Corn Disease Detection Based on an Improved YOLOX-Tiny Network Model

Shanni Li, Digital Grid Research Institute, China*
Zhensheng Yang, South China Agricultural University, China
Huabei Nie, Dongguan City University, China
Xiao Chen, Shenzhen Institute of Information Technology, China

ABSTRACT

In order to detect corn diseases accurately and quickly and reduce the impact of corn diseases on yield and quality, this paper proposes an improved object detection network named YOLOX-Tiny, which fuses convolutional attention module (CBAM), mixup data enhancement strategy, and center IOU loss function. The detection network uses the CSPNet network model as the backbone network and adds the CBAM to the feature pyramid network (FPN) of the structure, which re-assigns the feature maps’ weight of different channels to enhance the extraction of deep information from the structure. The performance evaluation and comparison results of the methods show that the improved YOLOX-Tiny object detection network can effectively detect three common corn diseases, such as cercospora grayspot, northern blight, and common rust. Compared with the traditional neural network models (90.89% of VGG-16, 97.32% of YOLOv4-tiny, 97.85% of YOLOX-Tiny, 97.91% of ResNet-50, and 97.31% of Faster RCNN), the presented improved YOLOX-Tiny network has higher accuracy.

KEYWORDS
Attention Mechanism, Corn Disease, IOU Loss Function, Object Detection, YOLOX-Tiny Network

1. INTRODUCTION

Corn is one of the important food crops in China. With the advantages of cold tolerance and drought resistance, corn is widely liked by farmers all over the country. However, in the process of corn growth and development, it is often infected with a variety of diseases, among which three kinds of diseases, Cercospora Grayspot, Northern Blight, and Common rust, are the most common. Traditional artificial identification of corn diseases not only takes a lot of time and effort but is also ineffective. Identifying the corn diseases accurately and rapidly is helpful for time rescue and has the great significance of reducing the yield and quality impact of corn diseases.

Computer vision has been applied to the intelligent identification of corn diseases and has achieved many achievements. Gong Ruikun used the canonical correlation analysis method to fuse the feature quantities (Gong et al. 2021), and then used the support vector machine for classification and
recognition, and got the highest recognition rate of 93.1%; based on the VGG-16 model, Xu Jinghui designed a new full connection network (Xu et al., 2020), in which the recognition rate of the layer module for three common corn diseases reaches 95.33%; Li Qingsheng proposed a residual network based on asymmetric attention mechanism (Li et al., 2021), in which the average recognition accuracy of corn disease images can reach 97.25%; and Liu Aoyu (Liu et al., 2021) proposed a corn disease recognition network based on deep residual network, in which the average recognition accuracy of corn disease images is high as 98.96% by introducing Focal loss function.

Although the above-mentioned deep learning methods have improved the recognition accuracy rates, the network parameters are increasing, and the network depth is deepening. Their redundancy and floating-point operation of the network are increasing, resulting in high delay and poor real-time detection effect. KuangShi technology, which is a famous AI company in China, proposed YOLOX-Tiny, which is used for plant protection UAV to detect corn leaf diseases with its advantages, such as small parameter quantity, fast floating-point operation, low delay, and easy deployment of mobile terminal. Due to the shallow network depth, the average accuracy of YOLOX-Tiny is much lower than that of the high network depth models on the market, resulting in a low recognition rate.

Based on the original YOLOX-Tiny deep convolution detection network structure, this paper presents an improved YOLOX-Tiny model with convolution block attention module (CBAM) (Woo et al., 2018), Mixup data enhancement strategy (Zhang et al., 2017), and CenterIOU loss function, which does not significantly increase the memory consumption of the network. Compared with YOLOX-Tiny, the average accuracy of corn leaf disease under a single background is increased from 97.85% to 99.36% by only increasing the recognition delay to 5%, which achieves the balance between the recognition accuracy and the recognition speed.

2. BACKGROUND

2.1 Convolutional Attention Module (CBAM)

The attention mechanism is a common trick used in deep learning models. It is a way to achieve adaptive network attention. The core focus is to let the network pay attention to the features and ignore the insignificant features, which are mainly divided into channel attention mechanism and spatial attention mechanism. The structure of the convolutional attention mechanism is shown in Figure 1.

2.2 Mixup Data Enhancement

Mixup is a data augmentation strategy for mixed-class augmentation that can mix images between different classes to augment the training dataset.

The realization principle of the Mixup data enhancement strategy is to assume that batch$_{x_1}$ is a batch of image samples, batch$_{y_1}$ is the label corresponding to this batch of image samples, batch$_{x_2}$ is another batch of image samples, batch$_{y_2}$ is the label corresponding to another batch of image samples,

Figure 1. Attention convolution mechanism CBAM
then $\lambda$ is determined by the parameter, the mixing coefficient calculated for the beta distribution of $\alpha$ and $\beta$, from which the Mixup principle formula can be obtained:

$$\lambda = \text{Beta} \left( \alpha, \beta \right)$$  \hfill (1)

$$\text{mixed}_{batchx} = \lambda \times batchx1 + (1 - \lambda) \times batchx2$$  \hfill (2)

$$\text{mixed}_{batchy} = \lambda \times batchy1 + (1 - \lambda) \times batchy2$$  \hfill (3)

According to the experimental introduction in the Mixup paper, when $\alpha = \beta = 0.5$, the algorithm works relatively well. The experiment randomly sets the Mixup parameters $\alpha \in [0.1, 0.9]$, $\beta \in [0.7, 0.9]$. During training, first read a sample image, fill both sides of the image, enlarge the image to 460×460 size, and then randomly select a sample image and also enlarge it to 460×460 size. Then set a fusion coefficient, such as 0.5, and fuse the sum and weight, finally getting the complete image.

### 2.3 Improvement of Regression Loss Function

A good loss function achieves twice the result with half the effort for network tuning parameters. YOLOX-Tiny still uses the GIOU loss function when regressing candidate boxes. Based on the characteristics of dataset picture samples, this paper designs a CoreoU loss function, which is good for regressing candidate boxes. The principle is as follows.

As shown in Figure 2, the center points of the candidate frame ($x_{pre}$, $y_{pre}$), the real frame ($x_{tag}$, $y_{tag}$), and the minimum circumscribed rectangle ($x_{ct}$, $y_{ct}$) are known.

**Figure 2. Diagram of CenterIOU loss function**
1. Calculate the IOU value of the candidate frame and the real frame according to formula (4).
2. Construct the general straight line equation of the center point of the candidate frame and the center point of the real frame \( Y = Ax + By + C \).
3. Calculate the straight distance from the center point of the minimum circumscribed rectangle \((x_{ct}, y_{ct})\) to the line according to formula (5).
4. Finally, substitute into formula (6) to get the candidate frame regression loss:

\[
IOU = \frac{BOX_{intersection}}{BOX_{union}}
\]

\[
Distance = \frac{|AX0 + BY0 + C|}{\sqrt{A^2 + B^2}}
\]

\[
CenterIOU\_Loss = 1 - \left( IOU - Distance \right)
\]

3. IMPROVED YOLOX-TINYOBJECT DETECTION NETWORK METHOD

YOLOX-Tiny is a simplified version of the YOLOX series. It reduces the width and depth of the network, cuts down the number of parameters in the model operation, and cancels improvements such as the Mixup data enhancement method during training, making the network simpler and faster.

Based on the above problems, this paper mixes the convolutional attention CBAM module in the YOLOX-Tiny network, and introduces the Mixup data enhancement method to enrich the sample set, so that the network can be trained more fully. Finally, the loss function in training is replaced by the CenterIoU function. The improved YOLOX-Tiny network structure is shown in Figure 3.

Figure 3. Improved YOLOX-Tiny network structure
CSPNet (Wang et al., 2020) is used as the backbone network by YOLOX-Tiny. The backbone feature extraction network is mainly composed of the CBS module, Focus module, Cross Stage Partial (CSP) module, and Spatial Pyramid Pooling (SPP) module.

4. EXPERIMENT RESULT ANALYSIS

4.1 Acquisition of Dataset

The corn leaf and disease picture data required in the experiment mainly comes from the Plant Village website (https://plantvillage.psu.edu/). The final dataset consists of 4,354 images, including 1,000 Northern Blight disease sample images, 1,000 CercosporaGrayspot disease sample images, 1,162 Commonrust disease sample images, and 1,192 healthy sample images, ensuring a balanced number of samples. Example of above-mentioned images in the experimental dataset are shown in Figure 4.

4.2 Evaluation Indicators

For the detection of corn leaf diseases in a single background, the network needs to have high accuracy and real-time performance. The mean average precision (mAP) is used as an indicator of the model identification accuracy. The mAP calculation formula is given in formulas (7)–(10):

\[
P = \frac{TP}{TP + FP} \times 100\%
\]

(7)

\[
R = \frac{TP}{TP + FN} \times 100\%
\]

(8)

\[
AP = \int_0^1 P(R) dR
\]

(9)

\[
mAP = \frac{1}{M} \sum_{k=1}^{M} AP(K) \times 100\%
\]

(10)

Figure 4. Corn Leaf Data Set

(a) Comgrayspot  (b) Cornrust  (c) Healthy  (d) NorthernBlight
In the formula, TP represents the number of positive samples correctly predicted by the model, FP is the number of positive samples incorrectly predicted by the model, FN is the number of negative samples wrongly predicted by the model, M is the total number of categories, and AP (k) is the AP value of class k.

The model hyperparameter is set to 32 samples per batch. Traversing the entire training set data once is called an iteration, and the batch is set to 100. The SGD stochastic gradient descent optimization algorithm described is used to optimize the network parameters.

### 4.3 Performance Analysis

It can be seen from Table 1 that the YOLOX-Tiny network has an accuracy of 98.46% for the identification of corn leaf diseases. After the introduction of the CBAM attention mechanism, the network extracts more fine-grained features for leaves, and the average accuracy of sample identification is improved to 1.04%. After introducing the Mixup data enhancement strategy, the average accuracy of object identification increased to 0.98 percentage points. By replacing the loss function of the regression box with the CenterIOU loss function designed in this paper, and the accuracy of disease identification is improved by nearly 0.91 percentage points.

It can be seen from Table 2 that YOLOX-Tiny has poor detection accuracy for leaves with CercosporaGrayspot and leaves with great NorthernBlight disease. The experimental results show the effectiveness of the CBAM module for differentiating similar object detection. The Mixup data enhancement module is introduced to add more training samples to the network, so that the network can be trained more fully. For the object detection network without an anchor frame, the regression loss function replaces the CenterIOU loss function, which is more in line with the characteristics of this dataset and makes the regression of the annotation frame more accurate.

This paper is also compared with the relatively mature lightweight object detection networks (resnet-50, mobilenet-v2, YOLOv4-Tiny) at this stage. The experimental results are shown in Table 3. YOLOX-Tiny is not only having small parameters, but also greatly reduces the amount of network calculations, thus improving the training of the model. The time-consuming of single image detection is only 2.41ms, while ensuring a high recognition accuracy of 99.36%. Its comprehensive performance is significantly better than that of YOLOX-Tiny and other lightweight networks before improvement.

### Table 1. Average accuracy comparison before and after detection network improvement

| Network                | Average Precision (mAP, %) |
|------------------------|---------------------------|
| YOLOX-Tiny             | 97.85                     |
| CBAM-YOLOX-Tiny        | 98.89                     |
| Mixup-YOLOX-Tiny       | 98.83                     |
| CenterIOU-YOLOX-Tiny   | 98.76                     |
| The proposed improved YOLOX-Tiny | 99.36                   |

### Table 2. Comparison of the average accuracy of the detection network before and after improvement for various of disease identification

| Network         | Healthy (mAP, %) | Grayspot Disease (mAP, %) | NorthernBlight (mAP, %) | Commonrust (mAP, %) |
|-----------------|------------------|---------------------------|------------------------|---------------------|
| YOLOX-Tiny      | 98.95            | 97.65                     | 95.63                  | 99.73               |
| ImprovedYOLOX-Tiny | 99.52          | 99.13                     | 99.01                  | 99.86               |
5. CONCLUSION

This paper introduces a convolutional attention mechanism in the YOLOX-Tiny object detection network, replaces the regression loss function with CenterIOU designed in this paper, and uses Mixup data enhancement strategy in training. In the corn disease dataset, the detection accuracy of CercosporaGrayspot, Commonrust, NorthernBlight, and healthy leaf are 99.13%, 99.86%, 99.01% and 93.52%, respectively. The overall average accuracy is 99.36%, which is higher than that of YOLOX-Tiny.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This work was supported by the Key Field Special Project of Guangdong Provincial Department of Education with No.2021ZDZX1029, and supported by characteristic innovation projects of ordinary colleges and universities of Guangdong Province under No. Grant 2021KTSCX281.
REFERENCES

Ge, Z., Liu, S., Wang, F., & Li, Z. S. J. (2021). YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430.

Gong, R., & Jia, L. (2021). Research on maize disease recognition based on image processing. Modern Electronic Technology, 44(24), 149–152.

He, K., Zhang, X., & Ren, S. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1904–1916.

Li, Q., Nan, M., & Xin, Z. (n.d.). Corn disease image recognition based on attention mechanism asymmetric residual network and transfer learning. Kexue Jishu Yu Gongcheng, 21(15), 6249–6256.

Liuaoyu, Wuyunzhi, & Zhuxiaoning. (2021). Corn disease identification based on depth residual network. Jiangsu Agricultural Journal, 37(1), 67-74.

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 390–391). IEEE.

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. Computer Vision-ECCV, 11211, 3–19.

Xujinghui, Shaomingye, & Wangyichen. (2020). Convolution neural network image recognition of maize disease based on transfer learning. Journal of Agricultural Machinery, 51(2), 230-236+253.

Zhang, H., Cisse, M., & Dauphin, Y. N. (2017). Mixup: Beyond empirical risk minimization. arXiv preprintarXiv.:1710.09412.