

Original papers

Research on wheat broken rate and impurity rate detection method based on DeepLab-EDA model and system construction



ZHANG Qi, WANG Ling*, NI Xindong, WANG Faming, CHEN Du, WANG Shumao

College of Engineering, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Keywords:

Wheat
Broken rate and impurity rate
DeepLab-EDA model
Semantic segmentation algorithm
Combine harvester

ABSTRACT

The broken rate and impurity rate of wheat are important indicators for assessing the quality of combine harvester operations. In view of the overlapping, occlusion and dense adhesion between the scattered grains during the operation of the combine harvester, it is difficult to obtain the grain crushing characteristics and impurity mass, which leads to low detection accuracy. In this paper, a method for detecting wheat broken rate and impurity rate based on DeepLab-EDA semantic segmentation model was proposed, and a detection system was built. In the detection system, an image acquisition device was designed and developed based on the principle of electromagnetic vibration, and the deep learning model was deployed in the embedded processor. Through the human-computer interaction interface design, the online processing and analysis of wheat image data and the display of the detection results of broken rate and impurity rate were realized. Comparative experiments with traditional semantic segmentation models showed that the MIoU, MP and MR of the DeepLab-EDA model were 89.41%, 95.97% and 94.83%, respectively, representing improvements of 9.94%, 7.41%, and 7.52% over the baseline model, and indicating a significant enhancement in the accurate identification and segmentation of broken grain and impurities. Based on this, indoor group matching experiments were conducted with three groups of broken rate and impurity rate levels set at 0.5%, 1.5%, and 2.5%, showing the average errors of 7.54% and 6.30% for broken rate and impurity rate detection systems, respectively. Furthermore, the detection device was installed under the grain outlet of the GM80 combine harvester for field experiments, which showed average errors of 13.32% and 9.77% for wheat broken rate and impurity rate, respectively. The effectiveness and accuracy of the wheat broken rate and impurity rate detection system meet the requirements, which can provide a data basis for intelligent control of combine harvester operation parameters by the operator.

1. Introduction

The broken rate and impurity rate of wheat are important indicators for assessing the quality of combine harvester operations. The monitoring of broken rate and impurity rate during the process of combine harvesting is helpful for the operator to understand the operation situation in time and make intelligent adjustments to the operation parameters, which is of great significance to improve the quality of harvesting operations and ensure food security (Zhang, 2021; Zhang, 2021).

In developed countries like Europe and the United States, intelligent combine harvesters are equipped with fully automatic harvesting control systems. These systems use industrial cameras installed inside the combine harvester's elevator to capture images of the harvested material and utilize image segmentation technology to identify broken grains

and impurities. Operators can monitor various operating parameters of the combine harvester and crop quality characteristics such as moisture content, broken rate, and impurity rate through a touch-screen monitor in the cabin, which allows them to make timely and accurate adjustments, significantly improving the quality and efficiency of harvesting operations. However, in China, the detection of wheat broken rate and impurity rate still relies mainly on manual methods, which are time-consuming, labor-intensive, and poor repeatability. Moreover, these manual methods require the machine to be stopped for testing, making it impossible to reflect the actual operation situation of the combine harvester in a timely manner, which lead to significant fluctuations in the quality of harvesting and fails to meet the demands of field operations (Lu, 2022; Chen et al., 2022).

In recent years, with the widespread application of machine vision technology and deep learning methods in agricultural product quality

* Corresponding author.

E-mail address: wangling.0928@163.com (W. Ling).

inspection, yield estimation, and grain identification (Liu et al., 2022; Ma et al., 2023; Rong et al., 2017), researchers have proposed methods for detecting broken grains and impurities in grains based on traditional digital image processing. Its main research contents are feature selection, threshold setting and target adhesion. It has the advantages of small calculation and fast detection speed. However, this method heavily relies on image quality and the analyst's understanding of target features. In addition, it needs to go through tedious trial and error combination, and the generalization ability and robustness of different scenes are poor (Ramirez-Paredes and Hernandez-Belmonte, 2020; Basati et al., 2018; Chen et al., 2020). The segmentation algorithm based on deep learning has strong learning ability for abstract features and fine segmentation ability at the pixel level, making it flexible to adapt to the numerous complex issues in crop quality inspection (Ni et al., 2023; Lin et al., 2022; Zhu et al., 2023). Guan et al. (Guan et al., 2022) proposed an impurity recognition algorithm based on Mask R-CNN to overcome misidentification caused by similar colors and shapes. Then the conversion between the number of impurity pixels and the actual mass was achieved based on the pixel density calibration test and the impurity rate correction factor. Han et al. (Chen et al., 2020) proposed a rice broken rate and impurity rate online detection system based on the U-Net semantic segmentation model. During the detection process, the U-Net model was first used to segment impurities and grains, followed by further screening of broken grains using HSV three-channel thresholds and geometric features. Finally, the corresponding target pixels are substituted into the pixel-mass model to calculate the corresponding broken rate and impurity rate. Chen et al. (Chen et al., 2022) designed a wheat impurity recognition and segmentation visual system based on the DeepLabV3 + semantic segmentation model to achieve online detection of impurity rates during mechanized wheat harvesting. They found that using ResNet-50 as the backbone network yielded the best impurity rate detection results. And they also construct the pixel-mass model to calculate the impurity rate. In summary, the construction of deep learning segmentation algorithm and pixel-mass model have been widely used in grain quality detection, and achieved certain effectiveness. However, in the complex agricultural production environment, there are different degrees of overlapping occlusion and dense adhesion between grains, resulting in low detection accuracy of segmentation algorithm.

In view of the above problems, relevant researchers have tried to solve them from two aspects: algorithm and detection device. In the aspect of algorithm, Wu et al. (Wu et al., 2024) introduced instance segmentation algorithm into the recognition of broken grains and impurities to realize segmentation and counting at the same time, and then solved the problem of inaccurate calculation of grain broken rate caused by occlusion through the fusion algorithm of thousand-grain weight and pixel-mass model. Liu et al. (Liu et al., 2022) introduced a CPU-Net semantic segmentation model specifically for maize impurity detection. The model improves the difficulty of feature extraction caused by grain occlusion through attention module and pyramid module. Chen et al. (Chen et al., 2022) collected the image sequence during the falling process of soybean seeds, and used SVM combined with multi-view shape features to classify the seeds completely and broken. In terms of detection devices, in order to realize the single layer of grains and reduce the occlusion between grains, the above-mentioned related studies are mostly based on conveyor belt structure or scraper structure. However, in practical applications, the feeding amount of grains and the speed of conveyor belt are difficult to cooperate to form local accumulation of grains, while the scraper structure is very easy to cause blockage. Meanwhile, the application of electromagnetic vibrator in agriculture has led to new thinking on the design scheme of detection device in this study. The electromagnetic vibrator has the transportation function of conveyor belt, and the adjustability of vibration frequency and amplitude makes it have more space for expansion. For example, Wang et al. (Wang et al., 2015) proposed a method to realize the directional arrangement and transportation of corn seeds according to the feeding

principle of the electromagnetic vibrator. By setting the working parameters such as the vibration frequency, the attitude control of the seeds is realized, and the seeding quality of the seeder is greatly improved. Zhang et al. (Wang et al., 2023) realized the uniform flattening of wheat grains and high-throughput detection of the number of grains by debugging the parameters such as vibration frequency and amplitude of the electromagnetic vibrator. Xia et al. (Xia et al., 2020) designed a guided vibration seed supply device with a Y-shaped guide groove. By analyzing the motion characteristics of the seeds on the vibration-guided seed-filling plate, the problem of high multi-grain rate of flat eggplant seeds under the vibration supply was solved. It can be seen that the single-layer motion control of grains can be realized by debugging the working parameters of the electromagnetic vibrator, which is more conducive to the high-throughput detection of large quantities of grains. This is very important to improve the detection effect, because other related studies are mostly aimed at the low-throughput detection of single target of broken grains or impurities in a small number of samples. The number of grains in a single sampling is only about 200, approximately 7.5 g, while the artificial multi-point sampling detection method usually requires a single sampling of not less than 100 g. The difference in sampling quality between the two leads to poor representativeness of sampling results and low detection efficiency. In the context of this problem, the detection of large quantities of wheat grains is also accompanied by the increase of image size and the aggravation of overlapping occlusion, and the related algorithms are no longer applicable.

Synthesize the above analysis, to achieve rapid identification of broken wheat grains and impurities, as well as the calculation and analysis of broken and impurity rates, this paper proposes an image-based detection method for broken and impurity rates. It utilizes the principle of electromagnetic vibration to achieve rapid and uniform spreading of high-throughput wheat grains. Building on this, the semantic segmentation model is improved to achieve precise identification and segmentation of broken grains and impurities, and deployed on AI embedded devices to realize the detection and display of wheat broken and impurity rates during combined harvester operations.

2. Materials and methods

2.1. Detection methods

The traditional manual detection is to use the five-point sampling method to pick up wheat samples from the grain outlets, and to clean the broken grains and impurities in the wheat samples for weighing. The mass percentage of broken grains and impurities is calculated as the broken rate and impurity rate (Lu, 2022). However, this method is time-consuming, labor-intensive, and poor-repeatability. Moreover, when the sample size is small, the objectivity of the detection results is compromised, making it challenging to meet the testing requirements for large quantities of wheat. Therefore, this paper proposes an image-based method for detecting wheat broken rate and impurity rate, and the core of this method lies in establishing a pixel-mass mathematical model, where the mass of each target object is calculated based on the number of pixels, allowing the determination of wheat broken rate and impurity rate.

In order to establish the mathematical relationship between the number of image pixels and the mass of each target object, a sample of wheat is collected during the operation of combine harvester. This sample is used to conduct a calibration experiment to determine the correlation between the number of image pixels and the mass of each target object. The experiment samples include intact grains, broken grains, straw, and husks. During the experiment, 50 samples are randomly selected from the experimental group, and the mass of each target object is measured using an electronic scale with an accuracy of 0.001 g. Then, the image processing method is used to segment the target object and count the number of pixels. The segmentation effect of

each target object is shown in Fig. 1. The sample mass of each target object and the corresponding number of pixels are saved and recorded in Excel. The least square method is used to fit the data, resulting in the linear regression equation.

As shown in Fig. 1, the correlation coefficients R^2 between the pixel number and the mass of the intact grain, broken grain, straw, and husk are 0.999, 0.957, 0.960, and 0.963, respectively. These values are significantly higher than 0.95, indicating a strong correlation between the mass and the number of pixels. This suggests that the regression equation accurately fit the observed values (Liang et al., 2016).

On the basis of the above mathematical model, considering the differences in moisture content and maturity of wheat in different scenarios, the grain correction coefficient μ_1 and impurity correction coefficient μ_2 are introduced. The μ_1 is calculated according to the ratio of thousand-grain weight, while μ_2 mainly considers the difference of moisture content of impurities, which is obtained by the ratio of moisture content. The final calculation formula of wheat broken rate and impurity rate is as follows:

$$\begin{aligned}\mu_1 &= \frac{m'_k}{m_k} \\ \mu_2 &= \frac{w'}{w_0} \\ r_b &= \frac{m_b}{m_g + m_b} = \frac{f_b(n_b) \times \mu_1}{[f_g(n_g) + f_b(n_b)] \times \mu_1} \\ r_i &= \frac{m_s + m_h}{m_g + m_b + m_s + m_h} = \frac{[f_s(n_s) + f_h(n_h)] \times \mu_2}{[f_g(n_g) + f_b(n_b)] \times \mu_1 + [f_s(n_s) + f_h(n_h)] \times \mu_2}\end{aligned}\quad (1)$$

In the formula, m_k and w_0 represent the original thousand-grain weight and moisture content, respectively, while m'_k and w' represent the thousand-grain weight and moisture content under specific scenarios. r_b and r_i represent the broken rate and impurity rate of wheat, respectively. g , b , s , and h represent intact grain, broken grain, straw, and husk, respectively. m , $f()$, and n represent the mass, linear regression model, and pixel number of each target object, respectively.

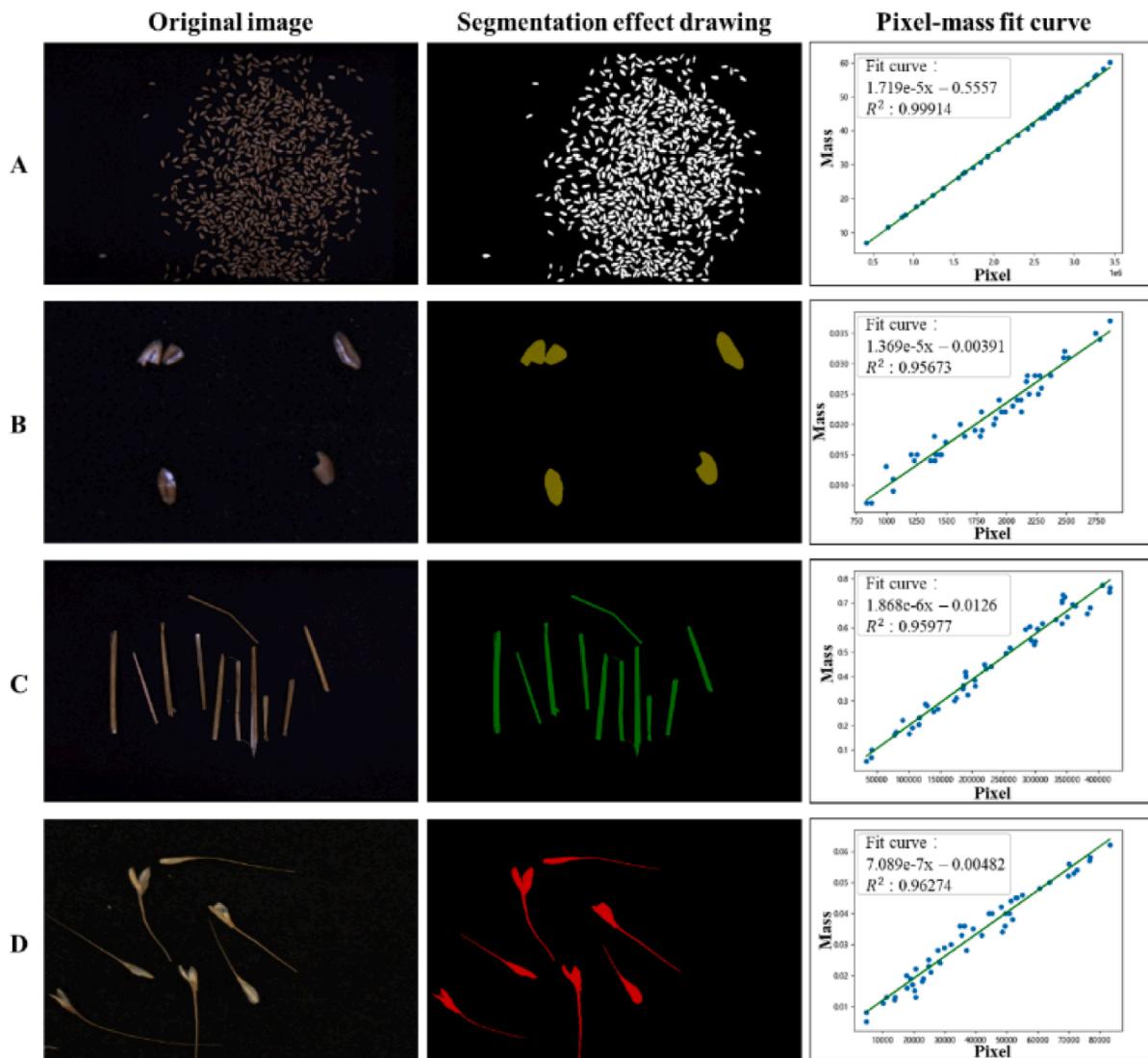


Fig. 1. Sample segmentation effect and pixel-mass fitting curve:(A) intact grain; (B) broken grain; (C) straw; (D) husk.

2.2. Construction of detection system

During the operation of the combine harvester, the grain is lifted to the outlet by the elevator auger and deposited in the silo. Due to the accumulation of scattered grains, there are varying degrees of overlap, occlusion, and dense adhesion between grain kernels. As a result, it becomes challenging to obtain information regarding grain broken characteristics and impurity mass. Therefore, in order to improve the detection accuracy and efficiency of wheat broken rate and impurity rate, this study established a broken rate and impurity rate detection system, which mainly includes wheat image acquisition device, AI embedded processor and upper computer human-computer interaction interface (Fig. 2). The wheat image acquisition device is designed based on the principle of electromagnetic vibration. When installed under the grain outlet of the combine harvester, the automatic sampling of wheat can be completed. This installation also eliminates the accumulation of wheat grains, enables the rapid and uniform distribution of large quantities of grains, and facilitates image acquisition. The AI embedded processor utilizes Nvidia's Jetson Orin NX development board. This board enables hardware control of the image acquisition device and also allows for processing and analysis of the image using a deployed deep learning model in an independent thread. The results are then displayed on the human-computer interaction interface of the upper computer in the combine harvester's cockpit.

As shown in Fig. 2(A) and 2(B), the image acquisition device is mainly composed of hopper, industrial camera, COB light source, electromagnetic vibrator, conveyor plate, discharge opening, steering gear and camera fixed bracket, among others. When the hopper is filled with wheat, the wheat grains are completely poured onto the conveyor plate under the drive of the coaxial steering gear. An electromagnetic vibrator is installed at the bottom of the conveyor plate. Based on the exciting force generated by the electromagnetic vibrator, the wheat grains can

vibrate and undergo linear movement on the conveying plate. Based on the previous research foundation (Wang et al., 2023), in order to ensure that the wheat grain can be evenly flattened under the action of the electromagnetic vibrator, the motion and force analysis of the grains was conducted to study the influence of the working parameters of the detection device such as vibration frequency, amplitude, conveying angle and friction factors on the single-layer and dispersion degree of the grain is studied, and the relevant evaluation indexes are introduced to measure the distribution characteristics of the grain. A multi-factor and multi-level orthogonal test is established, and the optimal working parameters for uniform flatten of the grain are determined according to the results of the orthogonal test. After the grains are evenly flattened, the industrial camera completes the collection of wheat static images through the visual window, and under the continuous vibration of the electromagnetic vibrator, the wheat grains are quickly translated to the discharge opening to complete the sample discarding operation. The multi-sample continuous acquisition of wheat images can be achieved by utilizing the sampling-discarding structure.

2.3. Dataset construction

The wheat samples used in the experiment are obtained from the trial field of the National Precision Agriculture Research Demonstration Base in Xiaotangshan Town, Beijing, in June 2021. The thousand-grain weight of the wheat samples is approximately 36 g, and the moisture content is 6.5 % after drying. 50 g of wheat is taken in each experiment, about 1400 grains. In the process of collecting the dataset, the intact grains, broken grains, and impurities are first roughly divided, and the samples are proportionally allocated according to different broken rates and impurity rates, so as to ensure that the image data of broken grains and impurities are diverse and the sample size is sufficient. Through adjusting the industrial camera lens and camera parameters, as well as

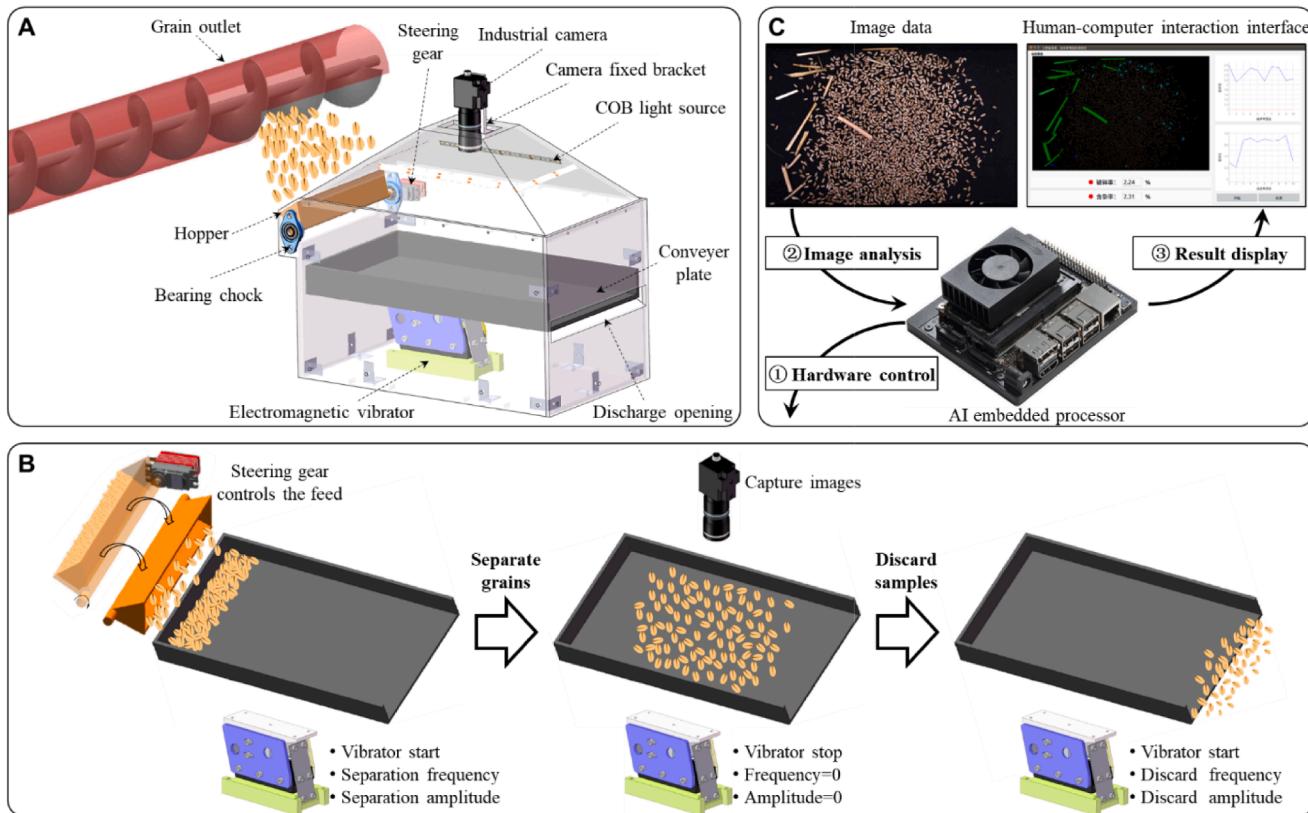


Fig. 2. Wheat broken rate and impurity rate detection system: (A) wheat image acquisition device; (B) wheat grain sampling-discarding process; (C) system architecture.

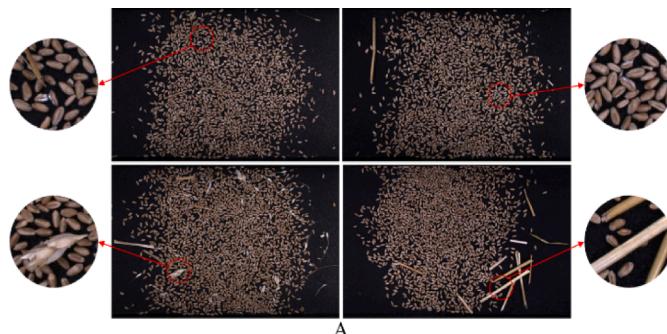
implementing a reasonable lighting design, the wheat image acquisition device constructed in this study is capable of capturing clear and easily processable image data, as depicted in Fig. 3(A). The image resolution is 4024 pixels \times 2600 pixels. A total of 137 original images are collected, and they are divided into training set, validation set, and test set according to a ratio of 7:1.5:1.5.

Then the Labelme software is used to classify and label all kinds of targets in the image. However, due to the cumbersome image labeling process of the semantic segmentation model, and the fact that there are too many grains in a single image which contains about 1400 intact grains, one by one labeling will undoubtedly cause huge manual workload. Therefore, in the semantic segmentation model, the intact grains are regarded as the background, and only the three types of targets of broken grains, straw and husk are labeled. The semantic segmentation model is used to segment these three types of targets, and the corresponding pixels are counted. The statistics of the number of intact grain pixels adopts the image processing method. The original image and the prediction results of the semantic segmentation model are binarized respectively, and subtracted to obtain a binary image which only retains the intact grain. Thus the number of intact grain pixels is counted. This method can not only reduce the workload of manual labeling, but also greatly reduce the segmentation objects of the model, which can be used to improve the segmentation accuracy. The color label masks generated by classifying and labeling broken grains, straws and husks are shown in Fig. 3(B).

Due to the harsh field environment, the actual harvesting operation process may face changes in environmental conditions, such as image luminosity distortion caused by changes in lighting conditions, image blur and complex image background caused by dust attached to the camera lens or conveying plate. In order to improve the generalization ability of the model in different scenes when the original image sample size is limited, the image is enhanced by processing methods such as brightness change, random rotation, Gaussian noise, filtering, and atomization. The enhancement effect is shown in Fig. 4. In addition, although the images before and after enhancement are different, there are still many similar features. Therefore, using one of the images as the training set to learn and then testing the corresponding enhanced image will not accurately reflect the true generalization ability of the model. In order to prevent the test set images and the training set images from being rooted in the same original image after data expansion, only 20 test set images are enhanced, but not expanded. Only the remaining 117 training set images are expanded, and the total number of training set images after expansion is 234.

2.4. DeepLab-EDA network model design

Since the image data collected in this paper contains both large-sized straw and small-sized broken grains, there are high requirements for the model's ability to capture multi-scale features. In addition, the image background contains a large number of densely distributed intact grains, which increases the difficulty for the model to identify broken grains.



Therefore, considering the characteristics of image data and the features of various classical semantic segmentation models, the DeepLabV3 + model is selected as the baseline for the algorithm model. This model has a strong ability to capture multi-scale features and a relatively simple structure (Chen et al., 2018).

2.4.1. Expansion prediction strategy

Since DeepLabV3 + requires the input size to be an integer multiple of 512 pixels \times 512 pixels, and the original dataset image size is 4024 pixels \times 2600 pixels, forcing the input will result in poor image segmentation. There are typically two methods for adjusting the size of an image: scaling-down and cropping. However, forced scaling and downsampling can cause image distortion and significantly impact the accuracy of segmentation. The cropping method requires multiple predictions during the test phase, and excessive cropping will significantly increase the prediction time. Therefore, the scaling-down and cropping methods are considered comprehensively. The original image is cropped into 15 blocks using a sliding window of 1024 pixels \times 1024 pixels, and it is then scaled down to 512 pixels \times 512 pixels. This process ensures that the number of cropped blocks is minimized, the height and width scaling ratio remains consistent, and the image is only scaled down by 0.5 times while maintaining a high resolution. The images of the training and validation sets are processed using the above approach, and the training and validation sets consist of 2808 images each of 512 pixels \times 512 pixels after processing.

In the model testing stage, the image also needs to be scaled and cropped. However, the common sliding window cropping technique is prone to causing incorrect segmentation due to the correlation between each cropped block in the predicted image. Additionally, the overlapping sliding window cropping method is inaccurate when it comes to identifying the edge area (Wang et al., 2023; Wang et al., 2021). Therefore, this paper proposes an expansion prediction strategy. For the strategy, only the central region of the prediction results is retained, and the image edge part with inaccurate prediction results is discarded on the basis of overlapping sliding windows. And then in the post-processing stage, the central region of each predicted result image is spliced, which effectively addresses the issue of inaccurate segmentation in large-sized images. The specific operational steps are as follows:

- (1) Fill the right and lower boundaries of the image (highlighted in yellow in Fig. 5). After filling, the length and width of the image should be integer multiples of the central area, which is convenient for evenly dividing the image.

- (2) Fill all sides of the image (blue part of Fig. 5), and the width of each side fill = $1/2 \times (\text{window size} - \text{center size})$.

- (3) After completing the filling process, the image was cropped into 15 blocks using a sliding window of 1024 pixels \times 1024 pixels and a step of 870 pixels, and further scaled down to 512 pixels \times 512 pixels, and perform the prediction. Then, the prediction result was enlarged to 1024 pixels \times 1024 pixels, and only the central area of 870 pixels \times 870 pixels was retained.

- (4) The final prediction result of 4024 pixels \times 2600 pixels can be

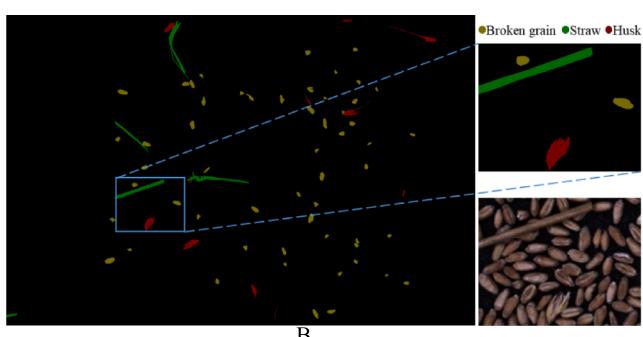


Fig. 3. Image dataset: (A) wheat image sample; (B) image labeling effect.

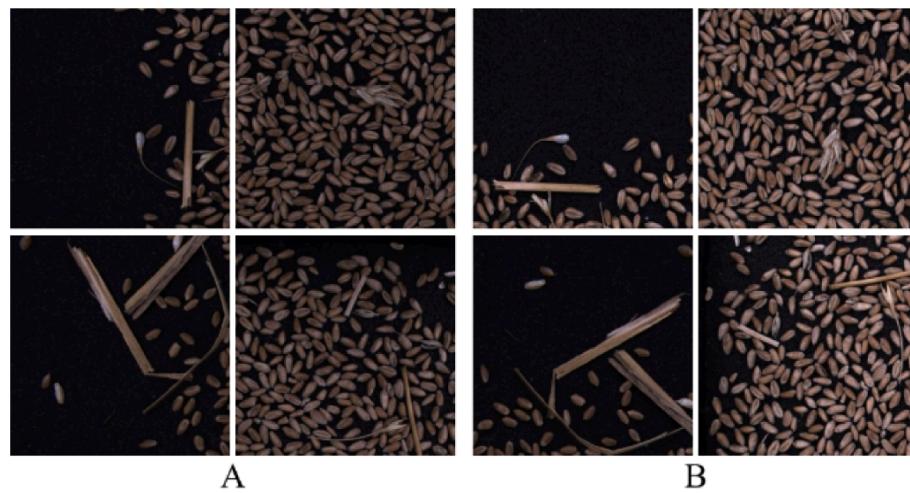


Fig. 4. Comparison of image enhancement effects: (A) pre-enhancement; (B) post-enhancement.

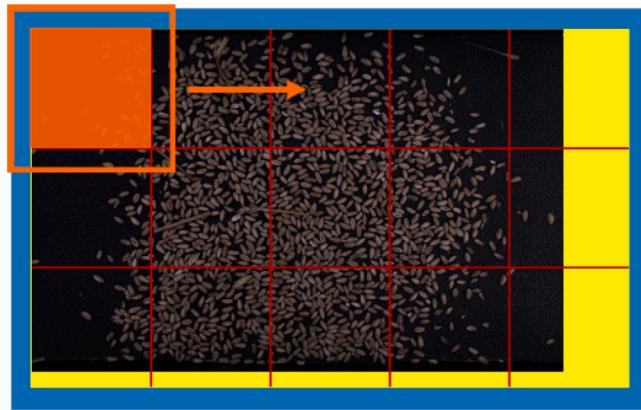


Fig. 5. Schematic diagram of expansion prediction.

obtained by concatenating the central prediction results of the 15 images and discarding the yellow filling area.

2.4.2. Detail detection head

The low-level feature map introduced by DeepLabV3+ in the Decoder part provides additional spatial position information and effectively guides the upsampling process. However, some spatial details are still missing, which leads to the problem of confusing segmentation boundaries of different objects with similar colors and textures in the segmentation results (Fan et al., 2021). Therefore, to further enrich the detail information of the segmentation results, this paper introduces the detail detection head, which models the prediction of details as a binary segmentation task and calculates the loss value between the low-level detail feature map and the edge map of label mask. This loss value is introduced to auxiliary training, which strengthens the model's learning of edges, corners and other detailed information in the Decoder stage of the model. The overall structure of the detail detection head is shown in Fig. 6.

Firstly, the low-level feature map output by the backbone network is extracted, and its edge features are extracted by using 3×3 convolution kernel, followed by normalization and activation operations. Finally, the 1×1 convolution is used to adjust the number of channels in order to obtain the detail feature map. In the process of obtaining the edge map of label mask, the Laplacian (Tong et al., 2021) convolution operator with strides of 1, 2 and 4 are used to obtain the multi-scale detail information of the label mask. The detail features are upsampled to the original size, and a trainable 1×1 convolution is used to fuse the multi-

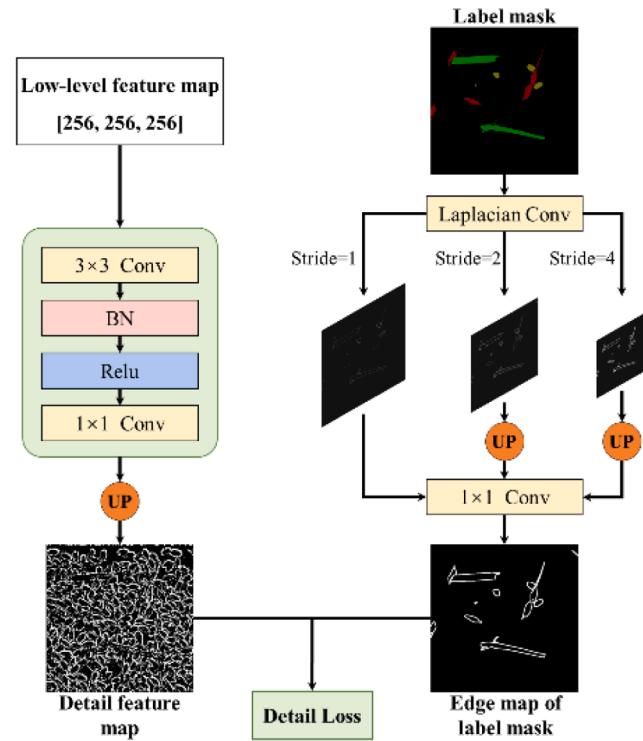


Fig. 6. Detail detection head.

scale information. Finally, the predicted details are transformed into a binary map with a threshold of 0.1. By calculating the detail loss value of the detail feature map and the edge map of label mask, the loss value is used for auxiliary training, and more detail information can be learned during the upsampling process. The detail detection head only guides the gradient descent direction during the training phase and serves as an auxiliary component in training. The test phase does not contain this part, so the detail detection head does not increase the complexity of the model, and improves the segmentation accuracy while maintaining the model inference time unchanged.

2.4.3. Convolutional block attention module

The deep feature map output by the backbone feature extraction network contains abundant semantic information. In order to further

improve the segmentation accuracy of the model, this paper utilizes the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) to optimize the spatial information and channel information of the deep feature map to reduce the interference of invalid features and improve the utilization of key features. CBAM is a module that incorporates spatial and channel dual attention mechanisms based on convolution. By giving an input, the module will successively infer the feature map using the channel attention module and the spatial attention module, and then multiply the attention weight with the input feature map to optimize the adaptive features. The overall structure of the Convolutional Block Attention Module is shown in Fig. 7.

The channel attention module utilize average pooling (AvgPool) and maximum pooling (MaxPool) to compresses the spatial dimension and obtain one-dimensional vectors. The AvgPool can map each pixel in the feature map, while the MaxPool only maps the pixel with the highest response in the feature map. These one-dimensional vectors are sent to their respective shared fully connected layers. The outputs from both pooling operations are fused together to aggregate different spatial information. Finally, the attention weights are obtained through a Sigmoid activation operation. The weight is multiplied with each channel of the feature map to optimize the channel information. The spatial attention mechanism also utilizes average pooling and max pooling for compression operations, but it focuses on the channel dimension. After compression, two single-channel feature maps are generated, and then a 1×1 convolution is used to fuse the two feature maps and obtain the spatial information of the single channel. The spatial attention weights can be obtained through the Sigmoid activation operation, and the optimization of spatial information is achieved by multiplying the weight with feature map spatial elements.

2.4.4. Overall structure of DeepLab-EDA model

Based on the improved strategy mentioned above, this paper introduces the DeepLab-EDA model, which is depicted in Fig. 8 to illustrate its overall network structure. The original test set images are cropped and reduced to 512 pixels \times 512 pixels using an expanding sliding window as the input for the model. The backbone features are extracted by the Xception-65 network (Chollet, 2017) with 8 times down-sampling, and the low-level feature maps and deep feature maps are outputted. Through the CBAM module, the channel weights and spatial weights are multiplied by the corresponding dimensions of the deep feature maps to optimize the utilization of features in different dimensions. The atrous spatial convolutional pooling pyramid module is used to capture the multi-scale context information of the deep feature map. This allows for better adaptation to the prediction of segmentation objects of different sizes in the image, such as grains and impurities, and the encoder stage is completed after fusing the multi-scale information. The low-level feature map outputted by the backbone network is introduced into the Decoder to provide more spatial location information for the upsampling of the deep feature map. Additionally, the detail detection head captures the detailed features of the low-level feature

map, and the detail loss is calculated by combining the edge map of label mask. This loss value and the segmentation loss value are added to guide the upsampling process of the feature map. The prediction result of a single cropped block is obtained. The segmentation results of each clipping block are post-processed to generate the final predicted segmentation map.

2.4.5. Model training

The loss value of the DeepLab-EDA model is composed of two parts: segmentation loss and detail loss. Since the detailed features targeted by the object of interest or the detection head in the segmentation process are only a small portion compared to the background of the image, it leads to imbalanced data samples. Therefore, this paper chooses Focal Loss (Lin et al., 2017) and Dice Loss (Milletari et al., 2016) as loss functions for model training. The total loss function is shown in Formula (2):

$$Loss = F_s + D_s + F_d + D_d \quad (2)$$

where F_s and F_d are the segmentation loss and detail loss calculated by Focal Loss, D_s and D_d are the segmentation loss and detail loss calculated by Dice Loss, respectively.

Based on the cross-entropy Loss function (Chockler et al., 2007), Focal Loss assigns different coefficients to positive and negative samples, so that samples with higher loss values dominate. The calculation method is shown in Formula (3):

$$L_{Focal} = \begin{cases} -\alpha(1-p)^\gamma \log p, & \text{if } y = 1 \\ -(1-\alpha)p^\gamma \log(1-p), & \text{if } y = 0 \end{cases} \quad (3)$$

where α is the weight coefficient and p is the prediction probability; y is the specific tuning coefficient, y is the predicted label.

Dice Loss pays more attention to the prediction results related to the positive samples, and according to Formula (4), the calculation results are independent of the true negative TN. Therefore, no matter how large the TN value is, it is difficult to affect the loss value. TN represents the negative samples with high confidence, which are the background class targets. Despite the imbalance between positive and negative samples caused by the proportion of background class targets, Dice Loss still prioritizes the mining of positive samples.

$$L_{Dice} = 1 - \frac{2TP}{2TP + FP + FN} \quad (4)$$

where TP is a true positive, FP is a false positive and FN is a false negative.

Based on the model structure and loss function mentioned above, the model construction is completed using Python 3.8 and PyTorch 1.7.0, and the model is trained on the Nvidia A40 GPU (48G memory) platform. To expedite the model fitting process, transfer learning is employed to load the weight parameters of the backbone network that have been previously trained on the ImageNet dataset (Deng et al., 2009). The backbone network is then frozen using the freezing training method, and only the head network is trained. After 50 epochs, the backbone network is thawed, resulting in a total of 120 epochs. The batch size is set to 16 when freezing the training, and it is later changed to 8 after unfreezing. The initial learning rate is 0.00005, and the cosine annealing strategy is used to dynamically adjust the learning rate (Loshchilov and Hutter, 2016). The curves depicting the training loss value and validation loss value changes are shown in Fig. 9(A). The mean intersection over union (MIoU) results are verified for the model every 10 epochs, and the change curve of the MIoU calculation results is shown in Fig. 9(B). The model gradually converges as the number of iterations increase, and after 100 epochs, the loss value and MIoU value fluctuate to be stable. The training results at this point are used as the final weight parameters.

2.4.6. Model evaluation metrics

In order to objectively demonstrate the segmentation performance of

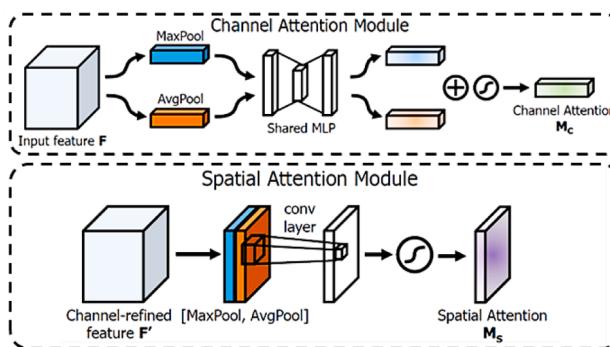


Fig. 7. Convolutional Block Attention Module.

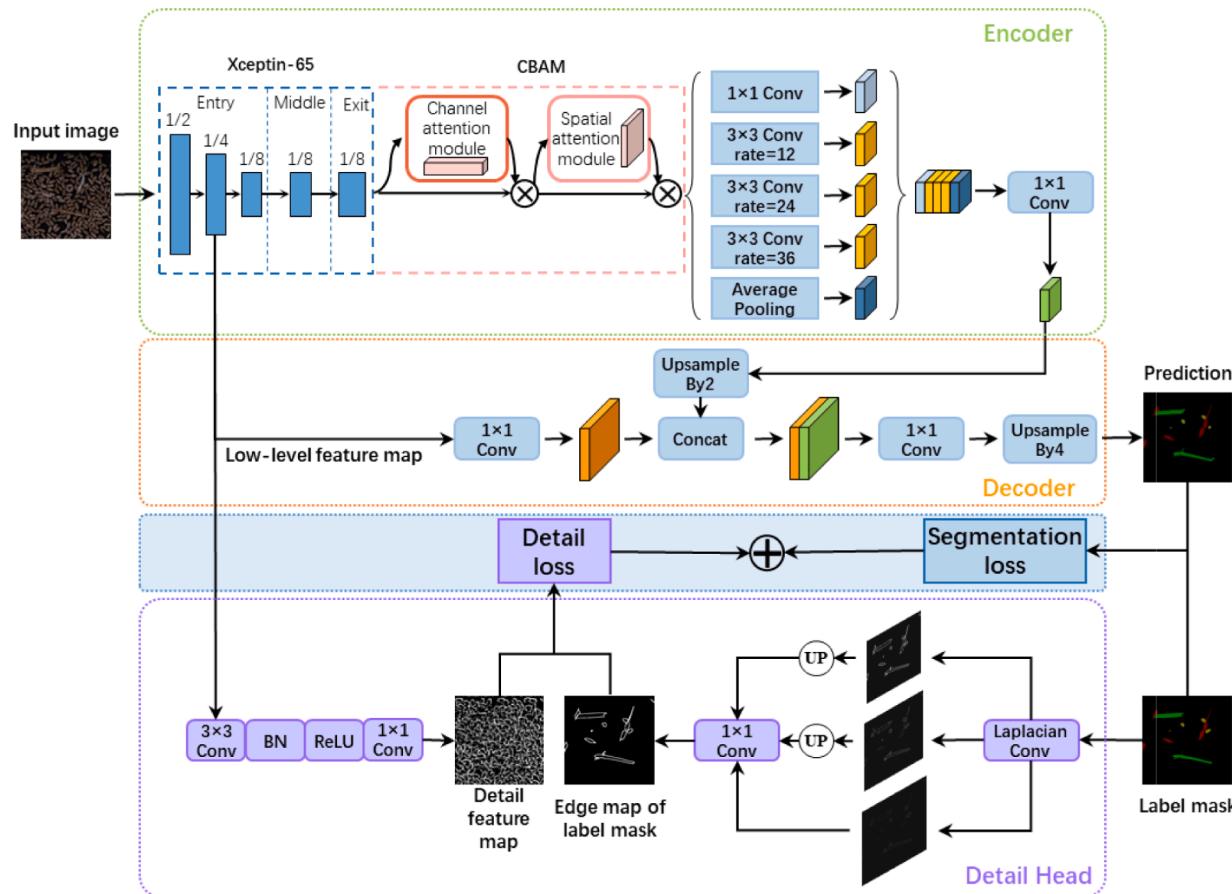


Fig. 8. Network structure of DeepLab-EDA model.

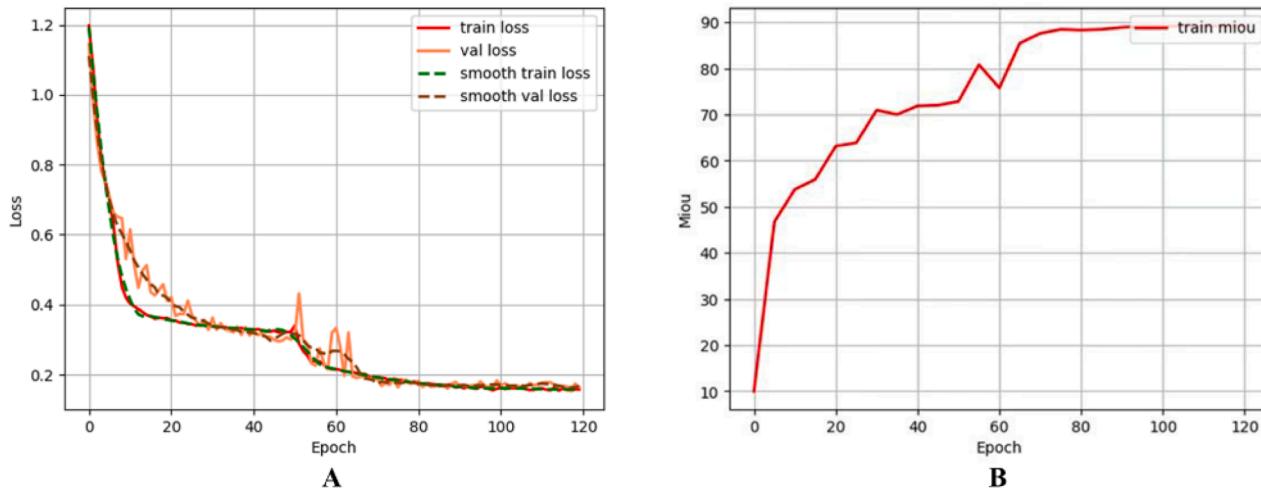


Fig. 9. Model training results: (A) change of training loss and validation loss; (B) validation result of Miou.

the proposed model, the model evaluation metrics adopts the mean intersection over union (Miou), category mean pixel accuracy (MPA) and category mean recall (MR) (Yu et al., 2023).

(1) Mean intersection over union

In the predicted segmentation result, the ratio of the intersection and union of the predicted segmentation map and the labeled segmentation map of a specific category is referred to as the intersection over union ratio. The mean of all categories is known as the mean intersection over union ratio. The formula is as follows:

$$Miou = \frac{1}{K+1} \sum_{i=0}^k \frac{p_{ij}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (5)$$

where p_{ii} is the number of pixels of class i predicted as class i, p_{ij} is the number of pixels of class i predicted as class j, p_{ji} is the number of pixels of class j predicted as class i, and K is the number of classes.

(2) Category mean pixel accuracy.

The pixel accuracy represents the proportion of correctly classified

pixels out of the total number of pixels. The category mean pixel accuracy, which is the mean of all category pixel accuracies, is calculated as follows:

$$MPA = \frac{1}{K+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (6)$$

(3) Category mean recall

Recall is the ratio of the number of pixels that are correctly predicted as class i to the total number of pixels that are labeled as class i . Recall is proposed so that the model can predict all desired classes. The category mean recall is calculated by taking the mean of the recall rates for all categories. The formula is as follows:

$$MR = \frac{1}{K+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ji}} \quad (7)$$

3. Results and discussion

3.1. Model performance evaluation test and result analysis

In order to evaluate the performance of the model in this paper, it is compared with the manual annotation results and the prediction results of other deep learning models. Firstly, a set of ablation experiments are set up to compare the improvement effects of the expansion prediction strategy, detail detection head, attention mechanism module and the DeepLab-EDA model with respect to the baseline model, and compare the error with the manual annotation results to analyze and evaluate the segmentation and recognition ability of the model for wheat broken grains and impurities. In the second group of experiments, DeepLab-EDA and three classical semantic segmentation models Unet (Zunair and Ben, 2021), PSPnet (Zhao et al., 2017) and HRnet (Sun et al., 2019) are used as control groups to compare the performance of the proposed model with other classical semantic segmentation models.

3.1.1. Analysis of ablation experimental results

The test results of the DeepLab-EDA model and various improvement measures are shown in Table 1. The DeepLab-EDA model reaches 89.41 %, 95.97 % and 94.83 % in MIoU, MP and MR Indicators, respectively, which are 9.94 %, 7.41 % and 7.52 % higher than the baseline model. After adding detail detection, attention mechanism and expansion prediction strategy separately with DeepLabV3 + as the baseline, MIoU is increased by 6.21 %, 2.01 % and 3.58 %, MPA is increased by 3.89 %, 1.54 % and 3.47 %, and MR is increased by 4.05 %, 1.40 % and 2.31 %, respectively. Therefore, the improvement measures added in this paper are effective in improving the segmentation accuracy.

As shown in Fig. 10, before the improved model, there are obvious traces of boundary lines at the location of the original image cropping and splicing (the location of the center line of the blue virtual frame line), and the wheat husk is missegmented. The two adjacent cropping blocks are both lack of complete features of this region, resulting in different recognition results of cropping blocks on both sides of the boundary line for this region. There are no cutting and splicing traces such as boundary lines in the blue dashed box of the improved model, and the recognition is accurate. Therefore, the expansion prediction

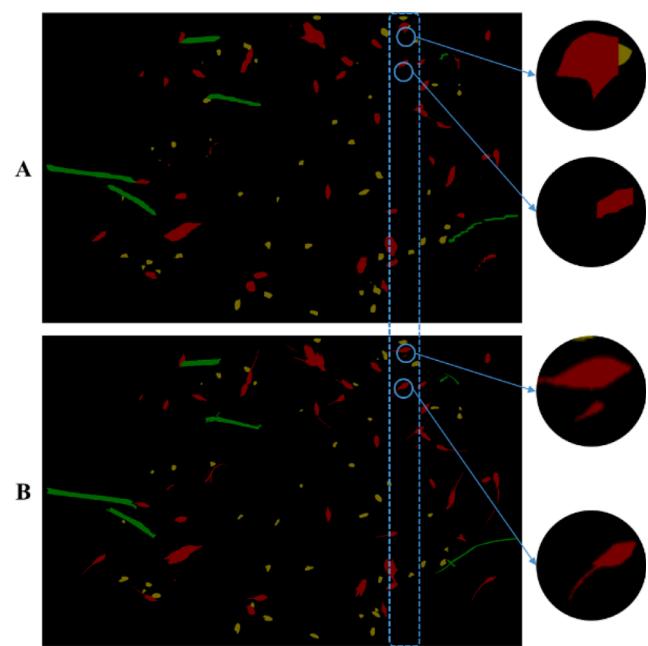


Fig. 10. Effect of comparison of expansion prediction strategies: (A) no expansion prediction; (B) using expansion prediction.

strategy has the best improvement effect in the ablation experiment, as the expansion prediction only retains the central region with the best prediction effect, which effectively solves the problem of low segmentation accuracy of large-size images.

Fig. 11 shows the learning differences of the model on the detail features of the low-level feature map before and after the improvement, and the comparison of the predicted segmentation results. By comparing Fig. 11(B) and (C), it can be found that the low-level detail feature map extracted by the DeepLab-EDA model contains more information, its edges and corners are clearer, and the segmentation results are more refined. Taking the slender tail of wheat husk, the small branches of wheat stalk and the narrow gaps between broken grains marked by blue circles in Fig. 11 as an example, the model before improvement is not sensitive to the boundary information, resulting in incomplete identification of the contour details of the target, while the improved model can accurately segment its detailed features. Compared with the purple circle marked part in Fig. 11(D), broken grains with unobvious characteristics are the difficulty of detection. The DeepLab-EDA model based on the convolutional block attention module benefits from the mining of key features, which improves the problem of missing recognition of broken grains. In terms of detection speed, the average detection time required by DeepLabV3 +, DeepLab-Detail, DeepLab-Attention, DeepLab-Expansion and DeepLab-EDA to complete a single image is 1.938 s, 1.938 s, 1.951 s, 2.471 s and 2.627 s, respectively. After DeepLab-EDA is accelerated by TensorRT, the model inference time is only 0.031 s, and its main time is spent in the image preprocessing and post-processing stages. This is due to the cumbersome image processing steps of large-scale high-resolution images, and TensorRT in the embedded development board cannot accelerate the optimization of OpenCV image processing, so it takes a lot of time in the image processing steps, but the overall time can meet the speed requirements of sampling detection.

3.1.2. Performance test results of different semantic segmentation models

The same dataset is used to train classical semantic segmentation models such as DeepLab V3+, Unet, PSPnet and HRnet and the DeepLab-EDA model proposed in this paper. The prediction results of each model on the test set are shown in Table 2. Compared with Unet, PSPnet, HRnet and DeepLabV3 + models, the DeepLab-EDA model has the MIoU

Table 1

Segmentation results predicted by the model.

Evaluation Metrics	MIoU /%	MPA /%	MR /%
Segmentation Algorithm			
DeepLabV3+	79.47	88.56	87.31
DeepLab- Detail	81.48	90.10	88.71
DeepLab- Attention	83.05	92.03	89.62
DeepLab- Expansion	85.68	92.45	91.36
DeepLab-EDA	89.41	95.97	94.83

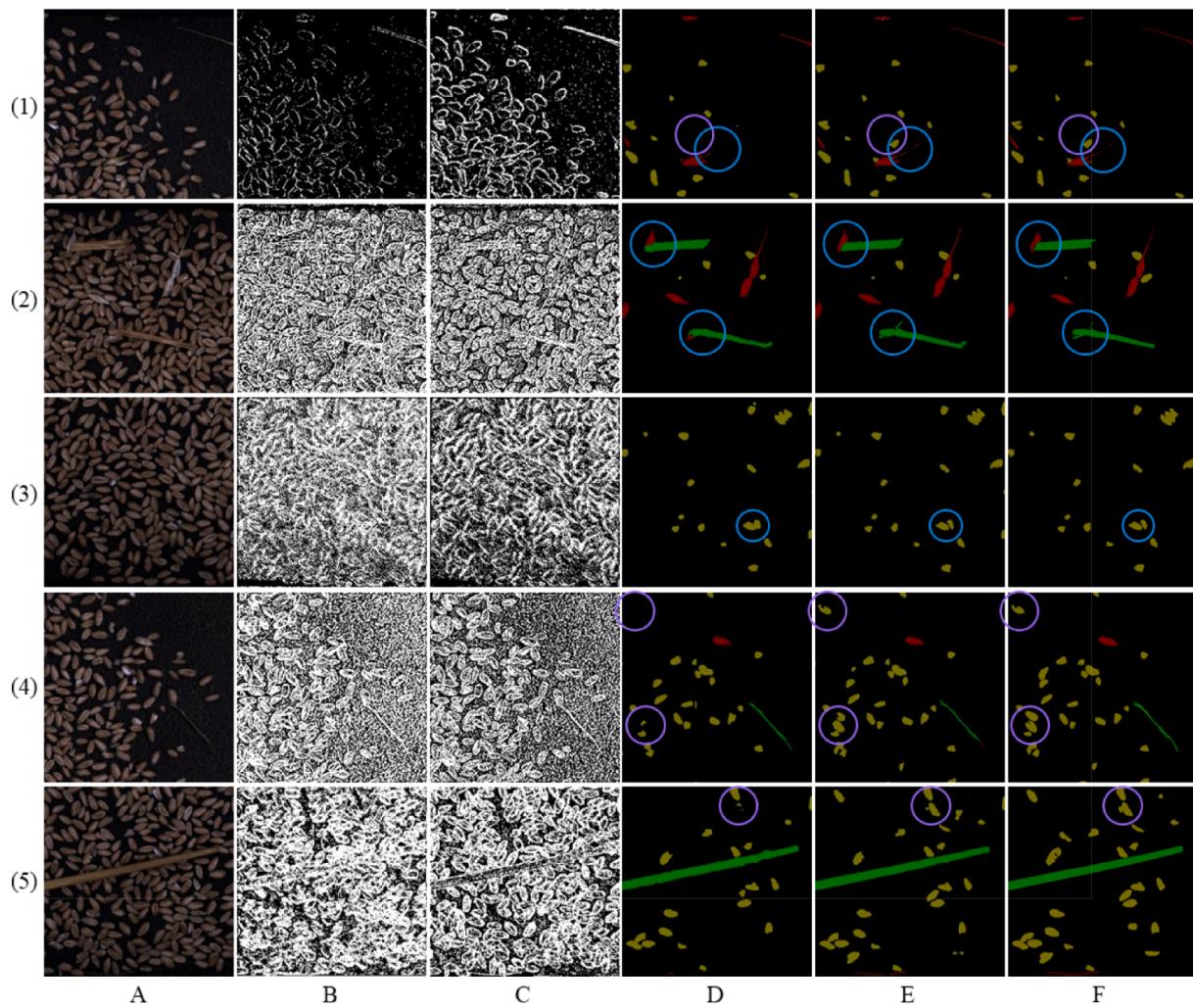


Fig. 11. Prediction effect of wheat broken grain and impurity segmentation: (1) ~ (5) are examples; (A) original image; (B) the low-level detail feature map before improvement; (C) the improved low-level detail feature map; (D) segmentation prediction results before improvement; (E) segmentation prediction results after improvement; (F) label mask;

Table 2
Comparison of test results of different semantic segmentation models.

Evaluation Metrics Segmentation Algorithm	MIoU /%	MPA /%	MR /%
Unet	78.26	88.51	86.50
PSPnet	70.66	80.99	82.58
HRnet	74.12	85.07	84.79
DeepLabV3+	79.47	88.36	87.42
DeepLab-EDA	89.41	95.97	94.83

increased by 11.15 %, 18.75 %, 15.29 % and 9.94 % respectively, and the MPA increased by 7.46 %, 14.98 %, 10.9 % and 7.61 % respectively. MR is increased by 8.33 %, 12.25 %, 10.04 % and 7.41 %, respectively. It can be seen that the DeepLab-EDA model proposed in this paper is significantly superior to other classical semantic segmentation models.

3.2. Analysis of group matching experiment results

The wheat harvested in the field experimental base is divided into four categories: intact grain, broken grain, straw and husk. A certain amount of broken grain, straw and husk is taken out and mixed with the intact grain evenly to form three groups of experiment samples with the broken rate and impurity rate of 0.5 %, 1.5 % and 2.5 %. The well-

portioned samples are placed into the hopper of the sampling device manually, and the single detection tasks such as automatic feeding, vibration translation of wheat grains, image acquisition and analysis, and material abandonment could be carried out through the control button of the human-computer interaction interface. In addition, the experiment results were displayed in the human-computer interaction interface. The experiment is repeated 10 times for each classification level and the average value is calculated, and then the error is calculated by comparing with the broken rate and impurity rate levels set in this group of experiments. The experiment device is shown in Fig. 12.

The relative error δ is introduced to judge the accuracy of the detection results of broken rate and impurity rate. The relative error calculation formula is as follows:

$$\delta = \frac{|T - P|}{T} \times 100\% \quad (8)$$

where T is the true value and P is the predicted value.

Table 3 shows the results of group matching experiment. When the level of broken rate and impurity rate is 0.5 %, the average error of broken rate and impurity rate are the smallest. With the increase of broken grains and impurities, due to some impurities covering the upper surface of the intact grain, the calculation of the pixel number of the intact grain is small, that is, the calculation of the mass of the intact grain is small, and the error of the broken rate and the impurity rate increase.

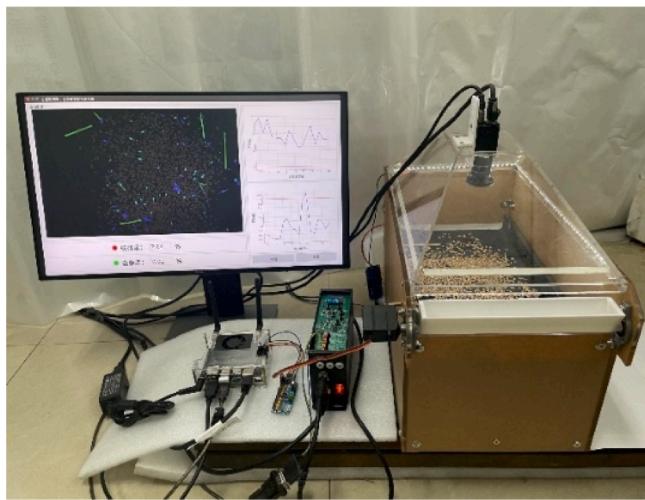


Fig. 12. Physical drawing of the detection device.

Table 3
Results of group matching experiment.

Broken rate and impurity rate grade	Broken rate δ	Impurity rate δ
0.5 %	5.64 %	4.76 %
1.5 %	7.27 %	6.02 %
2.5 %	9.72 %	8.12 %
General average	7.54 %	6.30 %

In addition, with the increase of broken grains and impurities, the probability of broken grains being occluded by impurities increases, resulting in missed detection of broken grains, which is also the main reason why the error of broken rate is generally higher than the error of impurity rate. Based on this, the variation curve of broken rate fluctuates more obviously than that of impurity rate. The general average error of broken rate is 7.54 %, and the general average error of impurity rate is 6.30 %, indicating that the detection system in this paper could accurately detect wheat broken rate and impurity rate.

3.3. Analysis of field experiment results

In order to verify the reliability of the wheat breaking rate and impurity rate detection system in the harvest operation of the combine harvester, a field experiment is carried out in Pinggu District of Beijing in June 2023. The wheat variety in the experiment base is 'Lunxuan 266', with plant height of 52 cm, moisture content of 10.5 %, and thousand-grain weight of 44.1 g.

During the experiment, the detection device is installed under the grain outlet of the Rewo Goshen GM80 combine harvester to realize wheat sampling and analysis, and the harvester is kept at a forward speed of 4 km/h for harvesting operations, as shown in Fig. 13. During the driving process of the harvester, the body vibration and road flatness may affect the grain single-layer effect. Therefore, a series of mitigation measures were adopted during the installation of the device to avoid the interference of external vibration. For example, all fixed connections are added with spring gaskets and shock mitigation rubber rings, and a large counterweight block is added to the bottom of the device to adjust and optimize the vibration characteristics of the device. Due to the increase of the overall mass of the device, the sensitivity of the device to external vibration can be reduced, which plays a role in shock mitigation, so as to better maintain its stability and work efficiency.

Five groups of field experiments are carried out, and each group is continuously detected for one minute. The average of multiple detection results within one minute is taken as the system detection result of this group of experiments. The broken grains and impurities in the samples are screened and weighed, and the broken rate and impurity rate are calculated as the manual detection results. The system detection results and manual detection results are compared to verify the effectiveness of the wheat broken rate and impurity rate detection system in this paper.

The images collected during the field experiment and their segmentation effects are shown in Fig. 14. It can be observed from the images that the grain still maintains a good single-layer effect, but there is a large amount of dust attached to the conveyor plate and camera lens, resulting in lower image brightness and lower clarity. Most of the broken grains and impurities in Fig. 14 have achieved good recognition and segmentation effects, and only some broken grains have unclear contour segmentation. The segmentation results show the strong generalization ability of the model in this paper. For the calculation results of the final broken rate and the impurity rate, the relative error δ is used as the evaluation index. The test results are shown in Table 4.. The general average errors of the system detection of wheat broken rate and impurity



Fig. 13. Field experiment.

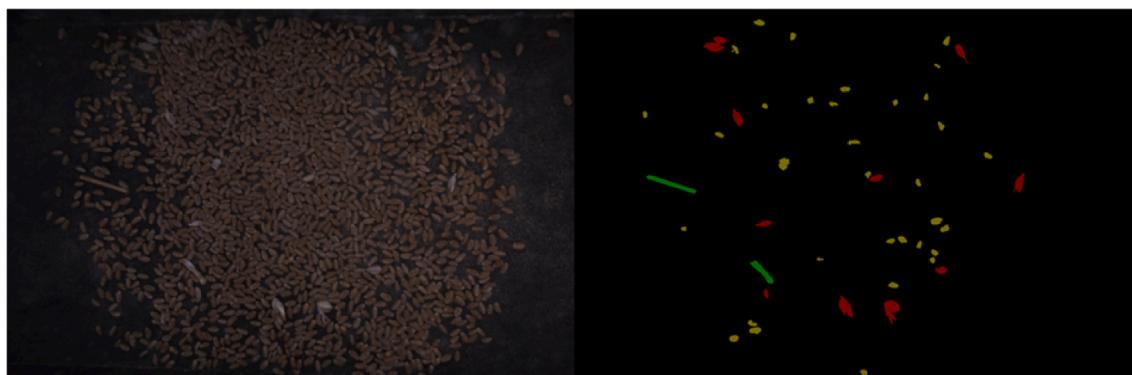


Fig. 14. Field experiment image segmentation effect.

Table 4
Results of field experiments.

Experimental groups	Broken rate manual detection /%	Broken rate system detection /%	Broken rate δ / %	Impurity rate manual detection /%	Impurity rate system detection /%	Impurity rate δ / %
1	1.812	2.094	15.58	1.041	1.177	13.09
2	1.792	1.614	9.96	1.554	1.405	9.65
3	2.189	1.939	11.41	1.383	1.511	9.31
4	0.839	0.714	14.90	2.589	2.782	7.45
5	1.577	1.810	14.77	1.908	2.087	9.36
General average			13.32			9.77

rate in the field experiment are 13.32 % and 9.77 %, respectively, which are higher than those in the indoor group matching experiment, and the main reasons for the increase of errors include two aspects: (1) The harsh field working environment and high dust environment affect the accuracy of image segmentation; (2) When the combine harvester passes through large pothole or stone, the grains are not completely single-layer, and there is a certain degree of accumulation, which leads to an increase in the error of image detection results. However, the cultivated land in the field environment is usually trimmed, and the extreme road conditions are relatively few, and the system error caused is kept within the acceptable range, which indicates that the wheat broken rate and impurity rate detection system in this paper is stable and effective.

4. Conclusion

In this paper, the broken rate and impurity rate of wheat is studied, and a wheat broken rate and impurity rate detection system based on the DeepLab-EDA semantic segmentation model has been developed to address the issues of time-consuming and laborious processes, poor repeatability, and low detection accuracy in current manual methods for detecting broken rate and impurity rate in wheat, as well as other related detection systems. A device for acquiring wheat images has been designed and developed using the principle of electromagnetic vibration, and it aims to achieve accurate identification and segmentation of broken grains and impurities in wheat. Besides, the deep learning model is deployed on the AI embedded processor to facilitate online processing and analysis of wheat image data, and through the human-computer interaction interface, the detection results of the broken rate and impurity rate is displayed. This provides a foundation for intelligent control of operational parameters of the combine and improves the operational quality of the combine.

The main conclusions are as follows:

- (1). The DeepLab-EDA semantic segmentation model is constructed by introducing the expansion prediction strategy, detail detection head, and convolutional attention mechanism module. The model achieves 89.41 %, 95.97 %, and 94.83 % in MIoU, MP, and MR indicators, respectively. Compared to the baseline model, the

proposed model shows improvements of 9.94 %, 7.41 %, and 7.52 %. Therefore, the proposed algorithm can effectively identify three types of wheat objects: broken grains, husks, and straw, and accurately segment them.

- (2). Three groups of samples with different rates of broken and impurity are set at 0.5 %, 1.5 %, and 2.5 % to conduct the group matching experiment of the detection system in this paper. The experimental results show that the system's general average error in detecting the broken rate is 7.54 %, and the general average error in detecting the impurity rate is 6.30 % for the experiment samples with different ratios of broken rate and impurity rate. In addition, the detection device is installed beneath the grain outlet of the combine harvester for field experiments, and the experimental results show that the general average errors of the broken rate and the impurity rate in the field experiment are 13.32 % and 9.77 %, respectively.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

Funding: This work was supported by Natural Science Foundation of China (Grant No. 32372592), and Horizontal Project of China Agricultural University (No.69193028).

References

- Wang Z, Zhou Y, Wang S, Wang F, Xu Z. House building extraction from high-resolution remote sensing images based on IEU-Net. National Remote Sensing Bulletin , 25 (11):2245-2254.

- Basati, Z., Rasekh, M., Abbaspour-Gilandeh, Y., 2018. Using different classification models in wheat grading utilizing visual features[J]. *Int. Agrophys.* 32 (2), 225–235.
- Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
- Chen, Z., Fan, W., Luo, Z., Guo, B., 2022. Soybean seed counting and broken seed recognition based on image sequence of falling seeds[J]. *Comput. Electron. Agric.* 196, 106870.
- Chen, J., Han, M., Lian, Y., 2020. Segmentation of impurity rice grain images based on U-Net model[J]. *Transactions of the Chinese Society of Agricultural Engineering (transactions of the CSAE)* 36 (10), 174–180.
- Chen, M., Jin, C., Ni, Y., Xu, J., Yang, T., 2022. Online detection system for wheat machine harvesting impurity rate based on DeepLabV3+[J]. *Sensors* 22 (19), 7627.
- Chen, J., Lian, Y., Li, Y., 2020. Real-time grain impurity sensing for rice combine harvesters using image processing and decision-tree algorithm[J]. *Comput. Electron. Agric.* 175, 105591.
- Chockler, H., Farchi, E., Godlin, B., et al., 2007. Cross-entropy based testing[C]//Formal Methods in Computer Aided Design (FMCAD'07). IEEE 101–108.
- Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1251-1258.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., ImageNet, F.-F., 2009. A large-scale hierarchical image database[C]//. IEEE 248–255.
- Fan, M., Lai, S., Huang, J., et al., 2021;. Rethinking Bisenet for Real-Time Semantic Segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9716–9725.
- Guan, Z., Li, H., Chen, X., Mu, S., Jiang, T., Zhang, M., Wu, C., 2022. Development of impurity-detection system for tracked rice combine harvester based on DEM and mask R-CNN[J]. *Sensors* 22 (23), 9550.
- Liang, J., Feng, C., Song, P., 2016. A Survey on Correlation Analysis of Big Data[J]. *Chinese Journal of Computers* 39 (01), 1–18.
- Jin, J., Chen, Y., Pan, R., Cao, T., Cai, J., Yu, D., Chi, X., Cernava, T., Zhang, X., Chen, X., 2022. CAMFFNet: a novel convolutional neural network model for tobacco disease image recognition[J]. *Comput. Electron. Agric.* 202, 107390.
- Lin, T.Y., Goyal, P., Girshick, R., et al., 2017. Focal Loss for Dense Object Detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2980–2988.
- Liu, L., Du, Y., Chen, D., Li, Y., Li, X., Zhao, X., Li, G., Mao, E., 2022. Impurity monitoring study for corn kernel harvesting based on machine vision and CPU-Net[J]. *Comput. Electron. Agric.* 202, 107436.
- Liu, Y., Zhang, T., Jiang, M., 2022. Review on non-destructive detection methods of grape quality based on machine vision [J]. *Transactions of the Chinese Society for Agricultural Machinery* 53 (s1), 299–308.
- Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.
- Lu L. Research on Impurity and Brokenness Detection Method of Machine Harvested Wheat Based on Deep Learning [D]. China Agricultural University, 2022.
- Ma, S., Li, Y., Peng, Y., 2023. Spectroscopy and computer vision techniques for noninvasive analysis of legumes: a review[J]. *Comput. Electron. Agric.* 206, 107695.
- Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee 565–571.
- Ni, X., Liu, K., Zhou, X., Mao, X., Chen, D., Wang, S., 2023. Unsupervised anomaly analysis-based manufacturing quality test and grading method for combine harvesters[J]. *Comput. Electron. Agric.* 210, 107898.
- Ramirez-Paredes, J.P., Hernandez-Belmonte, U.H., 2020. Visual quality assessment of malting barley using color, shape and texture descriptors[J]. *Comput. Electron. Agric.* 168, 105110.
- Rong, D., Rao, X., Ying, Y., 2017. Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm[J]. *Comput. Electron. Agric.* 137, 59–68.
- Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.2019: 5693-5703.
- Tong, S., Jiang, M., Jiao, C., 2021. Research on an improved edge detection method of workpiece [J]. *Journal of Electronic Measurement and Instrumentation* 35 (1), 128–134.
- Wang Y., Zhao X., Xu L., Li C., Lu X., Li S. Experiment and Directional Movement Technology of Corn Seed Based on Electromagnetic Vibration[J].*Transactions of the Chinese Society for Agricultural Machinery*,2015,46(1):79-88.
- Wang L, Zhang Q, Feng T, Wang Y, Li Y, Chen D. Wheat Grain Counting Method Based on YOLO v7-ST Model[J].*Transactions of the Chinese Society for Agricultural Machinery*,2023,54(10):188-197,204.
- Wang, Y., Liu, W., Tang, Y., 2023. Classification of Urban Functional Areas by Convolution neural network recognition combined with sliding window and semantic reasoning [J]. *Geomatics and Information Science of Wuhan University* 48 (06), 950–959.
- Woo, S., Park, J., Lee, J.Y., et al., 2018;. Cbam: Convolutional Block Attention Module [C]//proceedings of the European Conference on Computer Vision (ECCV). 3–19.
- Wu, Z., Chen, J., Ma, Z., Li, Y., Zhu, Y., 2024. Development of a lightweight online detection system for impurity content and broken rate in rice for combine harvesters [J]. *Comput. Electron. Agric.* 218, 108689.
- Xia, H., Zhou, S., Liu, Y., Zhao, K., Li, Z., 2020. Design and Test of Directional Vibrating Seed-feeding Device for Flat Solanaceous Vegetable Seeds[J].*Transactions of the Chinese Society for Agricultural Machinery* 51 (9), 82–88.
- Yu, Y., Wang, C., Fu, Q., Kou, R., Wu, W., Liu, T., 2023. Survey of Evaluation Metrics and Methods for Semantic Segmentation [J]. *Comput. Eng. Appl.* 59 (6), 57–69.
- Zhang, X., 2021. Research on online detection method of combine operation quality based on vision[D]. Hubei University of Technology.
- Zhang S. Monitoring Method and Device for Impurity Rate and Crushing Rate of Flowing Grain Based on Multi Threshold Segmentation [D]. Jiangsu University, 2021.
- Zhao, H., Shi, J., Qi, X., et al., 2017;. Pyramid Scene Parsing Network[c]//proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2881–2890.
- Zhu, S., Ma, W., Lu, J., Ren, B., Wang, C., Wang, J., 2023. A novel approach for apple leaf disease image segmentation in complex scenes based on two-stage DeepLabV3+ with adaptive loss[J]. *Comput. Electron. Agric.* 204, 107539.
- Zunair, H., Ben, H.A., 2021. Sharp U-Net: depthwise convolutional network for biomedical image segmentation[J]. *Comput. Biol. Med.* 136, 104699.