Research on citrus segmentation algorithm based on complex environment

Jia Jun Zhang[1], Peng Chao Zhang[2], Jun Lin Huang[3], Kai Yue[4], Zhi Miao Guo[5]
[1, 2, 3, 4, 5]School of Mechanical Engineering, Shaanxi University of Technology, Hanzhong, 723001, China
[2]Shaanxi Province Key Laboratory of Industrial Automation, Shaanxi University of Technology, Hanzhong, 723001, China
[2]Corresponding author
E-mail: [1]*862320515@qq.com*, [2]*snutzpc@126.com*, [3]*634915323@qq.com*, [4]*15667141329@163.com*, [5]*3207930878@qq.com*

Received 29 February 2024; accepted 9 April 2024; published online 21 April 2024
DOI https://doi.org/10.21595/jmai.2024.24040

Abstract. Aiming to address the low efficiency of current deep learning algorithms for segmenting citrus in complex environments, this paper proposes a study on citrus segmentation algorithms based on a multi-scale attention mechanism. The DeepLab V3+ network model was utilized as the primary framework and enhanced to suit the characteristics of the citrus dataset. In this paper, we will introduce a more sophisticated multi-scale attention mechanism to enhance the neural network's capacity to perceive information at different scales, thus improving the model's performance in handling complex scenes and multi-scale objects. The DeepLab V3+ network addresses the challenges of low segmentation accuracy and inadequate refinement of segmentation edges when segmenting citrus in complex scenes, and the experimental results demonstrate that the improved algorithm in this paper achieves 96.8 % in the performance index of MioU and 98.4 % in the performance index of MPA, which improves the segmentation effectiveness to a significant degree.

Keywords: citrus segmentation, deep learning, DeepLab V3+, attention mechanisms.

1. Introduction

Citrus is one of the most important fruit crops globally and is important to the global economy and food supply. It is important in terms of food supply, economic value, agro-diversity and ecology. However, fruit picking is a low-paying, seasonal and repetitive job with little prospect for growth. Moreover, picking is expensive, with labor costs for picking accounting for up to 50 percent of farmers' operating expenses. There are fewer and fewer workers engaged in fruit picking as the existing picking workers are aging and the younger generation is migrating to urban areas. Timing is crucial in fruit picking. Statistics show that fruit picked two weeks late loses 80 % of its value. Labor shortages cost fruit growers $3 billion in lost sales each year, and fruit growers around the world lose $30 billion in sales annually due to unpicked fruit. To prevent economic losses caused by harvesting delays, it is essential to invest in researching and developing automated harvesting technology.

With the increasing development of deep learning, the application of computer vision has become more widespread. Semantic segmentation in agriculture has started to develop gradually. In soil analysis [1], semantic segmentation algorithms can be utilized to segment and categorize farmland soils in high resolution. This can assist farmers in gaining a better understanding of the texture, water content, and nutrient composition of the soil, thereby guiding soil management practices and enhancing agricultural production; In terms of vegetation detection [2], farmland vegetation data are obtained through drones, satellite images and other technologies, and combined with semantic segmentation to segment and classify the vegetation, enabling the monitoring and analysis of the growth status of the crops, and to help farmers find out the problems of pests, diseases and droughts in a timely manner.; In crop identification [3], different crops in farmland can be identified and classified using semantic segmentation technology to assist farmers

with crop management and harvesting; in farmland planning, fields can be divided into blocks and organized using semantic segmentation technology to aid farmers in the rational layout of planting crops, irrigation and fertilization; In terms of pest and disease detection [4], the use of semantic segmentation technology can detect and analyze pests and diseases in farmland, assisting farmers in taking timely preventive and control measures to minimize losses.

In 2019 C. Senthilkumar et al. [5]. Proposed is an optimal segmentation method for diagnosing citrus leaf disease based on Backpropagation Neural Networks (BPNN), which firstly performs feature extraction of citrus lesions and then utilizes a weighted segmentation approach to segment the test image. In 2020 C. Senthilkumar et al. [6] proposed a Hough transform-based modeling method for citrus disease feature extraction and classification, which consists of preprocessing using bilateral filtering, optimal weighted segmentation (OWS) based, Hough Transform (HT) for feature extraction, and rough fuzzy artificial neural network (RFANN) for classification. In 2021 Akshatha Prabhu et al. [7]. proposed a fruit classification model for identifying citrus fruit defects using a computer vision system. 2022 Mohammed Ahmed Matboli et al. [8]. used a deep learning model-based approach to recognize and classify fruit diseases. In 2023 M. W. Hannan et al. [9]. proposed a machine vision algorithm for citrus fruit recognition. The algorithm includes segmentation, region labeling, dimensional filtering, perimeter extraction and perimeter-based detection. The segmentation above distinguishes between different crops using traditional segmentation methods and deep learning-based methods. However, there are still numerous challenges in the semantic segmentation of crops:

(1) Semantic segmentation requires a large amount of accurately labeled data to train models, but obtaining labeled data for crops is often challenging.

(2) The farmland environment is complex and variable, including factors such as vegetation cover, soil type, and lighting conditions, which can affect the quality and difficulty of the images and lead to a decrease in semantic segmentation accuracy.

(3) Existing semantic segmentation models tend to perform well in specific scenarios, but generalize poorly to other farmland environments and fail to adapt to different crop species and growth stages.

Based on this premise, this paper conducts experiments on homemade datasets and makes algorithmic improvements for citrus segmentation in complex environments on the basis of the existing model, using DeepLab V3+ network model as the primary structure and introducing a more intricate multi-scale attention mechanism to enhance the neural network's capacity to perceive information across various scales.

2. Data set production and evaluation indicators

2.1. Raw data set production

The dataset used in this paper is the citrus dataset, which was collected in the field at one of the orange groves in Chenggu County, Hanzhong City, Shaanxi Province. The sample images contain various citrus images, including multiple target images, target occlusion images, and single target images, captured under different weather conditions. A total of 1,000 citrus images were collected to meet the thesis research requirements. To facilitate dataset loading for the algorithm, the collected citrus images underwent sample screening. Initially, the photographed images were cropped to fit within the specified range, and then batch-named using Python scripts. Finally, 627 datasets were obtained for this experiment, and the dataset was named VOC_Orange. The ratio of 8:2 is used to divide the training set and test set. The final division results in 500 pictures in the training set and 127 pictures in the test set. Finally, the Labelme annotation tool is used to label the targets in the pictures. Fig. 1 shows a sample of the original graph of the collected dataset, and Fig. 2 shows a sample of data annotation using the Labelme annotation tool.

Fig. 1. Sample original map of the citrus dataset

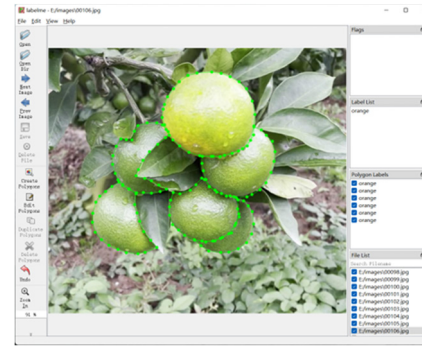
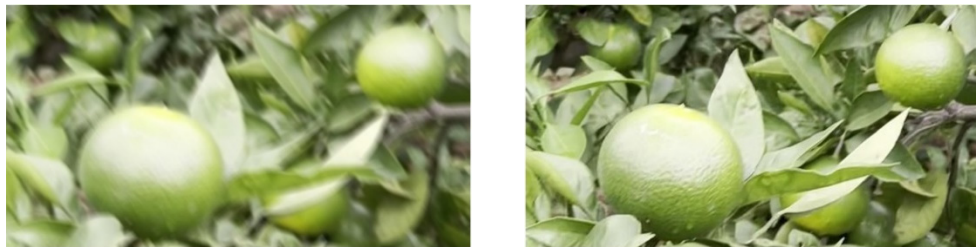


Fig. 2. Sample citrus dataset labeling

2.2. Data set preprocessing

In the photographed citrus dataset, due to the natural environment, there are many images where the target data is not obvious, including rainy day images, foggy day images, images with excessive light, and images with insufficient light. These factors reduce the algorithm's ability to generalize the model when processing images, leading to a significant decrease in the model's robustness and other related issues. Therefore, augmenting the dataset before processing it with the algorithm can effectively alleviate the imbalance problem of the data and improve the robustness and detection accuracy of the model.

(1) Motion blur processing: While capturing the citrus dataset, one of the frames extracted in the camera's motion state exhibits motion blur. This blurriness compromises the accuracy of the algorithm during dataset processing due to the indistinct target. Therefore, the Generative Adversarial Network DeblurGAN method [10] is used to eliminate motion blur from the image. The images before and after processing are shown in Fig. 3, where (a) represents the pre-processing image and (b) represents the post-processing image.



a) b)

Fig. 3. Motion fuzzy dataset before and after processing

(2) Defogging: Images captured under a foggy sky may experience issues like loss of semantic information and difficulty in extracting features. Therefore, the DRSDehazeNet algorithm [11] is used to de-fog the images. The images before and after processing are shown in Fig. 4, where Fig. 4(a) represents the pre-processing image and Fig. 4(b) represents the post-processing image.

(3) De-icing: The acquired dataset contains images with water droplets or raindrops, and the effect of these droplets needs to be removed to obtain a clear image. Therefore, the RCDNet algorithm [12] is used to de-rain the images. The images before and after processing are shown in Fig. 5, where Fig. 5(a) represents the pre-processing image and Fig. 5(b) represents the post-processing image.

(4) Brightness Enhancement Process: To address the presence of images in the dataset affected by low light and backlit conditions, a brightness enhancement process is applied to these datasets

using a Python script. The images before and after processing are shown in Fig. 6, where Fig. 6(a) represents the pre-processing image and Fig. 6(b) represents the post-processing image.

a) b)

Fig. 4. Comparison of the foggy day dataset before and after processing

a) b)

Fig. 5. Comparison of the rainy day dataset before and after processing

a) b)

Fig. 6. Low light dataset before and after processing

a) b)

Fig. 7. Comparison of the data set before and after color saturation processing

(5) Color Saturation Processing: By increasing the color saturation, the image can appear more vibrant and vivid, enhancing the visual effect. The robustness and generalization ability of the model can be enhanced. The images before and after processing are shown in Fig. 7, where Fig. 7(a) represents the pre-processing image and Fig. 7(b) represents the post-processing image.

2.3. Evaluation indicators

The following metrics are commonly used to evaluate the performance of semantic segmentation models:

1. Pixel Accuracy: That is, the ratio of the number of correctly classified pixels to the total number of pixels is an indicator for evaluating the overall classification accuracy of the model. However, this metric is susceptible to category imbalance and is not applicable to datasets with large differences in the number of category samples. The dataset category used in this paper is a single target, so it can be calculated using this evaluation index. The specific process is shown in Eq. (1):

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}}. \tag{1}$$

2. Mean Intersection over Union: The ratio of the intersection and concatenation of the predicted results to the real results is calculated, and the results of each category are averaged. MIoU is one of the most commonly used semantic segmentation evaluation metrics, which has good robustness and stability. The specific process is shown in Eq. (2):

$$MI\text{oU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}. \tag{2}$$

3. Algorithm design and improvement

3.1. Algorithm design ideas

Deeplabv3+ is the latest version of the Deeplab series [13], and the network structure is shown in Fig. 8. It improves on Deeplabv3 to achieve superior performance and increased efficiency. The key improvements of Deeplabv3+ are outlined below:

– Feature extraction using a fully convolutional network avoids the loss of resolution caused by pooling operations.

– Up-sampling using a Decoder allows the output resolution to reach the original size of the input image.

– The ASPP (Atrous Spatial Pyramid Pooling) module is used in the decoder, which is able to capture information at different scales and improve the accuracy of the model.

– Depthwise Separable Convolution is introduced to reduce the number of parameters and computation of the model, which improves the efficiency of the model.

In this paper, the DeepLab V3+ network model is utilized as the main structure, which is improved for the characteristics of citrus dataset. In order to solve the problems of low segmentation accuracy and inadequate refinement of segmentation edges when segmenting citrus in complex scenes using the DeepLab V3+ network the main reason for introducing a more intricate multi-scale attention mechanism in this chapter is to enhance the neural network's capacity to perceive information at different scales, and thus to improve the model's performance in handling complex scenes and multi-scale objects. The specific reasons are as follows:

(1) Multi-scale object processing: Complex scenes often contain objects of varying sizes, with some being small and others large. Simple neural networks may not be able to effectively capture and process such multi-scale information. Introducing a more complex multi-scale information

fusion mechanism can help the network better adapt to objects of different scales.

(2) Contextual information: For pixel-level tasks such as semantic segmentation, contextual information is crucial for accurately categorizing each pixel. By introducing an attention mechanism, the model can focus on the context around each pixel as it is processed, contributing to more accurate semantic segmentation.

(3) Mitigating information loss due to pooling: In neural networks, pooling operations typically result in a decrease in resolution, leading to information loss. By using an attention mechanism or multi-scale information fusion, important information can be better preserved and integrated after pooling operations.

(4) Improved perceptual range: The multi-scale attention mechanism expands the perceptual range of the model, enabling it to focus on both local details and global context. This is critical for understanding the overall structure and relationships in an image.

(5) Suppressing irrelevant information: In complex scenarios, there may be a significant amount of irrelevant information. By introducing the attention mechanism, the model can learn to suppress information that is irrelevant to the task, thus enhancing the robustness and generalization performance of the model.

(6) Adaptation to change: Objects in the scene may appear at different scales, and the relative scales between them may vary with camera distance, viewing angle, and other factors. Introducing a multi-scale attention mechanism can enhance the model's adaptability and improve its performance across various scales.

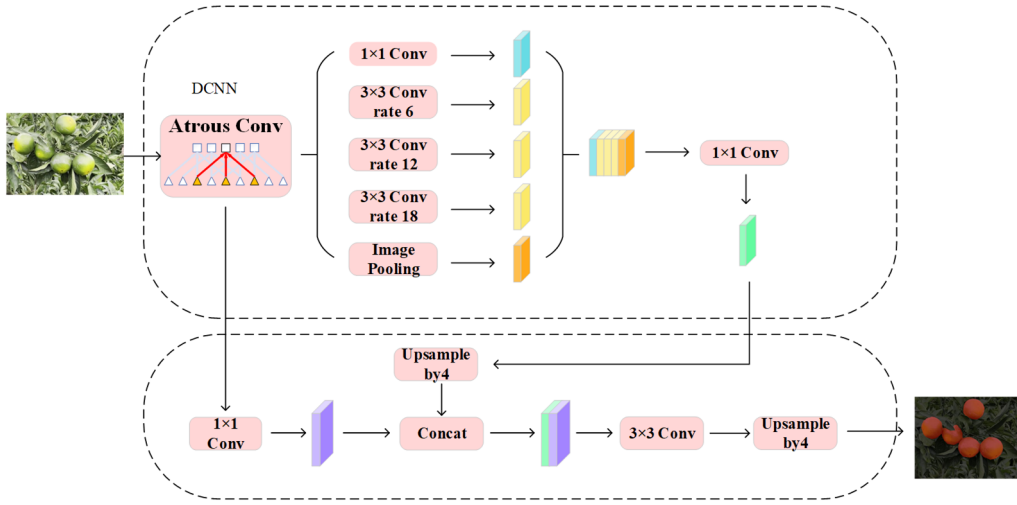


Fig. 8. Deeplabv3+ network architecture diagram

3.2. CBAM efficient attention module improvement

Convolutional Block Attention Module (CBAM) [14] is a convolutional neural network model based on the attention mechanism as shown in Fig. 9. Researchers at KAIST (Korea Advanced Institute of Science and Technology) proposed enhancing the learning ability of the convolutional neural network model for various regions in an image. The CBAM model comprises two sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). These modules are utilized to assign weights to the channel and spatial locations of the input feature maps, respectively. The channel attention module adjusts the weight of each channel's contribution to feature extraction, while the spatial attention module weights different regions of the input feature map to make the model focus more on important regions.

However, the citrus growth environment is quite complex. Rainy weather, insufficient light, and other factors can significantly impact the network’s ability to recognize citrus segmentation.

Therefore, by enhancing the structure of the CBAM module, the network's performance in complex environments can be further improved, leading to increased efficiency in citrus identification. The improved module was named CBAM module adapted for citrus segmentation, or Ado-CBAM for short.

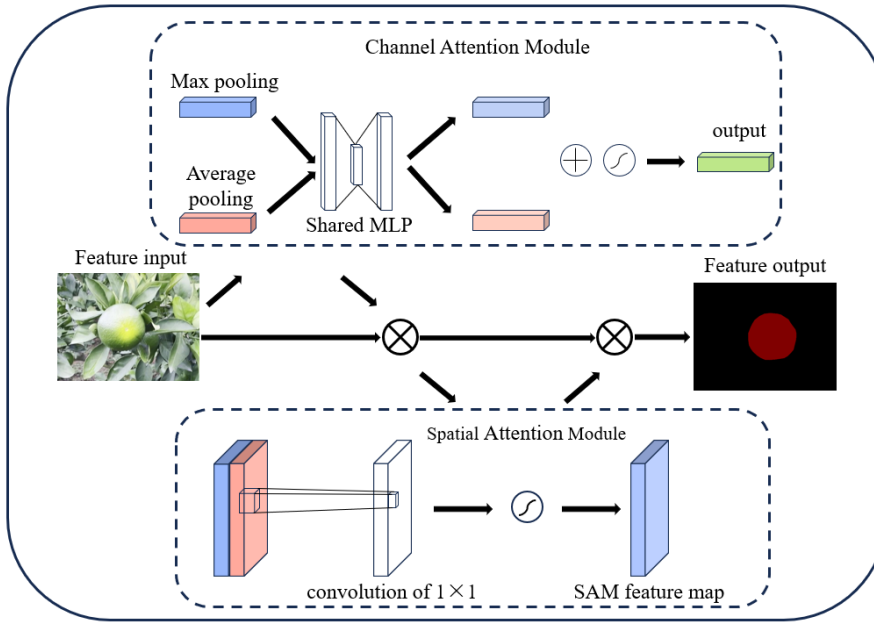


Fig. 9. Convolutional block attention module network diagram

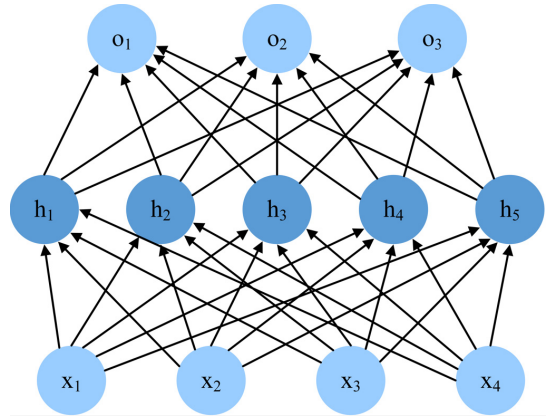


Fig. 10. Multilayer perceptron structure diagram

(1) A multilayer perceptron structure is incorporated into the attention module to enhance the depth and complexity of the network, enabling better capture of the correlations between channels. Multi-Layer Perceptron [15] (MLP) is a common artificial neural network structure. The structure is shown in Fig. 10. Consists of multiple fully connected layers (also known as dense layers), each containing multiple neurons. The basic components of the MLP structure include an input layer, a hidden layer, and an output layer. The hidden layer can contain multiple layers and each layer consists of multiple neurons. In each neuron, the input signal is weighted and summed and then nonlinearly transformed by an activation function to obtain the output of the neuron. The output of the hidden layer is used as input to the next layer until the final output layer receives the final

prediction. The local structures of the improved channel attention module and spatial attention module are shown in Fig. 11 and Fig. 12.

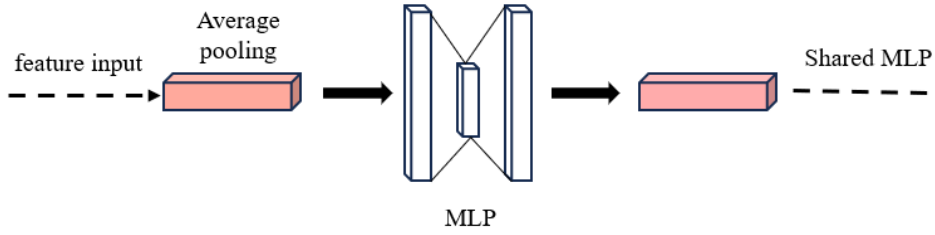


Fig. 11. Local structure of the improved channel attention module

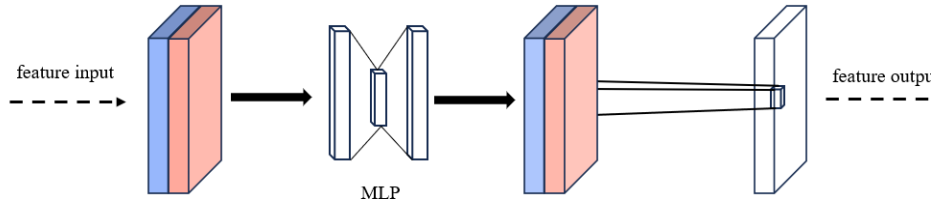


Fig. 12. Localized structure of the improved spatial attention module

(2) Multi-scale feature fusion. The main idea of the improvement in multi-scale feature fusion [16] in CBAM is that by incorporating multi-scale information in the channel attention module and spatial attention module, the network can focus on features of various scales simultaneously, thereby enhancing the network's ability to express features. Specifically, the channel attention module in CBAM is used to weigh the features across different channels to enhance the network's focus on channel-specific information. The spatial attention module is used to weight features at different spatial locations, enhancing the network's focus on spatial information. By incorporating multi-scale information into these two modules, the network can simultaneously focus on features at various scales to more effectively capture both global and local information in the image. This is done as follows:

– Channel Attention Module: First, global average pooling is performed on the input feature maps to obtain the global average features on the channel dimensions. Then, the global average features are processed through two fully connected layers (FC) separately to obtain two channel attention distribution vectors: one for enhancing the importance of the feature channels and one for weakening the importance of the feature channels. Second, these two channel attention distribution vectors are multiplied together to obtain the final channel attention weights. Finally, the channel attention weights are multiplied by the input feature maps to obtain the weighted feature maps.

– Spatial Attention Module: Initially, the input feature map undergoes processing in two branches: one for maximum pooling and the other for average pooling. Subsequently, the outcomes of these two branches are combined and processed through a convolutional layer to derive the spatial attention weights. Ultimately, these spatial attention weights are used to multiply the input feature map, resulting in the weighted feature map.

– The weighted feature maps obtained through the channel attention module and the spatial attention module are summed or concatenated.

3.3. Adaptive adjustment

Due to the complexity of the citrus growing environment, which results in a low recognition rate of citrus with severe leaf occlusion during segmentation, the CBAM module is implemented to achieve adaptive adjustment through the channel attention mechanism and the spatial attention mechanism[17], which can dynamically adjust the attention weights based on the content and

features of the input image, enabling better capture of important information in the image. The specific operation is as follows:

(1) Channel Attention Module: In the channel attention module, the channel attention weights are obtained through global average pooling and fully connected layers. The importance of each channel is adaptively adjusted according to the content of the input feature map through learning. Subsequently, the channel attention weights are dynamically adjusted based on the importance of each channel in the input feature map to better capture the key information of the image features. Finally, the channel attention weights are multiplied by the original feature map to adjust the contribution of each channel.

(2) Spatial Attention Module: In the spatial attention module, the spatial attention weights are obtained through maximum pooling and average pooling. The significance of various spatial locations is adaptively adjusted based on the content of the input feature map through learning. Subsequently, the spatial attention weights assist the network in concentrating more effectively on the location information of different citrus fruits in the image, thereby enhancing the comprehension of the image's structure. Ultimately, the spatial attention weights are multiplied by the original feature map to regulate the impact of each spatial location.

4. Experimental results and analysis

4.1. Experimental training environment

In this paper, the homemade citrus dataset is selected for experimentation, and 627 datasets, using the ratio of 8:2 to divide the training set and test set, and the final division of the training set of 500 and the test set of 127. The experimental training environment used NVIDIA Titanxp graphics card, NVIDIA Titanxp graphics card (CPU), Ubuntu 16.04 system, Python version 3.7, Pytorch framework, and programming environment of Tensorflow framework. During model training, the initial learning rate for network training was set to 0.001, the training batch size was 8, and Epoch was set to 100.

4.2. Network comparison experiment

The experiments in this section utilize the Ado-CBAM Efficient Attention Module as a fixed module. The DeeplabV3+ network is chosen for comparison with four classical networks: HrNet [18], U-Net [19], PSPNet [20], and SegNet [21], and the Mean Intersection-over-Union Ratio (MioU) and Pixel Accuracy (PA) are selected as the primary evaluation metrics, while Precision and Recall serve as the secondary metrics. After experimental comparison, the effect diagram of each network combined with Ado-CBAM efficient attention module is shown in Fig. 13. In the figure, Fig. 13(a) represents the original diagram of the citrus growing environment; Fig. 13(b) shows the calibration diagram of dataset processing; Fig. 13(c) displays the effect diagram of the HRNet network; Fig. 13(d) illustrates the effect diagram of the U-Net network; Fig. 13(e) demonstrates the effect diagram of the PSPNet network; Fig. 13(f) exhibits the effect diagram of the SegNet network; and Fig. 13(g) showcases the effect diagram of the improved network. The yellow box highlights the visual performance difference of the algorithm.

Table 1 presents the results of the experimental data comparison. It is evident from the table that the DeeplabV3+ algorithm attains 96.8 % in the MioU performance metric, which is 2.6 % higher than the second-highest MioU achieved by the U-net network. Additionally, it achieves 98.4 % in the MPA performance metric, which is 1.3 % higher than the second-highest MPA achieved by the U-Net network. It is on par with the HRNet network and the U-Net network in the performance metric precision, and on par with the U-Net network in the performance metric recall. Therefore, based on the comparison of experimental effect graphs and experimental data, under identical conditions, the enhanced Ado-CBAM high-efficiency attention module demonstrates superior detail effects when integrated with the DeeplabV3+ network. The

segmentation edges are more refined, leading to more accurate citrus segmentation in complex environments and with small targets.

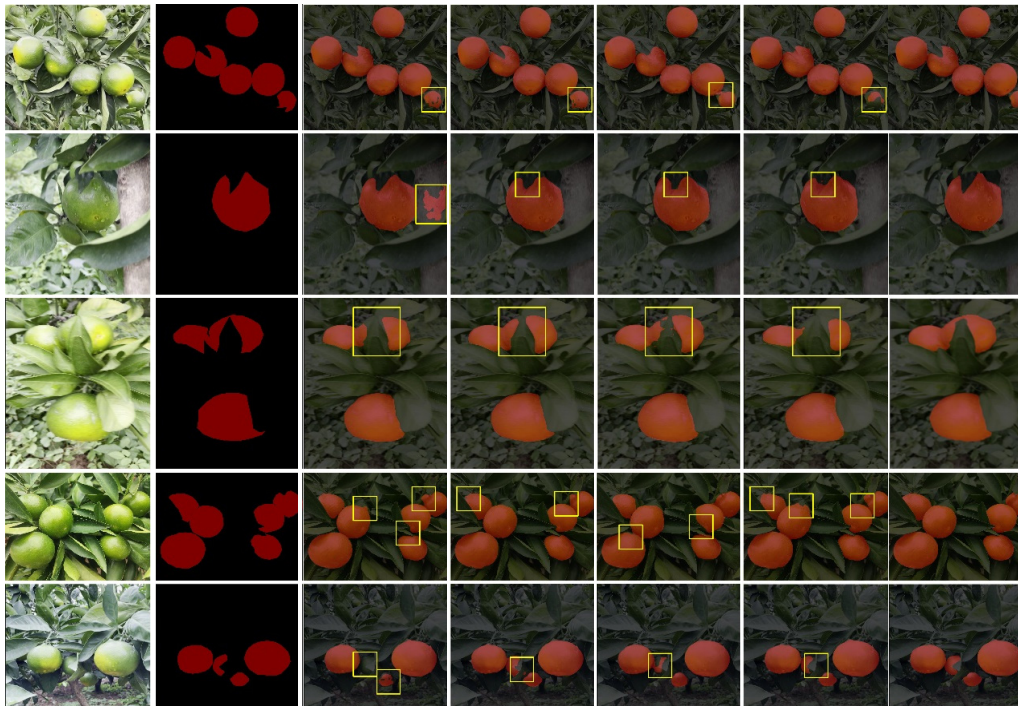


Fig. 13. Comparison chart of the effect of each network experiment

Table 1. Comparison results of experimental data for each algorithm

Algorithm	MioU (%)	MPA (%)	Precision (%)	Recall (%)
hrnet	92.2 %	96.3 %	96 %	96 %
U-net	94.2 %	97.1 %	96 %	98 %
pspnet	91.4 %	96.5 %	95 %	96 %
segnet	92.3 %	97.0 %	95 %	97 %
DeeplabV3+	96.8 %	98.4 %	96 %	98 %

4.3. Comparative tests of attentional mechanisms

In this section of experiments, the DeeplabV3+ network is used as the fixed module, and the Ado-CBAM efficient attention module is selected for comparison with four classical attention mechanisms, namely CBAM, SE [22], CA [23] and ECA [24]. Mean Intersection over Union (MioU) and Pixel Accuracy (PA) are selected as the primary evaluation metrics, while Precision and Recall are considered as secondary evaluation metrics. After experimental comparison, the impact diagram of DeeplabV3+ network combined with various efficient attention modules is illustrated in Fig. 14, in which Fig. 14(a) represents the original diagram of the citrus growing environment; Fig. 14(b) shows the calibration diagram of dataset processing; Fig. 14(c) displays the impact diagram of CBAM; (d) exhibits the impact diagram of SE; Fig. 14(e) demonstrates the impact diagram of CA; Fig. 14(f) illustrates the impact diagram of ECA; and Fig. 14(g) presents the impact diagram of improved Ado-CBAM. The yellow box highlights the visual performance difference.

Table 2 presents the comparison results of the experimental data. It is evident that the Ado-CBAM module achieves a performance index of 96.8 % in MioU, which is 2.2 % higher than

the MioU of the second-highest ECA module. Additionally, it achieves a performance index of 98.4 % in MPA, which is 0.7 % higher than the MPA of the second-highest CA module. It is on par with the CBAM module in terms of precision and improves by 1 % in terms of recall compared to the second-highest SE module. Therefore, based on the experimental effect graphs and comparisons of experimental data, the enhanced Ado-CBAM high-efficiency attention module demonstrates greater accuracy in segmenting large-area occluded targets and small targets when combined with the DeeplabV3+ network under identical conditions.

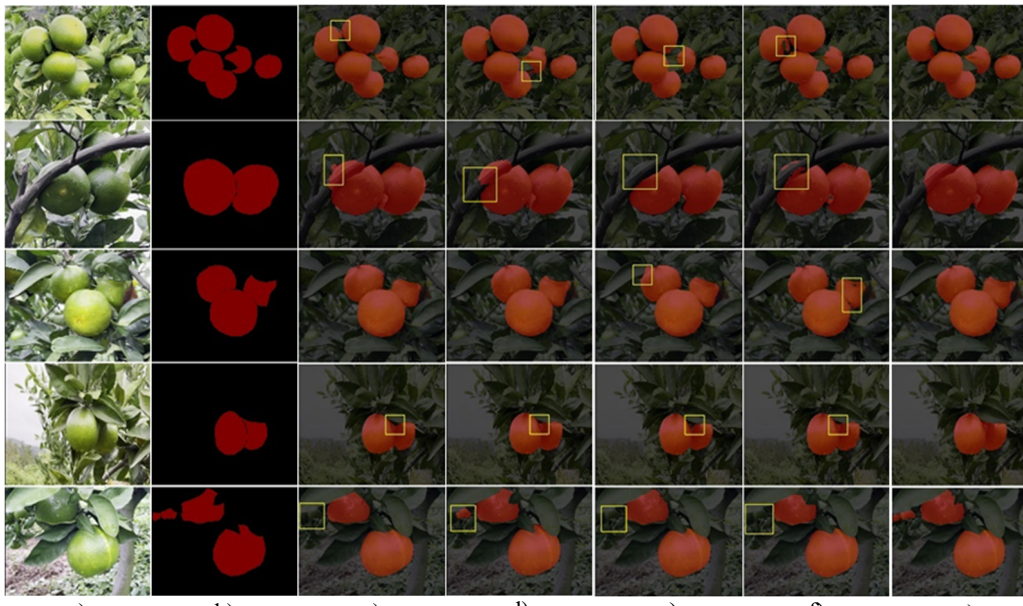


Fig. 14. Comparison chart of the experimental effects of each attention module

Table 2. Comparison results of experimental data of each attention module

Algorithm	MioU	MPA	Precision	Recall
CBAM	0.931	0.956	0.96	0.96
SE	0.943	0.962	0.95	0.97
CA	0.933	0.967	0.95	0.96
ECA	0.946	0.966	0.95	0.96
Ado-CBAM	0.968	0.984	0.96	0.98

5. Conclusions

In this paper, an algorithmic study of citrus segmentation in complex environments is conducted. Firstly, the DeeplabV3+ algorithm and CBAM attention mechanism module are selected based on the algorithmic design ideas compiled. Secondly, improvements are made according to the characteristics of the homemade citrus dataset used in this chapter, focusing on the two parts of the CBAM module: the channel attention module and the spatial attention module. Subsequently, Multi-scale feature fusion and adaptive adjustments are applied to the CBAM module to enable accurate segmentation of citrus. Finally, comparison experiments are conducted to evaluate different algorithms fused with the improved Ado-CBAM module and to compare the DeeplabV3+ algorithm fused with various attention mechanism modules. After experimental verification, the enhanced algorithm presented in this paper demonstrates superior performance in citrus segmentation.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Jia-Jun Zhang is primarily responsible for enhancing the algorithm, conducting experimental training of the model, planning the research content, and writing the paper. Peng-Chao Zhang is primarily responsible for the research findings and paper revisions. Jun-Lin Huang is responsible for resolving errors that occur during the execution of the algorithm. Yue Kai and Zhi-Miao Guo are responsible for shooting the dataset and preprocessing the images.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "Tomato leaf segmentation algorithms for mobile phone applications using deep learning," *Computers and Electronics in Agriculture*, Vol. 178, p. 105788, Nov. 2020, https://doi.org/10.1016/j.compag.2020.105788

[2] R. Ballesteros, D. S. Intrigliolo, J. F. Ortega, J. M. Ramírez-Cuesta, I. Buesa, and M. A. Moreno, "Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques," *Precision Agriculture*, Vol. 21, No. 6, pp. 1242–1262, May 2020, https://doi.org/10.1007/s11119-020-09717-3

[3] J. Ma et al., "Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation," *Computers and Electronics in Agriculture*, Vol. 176, p. 105662, Sep. 2020, https://doi.org/10.1016/j.compag.2020.105662

[4] O. Mzoughi and I. Yahiaoui, "Deep learning-based segmentation for disease identification," *Ecological Informatics*, Vol. 75, p. 102000, Jul. 2023, https://doi.org/10.1016/j.ecoinf.2023.102000

[5] C. Senthilkumar and M. Kamarasan, "Optimal segmentation with back-propagation neural network (BPNN) based citrus leaf disease diagnosis," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 78–82, Nov. 2019, https://doi.org/10.1109/icssit46314.2019.8987749

[6] C. Senthilkumar and M. Kamarasan, "An optimal weighted segmentation with Hough transform based feature extraction and classification model for citrus disease," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 215–220, Feb. 2020, https://doi.org/10.1109/icict48043.2020.9112530

[7] A. Prabhu, L. S., and S. K. V., "Identification of citrus fruit defect using computer vision system," in *2021 2nd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 1264–1270, Aug. 2021, https://doi.org/10.1109/icesc51422.2021.9532834

[8] M. A. Matboli and A. Atia, "Fruit disease's identification and classification using deep learning model," in *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pp. 432–437, May 2022, https://doi.org/10.1109/miucc55081.2022.9781688

[9] M. W. Hannan, T. F. Burks, and D. M. A. Bulanon, "A machine vision algorithm combining adaptive segmentation and shape analysis for orange fruit detection," *Agricultural Engineering International: the CIGR Journal*, Vol. 11, p. 1281, 2009.

[10] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8183–8192, Apr. 2018.

[11] F. Du, P.-Q. Jiang, S.-X. Song, and H.-Y. Xia, "Single-image defogging algorithm based on attention mechanism," *Advances in Lasers and Optoelectronics*, Vol. 60, No. 2, pp. 156–162, 2023.

[12] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3103–3112, Jun. 2020, https://doi.org/10.1109/cvpr42600.2020.00317

[13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision – ECCV 2018*, pp. 833–851, Oct. 2018, https://doi.org/10.1007/978-3-030-01234-2_49

[14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Computer Vision – ECCV 2018*, pp. 3–19, Oct. 2018, https://doi.org/10.1007/978-3-030-01234-2_1

[15] R. Kruse, S. Mostaghim, C. Borgelt, C. Braune, and M. Steinbrecher, "Computational intelligence: a methodological introduction," in *Texts in Computer Science*, Cham: Springer International Publishing, 2022, pp. 53–124, https://doi.org/10.1007/978-3-030-42227-1

[16] H. Zhang, S. Li, and J. Wang, "Multi-scale feature fusion: learning better semantic segmentation for road pothole detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[17] Wu, L., Zhang, Q., Li, and Y., "Adaptive adjustment in dynamic environments," *Journal of Adaptive and Dynamic Systems*, Vol. 10, No. 3, pp. 245–260, 2018.

[18] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 10, pp. 3349–3364, Mar. 2020.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28

[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, Jul. 2017, https://doi.org/10.1109/cvpr.2017.660

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481–2495, Dec. 2017, https://doi.org/10.1109/tpami.2016.2644615

[22] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018, https://doi.org/10.1109/cvpr.2018.00745

[23] J. Fu et al., "Dual attention network for scene segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, https://doi.org/10.1109/cvpr.2019.00326

[24] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, https://doi.org/10.1109/cvpr42600.2020.01155

Jun-Jia Zhang is a Master's degree student in Shaanxi University of Technology. Mainly engaged in computer vision semantic segmentation research.

Peng-Chao Zhang is Professor in Shaanxi University of Technology. Mainly engaged in robotics and control Engineering technology research.

Jun-Lin Huang is a Master's degree student, in Shaanxi University of Technology. Mainly engaged in computer vision semantic segmentation research.

Kai Yue is a Master's degree student, in Shaanxi University of Technology. Mainly engaged in computer vision semantic segmentation research.

Zhi-Miao Guo is a Master's degree student, in Shaanxi University of Technology. Mainly engaged in research on robot perception and control.