RESEARCH ARTICLE

Crop unmanned aerial vehicle image classification based on improved DeepLabv3+

Zhu Yuan*, Mingli Li

School of Automation Engineering, Henan Polytechnic Institute, Nanyang, Henan, China.

Received: March 31, 2026; **accepted:** July 18, 2026.

Agricultural automation has developed rapidly with the advancement of information technology and precision agriculture. Unmanned aerial vehicle (UAV) has been largely adopted in crop photography and classification due to its low cost and rapid response. However, current UAV image crop classification methods still suffer from low recognition accuracy. To address this issue, this study optimized the classification accuracy in UAV-captured images by developing an improved DeepLabv3+ model and constructing a crop classification system. An improved DeepLabv3+ model combining attention mechanism and transformer was designed. Spatial attention and channel attention were introduced to enhance feature extraction, and transformer was integrated to capture long-distance dependencies. A crop classification system was established based on the improved DeepLabv3+ and ResNet50. The improved DeepLabv3+ achieved a mean intersection over union of 90.2% and an average mean square error of 1.7. In the classification system application, the accuracy for corn, sugarcane, rice, and citrus reached 98.4%, 98.6%, 97.9%, and 98.7%, respectively. The system also performed better in classification time, CPU usage, and memory usage compared with other models. The proposed improved DeepLabv3+ model and crop classification system demonstrated good performance and practical value, providing a theoretical basis and technical support for intelligent agricultural development and UAV image classification in related fields.

Keywords: improved DeepLabv3+; UAV; images; remote sensing; crops; image classification; transformer; attention mechanism.

***Corresponding author:** Zhu Yuan, School of Automation Engineering, Henan Polytechnic Institute, Nanyang, Henan 473000, China. Email: ranzo@126.com.

Introduction

Influenced by information technology and precision agriculture, digitalization and intelligence of agricultural production have become important ways to increase crop yields [1, 2]. Precision agriculture is a modern agricultural production system that is supported by information technology and implements agricultural operations and management in a targeted, timed, and quantitative manner based on field spatial variation [3]. As one of the key tools, unmanned aerial vehicle (UAV) technology automatically obtains geospatial information and analyzes it. It has high resolution, fast response, low cost and has been largely applied in low-altitude economy and intelligent identification [4]. Obtaining images of crops in the region through UAV remote sensing technology and using computer technology to classify and identify crops can provide agricultural development guidance [5].

Many studies have been done to explore more accurate classification methods. Scientists proposed a classification method based on gray-level co-occurrence matrix to solve the crop classification accuracy and obtained accuracy of 0.87, but the computational complexity was high [6]. In response to the problem that features are difficult to fully extract when classifying crops from aerial images, Kalita *et al*. built a deep convolutional neural network (DCNN) classification method. The method had better performance but was highly dependent on the data used for training [7]. To optimize the crop classification accuracy, Eladl *et al*. designed a UAV classification method based on a classification framework and compared its performance with other baseline models. The results showed that the framework had the best performance, but its real-time performance needed further verification [8]. To solve the low crop type resolution accuracy, Allu *et al*. designed a classification method using principal component analysis and random forest. The method had good classification effect, but the generalization ability had not been verified [9]. Ajayi *et al*. built a ResNet architecture to solve the poor classification ability of UAV crop images. The ResNet-50 had the best performance, but there were few data samples, and the generalization ability needed further verification [10]. Although the previous research results have achieved certain effects, the current UAV image-based crop classification method still has problems such as low recognition accuracy. DeepLabv3+ is a semantic segmentation model with high accuracy and robustness, which has been largely applied in remote sensing image processing [11]. Cheng *et al*. designed a recognition method using DeepLabv3+ to solve the passable area identification in the intelligent navigation of unmanned surface vessels with the mean intersection over union (MIoU) as 89.7% [12]. In response to the need for much high-quality training data for deep learning algorithms in segmentation tasks, Edpuganti *et al*. proposed an image recognition model based on Sentinel-2A/B RGB images and DeepLabV3+ and obtained the accuracy of 94.3% [13]. Chen *et al*. built a DeepLabv3+ lightweight model to optimize the semantic segmentation accuracy with strong robustness [14]. In response to the low plant disease identification accuracy, Pal *et al*. developed an effective identification framework using uncertainty meta-learning and DeepLabV3+ [15]. Pan *et al*. designed an improved DeepLabV3+ to solve the low detection accuracy of tunnel lining cracks and confirmed that the method was better than the comparative framework [16].

Although the current research applied to UAV image crop classification has achieved certain results, they still have problems such as high computational complexity, low classification accuracy, and poor real-time performance. This study introduced an improved DeepLabv3+ model by combining the transformer model, spatial attention (SA), and channel attention (CA). The improvements better captured long-distance dependencies in UAV images. The attention mechanism improved the attention to key features, improving the ability to classify crop features. A UAV image crop classification method integrating ResNet50 and improved DeepLabv3+ was then constructed to improve the accuracy of UAV images, which not only provided an effective technical means for more accurately identifying crop types in complex agricultural environments, but also helped promote the development of agricultural intelligence and automation. This research provided a theoretical basis and technical support for intelligent agricultural development and precision crop management in related fields.

Materials and methods

Improved DeepLabv3+ model

The proposed improved DeepLabv3+ primarily consists of SA, CA, transformer, and the basic DeepLabv3+ (https://github.com/tensorflow/models/tree/master/research/deeplab) [17], which included an encoder and a decoder. In the encoder part, DCNN and Atrous convolution (Atrous Conv) were first used to extract image features. Then, through the Atrous spatial pyramid pooling (ASPP), 3 × 3 Conv with various expansion rates were used to get multi-scale features. A pooling layer was used to further aggregate the features, and the features were mapped to a new space through a linear projection layer. The captured features were fed into a module composed of multiple transformer layers to capture long-distance dependencies between features [18, 19]. In addition, the CA mechanism was used to enhance important features. The SA mechanism further extracted spatial features. In the decoder part, the channel was first adjusted through a 1 × 1 Conv. An upsampling operation was conducted to restore the spatial resolution. The features that had been upsampled were added to the features that the encoder had produced, and the features were further refined through a 3 × 3 Conv. The upsampling operation was performed again to generate the final crop classification results. The encoder of transformer was stacked by N identical layers, and each layer included several components. The encoder converted the input sequence into a vector representation rich in contextual information for use by the decoder. The core components of the encoder included input layer, position encoding, multi-head attention mechanism (MAM), residual link and layer normalization (Add&Norm) layer, and feed-forward neural network. The position coding was expressed as below.

$$z_0 = (a_p^1 E, a_p^2 E, \mathrm{K}, a_p^N E) + E_{pos} \quad (1)$$

where z_0 was the code. a_p was each image block. E was the embedded position code. E_{pos} was the embedding position. Position encoding allowed the pixel categories of the input sequence to be located. After the positioning was completed, MAM learned the relationship between different subspaces of the input sequence, thereby helping the model capture the contextual relationship of the input sequence. The output of MAM was expressed as follows.

$$Multihead(Q,K,V) = Concat(head_1, \mathrm{K}, head_h)W^O \quad (2)$$

where $Multihead(Q,K,V)$ was the output of MAM. h was the number of attention heads. $Concat$ was splicing. W^O was the parameter matrix. The output of the i -th attention head was expressed as follows.

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

where W_i^Q , W_i^K , and W_i^V were parameter matrices. The output of the l -th layer of the encoder output module processed by MAM was presented as below.

$$z_l = Multihead(LN(z_{l-1})) + z_{l-1} \quad (4)$$

where z_l and z_{l-1} were the outputs of the l and $l-1$ layers processed by MAM, respectively. $LN(g)$ was layer normalization. After MAM captured contextual information, the training efficiency and performance of the model were improved through residual connections and normalization layers. A fully connected layer was taken to perform nonlinear transformation on the features obtained by MAM, thereby improving the expression ability. The output of the fully connected layer was expressed as follows.

$$FNN(x) = max(0, xW_1, b_1)W_2 + b_2 \quad (5)$$

where $FNN(x)$ was the output of the fully connected layer. b_1 and b_2 were paranoia vectors. x was the input vector. W_1 and W_2 were weight matrices. In addition to the core components of the encoder, the decoder also included a masked self-attention mechanism (SAM) to prevent future information from being seen during decoding to avoid information leakage. The SAM better understood contextual relationships by calculating the weight of each position and dynamically assigning the importance of different positions. The output of the SAM was calculated as below.

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (6)$$

where K was all the indexes available for query in the data set, which was the key.

$Attention(Q, K, V)$ was the SAM output. d_k was the dimension of the key vector. Q was the request of the current query. V was the actual content value corresponding to K.

The SA and CA modules in improved DeepLabv3+ model

SA and CA were introduced to optimize the expression ability and segmentation accuracy of DeepLabv3+. SA emphasized important areas and suppressed irrelevant information by assigning learnable weights to different spatial locations such as pixels or regions, improving the performance of deep learning models in tasks. CA enhanced the feature representation capability. By adaptively recalibrating the weight of each channel, it highlighted important features [20]. The SA module first performed global maximum pooling (MP) and global average pooling (AP) of the channel dimension on an input feature map of size $H \times W \times C$ to obtain two $H \times W \times 1$. H and W were the height and width. C was the quantity of channels. The MP and AP were spliced. The feature map size was obtained as $H \times W \times 2$. Based on the $Sigmoid$, the SA weight matrix M_S was obtained as below.

$$M_s \in R^{H,W} \tag{7}$$

where R was a set of real numbers. The SA was then shown as follows.

$$\begin{aligned} M_s(F) &= \sigma(f^{7\times7}([Avpool(F); Maxpool(F)])) \\ &= \sigma(f^{7\times7}[F^s_{avg}; F^s_{max}] \end{aligned} \tag{8}$$

where $M_S(F)$ was the output of SA. $Maxpool$ was maximum average pooling. $Avgpool$ was global pooling. F was the input feature map. σ was the $Sigmoid$. $f^{7\times7}$ was a 7 × 7 convolution kernel. F^s_{avg} and F^s_{max} were the feature vectors after MP and AP. The CA module used MP and AP to aggregate the spatial dimension information, thereby obtaining two channel statistics. In this way, the spatial dimension information was compressed while retaining channel-level features. The obtained channel statistics were then input into the multi-layer perceptron to generate CA weights and adjust the feature maps of different channels to enhance important features and suppress unimportant features. The multi-layer perceptron contained two fully connected layers with the first layer compressing the number of channels from C to C/r, where r was the compression ratio. $ReLU$ was the activation function. The second layer was to restore the number of channels to C. The output results of the multi-layer perceptron were added. After being processed by the $Sigmoid$ activation function, the CA weight matrix M_c was finally obtained as below.

$$M_c \in R^{C\times1\times1} \tag{9}$$

The CA was then presented as follows.

$$\begin{aligned} M_c(F) &= \sigma(MLP(Avgpool(F)) + MLP(Maxpool(F))) \\ &= \sigma(\omega_1(\omega_0(F^c_{avg})) + \omega_1(\omega_0(F^c_{max}))) \end{aligned} \tag{10}$$

where F^c_{avg} was the global pooling feature. F^c_{max} was the maximum pooling feature. MLP was a multi-layer perceptron. ω_1 and ω_0 were the weight matrices of the multi-layer perceptron.

Design of crop UAV image classification system based on improved DeepLabv3+

The field experiments were conducted in Nanning, Guangxi Zhuang Autonomous Region, China. The total testing area was approximately 50 hectares, divided into multiple farmland plots. DJI 4-RTK drone (DJI, Shenzhen, Guangdong, China) was employed for this research with a centimeter-level navigation and positioning system and post-processing kinematic positioning technology to provide high-precision position data. The UAV flight height was set at 100 m above ground level. Based on the crop growth and maturity period, image shooting was performed during daytime in September to October 2023. The tested crops included corn (*Zea mays* L.) and sugarcane (*Saccharum officinarum* L.) in mature stages, rice (*Oryza sativa* L.) in heading stage, and citrus (*Citrus reticulata* Blanco) in fruit development stage.

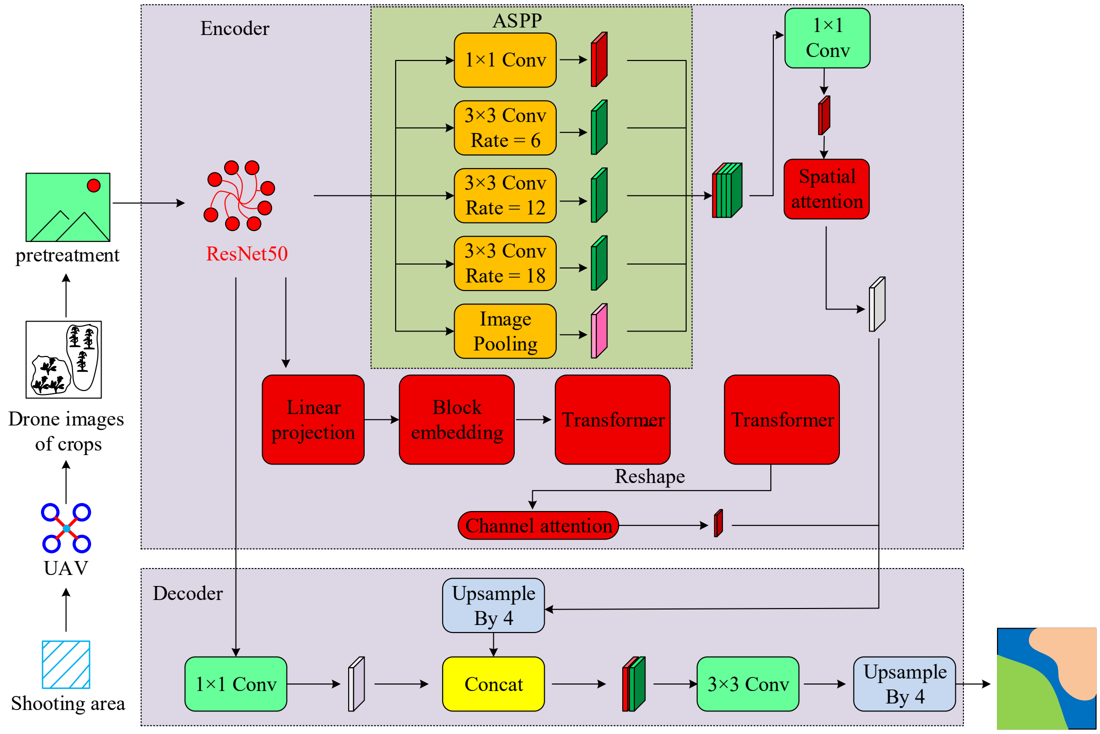

Figure 1. Crop UAV image classification model based on improved Deep Labv3+.

Images captured by UAV were clear and featured high-precision positioning capabilities. After the image acquisition was completed, the image data was preprocessed with all images being unified into a pixel size of 0.059. The ArcGIS (https://www.esri.com) was used to label the crops in the images. The image data was cropped, and the size was unified to 512 × 512. Excluding blank pictures, a total of 19,283 valid data were obtained. 80% of the data was used for model training and 20% of the data was used for testing. To further extract the characteristics of different crops, the study introduced the ResNet50 neural network combined with improved DeepLabv3+ model to establish a UAV image crop classification model (Figure 1). The model took preprocessed UAV crop images as input. After passing through ResNet50 as the encoder for feature extraction, the features extracted were input into the ASPP to capture multi-scale features. The information was then adopted to extract spatial features through SA before inputting into the 12-layer transformer after linear mapping and block embedding mapping to capture long-distance dependencies. The CA module was used to enhance feature representation. Subsequently, the feature maps processed by SA and CA entered the encoder module. After passing four layers of upsampling and 1 × 1 convolution operations, the feature maps of the decoder and encoder were spliced together, and the classification result was output after 3 × 3 convolution operations and four layers of upsampling. The system architecture mainly consisted of a database, data access layer, business layer, and user interaction interface. The database was responsible for storing data such as UAV images, user lists, operation records, and file tables. The database performed data access through MyBatis-Plus (https://baomidou.com) and transmitted the data to the business layer. The business layer included file management, user management, and improved DeepLabv3+, which was responsible for processing business logic such as file management, user management, and image classification. Element-UI (https://element.eleme.io) was used to provide UI components, while Vue.js (https://vuejs.org) was the front-end framework, and Axios (https://axios-http.com) was used to send HTTP requests. The user interaction interface used Nignx (https://nginx.org) reverse proxy to retrieve various data from the business layer for display. The user interaction interface was implemented through Element-UI, Vue.js, and Axios technologies.

Experimental setup and evaluation metrics

To verify proposed improved DeepLabv3+, the research conducted a performance comparison analysis with other exist models including DeepLabv3+, Swin transformer (https://github.com/microsoft/Swin-Transformer), and UNet++ (https://github.com/MrGiovanni/UNetPlusPlus). The experiments were conducted on a workstation equipped Intel Core i9-13900K processor with a main frequency of 5.8 Hz, 32 GB RAM, 500 TB hard drive, and an NVIDIA GeForce RTX 3080Ti graphics card under Windows 10, 64-bit. The deep learning framework was PyTorch 1.12.0 (Meta AI, Menlo Park, CA, USA). Matlab 2023b (MathWorks, Natick, MA, USA) was used for data preprocessing and analysis. The initial learning rate and training batch size of model training were 8 and 1×10^{-3}, respectively, and the number of training rounds was 300. The cross-loss function and the Dice similarity coefficient loss function were adopted, and the total loss was the weighted average of them. The learning rate adopted an exponential decay strategy. The optimization algorithm was Adam, and the momentum parameter and weight decay parameter were 0.8 and 5×10^{-4}. During training, the five-fold cross-validation method was used for parameter tuning. The input size of the transformer model was 16 × 16. The output channel was 512. The encoder was 24. The number of MAM was 16. The embedding sequence length was 1,024. The multi-layer perceptron expansion sequence length was 4,096. The number of output channels of the ASPP module was 256, and the model input image resolution was 512 × 512. The indicators compared included MIoU, classification precision, accuracy, and F1 score. In addition, mean square error (MSE), root mean square error (RMSE), and loss value were also used as comparative indicators to evaluate the regression performance and convergence behavior of the models. MSE measured the average squared difference between the predicted and actual values, while RMSE provided the standard deviation of the residuals. The loss value reflected the optimization degree of the model during training with lower values indicating better convergence.

Ablation experiment on the improved DeepLabv3+

The ablation experiment was conducted with model 1 using the basic DeepLabv3+ model, model 2 combining DeepLabv3+ and transformer, model 3 combining DeepLabv3+, transformer, and SA, model 4 combining DeepLabv3+, transformer, and CA, and model 5 using the improved DeepLabv3+ model.

Statistical analysis

The statistical analysis was performed using SPSS 26.0 (IBM, Armonk, NY, USA). The paired t-test was used to compare the performance differences between the improved DeepLabv3+ and other models. *P* value less than 0.05 was defined as statistically significant difference.

Results

Performance analysis of improved DeepLabv3+ model

Among the five models of ablation experiment, the improved DeepLabv3+ model (model 5) had the highest accuracy, reaching 97.3%, while the accuracy of model 1 was only 86.4%. The accuracy of model 2, model 3, and model 4 was gradually improving. The transformer and CA introduced to improved DeepLabv3+ improved the accuracy (Figure 2a). The precision and F1 score of the model 5 were also much higher than those of other models, which further proved that the introduced algorithm could optimize the performance (Figures 2b and 2c). The comparison results of improved DeepLabv3+

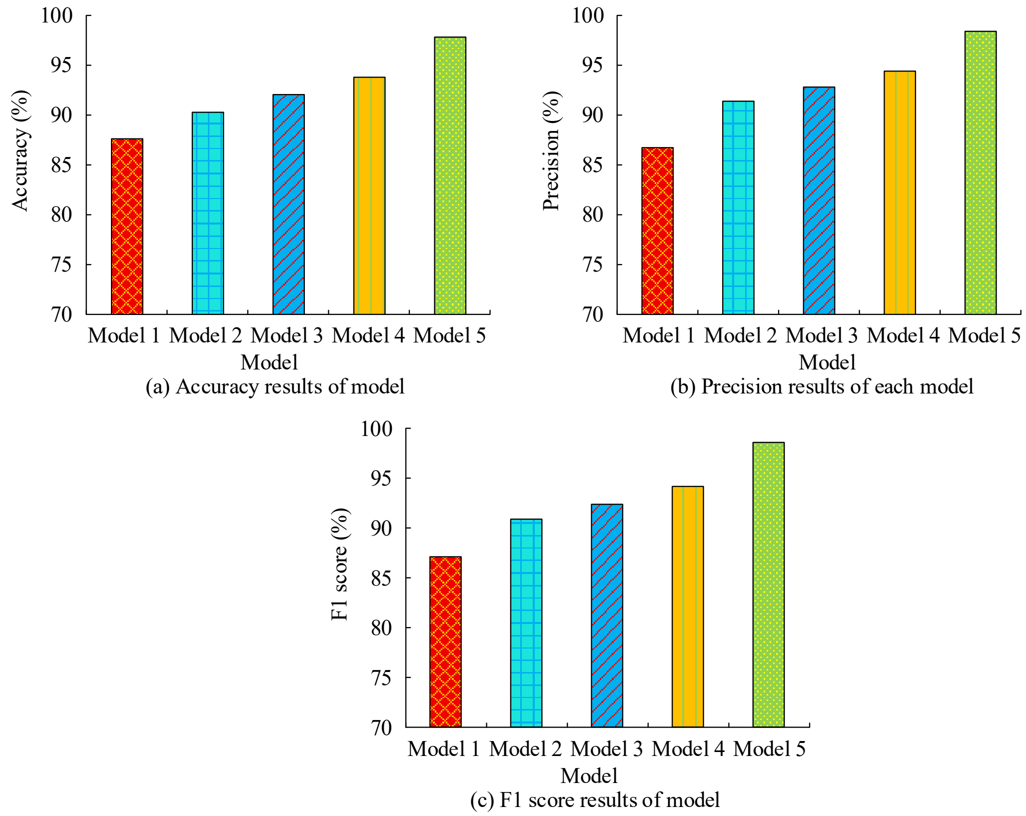

Figure 2. Results of ablation experiment.

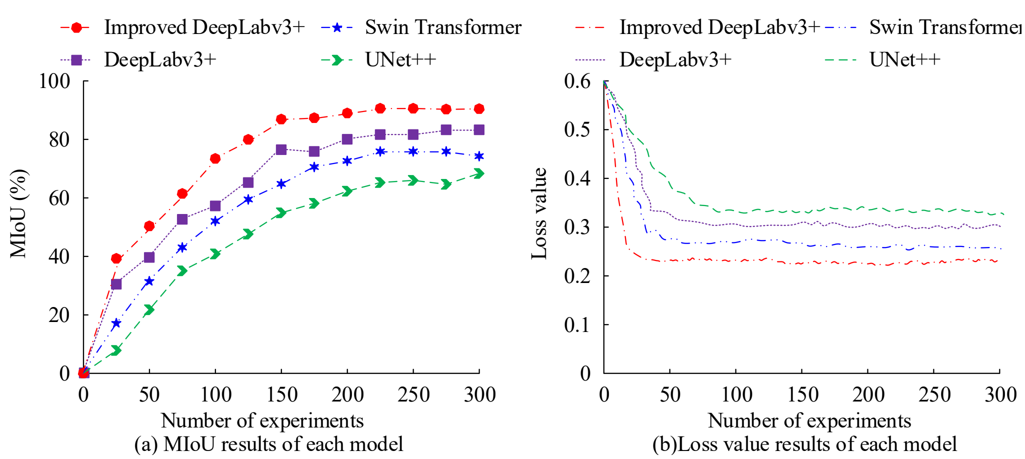

Figure 3. Comparison of loss value and MioU.

with Swin Transformer, UNet++, and DeepLabv3+ showed that, among the four comparison models, the MIoU value of the improved DeepLabv3+ was the highest one, reaching 90.2%. The UNet++ model had the lowest MIoU value, only 62.1%, while the MIoU values of the DeepLabv3+ model and Swin transformer model were 87.6% and 72.1%, respectively (Figure 3a). The loss value of the improved DeepLabv3+ was 0.23, which was lower than that of the other three models (Figure 3b). Statistical analysis confirmed that the differences between the

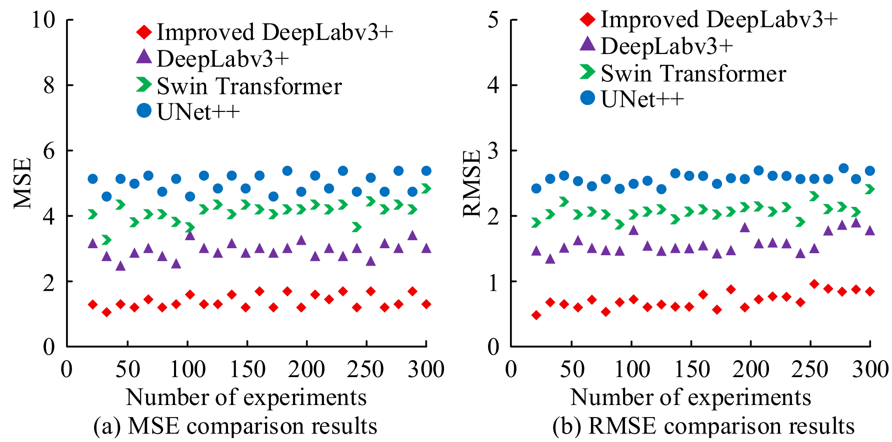

Figure 4. Comparison of MSE value and RMSE value.

improved DeepLabv3+ and all comparison algorithms were statistically significant (*P* <0.05). The MSE and RMSE results demonstrated that the average MSE values of the improved DeepLabv3+ model, DeepLabv3+ model, Swin transformer model, and UNet++ model were 1.7, 2.8, 3.9, and 4.3, respectively, and the improved DeepLabv3+ model had the smallest MSE value (Figure 4a). The improved DeepLabv3+ model had the lowest RMSE value with an average of 1.5, while the RMSE values of the other three models were all greater than 1 (Figure 4b). Statistical analysis showed significant differences in both MSE and RMSE between the improved DeepLabv3+ and compared models (*P* < 0.05). The improved DeepLabv3+ model could optimize the basic DeepLabv3+ and had sufficient advantages.

Classification system application effect analysis

The application effects of the proposed UAV image crop classification system showed that the classification accuracy of improved DeepLabv3+ model for corn, sugarcane, rice, and citrus crop images was 98.4%, 98.6%, 97.9%, and 98.7%, respectively. The results suggested that the proposed model had a better classification effect on actual crop images (Figure 5a). When the unimproved DeepLabv3+ model was used to classify four crops, the classification accuracy was 90.9%, 91.2%, 90.8%, and 92.4%, respectively (Figure 5b). When using the Swin transformer model and the UNet++ model to classify four crop images, the classification accuracy was lower than 90%, which was much lower than that of the improved DeepLabv3+ (Figures 5c and 5d). The improved DeepLabv3+ had higher classification accuracy when classifying actual crop images. All accuracy differences proved statistically significant (*P* < 0.05). The classification time, CPU usage, model size, memory usage, and AUC demonstrated that improved DeepLabv3+ model took 28.2 ms to classify images, while the CPU and the memory usages were 42.8% and 38.2%, respectively, the AUC value was 0.93, the average precision was 0.92, and the model misclassification rate was only 2.4%. The classification time, misclassification rate, CPU usage, and memory usage of other models were higher. The AUC value and average precision value of other models were lower than those of the improved DeepLabv3+ (*P* < 0.05) (Table 1). The improved DeepLabv3+ still had greater advantages than other models when classifying actual crop UAV images.

Discussion

This study verified the performance of the improved DeepLabv3+ model and the UAV image crop classification system. The improved DeepLabv3+ had significant advantages in MIoU, MSE, RMSE, and classification accuracy. In the

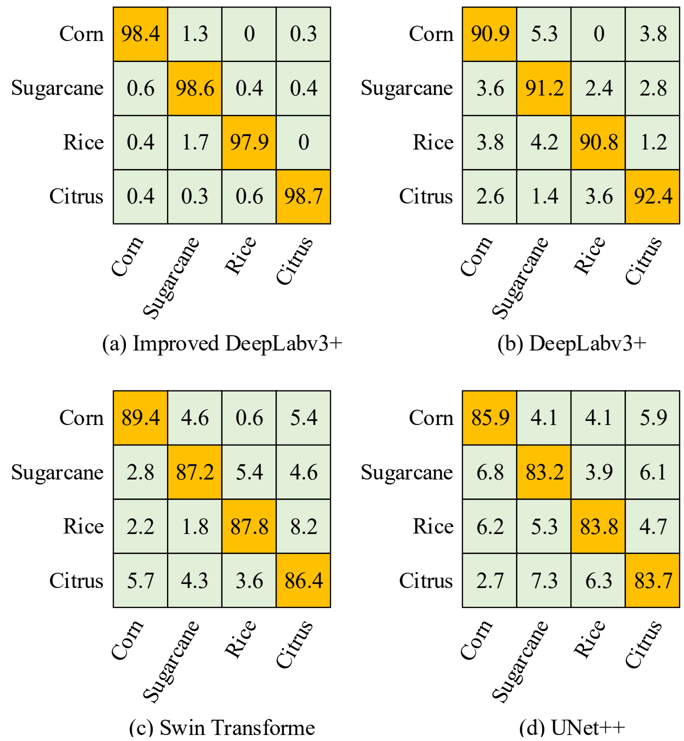

Figure 5. Classification results of four crops by different models.

Table 1. Results of system performance experiments.

Indicator	Improved DeepLabv3+	DeepLabv3+	Swin transformer	UNet++
Classification time	28.2 ms	34.3 ms	40.2 ms	45.3 ms
CPU usage rate	42.8%	47.3%	52.2%	63.1%
Memory usage rate	38.2%	40.3%	46.9%	51.2%
AUC value	0.93	0.91	0.87	0.83
Misclassification rate	2.4%	3.7%	4.8%	5.2%
Average precision	0.92	0.89	0.83	0.81

MIoU comparison, the improved DeepLabv3+ achieved higher value than the DeepLabv3+ model, Swin transformer model, and UNet++ model. These results were similar to the cascaded transformer results proposed by Wang *et al*. who aggregated global and local features through the multi-head SAM to enhance boundary information extraction in remote sensing image semantic segmentation [21]. In the MSE and RMSE comparison, the improved DeepLabv3+ achieved lower results than those of the comparison models, which coincided with the adaptive masking attention network results proposed by Wang *et al*. who demonstrated superior feature discrimination through selective elimination of redundant background information and fine-grained feature modeling for agricultural crop classification from UAV images [22]. In the ablation experiment, the accuracy of the improved DeepLabv3+ was higher than that of other models. The result was consistent with the semantic segmentation strategy proposed by Nuradili *et al*. who verified the effectiveness of multi-source feature fusion through progressive module addition for UAV remote-sensing image analysis [23]. In addition, the improved DeepLabv3+ achieved superior classification accuracy for corn, sugarcane, rice, and citrus, respectively, compared to the unimproved DeepLabv3+, Swin transformer, and UNet++ models. The results were similar to the low-light scene segmentation results proposed by Nuradili *et al*. who achieved robust semantic segmentation performance through deep learning and thermal infrared image features under challenging illumination conditions [24]. In the system application effect analysis, the improved DeepLabv3+ model achieved superior results compared to those of the other models. The result indicated that the improved DeepLabv3+ enhanced the model's adaptability to complex agricultural environments. Furthermore, the proposed system achieved efficient crop classification, low resource consumption, and stable operation performance, indicating that the UAV image crop classification system based on improved DeepLabv3+ accurately and effectively completed crop identification tasks in complex farmland environments.

This research also had limitations that the experiments were conducted under specific climatic conditions in Guangxi, China from September to October, and the performance of the improved DeepLabv3+ in other regions or seasons might be affected by unpredictable factors such as crop growth stages, lighting variations, and weather conditions. Further, the current attention mechanism modules were designed for specific crop types, and their generalization to other crop classification tasks required further validation. Furthermore, the computational complexity of the transformer layers increased with image resolution, which might limit the real-time performance for large-scale farmland monitoring. In future research, multi-regional and multi-seasonal experiments should be conducted to validate the model's effectiveness under diverse environmental conditions. Additionally, the integration of image enhancement algorithms such as the Retinex model should be explored to improve the model's classification performance in complex weather conditions.

Conclusion

This study introduced an improved DeepLabv3+ model to optimize UAV image crop classification. By integrating SA, CA, and Transformer, the model enhanced feature extraction capability, long-distance dependency capture, and segmentation accuracy in complex agricultural environments. The improved DeepLabv3+ effectively improved the classification accuracy and operational efficiency of crop identification, providing a theoretical foundation and practical guidance for research in precision agriculture and UAV remote sensing applications. However, this research had limitations. In practical applications, UAV image acquisition was affected by unpredictable weather conditions such as low light, strong light, and cloudy days, which impacted image clarity and classification performance. Therefore, employing image enhancement algorithms such as Retinex model to improve image quality and combining principal component analysis to reduce dimensionality and noise interference might be the directions for further research. Additionally, multi-regional and multi-seasonal experiments should be conducted to validate the model's effectiveness under diverse environmental conditions and expand its generalization to other crop types.

Acknowledgements

The research was supported by the National Natural Science Foundation of China (Grant No. 62271193), Henan Province Science and Technology Research Project (Project No. 242102220020), Key Research Project of Higher Education Institutions in Henan Province (Project No. 24B470003).

References

1. Zelisko N, Raiter N, Markovych N, Matskiv H, Vasylyna O. 2024. Improving business processes in the agricultural sector considering economic security, digitalization, risks, and artificial intelligence. Ekon APK. 31(3):10-21.

2. Faruk M, Ejarder S, Jannat R, Hossain MI, Rahman MH, Islam MS, *et al*. 2025. Marketing strategies and tactics in the metaverse: A systematic literature review and future research agenda. FSI. 2025:1-17.
3. Phang SK, Chiang THA, Happonen A, Chang MML. 2023. From satellite to UAV-based remote sensing: A review on precision agriculture. IEEE Access. 11(1):127057-127076.
4. Guo W, Gong Z, Gao C, Yue J, Fu Y, Sun H, *et al*. 2024. An accurate monitoring method of peanut southern blight using unmanned aerial vehicle remote sensing. Precis Agric. 25(4):1857-1876.
5. Sadenova MA, Beisekenov NA, Anuarbekov TB, Kapasov AK, Kulenova NA. 2023. Study of unmanned aerial vehicle sensors for practical remote application of earth sensing in agriculture. Chem Eng Trans. 98(1):243-248.
6. Alzhanov A, Nugumanova A. 2024. Crop classification using UAV multispectral images with gray-level co-occurrence matrix features. Procedia Comput Sci. 231(1):734-739.
7. Kalita I, Singh GP, Roy M. 2023. Crop classification using aerial images by analyzing an ensemble of DCNNs under multi-filter and multi-scale framework. Multimed Tools Appl. 82(12):18409-18433.
8. Eladl SG, Haikal AY, Saafan MM, Zaineldin HY. 2024. A proposed plant classification framework for smart agricultural applications using UAV images and artificial intelligence techniques. Alex Eng J. 109(1):466-481.
9. Allu AR, Mesapam S. 2024. Fusion of different multispectral band combinations of Sentinel-2A with UAV imagery for crop classification. J Appl Remote Sens. 18(1):016511.
10. Ajayi OG, Olufade OO. 2023. Drone-based crop type identification with convolutional neural networks: An evaluation of the performance of RESNET architectures. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. 10(1):991-998.
11. Gao X, Chen C, Gan Y, Liu Y. 2024. Comparison of extraction accuracy of sugarcane from different resolution satellite images using DeepLabV3+ mode. Int Arch Photogramm Remote Sens Spat Inf Sci. 48(1):179-184.
12. Cheng L, Xiong R, Wu J, Yan X, Yang C, Zhang Y, *et al*. 2024. Fast segmentation algorithm of USV accessible area based on attention fast DeepLabV3. IEEE Sens J. 24(15):24168-24177.
13. Edpuganti A, Akshaya P, Gouthami J, Sajith Variyar VV, Sowmya V, Sivanpillai R. 2023. Effect of data quality on water body segmentation with deeplabv3+ algorithm. Int Arch Photogramm Remote Sens Spat Inf Sci. 48(1):81-85.
14. Chen H, Qin Y, Liu X, Wang H, Zhao J. 2024. An improved DeepLabv3+ lightweight network for remote-sensing image semantic segmentation. Complex Intell Syst. 10(2):2839-2849.
15. Pal A, Kumar V, Hassan KL, Singh BK. 2025. A framework for leaf disease analysis and estimation using MAML with DeepLabV3. Microsyst Technol. 31(3):715-733.
16. Pan Z, Zhang X, Jiang Y, Li B, Golsanami N, Su H, *et al*. 2025. High-precision segmentation and quantification of tunnel lining crack using an improved DeepLabV3+. Undergr Space. 22:96-109.
17. Pan S, Li J, Jiang J. 2023. A street view semantic segmentation algorithm based on DeeplabV3+ architecture. Proc SPIE. 12717:602-609.
18. Sun Q, Pickett M, Nain AK, Jones L. 2025. Transformer layers as painters. Proc AAAI Conf Artif Intell. 39(24):25219-25227.
19. Yuan Z, Zhang H, Pu L, Ning X, Zhang L, Zhao T, *et al*. 2024. Ditfastattn: Attention compression for diffusion transformer models. Adv Neural Inf Process Syst. 37(1):1196-1219.
20. Wang Y, Wang W, Li Y, Jia Y, Xu Y, Ling Y, *et al*. 2024. An attention mechanism module with spatial perception and channel information interaction. Complex Intell Syst. 10(4):5427-5444.
21. Wang F, Ji J, Wang Y. 2024. Remote sensing image semantic segmentation based on cascaded transformer. IEEE Trans Artif Intell. 5(8):4136-4148.
22. Wang X, Wang D, He Z, Lin Z, Xie S. 2025. AMA-Net: Adaptive masking attention network for agricultural crop classification from UAV images. IEEE Trans AgriFood Electron. 3(1):246-253.
23. Nuradili P, Zhou J, Zhou X, Ma J, Wang Z, Meng L, *et al*. 2023. UAV remote-sensing image semantic segmentation strategy based on thermal infrared and multispectral image features. IEEE J Miniatur Air Space Syst. 4(3):311-319.
24. Nuradili P, Zhou G, Zhou J, Wang Z, Meng Y, Tang W, *et al*. 2024. Semantic segmentation for UAV low-light scenes based on deep learning and thermal infrared image features. Int J Remote Sens. 45(12):4160-4177.