SAB-DeepLabV3+: A Semantic Segmentation Framework for Mapping Maize Waterlogging from Single-Date Multispectral Imagery

Jiahao An , Qingxue Wang * , Chunshan Wang * , Xiang Sun , Qingwei Tian , Jin Yuan

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1571.v1

Keywords: maize waterlogging; single-date multispectral imagery; semantic segmentation; DeepLabV3+; cross-regional generalization; agricultural disaster remote sensing

 doi:10.20944/preprints202604.1571.v1

Article

SAB-DeepLabV3+: A Semantic Segmentation Framework for Mapping Maize Waterlogging from Single-Date Multispectral Imagery

Jiahao An [1], Qingxue Wang [2,3,*], Chunshan Wang [2,3], Xiang Sun [4,5], Qingwei Tian [2,3] and Jin Yuan [1]

[1] North Alabama International College of Engineering and Technology, Guizhou University, Guiyang, 550025, China
[2] College of Information Science and Technology, Hebei Agricultural University, Baoding, 071001, China
[3] Agricultural Remote Sensing Application Hebei Engineering Research Center, Baoding, 071001, China
[4] Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[5] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
* Correspondence: 20247060930@pgs.hebau.edu.cn; Tel.: +0312-7626815

Abstract

Rapid identification of maize waterlogging is essential for post-disaster agricultural assessment, but most existing methods rely on multi-temporal imagery that is often unavailable immediately after extreme rainfall events. This study proposes SAB-DeepLabV3+, a semantic segmentation model for mapping waterlogged maize from single-date multispectral imagery within pre-extracted maize planting areas. Built on DeepLabV3+, the model integrates three task-specific modules: a Spectral-Spatial Information Enhancement Module to improve feature discrimination under spectral mixing, an Adaptive Multi-Scale Pooling Module to capture heterogeneous patch sizes, and a Boundary Enhancement Module to refine transition zones. A pixel-level dataset containing 12,198 image patches was constructed from 62 multispectral scenes collected across five major maize-producing cities in Heilongjiang Province, China, during 2022–2024. On the test set, SAB-DeepLabV3+ achieved a waterlogged-class IoU of 68.30%, mIoU of 80.37%, mF1 of 88.62%, and OA of 93.49%, outperforming DeepLabV3+. Leave-one-city-out evaluation further produced an average mIoU of 76.56% and a waterlogged-class IoU of 63.45%. These results indicate that single-date high-resolution multispectral imagery can support rapid and reliable maize waterlogging mapping.

Keywords: maize waterlogging; single-date multispectral imagery; semantic segmentation; DeepLabV3+; cross-regional generalization; agricultural disaster remote sensing

1. Introduction

Agricultural waterlogging stress is one of the major water stress types restricting stable and high yields of dryland crops such as maize[1]. It is usually driven by factors including short-duration heavy rainfall, rising groundwater levels, and insufficient farmland drainage. After waterlogging stress occurs, the root of maize remains in a state of high water content or hypoxia for a long time[2], thus inhibiting root respiration and nutrient uptake[2]. This can further lead to stunted growth, increased lodging risk, and even yield reduction. In high-latitude, flat, and large-scale agricultural areas, such disasters often occur rapidly, affect a wide range, and exhibit a highly heterogeneous spatial distribution[4]. In recent years, the frequent occurrence of extreme precipitation events in Northeast China has placed continuous pressure of waterlogging stress on regional food security and agricultural risk management[5]. Therefore, it is of great practical significance to develop a high-precision and rapidly deployable waterlogging disaster identification method.

 doi:10.20944/preprints202604.1571.v1

Remote sensing technology provides key support for agricultural disaster monitoring with its wide coverage and rapid acquisition capabilities. Existing studies mostly rely on multi-temporal optical images to characterize crop stress through time-series changes in vegetation indices, water indicators, or soil moisture indicators, and have made remarkable progress in disaster evolution analysis[6,7]. However, multi-temporal methods usually depend on continuous observation data[8] and are susceptible to cloud cover, image acquisition intervals, and phenological differences[9], making it difficult to meet the demands of high timeliness and fine mapping in post-disaster emergency assessment. In practical production scenarios, only single-scene images can usually be obtained in a timely manner, which raises a scientific question: in the absence of time-series information, does single-date, high-resolution multispectral imagery contain sufficient spectral-spatial discriminative information to support reliable identification of agricultural waterlogging stress[10–12]? This issue essentially involves judging the information sufficiency and feature separability under single-date conditions, that is, whether the spectral response differences and spatial structural features in multispectral imagery can be fully exploited to form stable discriminative boundaries without temporal information support. Specifically, the key lies not in superficial improvements to model structures, but in constructing deep feature representation mechanisms for spectral-spatial synergistic enhancement[13].

Existing methods for identifying agricultural waterlogging stress can be broadly divided into three categories[14]. The first category includes methods based on spectral indices or threshold rules, which are simple to implement and highly interpretable, but sensitive to threshold settings and susceptible to soil background, crop coverage, and phenological differences, resulting in limited stability in complex farmland environments[15]. The second category consists of change detection methods using multi-temporal or multi-source data (e.g., SAR), which can strengthen disaster signals but usually rely on continuous observations or multi-source data collaboration, imposing certain constraints on data timeliness and cross-sensor consistency[16,17]. The third category involves deep learning-based semantic segmentation methods, which significantly improve object extraction accuracy from high-resolution images through multi-scale context modeling and have been widely applied in agricultural remote sensing scenarios[18]. However, in the task of single-date maize waterlogging stress identification, insufficient attention has been paid to spectral-spatial joint discrimination mechanisms, scale adaptability, and fine boundary representation. Overall, current research lacks systematic verification of the information sufficiency of single-date multispectral imagery, as well as spectral-spatial collaborative modeling schemes for mixed spectral responses under vegetation-covered conditions.

To address the above issues, this study proposed a spectral-spatial and boundary-enhanced semantic segmentation framework, namely SAB-DeepLabV3+ (Spectral-spatial and Boundary-enhanced DeepLabV3+), based on DeepLabV3+. This method improves the identification accuracy and spatial integrity of waterlogged regions in complex farmland scenarios via spectral-spatial information enhancement, adaptive multi-scale context modeling, and boundary-aware feature optimization. The main contributions of this paper are as follows:

(1) A single-date multispectral dataset for maize waterlogging stress was constructed, covering five prefecture-level cities in Heilongjiang Province of China, with typical events from 2022 to 2024, including pixel-level annotation, conventional testing, and cross-regional generalization validation.

(2) A maize waterlogging stress identification framework, namely SAB-DeepLabV3+, was proposed, with three customized modules: SSIEM (Spectral-Spatial Information Enhancement Module), AMSP (Adaptive Multi-Scale Pyramid Module), and BEAM (Boundary Enhancement Attention Module), targeting spectral aliasing, scale heterogeneity, and boundary blurring, respectively.

(3) The effectiveness of the proposed method was systematically evaluated through conventional tests and cross-regional independent tests, confirming the application potential of single-date RGB-NIR imagery for rapid maize waterlogging mapping.

 doi:10.20944/preprints202604.1571.v1

2. Materials and Methods

2.1. Study Area

In this study, five typical major maize-producing prefecture-level cities in Heilongjiang Province were selected as the study area, including Daqing, Qiqihar, Mudanjiang, Jixi and Hegang (Figure 1). All these cities are located in the core commodity grain production area of Northeast China, with large maize planting areas. In recent years, these cities have been affected by varying degrees of heavy rainfall and field waterlogging during the maize growing season, making them ideal for regional-scale research on maize waterlogging monitoring.

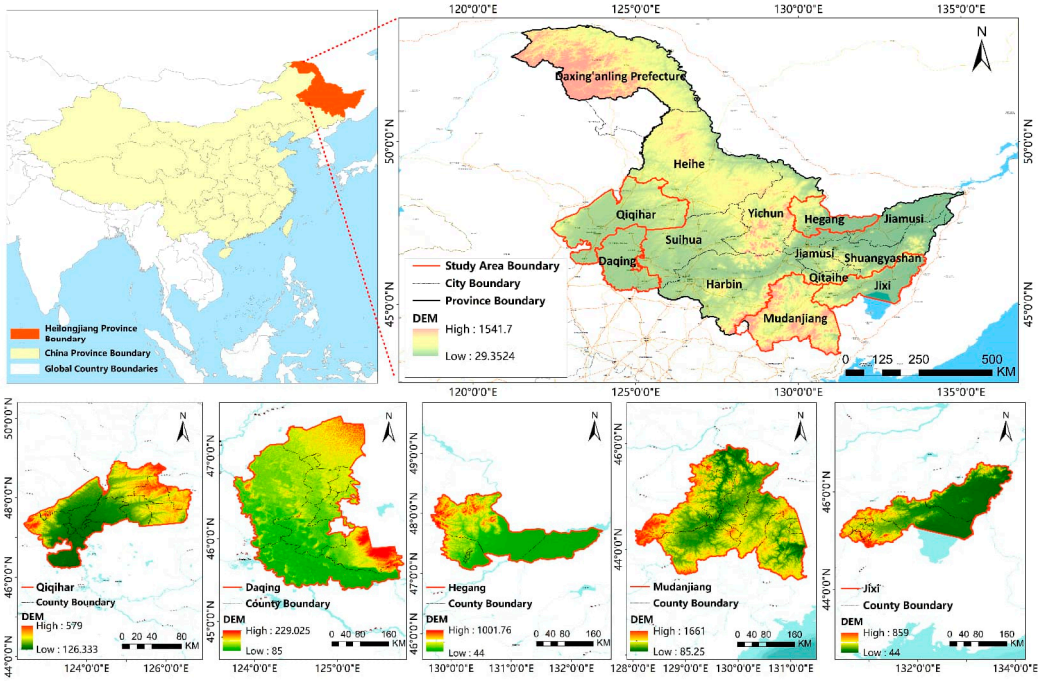

Figure 1. Distribution map of the study areas

The study area features diverse landform types, covering typical agricultural ecological units such as the low-lying agricultural area in the central-western Songnen Plain, the western margin of the Sanjiang Plain, and the piedmont hilly transition zone of the Lesser Khingan Mountains. Specifically, among the five cities included, Daqing and Qiqihar have relatively small topographic relief and are prone to persistent plain waterlogging in years with excessive precipitation; Mudanjiang, Jixi, and Hegang are dominated by mountain-hill-plain composite landforms, with waterlogging processes mostly manifested as slope runoff convergence, stagnant water in low-lying areas, and poor farmland drainage conditions.

Climatically, the study area has a temperate continental monsoon climate with uneven annual precipitation distribution, concentrated summer rainfall, and prominent interannual fluctuations. In the middle and late growth stages of maize, short-duration heavy rainfall or persistent overcast rain can easily induce field waterlogging and rhizosphere stress, exerting cumulative effects on crop physiological activities and canopy structure[19]. Variations in precipitation conditions, topographic convergence patterns, and farmland water conservancy infrastructure among different cities lead to significant heterogeneity in the spatial distribution, patch morphology, and damage severity of waterlogging impacts[20].

Overall, the study area can well represent the typical environmental conditions where maize waterlogging stress mainly occurs in Heilongjiang Province, providing a representative application background for subsequent construction of remote sensing samples and methodical validation.

 doi:10.20944/preprints202604.1571.v1

2.2. Data Sources and Dataset Construction

This study constructed a cross-regional remote sensing dataset for maize waterlogging stress, covering single-date multispectral images acquired after typical waterlogging events during the growing seasons from 2022 to 2024 in the study area. The image acquisition time was mainly concentrated from mid-to-late August to early September. During this period, maize is generally in the middle and late growth stages, and waterlogging stress has exerted obvious effects on canopy structure, leaf pigment status, and field background. Specifically, waterlogged and non-waterlogged regions exhibit distinct separability in visible and near infrared (NIR) reflectance characteristics and spatial texture, providing a solid data foundation for pixel-level waterlogging identification.

The remote sensing data used in this study were all obtained from the Power Engineering Satellite of the National Grid special remote sensing satellite constellation, with a spatial resolution of 4 m and four spectral bands (Blue, Green, Red and NIR). After screening for cloud cover, imaging quality and disaster coverage, 62 scenes of images were finally selected for dataset construction, including 11 scenes in Daqing, 15 in Hegang, 15 in Jixi, 12 in Mudanjiang and 9 in Qiqihar. Image preprocessing was performed on the ENVI 5.6 platform, mainly including radiometric calibration and atmospheric correction, to enhance the reliability of quantitative analysis and spatial consistency of the images.

Sample labeling was performed via manual fine interpretation based on multispectral images. First, boundaries of maize planting regions were extracted by combining agricultural survey data to further generate planting region masks. Then, non-waterlogged maize and waterlogging-affected maize were distinguished according to the typical spectral anomalies, texture changes and field morphological characteristics under waterlogging stress. Therefore, the task in this paper was not full-scene land cover classification, but pixel-level binary semantic segmentation between non-waterlogged maize and waterlogging-affected maize samples within pre-extracted maize planting regions. Initial labeling was independently completed by five annotators and reviewed by three experts in related fields to ensure label consistency and reliability. To further ensure labeling quality, about one-third of typical sample regions in each city were selected for verification through drone aerial photography and field sampling within 3 days after waterlogging occurrence, and the initial interpretation labels were revised pixel by pixel according to the verification results. The final labeling system was defined as two categories: non-waterlogged maize (assigned 0) and waterlogging-affected maize (assigned 1). After rasterization, the vector labels were registered with the corresponding multispectral images to generate standardized semantic segmentation labels. The specific process is shown in Figure 2. Figure 3 displays the overlay results of multispectral images and pixel-level labels for the visual verification of the consistency between label boundaries and image features.

 doi:10.20944/preprints202604.1571.v1

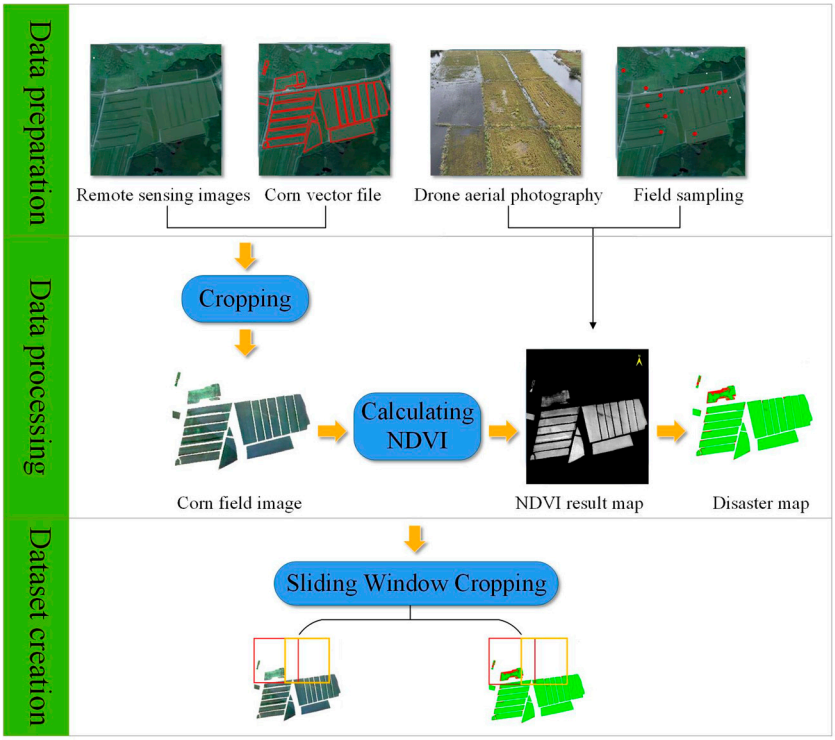

Figure 2. Workflow of dataset construction and labeling

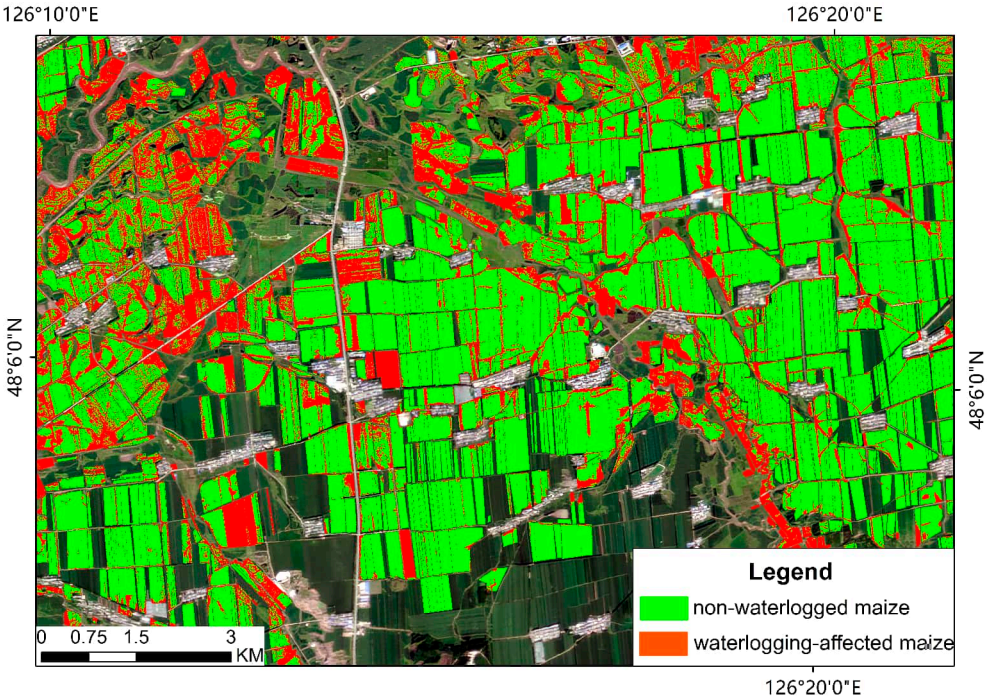

Figure 3. Overlay of multispectral imagery and pixel-wise waterlogging labels.

In the sample construction stage, the original images were first cropped using maize planting region masks to retain only the target crop regions. The images were then uniformly sliced into sample patches of 512×512 pixels, yielding a total of 12,198 valid samples. Considering the obvious spatial autocorrelation of remote sensing samples, a spatially non-overlapping tiling strategy was adopted in this study to construct the training, validation and test sets with a split ratio of 7:1:2. This is to ensure spatial non-overlapping among different subsets and reduce evaluation bias caused by

 doi:10.20944/preprints202604.1571.v1

spatial information leakage of adjacent samples. It should be noted that this splitting method belongs to spatially non-overlapping patch partitioning rather than fully scene-independent partitioning. Therefore, a leave-one-city-out cross-regional experiment was designed to verify the model's transferability from a stricter region-independent perspective. During the training phase, multiple data augmentation operations, including random scaling, flipping, brightness perturbation and contrast perturbation, were introduced to improve the model's adaptability to scale variations and illumination differences, thereby enhancing its generalization performance in complex regional scenarios.

2.3. Research Methods

2.3.1. DeepLabV3+

DeepLabV3+ is a classic encoder–decoder semantic segmentation framework that has been widely used in remote sensing feature extraction and agricultural scene segmentation tasks. The model expands the receptive field through dilated convolutions and aggregates multi-scale contextual information using the Atrous Spatial Pyramid Pooling (ASPP) strategy, thus enhancing semantic representation while maintaining relatively high feature resolution[21]. Subsequently, the decoder fuses the up-sampled high-level semantic features with low-level detail features to recover spatial resolution and improve the representation of boundary transition zones[22]. The overall network architecture of DeepLabV3+ is shown in Figure 4.

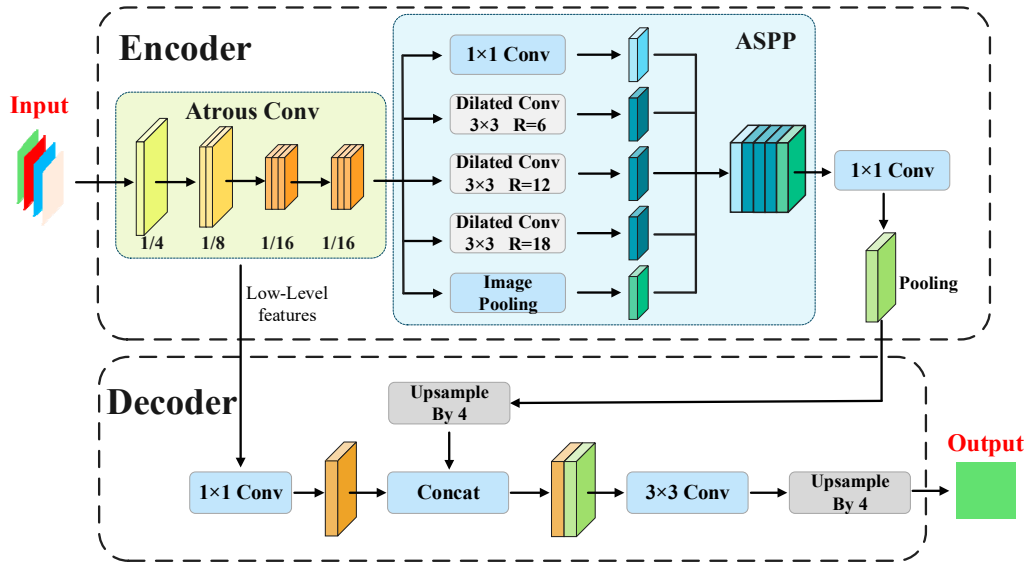

Figure 4. Architecture of the DeepLabV3+ network.

To balance segmentation accuracy and computational efficiency, we selected DeepLabV3+ as the baseline framework, with MobileNetV2 as the backbone network. Although DeepLabV3+ exhibits strong advantages in multi-scale feature modeling, it still suffers from three limitations in the task of single-date multispectral maize waterlogging mapping:

(1) Insufficient spectral discriminative modeling. The original DeepLabV3+ framework is designed mainly for natural images and cannot explicitly model the differential contributions of RGB-NIR bands in waterlogging identification.

(2) Relatively fixed scale response. The multi-branch receptive fields of the standard ASPP are determined by preset dilation rates, making it difficult to fully adapt to the significant differences in the area, shape, and spatial organization of waterlogged patches in farmland.

(3) Limited boundary recovery capability. When disaster-affected and non-disaster-affected regions show gradual transitions, relying solely on conventional decoder fusion makes it challenging to accurately delineate field boundaries and fragmented patches[23,24].

 doi:10.20944/preprints202604.1571.v1

To address the above issues, we introduced targeted enhancement modules at key feature flow nodes of DeepLabV3+ to form an improved framework.

2.3.2. SAB-DeepLabV3+

To tackle the common problems like spectral aliasing, large patch-scale variations, and boundary blurring in single-date multispectral maize waterlogging imagery, we proposed the SAB-DeepLabV3+ (Spectral-spatial and Boundary-enhanced DeepLabV3+) framework based on DeepLabV3+ (overall structure shown in Figure 5). The model retains the original encoder–decoder backbone and only embeds lightweight enhancement modules at three key positions to improve the model's adaptability to complex farmland scenes.

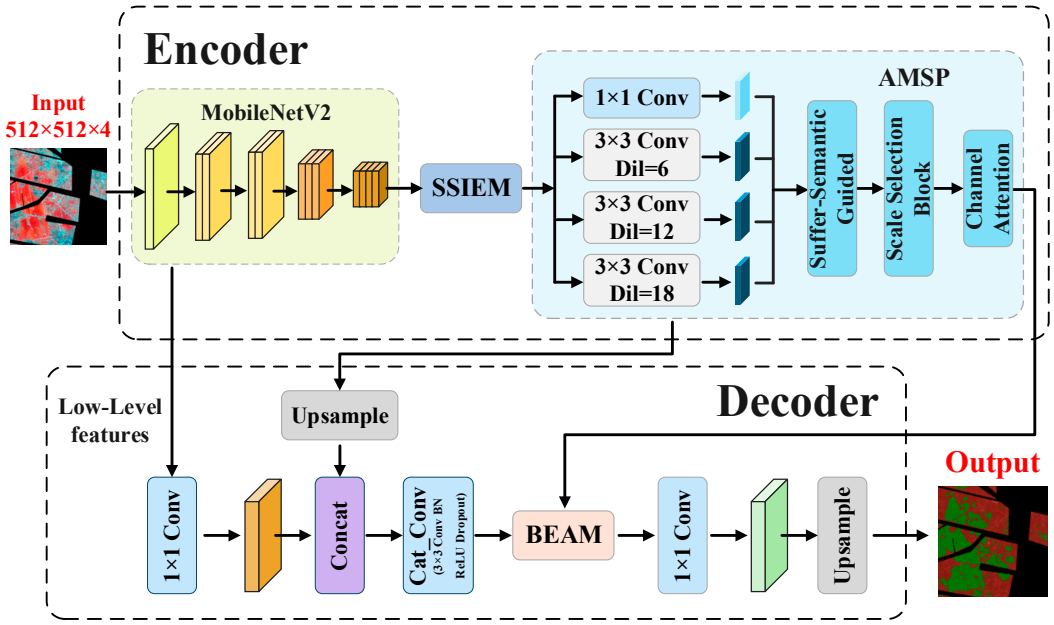

Figure 5. Overall architecture of the proposed SAB-DeepLabV3+ framework.

The SAB-DeepLabV3+ model still adopts MobileNetV2 as the backbone network. Specifically, the input image first passes through the encoder to extract low-level spatial detail features and high-level semantic features. Subsequently, the high-level semantic features sequentially pass through the Spectral-Spatial Information Enhancement Module (SSIEM) and the Adaptive Multi-Scale Pooling Module (AMSP) to improve discriminative representation and contextual modeling for disaster-affected regions. In the decoding stage, the up-sampled high-level features are fused with low-level detail features and further refined by the Boundary Enhancement Attention Module (BEAM) to sharpen the boundaries of waterlogged regions, finally outputting pixel-level segmentation results.

The three modules aforementioned correspond to three key challenges in single-date maize waterlogging identification:

(1) SSIEM operates on high-level encoder features and adaptively recalibrates different spectral channels and their spatial responses using global context, thereby strengthening the effective information related to waterlogging discrimination while suppressing redundant or interfering responses.

(2) AMSP replaces the standard ASPP module and dynamically learns the weights of different receptive field branches, thereby improving the model's ability to jointly adapt to large-scale contiguous waterlogged regions and small-scale fragmented disaster patches.

(3) BEAM operates on decoded fused features to strengthen the representation of transition zones between waterlogged and non-waterlogging-affected regions using boundary-sensitive information, thus improving boundary expression and shape recovery.

 doi:10.20944/preprints202604.1571.v1

Notably, the design goal of SAB-DeepLabV3+ is not simply to stack generic attention structures, but to implement task-driven structural improvements targeting specific identification challenges in single-date maize waterlogging mapping. While maintaining the stability of the overall network, the model balances segmentation accuracy, boundary quality, and computational efficiency. The detailed structures and mathematical formulations of SSIEM, AMSP, and BEAM are described in detail in the following subsections.

2.3.3. SSIEM Module

In the SAB-DeepLabV3+ framework, the high-level features output by the encoder carry not only spatial structural information but also implicit response differences of multispectral bands under varying crop physiological status and waterlogging stress conditions. However, in single-date agricultural remote sensing images, land cover types such as waterlogged vegetation, healthy vegetation, water bodies, and bare soil often exhibit significant overlaps in the spectral space, weakening the discriminative ability of some spectral channels for waterlogging identification and making them vulnerable to background interference.

Traditional convolutional neural networks (CNNs) generally treat multispectral channels as equivalent and deterministic input features, lacking explicit modeling of the differences in discriminative contribution among spectral channels. This tends to induce disaster signals to be masked by redundant background or ambiguous spectral information. In view of this, we introduced SSIEM at the output end of the encoder's high-level features to adaptively recalibrate different spectral channels and their spatial responses. This strategy can strengthen effective features related to waterlogging discrimination while suppressing redundant or interfering responses. The overall structure of SSIEM is shown in Figure 6.

Let the high-level features output by the encoder be expressed as:

$$F \in \mathbb{R}^{C \times H \times W} \tag{1}$$

Where C, H, and W represent the number of channels, feature map height, and width, respectively.

First, Global Average Pooling (GAP) is performed on the spatial dimension to obtain channel-wise spectral response description vectors as follows:

$$z = GAP(F), z \in \mathbb{R}^{C} \tag{2}$$

Then, an efficient channel attention mechanism is used to model inter-channel dependencies. This process can be expressed as:

$$w = \sigma(Conv1D(z)) \tag{3}$$

Where $w \in \mathbb{R}^{C}$ represents the confidence estimate of channel-wise spectral responses; $Conv1D(\cdot)$ denotes a one-dimensional convolution operator used to capture local correlations between adjacent spectral channels; $\sigma(\cdot)$ is the Sigmoid activation function.

Based on this, channel recalibration is performed on the original features as follows:

$$F' = w \odot F \tag{4}$$

Where $\odot$ denotes element-wise multiplication.

On the basis of channel enhancement, the spatial heterogeneity of waterlogged regions and the gradual characteristics of their boundary transition zones are taken into account. Specifically, SSIEM incorporates an adaptive response enhancement mechanism in the spatial dimension to generate a spatial weight map by aggregating channel-wise information and applying spatial convolution as follows:

$$M_s = \sigma(f^{3\times3}(Concat(AvgPool(F'), MaxPool(F')))) \tag{5}$$

Where $f^{3\times3}(\cdot)$ denotes a 3×3 convolution operator, and $M_s \in \mathbb{R}^{H\times W}$.

High-weight regions generally correspond to core waterlogged regions with distinct discrimination, while low-weight regions are mostly distributed in transition zones between waterlogged regions and the background. Finally, the module output can be expressed as:

$$F_{out} = M_s \odot F' \tag{6}$$

 doi:10.20944/preprints202604.1571.v1

Through the above joint channel-spatial recalibration process, SSIEM can highlight the spectral-spatial responses related to waterlogging identification from high-level features, thus improving feature discriminability in complex farmland scenarios. This provides more stable semantic representations for subsequent multi-scale contextual modeling.

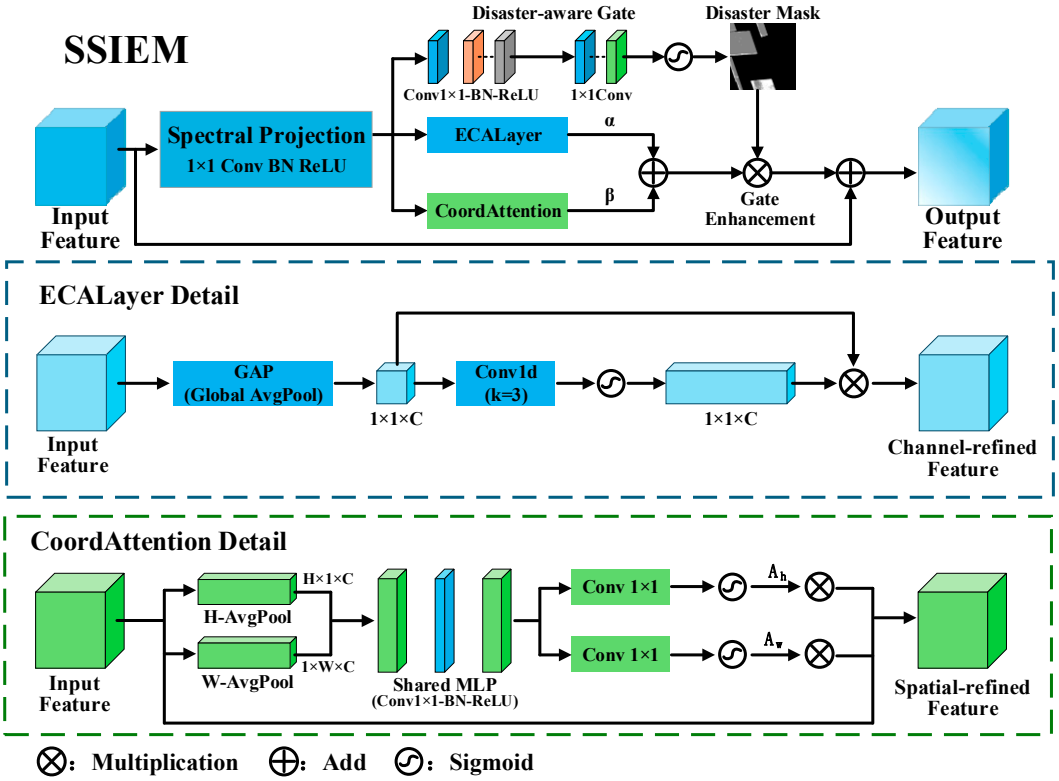

Figure 6. Network structure of SSIEM.

2.3.4. AMSP Module

In a standard DeepLabV3+ network, the ASPP module extracts multi-scale contextual information through the Atrous Convolution branches with different preset dilation rates, based on the implicit assumption that the target scale distribution is relatively stable. However, in agricultural disaster remote sensing scenarios, disaster-affected regions often exhibit highly uneven scales and fragmented spatial morphologies, such that fixed dilation rates cannot effectively respond to disaster patches of different scales.

To address this problem, we proposed the AMSP module, which dynamically adjusts the contribution of different receptive field branches by introducing a scale weight learning mechanism. Its structural diagram is shown in Figure 7. Let the output features of SSIEM be F''. AMSP performs multi-scale feature extraction through K parallel branches as follows:

$$F_k = f_k(F''), k = 1, \ldots, K \tag{7}$$

Where $f_k(\cdot)$ denotes convolution operations under different dilation rates or receptive field configurations.

To avoid scale information redundancy caused by simple concatenation or averaging, a global feature-guided weight generation mechanism was introduced. First, global pooling is performed on features of each scale and the results are fused to obtain a scale description vector as follows:

$$s = \sum_{k=1}^{K} GAP(F_k) \tag{8}$$

Then, scale weights are generated through linear channel mapping (1×1 convolution) and an activation function:

 doi:10.20944/preprints202604.1571.v1

$$\alpha = Softmax(f^{1\times1}(s)) \quad (9)$$

Where $\alpha = [\alpha_1, \ldots, \alpha_K]$ and $\sum_k \alpha_k = 1$.

Finally, multi-scale features are fused according to their adaptive weights:

$$F_{AMSP} = \sum_{k=1}^{K} \alpha_k \cdot F_k. \quad (10)$$

Unlike channel or kernel selection mechanisms, the weight learning of AMSP acts directly on the scale-level context branch level, enabling the model to autonomously emphasize more discriminative receptive field branches according to the scale characteristics of disaster patches in the input image, thus improving the flexibility and robustness of multi-scale context modeling.

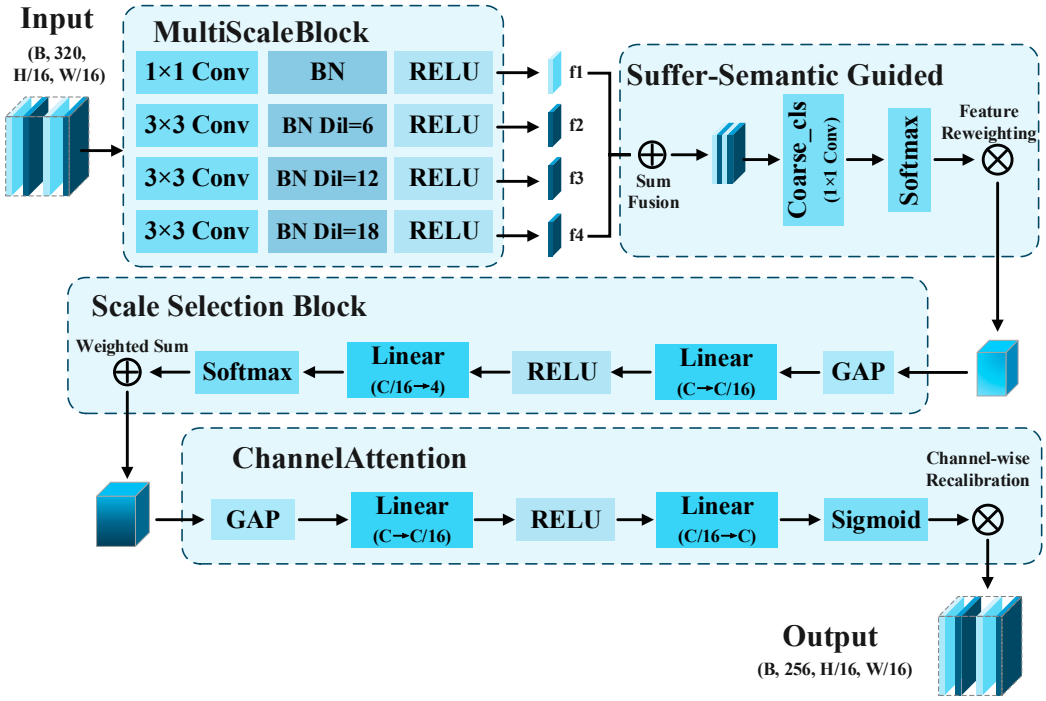

Figure 7. Network structure of AMSP.

2.3.5. BEAM Module

In single-date maize waterlogging identification tasks, wide transition zones often exist between waterlogged and non-waterlogged regions. Especially in regions with mild waterlogging, fragmented fields, or complex background textures, boundaries usually exhibit gradual rather than sharp transition characteristics. Therefore, relying solely on high-level semantic features easily leads to local over-smoothing and a certain degree of misclassification in boundary transition zones. To address this issue, we introduced BEAM after the fusion of low-level and high-level features, which integrates local edge cues and high-level semantic information to improve the representation quality of boundary transition zones and local classification consistency. The overall structure of BEAM is shown in Figure 8.

 doi:10.20944/preprints202604.1571.v1

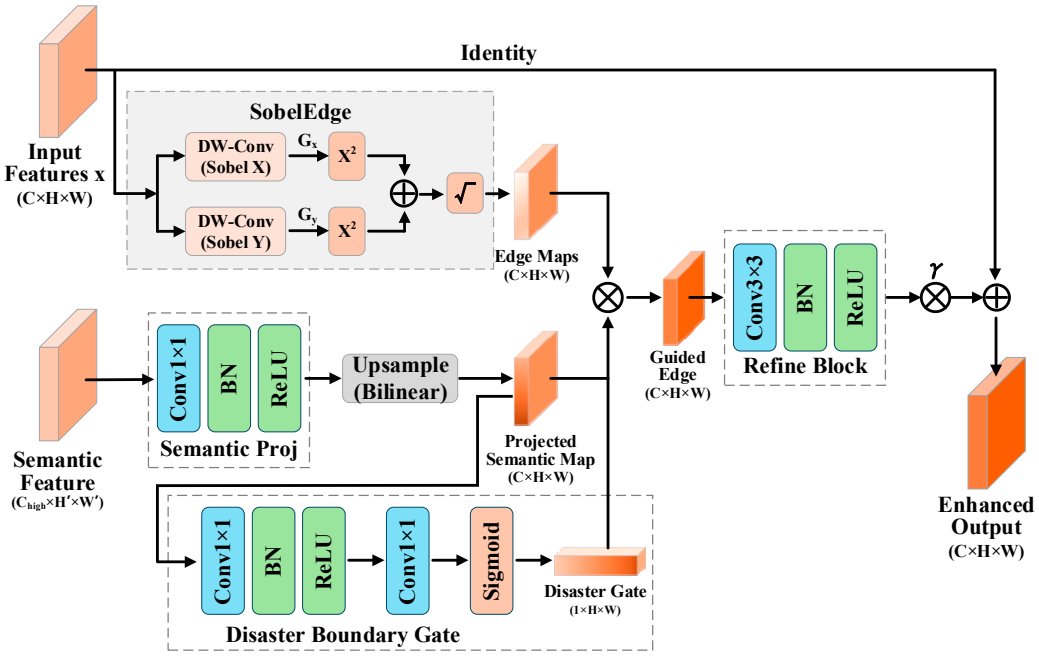

Figure 8. Network structure of BEAM.

Let the fused feature in the decoding stage be denoted as F_d. First, edge responses are extracted through a gradient operator as follows:

$$E = g(F_d) \tag{11}$$

Where g(·) denotes a differentiable Sobel operator used to characterize potential boundary zones. This operator can participate in end-to-end training under the backpropagation framework.

Subsequently, a semantically guided attention mechanism is introduced to map high-level semantic information into boundary weights:

$$A_b = \sigma(f^{1\times1}(F_d)) \tag{12}$$

Where $f^{1\times1}(\cdot)$ denotes a channel mapping operator used to generate attention weights related to boundaries.

Finally, the boundary-enhanced feature is expressed as:

$$F_{BEAM} = F_d + A_b \odot E \tag{13}$$

Through the above design, BEAM can maintain the dominance of high-level semantic information while appropriately introducing local structural constraints. This is conducive to refining the representation of boundary transition zones and improving the local integrity of waterlogged patches.

3. Results and Analysis

3.1. Experimental Environment and Configuration

All experiments were performed in a 64-bit Linux system (Kernel 5.15.0) with Python 3.8.20 installed. The deep learning framework employed was PyTorch 2.4.1 (built on CUDA 12.1), combined with cuDNN 9.1.0 for accelerated computing and operator optimization. The hardware platform was equipped with 4 NVIDIA GeForce RTX 3090 graphics processing units (each with 24GB of video memory and a computing capability of 8.6), which facilitate multi-GPU parallel training to meet the video memory capacity and computational throughput requirements of semantic segmentation tasks for high-resolution remote sensing images.

The Adam optimizer was uniformly adopted during model training, with an initial learning rate of 1×10^{-4}, a batch size of 16, and total training epochs of 100. The loss function adopted a combination

 doi:10.20944/preprints202604.1571.v1

of cross-entropy loss and Dice loss to balance pixel-level classification accuracy and regional overlap consistency. To ensure experimental reproducibility, the random seed was uniformly set to 11, and all comparative experiments and ablation experiments were conducted under the same software and hardware environment and training configuration.

3.2. Evaluation Indicators

To comprehensively evaluate the segmentation performance of the proposed model in the task of maize waterlogging identification, we adopted Intersection over Union (IoU), mean Intersection over Union (mIoU), Overall Accuracy (OA), mean Pixel Accuracy (mPA), and mean F1-score (mF1) as evaluation indicators. Considering that waterlogged regions usually account for a small proportion in pixel distribution, mIoU, mPA, and mF1 were used simultaneously to alleviate the influence of class imbalance on evaluation results.

Single-class IoU is defined as the degree of overlap between the prediction result and the ground truth, with the calculation formula as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{14}$$

Where TP, FP, and FN denote the correctly classified target pixels, background pixels misclassified as targets, and missed target pixels, respectively. The value of mIoU is calculated as the average IoU over all classes.

OA was used to measure the overall classification correctness across the entire image, defined as:

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \tag{15}$$

Where TN denotes the number of pixels correctly classified as background.

To further measure the recognition balance among classes, we adopted mPA, defined as the average of single-class pixel accuracies:

$$mPA = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FN_i} \tag{16}$$

Where K denotes the number of classes.

F1-score comprehensively takes into account the precision and recall, with the single-class F1 defined as:

$$F1 = \frac{2TP}{2TP + FP + FN} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

Subsequently, we further calculated the average F1-score over all classes as mF1 to characterize the model's comprehensive recognition ability.

In summary, this study comprehensively evaluated model performance from the perspectives of spatial overlap, overall classification accuracy, and class balance, providing a basis for performance comparison among different methods in maize waterlogging segmentation tasks.

3.3. Spectral Separability Analysis

To evaluate the spectral separability between waterlogged and non-waterlogging-affected maize in single-date multispectral images, we randomly selected 10,000 pixels from each of the two classes and quantitatively analyzed the single-band feature, four-band combined feature, and typical vegetation indices using the Jeffries-Matusita (JM) distance. The JM distance ranges from 0 to 2, with larger values indicating stronger inter-class separability.

The results show that the JM distance of the four-band combined feature (RGB-NIR) was 1.3291, indicating that under single-date conditions, multispectral images can provide a favorable discriminative basis for waterlogging identification, although the two classes still exhibit certain overlap in the original spectral space. Single-band results (Figure 9) show that the Red band had the highest separability (0.6589), followed by NIR (0.5494) and Green (0.5288), with Blue being the lowest

 doi:10.20944/preprints202604.1571.v1

(0.3741). This suggests that within the study area and observation period, the Red and NIR bands are more sensitive to maize waterlogging responses, while the Blue band has relatively limited individual discriminative ability.

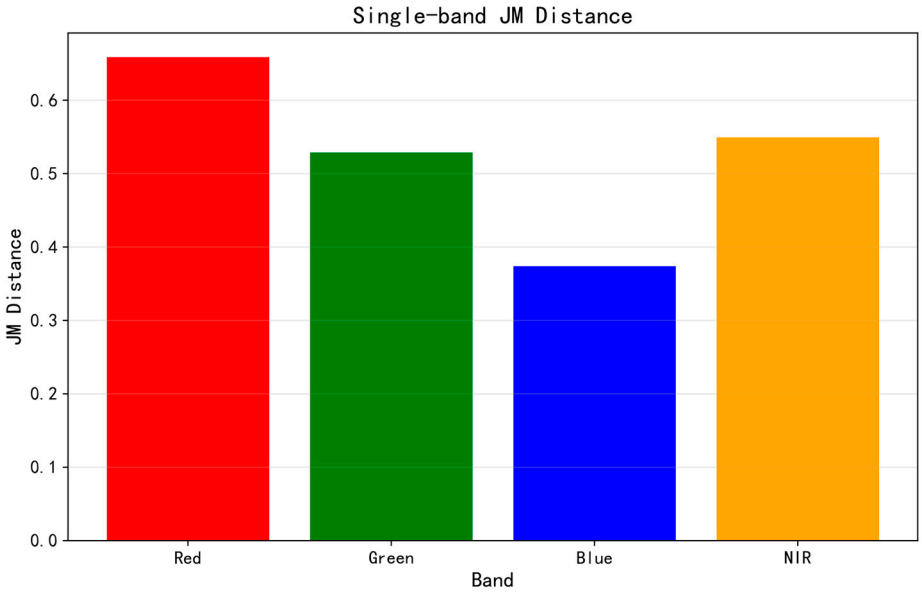

Figure 9. Single-band JM distances between non-waterlogged and waterlogging-affected maize samples.

Analysis of typical vegetation indices shows that the JM distances of NDVI and GNDVI were 0.7229 and 0.8845, respectively, both superior to those of single-band features. This indicates that the band combination strategy can enhance the differential expression between waterlogged and non-waterlogging-affected maize samples, with GNDVI performing better than NDVI. However, the separability of both indices was lower than that of the four-band combined feature, suggesting that a single index cannot fully represent waterlogging information in complex farmland scenarios.

Overall, single-date RGB-NIR images exhibit the characteristic of "separable but not fully separated" for maize waterlogging identification. This demonstrates that single-date multispectral data hold potential for rapid waterlogging identification, yet further integration of spatial context and deep feature modeling is still required to improve segmentation accuracy in complex scenes.

3.4. Comparative Experiments

3.4.1. Comparison with Traditional Single-Date Methods

To further verify the feasibility of applying single-date multispectral imagery for waterlogging identification within maize planting regions, we conducted a comparative experiment with three types of traditional single-date baseline methods: the NDVI threshold method, GNDVI threshold method, and Random Forest (RF) classification method. Among them, the NDVI and GNDVI thresholds were automatically determined on the training set, with the optimal threshold selected based on the highest F1-score for the waterlogged maize class. The RF classifier used a total of 22-dimensional input features, including original bands, vegetation index/ratio features, and local spatial statistical features. Class-balanced sampling was employed during training, and the number of trees was set to 500. Meanwhile, class_weight = balanced_subsample was adopted to alleviate the impact of class imbalance. To ensure consistency for comparison with deep learning models, the threshold optimization and classifier training for traditional methods were only performed on the training set, with the final results uniformly reported on the test set. Relevant results are presented in Table 1.

 doi:10.20944/preprints202604.1571.v1

Table 1. Comparison results of traditional single-date methods for maize waterlogging identification.

Method	IoU-WM/%	Recall-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
NDVI	16.94	53.85	31.81	52.93	46.31	51.92
GNDVI	18.51	57.44	33.55	55.34	48.32	53.96
Random Forest	24.96	34.38	52.45	57.88	64.40	81.18

Note: IoU-WM denotes the IoU of the waterlogged maize class, and Recall-WM denotes the recall of the waterlogged maize class.

3.4.2. Comparison with Deep Learning Segmentation Models

To systematically evaluate the performance of the proposed SAB-DeepLabV3+ model in remote sensing identification of maize waterlogging stress, we selected several representative semantic segmentation models for comparative experiments, including encoder-decoder CNNs, lightweight real-time models, multi-scale context modeling methods, high-resolution or attention-enhanced models, and Transformer-based semantic segmentation models. To ensure a fair comparison among different models on the same agricultural remote sensing dataset, all deep learning models were trained from scratch with unified data augmentation strategies, input size, optimizer settings, and training epochs, without using extra pre-trained weights. Quantitative evaluation results of each model are summarized in Table 2.

Table 2. Quantitative comparison among different semantic segmentation models on maize waterlogging stress dataset.

Model Category	Model	IoU-NM/%	IoU-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
Encoder-Decoder (CNN)	SegNet[25]	87.25	45.78	66.52	74.84	78.00	88.49
	UNet[26]	90.77	62.14	76.46	84.32	85.91	91.98
	UNet++[27]	89.79	56.19	72.99	80.33	83.28	90.97
	DoubleUNet[28]	90.25	58.73	74.49	81.91	84.44	91.44
Lightweight / Real-time	BiSeNetV2[29]	88.22	48.60	68.41	76.04	79.58	89.40
Multi-scale Context	PSPNet[30]	91.11	63.65	77.38	85.18	86.57	92.31
	FPN[31]	90.94	63.63	77.29	85.55	86.51	92.18
	DeepLabV3+[32]	91.24	62.23	76.74	84.05	86.07	92.35
High-resolution /Attention	HRNet[33]	90.00	57.88	73.94	81.54	84.03	91.21
	DCSA-UNet[34]	90.52	58.71	74.61	81.39	84.50	91.64
Transformer-Based	SegFormer[35]	88.23	50.27	69.25	77.39	80.33	89.49
	TransUNet[36]	89.91	55.65	72.78	79.65	83.10	91.04
	SwinUNet[37]	90.93	62.06	76.50	83.87	85.92	92.11
Proposed	SAB-DeepLabV3+	92.43	68.30	80.37	87.87	88.62	93.49

Note: IoU-NM and IoU-WM denote the IoU of non-waterlogged maize and waterlogged maize classes, respectively.

 doi:10.20944/preprints202604.1571.v1

As can be seen from Table 2, the IoU differences among different models for the non-waterlogged maize class were relatively small, with most methods exceeding 88%, indicating that this class has generally good separability in single-date multispectral imagery. In contrast, the IoU of the waterlogged maize class fluctuated much more significantly, suggesting that waterlogged regions are more complex in spatial morphology and spectral response, making them the key for distinguishing model performance.

From the perspective of model architecture, traditional encoder-decoder structures (e.g., UNet, DoubleUNet) performed well in preserving local details well but had limited ability to model large-scale continuous waterlogged regions holistically. Although the lightweight model BiSeNetV2 achieved high efficiency, it was weak in depicting fine-grained waterlogged regions in complex farmland scenes. In contrast, models with multi-scale context modeling capabilities (e.g., PSPNet, FPN, and DeepLabV3+) performed more stably overall, indicating that multi-scale semantic fusion plays an important role in waterlogging identification. High-resolution preservation or attention-enhanced models (e.g., HRNet, DCSA-UNet) showed certain advantages in spatial detail representation, but their overall improvement on our dataset was limited. Transformer-based models (e.g., SegFormer, TransUNet, and SwinUNet) showed no obvious advantages without pre-trained weights, implying a strong dependence on prior knowledge when dealing with moderate-scale agricultural remote sensing samples.

To further analyze the prediction variations among different models in complex scenes, Figure 10 shows the qualitative segmentation results of three representative test regions, including (a) mildly waterlogged regions, (b) severely waterlogged regions, and (c) fragmented/complex boundary regions. It can be observed from Figure 10 that in (a) and (c), some models exhibited boundary shrinkage or patch fragmentation in waterlogging-affected regions; in (b), some methods suffered from over-smoothing and misclassified local waterlogging-affected regions as the non-waterlogged class. In contrast, SAB-DeepLabV3+ performed more stably in terms of boundary integrity and patch continuity, significantly reducing the phenomena of omission and misclassification in confusing regions.

In summary, SAB-DeepLabV3+ achieved the best results in IoU-WM, mIoU, and mF1 for the waterlogging-affected maize class, further supporting that incorporating spectral-spatial enhancement and boundary enhancement mechanisms on the basis of the multi-scale context framework of DeepLabV3+ can effectively alleviate class confusion and boundary blurring in waterlogging identification, thereby improving the model's stability and robustness in complex farmland scenes.

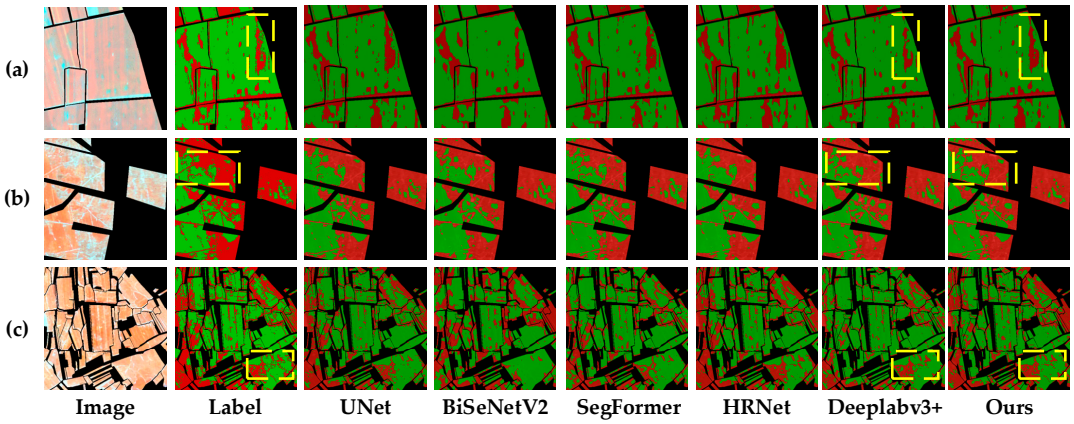

Figure 10. Qualitative segmentation results in representative waterlogging-affected regions.

3.5. Ablation Experiments

To systematically evaluate the independent contribution and synergistic mechanism of each functional module in the proposed SAB-DeepLabV3+ framework, progressive ablation experiments

 doi:10.20944/preprints202604.1571.v1

were conducted on the basis of the DeepLabV3+ baseline model. Specifically, the single-module, dual-module, and complete model configurations were implemented stepwise by sequentially incorporating SSIEM, AMSP, and BEAM. The quantitative results are summarized in Table 3.

Table 3. Quantitative results of ablation experiments.

Exp	SSIEM	AMSP	BEAM	IoU-NM/%	IoU-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
0	×	×	×	91.24	62.23	76.74	84.05	86.07	92.35
1	√	×	×	91.22	65.14	78.18	87.40	87.15	92.46
2	×	√	×	91.21	64.96	78.08	87.21	87.08	92.44
3	×	×	√	91.50	63.99	77.74	85.36	86.80	92.61
4	√	√	×	92.04	67.42	79.73	87.93	88.20	93.17
5	√	√	√	92.43	68.30	80.37	87.87	88.62	93.49

Note: √ indicates the use of the module, and × indicates the non-use of the module. IoU-NM and IoU-WM represent the IoU of the non-waterlogged maize class and waterlogging-affected maize class, respectively.

As shown in Table 3, the baseline model already achieved a high IoU (91.24%) for the non-waterlogged maize class, but only 62.23% for the waterlogging-affected maize class, indicating that waterlogging identification in complex farmland scenes remains the main bottleneck limiting overall performance.

Under single-module settings, both SSIEM and AMSP delivered significant improvements. Specifically, after incorporating SSIEM, the IoU and mIoU of the waterlogging-affected maize class increased to 65.14% and 78.18%, respectively; after incorporating AMSP, the IoU and mIoU of the waterlogging-affected maize class reached 64.96% and 78.08%, respectively. This demonstrates that both spectral-spatial enhancement and multi-scale feature modeling strategies are conducive to improving the discriminability and spatial representation of waterlogging-affected regions. In contrast, BEAM produced a more noticeable gain in overall OA, while its direct contribution to the IoU of the waterlogging-affected class was limited, suggesting that it functions more in refining boundary transition zones and enhancing local spatial consistency rather than simply strengthening the discriminative power for waterlogging-affected regions.

When SSIEM and AMSP were jointly incorporated, model performance was further improved, with the IoU and mIoU of the waterlogging-affected class increasing to 67.42% and 79.73%, respectively, indicating their good complementarity in class discrimination and spatial structure modeling. Upon further incorporation of BEAM, the complete model achieved the best results: for the waterlogging-affected class, IoU increased to 68.30% (6.07 percentage points higher than the baseline), while mIoU, mF1, and OA reached 80.37%, 88.62%, and 93.49%, respectively. This reveals that the three modules can achieve synergistic improvement in spectral discrimination enhancement, multi-scale structure reconstruction, and boundary regularization.

Figure 11 shows three types of representative scenes with high discrimination difficulty: (a) fragmented patches, (b) fuzzy boundaries, and (c) spectral aliasing regions. For each example, the original image, ground truth label, baseline model prediction result, dual-module (SSIEM+AMSP) prediction result, and tri-module prediction result are provided. The performance evolution process can be clearly observed from the visual results. In fragmented patches, the baseline model was prone to fragmentation and missed detection, the dual-module model significantly restored spatial continuity, while the final model further improved the intra-patch consistency. After incorporating BEAM, the misclassification phenomenon in regions with blurred boundaries was alleviated, and the prediction results showed a more complete local transition expression visually.

Overall, the visual comparison results are highly consistent with the quantitative trends in Table 3, demonstrating a clear performance progression path: from single-module enhancement to dual-module collaboration, and then to the complete design, the model's recognition integrity for disaster-affected regions, ability to express boundary regions, and class consistency were continuously improved. This synergistic mechanism (i.e., spectral discrimination enhancement – multi-scale

 doi:10.20944/preprints202604.1571.v1

structural reconstruction – boundary consistency constraint) offers structural support on the high-precision disaster recognition ability of SAB-DeepLabV3+ in complex agricultural remote sensing scenarios and provides an interpretable performance improvement path for subsequent modular design.

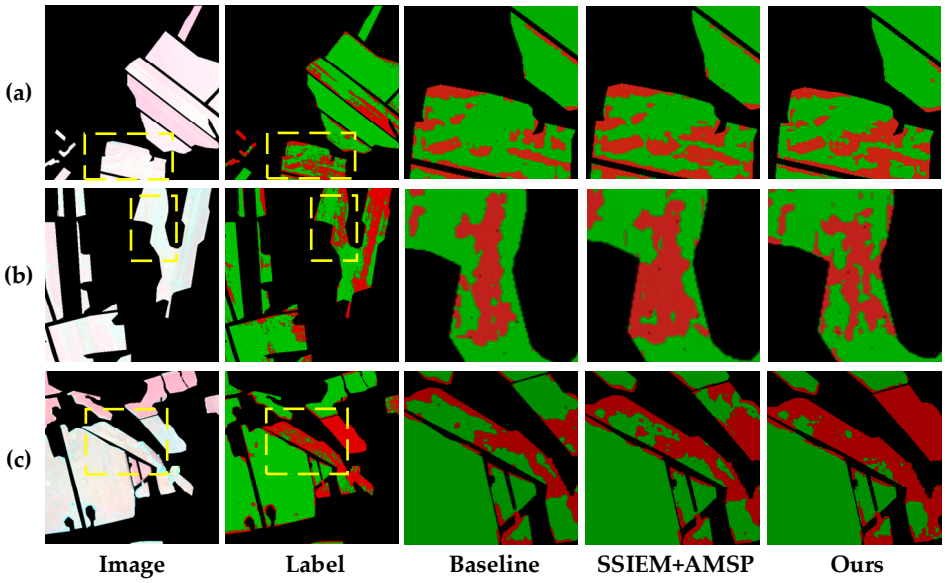

Figure 11. Visual comparison of ablation results in representative waterlogging-affected maize regions.

3.6. Comparative Experiments with Representative Modules

To further verify the effectiveness of SSIEM, AMSP, and BEAM against existing representative structures, we adopted a univariate replacement strategy to perform comparative experiments on the high-level feature enhancement module, multi-scale context module, and boundary enhancement module separately under consistent backbone network, decoder architecture, training settings, and data partitioning.

3.6.1. Comparison Between SSIEM and Generic Attention Modules

To verify the effectiveness of SSIEM in single-date maize waterlogging identification, we separately introduced SE, ECA, CBAM, and SSIEM at the high-level feature output of the encoder and compared with the baseline DeepLabV3+ without incorporating an attention module. All experiments retained the original ASPP context module and decoder structure, with only the high-level feature enhancement method replaced to ensure fair comparison.

As shown in Table 4, all generic attention modules could improve waterlogging identification to a certain extent, confirming the effectiveness of adaptive enhancement on high-level semantic features. Compared with the baseline model, CBAM, SE, and ECA increased the IoU for the waterlogging-affected class to 64.37%, 64.51%, and 63.93%, respectively; SSIEM achieved the best performance, with the IoU, mIoU, and mF1 for the waterlogging-affected class reaching 65.14%, 78.18%, and 87.15%, respectively. In terms of class statistics, the advantage of SSIEM was mainly reflected in the improved Recall of the waterlogging-affected class, indicating that this module can more effectively enhance the discriminability of weakly responsive waterlogging-affected regions and reduce missed detections. Overall, SSIEM is more suitable for modeling subtle differences between waterlogged and non-waterlogging-affected maize samples under single-date multispectral conditions compared with generic attention modules.

 doi:10.20944/preprints202604.1571.v1

Table 4. Comparison of SSIEM with representative attention modules.

Attention module	IoU-NM/%	IoU-WM/%	Recall-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
None	91.24	62.23	71.21	76.74	84.05	86.07	92.35
CBAM	90.95	64.37	77.18	77.66	86.37	86.79	92.22
SE	91.12	64.51	76.32	77.82	86.12	86.89	92.36
ECA	91.08	63.93	74.98	77.50	85.56	86.66	92.30
SSIEM	91.22	65.14	79.57	78.18	87.40	87.15	92.46

Note: IoU-NM and IoU-WM represent the IoU of non-waterlogged maize and waterlogging-affected maize classes, respectively; Recall-WM denotes the recall of the waterlogging-affected maize class.

3.6.2. Comparison Between AMSP and Multi-Scale Context Modules

To verify the effectiveness of the adaptive multi-scale modeling strategy in AMSP, we further compared among the original ASPP, the static multi-scale structure without dynamic scale selection (Static-MSP), and the full AMSP. All three modules operated on high-level encoder features, with other network architectures and training configurations unchanged. Static-MSP retained the same multi-scale branch settings as AMSP but adopted fixed concatenation fusion instead of dynamic scale weight learning. This was designed to distinguish whether performance gains originate from the explicit multi-scale structure itself or the adaptive scale selection mechanism.

As shown in Table 5, Static-MSP already outperformed the original ASPP, raising the IoU of the waterlogging-affected class from 62.23% to 64.17%, indicating that explicit multi-scale branches are conducive to enhancing representation for waterlogged patches of different sizes. On this basis, AMSP further achieved the best performance, with the IoU, mIoU, and mF1 of the waterlogging-affected class reaching 64.96%, 78.08%, and 87.08%, respectively. Compared with Static-MSP, AMSP yielded additional gains in Recall-WM and overall mIoU, demonstrating that its improvement stems not only from the multi-scale structure but also from adaptive selection of scale responses. In general, AMSP can better adapt to variations in the area, shape, and spatial organization of waterlogged patches, thus delivering stronger context modeling capability.

Table 5. Comparison of AMSP with different multi-scale context modules.

Context module	IoU-NM/%	IoU-WM/%	Recall-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
ASPP	91.24	62.23	71.21	76.74	84.05	86.07	92.35
Static-MSP	90.97	64.17	76.42	77.57	86.08	86.72	92.23
AMSP	91.21	64.96	79.10	78.08	87.21	87.08	92.44

Note: IoU-NM and IoU-WM represent the IoU of non-waterlogged maize and waterlogging-affected maize classes, respectively; Recall-WM denotes the recall of the waterlogging-affected maize class.

3.6.3. Comparison Between BEAM and Boundary Enhancement Variants

To verify the roles of each component in BEAM, we constructed two simplified variants for comparative analysis while keeping the full BEAM structure unchanged, namely Edge-only and Semantic-no-gate. The former only uses Sobel edge extraction and residual enhancement, whereas the latter further incorporates semantic guidance without the disaster boundary gate. The complete BEAM integrates three components simultaneously: edge extraction, semantic guidance, and disaster boundary gating. All boundary enhancement modules were inserted after the fusion of low-level and high-level features, with other network architectures and training configurations remaining consistent.

As shown in Table 6, all boundary enhancement strategies consistently improved segmentation performance for the waterlogged maize class relative to the no-boundary-module baseline. Specifically, Edge-only and Semantic-no-gate achieved waterlogged-class IoU values of 63.32% and

 doi:10.20944/preprints202604.1571.v1

63.51%, respectively, both exceeding the baseline by a clear margin. These results indicate that edge cues alone can partially sharpen boundary delineation, while the addition of semantic guidance provides complementary contextual information that further refines class transition regions. The full BEAM configuration, which integrates edge enhancement, semantic context, and a boundary gating mechanism, achieved the best overall performance, with a waterlogged-class IoU of 63.99%, an mIoU of 77.74%, and an OA of 92.61%. Notably, the performance gain of the full BEAM over Edge-only and Semantic-no-gate is modest in absolute terms but consistent across all metrics, suggesting that the boundary gating mechanism plays a complementary rather than dominant role. These findings collectively indicate that the performance improvement brought by BEAM arises from the synergistic interaction among edge information, semantic context, and adaptive boundary gating, rather than from any single component in isolation.

Table 6. Comparison between BEAM and different boundary enhancement variants.

Boundary strategy	IoU-NM/%	IoU-WM/%	Recall-WM/%	mIoU/%	mPA/%	mF1/%	OA/%
None	91.24	62.23	71.21	76.74	84.05	86.07	92.35
Edge-only	90.81	63.32	75.17	77.07	85.50	86.36	92.07
Semantic-no-gate	91.06	63.51	73.92	77.29	85.14	86.5	92.28
BEAM	91.50	63.99	74.14	77.74	85.36	86.8	92.61

Note: IoU-NM and IoU-WM represent the IoU of non-waterlogged maize and waterlogging-affected maize classes, respectively; Recall-WM denotes the recall of the waterlogging-affected maize class.

Overall, all three groups of comparative experiments demonstrate that SSIEM, AMSP, and BEAM have better task adaptability than their representative alternative structures. Collectively, they work synergically to improve the performance of single-date maize waterlogging identification from three aspects: high-level discriminative enhancement (SSIEM), multi-scale context modeling (AMSP), and boundary refinement (BEAM).

3.7. Generalization Experiment

To evaluate the model's transferability in unseen regions, we conducted cross-regional generalization experiments by employing a leave-one-city-out strategy. In each trial, one city was selected as the independent test area, while the remaining four cities were used for training and validation, with no retraining or fine-tuning during the testing phase. This setup can realistically simulate the cross-regional deployment of the model in practical operational scenarios. To further prove that the performance improvement stems from the three proposed modified modules in this study rather than data partitioning differences, cross-regional generalization tests were also performed on the baseline DeepLabV3+ under identical data partitioning, training strategy, and testing procedures.

Table 7 presents the results of SAB-DeepLabV3+ across the five independent cities. Overall, the model performed stably across all cities, with mIoU ranging from 75.96% to 77.15% (mean = 76.56%) and OA ranging from 88.97% to 93.62% (mean = 91.38%). The IoU of the waterlogging-affected maize class ranged from 60.53% to 68.89% (mean = 63.45%), indicating that the model maintains favorable waterlogging identification ability under cross-regional conditions. In contrast, the IoU of the non-waterlogged maize class exceeded 85% in all cities, suggesting that the main challenge of cross-regional generalization lies in the waterlogging-affected class, especially in regions with blurred boundaries, fragmented patches, or mild waterlogging.

 doi:10.20944/preprints202604.1571.v1

Table 7. Cross-regional generalization performance of SAB-DeepLabV3+ under the leave-one-city-out setting.

City	IoU-NM/%	IoU-WM/%	Recall-WM/%	F1-WM/%	mIoU/%	OA/%
Jixi	92.93	60.53	71.59	75.41	76.73	93.62
Daqing	85.41	68.89	80.52	81.58	77.15	88.97
Hegang	90.52	63.00	72.68	77.30	76.76	91.84
Mudanjiang	89.83	62.09	72.54	76.61	75.96	91.28
Qiqihar	89.62	62.74	68.96	77.10	76.18	91.17
Average	89.66	63.45	73.26	77.60	76.56	91.38

Note: IoU-NM and IoU-WM represent the IoU of non-waterlogged maize and waterlogging-affected maize classes, respectively; Recall-WM denotes the recall of the waterlogging-affected maize class; F1-WM denotes the F1-score of the waterlogging-affected maize class.

Among different cities, Daqing achieved the best performance, with the IoU of the waterlogging-affected class reaching 68.89%. Qiqihar exhibited high Precision (87.43%) but relatively low Recall (68.96%). These results indicate that the model controlled false positives well in unseen regions but still showed certain omissions in boundary transition zones or mildly waterlogged regions. Overall, SAB-DeepLabV3+ maintained stable segmentation performance across all cities, demonstrating strong cross-regional generalization ability and practical deployment potential.

Furthermore, Table 8 presents the comparison results between DeepLabV3+ and SAB-DeepLabV3+ under the cross-regional setup. As can be seen, the proposed method outperformed the baseline in all five independent cities. On average, SAB-DeepLabV3+ increased mIoU from 74.43% to 76.56% (a gain of 2.13 percentage points) and OA from 90.71% to 91.38% (a gain of 0.67 percentage points) compared to DeepLabV3+. More importantly, the improvement gained was concentrated on the waterlogged maize class (the more challenging class), with IoU-WM increasing from 59.82% to 63.45% (a gain of 3.63 percentage points), Recall-WM from 67.78% to 73.26% (a gain of 5.48 percentage points), and F1-WM from 74.77% to 77.60% (a gain of 2.83 percentage points). In comparison, the IoU of the non-waterlogged maize class improved by only 0.62 percentage points, demonstrating that the spectral-spatial enhancement, adaptive multi-scale fusion, and boundary enhancement modules can effectively strengthen the model's discriminability for cross-regional waterlogging patterns rather than merely benefiting easily classified samples.

Table 8. Cross-regional comparison between DeepLabV3+ and SAB-DeepLabV3+ under the leave-one-city-out setting.

City	IoU-WM (DL)/%	IoU-WM (SAB)/%	ΔIoU-WM/%	mIoU (DL)/%	mIoU (SAB)/%	ΔmIoU/%	OA (DL)/%	OA (SAB)/%	ΔOA/%
Jixi	57.69	60.53	+2.84	75.18	76.73	+1.55	93.34	93.62	+0.28
Daqing	66.89	68.89	+2.00	75.95	77.15	+1.20	88.49	88.97	+0.48
Hegang	59.79	63.00	+3.21	74.92	76.76	+1.84	91.33	91.84	+0.51
Mudanjiang	60.92	62.09	+1.17	75.28	75.96	+0.68	91.07	91.28	+0.21
Qiqihar	53.81	62.74	+8.93	70.82	76.18	+5.36	89.34	91.17	+1.83
Average	59.82	63.45	+3.63	74.43	76.56	+2.13	90.71	91.38	+0.67

Note: DL denotes DeepLabV3+; SAB denotes SAB-DeepLabV3+; WM-IoU represents the IoU of the waterlogged maize class; Δ represents the improvement gain of SAB over DL.

At the city level, the most significant improvement was observed in Qiqihar, where IoU-WM increased from 53.81% to 62.74% (a gain of 8.93 percentage points) and mIoU increased by 5.36 percentage points. Combined with regional test results, the baseline model tends to be conservative in identifying waterlogged regions in this city, achieving high Precision but low Recall. In contrast, SAB-DeepLabV3+ not only maintained high Precision but also significantly reduced the under-segmentation of waterlogging-affected regions, showing stronger cross-regional adaptability. In

 doi:10.20944/preprints202604.1571.v1

other cities, including Daqing, Hegang, Mudanjiang, and Jixi, the proposed method also achieved consistent improvement gains, indicating that the model was effective not only in individual regions but also robust under diverse geographical environments and disaster backgrounds.

In summary, SAB-DeepLabV3+ not only achieved superior results in within-region testing but also consistently outperformed the baseline in rigorous cross-city generalization experiments, especially showing stronger transferability in waterlogging-affected maize class identification. This demonstrates that the three proposed modules can effectively alleviate spectral aliasing, scale heterogeneity, and boundary blurring under single-date multispectral conditions, supporting cross-regional deployment of the model in practical agricultural disaster monitoring.

4. Discussion

4.1. Feasibility of Single-Date Multispectral Imagery for Maize Waterlogging Identification

A core question this study aimed to address is whether single-date, high-resolution multispectral imagery can support maize waterlogging identification in the absence of time-series information. Spectral separability analysis shows that the JM distance of the four-band combined feature was 1.3291, higher than that of single bands and typical vegetation indices, indicating that single-date RGB-NIR imagery provides a reasonable discriminative basis for waterlogging identification. However, this value did not reach full separability, implying significant overlap between non-waterlogged and waterlogging-affected maize samples in the original spectral space, making stable identification difficult using only shallow spectral features.

In terms of band and index performance, Red and NIR exhibited higher separability than Green and Blue, and GNDVI outperformed NDVI. These results were consistent with remote sensing responses to changes in canopy structure, leaf pigment status, and moist backgrounds after waterlogging. Comparisons with traditional single-date methods further confirm that single-date imagery contains usable waterlogging information, yet single-index thresholding and shallow machine learning methods cannot meet the demands of high-precision mapping in complex farmland scenes.

From an application perspective, the single-date scheme offers advantages of timeliness and simplicity, making it suitable for rapid post-disaster verification and regional mapping. Overall, for major maize-producing areas in Northeast China, single-date, high-resolution multispectral imagery acquired in the middle and late growing seasons shows strong application potential.

4.2. Performance Sources and Module Synergy Mechanism of SAB-DeepLabV3+

Experimental results show that the improvement gains of SAB-DeepLabV3+ mainly stem from targeted modeling of three key issues in single-date scenes: spectral aliasing, scale heterogeneity, and boundary blurring. Compared with the baseline DeepLabV3+, the proposed model increased the IoU and mIoU of the waterlogged class from 62.23% to 68.30% and 76.74% to 80.37%, respectively, indicating that the waterlogged class, which is of higher difficulty for identification, achieved more significant improvements.

Regarding module functions, SSIEM and AMSP provided more direct gains for the waterlogged class, while BEAM focused on refining boundary transition zones and improving local spatial consistency. The superiority of SSIEM over SE, ECA, and CBAM demonstrates that spectral-spatial collaborative enhancement designed for single-date, multispectral scenes is more suitable for waterlogging identification than generic attention mechanisms. On the other hand, the advantage of AMSP over ASPP and Static-MSP reveals that dynamic scale selection helps the model better adapt to variations in area, shape, and spatial organization of waterlogged patches. Although BEAM yielded relatively modest absolute gains, it improved boundary quality via semantic constraints and reduced misclassification in fuzzy transition zones.

Overall, the performance improvement of SAB-DeepLabV3+ can be summarized as a synergistic process of “discrimination enhancement – scale adaptation – boundary refinement”. SSIEM

 doi:10.20944/preprints202604.1571.v1

strengthens class separability, AMSP enhances contextual representation of complex patches, and BEAM further improves local consistency in boundary zones. Collectively, the three support the stable extraction of waterlogged maize-planting regions in complex farmland scenes.

4.3. Cross-Regional Generalization Ability and Agricultural Application Significance

Leave-one-city-out experiments show that SAB-DeepLabV3+ achieved an average mIoU of 76.56%, average IoU-WM of 63.45%, and average OA of 91.38% across five independent cities. Although cross-regional results were lower than within-region results, performance fluctuations across cities were generally small, indicating favorable regional transferability.

Regarding error distribution, cross-regional errors mainly concentrated on the waterlogged maize class, especially in mildly waterlogged regions, boundary transition zones, and fragmented patches. This suggests that regional differences in rainfall processes, topographic drainage conditions, field management practices, and disaster severity jointly affect the remote sensing responses of waterlogged maize, leading to domain shift.

From an agricultural application perspective, the strategy of single-date, high-resolution multispectral imagery combined with a lightweight segmentation model can provide spatial distribution information of waterlogged regions shortly after disasters, supporting key field targeting, field verification, drainage scheduling, agricultural insurance surveys, and disaster loss statistics. For major maize-producing areas in Northeast China, this method effectively balances accuracy with timeliness and deployment potential.

4.4. Research Limitations and Future Work

Although this study confirmed the application potential of single-date multispectral imagery for rapid maize waterlogging mapping, several limitations remain and should be treated with caution. First, the study area was concentrated in Heilongjiang Province, and applicability across wider climate zones and cropping systems has not been validated. Second, only RGB-NIR bands were used, without auxiliary information such as SAR, SWIR, thermal infrared, or terrain data, leaving room for improvement in identifying latent damaged areas under persistent cloudy and rainy conditions. Third, the current method relied on pre-extracted maize planting region masks. The task was essentially binary segmentation within maize fields rather than end-to-end full-scene identification. Lastly, only binary classification was performed, without further characterizing disaster levels or yield impacts.

Future research can be extended in the following directions: (1) Expanding cross-regional, cross-year, and cross-sensor samples to improve model robustness; (2) Fusing multi-source data including optical, radar, and terrain information to enhance identification performance under complex conditions; (3) Exploring an integrated framework for maize planting region extraction and waterlogging identification to reduce dependence on external masks; (4) Extending from binary classification to disaster level assessment and yield risk analysis, enabling the results to directly support agricultural management and disaster decision-making.

5. Conclusions

Aiming at rapid maize waterlogging mapping using single-date, high-resolution multispectral imagery, this study proposed SAB-DeepLabV3+, a spectral-spatial collaboratively enhanced semantic segmentation model based on DeepLabV3+. Subsequently, systematic experiments were conducted on a dataset constructed from 62 scenes covering five typical major maize-producing cities in Heilongjiang Province of China from 2022 to 2024. The main conclusions are as follows:

(1) Single-date RGB-NIR imagery provides a discriminative basis for maize waterlogging identification. Spectral separability analysis shows that the JM distance of the four-band combined feature was 1.3291, higher than that of single bands and typical vegetation indices. Traditional single-date baseline experiments further confirm that single-date multispectral imagery contains usable

 doi:10.20944/preprints202604.1571.v1

waterlogging information, yet simple thresholding and shallow machine learning methods cannot meet high-precision identification requirements.

(2) SAB-DeepLabV3+ effectively improves the identification accuracy of the waterlogged maize class. Compared with the baseline DeepLabV3+, the proposed model increased the IoU, mIoU, mF1, and OA of the waterlogged maize class from 62.23% to 68.30%, from 76.74% to 80.37%, from 86.07% to 88.62%, and from 92.35% to 93.49%, respectively.

(3) The SSIEM, AMSP, and BEAM modules were used to model spectral aliasing, scale heterogeneity, and boundary blurring, respectively, and achieved complementary gains. Ablation experiments and module comparisons show that all three modules outperformed their corresponding alternative structures, and the combination achieved the optimal performance.

(4) The proposed method exhibits favorable cross-regional transfer potential. In leave-one-city-out experiments, SAB-DeepLabV3+ achieved average mIoU, IoU-WM, and OA of 76.56%, 63.45%, and 91.38%, respectively, indicating stable identification performance in unseen regions. Given pre-determined maize planting regions, this method can support rapid post-disaster waterlogging mapping, field verification, drainage scheduling, and agricultural insurance surveys.

Overall, this study demonstrates that, in the absence of time-series information, task-driven design targeting spectral aliasing, scale heterogeneity, and boundary blurring can significantly improve pixel-level accuracy and cross-regional stability of the identification of waterlogged regions within maize fields, providing a feasible solution for single-date remote sensing identification of agricultural disasters.

Author Contributions: Jiahao An: writing-original draft; Qingxue Wang: Software, writing-review & editing; Chunshan Wang: methodology; Xiang Sun: data curation; Qingwei Tian: validation; Jin Yuan: visualization.

Funding: This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFD1601304, and in part by the the National Natural Science Foundation of China (Grant No. 62472012).

Data Availability Statement: Due to the interests of partners, the dataset used in this study is currently not publicly available. Consideration will be given to making the dataset open after the completion of the collaborative project.

Acknowledgments: We are grateful to our colleagues at Hebei Key Laboratory of Agricultural Big Data and National Sub-center for Digital Agriculture Innovation (Beijing-Tianjin-Hebei Region) for their help and input, without which this study would not have been possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Acharya, B.; Dodla, S.; Tubana, B.; Gentimis, T.; Rontani, F.; Adhikari, R.; Duron, D.; Bortolon, G.; Setiyono, T. Characterizing Optimum N Rate in Waterlogged Maize (Zea mays L.) with Unmanned Aerial Vehicle (UAV) Remote Sensing. *Agronomy*. **2025**, 15, 434.
2. Li, X.; Feng, Y.; Sun, X.; Liu, W.; Yang, W.; Ge, X.; Jia, Y. Effects of Various Levels of Water Stress on Morpho-Physiological Traits and Spectral Reflectance of Maize at Seedling Growth Stage. *Agronomy*. **2024**, 14, 2173.
3. Min, C.; Yoon, I.; Kim, M.; Jung, J.; Rahman, M.A.; Lee, B. Effects of Waterlogging at Different Developmental Stages on Growth, Yield and Physiological Responses of Forage Maize. *Agronomy*. **2025**, 15, 2389.
4. Zhi, F.; Zhang, J.; Bao, Y.; Bao, Y.; Dong, Z.; Tong, Z.; Liu, X. Assessment of waterlogging hazard during maize growth stage in the Songliao plain based on daily scale SPEI and SMAI. *Agr. Water Manage.* **2024**, 304, 109081.
5. Chen, H.; Liang, Q.; Liang, Z.; Liu, Y.; Xie, S. Remote-sensing disturbance detection index to identify spatio-temporal varying flood impact on crop production. *Agr. Forest Meteorol.* **2019**, 269-270, 180-91.

 doi:10.20944/preprints202604.1571.v1

6. Teixeira, A.C.; Bakon, M.; Lopes, D.; Cunha, A.; Sousa, J.J. A systematic review on soil moisture estimation using remote sensing data for agricultural applications. *Science of Remote Sensing*. **2025**, 12, 100328.
7. Badarneh, O.; Hazaymeh, K.; Almagbile, A.; Shogoor, S.A. Remote sensing-based agricultural drought mapping in Northern Jordan using Landsat and MODIS data. *Environmental Advances*. **2024**, 18, 100602.
8. Thapa, A.; Horanont, T.; Neupane, B. Parcel-Level Flood and Drought Detection for Insurance Using Sentinel-2A, Sentinel-1 SAR GRD and Mobile Images. *Remote Sens.* **2022**, 14, 6095.
9. Andrade, J.; Cunha, J.; Silva, J.; Rufino, I.; Galvão, C. Evaluating single and multi-date Landsat classifications of land-cover in a seasonally dry tropical forest. *Remote Sensing Applications: Society and Environment*. **2021**, 22, 100515.
10. Karmakar, P.; Teng, S.W.; Murshed, M.; Pang, S.; Li, Y.; Lin, H. Crop monitoring by multimodal remote sensing: A review. *Remote Sensing Applications: Society and Environment*. **2024**, 33, 101093.
11. Mazzia, V.; Khaliq, A.; Chiaberge, M. Improvement in Land Cover and Crop Classification based on Temporal Features Learning from Sentinel-2 Data Using Recurrent-Convolutional Neural Network (R-CNN). *Appl. Sci*. **2020**, 10, 238.
12. Galieni, A.; D'Ascenzo, N.; Stagnari, F.; Pagnani, G.; Xie, Q.; Pisante, M. Past and Future of Plant Stress Detection: An Overview From Remote Sensing to Positron Emission Tomography. *Front. Plant Sci.* **2021**, 11, 609155.
13. Cho, S.B.; Soleh, H.M.; Choi, J.W.; Hwang, W.; Lee, H.; Cho, Y.; Cho, B.; Kim, M.S.; Baek, I.; Kim, G. Recent Methods for Evaluating Crop Water Stress Using AI Techniques: A Review. *Sensors* **2024**, 24, 6313.
14. Omia, E.; Bae, H.; Park, E.; Kim, M.S.; Baek, I.; Kabenge, I.; Cho, B. Remote Sensing in Field Crop Monitoring: A Comprehensive Review of Sensor Systems, Data Analyses and Recent Advances. *Remote Sens.* **2023**, 15, 354.
15. Ding, Y.; Zheng, X.; Zhao, K.; Xin, X.; Liu, H. Quantifying the Impact of NDVIsoil Determination Methods and NDVIsoil Variability on the Estimation of Fractional Vegetation Cover in Northeast China. *Remote Sens.* **2016**, 8, 29.
16. Ren, J.; Shao, Y.; Wan, H.; Xie, Y.; Campos, A. A two-step mapping of irrigated corn with multi-temporal MODIS and Landsat analysis ready data. *ISPRS J. Photogramm.* **2021**, 176, 69-82.
17. Zhang, J.; Pan, B.; Shi, W.; Zhang, Y. Monitoring Waterlogging Damage of Winter Wheat Based on HYDRUS-1D and WOFOST Coupled Model and Assimilated Soil Moisture Data of Remote Sensing. *Remote Sens.* **2023**, 15, 4133.
18. Zheng, G.; Jiang, Z.; Zhang, X.; Jiang, D. Multi-scale feature fusion-based semantic segmentation network for agricultural remote sensing images. *Chemical and Biological Technologies in Agriculture*. **2025**, 12, 126.
19. Nazir, A.; Ahmad, A.; Ramzan, M.; Gilani, H.; Mobeen, M.; Tarer, S.; Hanan, N.P. Flood-Induced Agricultural Damage Assessment: A Case Study of Pakistan. *Water.* **2025**, 17, 3060.
20. Wei, P.; Ye, H.; Nie, C.; Qin, M.; Zhang, Y.; Wang, H.; Huang, S.; Liu, R. Evaluating the impact of crop waterlogging and flood disasters using multi-source data: a case study of the Sanjiang Plain. *Climate Services*. **2025**, 39, 100596.
21. Sun, J.; Zhou, J.; He, Y.; Jia, H.; Liang, Z. RL-DeepLabv3+: A lightweight rice lodging semantic segmentation model for unmanned rice harvester. *Comput. Electron. Agr.* **2023**, 209, 107823.
22. Wang, Y.; Gao, X.; Sun, Y.; Liu, Y.; Wang, L.; Liu, M. Semantic segmentation-based conservation tillage corn straw return cover type recognition. *Comput. Electron. Agr.* **2025**, 229, 109792.
23. Lai, P.; Lv, C.; Zhou, L.; Yang, S.; Xu, J.; Dong, Q.; He, M. Improved lightweight DeepLabV3+ for bare rock extraction from high-resolution UAV imagery. *Ecol. Inform.* **2025**, 89, 103204.
24. Zhang, X.; Zhang, S.; Meng, X.; Zhang, G.; Zang, D.; Han, Y.; Ai, H.; Liu, H. Remote sensing image segmentation of gully erosion in a typical black soil area in Northeast China based on improved DeepLabV3+ model. *Ecol. Inform.* **2024**, 84, 102929.
25. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE T. Pattern Anal.* **2017**, 39, 2481-95.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*. **2015**,arXiv: 1505.04597.

 doi:10.20944/preprints202604.1571.v1

27. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *arXiv*. **2018**,arXiv: 1807.10165.
28. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. *arXiv*. **2020**.
29. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation. *arXiv*. **2020**,arXiv: 2004.02147.
30. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017,6230-9.
31. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017,936-44.
32. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Hartwig, A. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv*. **2018**,arXiv: 1802.02611.
33. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.*; et al.*. Deep High-Resolution Representation Learning for Visual Recognition. *arXiv*. **2021**,arXiv: 1802.02611.
34. Xu, Q.; Ma, Z.; HE, N.; Duan, W. DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation. *Comput. Biol. Med.* **2023**, 154, 106626.
35. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *arXiv*. **2021**,arXiv: 2105.15203.
36. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv*. **2021**,arXiv: 2102.04306.
37. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Manning, W. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv*. **2021**,arXiv: 2105.05537.