Identification and Classification of Wheat Quality Grading Based on Deep Learning

Zhaohui Zhang [*1,2], Zengyang Zuo [*2], Zhi Li [*3], Tianyao Zhang [*1] and Xiaoyan Zhao [*1,2]

[*1] School of Automation and Electrical Engineering, University of Science and Technology Beijing 30# Xueyuan Road, Haidian District, Beijing 100083, China
E-mail: zhangzhaohui@ustb.edu.cn
[*2] Shunde Graduate School, University of Science and Technology Beijing Fo Shan 528399, China
E-mail: zzyzzy981111@163.com
[*3] College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China,
E-mail: lizhi@haut.edu.cn

Abstract. China is one of the world's largest producers of wheat. Because of its storability, wheat has become one of the country's three largest grain reserves. However, the detection methods of wheat quality are mainly manual detection. There are many problems with this identification and classification method. Initially, different identification personnel will produce different identification results. Secondly, the identification process is long, the identification personnel are prone to visual fatigue. Finally, the identification takes a lot of time and money, and the classification results are difficult to reproduce.

Based on the above background, this paper designs a wheat unsound kernel identification method based on YOLO algorithm. Image acquisition is carried out with sound kernel, injured kernel, broken kernel, sprouted kernel, spotted kernel, and foreign material in wheat kernels. And image data enhancement methods such as image cropping, rotation, scaling, and segmentation are used to determine the number of data sets expanded. The deep learning recognition model is used to identify wheat quality categories. The wheat quality detection experiment of YOLOv5 recognition model is realized, and the performance indicators of the model are evaluated.

The recognition results show that the overall recognition rate of YOLOv5 for wheat is about 93%. And the detection can be completed quickly. The average accuracy of YOLOv5 is evaluated. The experimental results show that YOLOv5 is more suitable for the recognition of unsound wheat kernels.

Keywords: Wheat, Unsound Kernels, Deep Learning, YOLOv5

1. INTRODUCTION

According to the wheat quality grade standard, the unsound kernel and the foreign material content in wheat kernels are important factors affecting the quality classification of wheat[1]. My country has large reserves of wheat, and long-term mixed storage of wheat of different qualities can easily lead to qualitative changes such as mold and insects. At present, the main method of wheat quality identification in my country is artificial visual identification[2]. This method is highly subjective, labor-intensive, requires a lot of time and money , and the results are unstable and prone to deviations. Therefore, it is very important to develop a method that can stably identify unsound kernels in wheat.

At present, there are traditional machine vision methods for the identification of unsound kernels. The unsound kernels of wheat are identified by analyzing the appearance differences, color differences or extracting morphological features of wheat[3]. However, the traditional machine vision methods mainly focus on the results. In the identification of wheat varieties and the identification of unsound kernels, the identification results of wheat quality are relatively lacking[4]. The hyperspectral image recognition method, the spectral information of different types of unsound kernels is not the same, and the spectral information of different wheat kernels can be extracted for identification. But the analysis method using full spectrum is too large in data volume and time-consuming, and has poor universality for the identification of unsound wheat kernels[5].

The method used in this paper is mainly based on the deep learning method, using the Pytorch framework in the field of deep learning. Under this framework, the YOLOv5 target detection algorithm is used to identify the unsound wheat kernels, and the recognition effect is tested. Performance indicators were evaluated to determine whether it was suitable for unsound kernels identification in wheat.

2. METHODOLOGY

The overall structure of the system is shown in the "**Fig.1**". The first step is to create a dataset based on the multiple types of wheat images captured by the HD camera, followed by a labeling operation on the images obtained. After completing the production of the data set, training can be carried out according to the parameters configured by yourself. After multiple rounds of training, the corresponding weights are obtained, and the obtained weights are used to test the wheat images outside the data set to check the detection effect. Qualitative analysis of target detection is carried out for the weight file, its performance indicators are analyzed, and the performance of the model algorithm is evaluated.

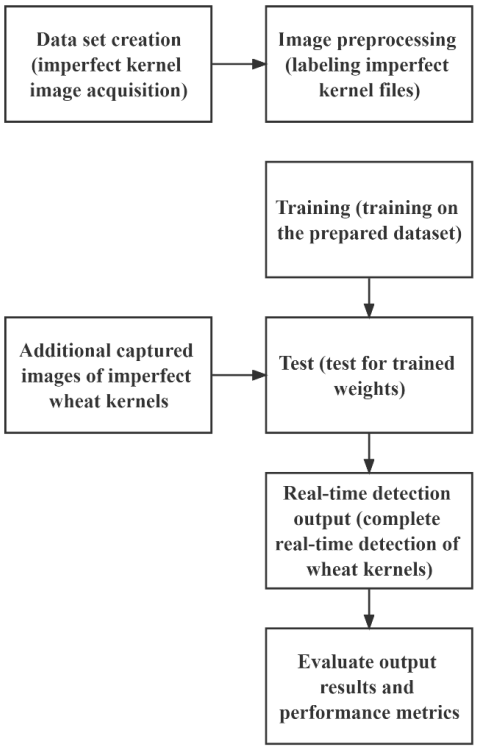


Fig.1 The overall structure of the system**.**

2.1. Data set collection and labeling

The wheat kernels used in this paper are all the wheat kernels of the same year and the same variety collected from the grain depot. The classification is as follows shown in the table 1.

Table.1 Six classes used for identification.

Class number	Class
0	Spotted kernel
1	Sprouted kernel
2	Sound kernel
3	Broken kernel
4	Injured kernel
5	Foreign material

When collecting images, a mobile phone with high pixels and clear picture quality is used. And the wheat is placed on a black base plate with an independent light source, which can highlight the characteristic information of wheat kernels. "**Fig.2**" shows some images of the wheat kernels collected.



Fig.2 Partial wheat kernel images.

Since the collected wheat kernel images are multi-row and multi-column with fixed spacing, in order to obtain a separate wheat kernel image, the method adopted is to divide the image according to the same proportion of the image, and input the number of cropping times of width and height to obtain a separate wheat kernel image. The wheat kernel images and their orderly numbers are shown in "**Fig.3**"



Fig.3 Training dataset.

In order to obtain better training results, the number of data sets is usually required to be sufficient, so that the trained model has strong generalization ability. So data enhancement operations are usually taken for data sets with insufficient number. In order to obtain more wheat kernel data sets, the following data enhancement methods are adopted in this paper, as shown in "**Fig.4**"

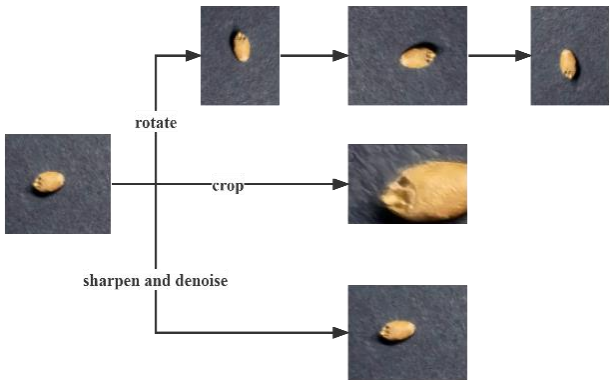


Fig.4 Data augmentation methods.

After completing the acquisition of the dataset, the individual wheat grain images were subjected to a labelling operation, using Colabeler software. Since the data set format required for training is in VOC format, after the data set is marked, export it as an xml file in VOC format, and store it in the folder required for training. The image file corresponding to the file is stored in the corresponding image folder, as shown in "**Fig.5**".

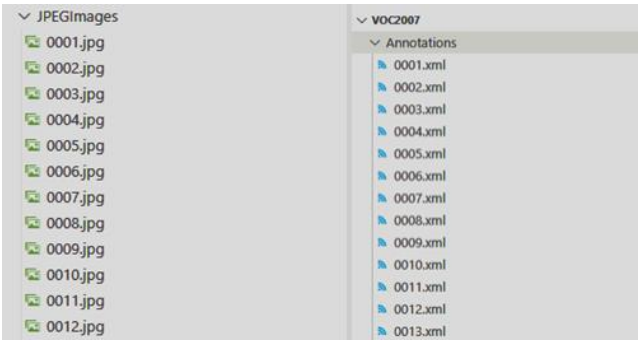


Fig.5 Index file location.

2.2. Model training of yolov5 based on Pytorch

At present, the more mainstream open source deep learning frameworks include Tensorflow, Pytorch, and Darknet. Among them, Pytorch is a python implementation of the torch framework. It is intuitive to operate, easy to implement a flexible network architecture, and has a good visualization effect. Therefore, we choose to build a recognition model under Pytorch. The "**Fig.6**" shows the framework structure of Pytorch[6].

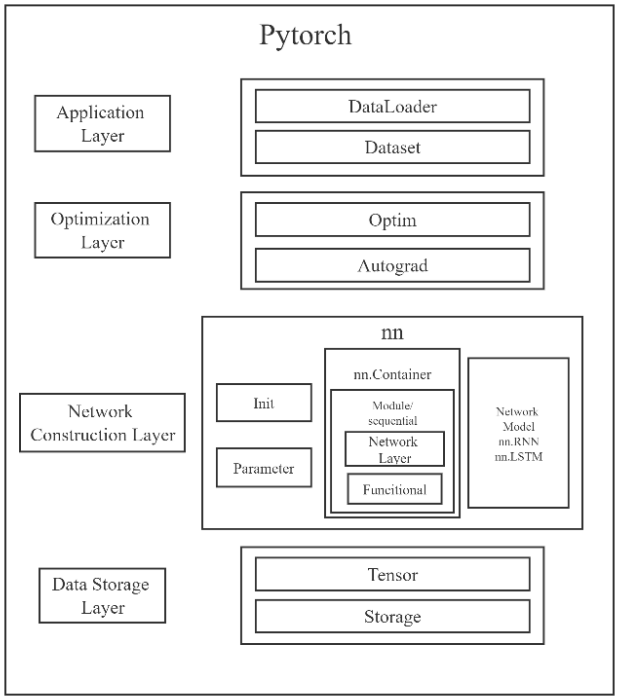


Fig.6 Pytorch Structure diagram.

At present, the mainstream deep learning recognition models for image object detection include Faster R-CNN, YOLO series, and SSD algorithms. These algorithm models have their own advantages, as shown in the following Table 2, which are the advantages and disadvantages of these algorithms[7]. It is necessary to have a good recognition effect in both target recognition and classification, and consider that the detection speed of the model should be able to meet the requirements of real-time detection. Therefore, in this article, the YOLOv5 algorithm is adopted.

Table.2 Comparison of Object Detection Algorithms.

Algorithm	Advantage	Disadvantage
Faster R-CNN	Take the multi-task loss function bounding box	Detection speed is slow
YOLO	The image is divided into grid cells, and the model detection speed is fast	Poor detection of dense objects and small objects
YOLOv2	Use clustering to generate anchor boxes to improve classification accuracy	Using pre-training, transfer is difficult
YOLOv3	Learning from residual network to realize multi-scale detection	The model is complex, and the detection effect of medium and large targets is poor
YOLOv4	A trade-off between detection accuracy and speed	Detection accuracy needs to be improved
YOLOv5	Small model size, low deployment cost, high flexibility, and high detection speed	Performance needs to be further improved
SSD	The algorithm model has strong stability	slower to train

The core idea of the YOLO algorithm is to divide the original image into N×N small squares, and then determine which small square the center of the target object on the image falls in. The preset bounding box of this small square is used to detect and identify the target object. During training, the model is told through the small square that the object A in the image is composed of some pixels around the specific small square where the center is located, then the model searches for all the pixels that satisfy the characteristics of the A object around the specific small square. The model continuously changes the position of the center point in a specific small square to determine the most accurate bounding box containing all the features of object A. Therefore, the YOLO algorithm combines position detection and category recognition into a convolutional network to complete the prediction, that is, you only need to scan the picture once (you only look once) to know the position information and category[8]. The core idea of the YOLO algorithm is shown in the "**Fig.7**":

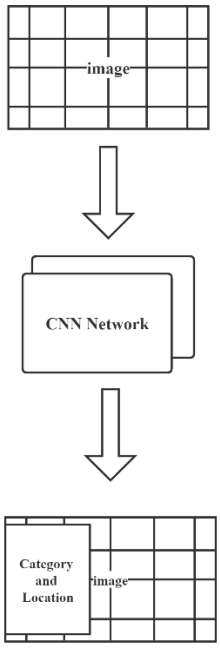


Fig.7 Algorithm principle.

The YOLOv5 used in this article can be divided into three parts, namely Backbone, FPN and Yolo Head.

Backbone can be called the backbone feature extraction network of YOLOv5. According to its structure and the name of the previous Yolo backbone, it is usually called CSPDarknet. The input image will first be feature extracted in CSPDarknet, and the extracted features can be called as the feature layer. It is the feature set of the input image. In the backbone part, we obtain three feature layers for the next network construction, these three feature layers I call it the effective feature layer.

FPN can be called the enhanced feature extraction network of YOLOv5. The three effective feature layers obtained in the backbone part will be feature fusion in this part. The purpose of feature fusion is to combine feature information of different scales. In the FPN part, the effective feature layers that have been obtained are used to continue extracting features. The Panet structure is still used in YOLOv5. We not only upsample the features to achieve feature fusion, but also downsample the features again to achieve feature fusion.

Yolo Head is the classifier and regressor of YOLOv5. Through CSPDarknet and FPN, we have obtained three enhanced effective feature layers. Each feature layer has width, height and number of channels. At this time, we can regard the feature map as a collection of feature points one after another, and each feature point has a number of channels[9]. What Yolo Head actually does is to judge the feature points and determine whether the feature points have objects corresponding to them. Like previous versions of Yolo, the decoupling head used by YOLOv5 is together, that is, classification and regression are implemented in a 1X1 convolution

Therefore, the work done by the entire YOLOv5 network is feature extraction-feature enhancement-prediction of the object corresponding to the feature point. The "**Fig.8**" shows the frame diagram of YOLOv5.

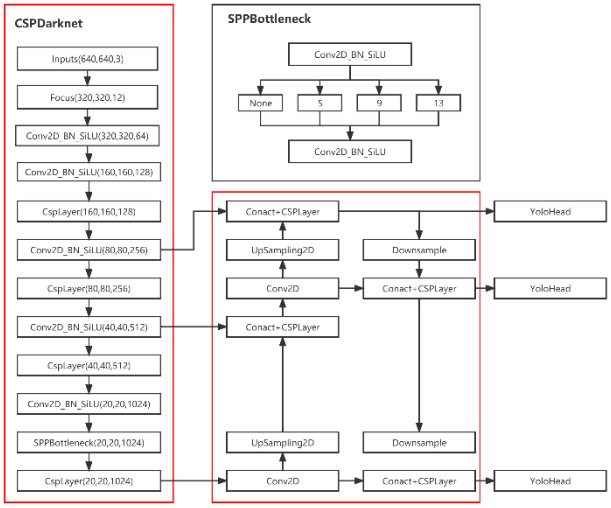


Fig.8 The frame diagram of YOLOv5.

As shown in the figure above, the part marked in red is the most special part of the YOLOv5 structure. YOLOv5 has designed two CSP structures. CSPNet solves the problem of repeating gradient information in network optimization in other large-scale convolutional neural network frameworks. The method is to integrate the gradient changes into the feature map from beginning to end, reducing the parameter quantity and FLOPS value of the model, which not only ensures the inference speed and accuracy, but also reduces the model size.

This training uses the YOLOv5 algorithm to divide the wheat quality categories into 6 categories, namely Sound kernel, Injured kernel, Broken kernel, Sprouted kernel, Spotted kernel, Foreign material. In order to obtain a better training effect, the training samples should be as many as possible. Therefore, this training will divide the data set into 90% training set and 10% test. The training data set is divided as shown in the following Table 3.

Table.3 Training dataset partition.

Class	Training samples	Testing samples	Total samples
Spotted kernel	2057	228	2285
Sprouted kernel	1899	211	2110
Sound kernel	1809	201	2010
Broken kernel	2043	227	2270
Injured kernel	1931	214	2145
Foreign material	1512	168	1680

After determining the model and algorithm, the next step to confirm is the training parameters. Considering the influence of computer computing power, firstly adjust the batch number of images needed to train each

iteration to update the parameters, and modify it to 16. That is, 16 images can be trained each time. Secondly, set the image size input of the network, in order to make the training accuracy of the network more accurate and higher. Set the image resolution of the network to 640×640; modify the categories classes, this training divides the wheat kernels into 6 categories, so the classes are set to 6. Load the pre-training weight YOLOv5_s.pth, the pre-training weight of the model is paired with different datasets are generic because features are generic. The training adopts a combination of freezing training and unfreezing training. In the freezing stage, the backbone of the model is frozen, and the feature extraction network does not change. The memory occupied is small, and only the network is fine-tuned. In the unfreezing phase, the backbone of the model is not frozen, and the feature extraction network will change. The occupied video memory is large, and all the parameters of the network will be changed, which can speed up the training efficiency. The training process is shown in the "**Fig.9**":

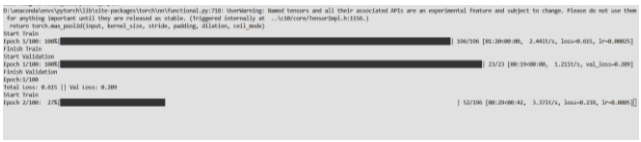


Fig.9 Training process.

At the end of each round of training, corresponding weights will be generated. The main determinant of weight selection is whether the training loss function tends to converge. As shown in the "**Fig.10**", it is the loss value change curve during training. The loss value converges and the training is completed.

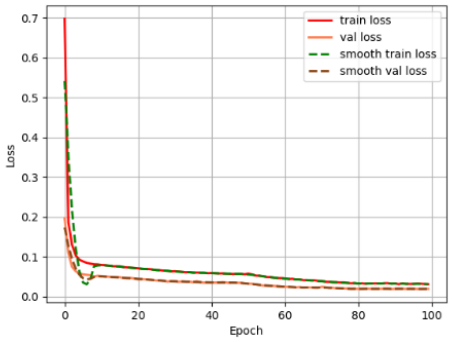


Fig.10 Loss function curve.

3. EXPERIMENT AND RESULTS

After the training is completed, use the converged weights to predict the samples outside the data set. The prediction results are shown in the "**Fig.11**":



Fig.11 Recognition result.

After completing the prediction of unsound wheat kernels, the YOLOv5 algorithm under Pytorch used this time should be evaluated. This paper uses the three criteria of Recall, Precision and mAP to evaluate the model.

The calculation method of Precision and Recall consists of the following parameters, as shown in the following Table 4:

Table.4 The parameters of Precision and Recall.

	Predict positive examples	**Predict Negatives**
True positive example	TP (True Positive)	FN (False Negative)
True negative example	FP (False Positive)	TN (True Negative)

The formula for calculating Precision is in (1).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

TP is an example that the classifier considers to be a positive sample and is indeed a positive sample. FP is an example that the classifier considers to be a positive sample but is not actually a positive sample. And Precision is the part that the classifier considers to be a positive class and is indeed the proportion of the positive class.

The calculation formula of Recall is in (2).

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

TP is an example that the classifier thinks is a positive sample and is indeed a positive sample. FN is an example that the classifier thinks is a negative sample but is not actually a negative sample. And Recall is the part that the classifier thinks is a positive class and is indeed the proportion of the positive class.

Because it is very one-sided to examine the two indicators of Recall and Precision when examining the performance indicators of the model. The mAP value is introduced here. mAP is the average of the AP values of all classes, and AP is the precision and the Recall curve and the area enclosed by the horizontal axis. The following "**Fig.12**" shows the mAP value of this model.

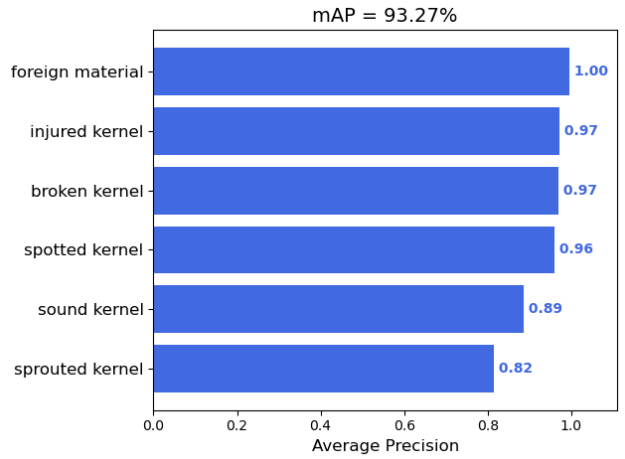


Fig.12 The detection's mAP.

As can be seen from the figure above, foreign materials are best identified at approximately 100%. The sprouted kernels, on the other hand, are the least well identified, at approximately 82%. This is because the foreign materials' features are the most obvious in the dataset, and the size of the annotation box at the time of annotation is more different from the other categories, so they are better identified. The features of sprouted kernels, on the other hand, were less different for the other types of unsound kernels and had average results during training, so the average accuracy was lower. The recognition of all other types of unsound kernels was better. The overall average accuracy of the system was 93.27%, which is positive for the classification and identification of wheat kernels.

4. CONCLUSION

This paper mainly studies the YOLOv5 algorithm under the framework of Pytorch for the detection of unsound wheat kernels. First, a large number of pictures of unsound wheat kernels were collected, and the labeling software was used to create the data set required for training. Secondly, after the collection, the YOLOv5 algorithm was used for training and prediction under the Pytorch framework. Finally, the performance of the trained model was evaluated, and its average accuracy was evaluated. The evaluation results showed that the trained model was suitable for wheat unsound kernel detection. And the recognition effect is fairly good. In conclusion, the recognition model and algorithm studied in this paper have a positive effect on the classification and recognition of wheat kernels.

Acknowledgements

This work was supported by Scientific and Technological Innovation Foundation of Shunde Graduate School (BK19BE020).

REFERENCES:

[1] O. Acar, T. Sanal, H. Köksel. Effects of Wheat Kernel Size on Hardness and Various Quality Characteristics[J]. Quality Assurance and Safety of Crops & Foods,2019,11(5).

[2] Zhou Lei, Zhang Chu, Taha Mohamed Farag, Wei Xinhua, He Yong, Qiu Zhengjun,Liu Yufei. Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method[J]. Frontiers in Plant Science,2020. 022,194

[3] Zhu Yuchen. Target Detection Based on Deep Learning[J]. Journal of Physics: Conference Series,2022,2181(1).

[4] Shivali Amit Wagle,Harikrishnan R,Jahariah Sampe,Faseehuddin Mohammad,Sawal Hamid Md Ali. Effect of Data Augmentation in the Classification and Validation of Tomato Plant Disease with Deep Learning Methods[J]. TS,2021,38(6).

[5] Bird Jordan J., Barnes Chloe M., Manso Luis J., Ekárt Anikó,Faria Diego R.. Fruit Quality and Defect Image Classification with Conditional GAN Data Augmentation[J]. Scientia Horticulturae,2022,293.

[6] Jiang Lizheng, Zhang Zizhao. Research on Image Classification Algorithm Based on Pytorch[J]. Journal of Physics: Conference Series,2021,2010(1).

[7] Jiang Peiyuan, Ergu Daji, Liu Fangyao, Cai Ying, Ma Bo. A Review of Yolo Algorithm Developments[J]. Procedia Computer Science,2022,199.

[8] Qi Jiangtao, Liu Xiangnan, Liu Kai, Xu Farong, Guo Hui, Tian Xinliang, Li Mao, Bao Zhiyuan, Li Yang. An Improved YOLOv5 Model Based on Visual Attention Mechanism: Application to Recognition of Tomato Virus Disease[J]. Computers and Electronics in Agriculture,2

[9] Zhao Ziyu, Yang Xiaoxia, Zhou Yucheng, Sun Qinqian, Ge Zhedong, Liu Dongfang. Real-time Detection of Particleboard Surface Defects Based on Improved YOLOV5 Target Detection[J]. Scientific Reports,2021,11(1).