted from remote sensing images with different resolutions and complex scenes. Through many experiments, the performance is compared with that of

several commonly used methods. The main contributions are as follows:

1) We constructed a two-path multi-aware network structure to acquire features at different levels in two paths to achieve better global multi-level feature extraction.

2) We proposed a pyramidal multiscale attention module (PMAM) and a spatial directional attention module (SDAM). PMAM combined a multiscale global averaging pool and a multi-attention mechanism to improve the attention to different scales of cropland information and the ability to extract critical information. The SDAM analyzes detailed features of cropland from different directions in pixel space, thus enabling the network to focus better on the cultivated region and its surrounding areas with structural information.

3) We designed a triangular self-supervised training strategy based on feature perturbation to train a small number of samples and designed boundary loss using the Roberts cross operator to optimize the edge details so that the model can focus on the boundary pixels.

II. MATERIALS AND METHODS

A. EXPERIMENTAL SETTINGS

The proposed method is performed on an Ubuntu 18.04 system with NVIDIA GeForce RTX 3090 GPU and 16 GB RAM. We set up a training process with 100 epochs and use the Adam algorithm to improve the training, The batch size is set to 4. If the epoch is too small, the training parameters cannot be optimized, and if the epoch is too large, the training parameters can be overfitted. We dynamically observe the accuracy of the model on the validation set after each epoch and find that the highest accuracy is achieved when epoch is around 100. Set up a batch size of 4 to match epoch strike a balance between the efficiency of memory and memory capacity. Using a low learning rate can ensure that we do not miss any local minima, and the learning decay rate is usually more than 100 times the learning rate, so we choose 0.00015 as the initial learning rate, and the learning decay rate is 0.92.

B. STUDY AREA AND DATASET

Denmark's agricultural region is dominated by arable land [41], accounting for about 91% of the total arable land, mainly cereals, oilseeds, and protein crops, especially cereals (1.5 million hectares), which accounts for more than half (55%) of the total agricultural area. The cropland label for the selected area was derived from the Danish 'Marker' open dataset in the 2016 Land Parcel Identification System (LPIS), taken during the May 2016 growing season and consisted of already-trimmed Sentinel-2A images with a spatial resolution of 10 m. We randomly selected 2000 images from the cultivated data of LPIS for the experiment, where the ratio of the training set, test set, and validation set is 6:3:1 and the size of each image is 640 × 640.

Hebei Province is located in the north of China, close to the Yellow River Valley, and the central and southern parts belong to the North China Plain, with good lighting conditions and more planting types [42]. Compared with other provinces such as Guizhou, Hebei Province has superior cropland conditions, obvious structure distribution, and a high degree of intensification. The cropland area accounts for 41.8% of the total land area in Shijiazhuang. We collected 10 m resolution cropland images (Sentinel-2A dataset) of Shijiazhuang City from Sentinel-2A and mixed them with 4m spatial resolution images (Google Earth dataset) collected from Google Earth to obtain 1400 images. The two groups of data account for 50% each. We flipped and cropped some images to increase sample diversity. The ratio of the training set, verification set, and test set is 5:1:1. At the same time, we collected 700 images with spatial resolution of 0.5-2 m from the GaoFen-2 (GF-2 dataset) of complex, irregular mountain cropland in Hebei Province, China, in 2019 and 2020. The ratio of training set, verification set, and test set is 3:1:1. We resample downloaded Sentinel 2A data into ENVI format through SNAP and then use the Layer Stacking tool in ENVI 5.3 to fuse the converted data and perform band stacking. The ROI tool was used to crop the region of interest of the fused image to 512 × 512 size. Each sample set consists of two 512 × 512 images and corresponding binary labels of cropland.

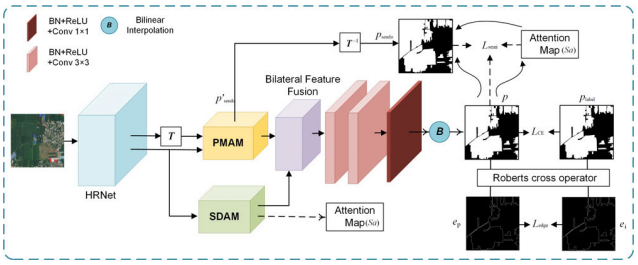


FIGURE 1. The structure of the MBSDANet for cropland extraction.

C. METHODS

We proposed a two-path fused multiscale attention network with four spatial directions for extracting cropland, called a multiscale bilateral spatial direction-aware network (MBSDANet), which has the structure shown in Figure 1. The MBSDANet performs cropland segmentation and extraction in an end-to-end way. The model adopts HRNet [43] as the backbone network for feature extraction, maintains a high resolution, and performs down-sampling and fusion to generate feature maps with multiple resolutions in the end. After high-resolution feature extraction, two branches are formed, namely the multi-scale sensing path and spatial direction path. In multiscale awareness paths, features with different resolutions are extracted by PAMA, which can increase the perceptual field to obtain effective global contextual a priori information. Spatial direction paths extract the output of the high-resolution features from the previous stage from different pixel directions to refine the contextual feature extraction and obtain detailed information about the

cropland in the image and use spatial attention to achieve a contextual correlation of the spatial scope of pixels and depict various complex scenes in remote sensing images. In order to make better use of the complementarity between the two path features, the feature fusion of the two paths is carried out to obtain the final feature map. Two convolution layers are added to the output end after feature fusion, and bilinear interpolation is used to restore the original input size image.

We compute the supervised loss L_{CE} using the image label P_{label} and the segmentation result of the main network feature fusion p. In addition, we perform a direct transformation of the encoded features for the unlabeled images with a training strategy based on feature perturbation triangulation. We use the Roberts cross operator to obtain the corresponding boundary mapping as a boundary loss to enhance the boundary detail pixel prediction.

1) MULTISCALE AWARENESS PATHS

There are significant differences between cropland in different areas in remote sensing images, such as cropland plots between urban and rural areas having uniform tones, moderate brightness and texture, and regular shapes. However, cropland in complex areas has irregular shapes and a wide range of land types, and the information features in the feature maps are mixed with a large amount of redundant information. The PMAM integrates local features and non-local semantic associations of cultivated plots in complex scenes in the deep feature map by coupling multiscale global average pooling and multi-attention mechanisms.

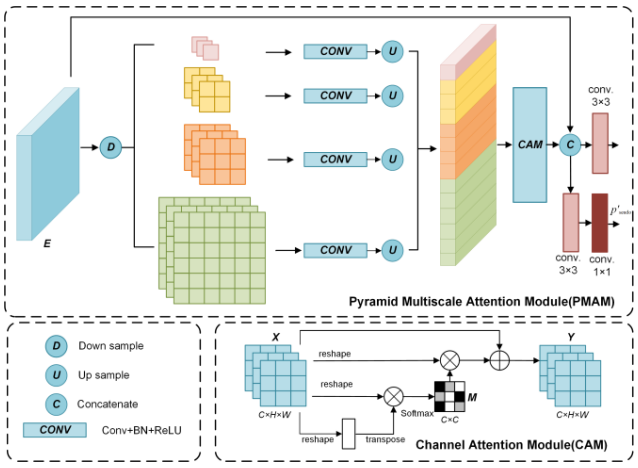


FIGURE 2. The structure of Pyramid Multiscale Attention Module (PMAM).

The structure of PMAM is shown in Figure 2, where the high-resolution features $E \in i^{C\times H\times W}$ extracted from the previous stage are multi-scale pooled with a pooling size of 1×1, 2×2, 3×3 and 6×6. The multi-scale pooling can reduce the overhead of model computation and capture the multi-scale features of cropland information and information that varies between different subregions. The dimensions of the pooled features are directly up-sampling by bilinear interpolation to obtain a feature map $X \in i^{C\times H\times W}$ with the same scale as the original feature map, and the features of different levels are concatenated into pyramidal pooled global features.

The semantic association of pooled centers with single-point pixels is established by the channel attention module (CAM) [44]. X_p is projected onto three different spaces and reshaped directly into $X_1 \in i^{C\times N}$, $X_2 \in i^{C\times N}$ and $X_3 \in i^{C\times N}$, where $N = H \times W$ is the number of pixels.

The correlation matrix M between the reshaping matrices X_1 and X_2 is calculated and named the channel attention map.

$$M_{b,a} = \frac{\exp\left(X_{2a}\cdot X_{1b}^{T}\right)}{\sum_{a=1}^{C}\exp\left(X_{2a}\cdot X_{1b}^{T}\right)} \tag{1}$$

where $M_{b,a}$ measures the a channel's impact on the b channel. The reshaping matrix X is weighted and added to E to obtain the aggregated features $Y \in i^{C\times H\times W}$:

$$Y_i = \beta\sum_{a=1}^{C}\left(M_{b,a}\cdot X_{3a}\right)+E \tag{2}$$

where β is the scaling factor, and the parameter is updated adaptively during the training.

PMAM uses the attention mechanism to focus on the feature retrieval and sifts out a small amount of truly valuable information from the images by establishing spatial or channel dimensional semantic associations to obtain a more consistent global semantic response feature. Finally, the original input E_1 is connected with the final attentional aggregated features Y by a connection layer and output by a 3×3 convolutional layer.

$$O = W_1\left(concat\left(E_1, Y\right)\right) \tag{3}$$

where W_1 denotes the linear transformation.

2) SPATIAL DIRECTION PATHS

The existing cropland extraction methods have the problems of fuzzy cropland edge information extraction and difficulty to distinguish forest land and cropland in satellite images. In order to solve the above problems, we added the SDAM to comprehensively extract large areas of cropland, understand the context information of the global image, analyze the details of different spatial directions, and further extract the texture features between different croplands, improve the extraction accuracy of edge information, and strengthen the extraction ability of context information and detail features. The structure of the spatial directional attention module is shown in Figure 3. The three-layer residual block (RB) in SDAM further extracted the high-resolution features E_a of the previous stage, and the DSC summarized the results of the four directions (the row and column information of the pixel) to obtain the spatial attention map S_a.

The DSC module produces an attention-aware map of the cropland from the input feature. The S_a is coupled with the three-layer RB to obtain the first global spatial context

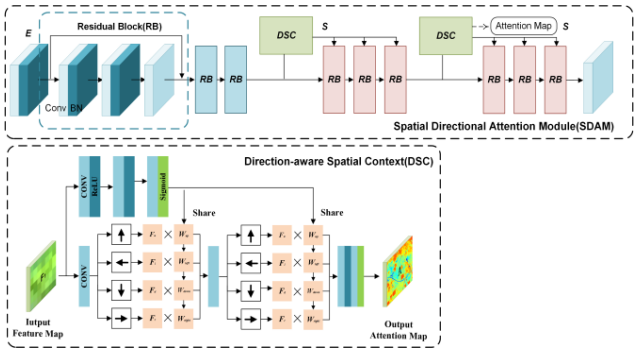


FIGURE 3. The structure of Spatial Directional Attention Module (SDAM). Blocks of the same color are the same operation module.

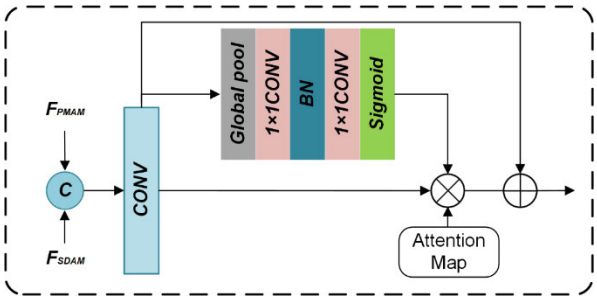


FIGURE 4. The Structure of two-path feature fusion.

feature E_{a1}, where each spatial attention maps coupling operation is expressed as follows:

$$E_{a1} = W_{res}(x) \cdot S_{a1}(x) + x \tag{4}$$

where $W_{res}(x)$ is the feature input to the RB through the three convolutional layers parameter. In the process of obtaining the spatial attention map of the DSC module, the global spatial background feature is obtained after two calculations by convoluting the feature map in four directions: up, down, left and right. The local features $f_{i,j}$ at the location (i, j) are as follows:

$$f_{i,j}^{direction} = \max(\alpha_{direction} f_{i,j-1} + f_{i,j}, 0) \tag{5}$$

where α denotes the weight parameter in the convolution operation, and i, j denotes the spatial location index, $i = 1, 2, \cdots, H \times W$ and $j = 1, 2, \cdots, N$. The global spatial contextual features S_a. are calculated as:

$$W_{i,j}^{direction} = f_{i,j}^{direction} \times W_{i,j}^{direction} \tag{6}$$

$$S_a = W^{total} \times f(x) + x \tag{7}$$

where x is the input feature map and W_{total} is the attention weight map. Concatenating the four direction weights gives $W_{total} = concat(W_{right}, W_{left}, W_{up}, W_{down})$. The value of each element in the attention map indicates how much attention should be assigned to the pixel.

Spatial four-directional contextual attention helps to determine the characteristics of the spatial relationship and location of cropland, making it easier to distinguish spectrally similar plots to enhance texture detection between cultivated and non-cropland.

3) BILATERAL FEATURE FUSION

In order to get the highest-level feature map, we fuse the features extracted from the multi-scale perceptual path and the spatial direction path. The structure of two-path feature fusion is shown in Figure 4. The former parts of F_{PMAM} contain different scales of contextual semantic features, and the latter features of F_{SDAM} provide spatial detail information. We used the attention module in squeeze-and-excitation network (SENet) [45] to multiply with the input features and the spatial attention map S_a to complete the information filtering and guide the attention of MBSDANet to the cropland information so that the spatial dimensional feature responses that are more useful for cropland information identification can be retrieved.

4) OPTIMIZATION OF EDGE DETAILS AND TRAINING

To calibrate the accuracy of the segmentation for cropland, we utilized the cropland edge information as a constraint term, which optimizes the boundaries between cultivated and non-cropland. We used the edge loss function as an auxiliary supervision term to optimize the training process for effective boundary prediction.

Since the segmented images are binary, i.e., cropland non-cropland, we choose the most straightforward and fastest edge detection operator [46], the Roberts cross operator, to extract the edge information. The Roberts cross operator is the first-order derivative edge detection operator, and its computational form is equivalent to the convolution operation of each image pixel with two pseudo-convolution kernels. The pseudo-convolution kernels of the Roberts cross operator are shown as follows:

$$g_i = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad g_j = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{8}$$

The model output of the cropland binary segmentation result is convolved with the Roberts operator to get the edge prediction value, and binary edge cross-entropy loss is applied as an auxiliary function to optimize the training.

$$g(p_{i,j}) = \left|g_i * p_{i,j}\right| + \left|g_j * p_{i,j}\right| \tag{9}$$

$$L_{edge} = -\sum_{i=1}^{N} \left[e_i \log(g(p_i)) + (1 - e_i) \log(1 - g(p_i))\right] \tag{10}$$

where e_i is the binarized edge label of the extracted cropland and $p_{i,j}$ is the model prediction output.

$$L_{CE}(p_{\text{label}}, p) = -\sum_{i=0}^{1} p_{\text{label}} w_i \log(\text{p}_i) \tag{11}$$

We use the results of the main network output $p = \{p^l, p^{ul}\}$ to supervise the pseudo label p_{seudo} of the subnetwork PMAM sub-output and the spatial attentional feature

map $S_a = \{S_a^l, S_a^{ul}\}$ derived from the labeled data to supervise the pseudo label p_{seudo}, where l represents the dataset consisting of labeled images, ul represents the dataset consisting of pseudo images. The triangular self-supervised cross loss (TCSLoss) consists of the attention supervision loss L_{s-1} and the pseudo-supervised loss L_{s-2}.

$$L_{\mathrm{s}-1} = \frac{1}{|N_l|}\sum_{l}\frac{1}{H\times W}\sum_{i=0}^{H\times W}\Big(L_{\mathrm{CE}}(p_{label}, S_a^l) + L_{\mathrm{CE}}(S_a^l, p^l)\Big) \tag{12}$$

$$L_{\mathrm{s}-2} = \frac{1}{|N_{ul}|}\sum_{ul}\frac{1}{H\times W}\sum_{i=0}^{H\times W}\Big(L_{\mathrm{CE}}(p^{ul}, p_{\mathrm{seudo}}) + L_{\mathrm{CE}}(S_a^{ul}, p_{\mathrm{seudo}})\Big) \tag{13}$$

$$L_{TCS} = L_{\mathrm{s}-1} + \lambda L_{\mathrm{s}-1} \tag{14}$$

where λ is the trade-off weight.

$$L_{total} = L_{CE} + L_{edge} + L_{TCS} \tag{15}$$

where p_{seudo} is the binarized mask of the corresponding true label and $|N_l|$ is the number of output pixels. L_{CE} takes the form of a cross-entropy loss function as the main supervised term for training. The model training integrated loss function is L_{total}.

III. EXPERIMENTS AND RESULTS

A. ACCURACY ASSESSMENT INDEXES

The performance of the model was estimated in the experiments with the evaluation metrics of the binary classification task, which included precision P_{re}, recall R_e, F_1 score, and intersection over union (IoU) I_{ou}.

$$P_{re} = \frac{TP}{TP+FP}\times 100\% \tag{16}$$

$$R_e = \frac{TP}{TP+FN}\times 100\% \tag{17}$$

$$I_{ou} = \frac{TP}{TP+FN+FP}\times 100\% \tag{18}$$

$$F_1 = \frac{2\times P_{re}\times R_e}{P_{re}+R_e}\times 100\% \tag{19}$$

where TP and TN are the pixel counts of successfully classified cropland information and other non-cultivated.

Land information, respectively, and FP and FN are the pixel counts of the cropland and other non-cropland that are not detected, respectively. P_{re} denotes the proportion of the examples classified as positive cases that are positive cases, and R_e evaluates the proportion of TP over entire cropland pixels in the actual land distribution. The F_1 score is a weighted numerical evaluation that takes into account precision and recall. The I_{ou} represents the ratio of intersection and union of predicted cropland and indeed cropland. To further assess the performance of different segmentation networks in extracting cropland information and measure the accuracy of classification, the separated kappa coefficient K_a (SeK) [47] and I_{ou} are used to construct a comprehensive extraction evaluation score S_c with the following equation.

$$K_a = \frac{2e^{(I_{ou}-1)}\times(TP\times TN - FN\times FP)}{(TP+FP)(FP+TN)+(TP+FN)(FN+TN)} \tag{20}$$

$$S_c = \gamma I_{ou} + (1-\gamma)K_a \tag{21}$$

where γ represents the influence proportion of each item in the composite score, $\gamma = 0.3$.

B. EXPERIMENTAL RESULTS

We evaluated the proposed approach using experiments on datasets from Denmark and parts of China. Several models with excellent performance in semantic segmentation and cropland classification are selected as comparison methods, which are PSPNet [48], DeepLab-v3+ [49], U-Net [26], SPANet [36], and MAENet [37] The superiority of our model is verified by comparison and quantitative evaluation of cropland extraction results obtained by different models.

1) DIFFERENT METHODS IN PARTS OF DENMARK

We first tested the dataset for Denmark using various methods. Figure 5 illustrates the extracted cropland results from DeepLab-v3+, MAENet, U-Net, PSPNet, and SPANet. While DeepLab-v3+ extracted a large area of cropland, the results were imprecise. MAENet and U-Net maintained the forecast integrity of the cropland distribution and had advantages in processing marginal details. However, all three methods, including SPANet, had a high missed detection rate. Although all six methods achieved good results in cropland extraction, our proposed MBSDANet model outperformed the others by identifying cropland and refining its edge information more completely. This was due to our model's enhanced boundary constraints and attention to detailed texture information.

The comparative results of quantitative evaluation on the Danish dataset are shown in Table 1, which shows the cropland extraction results of different models in terms of precision, recall, IoU, Kappa coefficient, and comprehensive classification evaluation score S_c. As can be seen from the table, most indices of SPANet and MBSDANet outperform other models. Although SPANet also achieved good results under the dual-path fusion structure, our proposed method improved both accuracy and recall by more than 1%. Compared with the SPANet index, the accuracy, F1 score, and IoU of the proposed method increased by 1.33%, 1.69%, and 2.75%, respectively.

2) DIFFERENT METHODS IN PARTS OF HEBEI

We employed a total of 6 highly effective methods to extract cropland from images captured by Sentinel-2A, GF-2, and Google Earth. We meticulously examined and evaluated the obtained results and quantitatively assessed the extracted data from a mixed dataset in Hebei Province. The cropland

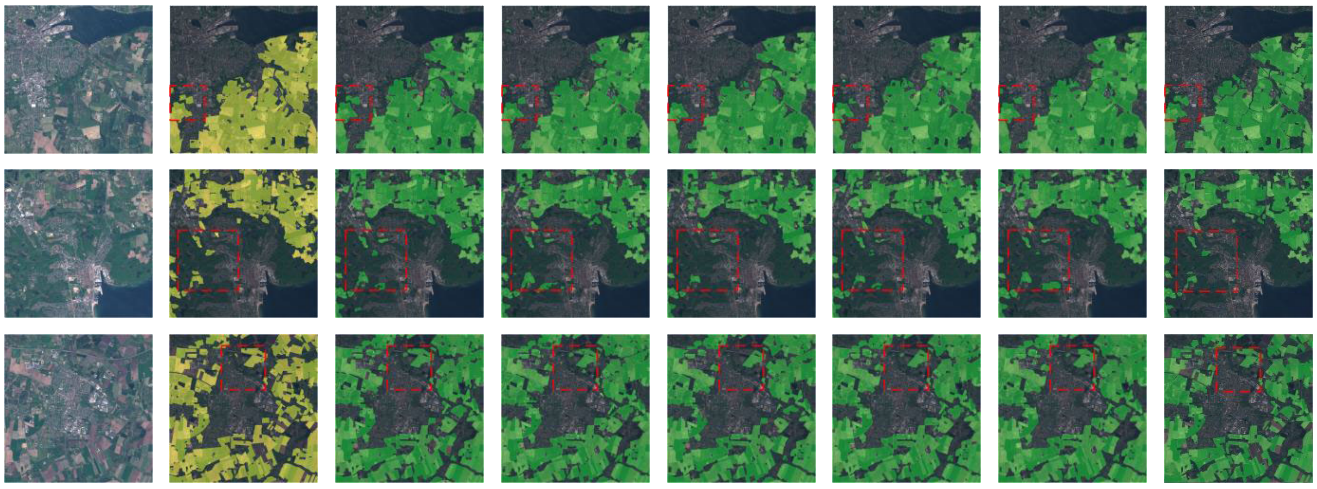


FIGURE 5. Examples of cropland extraction results obtained by different methods in the region of Denmark(The red dotted line box in the figure is used to emphasize the differences of cropland extraction results by different methods): (a) Original image; (b) Area of real cropland; (c) results of the SPANet; (d) results of the MAENet; (e) results of the U-Net; (f) results of the PSPNet; (g) results of the DeepLab-v3+; (h) results of the proposed method (MBSDANet).

TABLE 1. Quantitative evaluation for different methods in the Denmark.

Methods	Precision (%)	Recall (%)	IoU (%)	F1 (%)	Score (S_c)
SPANet	88.78	87.96	79.16	88.37	74.16
MAENet	87.59	86.78	77.28	87.18	72.27
U-Net	85.46	84.99	74.27	85.23	69.62
PSPNet	85.90	85.62	75.07	85.76	70.99
DeepLab-v3+	84.56	84.62	73.30	84.59	67.53
MBSDANet(ours)	**90.11**	**90.02**	**81.91**	**90.06**	**77.47**

extraction outcomes of the five methods in Shijiazhuang City, China, are presented in Figure 6 and Figure 7, the red dotted line box in the figure is used to emphasize the differences of cropland extraction results in Shijiazhuang by different methods. Figure 6 displays the cropland extraction results obtained from the Sentinel-2A satellite (resolution 10m), while Figure 7 shows the results from the Google Earth satellite (resolution 4m). It can be seen from Figure 6 and Figure 7 that SPANet has a poor effect on extracting details. Compared with the real label b, some of the ridges of the field are wrongly classified as cropland, as shown in the small red box below the subgraph in the first row of Figure 6. MAENet's prediction results for cropland extraction were incomplete, indicating poor contextual detail acquisition. U-Net and PSPNet's predictions were similar, and despite their internal model structures with contextual information aggregation modules, they still could not fully detect large areas of cropland. DeepLab-v3+ failed to accurately predict divided land types and could not completely extract cropland. Compared with other methods, our method uses edge constraints and multi-scale spatial context information extraction, which is conducive to paying attention to the aggregation of bilateral multi-sensory information, making the plot edge more refined, and extracting cultivated land information more complete.

Figure 8 presents the outcomes of five different extraction methods in the analysis of complex mountain-cropland data from the GF-2 satellite. Compared with the real cropland (b), SPANet, MAENet, and U-Net cropland extraction results have higher missing and error detection rates. Although PSPANet reduces the missing detection rate, the error detection rate and edge information loss are serious. Our model pays more attention to detailed information, thus enhancing the accuracy of extracting cropland information. While Deeplab-v3+ uses extended Convolution and extended space pyramid pool modules, its single inverse convolution in feature extraction disregards the correlation between adjacent pixels, leading to errors in edge prediction. Additionally, the simple global pool in the multi-scale feature fusion module overlooks the spatial context information of some feature maps. Our model addresses these issues by enhancing boundary constraints and focusing on texture details, resulting in significantly improved identification and refinement of complex farmland edges. Through the above analysis, the proposed method can locate the cultivated land boundary more accurately under different scenarios, which proves the effectiveness of this method in extracting cropland.

Table 2 and Table 3 show the quantitative evaluation results of the Hebei mixed dataset and GF-2 dataset under different methods respectively. PSPNet uses a pyramid pool module to combine global features and context information and obtains better prediction results. However, PSPNet and U-Net have some classification errors on roads and cropland edges. SPANet uses bidirectional continuous pooled attention

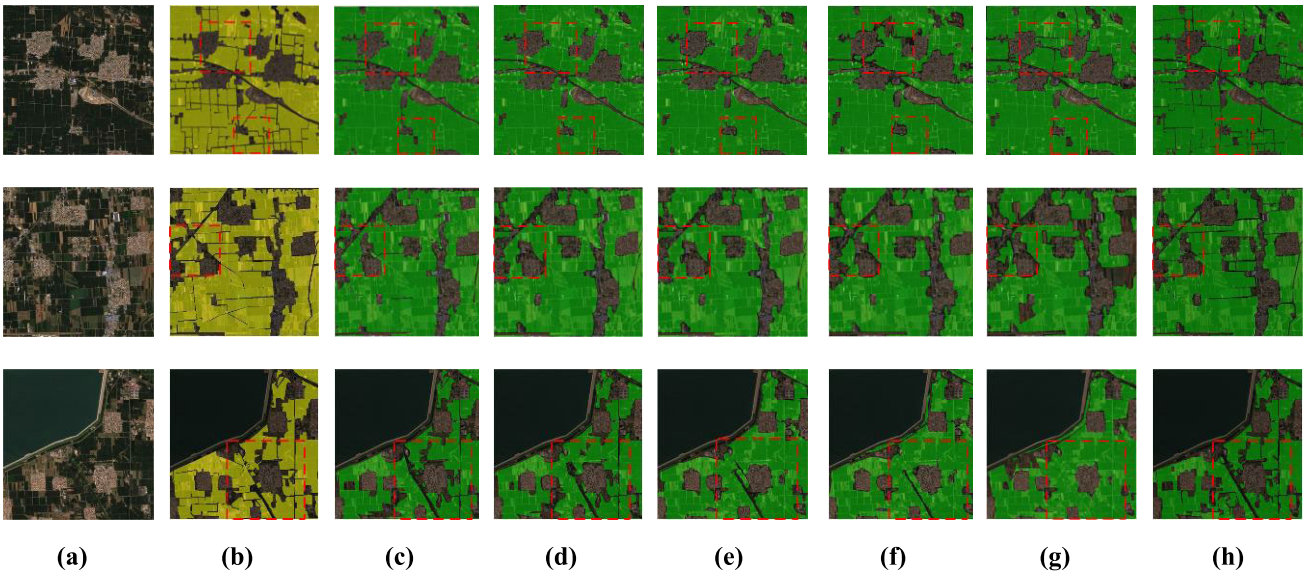


FIGURE 6. Comparative results of different methods to obtain examples of cropland extraction in the region of Hebei, China from Sentinel-2A (The red dotted line box in the figure is used to emphasize the differences of cropland extraction results by different methods): (a) Original image; (b) Area of real cropland; (c) results of the SPANet; (d) results of the MAENet; (e) results of the U-Net; (f) results of the PSPNet; (g) results of the DeepLab-v3+; (h) results of the proposed method (MBSDANet).

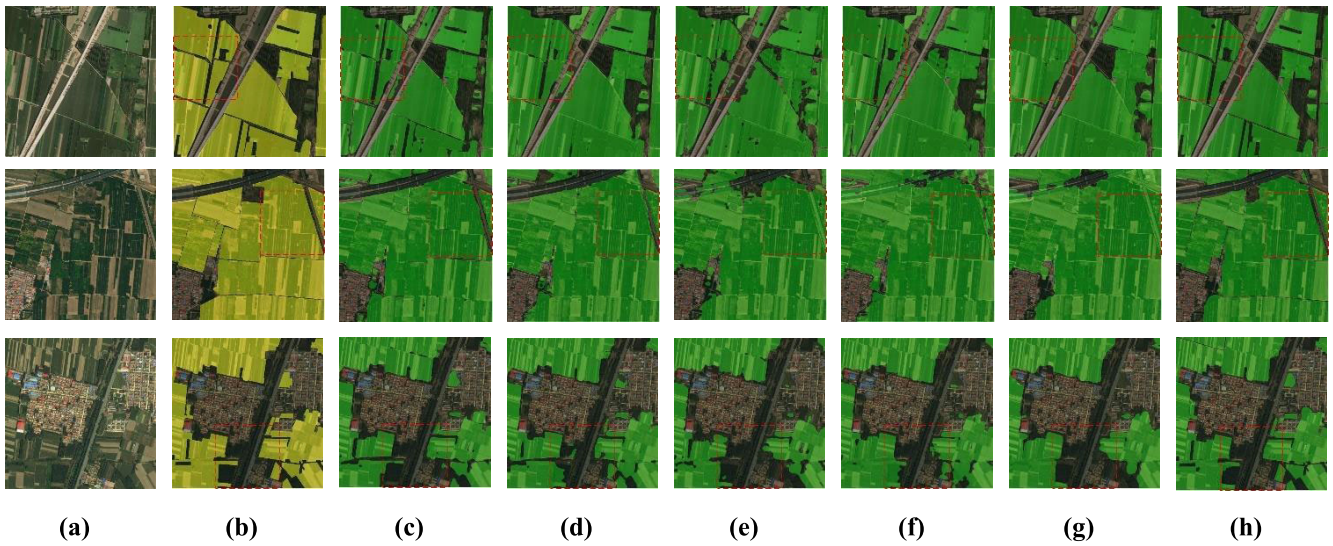


FIGURE 7. Comparative results of different methods to obtain examples of cropland extraction in the region of Hebei, China from Google Earth (The red dotted line box in the figure is used to emphasize the differences of cropland extraction results by different methods): (a) Original image; (b) Area of real cropland; (c) results of the SPANet; (d) results of the MAENet; (e) results of the U-Net; (f) results of the PSPNet; (g) results of the DeepLab-v3+; (h) results of the proposed method (MBSDANet).

modules to merge low-level and high-level features to achieve target extraction, which is superior to MAENet in small-scale target segmentation. MBSDANet makes the segmentation of cultivated and non-cropland more accurate by enhancing the extraction of spatial pixel context features. MBSDANet reduces the segmentation error rate by cross-supervision and edge refinement. Under the Hebei Province mixed (GF-2 dataset), our method obtained 94.81% (90.85%) accuracy, 89.37% (84.07%) IoU, and 86.51 (79.92) scores. Compared with SPANet with superior performance, F1 scores were 1.73% (1.41%) higher. IoU increased by 3.06% (1.27%). Our model has better ability to extract common cropland than complex cropland under GF-2 dataset. However, compared with other models, it can still extract complex cropland information more thoroughly. At the same time, our model also has better prediction results than other models under different-resolution mixed datasets. These data show that our model has stronger generalization ability and robustness.

In order to show the superiority of our model more intuitively. We used root mean square error (RMSE) to represent the extraction accuracy of different models under Sentinel-2A and GF-2 data sources, as shown in Figure 9. The figure

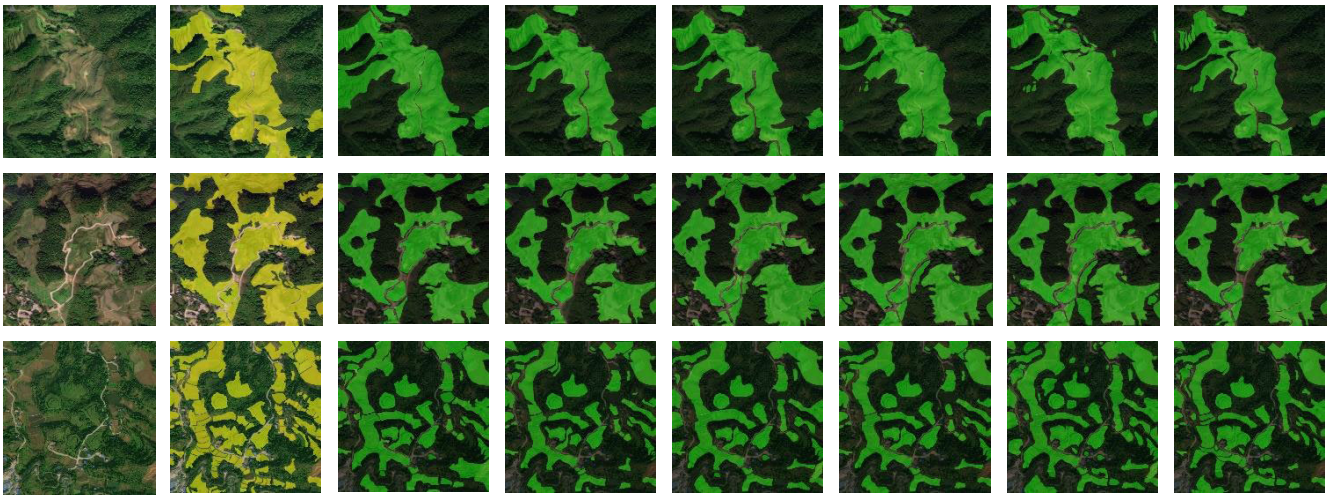


FIGURE 8. Comparative results of different methods to obtain examples of cropland extraction in the mountainous region from GF-2: (a) Original image; (b) Area of real cropland; (c) results of the SPANet; (d) results of the MAENet; (e) results of the U-Net; (f) results of the PSPNet; (g) results of the DeepLab-v3+; (h) results of the proposed method (MBSDANet).

TABLE 2. Quantitative evaluation for different methods in the HeBei.

Methods	Precision (%)	Recall (%)	IoU (%)	F1 (%)	Score (S_c)
SPANet	92.69	92.61	86.31	92.65	82.67
MAENet	93.05	91.96	86.14	92.55	82.60
U-Net	89.89	90.82	82.41	90.35	77.97
PSPNet	90.29	89.21	81.08	89.75	76.82
DeepLab-v3+	87.81	88.59	78.89	88.20	74.24
MBSDANet(ours)	**94.81**	**93.96**	**89.37**	**94.38**	**86.51**

TABLE 3. Quantitative evaluation for different methods on the GF-2.

Methods	Precision (%)	Recall (%)	IoU (%)	F1 (%)	Score (s_C)
SPANet	89.89	89.84	82.80	89.93	76.88
MAENet	89.38	88.80	80.41	89.14	75.62
U-Net	88.42	87.94	79.20	88.39	74.15
PSPNet	88.40	87.31	78.38	87.85	73.21
DeepLab-v3+	86.53	86.23	76.03	86.38	70.95
MBSDANet(ours)	**90.85**	**91.83**	**84.07**	**91.34**	**79.92**

shows that the model proposed in this paper has the lowest RMSE (36.96 and 47.77) under different data sources, indicating that its extracted area value is closest to the measurement label.

3) RESULTS OF CROPLAND EXTRACTION

We used the proposed method to extract uncropped cropland data in Denmark and Hebei Province, China, and visually

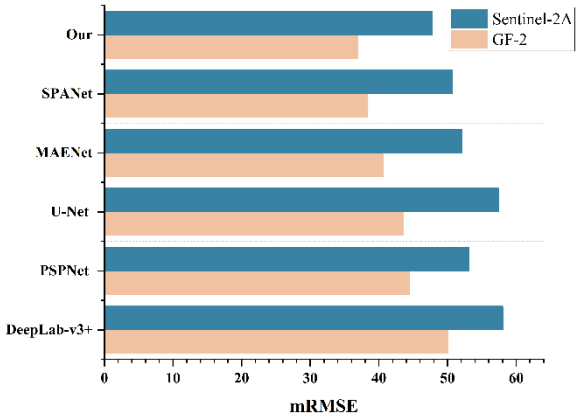


FIGURE 9. Comparison of geometric accuracy of different models.

demonstrated the cropland extraction capability of our proposed method. Figure 10 shows the extraction results of farmland near Odense, Denmark (the central city of Fyn Island) with uneven distribution and irregular edges, and cropland near Wuji County, Shijiazhuang, Hebei Province, with uniform distribution and regular edges extracted by our method, where (a) is cropland near Odense, Denmark (the central city of Fyn Island); (b) The results of MBSDANet method in field extraction in Odense, Denmark; (c) Cropland in Wuji County, Shijiazhuang City, Hebei Province; (d) Extraction results of MBSDANet from cropland in Wuji County. From the extraction results, our model can not only completely extract cropland with complex distribution and fuzzy edge information, but also accurately extract cultivated land with complex distribution and regular edge information. This also verifies that our method

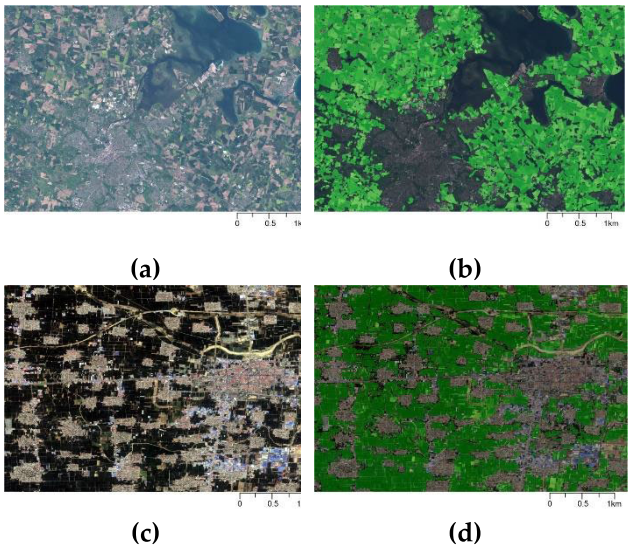


FIGURE 10. The results of cropland extraction in the study area: (a) The area around Odense, Denmark; (b) Results of the MBSDANet method on cultivated land extraction in Odense, Denmark; (c) The area around Wuji County, Shijiazhuang, Hebei Province, China; (d) Results of the MBSDANet method on cropland extraction in Wuji County.

has great application prospects for the efficient and accurate acquisition of cultivated land information, as well as farmland protection and agricultural yield estimation.

IV. DISCUSSION

A. CHOICE OF BACKBONE

The current excellent methods for extracting cropland, such as SPANet, MAENet, and PSPNet, utilize ResNet [50] as the backbone network. We conducted a comparison of ResNet and HRNet [51] as the encoder to determine a superior backbone for our model. Even though ResNet has strong feature encoding capabilities, HRNet performs than ResNet when parameters are similar. HRNet makes use of high-dimensional semantic information and low-dimensional semantic information that contains spatial information and integrates features of different levels to obtain feature maps with rich spatial information.

In order to demonstrate the superiority of our backbone network HRNet in cropland extraction, we compare it with ResNet under different datasets. Table 4 summarizes the evaluation results obtained by ResNet and HRNet under two datasets. The bolded scores are the optimal results. HRNet-w32 had an accuracy of 0.82% higher than ResNet-18 and 1.42% higher than IoU in Denmark. In China, HRNet-w32 had an accuracy of 1.85% higher than ResNet-18 and 2.62% higher than IoU. HRNet (HRNet-w32) had a precision of 94.81% in the last three stages, IoU reached 89.37%, and the overall evaluation score was 86.51 points. HRNet has better cropland extraction capability than ResNet in our network model, and HRNet-w32 is more suitable for our backbone network.

TABLE 4. Quantitative results of different backbone.

Region	Methods	Precision (%)	Recall (%)	IoU (%)	Score (Sc)
Denmark	ResNet-18	89.29	89.01	80.49	75.79
	ResNet-34	89.54	89.36	80.90	76.29
	HRNet-W18	90.03	89.95	81.54	77.03
	HRNet-W32	**90.11**	**90.00**	**81.91**	**77.47**
China	ResNet-18	92.96	92.84	86.75	83.32
	ResNet-34	93.45	93.13	87.47	85.58
	HRNet-W18	94.19	93.72	88.97	85.58
	HRNet-W32	**94.81**	**93.96**	**89.37**	**86.51**

TABLE 5. Accurate assessment of the comparison of different modules in ablation experiment.

Backbone	PMAM	SDAM	CAM	BeLoss	TCSLoss	S_c
√	√					76.99
√	√		√			78.56
√	√	√				81.37
√	√		√	√		79.55
√	√	√	√			82.33
√	√	√	√	√		83.95
√	√	√	√	√	√	**86.48**

B. ABLATION EXPERIMENT

To verify the design of MBSDANet, HRNet was selected as the baseline model and kept the same number of labeled samples as the Hebei Province dataset. The comprehensive extraction evaluation score S_c was adopted to assess the effectiveness quantitatively. The detailed evaluation results of the ablation experiment are summarized in Table 5, The bolded scores are the optimal results. The model performance was significantly improved by adding the attention mechanism module to PAMA. The addition of the SDAM module for the parallel fusion of features to capture the multiscale background significantly enhanced a score increase of 4.38%, implying that fusing cross-level elements is more advantageous than single-feature extraction. The binary edge cross-entropy loss (BeLoss) significantly improves the performance compared to the commonly used cross-entropy loss, with a 0.99% increase in score. In addition, with the triangle self-supervised cross-loss strategy, the model score was increased by 3.75%.

C. SMALL SAMPLE TRAINING

We validated the feasibility of small samples through experiments. Table 6 shows the quantitative evaluation results of the datasets composed of different proportions of labels 1/2, 1/4, 1/8, and 1/16 of the 1000 images in the Hebei mixed data training set were selected as the labeled data to construct

TABLE 6. Quantitative evaluation of label data of different proportions.

Methods	1/2		1/4		1/8		1/16	
	F1 (%)	Kappa (%)	F1 (%)	Kappa (%)	F1 (%)	Kappa (%)	F1(%)	Kappa (%)
MAENet	88.33	76.64	82.42	68.25	78.10	67.58	69.90	54.30
SPANet	83.25	76.53	83.26	69.84	79.01	68.44	70.61	56.42
SDLED	89.33	78.84	84.23	71.22	79.34	69.84	73.32	61.23
MBSDANet(ours)	**90.01**	**80.43**	**85.43**	**72.73**	**80.28**	**70.91**	**74.60**	**63.04**

the small sample dataset. The bolded scores are the optimal results. That is to say, the number of {labeled, unlabeled} samples in the Hebei mixed data training set are respectively {500, 500}, {250, 750}, {125, 875}, {63, 937}.

We compare supervised learning methods MAENet and SPANet with semi-supervised learning methods SDLED [51] and MBSDANet. The mentioned appeal methods are evaluated quantitatively to verify the superiority of the semi-supervised learning methods and our model. SPANet extracts significant features through continuous pooling and highlights edge information more than MAENet. Compared with SPANet, F1 (Kappa) of the proposed model increased by 6.76% (3.90%), 2.17% (2.89%), 1.27% (2.47%) and 3.99% (6.62%) respectively under the ratio of 1/2, 1/4, 1/8 and 1/16 data in Hebei province. Compared with SDLED, the model of F1 (Kappa) on four ratios was increased by 0.68% (1.59%), 1.20% (1.51%), 0.94% (1.07%), and 1.28% (1.81%). The smaller the sample size, the more obvious the superiority of our model over other models. When 1/2 of the labeled data is used, the F1 of the model can still reach 90.01%. The proposed method can effectively use unlabeled data for training and improve the ability of cropland extraction. The feasibility of the model's semi-supervised learning method and its strong robustness and generalization ability to small samples are verified.

V. CONCLUSION

We proposed a deep learning segmentation model MBSDANet for cropland extraction, which structurally extracts different levels of features through two different paths. PMAM enhances the multiscale representation and global semantic consistency of the model. The multiscale perception determines the generalizability of the model to different sizes of cropland, and the attention mechanism focuses on the retrieval of features to filter a small amount of meaningful information from a large amount of information by establishing channel-dimensional semantic correlations. SDAM aggregates a single pixel's non-local target spatial contextual features by calculating the correlations among all pixels in four directions in each pixel space to obtain detailed information on the cropland. It also changes the attention scope of cropland feature extraction from local to global so that the high-value information in the high-dimensional depth features of cropland in the image can be extracted quickly. We use a semi-supervised approach to reduce reliance on large samples. Compared with other methods, our model has the best performance for image extraction with different resolutions in complex scenes. In conclusion, MBSDANet can automatically extract cropland from high-resolution images efficiently and accurately. With the development of urban construction, the cropland area has been greatly reduced in Hebei Province, China, because of construction occupation and agricultural structure adjustment, and the cropland area has changed over the years. Rapid and accurate extraction of cultivated land based on MBSDANet shows potential for applications in the classification of cropland utilization, yield prediction of agricultural products, and agricultural land resource conservation.

REFERENCES

[1] X.-L. Chen, H.-M. Zhao, P.-X. Li, and Z.-Y. Yin, "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote Sens. Environ.*, vol. 104, no. 2, pp. 133–146, Sep. 2006, doi: 10.1016/j.rse.2005.11.016.

[2] H. Li, P. Xiao, X. Feng, Y. Yang, L. Wang, W. Zhang, X. Wang, W. Feng, and X. Chang, "Using land long-term data records to map land cover changes in China over 1981–2010," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1372–1389, Apr. 2017, doi: 10.1109/JSTARS.2016.2645203.

[3] J. Huang, J. L. Gómez-Dans, H. Huang, H. Ma, Q. Wu, P. E. Lewis, S. Liang, Z. Chen, J.-H. Xue, Y. Wu, F. Zhao, J. Wang, and X. Xie, "Assimilation of remote sensing into crop growth models: Current status and perspectives," *Agricult. Forest Meteorol.*, vols. 276–277, Oct. 2019, Art. no. 107609, doi: 10.1016/j.agrformet.2019.06.008.

[4] J. Huang, H. Ma, F. Sedano, P. Lewis, S. Liang, Q. Wu, W. Su, X. Zhang, and D. Zhu, "Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model," *Eur. J. Agronomy*, vol. 102, pp. 1–13, Jan. 2019, doi: 10.1016/j.eja.2018.10.008.

[5] J. Huang, F. Sedano, Y. Huang, H. Ma, X. Li, S. Liang, L. Tian, X. Zhang, J. Fan, and W. Wu, "Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation," *Agricult. Forest Meteorol.*, vol. 216, pp. 188–202, Jan. 2016, doi: 10.1016/j.agrformet.2015.10.013.

[6] F. Waldner and F. Diakogiannis, "Extracting field boundaries from satellite imagery with a convolutional neural network to enable smart farming at scale," in *Proc. EGU General Assembly Conf. Abstr.*, 2020, p. 102, doi: 10.5194/egusphere-egu2020-102.

[7] J. Huang, H. Ma, W. Su, X. Zhang, Y. Huang, J. Fan, and W. Wu, "Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4060–4071, Aug. 2015, doi: 10.1109/JSTARS.2015.2403135.

[8] J. Huang, L. Tian, S. Liang, H. Ma, I. Becker-Reshef, Y. Huang, W. Su, X. Zhang, D. Zhu, and W. Wu, "Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model," *Agricult. Forest Meteorol.*, vol. 204, pp. 106–121, May 2015, doi: 10.1016/j.agrformet.2015.02.001.

[9] R. Harb and P. Knöbelreiter, "InfoSeg: Unsupervised semantic image segmentation with mutual information maximization," in *Pattern Recognition*, C. Bauckhage, J. Gall, and A. Schwing, Eds. Cham, Switzerland: Springer, 2021, pp. 18–32.

[10] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S^4L: Self-supervised semi-supervised learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1476–1485, doi: 10.1109/ICCV.2019.00156.

[11] D. Zhu, A. Ge, X. Chen, Q. Wang, J. Wu, and S. Liu, "Supervised contrastive learning with angular margin for the detection and grading of diabetic retinopathy," *Diagnostics*, vol. 13, no. 14, p. 2389, Jul. 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/14/2389

[12] Y. Xue, J. Zhao, and M. Zhang, "A watershed-segmentation-based improved algorithm for extracting cultivated land boundaries," *Remote Sens.*, vol. 13, no. 5, p. 939, Mar. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/5/939

[13] A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2514–2520, Nov. 2001, doi: 10.1109/36.964989.

[14] T. Su, H. Li, S. Zhang, and Y. Li, "Image segmentation using mean shift for extracting croplands from high-resolution remote sensing imagery," *Remote Sens. Lett.*, vol. 6, no. 12, pp. 952–961, Dec. 2015, doi: 10.1080/2150704X.2015.1093188.

[15] 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. [Online]. Available: https://www.mdpi.com/2072-4292/13/6/1167

[16] M. A. Peña and A. Brenning, "Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile," *Remote Sens. Environ.*, vol. 171, pp. 234–244, Dec. 2015, doi: 10.1016/j.rse.2015.10.029.

[17] 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.

[18] P. Teluguntla, P. S. Thenkabail, A. Oliphant, J. Xiong, M. K. Gumma, R. G. Congalton, K. Yadav, and A. Huete, "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.*, vol. 144, pp. 325–340, Oct. 2018, doi: 10.1016/j.isprsjprs.2018.07.017.

[19] F. Waldner, G. S. Canto, and P. Defourny, "Automated annual cropland mapping using knowledge-based temporal features," *ISPRS J. Photogramm. Remote Sens.*, vol. 110, pp. 1–13, Dec. 2015, doi: 10.1016/j.isprsjprs.2015.09.013.

[20] G. Chen, X. Tan, B. Guo, K. Zhu, P. Liao, T. Wang, Q. Wang, and X. Zhang, "SDFCNv2: An improved FCN framework for remote sensing images semantic segmentation," *Remote Sens.*, vol. 13, no. 23, p. 4902, Dec. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/23/4902

[21] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3051053.

[22] M. Yuan, Q. Zhang, Y. Li, Y. Yan, and Y. Zhu, "A suspicious multi-object detection and recognition method for millimeter wave SAR security inspection images based on multi-path extraction network," *Remote Sens.*, vol. 13, no. 24, p. 4978, Dec. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/24/4978

[23] Z. Zhang, S. Liu, Y. Zhang, and W. Chen, "RS-DARTS: A convolutional neural architecture search for remote sensing image scene classification," *Remote Sens.*, vol. 14, no. 1, p. 141, Dec. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/14/1/141

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[25] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, Mar. 2019, Art. no. 014006, doi: 10.1117/1.JMI.6.1.014006.

[26] W. Weng and X. Zhu, "INet: Convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[27] L. Yan, D. Liu, Q. Xiang, Y. Luo, T. Wang, D. Wu, H. Chen, Y. Zhang, and Q. Li, "PSP net-based automatic segmentation network model for prostate magnetic resonance imaging," *Comput. Methods Programs Biomed.*, vol. 207, Aug. 2021, Art. no. 106211.

[28] 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.

[29] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[30] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, "A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net," *Comput. Electron. Agricult.*, vol. 189, Oct. 2021, Art. no. 106373, doi: 10.1016/j.compag.2021.106373.

[31] 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.

[32] 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.

[33] R. Lu, N. Wang, Y. Zhang, Y. Lin, W. Wu, and Z. Shi, "Extraction of agricultural fields via DASFNet with dual attention mechanism and multi-scale feature fusion in South Xinjiang, China," *Remote Sens.*, vol. 14, no. 9, p. 2253, May 2022, doi: 10.3390/rs14092253.

[34] Q. Yang, B. She, L. Huang, Y. Yang, G. Zhang, M. Zhang, Q. Hong, and D. Zhang, "Extraction of soybean planting area based on feature fusion technology of multi-source low altitude unmanned aerial vehicle images," *Ecological Informat.*, vol. 70, Sep. 2022, Art. no. 101715, doi: 10.1016/j.ecoinf.2022.101715.

[35] D. Zhang, Y. Pan, J. Zhang, T. Hu, J. Zhao, N. Li, and Q. Chen, "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.

[36] L. Sun, S. Cheng, Y. Zheng, Z. Wu, and J. Zhang, "SPANet: Successive pooling attention network for semantic segmentation of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4045–4057, 2022, doi: 10.1109/JSTARS.2022.3175191.

[37] H. Huan, Y. Liu, Y. Xie, C. Wang, D. Xu, and Y. Zhang, "MAENet: Multiple attention encoder–decoder network for farmland segmentation of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3137522.

[38] K. Cao and X. Zhang, "An improved res-UNet model for tree species classification using airborne high-resolution images," *Remote Sens.*, vol. 12, no. 7, p. 1128, Apr. 2020, doi: 10.3390/rs12071128.

[39] Z. Xu, W. Zhang, T. Zhang, and J. Li, "HRCNet: High-resolution context extraction network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 13, no. 1, p. 71, Dec. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/13/1/71

[40] X. Hu, C.-W. Fu, L. Zhu, J. Qin, and P.-A. Heng, "Direction-aware spatial context features for shadow detection and removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2795–2808, Nov. 2020, doi: 10.1109/TPAMI.2019.2919616.

[41] C. Andreasen and I. M. Skovgaard, "Crop and soil factors of importance for the distribution of plant species on arable fields in Denmark," *Agricult., Ecosyst. Environ.*, vol. 133, nos. 1–2, pp. 61–67, Sep. 2009, doi: 10.1016/j.agee.2009.05.003.

[42] Y. Rong, P. Du, F. Sun, and S. Zeng, "Quantitative analysis of economic and environmental benefits for land fallowing policy in the Beijing-Tianjin-Hebei Region," *J. Environ. Manag.*, vol. 286, May 2021, Art. no. 112234, doi: 10.1016/j.jenvman.2021.112234.

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

[44] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[46] G. M. H. Amer and A. M. Abushaala, "Edge detection methods," in *Proc. 2nd World Symp. Web Appl. Netw. (WSWAN)*, Mar. 2015, pp. 1–7, doi: 10.1109/WSWAN.2015.7210349.
[47] K. Yang, G.-S. Xia, Z. Liu, B. Du, W. Yang, M. Pelillo, and L. Zhang, "Semantic change detection with asymmetric Siamese networks," 2020, *arXiv:2010.05687*.
[48] R. Zhang, J. Chen, L. Feng, S. Li, W. Yang, and D. Guo, "A refined pyramid scene parsing network for polarimetric SAR image semantic segmentation in agricultural areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3086117.
[49] W. Liu, A. Yue, W. Shi, J. Ji, and R. Deng, "An automatic extraction architecture of urban green space based on DeepLabv3plus semantic segmentation model," in *Proc. IEEE 4th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2019, pp. 311–315, doi: 10.1109/ICIVC47709.2019.8981007.
[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[51] L. Xia, X. Zhang, J. Zhang, H. Yang, and T. Chen, "Building extraction from very-high-resolution remote sensing images using semi-supervised semantic edge detection," *Remote Sens.*, vol. 13, no. 11, p. 2187, Jun. 2021, doi: 10.3390/rs13112187.

WEIMIN HOU received the Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2007.

He is currently a Professor with the School of Information Science and Engineering, Hebei University of Science and Technology. He is also a Distinguished Researcher with the Hangzhou Research Institute, Beihang University. His research interests include array signal processing, wireless communication, remote sensing image processing, and artificial intelligence.

YANXIA WANG received the B.S. degree from the North China Institute of Aerospace Engineering, Hebei, China, in 2021. She is currently pursuing the M.E. degree with the Hebei University of Science and Technology.

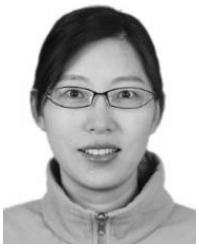
Her research interests include deep learning, computer vision, remote sensing image processing, and object detection.

JIA SU received the Ph.D. degree in communication and information systems from Harbin Engineering University, Heilongjiang, China, in 2010. She is currently a Professor with the School of Information Science and Engineering, Hebei University of Science and Technology. She has presided over One Key Research and Development Project in Hebei Province and One Key Project of the Hebei Provincial Department of Education. Her research interests include multi-antenna array, image processing, object detection, and object tracking.

YANLI HOU received the Ph.D. degree in signal and information from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China, in 2008. She is currently an Associate Professor with the School of Information Science and Engineering, Hebei University of Science and Technology. Her research interests include wireless communication technology, radio direction finding technology, and image processing.

MING ZHANG received the B.S. and Ph.D. degrees from the College of Optoelectronics Engineering, Chongqing University, Chongqing, China, in 2010, and 2019, respectively.

From 2016 to 2019, he underwent a joint Doctoral Training Program with the Institute of Optics and Electronics, Chinese Academy of Sciences. In 2019, he joined the School of Information Science and Engineering, Hebei University of Science and Technology, where he is currently an Associate Professor. His research interests include new optoelectronic devices, wireless communications, and subwavelength electromagnetics.



YAN SHANG received the B.S. degree in electric information engineering and the master's degree in circuits and systems from Yanshan University, Qinhuangdao, China, in 2004 and 2007, respectively.

She joined the School of Information Science and Engineering, Hebei University of Science and Technology, in 2007, where she is currently a Lecturer. Her research interests include signal processing and image processing.