Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v44e20230110/2024

ENHANCED U-NET ALGORITHM FOR TYPICAL CROP CLASSIFICATION USING GF-6 WFV REMOTE SENSING IMAGES

Yinjiang Jia[1], Hao Lan[1], Renshan Jia[2], Kang Fu[1], Zhongbin Su[1*]

[1*]Corresponding author. Northeast Agricultural University/Harbin, China.
E-mail: suzb001@163.com | ORCID ID: https://orcid.org/0000-0002-8966-8933

KEYWORDS

multi-scale features; spatial attention; U-Net.

ABSTRACT

Accurate crop classification, crucial for a macro-level understanding of food production, formulating relevant agricultural policies, and predicting comprehensive agricultural productivity, enables precise crop distribution. In remote sensing image classification, feature selection and representation play a pivotal role in accuracy. An augmented U-Net algorithm, named ASPP-SAM-UNet, integrating spatial attention mechanisms and multi-scale features is proposed for the enhancement of typical crop classification accuracy in remote sensing. The ASPP-SAM-UNet design integrates features over multiple scales, boosts the representational capacity of shallow features, and expands the neural network's receptive field by incorporating Atrous Spatial Pyramid Pooling (ASPP) into the convolutional components of the standard U-Net encoder via residual connections. The integration of the residual module allows for a profound fusion of deep and shallow features, thereby enhancing their utility. The spatial attention mechanism amalgamates spatial and semantic information, empowering the decoder to reclaim more spatial information. This study focused on Bayan County, Harbin City, Heilongjiang Province, China, employing GF-6 WFV remote sensing images for crop classification. Empirical outcomes showed a significant improvement in classification accuracy with the advanced algorithm, boosting the overall accuracy (OA) from 89.49 to 92.80%. Specifically, the segmentation accuracy for maize, rice, and soybean increased from 89.90, 89.96, and 87.37% to 93.47, 94.82, and 89.35%, respectively. The suggested algorithm offers a pioneering performance standard for crop classification leveraging GF-6 WFV remote sensing imagery.

INTRODUCTION

Remote sensing technology significantly contributes to fields such as crop monitoring, geological surveying, and precision agriculture. Wide-scale crop classification, an application of remote sensing technology, is a pressing research challenge (Jia et al., 2022). To estimate crop productivity, it is essential to classify crops accurately and quickly, enhance crop production management, and guide agricultural insurance policies. Remote sensing technology enables accurate national-scale monitoring of crop growth and yield, furnishing agricultural producers and managers with detailed information on farmland and crops, and timely insights into crop distribution, growth status, and production. Such information can subsequently guide agricultural production and decision making, thereby enhancing grain yield and quality (Kang et al., 2021). An improved understanding of land use aids in optimising the structure of agricultural production and land resource utilisation, consequently boosting agricultural production efficiency.

Predominantly, conventional crop type mapping techniques resort to machine learning methodologies, such as the random forest algorithm (Yang et al., 2019; Pott et al., 2021). Nonetheless, deep learning has developed throughout time, and classification methods leveraging this technology, including Convolutional Neural Networks (CNN) and enhanced Transformers, have found extensive application in crop classification. Yang et al. (2020)

[1] Northeast Agricultural University/Harbin, China.
[2] Heilongjiang Polytechnic/Harbin, China.

Area Editor: Gizele Ingrid Gadotti
Received in: 7-31-2023
Accepted in: 1-25-2024

employed a combination of Convolutional Neural Networks and Random Forests, classifying multitemporal optical remote sensing picture crops by first extracting features using CNNs and then utilising those features in a Random Forest classifier (Wang et al., 2021). U-Net, a precursor to CNN, was initially deployed for medical image segmentation due to its ability to generate precise segmentation results with minimal data. Many studies have employed U-Net for land object classification projects, suggesting multiple advanced U-Net network architectures for improved semantic segmentation performance. Since U-Net loses extensive detailed data during downsampling, the inclusion of ASPP in U-Net assists in maintaining this data and expanding the receptive field. An ASPP module was included in the lower layers of the U-Net by Zhang et al. (2018), facilitating the use of multi-scale contextual information for extraction by feature maps and diminishing confusion between adjacent pixels of different types. Cao & Zhang (2020) proposed Res-UNet, a hybrid of ResNet and U-Net. In this model, ResNet's residual units supersede U-Net's convolutional layers, thereby easing the propagation of shallow features to deeper ones and enhancing the differentiation between features with minor spectral differences (Dave et al., 2022). The attention mechanism demonstrates its effectiveness in computer vision assignment by emphasising significant representation features and minimising irrelevant ones. Consequently, many studies have integrated attention mechanisms into the classification of remote sensing images (John & Zhang, 2022). A U-Net network was used by Bian et al. (2022), who added a channel attention mechanism to further the model's capacity to abstract spectral characteristics. In their studies of the VAIHINGEN dataset, ISPRS (2018) and Li et al. (2022) presented MAResU-Net, a model that includes a multi-stage CAM attention module based on the U-Net network. A spectral and spatial dual attention network was presented by Zhu et al. (2021) for the classification of high-spectral (HIS) images, and the results were promising.

The aforementioned algorithms offer a variety of enhancement and optimisation strategies for diverse classification tasks, providing research insights for this study. According to empirical evidence, the deepening of a network can result in the fragmentation of spatial information and a subsequent decrease in spatial resolution. This research presents an enhanced U-Net algorithm to improve the accuracy of remote sensing image classification. This algorithm synergistically combines spatial attention and multi-scale features (Baesso et al., 2023). The proposed methodology chiefly employs dilated convolutions across diverse scales to enlarge the receptive field. This feature allows the network to integrate multi-scale characteristics proficiently, thereby enhancing the role of shallow information (Shao et al., 2022). Furthermore, the fusion of residual modules allows for the efficient integration of shallow and deep features, thus effectively utilising the advantages of both types (Wang et al., 2022a). Additionally, SAM is used to strengthen the fusion of semantic and spatial information by embedding additional spatial data into the upsampled feature maps. SAM integrates the feature maps obtained from skip connections with the upsampled feature maps.

This study aimed to strengthen and refine the elements involved in crop classification utilising GF-6 WFV remote sensing imagery by integrating U-Net with ASPP and SAM. The primary contributions of this study involve: (1) an examination of the impacts of different U-Net feature levels on crop classification using 16-m resolution GF-6 WFV imagery; (2) the integration of the original U-Net with the ASPP module, thereby enhancing the combination of multi-scale features and representation of shallow information; and (3) utilising SAM to generate a spatial weight matrix for feature maps rich in spatial information. This approach allows the spatial weight matrix to interact with appropriate semantic feature maps, yielding feature maps that integrate both spatial and semantic information.

MATERIAL AND METHODS

Overview of the study area

The research site lies in Bayan County, Harbin, Heilongjiang Province, China (Fig. 1), nestled in the core of the Songnen Plain, along the northern bank of the Songhua River's midsection. The geographic coordinates of the research site span from 46°8'10" to 46°18'58" N and from 127°6'42" to 127°21'41" E. A moderate continental monsoon climate dominates the region, with windy, dry springs, warm, wet summers, cool, damp autumns, and chilly winters. The region experiences extended annual sunshine durations and brief frost-free periods, maintaining an average annual temperature of 2.6℃. This region predominantly grows three major crops: soybean, corn, and rice, while other crops and vegetation are less prevalent.

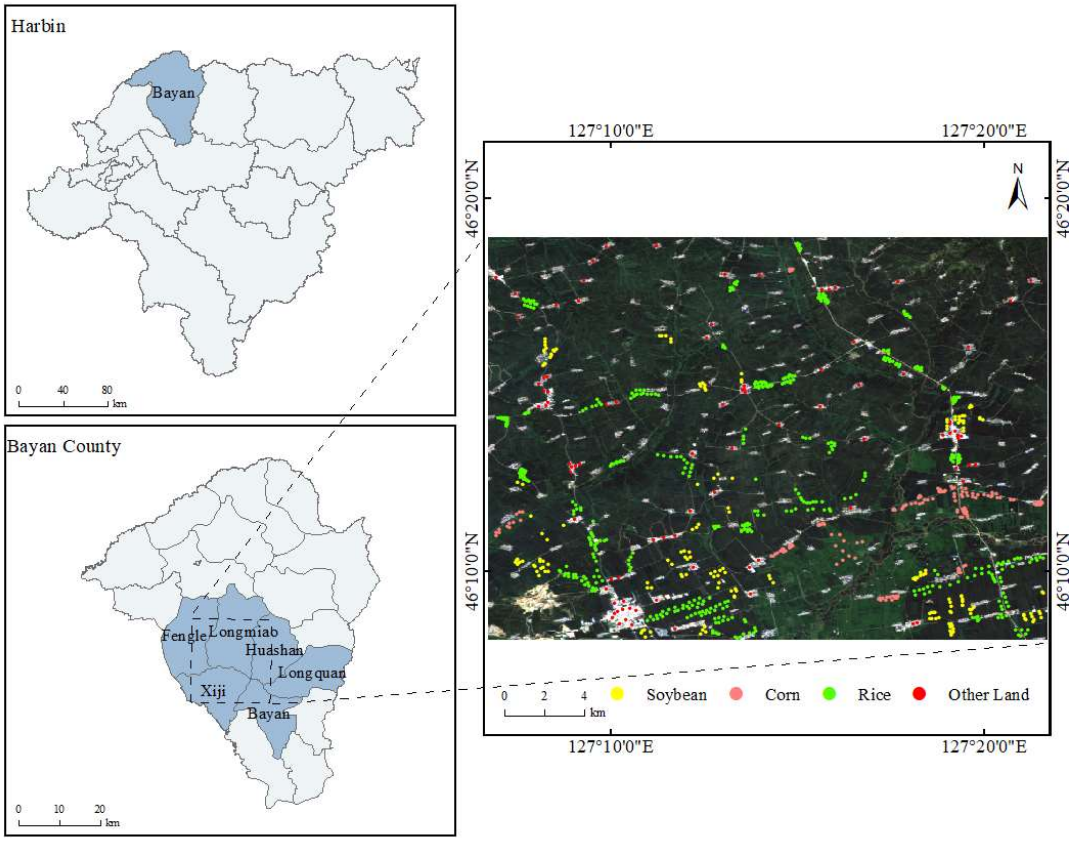

FIGURE 1. Study area.

Data

Field investigations were carried out in July 2022, coinciding with the crops' critical growth stage. An HCE320 from Huace facilitated on-site surveying, leading to the collection and observation of 885 sample points with diverse feature types, as demonstrated in Table 1 and annotated in Fig. 1. Other prevalent land uses encompassed urban buildings, villages, and roads.

TABLE 1. Number of field measurements per land cover type.

No.	1	2	3	4	Total
Type	Soybean	Corn	Rice	Other Land	
Number	179	473	143	90	885

The GF-6 spacecraft had a medium resolution multispectral camera (WFV) (Kang et al., 2021) with a 16-m wide field and a 2-m panchromatic/8-m high resolution multispectral camera (PMS).The WFV data included two red-edge bands, a purple band, and a yellow band. The different spectral characteristics of the crops were vividly revealed by the red-edge bands. The WFV data were distinguished by their high resolution and extensive coverage, boasting an observational swath extending up to 800 km. Table 2 delineates the principal parameters of the WFV data.

TABLE 2. GF-6 WFV main parameter information.

Band No.	Band Name	Centre Wavelength (nm)
B1	Blue	485
B2	Green	555
B3	Red	660
B4	Near-infrared	830
B5	Red Edge 1	710
B6	Red Edge 2	750
B7	Purple	425
B8	Yellow	610

In this study, three unobstructed GF-6 WFV images (23/7/2022, 1/8/2022, 3/9/2022) were gathered during the crops' critical growth stage. For more effective image information extraction, ENVI facilitated preprocessing operations, including geolocation, radiometric correction, atmospheric correction, geometric correction, latitude and longitude conversion, mosaicking, and cropping. Field survey samples were visually interpreted (Wang et al., 2022c), complemented by higher resolution optical remote sensing images (GF-2). Utilising ArcGIS Pro, four labels corresponding to land cover types were generated to constitute a reference dataset, as depicted in Fig. 2. The data were set aside for testing, validation, and training. By setting the classifier parameters, the classification model was trained using a training dataset (Zhu et al., 2021). A test dataset was used to assess the final classification performance, while a validation dataset was used to identify the model's ideal parameters. The study's sample database contained a total of 4,661,250 samples, with a distribution of 60% for training, 20% for validation, and 20% for testing. The sliding window technique was applied to enhance the diversity of the training samples by segmenting the image into 256 × 256 pixel sizes and sliding 32 pixels at each step, thereby ensuring 224 overlapping pixels between neighbouring image blocks. The training, validation, and test sets each contained a tiny portion of the intersecting region.

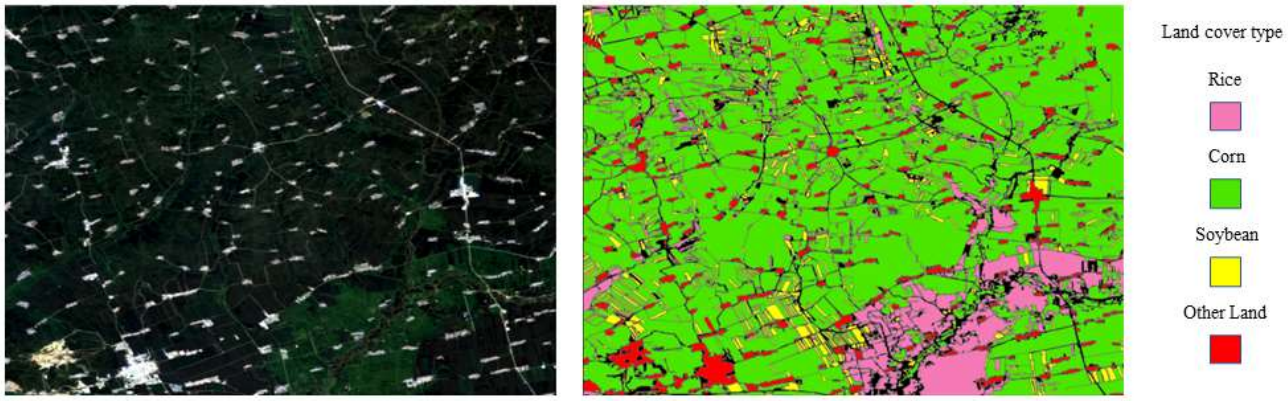

FIGURE 2. Reference to labels in the dataset.

Impact of varied level features on crop categorisation in GF-6 WFV images

Fig. 3 demonstrates the use of skip connections by U-Net to integrate features from several levels, enhancing the integration of encoder and decoder characteristics and obtaining more precise information from the picture.

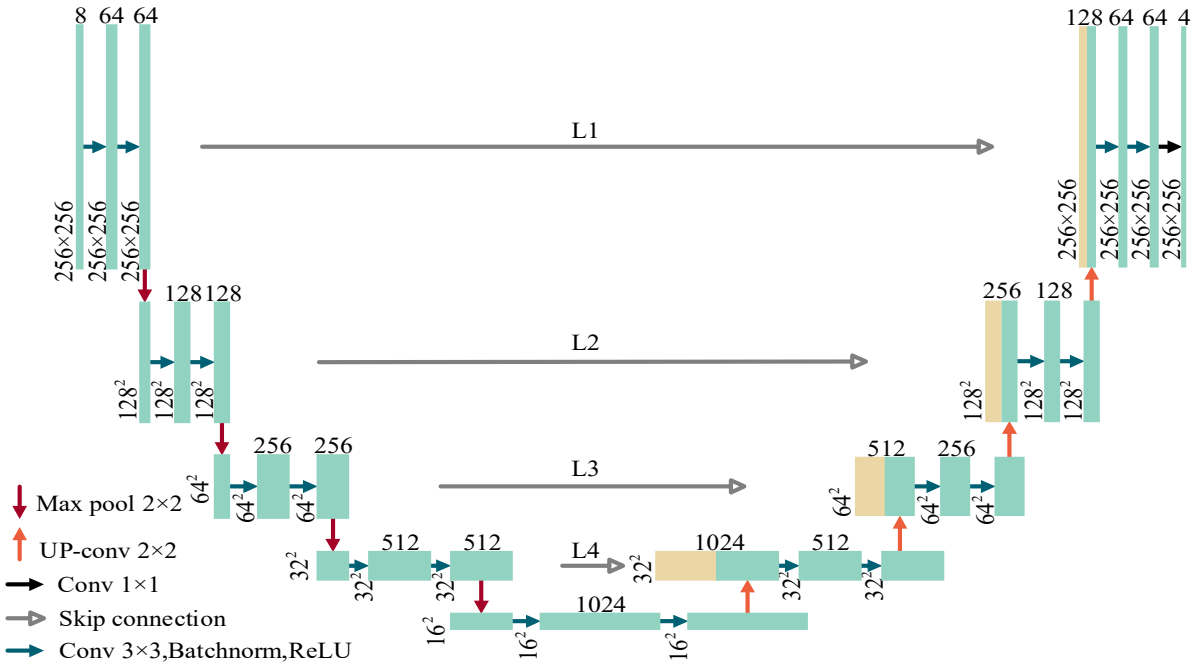

FIGURE 3. U-Net structure.

U-Net++ constructs a full-size U-Net out of a set of dense skip connections. Nonetheless, in the context of GF-6 WFV remote sensing images, classification accuracy is not always improved by denser skip connections. In some instances, skip connections can potentially trigger adverse effects. Therefore, utilising the study area data, a thorough analysis of U-Net's effects on the classification outcomes of GF-6 WFV images at various feature levels was conducted. The outcomes are displayed in Fig. 4.

(1) The U-Net version without skip connections accomplished the poorest accuracy, with a 17.16% drop in overall accuracy when contradistinguished to the original U-Net (Bian et al., 2022). The "U-Net-None" decoder merely upscales the feature map to recover the input size without integrating the encoder's data, which results in a significant loss of information.

(2) The classification results were affected differently by features from different levels: U-Net showed an accuracy of 89.49%; U-Net-L1 showed 89.13%; and U-Net-w/o L1 produced 85.72%. In comparison to L2, L3, and L4, L1 contributed the most significantly to the U-Net, demonstrating that shallow characteristics had a substantial role in the results of categorisation.

(3) U-Net-w/o L1 was more accurate than U-Net-L2, U-Net-L3, and U-Net-L4. However, compared to U-Net-L1, U-Net-w/o L2, U-Net-w/o L3, and U-Net-w/o L4 were all less accurate. This observation suggests that a combination of shallow and deep characteristics is not desirable.

The contribution of features from disparate levels to the results exhibits variability. Consequently, the expression capability of features with more substantial contributions should be augmented. These outcomes underscore the criticality of enhancing the manifestation of shallow features and enriching semantic information.

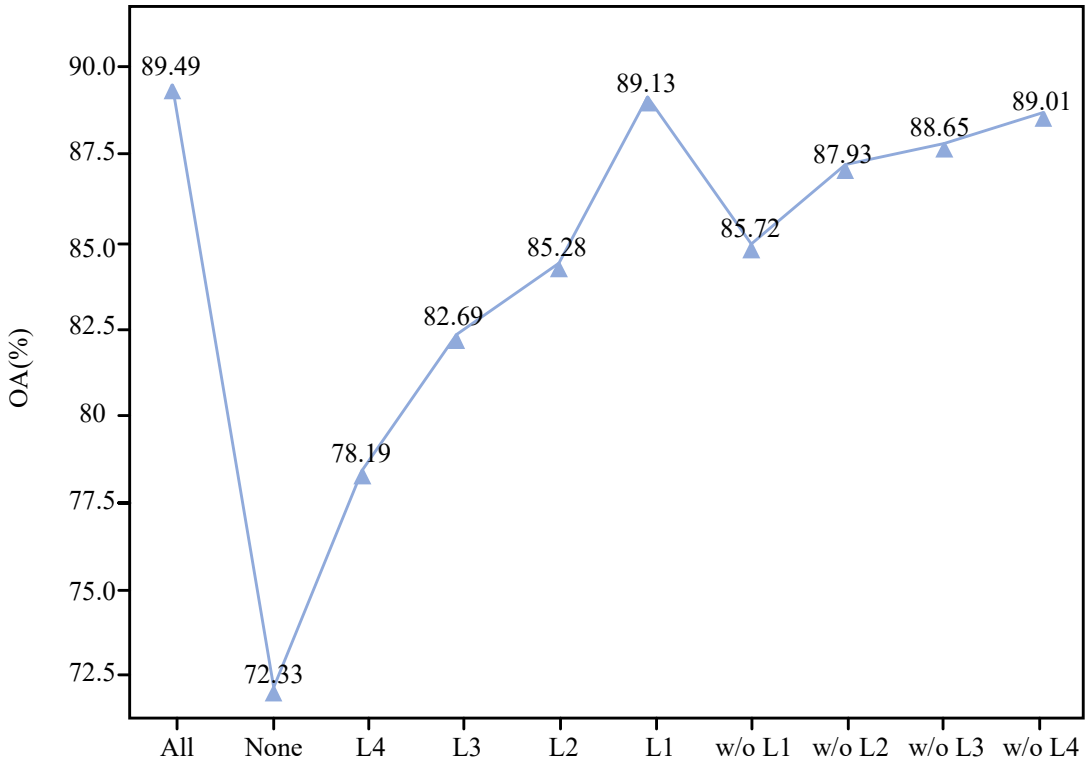

FIGURE 4. Analysis of the characteristics of the different levels of U-Net fusion. The term "All" denotes the original U-Net network; "None" signifies the absence of skip connections; "L1" embodies the retention of only the first-level skip connection; "w/o L1" designates removal of only the first-level skip connection.

ASPP-SAM-UNet

The ASPP-SAM-UNet network structure was created to overcome the difficulties mentioned in the previous section. Fig. 5's representation of the ASPP-SAM-UNet's overall structure consists of two interconnected components: an encoder and a decoder. Five ASPP modules, four SAMs, and five residual modules make up the total architecture. Each convolutional layer in the backbone network is combined with a batch normalisation layer and a ReLU layer (Chamundeeswari et al., 2022). A ReLU layer in the ASPP module is placed after the atrous convolution and is identified by its pooling size of 2×2, convolution kernel size of 2×2, and transposed convolution stride of 2×2.

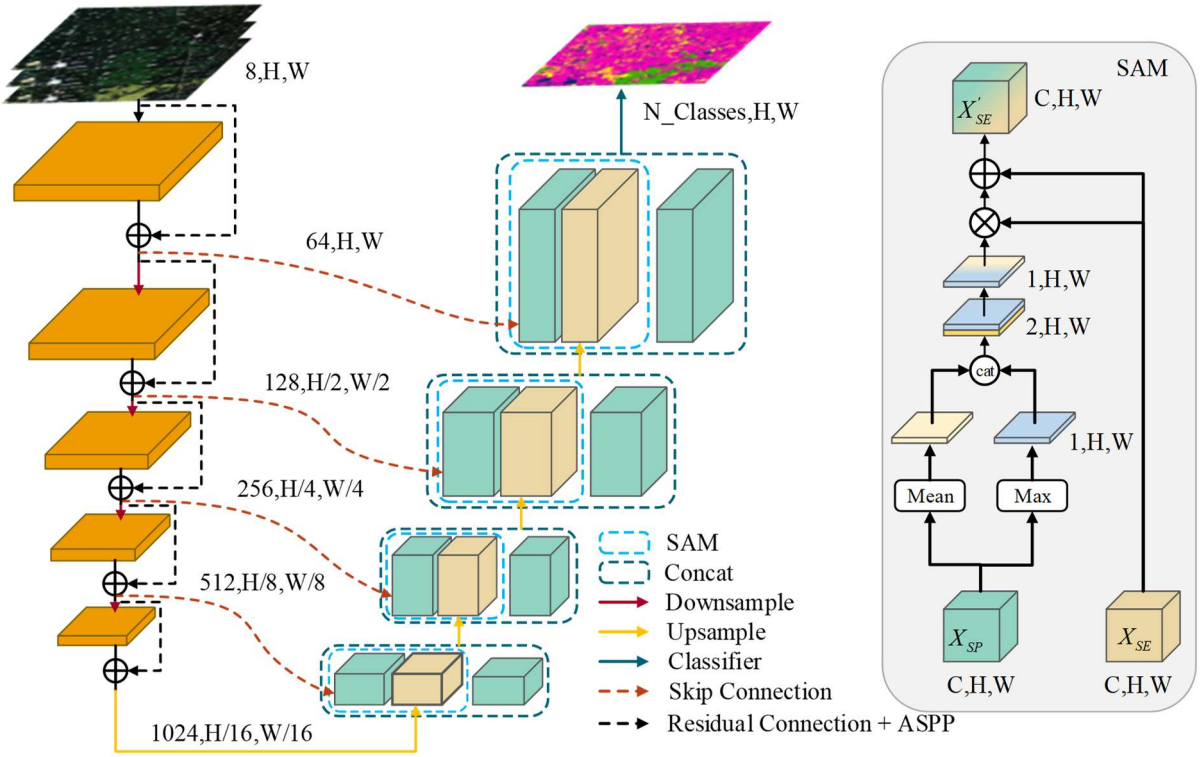

FIGURE 5. ASPP-SAM-UNet structure.

ASPP

As shown in Fig. 6, the ASPP module consists of five parallel branches. The first branch embodies a standard 1×1 convolution layer, while the second, third, and fourth branches utilise a 3×3 atrous convolution with rates of 18, 12, and 6, respectively. In the fifth branch, the global average pooling is used, resulting in an output of (batchsize, in_channel, 1,1), followed by a 1×1 convolution to regulate the number of channels. The feature map is ultimately returned to its original input size using bicubic interpolation. Dimensionally aligned features have been drawn from the five branches. The number of output channels is five times greater than the number of input channels. To produce the final result, a 1×1 convolution is used to alter the channel count. By adding additional gaps to the conventional convolution, the ASPP module inflates the receptive field and reduces the spatial resolution degradation caused by the maximum pooling layer. As a result, by combining the multi-scale features and receptive field (Sykas et al., 2022; Wang et al., 2022b), the expressivity of shallow qualities may be increased. Furthermore, a successful fusion of shallow and deep features is accomplished by connecting the ASPP module to the backbone network and integrating the remaining structure, enhancing the application of both shallow and deep feature characteristics (Menon et al., 2021).

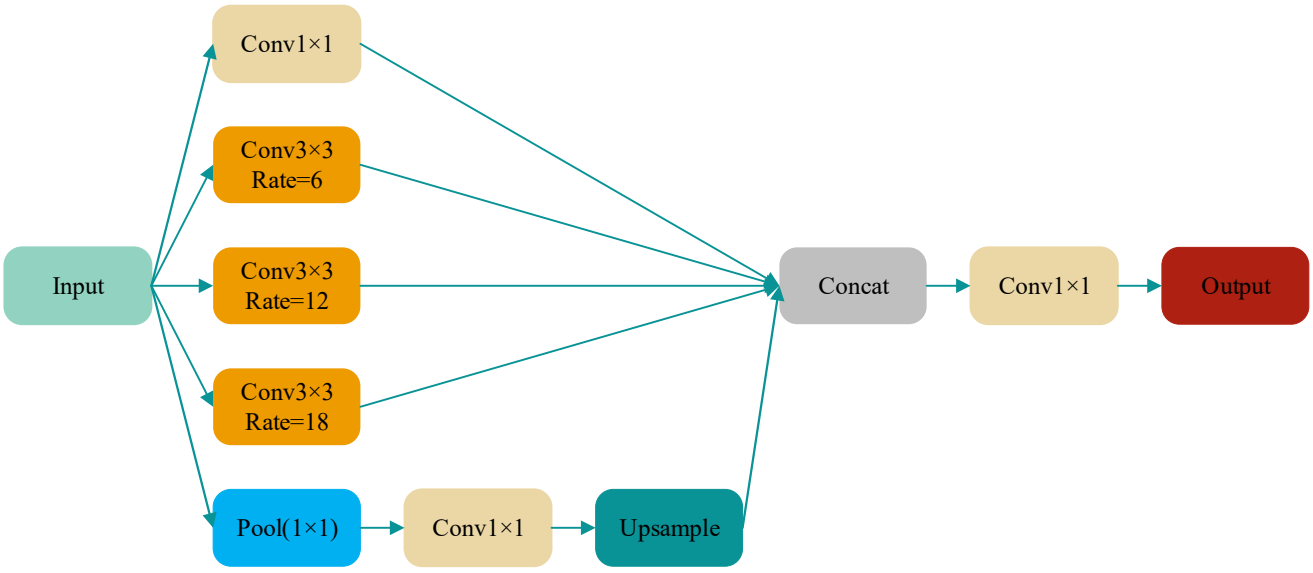

FIGURE 6. ASPP module.

SAM

The mechanism of spatial attention concentrates on the data's position within the current task. The use of remote sensing technology to obtain semantic and spatial information on remote sensing pictures, where land item types are many and intricately distributed, describes a way to enhance land object identification competency. The SAM borrows from the concept of CBAM (Zhang et al., 2022). To secure the spatial feature weight map W_S, the SAM initially makes use of the feature map of the spatial information pathway (X_{SP}). Subsequently, to produce representative features, it multiplies the semantic information pathway's feature map (X_{SE}) by the appropriate spatial position, which are appended to the corresponding spatial position of X_{SE} to yield X'_{SE}.

To append spatial weights, two channel information feature descriptors, $X^S_{SP_{\text{avg}}} \in R^{1\times H\times W}$ and $X^S_{SP_{\max}} \in R^{1\times H\times W}$, are initially procured. These are concatenated using average and maximum sets along the feature map's channel axis, and the spatial attention map is produced using a 7×7 convolution process. Ultimately, the spatial feature weight map W_S is ultimately produced by rescaling the spatial attention map using the sigmoid function to a range from 0 to 1.

Spatial attention is calculated as follows:

$$\begin{aligned} X'_{SE} &= \sigma\left(f^{7\times7}([\text{AvgPool}(X_{SP}); \text{MaxPool}(X_{SP})])\right) \times X_{SE} + X_{SE} \\ &= \sigma\left(f^{7\times7}\left(\left[X^S_{SP_{avg}}, X^S_{SP_{max}}\right]\right)\right) \times X_{SE} + X_{SE} \\ &= W_S \times X_{SE} + X_{SE} \end{aligned} \quad (1)$$

Where:

X_{SE}, X_{SP} — the feature maps of the semantic and spatial information paths, respectively;

$f^{7\times7}$ — the convolution process using a 7×7 grid size;

σ — sigmoid function;

MaxPool — greatest pooling along the channel axis for each pixel,

AvgPool — average pooling along the channel axis for every pixel.

Loss function

Because of its superior performance, the cross-entropy loss function (Zhang et al., 2021) is usually used as the loss function for multi-classification problems. The following formula shows how the cross-entropy loss function compares the expected and target classes for each pixel:

$$L = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log(p_{ic}) \quad (2)$$

Where:

M — the number of classes;

y_{ic} — functions as the indicator (0 or 1), such that y_{ic} equals 1 when the actual class of the i-th sample equals c, and 0 otherwise,

p_{ic} — the predicted probability of the observed i-th sample being classified under class c.

Evaluation indicators

The accuracy of crop classification underwent evaluation using the test dataset. This research incorporated four evaluation parameters predicated on pixel assessment, namely overall accuracy (OA), Kappa coefficient (Yang et al., 2020), precision, and average accuracy (AA), to quantify the accuracy of crop classification.

OA, the ratio of correctly classified crop pixels to all crop pixels, was calculated as follows:

$$OA = \frac{\sum_{i=1}^{n} p_{i,i}}{\sum_{j=1}^{n}\sum_{i=1}^{n} p_{i,j}} \quad (3)$$

Where:

$p_{i,j}$ — total number of pixels categorised as class j and class i,

n — quantity of classes.

The Kappa coefficient was computed based on the confusion matrix, delineated as follows:

$$Kappa = \frac{N^2 \times OA - \sum_{i=1}^{n} a_i b_i}{N^2 - \sum_{i=1}^{n} a_i b_i} \quad (4)$$

Where:

N — total count of samples;

a_i — actual sample number for each category,

b_i — predicted sample number of each category.

In comparison with the ground truth reference data, a True Positive (TP) occurs when both the classifier's prediction and the actual are positive, signifying the count of positive samples accurately identified. False Positive (FP) arises when the classifier's prediction is positive but the actual is negative, denoting the quantity of negative samples incorrectly reported (Guo et al., 2019). Precision was employed to depict the proportion of accurately classified categories within the results extracted from a given category. It was computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

AA was the average derived from the sum of accuracies for all categories.

Experimental environment

During the training phase, the following were chosen: the cross-entropy loss function, an 8-batch size, a 20-epoch maximum, and Adam as the optimisation algorithm. To lessen the effects of gradient fluctuations, in addition to momentum gradient descent, the Adam optimiser uses a gradient descent technique with a variable learning rate. The learning rate was initially set at 0.01, and after every fifth iteration, it fell to 0.1 times the starting rate. The model was trained on a computer running Windows 10 with an Intel Xeon Gold 6244 processor, an NVIDIA Quadro RTX 5000 GPU, Python 3.9, and Pytorch 1.10.2.

RESULTS AND DISCUSSION

In the research area, the sample pool was distributed precisely, as shown in Table 3, with 60% designated for training, 20% for validation, and the remaining 20% for testing. The sample pool information listed in Table 3 is used in the studies that follow. Due to space limitations, other land use types were shortened to 'OL'.

TABLE 3. Information on the categories of data in the sample pool for the study area.

No.	Class	Train	Val	Test
1	Corn	1,957,726	702,547	603,668
2	Rice	307,642	107,147	94,810
3	Soybean	251,701	63,902	104,796
4	OL	289,512	119,837	97,965
	total	2,806,581	993,433	901,239

To ascertain a suitable learning rate, we conducted tests with the proposed model at learning rates of 0.001, 0.01, and 0.1. The outcomes, displayed in Table 4, reveal that the OA, AA, and Kappa values peaked when the learning rate was fixed at 0.01.

TABLE 4. Learning rate options.

LR	Precision（%）				Evaluation Metrics (%)		
	Corn	Rice	Soybean	OL	OA	AA	Kappa
0.1	92.37	87.51	90.66	90.02	91.40	90.14	84.37
0.01	93.47	94.82	89.35	90.41	92.80	92.01	86.49
0.001	92.61	92.26	89.22	90.93	92.00	91.26	85.72

Ablation experiment

To evaluate the efficacy of the proposed network design, along with its two key modules, we performed ablation investigations utilising U-Net as the basis network on the test set inside the study domain.

(1) The influence of ASPP is displayed in Table 5. This module, incorporated as residuals into the U-Net, enabled segmentation of the images in the test set. We observed improvements in overall accuracy by 3.02%, AA by 2.3%, and the Kappa coefficient by 4.15%. There was an enhancement in the recognition accuracy for corn (+3.59%), rice (+2.29%), soybean (+2.02%), and other land use types (+1.32%).

This substantiated the efficacy of assimilating ASPP into U-Net in residual form. Upon incorporating ASPP into U-Net, we noted a more precise segmentation of corn and soybean compared to the use of U-Net alone. According to the findings, the network focused on both general information and minute details when ASPP and residual units were used together (Wang et al., 2022 a).

(2) The influence of SAM: We noted an enhancement in overall accuracy by 2.81%, AA by 2.60%, and the Kappa coefficient by 3.91% in the images of the test set. SAM amalgamated spatial and semantic data, thus improving the recognition precision for corn (+3%), rice (+4.55%), soybean (+1.32%), and other land use types (+1.56%).

Following the addition of SAM, we noted a higher accuracy in rice segmentation compared to using U-Net alone. The network demonstrated an enhanced capability to distinguish between categories with larger inter-group variations, while categories with smaller intra-group differences proved less discernible. The findings indicate that SAM accentuates inter-group disparities and makes it easier to separate characteristics from big disparities.

(3) The impact of ASPP+SAM: ASPP-SAM-UNet reduced misclassification within and between various groups. As depicted in Table 4, there were improvements in overall accuracy by 3.31%, AA by 3.01%, and the Kappa coefficient by 4.97%. We noted an increase in the recognition accuracy for corn (+3.57%), rice (+4.86%), soybean (+1.98%), and other land uses (+1.66%). Thus, ASPP-SAM-UNet concentrated on information across varying scales, thereby enhancing the accuracy of land object classification outcomes.

TABLE 5. Ablation experiments of the proposed module on the test set.

Model Name	Modules		Precision（%）				Evaluation Metrics (%)		
	ASPP	SAM	Corn	Rice	Soybean	OL	OA	AA	Kappa
U-Net			89.90	89.96	87.37	88.75	89.49	89.00	81.52
ASPP-UNet	√		93.49	92.25	89.39	90.07	92.51	91.30	85.67
SAM-UNet		√	92.90	94.51	88.69	90.31	92.30	91.60	85.43
ASPP-SAM-UNet	√	√	93.47	94.82	89.35	90.41	92.80	92.01	86.49

Comparative analysis of different methods

The efficacy of ASPP-SAM-UNet in crop classification was assessed via comparative studies and analyses alongside U-Net, U-Net++, Attention-UNet, SAR-UNet, Swin-Unet (Cao et al., 2023), UCTransNet (Chen et al., 2021), CAM-UNet+SVM (Yan et al., 2022), and Res-UNet++. U-Net++ enhanced U-Net through a refined fusion of comprehensive features. An attention mechanism built on the U-Net framework was integrated into Attention-UNet. ResUNet++ bolstered segmentation performance by integrating residual and dense connections into the U-Net structure. UCTransNet, a transformer-based neural network architecture, merged convolutional neural networks and a self-attention mechanism to manage spatial information and contextual relationships in images. Swin-UNet represented an advancement and expansion of the U-Net architecture, adopting a Swin Transformer to amplify image segmentation performance. CAM-UNet+SVM replaced the classification layer in the original U-Net network with a support vector machine by adding a channel attention module to the original U-Net framework and fused the multi-feature classification results using a majority voting game-theoretic algorithm.

Table 6 illustrates the segmentation outcomes of the test set, utilising different algorithms. Using OA, AA, and KAPPA evaluation metrics, the proposed ASPP-SAM-Unet network surpassed other algorithms, registering OA (92.80%), AA (92.01%), and KAPPA (86.49%). Compared to other algorithms, ASPP-SAM-Unet reached the utmost segmentation precision for rice (94.82%), soybean (89.35%), and other land uses (90.41%). Notably, its corn segmentation accuracy (93.47%) was slightly outperformed by SAR-Unet's accuracy rate (93.60%). Compared to the network proposed in this study, SAR-Unet yielded satisfactory segmentation outcomes for corn (93.60%), soybean (89.19%), and other land uses (90.15%). However, its rice segmentation fell short, achieving only a 90.56% success rate. Despite U-Net++ adopting a full-scale fusion strategy, its capacity for generalisation experienced a drop compared to U-Net. Attention-UNet, Res-UNet++, and Swin-UNet secured satisfactory segmentation outcomes and heightened precision compared to U-Net. However, Res-UNet++ underperformed in soybean segmentation, and Swin-UNet revealed a marginal dip in the segmentation accuracy for other land uses. In comparison to U-Net, UCTransNet enhanced segmentation accuracy for corn and rice but exhibited a reduction in precision for soybean and other land uses. The CAM-UNet+SVM performed well overall but was slightly inferior to the ASPP-SAM-UNet in every aspect. This implies that ASPP-SAM-UNet possesses a competitive advantage in crop classification within GF-6 WFV remote sensing imagery.

TABLE 6. Comparison of test set segmentation results by different methods.

Model Name	Precision（%）				Evaluation Metrics (%)		
	Corn	Rice	Soybean	OL	OA	AA	Kappa
U-Net	89.90	89.96	87.37	88.75	89.49	89.00	81.52
U-Net++	89.09	85.76	87.97	88.43	88.54	87.81	76.38
Attention-UNet	92.15	91.69	88.84	89.02	91.38	90.43	84.17
SAR-UNet	93.60	90.56	89.19	90.15	92.39	90.88	85.03
Res-UNet++	93.10	92.95	86.83	90.28	92.05	90.79	84.55
UCTransNet	90.76	91.03	85.60	87.49	89.83	88.72	80.69
Swin-UNet	91.83	92.91	88.24	87.16	91.02	90.04	83.86
CAM-UNet+SVM	93.17	93.28	88.59	89.80	92.34	91.21	84.92
ASPP-SAM-UNet	93.47	94.82	89.35	90.41	92.80	92.01	86.49

The confusion matrix of the algorithm proposed in this study is presented in Table 7. Corn accounted for the most missing pixels, primarily misclassified as soybean (29,870 pixels). Fewer misclassified pixels were observed in rice and other land uses, with the majority of soybean misinterpreted as corn (7646 pixels). Field surveys showed a more regimented pattern in corn cultivation in the study area, in contrast to the more scattered soybean planting. Some areas exhibited mixed corn planting, resulting in mixed pixel phenomena. Additionally, the relatively high spectral reflectance of soybean may induce a shift in the reflectance of mixed pixels towards soybean, subsequently leading to instances of corn being misclassified as soybean. The 16-m resolution of GF-6 WFV remote sensing images rendered the distinction between scattered corn and soybean cultivation areas challenging. A pixel in an image with a 16-m resolution may encompass other land uses and farmland, potentially impacting the accuracy of the experiment.

TABLE 7. Test set confusion matrix based on the algorithm proposed in this paper.

Class	Corn	Rice	Soybean	OL	Total
Corn	564,248	1342	7646	1261	574,497
Rice	2472	89,899	347	6960	99,678
Soybean	29,870	2312	93,636	1174	126,992
OL	7078	1257	3167	88,570	100,072
total	603,668	94,810	104,796	97,965	901,239
User's Accuracy (%)	98.22	90.19	73.73	88.51	-
Producer's Accuracy (%)	93.47	94.82	89.35	90.41	-
OA (%)			92.80		
kappa			0.8649		

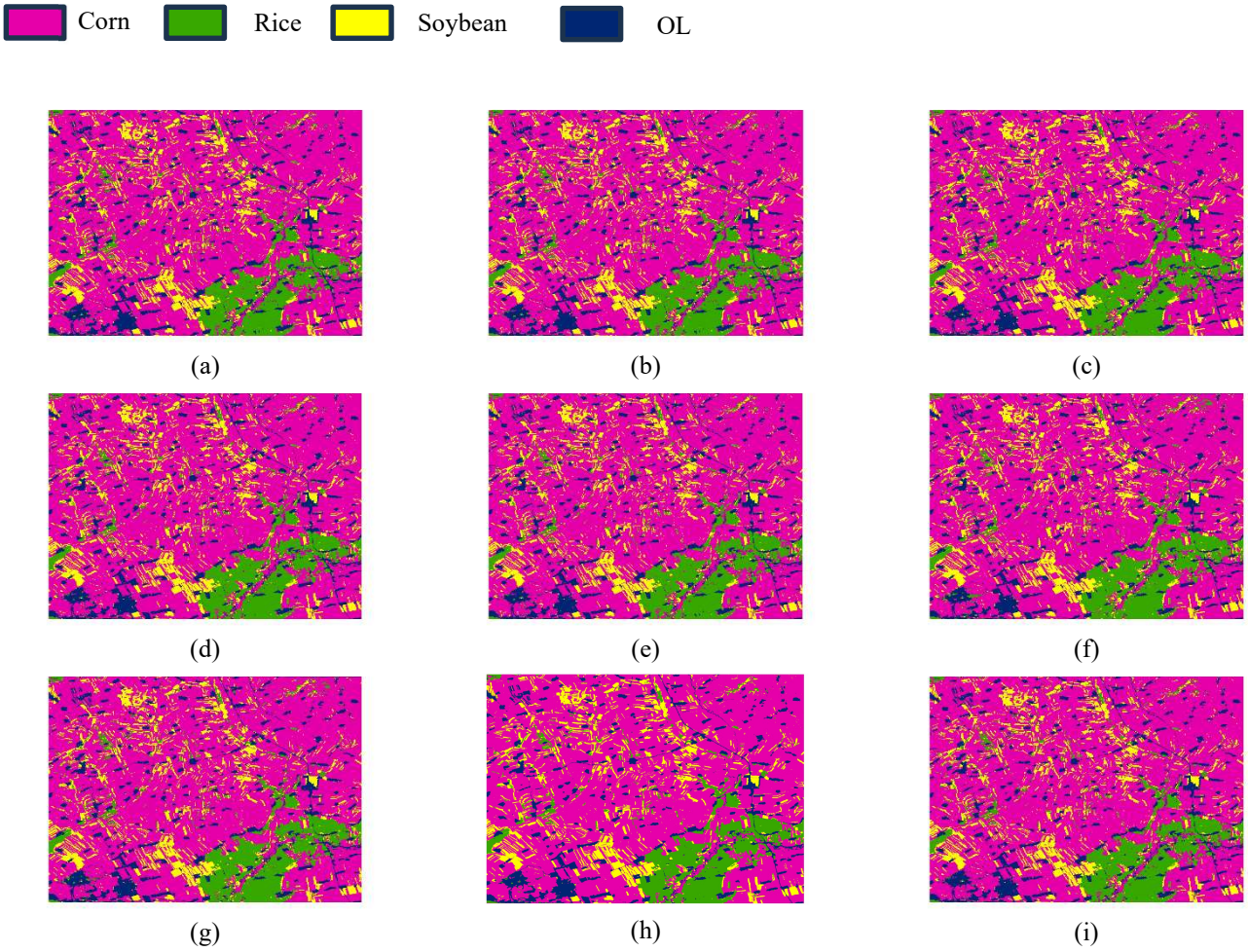

FIGURE 7. Comparison of remote sensing image categorisation outcomes from various approaches, where (a–i) are, in order, U-Net, U-Net++, Attention-UNet, SAR-UNet, Res-UNet++, UCTransNet, Swin-UNet, CAM-UNet+SVM, and ASPP-SAM-UNet.

The results of remote sensing image segmentation in the study area, employing diverse methods, are presented in Fig. 7. In general, image segmentation outcomes across different algorithms exhibited substantial similarity. As shown in Fig. 7, the towns and villages in the study area were very dispersed, with maize having the largest cultivated area, soybean and rice being cultivated in smaller areas, soybean cultivation being dispersed, and rice fields being more concentrated.

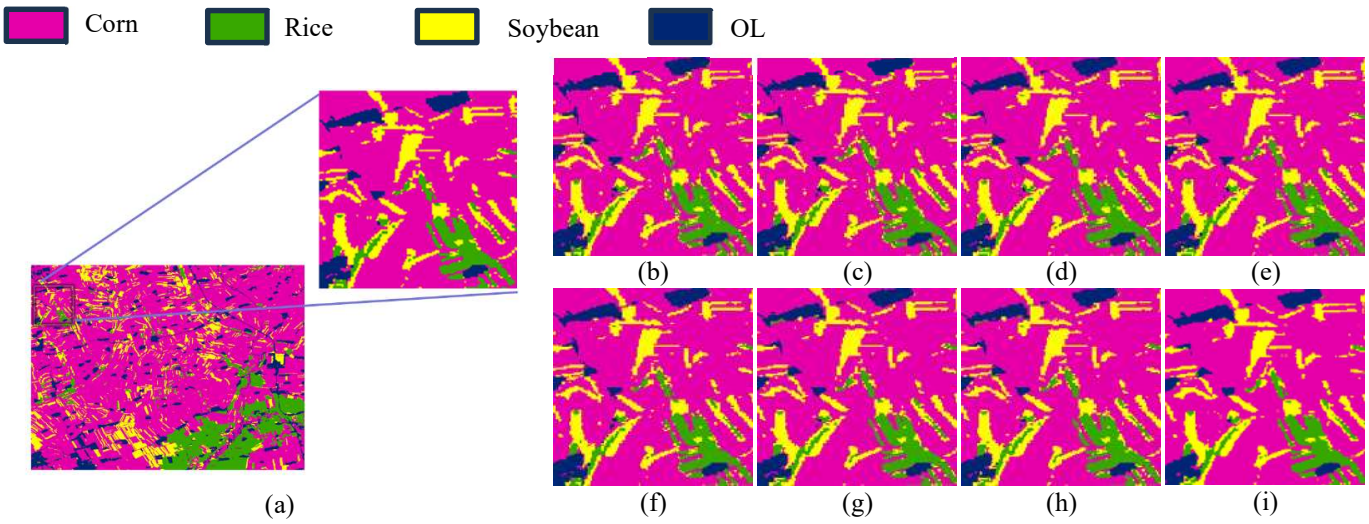

FIGURE 8. Results of the test set's local semantic segmentation, where a shows the local position and (a–i) shows the segmentation of ASPP-SAM-UNet, U-Net, U-Net++, Attention-UNet, SAR-UNet, Res-UNet++, UCTransNet, Swin-UNet, and CAM-UNet+SVM in that order.

Algorithmic differences were primarily discernible in local areas. As indicated in Fig. 8, soybean cultivation was notably dispersed, with instances of mixed cropping evident. Comparative observation revealed that the algorithm proposed in this study produced superior segmentation results, closely mirroring actual ground conditions. The efficacy of the proposed algorithm was corroborated by amalgamating the data from Table 6 and Figs. 7 and 8.

In crop classification utilising GF-6 WFV remote sensing imagery, U-Net surpassed U-Net++ in accuracy (Yan et al., 2022). This performance difference can be traced back to U-Net++'s full-scale fusion strategy, which is advantageous in enhancing inter-group differentiation and bolstering classification accuracy when feature type disparities are substantial. However, with minimal feature type variations, it is vital to perform ablation studies on skip connections and to choose optimal features for fusion. As demonstrated in Section 2.3, features at diverse levels contribute differently to the outcomes, with shallow features playing a crucial role in the classification results. In crop classification using 16-m high-resolution remote sensing imagery, shallow features considerably impacted the experimental outcomes. However, as the network's depth increases, the spatial resolution diminishes, leading to a dispersion of spatial information. Consequently, the proposed ASPP-SAM-UNet in this research employed dilated convolutions at multiple scales to expand the receptive field, enabling multi-scale fusion of features and, consequently, amplifying the representational capacity of shallow features. Moreover, by incorporating a residual module, ASPP-SAM-UNet fuses shallow and deep features, thereby effectively capitalising on the properties of both feature levels. The spatial attention module combines the feature map derived from skip connections with the upsampled feature map to provide more spatial information to the upsampled feature map (Ge et al., 2021), strengthening the merging of spatial and semantic data. Relative to U-Net, ASPP-SAM-UNet enhanced classification accuracy for all typical crop types. Rice, corn, and soybean constitute the main food crops in our country, making the accurate extraction of their planting conditions from remote sensing images essential. However, utilising this approach still results in the subpar segmentation of maize and soybean, demanding further accuracy improvements.

CONCLUSIONS

The main objective of this research was to strengthen the integration of semantic and spatial information and to enhance the shallow features' capability to be represented in remote sensing images, capturing a global context and improving the segmentation effectiveness of terrestrial features. In this research, residual fusion was employed to combine U-Net and the ASPP module. This fusion not only broadens the perceptual field through variable-sized dilated convolutions, facilitating multi-scale feature fusion and augmenting shallow features' representational capacity but also accomplishes deep integration of shallow and semantic features, mitigating interferences from intricate local feature types. Furthermore, the spatial attention module helps to merge the upsampled feature map with the feature map produced from skip connections, resolving the issue of insufficient spatial data consumption during the upsampling process. The findings suggest that compared to U-Net, U-Net++, Attention-UNet, SAR-UNet, Res-UNet++, UCTransNet, and Swin-UNet, ASPP-SAM-UNet offers superior accuracy in the crop classification of GF-6 WFV remote sensing imagery within the study region. The algorithm substantially enhances the classification precision of remote sensing imagery, presenting new technical benchmarks for crop classification within GF-6 WFV remote sensing images.

ACKNOWLEDGEMENTS

This work was supported by Science and Technology Innovation 2030 – "new generation artificial intelligence" major project (No. 2021ZD0110904).

REFERENCES

Baesso M, Leveghin L, Sardinha E, Oliveira G, Sousa R (2023) Deep learning-based model for classification of bean nitrogen status using digital canopy imaging. Engenharia Agrícola 43(2):e20230068. http://dx.doi.org/10.1590/1809-4430-eng.agric.v43n2e20230068/2023

Bian Y, Li L, Jing W (2022) CACPU-Net: channel attention U-net constrained by point features for crop type mapping. Frontiers in Plant Science 13: 1030595. http://dx.doi.org/10.3389/fpls.2022.1030595

Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2023) Swin-Unet: unet-like pure transformer for medical image segmentation. In: Tel Aviv, European Conference on Computer Visión.

Cao K, Zhang X (2020) An improved Res-UNet model for tree species classification using airborne high-resolution images. Remote Sensing 12: 1128. https://doi.org/10.3390/rs12071128

Chamundeeswari G, Srinivasan S, Bharathi SP, Priya P, Kannammal GR, Rajendran S (2022) Optimal deep convolutional neural network based crop classification model on multispectral remote sensing images. Microprocessors and Microsystems 94:104626. http://dx.doi.org/10.1016/j.micpro.2022.104626

Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y (2021) TransUNet: transformers make strong encoders for medical image segmentation. East China, Pii. 13p.

Dave K, Vyas T, Trivedi YN (2022) Band selection technique for crop classification using hyperspectral data. Journal of the Indian Society of Remote Sensing 50(8): 1487-1498. http://dx.doi.org/10.1007/s12524-022-01545-4

Ge Z, Cao G, Shi H, Zhang Y, Li X, Fu PJRS (2021) Compound multiscale weak dense network with hybrid attention for hyperspectral image classification. Remote Sensing 13(16):3305

Guo Y, Cao H, Bai J, Bai Y (2019) High efficient deep feature extraction and classification of spectral-spatial hyperspectral image using cross domain convolutional neural networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12(1): 345-356. http://dx.doi.org/10.1109/JSTARS.2018.2888808

Jia Y, Zhang X, Zhang H, Su Z (2022) Crop classification based on a gaofen 1/wide-field-view time series. Engenharia Agrícola 42(2):e20210184. http://dx.doi.org/10.1590/1809-4430-eng.agric.v42n2e20210184/2022

John D, Zhang C (2022) An attention-based U-Net for detecting deforestation within satellite sensor imagery. International Journal of Applied Earth Observation and Geoinformation 107: 102685. http://dx.doi.org/10.1016/j.jag.2022.102685

Kang Y, Hu X, Meng Q, Zou Y, Zhang L, Liu M, Zhao M (2021) Land cover and crop classification based on red edge indices features of GF-6 WFV time series data. Remote Sensing 13(22):4522. http://dx.doi.org/10.3390/rs13224522

Li Z, Duan P, Hu S, Li M, Kang X (2022) Fast hyperspectral image dehazing with dark-object subtraction model. IEEE Geoscience and Remote Sensing Letters 19: 1-5. http://dx.doi.org/10.1109/LGRS.2022.3217766

Menon RV, Kalipatnapu S, Chakrabarti I (2021) High speed VLSI architecture for improved region based active contour segmentation technique. Integration 77: 25-37. http://dx.doi.org/https://doi.org/10.1016/j.vlsi.2020.11.004

Pott LP, Amado TJC, Schwalbert RA, Corassa GM, Ciampitti IA (2021) Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. ISPRS Journal of Photogrammetry and Remote Sensing 176: 196-210. http://dx.doi.org/https://doi.org/10.1016/j.isprsjprs.2021.04.015

Shao Y, Lan J, Niu B (2022) Dual-channel networks with optimal-band selection strategy for arbitrary cropped hyperspectral images classification. IEEE Geoscience and Remote Sensing Letters 19: 1-5. http://dx.doi.org/10.1109/lgrs.2020.3023103

Sykas D, Sdraka M, Zografakis D, Papoutsis I (2022) A Sentinel-2 multiyear, multicountry benchmark dataset for crop classification and segmentation with deep learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15: 3323-3339. http://dx.doi.org/10.1109/jstars.2022.3164771

Wang J, Lv P, Wang H, Shi C (2021) SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in Computed Tomography. Computer Methods and Programs in Biomedicine 208: 106268. http://dx.doi.org/https://doi.org/10.1016/j.cmpb.2021.106268

Wang L, Wang J, Liu Z, Zhu J, Qin F (2022a) Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. The Crop Journal 10(5): 1435-1451. http://dx.doi.org/10.1016/j.cj.2022.01.009

Wang L, Wang J, Zhang X, Wang L, Qin F (2022b) Deep segmentation and classification of complex crops using multi-feature satellite imagery. Computers and Electronics in Agriculture 200:107249. http://dx.doi.org/10.1016/j.compag.2022.107249

Wang S, Feng W, Quan Y, Li Q, Dauphin G, Huang W, Li J, Xing M (2022c) A heterogeneous double ensemble algorithm for soybean planting area extraction in Google Earth Engine. Computers and Electronics in Agriculture 197:106955. http://dx.doi.org/10.1016/j.compag.2022.106955

Yan C, Fan X, Fan J, Wang N (2022) Improved U-Net remote sensing classification algorithm based on multi-feature fusion perception. Remote Sensing 14(5):1118. http://dx.doi.org/10.3390/rs14051118

Yang N, Liu D, Feng Q, Xiong Q, Zhang L, Ren T, Zhao Y, Zhu D, Huang J (2019) Large-scale crop mapping based on machine learning and parallel computation with grids. Remote Sensing 11(12): 1500.

Yang S, Gu L, Li X, Jiang T, Ren R (2020) Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. Remote Sensing 12(19):3119. http://dx.doi.org/10.3390/rs12193119

Zhang H, Liu M, Wang Y, Shang J, Liu X, Li B, Song A, Li Q (2021) Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. International Journal of Applied Earth Observation and Geoinformation 105: 102557. http://dx.doi.org/https://doi.org/10.1016/j.jag.2021.102557

Zhang L, Gao L, Huang C, Wang N, Wang S, Peng M, Zhang X, Tong Q (2022) Crop classification based on the spectrotemporal signature derived from vegetation indices and accumulated temperature. International Journal of Digital Earth 15(1): 626-652. http://dx.doi.org/10.1080/17538947.2022.2036832

Zhang P, Ke Y, Zhang Z, Wang M, Li P, Zhang S (2018) Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. Sensors 18(11): 3717.

Zhu M, Jiao L, Liu F, Yang S, Wang J (2021) Residual spectral–spatial attention network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 59(1): 449-462. http://dx.doi.org/10.1109/TGRS.2020.2994057