IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 17, 2021, accepted January 20, 2022, date of publication January 31, 2022, date of current version February 25, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3147838

Detection of Wheat Unsound Kernels Based on Improved ResNet

HUI GAO, TONG ZHEN, AND ZHIHUI LI
College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
Key Laboratory of Grain Information Processing and Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China

Corresponding author: Zhihui Li (653081735@qq.com)

This work was supported by the National Key Research and Development Project under Project 2018YFD0401404.

ABSTRACT In the process of grain acquisition, the unsound kernels of wheat are traditionally detected manually. The determination method based on computer vision typically requires expensive equipment for image acquisition and has disadvantages of low recognition efficiency and difficulties in adhesion segmentation, which strongly limit the application in routine detections. In this paper, six kinds of wheat including sound kernel, broken kernel, sprouted kernel, injured kernel, moldy kernel and spotted kernel are considered as the samples. An image acquisition platform is built with low cost to capture wheat pictures. The designed two-kernels adhesion wheat segmentation algorithm based on concave-mask exhibits high accuracy, with the error rate of 0.93% for total 9988 wheat grains. By comparing the advantages and disadvantages of GoogleNet, DenseNet, IX-ResNet, Res2Net, this paper studies the optimization of depth, width, downsampling mode, convolution order, attention mechanism, receptive field. Finally a wheat unsound kernel detection method is proposed based on Res24_D_CBAM_Atrous. The Macro avg values of Precision, Recall and F1 are respectively 94%, 95% and 94%, which are increased by 3%-4% based on the original Res34. The prediction time is reduced by 220s, which can meet the rapid and accurate evaluation of wheat appearance quality. The method shows important theoretical significance and practical application value for wheat unsound kernel routine detection.

INDEX TERMS Unsound kernels, image acquisition platform, adhesive segmentation, improved ResNet, classification.

I. INTRODUCTION

In 2021, the sown area of wheat is about 23 million hectares and the output reaches 134 million tons, which account for about 20% of the sown area and output of grains in China. So the quality of wheat is directly related to the national economy and people's livelihood [1]. Unsound kernel refers to damaged wheat kernels that are still valuable, including injured kernel, spotted kernel (gibberella damaged kernel and black germ kernel), broken kernel, sprouted kernel and moldy kernel. The appearance features are as follows: Injured kernels have many insect-eroded holes on the surface; Gibberella damaged kernels are shriveled and stiff-white and the kernels surface is purple or accompanied by pink mildew; The embryo surface of black germ kernels is dark brown or black; Broken kernels are flattened and broken; Sprouted kernels show the seed coat of the embryo broken, obviously bulged or the bud and radicle break through the seed coat. The main effects are as follows: the changes of starch physicochemical properties, protein content, gluten structure and uric acid content result in peculiar smell, darkening color, pollution and quality deterioration of processed products [2], [3]; Even the continuous accumulation of mycotoxins seriously endangers the health of humans and animals [4], [5]; The increase of sprouted kernels and moldy kernels leads to the decrease of wheat test weight [6]; In the process of wheat acquisition and circulation, the unsound kernel is used as the limiting index of quality and the basis for increasing or deducting amount [7]. To sum up, the detection of wheat unsound kernels is of great significance to the evaluation of wheat quality.

At present, the detection of wheat unsound kernels mainly exists in grain depots, processing factories and quality inspection department and its technology mainly stays in manual detection. The deviation and fatigue in the understanding of unsound kernels is easy to cause grading errors; There is a risk of collusion and fraud among inspectors; The low

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Babu Thanikanti.

detection efficiency will directly affect the acquisition progress of the whole batch of grain [8], [9]. Hyperspectral imaging technology has the advantages of machine vision and spectral analysis, but the problem of large amount of data and expensive system cannot be avoided [10]. Acoustical detection technology can detect injured kernels according to different signals generated by wheat hitting metal targets, but it is easily interfered by noise [11]. Micro X-ray CT technology can observe the surface characteristic and internal structure of wheat kernels in an all round way, but it is necessary to find a more effective three-dimensional reconstruction method [12]. Traditional machine learning method can capture crop images through computer vision system, extract features according to color, shape, area and texture [13]. However, the feature extracted manually relies on human prior knowledge and the features are interrelated, resulting in redundancy. Therefore, it can no longer meet rapid and accurate evaluation of appearance quality [14].

So, the application of deep learning in the field of grain inspection has been widely concerned. In the detection of corn kernels, Huang *et al.* [15] first adopted VGG19 transfer learning. The result was better than the traditional machine learning algorithm, but the accuracy was still not ideal. Then they adopted more complex GoogLeNet, the recognition accuracy rate reached 96.5%; Zhao *et al.* [16] used electromagnetic vibration platform to arrange corn kernels in batches, built a visual classification system based on Faster R-CNN to achieve better recognition and pointed out that the smaller and faster model, double-sided detection of corn kernels need to be further studied; Ni *et al.* [17] designed a double-sided corn automatic detector, separated two adhesive kernels based on K-means clustering of curvature and embedded ResNet into corn grain evaluation system. The results showed that the prediction accuracy of 408 test images reaches 98.2%. Altuntas *et al.* [18] identified haploid and diploid maize seeds through transfer learning. By comparing with AlexNet, VGGNet, GoogleNet and ResNet, the most effective VGG-19 realized rapid and low-cost classification. In the detection of rice, Kiratiratanapruk *et al.* [19] found that the classification accuracy of InceptionResNet V2 was the highest after comparing VGG16, Xception, Inception V3 and InceptionResNet V2. At the same time, the research found that appropriate reduction of image size and redirection of rice seeds can improve the detection effect. To distinguish falsely marked rice, Gilanie *et al.* [20] designed an 18-layer CNN rice classification model, which had better classification effect than VGG19, ResNet50 and GoogleNet. In the detection of wheat unsound kernels, ZHANG *et al.* [21] constructed the ResNet50 by adding attention mechanism to the residual networks (18, 34, 50, 101) of different depths. The average recognition rate of 6 types of wheat kernels reached 96.5%; Zhang [22] constructed the ResNet with only three blocks, added a pooling operation with step size of 2 between residual blocks, made the downsampling independent, and used a 5 × 5 convolution layer to ensure that the input and output have the same number of channels, instead of the dotted connection. The recognition accuracy of 7 types of wheat kernels reached 96.3%. Zhu *et al.* [23] collected sound and broken kernels and established a wheat sound kernels detection system based on AlexNet with an average recognition rate of 96.67%. It can be seen that the grain crop recognition method based on deep learning has become the mainstream algorithm.

In practice, the detection of wheat unsound kernels based on deep learning is rare. The main reason is not only that the classification network of wheat unsound kernels still stays in a simple classification model, but also the image acquisition equipments and segmentation methods seriously restrict recognition efficiency. The mainstream image acquisition methods are divided into three ways: the grains fall down with oscillation [24], the single kernel is fixed in an independent hole [22]–[25], the grains are separated with oscillation and scanned linearly [16]. It can be seen that more unsound kernels detection method depends on equipment, which reduces the difficulty of segmentation, but it requires higher equipment precision and hardware cost. In addition, some scholars abandoned the research of equipment. They placed grains on the image acquisition platform one by one manually, which completely ignoring the adhesion and randomness of grains in actual scenes [21]–[23]. To sum up, combining the dual advantages of equipment and segmentation method, a fast and accurate multi-grain detection model is established, which provides a scientific basis for the quality evaluation in wheat market circulation.

II. MATERIALS AND METHODS

A. IMAGE ACQUISITION PLATFORM

In this study, a simple wheat image acquisition platform was built, as shown in Fig. 1. The platform included industrial camera bracket, electromagnetic vibration meter, tray, light source, industrial camera. The 900mm industrial camera bracket is fixed on the lab bench. At the bottom of it is placed a small electromagnetic vibration meter that vibrates vertically.The model is HTA-3000A, with a frequency of 50Hz and a size of 350 × 400 × 250mm. By controlling the vibration frequency, the uniform distribution of wheat kernels can be promoted; A black tray is fixed above the vibrating meter with a size of 350 × 350mm, which is made of non-reflective melamine to decrease the interference of background; The light source is erected above the tray. 16 led light bars with length of 40cm and 48W are combined into a square light source with 8 horizontal and 8 vertical beams. A square hole is left in the middle to ensure that the visual field of the camera is not obstructed. The above measures can provide sufficient brightness and uniform image background; Next to the light source, Hikvision industrial camera supported MVS software, model is MV-CE100-30GC with 10 million pixels, can adjust the image detail information and present clearly the characteristics of wheat color and shape; Finally, brush is used to clear the wheat kernels outside the visual field around the edge of the tray to prevent the incomplete grains from affecting the recognition result.

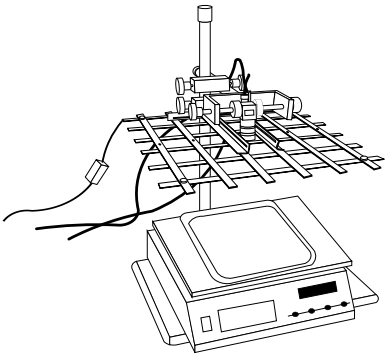


FIGURE 1. Wheat image acquisition platform.

B. IMAGE ACQUISITION

The wheat samples were provided by Xinglong National Grain Reserve in Zhengzhou, Henan Province, including unsound kernel, broken kernel, sprouted kernel, injured kernel, modly kernel and spotted kernel(the spotted kernel is mainly black germ kernel), which were screened by professional quality inspectors. The wheat kernels mixed fully were about 210g, 50g of mixed samples were randomly weighed as test set samples and the remaining 160g were taken as training set samples. We take about 5g of the mixed sample of 160g each time, put it on the tray, turn on the vibrating meter to make the samples uniformly distributed and then collect the images. After the first round of collection, we repeat the above operations and finally collect 100 mixed images. In order to fully mine the characteristics of wheat, we classify manually the mixed samples and take 5 images of each wheat sample. There are a total of 130 images as the original images for establishing the training set. For 50g mixed samples, we obtained a total of 13 mixed type images for establishing the test set through the same operation as above. Original images of wheat mixed kernels are shown in Fig 2.

C. IMAGE SEGMENTATION

In deep learning, the detection of wheat unsound kernels is based on single kernel recognition, so the original image must be single-grained. It can be seen from Fig. 2 that the adhesive kernels inevitably still exist widely, but the number of adhesive kernels and the difficulty of segmentation have been greatly reduced. On this basis, we need to combine the segmentation algorithm for further processing.

In the task of classification, counting of agricultural products such as wheat and corn, it is often necessary to combine the adhesion segmentation algorithm. The single segmentation methods based on threshold, edge detection, region, clustering, morphology and active contour, can not completely segment the adhesive kernels, but they often serve as an intermediate means and a preprocessing method [26]. The segmentation method based on watershed is easily affected by image noise, local irregularities and subtle gray changes on the grain surface, which results in over-segmentation [27]. Concave-based segmentation methods are generally summarized as three steps: concave detection, concave matching and concave segmentation. Concave detection often uses curvature, angle, convex hull, chain code, template search to obtain the location of concave, then we adopt concave matching, directional corrosion, circle fitting to segment, which is extremely suitable for the segmentation of adhesion kernels [28], [29].

On the basis of the original image, the segmentation algorithm of adhesive wheat with two-kernels based on concave-mask is adopted.

The first part is image preprocessing. The original image is gray-processed and in order to eliminate the low-frequency noise in the image, the bilateral filtering method is adopted to highlight the edge and texture characteristics of wheat; Because of the large field of view, it is difficult to illuminate the light evenly. The ideal segmentation effect can not be achieved only by fixed threshold, so the adaptive threshold method is used for binarization; The edge of binary image is smoothed by morphological open-close operation to separate wheat targets from the background.

The second part is the segmentation of adhesive kernels. By setting the threshold of wheat area, two or more adhesion kernels are screened out. Because the shape of wheat kernels is elliptical and the shape of defect area formed between adhered kernels is simple, so we choose the convex hull method to determine the concave position. As shown in Fig. 3, the area of adhered wheat is enlarged and displayed. The convex hull location is the smallest circumscribed polygon area of adhered wheat, which is surrounded by green line. The contour of adhesive wheat is surrounded by blue line. Subtracting the contour area of adhesive wheat from the convex hull area is the defect area. By judging the number of defects in the area, we can extract the two parts with the largest defect area, determine the concave position and connect the two parts with red line to realize the segmentation of two adhesive kernels.

The third part is cropping single kernel wheat image. The single kernel wheat images without adhesion or after adhesion segmentation are mainly used as the mask images. The original wheat image can be segmented through mask processing. Then, the best angle, height and width of wheat segmentation are found by rotating the smallest external rectangle. Finally, the wheat images are cropped into many single kernel images. As shown in Fig. 4.

The fourth part is to abandon the redundant contour. It can be seen from Fig. 5 that the cropped single-kernel wheat images have more or less redundant parts, mainly because the wheat is close to each other and the minimum rectangular cropping method cannot completely avoid neighboring wheat. The main method is to extract the largest contour from the binary image as a mask, remove the redundant contour, mask the cropped single-kernel wheat image and save them. As shown in Fig. 6.

By croping and segmenting 130 original images, the segmentation results of various wheat kernels are shown in Table 1.

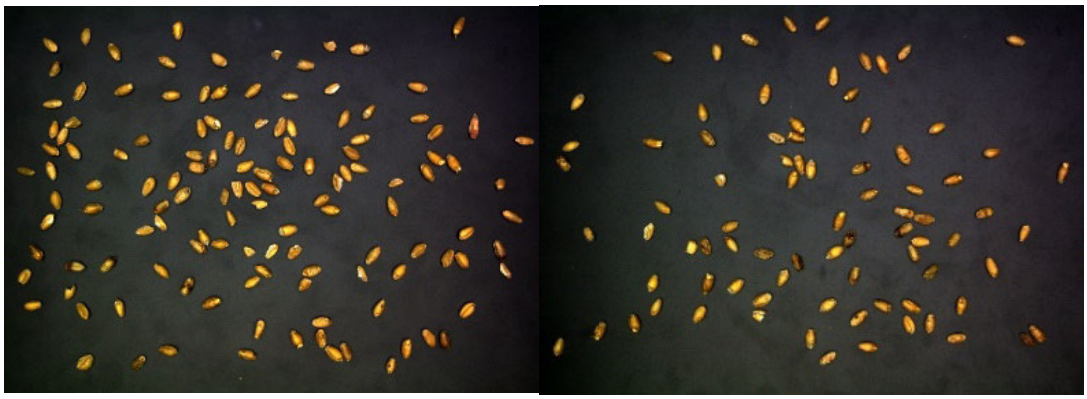


FIGURE 2. original images of wheat mixed kernels.

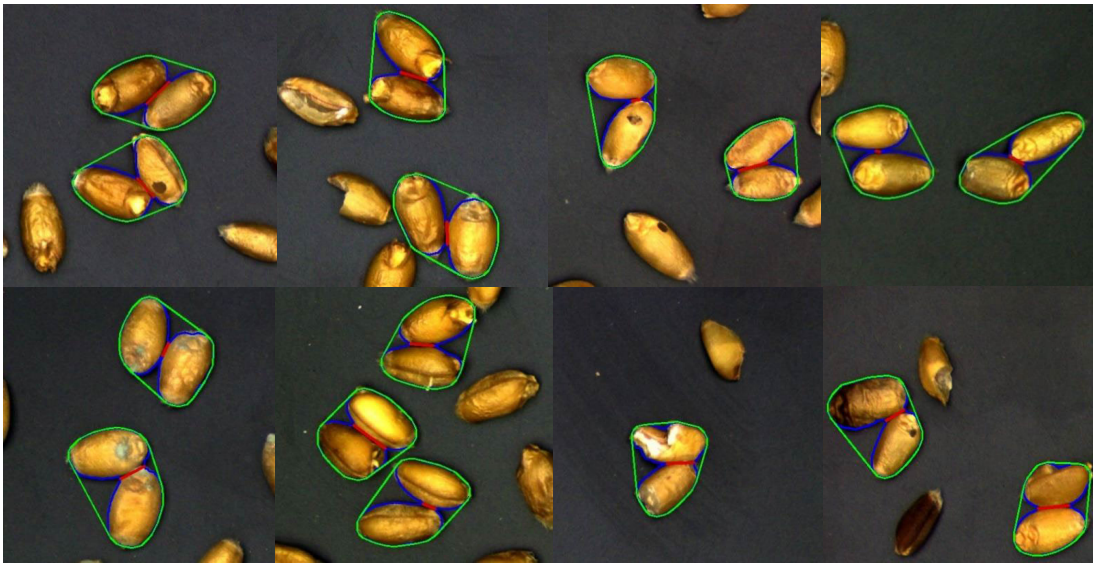


FIGURE 3. Segmentation images of adhesive wheat kernels.

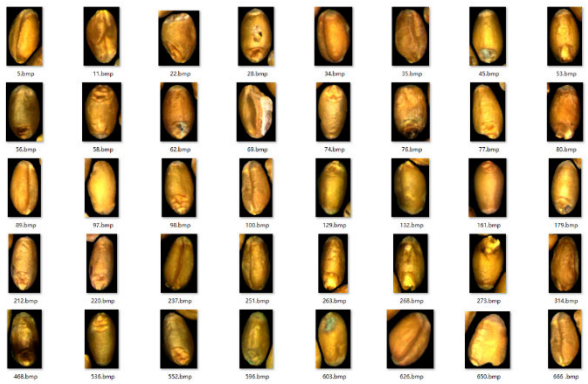


FIGURE 4. Cropped single-kernel image.

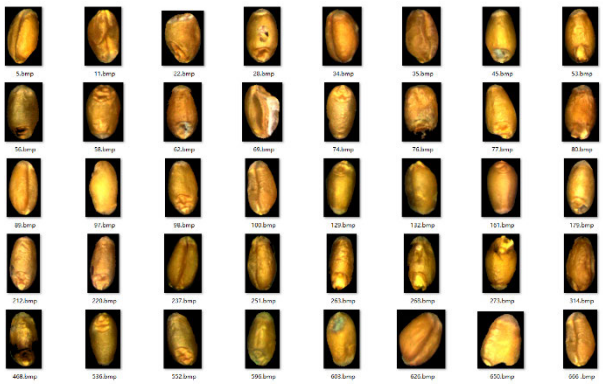


FIGURE 5. Single-kernel image without redundant contou.

To sum up, based on the image acquisition platform, combined with the segmentation method of adhesive wheat with two-kernels based on concave-mask, the error segmentation rate is 0.93% in a total of 9988 wheat images.

D. MAKING DATA SETS

For all kinds of wheat kernel images, we randomly select 800 pieces of each kind for a total of 4800 wheat images. In order to prevent network over-fitting, improve generalization ability and apply to more scenes, the image is enhanced by increasing brightness after rotating 180, adding noise and reducing brightness after mirroring [30]. Each type of wheat image is expanded to 3200 for a total of 19200 wheat images. The data sets are divided according to the ratio of training set to verification set of 8:2. A total of 15360 training

TABLE 1. Wheat kernels image segmentation results.

Category	Number of effective segmentation	Number of false segmentation	Rate of false segmentation	Reason of false segmentation
Sound	2827	93	0.93%	Circular adhesion of multi-sprouted; The area of two adhered-injured is less than that of single-sound; Error in concave detection of single-damaged; The black spots of spotted are lost after binarization.
Sprouted	1422			
Injured	1193			
Moldly	1606			
Broken	1778			
Spotted	1069			

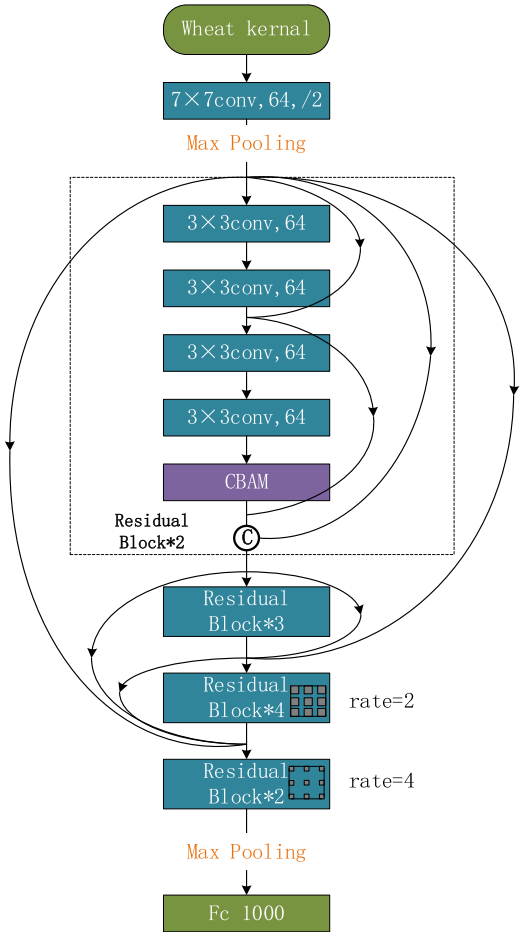


FIGURE 6. Res24_D_CBAM_Atrous model.

set images and 3840 verification set images are obtained. Through the above segmentation method, 13 original images collected from 50g mixed wheat samples are segmented and 1113 wheat images are obtained as test sets.

III. DEEP LEARNING MODELS

A. MODELS AND METHODS

In this paper, the typical GoogleNet, ResNet, DenseNet and ShuffleNet models are compared. The advantages and disadvantages of their respective network frameworks are combined to improve the ResNet and finally the optimized model for wheat unsound kernels detection is obtained.

The advantage of GoogleNet lies in the proposal of Inception module, which greatly increases the number of network channels. it uses sparse connection to separate channels, increases the network width and reduces the number of parameters. In order to increase the receptive field, a large-size convolution kernel is used. Convolution kernels of 1×1, 3×3, and 5×5 are used for dense connection to fuse multi-scale information. The disadvantages of this model are that the Inception module is limited to high-level applications and the underlying structure remains unchanged [31].

The advantage of ResNet lies in the proposal of residual module, which mainly builds deep network by stacking different residual blocks. it extracts abundant features; Skip connection is adopted to avoid model degradation caused by too deep network. 1×1 convolution is added to realize channel dimension reduction. BN is introduced to accelerate the training. The main disadvantages are its limited ability to capture multi-scale features and its downsampling with step size of 2 and 1×1 convolution resulting in the loss of feature information and long training time [32].

The advantage of DenseNet lies in the proposal of dense connection. The multiple concatenations in dense block make the best use of the feature map, which makes the accuracy comparable to ResNet. The 1×1 convolution guarantee network in transition layer is narrow and the sequence of BN, Relu and Conv greatly reduces the parameter quantities. The main disadvantages are that the dense connection may bring redundancy and the video memory is too large [33].

The advantage of ShuffleNet lies in the proposal of Channel shuffle, which reduces the amount of computation and enriches feature learning. It uses 1×1 group convolution, 3×3 depth separable convolution to replace the original convolution and concatenation to replace the original Add, which greatly reduce the parameters. The average pooling with convolution kernel of 3×3 and step size of 2 is used to replace the downsampling to increase the feature transfer. The guiding principle of ShuffleNet V2 indicates that the number of input and output channels must be balanced, the degree of network fragmentation and the use of Relu and Add should be reduced [34].

Park *et al.* [35] combined attention mechanism SENet with ResNet, then added SE to different depth layers of ResNet (18/34/50/101/152), which was suitable for 7 classification of electrocardiogram signals. Miao *et al.* [36] based

on ResNet29, introduced gating strategy into residual unit, which changed feature fusion from adding the same weight to weighted summation and pointed out that the performance improvement brought by gating strategy decreased with the increase of model complexity. Tao *et al.* [37] proposed a fragmented multi-scale feature fusion network IX-ResNet with reference to the Inception. It used $1\times1, 3\times3, 5\times5$ group convolution to fuse multi-scale information and promoted the performance with 1×3, 3×1, 1×5, 5×1 convolution. The Res2Net proposed by Gao *et al.* [38] inserted more hierarchical residual connection into the residual unit, increased the receptive field, which realized feature fusion at a finer granularity level. Lu *et al.* [39] proposed a ResNet embedded with pyramid-shaped atrous convolution. It used atrous convolution with rates of 1, 3 and 5 in the bottom residual block to fuse multi-scale information and improve classification performance. Woo *et al.* [40] added SE and CBAM to each residual block of ResNe18, ResNet34, ResNet50, ResNt101. The comparison shows the model with CBAM higher accuracy.

B. MODELS OPTIMIZATION

Obviously, the model is optimized mainly by changing depth and width, multi-scale feature fusion, adding attention mechanism and changing topological structure.

For the depth of network, it is considered that the wheat unsound kernel has few types and simple features. There must be redundancy in feature extraction of ResNet, so a 24-layer ResNet is designed. The number of residual modules is 2, 3, 4 and 2, respectively.The model is named Res24; In terms of width, dense connection is introduced into the first three layers and the concatenation is used to splice output of each layer with that of dense connection. Because the number of output channels of each layer is consistent with input channels of the next layer, the dashed line of skip connection does not need upgrade dimension by 1×1 convolution, instead, it is replaced by the average pooling with convolution kernel of 3×3 and step size of 2. The sequence of Conv, BN and Relu is changed to BN, Relu, Conv. The model is named Res24_D; Considering the small characteristics of unsound kernels, CBAM is added to the last Block of each layer. The model is named Res24_D_ CBAM; At last, the atrous convolutions with rates of 2 and 4 are added to the block in the third and fourth layers and the step size of the fourth layer is set to 1. The receptive field of 7×7 is converted to 14×14. The model is named Res24_D_ CBAM_Atrous. The main frame diagram of the improved Res24_D_CBAM_Atrous is shown in Fig. 6.

IV. RESULTS

A. MODELS AND METHODS

The experimental software and equipment configuration parameters are as follows: Intel (R) Xeon (R) E5-2620 V3 2.40GHz, two NVIDIA Telsa T4 GPU, two 16GB RAM, Ubuntu 18.04, pytorch1.7.1, Python3.6.13 and CUDA10.0.

B. EVALUATION INDEX

The evaluation indexes selected in this experiment are as follows: in the training process, the accuracy of the current model is expressed by Accuracy. The Loss function curve indicates the difference between the predicted and the actual. In the test set, the Precision indicates the proportion of correctly predicted positive samples to actually predicted positive samples. The Recall rate indicates the proportion of correctly predicted positive samples to actual positive samples. F1 value shows the harmonic average of precision and recall rate. Macro avg represents the arithmetic average of performance indicators of each category. Support is the number of times the real label appears. Confusion Matrix can directly display the final classification results of the model. Time means the prediction time for the same batch of unsound kernels under the same conditions.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

where TP (True Positive) stands for correctly predicting positive samples as positive, FP (False Positive) stands for incorrectly predicting negative samples as positive, TN (True Negative) stands for correctly predicting negative samples as positive, and FN (False Negative) stands for incorrectly predicting positive samples as negative.

C. MODELS AND METHODS

In this experiment, we compare six models, such as Res18, Res34, Res24, Res24_D, Res24_D_CBAM and Res24_D_ CBAM_Atrous. During training, the image size is uniformly scaled to 224×224. The batch size is 16. The number of iterations is 30 and the learning rate is 0.001.

The training results is shown in Fig. 7.

As can be seen from Fig. 7, for the Res18, Res24 and Res34 models, the convergence speed of the training loss function is roughly the same and slightly faster than the other three improved models, but in the verification set, the accuracy is significantly lower than the three improved models. Among the iterations of different epochs, the accuracy of Res18 and Res34 models is the lowest, even the accuracy of Res18 is extremely unstable. However, the accuracy of Res24 fluctuates very little and it shows the same accuracy as Res34 in shallow network. In the three improved models, Res24_D, Res24_D_CBAM and Res24_D_CBAM _Atrous, the convergence speed of loss function decreases slightly with the increase of the model complexity, but the accuracy increases with continuous optimization and the Res24_D_CBAM_Atrous model performs best.

In the test set, the classification evaluation indexes of different ResNet are shown in Table 2.

Based on more feature extraction layers, the performance of Res34 is slightly higher than that of Res34, but the

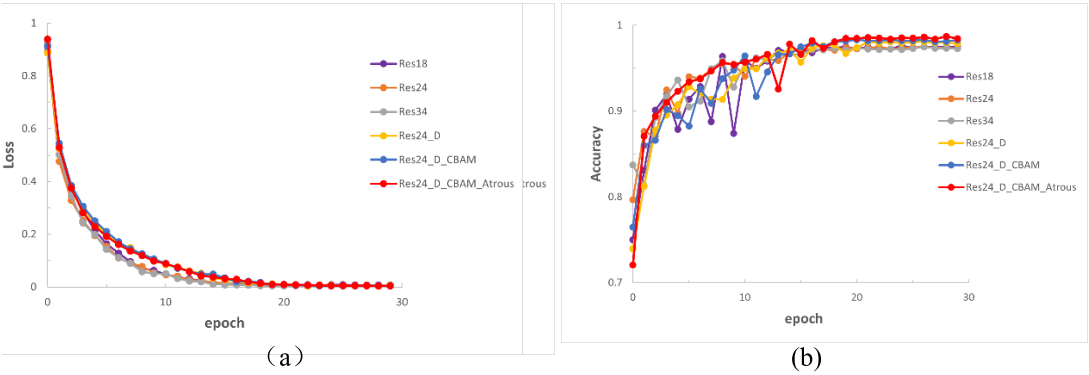


FIGURE 7. Training results:(a)Training loss;(b) Verification accuracy.

TABLE 2. Classification performance evaluation of different ResNet.

Model	Accuracy	Precision	Recall	F1	Time(s)
Res34	0.91	0.91	0.91	0.91	698
Res18	0.90	0.90	0.92	0.91	437
Res24	0.92	0.91	0.93	0.92	451
Res24_D	0.92	0.92	0.93	0.93	486
Res24_D_CBAM	0.93	0.93	0.94	0.93	453
Res24_D_CBAM_Atrous	0.94	0.94	0.95	0.94	478

classification time is greatly improved. It means Res34 has redundancy. Therefore, the Res24 is selected for improvement. After three times of optimization, the average indexs of wheat classification, Accuracy, Precision, Recall and F1 reach 94%, 94%, 95% and 94%, respectively, which is increased by 3%-4% based on Res34. The prediction time is reduced by 220s, which indicates that the proposed Res24_D_CBAM_Atrous has better generalization ability and greatly improves the overall performance.

Classification performance evaluation and confusion matrix of various wheat based on Res24_D_CBAM_ Atrous are shown in Table 3 and Fig. 8.

To sum up, in the test set of 1113 wheat kernels, the improved Res24_D_CBAM_Atrous shows that the F1 values of all kinds of wheat are above 92% and the F1 values of moldy kernels are the highest, reaching 96%. For the actual 206 broken kernels, 201 kernels are correctly detected. Only 5 broken kernels are misjudged as sound kerenls and spotted kernels, but the predicted broken kernels reach 216, so the recall value is greater than precision value; For the actual 149 injured kernels, 142 kernels are correctly detected. The main error source is that some insect-eroded holes of injured kernels are located at the edge of wheat. They are difficult to distinguish and misjudged as sound kernels, but the same number of other kinds of wheat is misjudged as injured kernels, so the recall value and precision value are the same. For the actual 126 moldy kernels, 120 kernels are correctly detected. The main error source is that the area of moldy parts is small and confused with sound kernels, so the precision value is similar to the recall value. For the actual 390 sound

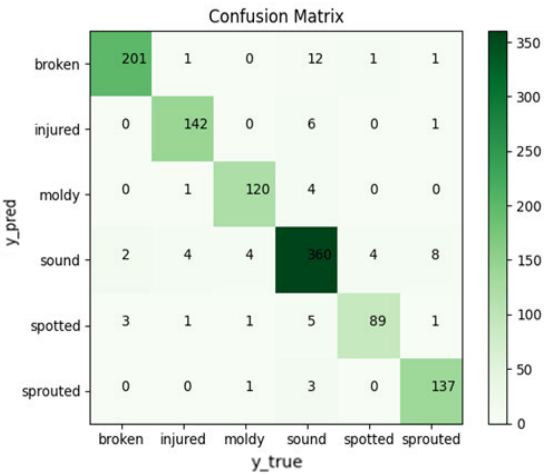


FIGURE 8. Classification confusion matrix based on Res24_D_CBAM_Atrous.

kernels, 360 kernels are correctly predicted, 12 sound kernels are misjudged as broken kernels because of their small area. 6 sound kernels have spots, irregular contours on the surface, so they are misjudged as injured kernels. 4 sound kernels' top hairs are misjudged as mildew. 5 sound kernels have black spots, so they are misjudged as spotted kernels. 3 sound kernels are misjudged as sprouted kernels because of irregular bulges on the embryo. For the actual 94 spotted kernels, 89 kernels are correctly detected, but other kinds of wheat are misjudged as spotted kernels in varying degrees,

TABLE 3. Classification performance evaluation based on Res24_D_CBAM_Atrous.

Category	Precision	Recall	F1	Support	Time(s)
Broken	0.93	0.98	0.95	206	
Injured	0.95	0.95	0.95	149	
Moldy	0.96	0.95	0.96	126	
Sound	0.94	0.92	0.93	390	478
Spotted	0.89	0.95	0.92	94	
Sprouted	0.97	0.93	0.95	148	

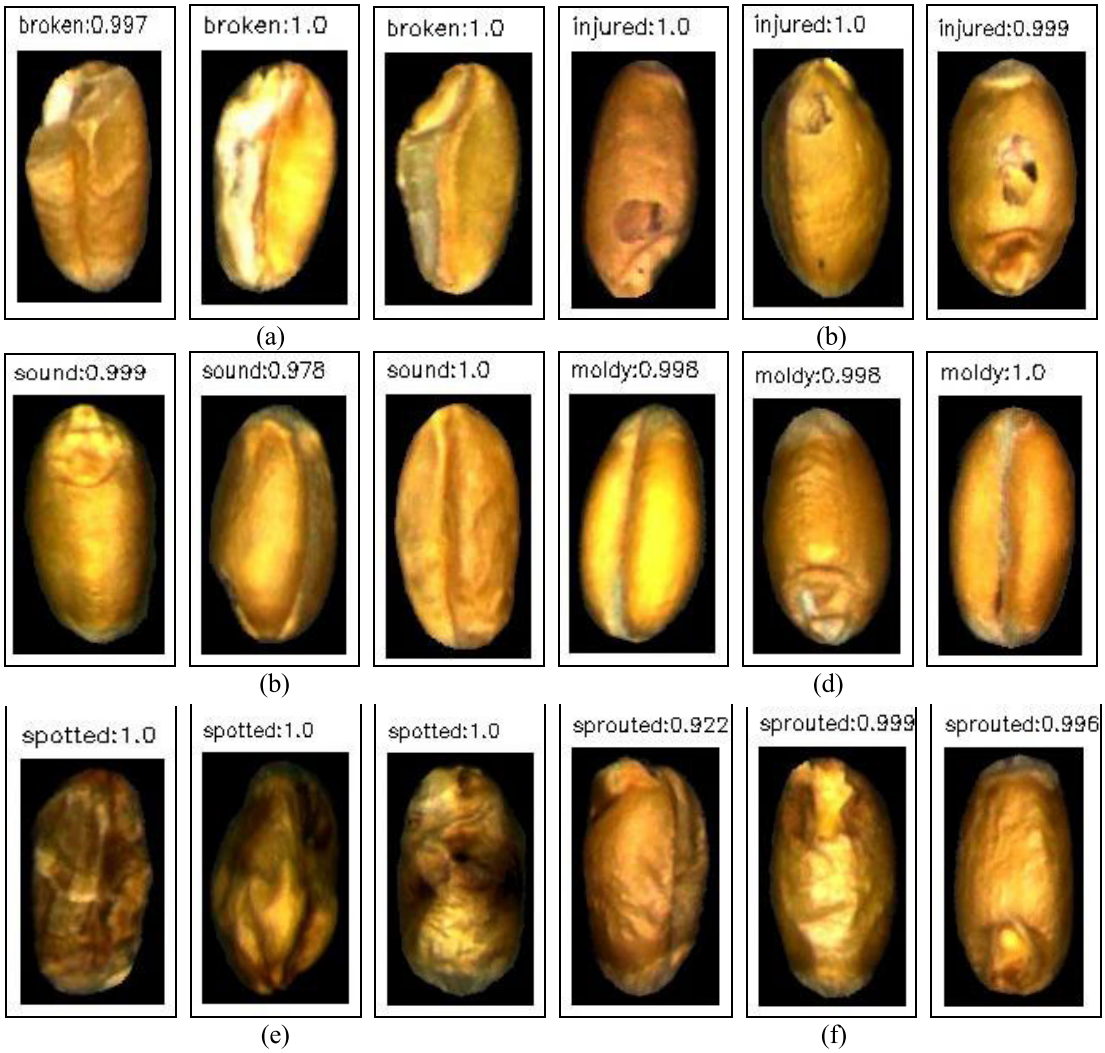


FIGURE 9. Types and probabilities of wheat prediction: (a) broken kernel; (b) injured kernel; (c) sound kernel; (d) moldy kernel; (e) spotted kernel; (f) sprouted kernel.

so the precision is lowest. For the actual 148 sprouted kernels, 137 kernels are correctly detected, 8 sprouted kernels are bulged in the embryo and misjudged as sound kernels, but the predicted sprouted kernels reach 141 kernels, so the precision value is greater than recall value.

In the results based on Res24_D_CBAM_Atrous, the softmax is used to change the pixel value of the output feature map into a probability distribution. The maximum probability value is indexed as the prediction probability of wheat kernel. Finally, the predicted types and probabilities of wheat kernels are written into the image. 4 images of each kind of wheat are displayed. As shown in Fig. 9.

Comparing Res24_D_CBAM_Atrous with advanced models in this field, the results are shown in Table 4.

TABLE 4. Classification performance evaluation of different models.

Paper	Model	Size	Accuracy	Precision	Recall	F1
Zhang et al	ResNet50+CBAM	224×224	0.88	0.88	0.88	0.88
Huang et al	GoogLeNet	224×224	0.85	0.84	0.87	0.85
Altuntas et al	VGG19	224×224	0.81	0.80	0.84	0.81
Kiratiratanapruk et al	InceptionResNetV2	299×299	0.91	0.90	0.93	0.91
Zhu et al	AlexNet	224×224	0.85	0.84	0.86	0.85
Our method	Res24_D_CBAM_Atrous	224×224	0.94	0.94	0.95	0.94

V. CONCLUSION

In the acquisition and market circulation of wheat, it is a general trend to use computer vision method to detect wheat unsound kernels, but this method is still in the experimental stage. The main reason is that it is difficult to solve the problem of adhesion segmentation of wheat unsound kernels. The use of equipment alone to realize single kernel of wheat increases the cost and it has many difficulties. In view of the above situation, an image acquisition platform of wheat unsound kernel is built. The wheat kernels are separated as more as possible by means of vibration. Then we adopt the segmentation algorithm of adhesion wheat with two-kernels based on concave-mask. The concave position is determined by the defect area formed between the adhesive kernels, then the masking segmentation is performed. By combining the equipment and algorithm, the segmentation error rate is 0.93% in a total of 9988 wheat kernels, which meets the recognition requirements.

In this paper, there are 15360 training sets, 3840 verification sets and 1113 test sets of wheat images. By comparing the advantages of GoogleNet, ResNet, DenseNet, ShuffleNet, IX-ResNet, Res2Net, combining with the simple characteristics of wheat unsound kernel, it is found that the depth, width, downsampling mode, attention mechanism, convolution mode and receptive field size of the model are closely related to the accuracy and prediction speed. The Res34 generates too many similar feature in dimension upgrading, resulting in redundancy. The improved Res24 not only reduces redundancy and computation, but also slightly improves classification performance. The introduction of dense connections in Res24_D increases the width of the network and enables the model to make full use of the shallow features. The reason why dense connections are not introduced in layer 4 is that too many dense connections will lead to network redundancy. In the order of BN, Relu and Conv, the BN carried out firstly can also reduce the amount of algorithm calculation. The average pooling with 3×3 convolution and step size of 2 can reduce the loss of characteristic information in the down sampling. The classification of wheat unsound kernel is based on small features such as holes and sprouts. The introduction of attention mechanism in Res24_D_CBAM can effectively enhance the ability of feature extraction. The atrous convolution and 14×14 feature output introduced in Res24_D_CBAM_Atrous increase the feature receptive field, which maximizes the efficiency of the model. In addition, by comparing with the advanced models in this field, our model is significantly superior in various evaluation indexes.

Obviously, the method proposed in this paper can realize the detection of wheat unsound kernels completely and efficiently, which has practical application value, but it still faces many difficulties when it is really applied to the market. There are mainly: the automation level of testing equipment needs to be improved from feeding, segmenting, classifying, calculating and unloading and it must be combined with multi-view image acquisition to improve the recognition accuracy. The image acquisition scene should be strictly fixed. A larger unsound kernel data set should be constructed to improve the standardization level; In-depth combination of traditional segmentation, semantic segmentation, instance segmentation to segment more complex adhesion situations; On the basis of classification, the detection standard of unsound kernel rate of wheat based on mass ratio should be established, which provides scientific basis for classification and grading of wheat.

REFERENCES

[1] S. Sun, X. Yang, X. Lin, G. F. Sassenrath, and K. Li, "Winter wheat yield gaps and patterns in China," *Agronomy J.*, vol. 110, no. 1, pp. 319–330, Jan. 2018.

[2] C. B. Singh, D. S. Jayas, J. Paliwal, and N. D. G. White, "Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging," *J. Stored Products Res.*, vol. 45, no. 3, pp. 151–158, Jul. 2009.

[3] J. G. A. Barbedo, E. M. Guarienti, and C. S. Tibola, "Detection of sprout damage in wheat kernels using NIR hyperspectral imaging," *Biosyst. Eng.*, vol. 175, pp. 124–132, Nov. 2018.

[4] S. Liu, X. Tan, C. Y. Liu, C. L. Zhu, and W. H. Li, "Recognition of fusarium head blight wheat grain based on hyperspectral data processing algorithm," *Spectrosc. Spectral Anal.*, vol. 39, no. 11, pp. 3540–3546, Nov. 2019.

[5] M. Masiello, S. Somma, A. Susca, V. Ghionna, A. F. Logrieco, M. Franzoni, S. Ravaglia, G. Meca, and A. Moretti, "Molecular identification and mycotoxin production by alternaria species occurring on durum wheat, showing black point symptoms," *Toxins*, vol. 12, no. 4, p. 275, Apr. 2020.

[6] F. N. Chen and F. Cheng, "Defective kernel detection using a linear colour CCD," *Imag. Sci. J.*, vol. 61, no. 4, pp. 361–368, May 2013.

[7] F. N. Chen, P. L. Chen, K. Fan, and F. Cheng, "Hyperspectral reflectance imaging for detecting typical defects of durum kernel surface," *Intell. Automat. Soft Comput.*, vol. 24, no. 2, pp. 351–357, Jun. 2018.

[8] H. Liu, Y. Q. Wang, X. M. Wang, D. An, and Y. G. Wei, "Study on detection method of wheat unsound kernel based on near-infrared hyperspectral imaging technology," *Spectrosc. Spectral Anal.*, vol. 39, no. 1, pp. 223–229, Jan. 2019.

[9] R. Choudhary, S. Mahesh, J. Paliwal, and D. S. Jayas, "Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples," *Biosyst. Eng.*, vol. 102, no. 2, pp. 115–127, Feb. 2009.

[10] L. Feng, S. Zhu, F. Liu, Y. He, Y. Bao, and C. Zhang, "Hyperspectral imaging for seed quality and safety inspection: A review," *Plant Methods*, vol. 15, no. 1, p. 91, Aug. 2019.

[11] M. Guo, Y. Ma, X. Yang, and R. W. Mankin, "Detection of damaged wheat kernels using an impact acoustic signal processing technique based on Gaussian modelling and an improved extreme learning machine algorithm," *Biosyst. Eng.*, vol. 184, pp. 37–44, Aug. 2019.

[12] A. Suresh and S. Neethirajan, "Real-time 3D visualization and quantitative analysis of internal structure of wheat kernels," *J. Cereal Sci.*, vol. 63, pp. 81–87, May 2015.

[13] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105507, doi: 10.1016/j.compag.2020.105507.

[14] S. Javanmardi, S.-H. M. Ashtiani, F. J. Verbeek, and A. Martynenko, "Computer-vision classification of corn seed varieties using deep convolutional neural network," *J. Stored Products Res.*, vol. 92, May 2021, Art. no. 101800, doi: 10.1016/j.jspr.2021.101800.

[15] S. Huang, X. Fan, L. Sun, Y. Shen, and X. Suo, "Research on classification method of maize seed defect based on machine vision," *J. Sensors*, vol. 2019, pp. 1–9, Nov. 2019.

[16] C. Zhao, L. Quan, H. Li, R. Liu, J. Wang, H. Feng, Q. Wang, and K. Sin, "Precise selection and visualization of maize kernels based on electromagnetic vibration and deep learning," *Trans. ASABE*, vol. 63, no. 3, pp. 629–643, 2020.

[17] C. Ni, D. Wang, R. Vinson, M. Holmes, and Y. Tao, "Automatic inspection machine for maize kernels based on deep convolutional neural networks," *Biosyst. Eng.*, vol. 178, pp. 131–144, Feb. 2019.

[18] Y. Altuntaş, Z. Comert, and A. F. Kocamaz, "Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach," *Comput. Electron. Agricult.*, vol. 163, pp. 1–11, Aug. 2019.

[19] K. Kiratiratanapruk, P. Temniranrat, W. Sinthupinyo, P. Prempree, K. Chaitavon, S. Porntheeraphat, and A. Prasertsak, "Development of paddy Rice seed classification process using machine learning techniques for automatic grading machine," *J. Sensors*, vol. 2020, pp. 1–14, Jul. 2020.

[20] G. Gilanie, N. Nasir, U. I. Bajwa, and H. Ullah, "RiceNet: Convolutional neural networks-based model to classify Pakistani grown Rice seed types," *Multimedia Syst.*, vol. 27, no. 5, pp. 867–875, Oct. 2021.

[21] W. Zhang, H. Ma, X. Li, X. Liu, J. Jiao, P. Zhang, L. Gu, Q. Wang, W. Bao, and S. Cao, "Imperfect wheat grain recognition combined with an attention mechanism and residual network," *Appl. Sci.*, vol. 11, no. 11, p. 5139, Jun. 2021.

[22] B. Zhang, "Machine vision inseption of wheat apprearance quality based on deep learning," Ph.D. dissertation, Dept. Mech. Elect. Eng., Northwest A&F Univ., ShanXi, China, 2019.

[23] S. P. Zhu, J. X. Zhuo, H. Huang, and G. L. Li, "Wheat grain integrity image detection system based on CNN," *Trans. Chin. Soc. Agricult. Machinery*, vol. 51, no. 5, pp. 36–42, Mar. 2020.

[24] Y. J. Heo, S. J. Kim, D. Kim, K. Lee, and W. K. Chung, "Super-high-purity seed sorter using low-latency image-recognition based on deep learning," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3035–3042, Oct. 2018.

[25] X. Yi, Z. Junxiong, L. Wei, and C. Weiguo, "Multi-objective dynamic detection of seeds based on machine vision," *Prog. Natural Sci.*, vol. 17, no. 2, pp. 217–221, Feb. 2007.

[26] X. K. Yu, Z. W. Wang, and C. L. Zhang, "Edge detection of agricultural products based on morphologically improved Canny algorithm," *Math. Problems Eng.*, vol. 2021, Jun. 2021, Art. no. 6664970, doi: 10.1155/2021/6664970.

[27] S. Tan, X. Ma, Z. Mai, L. Qi, and Y. Wang, "Segmentation and counting algorithm for touching hybrid Rice grains," *Comput. Electron. Agricult.*, vol. 162, pp. 493–504, Jul. 2019.

[28] X. P. Wang, L. J. Yao, H. T. Wen, and J. J. Zhao, "Wolfberry image segmentation based on morphological multiscale reconstruction and concave points matching," *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 2, pp. 212–218, Jan. 2018.

[29] X. Zhang, C. Zhong, T. H. Huang, and T. Ji, "Method of image analysis of aggregate particle group based on gray morphological reconstruction," *J. Building Mater.*, vol. 21, no. 6, pp. 886–891, Oct. 2018.

[30] Z. Lyu, H. Jin, T. Zhen, F. Sun, and H. Xu, "Small object recognition algorithm of grain pests based on SSD feature fusion," *IEEE Access*, vol. 9, pp. 43202–43213, 2021.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[32] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[33] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[34] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[35] J. Park, J.-K. Kim, S. Jung, Y. Gil, J.-I. Choi, and H. S. Son, "ECG-signal multi-classification model based on squeeze-and-excitation residual neural networks," *Appl. Sci.*, vol. 10, no. 18, p. 6495, Sep. 2020.

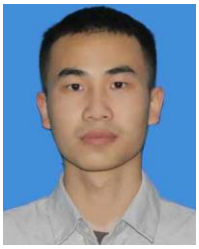
[36] J. Miao, S. Xu, and B. Zhou, "ResNet based on feature-inspired gating strategy," *Multimedia Tools Appl.*, vol. 2021, no. 5, pp. 1–18, Mar. 2021, doi: 10.1007/s11042-01-10802-6.

[37] T. Xue and Y. Hong, "IX-ResNet: Fragmented multi-scale feature fusion for image classification," *Multimedia Tools Appl.*, vol. 80, no. 18, pp. 27855–27865, Jul. 2021.

[38] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.

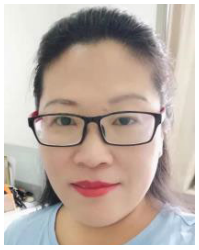
[39] Z. Lu, Y. Bai, Y. Chen, C. Su, S. Lu, T. Zhan, X. Hong, and S. Wang, "The classification of gliomas based on a pyramid dilated convolution ResNet model," *Pattern Recognit. Lett.*, vol. 133, pp. 173–179, May 2020.

[40] S. Y. Woo, J. C. Park, and J. Y. Lee, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, German, Sep. 2018, pp. 3–19.



HUI GAO is currently pursuing the master's degree. He majors in computer technology. He had published two Chinese core journals and accepted one invention patent. His research interests include image processing, deep learning, and grain information.

TONG ZHEN is currently a Professor and a Ph.D. Supervisor. He is also the Deputy Secretary General of the Information and Automation Branch of Chinese Cereals and Oils Association. He had presided over the formulation of 18 national and grain industry standards, compiled three books, published more than 100 articles, and more than 30 articles had been cited by SCI and EI. He is also a member of the Industrial Information Working Committee of the China Instrument and Control Society.



ZHIHUI LI received the Ph.D. degree. She is currently an Associate Professor. She had written many books and six articles had been cited in SCI and EI. Her research interests include intelligent detection and control, and grain information.