DROPPING EAR DETECTION METHOD FOR CORN HARVERSTER BASED ON IMPROVED Mask-RCNN
/
基于改进 Mask-RCNN 的玉米果穗损失检测方法

Geng AIJUN [1,2], Gao ANG [1], Yong CHUNMING [3], Zhang ZHILONG [1,2*], Zhang JI [1], Zheng JINGLONG [1]
[1] Shandong Agricultural University, College of Mechanical and Electrical Engineering/ China;
[2] Shandong Provincial Engineering Laboratory of Agricultural Equipment Intelligence/ China;
[3] Shandong Wuzheng Group, Agricultural Equipment Research Institute/ China
Tel: +86-0538-8242500; E-mail: sdauzzl@163.com
DOI: https://doi.org/10.35633/inmateh-66-03

***Keywords**: corn ear；convolution neural network；image recognition；Mask-RCNN*

ABSTRACT
In order to quickly and accurately identify the corn ears lost during the corn harvesting process, a corn ear loss detection method based on the improved Mask-RCNN model was proposed. The lost corn ears in the field were taken as research objects, the images of the lost corn ears were collected and the fallen ears data set was established. The size ratio of the Anchor Box of the area recommendation network was changed by changing the K-means algorithm to reduce the influence of artificial setting intervention. The group convolution was introduced into the residual unit and the channel dimension was divided into 3 equal parts to reduce the model parameters in the basic feature extraction network ResNet. A Convolutional Block Attention Module (CBAM) was introduced to improve the accuracy of the model in the last layer of the ResNet network. Results showed that the average target recognition accuracy of the method on the test set in this study was 94.3%, which was better than that of the previous model, and the average time to recognize a single image was 0.320 s. The proposed method could detect the lost corn ears during the harvesting process under the complicated background, and provide a reference for the corn ear loss detection of the corn harvester.

摘要
为快速准确识别玉米收获过程果穗损失，本文以玉米收获机田间收获掉落的果穗为研究对象，进行图像采集并建立掉落果穗数据集，提出一种基于改进 Mask-RCNN 的果穗损失检测方法。将 Mask-RCNN 深度学习模型引入玉米果穗图像识别中并提出一种优化方法，通过 K-means 算法改变区域建议网络 Anchor Box 尺寸比例以减少人为设置干预影响，在基础特征提取网络 ResNet 中将分组卷积引入残差单元并将通道分为 3 等份以减少模型参数，在 ResNet 网络层最后一层引入卷积注意力模块(CBAM)以提高模型的准确率。结果表明：本文方法在测试集上平均目标识别准确率为 94.3%，优于改进前的模型，识别单幅图像的平均耗时为 0.320 s。所提方法对复杂背景下玉米收获机掉落的果穗有较好的检测效果，可为玉米收获机掉穗损失检测提供参考。

INTRODUCTION
The North China Plain (NCP) is the main grain production area in China and the mechanized harvesting of corn in NCP has been dominated by ear harvesting. Due to the shocks and vibrations in the working process of corn harvesters, it is inevitable for a small number of ears to be omitted. The high corn ear loss rate indicates that there are some problems with the working quality of the corn harvester and the harvester should be inspected and adjusted in time to avoid greater economic losses. Therefore, corn ear loss rate is regarded as an important index to measure the working performance of the corn harvester, and it is of great significance to determine the ear loss rate of corn harvester in real time.

With the development of deep learning, more and more target detection methods have been developed and applied in agricultural production (*Kamilaris et al., 2018*). The role of target detection is to identify targeted targets and non-targeted targets. Compared with traditional algorithms, deep learning algorithms are highly adaptable, easy to migrate, and do not require feature engineering. As a mainstream model, Mask Region Convolutional Neural Network (Mask-RCNN) has been widely applied in the field of agriculture. *Jesus et al., (2020),* adopted the Mask-RCNN model to segment 10 different types of microalgae with an average accuracy of 85%. *Yu et al., (2019),* used Mask-RCNN with ResNet50 as the basic network to visually locate the strawberry picking points, and the average detection accuracy was 95.41%. *Xiong et al., (2020),* used the Mask-RCNN model to segment soybean leaves under a complex background, and used the migration learning method of the VGG network to extract the leaf features to obtain the leaf element classification method, and the average accuracy of the classification was 89.42%. *Li et al., (2019),* proposed a pig crawling behavior recognition algorithm based on Mask R-CNN, introduced a migration learning method to train the ResNet-FPN network, the segmentation accuracy rate of the pig segmentation network model was 94%, and the accuracy rate of the crawling behavior recognition algorithm was 94.5%. In spite of this, there were still few reports on the detection of missing ears in corn harvesting process based on the Mask-RCNN model.

In this paper, the lost corn ears in the working process of corn harvester were taken as the research object and the images of the lost corn ears were collected. The improved Mask-RCNN was adopted to recognize the corn ears of different sizes and shapes. A visual detection method was proposed to detect the lost corn ears in the working process of corn harvesting, which could provide reliable feedback to ensure the consistent performance of the corn harvester.

MATERIALS AND METHODS

IMAGE ACQUISITION AND PROCESSING

Image acquisition

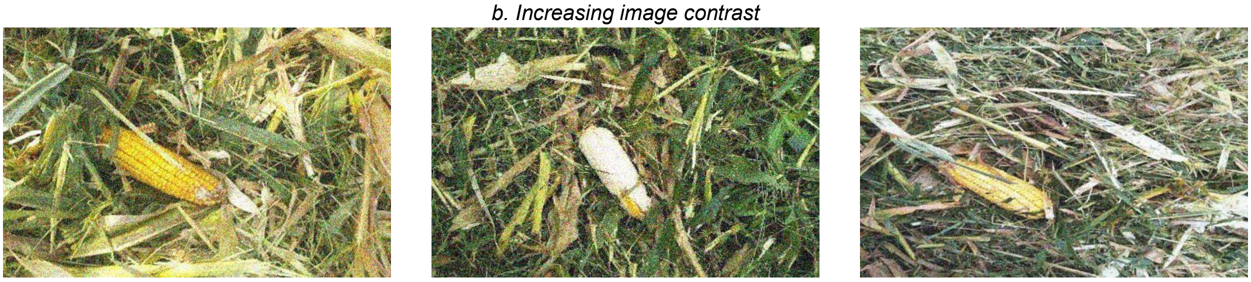
The images of corn ears were collected in October 2019 twice under the conditions of suitable corn harvesting and good weather from the experimental field which was located in Nanqiu village, Feicheng City. The collection tool was an iPhone 6S smartphone with a 12-million-pixel camera which was used to collect 1500 sample images at different angles and shooting distance ranging from 0.3 to 0.5m.

Image preprocessing

The data set was augmented to improve the generalization ability of the training model and prevent the occurrence of overfitting for the small number of images about the data set. The existing data enhancement methods could be classified into three different categories: spatial transformation, such as rotation, flip, etc.; color transformation, such as changing brightness, changing image color, etc.; message deletion, such as random deletion, hide-and-seek (HAS), etc.

a. Original pictures

b. Increasing image contrast



c. Gaussian noises

Fig. 1 - Data enhancement results

In order to facilitate the computer to process the data set, the collected images were uniformly scaled to 720*406 pixels, and the images were processed by means of rotation, brightness enhancement, Gaussian noise, and Hide-and-Seek (HAS) data enhancement (*Singh et al., 2019*). The results of data enhancement were shown in Fig. 1. Finally, there were 6000 images included in the expanded data set. The data set was split into training sets (4800 pieces) and verification sets (1200 pieces) based on the ratio of 8:2.

DETECTION METHOD FOR FALLING CORN EARS

Corn ear detection model based on Mask-RCNN

The flowchart of the corn ear detection model based on Mask R-CNN was shown in Fig. 2. The image data was firstly performed data preprocessing after collection, involving the scaling of images in data set, the data enhancement processing and the conversion of data sets into MS coco format. Then, the Mask R-CNN model was used to train the preprocessed data, and the initial weight and the number of training iteration were set. The corn ear detection was completed by using the images in the test set to test and assess the trained model after the model training was completed.

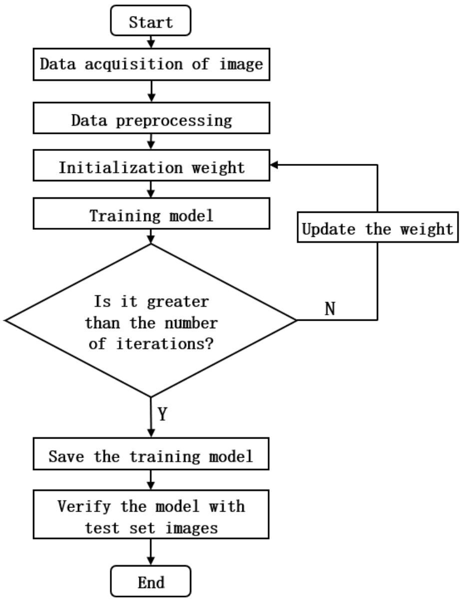


Fig. 2 - Flow chart of corn ear detection model

As an improvement to the Faster R-CNN model, Mask R-CNN consisted of target location, category and segmentation mask prediction. The structure and working process of the algorithm were illustrated in Fig. 3. In the Mask R-CNN model, the input images with length *P* and width *Q* was first scaled to an image with length *M* and width *N*, and then the feature maps were extracted by the convolutional neural network embedded in the backbone network. The feature maps were shared by the region proposal network (RPN) and Mask R-CNN network, and the candidate frame generated by the feature image and RPN was input to ROI pooling layer. The corresponding target features in the shared feature mapping were extracted after pooling processing, and then output to FC layer and FCN layer respectively for target classification and segmentation mask prediction.

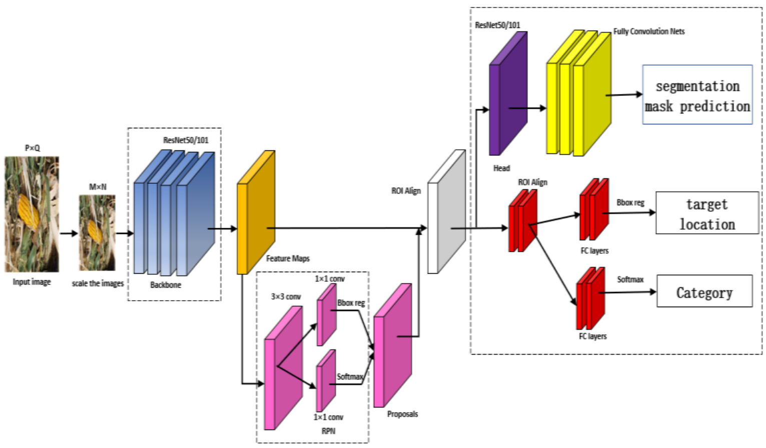


Fig. 3 - Network structure of Mask R-CNN

The convolutional neural network of the backbone network adopted the residual network (ResNet) in this paper *(He al., 2016*). In order to expand the main network at multiple scales, the feature pyramid network (FPN) was introduced (*Lin al., 2016*), and the top and bottom features of the feature pyramid network structure were fused through up-sampling to generate feature graphs of different levels (*Lin al., 2019*).

TEST SOFTWARE AND HARDWARE

The processing platform was a desktop computer, the processor was Intel Pentium G4560, the main frequency was 3.5GHz, and the GPU was GeForce GTX 2080. The operating environment was Ubuntu (18.04) Linux system, the machine learning library was Tensorflow1.5.0, and the parallel computing architecture was CUDA 9.0.

TEST EVALUATION INDEX

Average precision (AP) was treated as the evaluation index of corn ear recognition, so that the calculation formula is written as follows.

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{4}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{5}$$

$$AP = \int_0^1 P(R)\,\mathrm{d}\,R \tag{6}$$

where:

P –precision rate; R–recall rate;

T_P—the number of positive samples correctly predicted;

F_P—the number of negative samples predicted as positive;

F_N— the number of positive samples predicted as negative;

AP— average accuracy.

Improved model

Although Mask R-CNN was an advanced example segmentation model, there were still problems in the application of Mask R-CNN to the detection and segmentation of the fallen corn ears, because the model could not perform target detection and Mask segmentation well due to the influence of the outdoor environment and the complex background of images. Therefore, the model needs to be improved to enhance the detection accuracy and mask segmentation effect.

In order to solve the above problems, the K-means algorithm was used to cluster the targets in the regional recommendation network, and the reasonable recommendation size was set to reduce the interference of artificial design recommendation size. The attention mechanism was added in the feature extraction network to make the convolutional neural network pay Convolutional Block Attention Module (CBAM) to feature extraction and improve the classification ability. The residual structure was optimized to reduce the model parameters in the residual element.

(1) Size regression of area recommendation network Anchor Box based on K-means Algorithm

In the Mask R-CNN model, the RPN scanned the image through a sliding window to search the target area of the Anchor. Anchor was a rectangular box of different sizes and different proportions distributed on the image, and the Anchor Box was designed in 3 sizes and 3 aspect ratios of 1:2, 1:1, 2:1 in the original model. Image tests were carried out by using the Coco dataset, which has 80 categories and more than 330,000 images. The data set in this paper was self-made, and the category and target size were different from the original Coco data set. Therefore, modifying the proportional size of Anchor Box was helpful to improve the target detection accuracy.

K-means algorithm was a typical unsupervised learning algorithm, which could automatically divide different clusters. If the Euclidean distance in standard K-means was directly used as a measure, different clusters would produce different errors, and the larger the cluster, the greater the error. Therefore, the IOU value was used as the distance measure, and the measurement formula was shown below.

$$\mathrm{d}(box, anchor) = 1 - \mathrm{IOU}(box, anchor) \tag{1}$$

According to the location box of the target region of the data set in this paper, the values of *k* were clustered several times in the interval of 2-10 to obtain the change curves of the average IOU and *k* values, as shown in Fig. 4. The optimal *k* value was estimated using the Elbow method (*Dhanachandra, 2015*), that was, the optimal *k* value was the one with the most obvious change in the slope of the curve.

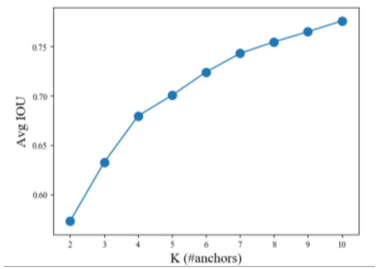


Fig. 4 - Variation curves of average IOU and *k* values

Finally, k =4 was selected for cluster analysis, and the clustering results were shown in Fig. 5. Four clustering centers were obtained (236.96, 80.12), (132.52, 157.78), (113.03, 66.96) and (544.55, 111.34). The aspect ratio of Anchor was determined to be (0.8, 1.6, 3.0, 4.9), and the size was not changed.

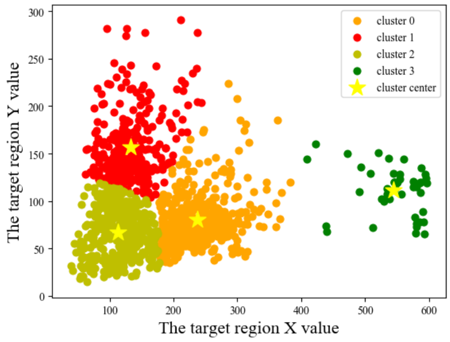


Fig. 5 - Clustering visualization diagram when *k*=4

(2) Residual network and attention mechanism

A large number of machine learning experiments showed that different feature extraction networks had a certain impact on the accuracy and speed of model detection, and the popular feature extraction networks were the VGG network and the ResNet network (*Simonyan, 2016*). As the number of network layers increased, however, the level of accuracy would drop even though a more complex process of feature extraction could be carried out. Based on the mapping relationship shown by the identity transformation, He et al. (2016) proposed a residual network structure (ResNet). Fig. 6 showed the residual unit, where X indicated the input of the neural network, while (X) denoted the expected output. Then, the function could be expressed as $F(X)= H(X)-X$, let $F(X)=0$, and constituted identity mapping $H(X)=X$, the learning goal of the ResNet network became the residual of input $H(X)$ and output X. In this way, it was easier to fit residuals and more sensitive to changes in the mapping output. Thus, the ResNet series were taken as the feature extraction network in this paper for the model.

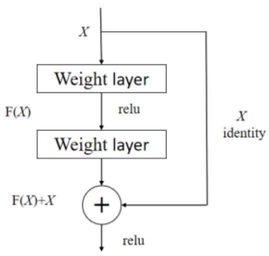


Fig. 6 - Schematic diagram of residual structure

In order to increase the recognition ability of the network and reduce the network parameters (*Li et al*), the residual module was optimized and the grouped convolution was introduced into the residual unit. After the feature map was dimensioned through the 1×1 convolutional layer, the channel dimension was divided into 3 equal parts, and the original 3×3 convolutional layer was replaced with a 3×1 convolution kernel and a 1×3 convolution layer. The optimized residual unit structure was shown in Fig. 7, the input X_1, X_2, X_3 corresponded to the output Y_1, Y_2, Y_3, and the output of the previous channel was used as the input K_2 and K_3 of the next channel. Assuming that the feature map of size $n \times n \times h$ passed through the 3×3 convolutional layer, the parameter A was $h \times 3 \times 3 \times h = 9h^2$, then the optimized parameter B was $3 \times (h/3 \times 3 \times 1 \times h/3 + h/3 \times 1 \times 3 \times h/3) = 2h^2$. It was pointed out that the optimized design greatly reduced the number of parameters of the convolutional layer, and could make up for the computational expenses of adding the CBAM module.

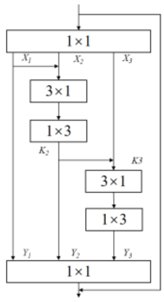


Fig. 7 - Structure diagram of residual after optimization

Inspired by the attention mechanism of deep learning, the CBAM module was introduced into ResNet of the Mask R-CNN feature extraction network to construct the channel attention module and the spatial attention module. Given an intermediate feature map, the CBAM module would infer the attention weight along the spatial and channel dimensions successively, and adjust the features flexibly through multiplication with the original feature map. The performance of the model was significantly improved and the cost incurred by the model was kept low in the meantime.

The channel attention module relied on maximum pooling and average pooling to compress the feature map in spatial dimensions, and two different spatial background descriptions could be obtained including *Fc max* and *Fc avg*. Then, it was evaluated and computed element-by-element through a shared network composed of MLPs to produce the channel attention map $M_c \in R^{c\times1\times1}$. The channel attention mechanism could be expressed as:

$$M_S(F)=\sigma(f^{7\times7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]))=\sigma(f^{7\times7}([F_{avg}^S;F_{max}^S])) \quad (2)$$

where, $f^{7\times7}$ represented a 7×7 convolution layer.

ResNet had four layers with different parameters. We only added channel attention module and spatial attention module in the last layer, and then connected with average pooling and full connection layer. The improved structure was shown in Fig. 8.

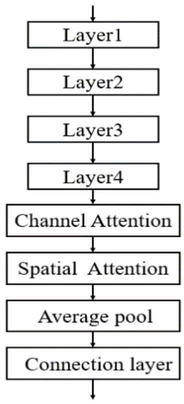


Fig. 8 - CBAM insert structure diagram

RESULTS

EXPERIMENTAL DESIGN AND ANSLYSIS

Experiment and analysis of different feature extraction networks

In order to validate the proposed method, the same data set would be used in this paper to conduct comparative experiments on the speed and accuracy of different feature extraction network models through the same processing platform. Resnet50 and resnet101 feature extraction networks were used to train and test the Mask R-CNN model. The recognition accuracy of Mask R-CNN -ResNet50 model was 88.5%, and the average recognition accuracy of Mask R-CNN -ResNet101 model was 90.0%. Therefore, the Mask-Rcnn-ResNet101 model was chosen as the object of further research in this paper.

Hyperparameter selection experiment and analysis

Through the above-mentioned experiments, the Mask R-CNN model based on ResNet101 was selected to further optimize the hyperparameters. Plenty of deep learning experiments showed that batch processing (minibatch), learning rate and learning decay rate had different impacts on the accuracy of model detection. Empirical batch processing value was set as 32 and 64, learning rate was set as 0.01 and 0.001, and learning decay rate was set as 0.001 and 0.0001. The test results were listed in Table 1. As shown in Table 1, *a* was batch processing value, *b* was learning rate and *c* was learning decay rate.

Comparison test results of different parameter combinations **Table 1**

Serial number	Parameter	Average accuracy / %
1	a= 32, b=0.01, c=0.001	87.6
2	a= 32, b=0.001, c=0.001	88.2
3	a= 32, b=0.01, c=0.0001	86.6
4	a= 32, b=0.001, c=0.0001	87.4
5	a= 64, b=0.01, c=0.001	88.1
6	a= 64, b=0.001, c=0.001	90.8
7	a= 64, b=0.001, c=0.001	89.6
8	a= 64, b=0.001, c=0.001	88.7

The effects that different combinations of learning rate and learning decay rate on the average accuracy of the model were tested based on the different batch processing values as 32 and 64. Compared with the results of the third group and the fifth group in Table 1, the average accuracy of the model test with batch processing value of 64 was higher than that of the model with batch processing value of 32. Compared with the first group and the second group, the sixth group and the eighth group, the model with a learning rate of 0.001 and a learning decay rate of 0.001 had a higher accuracy than the model with a learning rate of 0.01 and a learning decay rate of 0.0001. The average accuracy of the sixth group reached 90.8%, which was the highest among the six groups. Therefore, the model with batch processing value as 64, the learning rate as 0.001 and the learning decay rate as 0.001 was finally adopted.

Comparative test and analysis of the improved model

In order to further demonstrate the effectiveness of this algorithm, a comparison experiment was conducted between the Mask R-CNN model with batch processing value as 64, learning rate as 0.001, learning decay rate as 0.001 based on ResNet101 and the Mask R-CNN model with enhanced data set, improved Anchor Box ratio, residual optimization module and the CBAM. Table 2 shows that the improved Mask R-CNN model without enhanced datasets was 2.4% higher than before, and our model was 3.5% higher than before. In comparison with SharpMask and Scoring R-CNN, SharpMask had the lowest recognition accuracy and Mask Scoring R-CNN had higher recognition accuracy than Mask Scoring R-CNN, but 1.6 % lower than this method. However, with the increase of the average detection time when the convolutional Block Attention Module (CBAM) model was added, it can still meet the practical application requirements.

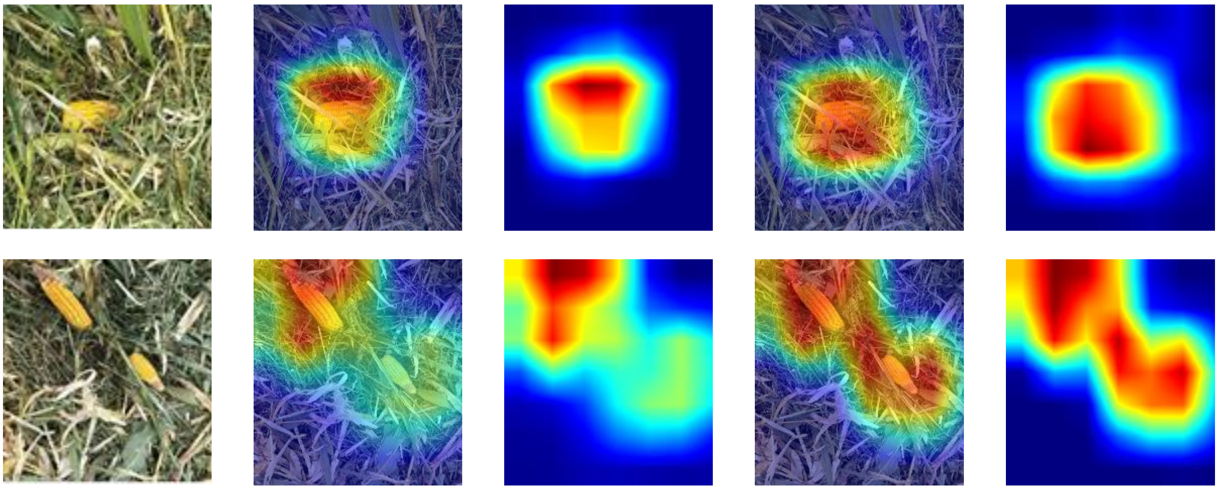
Comparison test results of improved model **Table 2**

Model	Data enhancement	Average accuracy /%	Average detection time /s
Mask R-CNN	No	90.8	0.282
SharpMask	Yes	87.6	0.262
Mask Scoring R-CNN	Yes	92.7	0.292
Improved model	No	93.2	0.317
Improved model	Yes	94.3	0.320

Thermal diagram analysis

The Gradient-weighted Class Activation Mapping (Grad-CAM) (*Selvaraju, 2018*) was performed to visualize the results of model classification in this paper, and the effect of Convolutional Block Attention Module (CBAM) on the model classification process was demonstrated. The global average of the gradient was used by Gard-Cam to calculate the weight, so that the local feature map was generated in the last layer of the convolutional neural network, and the important feature region of the target prediction was highlighted through the point-by-point multiplication of the thermal map and the back propagation.

The basic network of ResNet101 in the Mask R-CNN model and the improved method in this paper were selected to visualize the classification results. As shown in Figure 9, the red part indicated the high credibility of the feature area. Upon comparison, it was found out that the heat map of ResNet101 with the CBAM paid more attention to the feature area. When there are two objects in an image, the thermal map of the attention mechanism module with convolution focused on two objects, while the original thermal map only focused on one object, indicating that it was helpful for the classification and extraction of feature regions, making the feature extraction network more accurate and improving the accuracy of the overall model.



Original picture *Resnet101* *Resnet101* *Improved model* *Improved model*

Fig. 9 - Gard-CAM thermal diagram

Results analysis

a. *Mask R-CNN*

b. *Mask Scoring R-CNN*

c. *SharpMask*

d. *Improved model*

Fig. 10 - Comparison of corn ear recognition results

The recognition result of the fallen corn ear was shown in Fig. 10. It can be seen that the SharpMask method had an incomplete segmentation of the corn ear. The accuracy of the Mask R-CNN method was above 0.9, and the Mask contour of a single ear was clearly identified and accurate. However, the contour of partial Mask segmentation deviated from the actual target, and the Mask segmentation error occurred when there was ear peel next to the ear. The segmentation contour of Mask Scoring R-CNN was relatively accurate, but it was slightly inferior to the accuracy of the improved method in this paper. It could be seen that the improved model had a more complete mask segmentation wheel for incomplete ears and occluded ears. There was no mask segmentation error when there were ear peels nearby, and good results had been achieved in recognition and segmentation. Therefore, the improved method in this paper could be used to identify and segment the corn ears dropped by the harvester.

CONCLUSIONS

A method based on the improved Mask R-CNN model to detect the corn ears lost by the corn harvester was proposed in our study. Data processing for the recognition and segmentation of the ears lost by the corn harvester were enhanced, the Anchor Box size ratio of the RPN sliding window was improved, and the grouping convolution was introduced into the residual unit in this method. Moreover, the original 3×3 convolution layer was replaced by a 3×1 convolution kernel and a 1×3 convolution layer and the CBAM was added to the last layer of the feature extraction network ResNet101, which improved the detection and segmentation accuracy of the model.

Different advanced model tests were performed on the images of the lost corn ears in the test set. Results showed that the average recognition accuracy of the improved method in our study on the test set was 94.3%, which was 4.3% higher than that of the original model. In addition, the method was better than average accuracy of 87.6% by SharpMask and average accuracy of 92.7% by Mask Scoring R-CNN. The method not only had good recognition accuracy and contour segmentation for a single corn ear, but also could identify the interference and overlapping corn ears under the complicated background. It was predicted that the time required for each image was 0.320 s, which met the identification requirements of fallen corn ears.

The reliable recognition of corn ears could be ensured to assist the intelligent detection of the ears lost by corn combine harvester by the dropped corn ear detection method based on the improved Mask R-CNN model.

ACKNOWLEDGEMENT

This work was sponsored by the 13th Five-Year National Key Research and Development Program (2018YFD0300606) and the Shandong Provincial Key Science and Technology Innovation Engineering Project (2018CXGC0217) and the Shandong Provincial Natural Science Foundation (ZR2017BEE032).

REFERENCES

[1] Dhanachandra N., Manglem K., & Chanu Y. J. (2015). Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, Vol.54, pp.764-771, Manipur/India.

[2] He K., Gkioxari G., Piotr Dollár, & Girshick R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961-2969, Venice / Italy.

[3] Huang Z., Huang L., Gong Y., Huang C., & Wang X. (2019). Mask Scoring R-CNN. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *IEEE*, Long Beach / USA.

[4] Jesus R.S., Gloria B., Oscar D., Noelia V., & Gabriel C. (2020). Semantic versus instance segmentation in microscopic algae detection. *Engineering Applications of Artificial Intelligence*, Vol.87, pp.1-15, Ciudad Real / Spain.

[5] Kamilaris A., Prenafeta-Boldu F. X. (2018). Deep learning in agriculture: a survey. *Computers and Electronics in Agriculture*, Vol.147, pp. 70-90, Catalonia / Spain.

[6] Lin T. Y., Dollár P., Girshick R., He K., Hariharan B., Belongie S. (2017, July). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu.

[7] Li D., Zhang K., Li X., Chen Y., Li Z., & Pu D. (2019). Mounting behavior recognition for pigs based on Mask R-CNN. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.50, Issue (Supp), pp.261-266+275, Beijing/China.

[8] Ren S., He K., Girshick R., & Sun J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol.39, Issue 6, pp.1137-1149, Anhui / China.

[9] Sun H., Li S., Li M., Liu H., Qiao L., & Zhang Y. (2020). Research progress of image sensing and deep learning in agriculture. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.51, Issue 5, pp.1-17, Beijing / China.

[10] Singh K. K., Yu H., Sarmasi A., Pradeep G., & Lee Y. (2018). Hide-and-Seek: A data augmentation technique for weakly-supervised localization and beyond. *ArXiv*, 1811.02545, Davis / USA.

[11] Simonyan K., & Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. ArXiv, 1409.1556. https://doi.org/10.48550/arXiv.1409.1556. Cornel University / USA.

[12] Selvaraju R. R., Cogswell M., Das A., Vedantam R., & Batra D. (2020). Grad-cam: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*. Vol.128, issue 2, pp. 336-359, Atlanta / USA.

[13] Trullo, R., Petitjean, C., Ruan, S., Dubray, B., Nie, D., & Shen, D. (2017). Segmentation of organs at risk in thoracic CT images using a SharpMask architecture and conditional random fields. *IEEE International Symposium on Biomedical Imaging,* Vol.2017, pp.1003-1006. Melbourne / Australia.

[14] Wang C., Guo X., Wu S., Xiao B., & Du J. (2014). Three dimensional reconstruction of maize ear based on computer vision. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.45, Issue 9, pp.274-279, Beijing/China.

[15] Wang H., Sun Y., Zhang T., Zhang G., & Li Y. (2010). Appearance quality grading for fresh corn ear using computer vision. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.41, issue 8, pp.156-159, Changchun / China.

[16] Woo S., Park J., Lee J. Y., & Kweon, I. S (2018, September). CBAM: convolutional block attention module, *ECCV*, Vol.11211, pp. 3-19, Munich / Germany.

[17] Xiong J., Dai S., Ou J., Lin X., Huang Q., & Yang Z. (2020). Leaf deficiency symptoms detection method of soybean based on Deep learning. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.51, Issue1, pp.195-202, Guangzhou / China.

[18] Yu Y., Zhang K., Yang L., & Zhang D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask R-CNN. *Computers and Electronics in Agriculture*, Vol.163, Issue (2019), pp.1-9, Beijing/China.