Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning

JUN SUN[ID], YU YANG[ID], XIAOFEI HE[ID], AND XIAOHONG WU[ID]

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

Corresponding author: Jun Sun (sun2000jun@sina.com)

ABSTRACT Northern maize leaf blight is one of the major diseases that endanger the health of maize. The complex background of the field and different light intensity make the detection of diseases more difficult. A multi-scale feature fusion instance detection method, based on convolutional neural network, is proposed to detect maize leaf blight. The proposed technique incorporates three major steps of data set preprocessing part, fine-tuning network and detection module. In the first step, the improved retinex is used to process data sets, which successfully solves the problem of poor detection effects caused by high-intensity light. In the second step, the improved RPN is utilized to adjust the anchor box of diseased leaves. The improved RPN network identifies and deletes negative anchors, which reduces the search space of the classifier and provides better initial information for the detection network. In this paper, a transmission module is designed to connect the fine-tuning network with the detection module. On the one hand, the transmission module fuses the features of the low-level and high-level to improve the detection accuracy of small target diseases. On the other hand, the transmission module converts the feature map associated with the fine-tuning network to the detection module, thus realizing the feature sharing between the detection module and the fine-tuning network. In the third step, the detection module takes the optimized anchor as input, focuses on detecting the diseased leaves. By sharing the features of the transmission module, the time-consuming process of using candidate regions layer by layer to detect is eliminated. Therefore, the efficiency of the whole model has reached the efficiency of the one-stage model. In order to further optimize the detection effect of the model, we replace the loss function with generalized intersection over union (GIoU). After 60000 iterations, the highest mean average precision (mAP) reaches 91.83%. The experimental results indicate that the improved model outperforms several existing methods in terms of greater precision and frames per second (FPS).

INDEX TERMS Northern maize leaf blight, disease detection, transmission module, retinex, single shot multiBox detector (SSD).

I. INTRODUCTION

Maize is one of the major food crops in the world. The planting area and output of maize in the world are only lower than that of wheat and rice [1]. In addition to be an excellent feed for animal husbandry, maize is also an important raw material for the development of light industrial products.

However, maize is usually suffered from Northern leaf blight (NLB). In recent years, the decrease of maize yield caused by NLB has been steadily increasing [2] Therefore, it is extremely important to ensure the accurate detection and identification of maize leaf blight. The disease is not easy to be detected in the early period, showing water-stained cigar-shaped spots, gradually spread to leaf sheath. At the later period of the disease, the whole plant loses their vitality, leading to the reduction of maize yield. Traditional maize

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

disease detection mainly relies on the experience of agricultural experts and the expertise of plant pathology. Misjudgment of the disease often leads to a large number of inaccurate pesticide application, which not only leads to environmental pollution, but also increases the pesticide content of maize.

With the widespread use of machine vision, many scholars use the method of machine vision to study disease detection. Hyperspectral imaging combined with chemometrics [3] was successfully proposed to identify the rice sheath blight disease; Support vector machine (SVM) classifiers [4] were developed for distinguishing infected and healthy seedlings. Deep belief networks [5] in the construction of robust methods was applied to precision agriculture. All the above literatures used traditional target detection methods to manually select the characteristics of diseased leaves for segmentation and detection. Although the detection accuracy is relatively high, the color and texture of manual calibration are subjective that may still affect the objectivity of the disease detection.

Convolutional neural network (CNN), a popular method of target detection, has a wide application prospect in the field of crop disease detection [6]–[9]. As a kind of machine learning, CNN can achieve the purpose of accurate detection by training a large number of images. CNN does not depend on specific features, and has a good detection effect in the field of generalized identification, such as target detection [10], [11], target segmentation [12], and target recognition [13], [14]. Zhang *et al.* established a three-channel convolutional neural network for the detection of vegetable leaf diseases according to the different colors of diseased leaves, and the detection accuracy reached 87.15% [15]. Ma *et al.* [16] proposed a deep convolution neural network DCNN to identify and detect four cucumber diseases. In order to reduce the over-fitting of the model, the data enhancement method is used to expand the experimental data set. DCNN obtain good detection accuracy for anthracnose, downy mildew, powdery mildew, and target leaf spots from 14,208 images. Srdjan *et al.* [17] established a plant disease recognition model based on leaf image classification using deep convolutional neural network, and its detection accuracy reached 91%. The above literature proves that it is feasible to detect crop diseases by convolutional neural networks. However, the above data sets are all collected from the background of the laboratory (only a single leaf or a single background), and the detection performance is quite different from that of the images taken in the field. In addition, different from the detection of other diseases, the spot area of maize leaf blight at the initial pathological stage is small and difficult to detect, so the accuracy of detection requirements for small targets is relatively high.

To solve the problems that high-intensity light interfered with disease detection in the field and the traditional model is insensitive to small-target disease detection, this paper added retinex model with low-pass output to preprocess the data set, thus the data set has a higher degree of identification. Meanwhile, the multi-scale feature fusion and the fine-tuning network of anchor box was used in the detection network
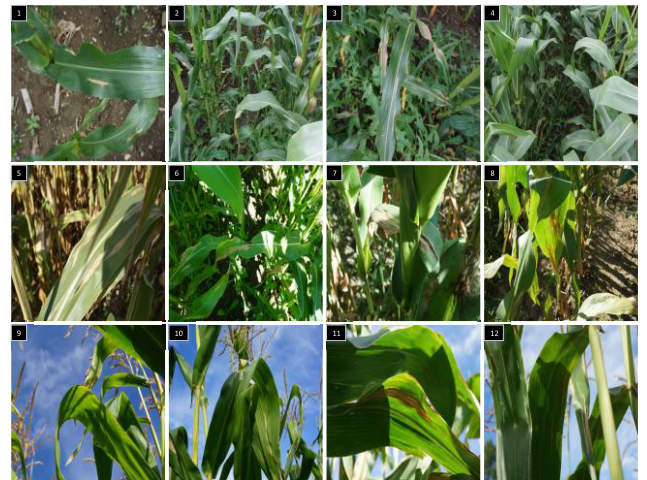

FIGURE 1. Examples of northern leaf blight. Images are under different backgrounds.

to improve the detection effect of small targets. Generalized Intersection over union (GIoU) [18] was adopted to redefine the original loss function for increasing detection accuracy. The images with maize leaf blight in the field are detected by the improved model, and the detection results was compared with the traditional single shot multiBox detector (SSD) model to provide reference for the accurate detection of maize leaf blight.

II. DATA SOURCE

The NLB data set (https://osf.io/p67rz/) is produced in response to the terrible disease, which is the largest open data set on NLB. Each image is calibrated by human plant pathologists and has high accuracy. The NLB data set includes three different parts. The first part is the hand-held set, which is taken by hand. The second part is the boom set, which is taken by mounting the camera on a 5 meters boom. The last data set is unmanned drone set, which is taken by mounting the camera on a DJI Matrice 600. The hand-held part of the data set has a higher clarity. Thus, this part is chosen as the data set in this paper, including 1019 images with different angles and backgrounds, 7669 annotations. Typical images are shown in Fig. 1.

The number of images in the hand-held data set is small, which may affect the training effect. Besides, the uneven classification of the disease sample labels may affect the stability of the model. Raw data sets mainly go through two aspects of data enhancement process. First, the photometric distortions, including random brightness, random lighting noise, and random contrast, hue, saturation. Second, the geometric distortions, including random crop, random expand, and random mirror. The above two categories of data enhancement methods are carried out with a probability of 50% and get 8152 images in total. The augmentation operation does not change the original number of annotations, which also ensures the integrity of the data set. The data set is divided into the training set, validation set, and testing set by 5:4:1.

III. MAIZE LEAF BLIGHT DETECTION MODEL STRUCTURE

Target detection includes one-stage and two-stage detection methods. The main idea of the one-stage method is to use the multi-scale method to perform intensive sampling on the image, and then take advantage of the convolutional neural network to extract features after the classification and regression. There is no extraction of the region proposal, so the speed advantage is demonstrated. As a representative of the single-stage target detection algorithm, SSD [19] improves the detection effect of targets through the anchor of different scales. But it uses low-level feature maps to detect small targets, making the detection of small targets not ideal. The uniform sampling of SSD also leads to the imbalance of positive and negative samples, which leads to a decrease in the accuracy of model detection.

To the best of our knowledge, most of the researches focus on the extension of the data set, whereas ignore the specific optimization to the data set problems. The data set studied in this paper is taken in the field with high light intensity, causing the appearance of 'reflection' phenomenon in some images. Therefore, it is difficult to detect the diseased position clearly. The improved retinex [20] is used to optimize the original data set, making the images adjusted to visual acceptance range [21] for better detection results.

The aim of this study is to solve the problems of poor detection effects caused by high-intensity light, poor detection effect for small targets [22], and inaccurate reflection of the loss function in SSD. Thus, this paper makes three improvements based on SSD:

- The data set is preprocessed by the improved retinex to deal with the problem of high-intensity light;
- A two-stage structure is used to deal with the problem of class imbalance while adding multi-scale feature fusion to improve the detection of small targets;
- GIoU is adopted to optimize the original loss to improve detection accuracy.

A. RETINEX WITH LOW-PASS OUTPUT

Data set preprocessing is an important part of deep learning. The filter function of the single-scale retinex [23] is modified to solve the problem of high-intensity light in this paper. The high-pass filter is used to instead of the original Gaussian low-pass filter to obtain a low-pass output image, which reduces the reflection of the image. Retinex theory points out that the color of an object is determined by the reflection ability of the object, rather than by the absolute value of the intensity of external reflected light [24]. In other words, the color of the object is not affected by the non-uniformity of reflection and has universal consistency [25]. The formula can be expressed as:

$$S(x, y) = R(x, y) * L(x, y) \tag{1}$$

where $S(x, y)$ represents a given image, $R(x, y)$ represents a reflected image, $L(x, y)$ represents a luminance image, and (x, y) represents each point in the image.
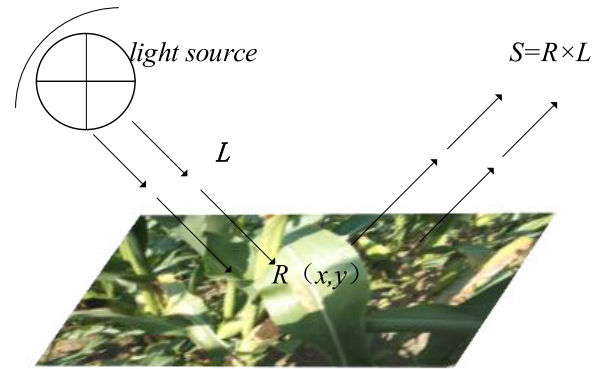
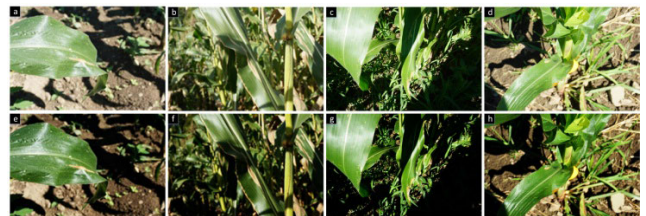

FIGURE 2. The schematic diagram of retinex.

FIGURE 3. Samples of original and optimized pictures. Images a-d are examples with strong reflection in the data set; Images e-h are the corresponding images processed by retinex.

The two variables are separated by taking the logarithm of the reflected image $R(x, y)$ and the luminance image $L(x, y)$. The formula is as follows:

$$logS(x, y) = logR(x, y) + logL(x, y) \tag{2}$$

Then, by calculating the weighted average of pixel points (x, y) in the image and pixel points (x, y) in the surrounding area, the change of luminance $L(x, y)$ is estimated and removed in the original image $S(x, y)$. Thus, the reflected value $R(x, y)$ in the original image $S(x, y)$ is preserved. The specific transformation process is as follows. Firstly, the original image is convolved by a high-pass filter function to obtain a high-pass filtered image $H(x, y)$, and $M(x, y)$ represents a high-pass filter function. It is defined as:

$$H(x, y) = S(x, y) \otimes M(x, y) \tag{3}$$

The high-pass filtered image $H(x, y)$ is subtracted from the original image $S(x, y)$ to obtain the low-frequency weakened image $D(x, y)$. The definition can be normalized as follows:

$$D(x, y) = logS(x, y) - logH(x, y) \tag{4}$$

Finally, an antilog is used to the low-frequency image, and the image $R(x, y)$ with appropriate reflection is obtained. The result is shown as follows:

$$R(x, y) = expD(x, y) \tag{5}$$

In this paper, the retinex is performed on the image with strong reflection in the data set. The specific effect is shown in Fig. 3.
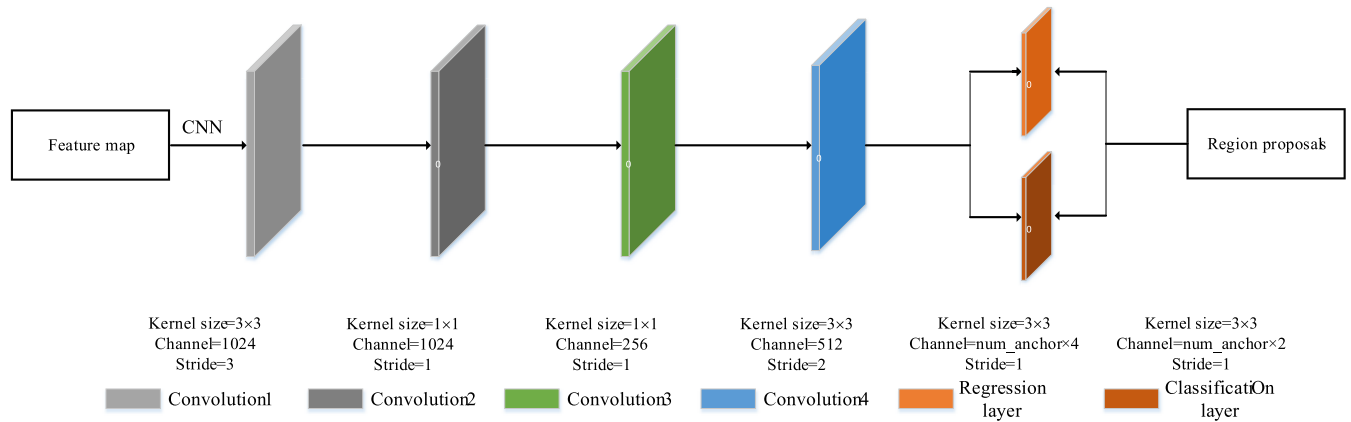
FIGURE 4. The improved RPN network.

B. MULTI-LAYER INPUT RPN NETWORK

With the addition of region proposal network (RPN), the division of anchor box is more detailed. However, in practical application, the efficiency and precision are not enough in the modified SSD with RPN network. Because a feature map generated more than 45,000 anchor boxes. A large number of anchor boxes are located in the background and need to be filtered in the next step. Therefore, it is necessary to adjust the structure of RPN to detect disease areas effectively.

The two-stage method has solved the problem of class imbalance well. As is shown in Fig. 4, the three-layer convolution of the original RPN [26] network is replaced by one Kernel (size = 3×3, Channel = 1024), two Kernel (size = 1 × 1, Channel is 1024 and 256), and a four-layer convolution of Kernel (size = 1 × 1, Channel = 512). Convolution calculation in RPN network is adopted to slide on the feature map. Meanwhile, a series of region proposals are sent out to provide better initial information for the detection network.

In this paper, the 320*320 size of feature map is taken as an example. To deal with the problem of diseased position with different scales, the anchor is extracted on four feature layers from the input feature map, stride sizes 5, 10, 20, and 40 pixels are chosen as four feature layers. Each feature layers are combined with four different scales (20, 40, 80, 160) and three kinds of aspect ratio (1:1, 1:2, 2:1). Finally, 12 anchors with different sizes are generated. We follow the design of anchor scales over different layers, which ensure that anchors of different sizes can have the same density on the image [27]. In this study, the largest IoU values and the samples with IoU > 0.5 are selected as positive samples. Meanwhile, all anchors with negative confidence > 0.99 are removed, that is to say, most background anchors are removed. As a result, the complexity of the model was reduced. The problem of class imbalance is alleviated and the testing time is shortened [28].

C. TRANSMISSION MODULE

In many researches, fusing features [29], [30] of different scales is an important measure to improve detection performance. Low-level features have higher resolution and contain more location and detail information. However, due to less convolution, they have lower semantics and more noise. High-level features have stronger semantic information, but their resolution is very low and their perception of details is poor. How to combine them efficiently is the key to improve the accuracy of detection model.

Transmission module (TM) is designed to improve the detection effect on small targets and detection efficiency in this paper. The feature map associated with the anchor is adopted to fused feature by transmission module. As shown in Fig. 5, firstly, two 3 × 3 convolutions are carried to the feature map and one 4×4 deconvolution is used to expand the high-level feature map, then they are subjected to elemental summation to achieve the purpose of feature fusion. In order to ensure the identifiability of the detection features, one 3×3 convolution is added to the summed feature map. The module refines the features and sum the corresponding elements with the deep features. The network takes the summation result as the feature of the current layer to the detection module, and it solves the problem that low-level feature used in the traditional SSD is insufficient. Thus, the detection accuracy of small target is improved. The fine-tuning network will only send the anchors judged as target disease to the detection module through the transmission module, thus realizing the feature sharing between the detection module and the fine-tuning network.

D. GENERALIZED IOU

Smooth-L1 is used to optimize the bounding box of the SSD. The loss measured by distance does not fully reflect the actual detection situation of the detection box. As shown in Fig. 6, when the three norm values reach the same value, there is a big difference in the actual detection effect (a big difference in the IoU). The phenomenon indicates that the distance norm cannot accurately reflect the real detection effect. The effect of the target detection directly affects by the accuracy of the bounding box regression. Thus, the IoU-based loss can not only accurately reflect the detection effect of the bounding box and the ground truth, but also has the scale invariance. Therefore, the accuracy of target detection can be effectively
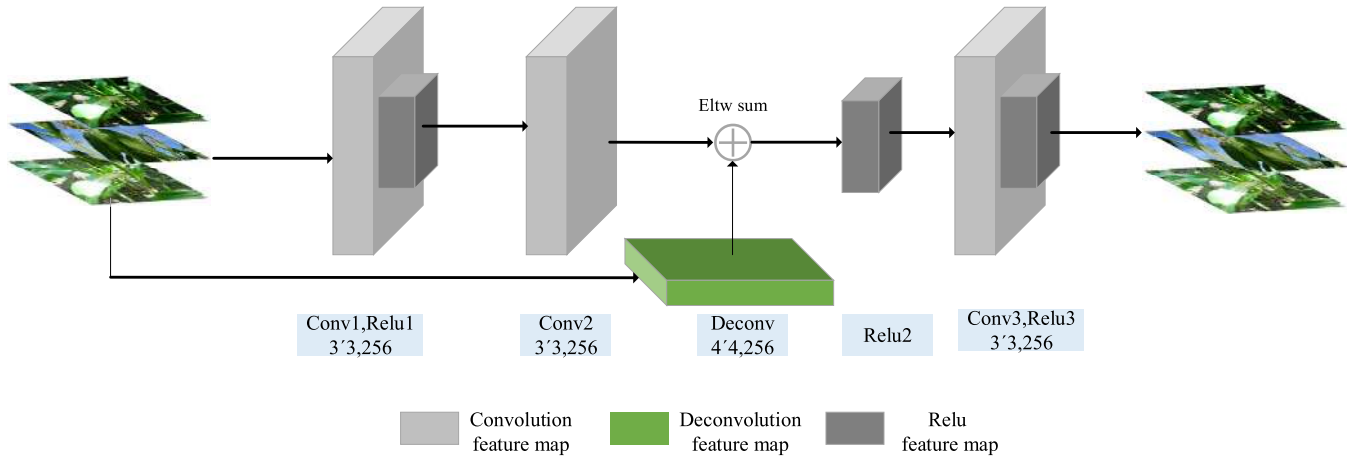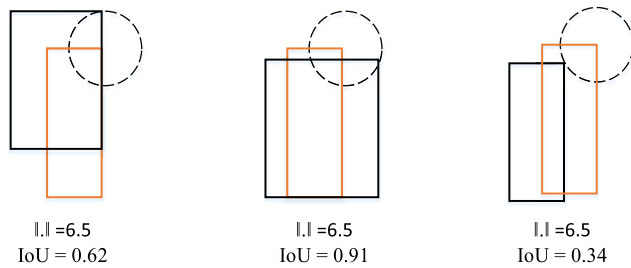
FIGURE 5. Transmission module.

| ‖.‖ =6.5 | ‖.‖ =6.5 | ‖.‖ =6.5 |
| IoU = 0.62 | IoU = 0.91 | IoU = 0.34 |

FIGURE 6. IoU of the same norm. The black box indicates the bounding box and the orange box indicates the ground truth. When the distance norm (‖.‖) is 6.5, the actual IoU is different (IoU is 0.62, 0.91, 0.34).

improved by using the IoU as a loss function instead of the original smooth-L1. Using the IoU [31] as a loss function requires solving the following two problems:

(1) When there is no coincidence between the bounding box and the ground truth, in other words, IoU = 0, the gradient is 0, it cannot be optimized.

(2) When the bounding box coincides with the ground truth, the detection effect is different.

Based on the excellent characteristics of IoU and its shortcomings as a loss function, GIoU is proposed to solve the problems. The loss function of original SSD is optimized by GIoU in this paper. First, the IoU is calculated by the conventional method.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

In the formula (6), A, B are the bounding box and the ground truth, and A, B belong to the set S (S is all the boxes) ($0 \leq IoU \leq 1$). A minimum closed shape C ($C \subseteq S$) is introduced.

$$GIoU = IoU - \frac{|C \backslash A \cap B|}{|C|} \quad (7)$$

According to the definition of GIoU, it can be seen that:

(1) GIoU also has scale invariance.

(2) There is a symmetric range for GIoU ($-1 \leq GIoU \leq 1$).

(3) When A and B do not intersect, the gradient is not 0 due to the introduction of variable C, and the optimization can be continued.

E. MULTI-INPUT RPN NETWORK COMBINED WITH MULTI-SCALE FUSION DISEASE DETECTION MODEL

In this section, Fig. 7 shows the whole NLB detection model based on multi-input RPN network and multi-scale fusion. The model consists of the improved RPN network [32], [33] and transmission module, a total of 11 layers, and the Softmax is adopted to be the classifier [34]. The RPN network replaces the original classification layer with the multi-scale feature input network (conv1, conv2, conv3, conv4) for anchor fine-tuning, and use the 4-layer convolution (conv5, conv6, conv7, conv8) as the detection layer. The transmission module includes two convolution layers and one deconvolution layer (Conv9, Conv10, Deconv11), which is not shown in the figure due to the limits of picture size. Considering that different parameter settings will affect the accuracy of the model, the mean Average Precision of one stage model (SSD) is compared with the mAP of the new models under images of different sizes and different loss evaluation indicators.

$$mAP = \frac{1}{|Q_R|} \sum_{q \in Q_R} AP(q) \quad (8)$$

where Q_R is the number of all categories; $AP(q)$ is the average precision value of the detection.

IV. MODEL TRAINING

A. EXPERIMENTAL PLATFORM

The experimental platform is the Ubuntu 16.04 system, which uses the Caffe as deep learning framework. The computer memory is 16GB, and it is equipped with Intel @CoreTMi7-7 700KCPU@4.00GHzzx8 processor. Two NVIDIA GTX1080Ti GPUs are used in the experiment. The memory type is GDDR5, with a capacity of 11GB, and core frequency is 1480-1582MHz.
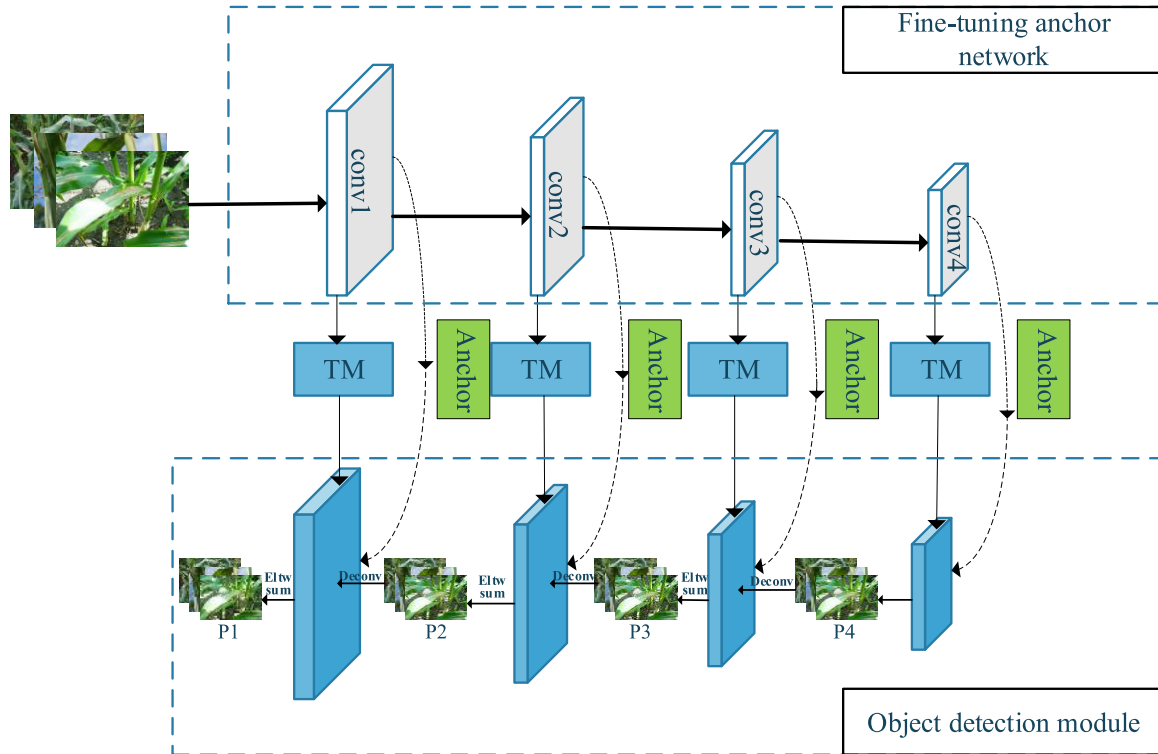
FIGURE 7. Detection model of NLB. Conv1, Conv2, Conv3, and Conv4 corresponding to different input image (512*512, 320*320) is 64*64, 32*32, 16*16, 8*8(40*40, 20*20, 10*10, 5*5), the number of channels is 512, 512, 1024, 512 (512, 512, 1024, 512). P4 is the highest-level input (no deconvolution) obtained from the feature map after three convolution cores (size 3*3, step size 1, channel 256) and pooling. P3 is obtained from the feature map after convolution, pooling, and the sum of elements after deconvolution with P4. P2 and P1 are the same process.

B. SETTING OF TEST PARAMETERS

Batch training combined with momentum factor method is used to divide the training set and test set into several batches, and 16 images are trained in each batch. The number of iteration is 60000. The stochastic gradient descent (SGD) is adopted [35]. The initial learning rate is 0.001, gradually reducing it to the previous 1/25 in stages, and the weight-decay is set to 0.0005 to prevent over fitting.

V. EXPERIMENTAL RESULTS

A. COMPARISON OF DISEASED POSITION

In this section, we not merely shows the conclusions obtained by training the images of the different sizes, but also compare with the result of traditional SSD, which has been trained from two different data set. The mean Average Precisions (mAP) and the Frames Per Second (FPS) of the models are listed in Table 1. As a result, these improvements have proved to be effective for improving the performance of the new model. In the following parts, the impact of these improvements in the overall network framework will be analyzed.

B. THE EFFECT OF IMPROVED RETINEX MODEL ON MAP

Comparing the mAP of Data set A with that of Data set B in each model, it can be concluded that the retinex greatly improves the problem of poor detection accuracy caused

TABLE 1. Model parameter setting and test accuracy.

number	Method	size	mAP (%) Data set A	mAP (%) Data set B	FPS
1	SSD	320*320	71.80	75.42	48
2	SSD	512*512	73.67	77.31	24
3	SSD+TM+RPN	320*320	81.72	85.65	45.2
4	SSD+TM+RPN	512*512	87.37	90.60	22.7
5	SSD+TM+RPN +GIoU	320*320	83.48	88.79	46.7
6	SSD+TM+RPN +GIoU	512*512	89.21	91.83	28.4

by high-intensity light. The mAP of SSD is improved from 71.8% to 75.42%. The accuracy of Data set B produces 5.31% higher than the accuracy of Data set A in model 5, and the mAP improves by 2.26% in model 6. In general, the accuracy of Data set B is higher than that of Data set A. In Fig. 8, part of the detection accuracies (model 6) are shown in Data set A and Data set B. In view of the specific problems of the data set in this study, the improved retinex model effectively solve the problem that the disease position is not obvious.
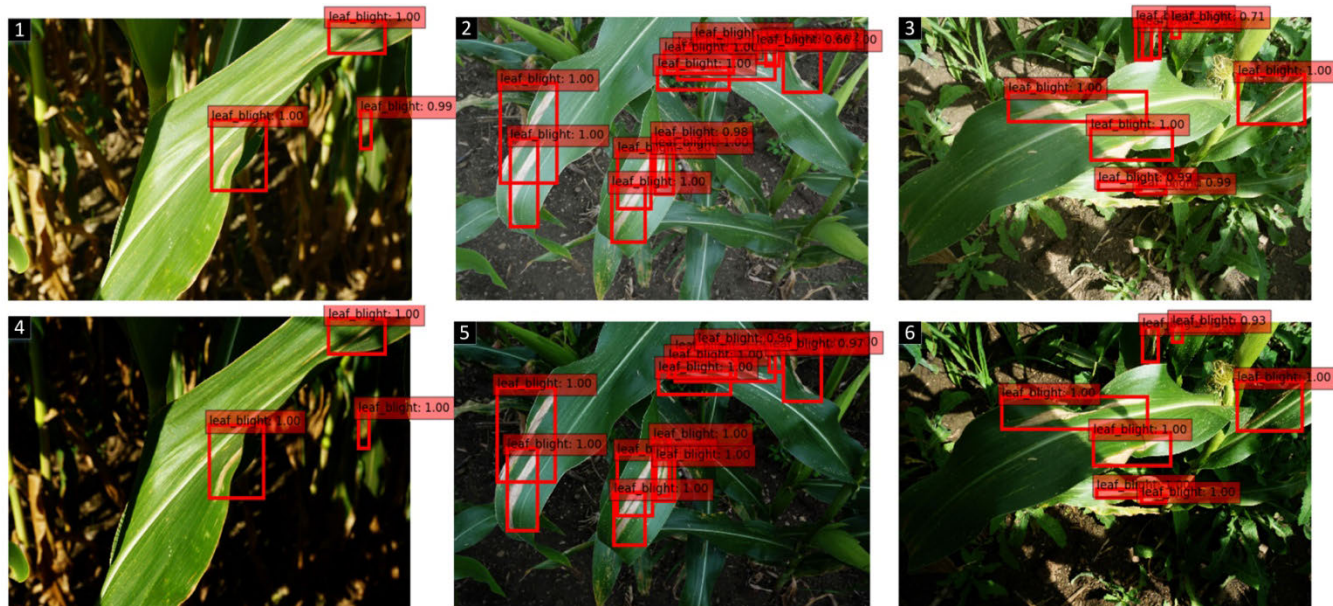
FIGURE 8. Detection accuracy of Data set A and Data set B under model 5. Images 1-3 and Images 4-5 show the detection effects of Data set A and Data set B under model 6 respectively.

FIGURE 9. Comparison of detection accuracy between Model 2 and Model 4.

C. THE EFFECT OF TRANSMISSION MODULE COMBINED WITH RPN NETWORK ON MAP

It is clear that the proposed architecture for the detection of maize leaf blight under complex background is more effective than SSD model in Fig. 9. The multi-layer input RPN network improves the initial information by adjusting the position of region proposal for the classification and precise adjustment of the detection network. Compared with the original SSD model, the mAP of model 3 (320*320) is improved to 85.65%, but its FPS reduce from 48 to 45.2. Compared with

the model 2 (512*512) in Data set B, the model 4 (512*512) achieves 13.29% mAP. The transmission module performs feature layer fusion and combines the high-level semantic features with the previous layer features by deconvolution, which improves the semantic information of the bottom feature layer. Therefore, the detection effect of model 3 and 4 on small targets is improved. A partial visualization of the model 3 and model 4 under Data set B is shown in Fig. 9. It can be clearly seen from Fig. 9 that the models are more effective than the original SSD model under Data set B.

FIGURE 10. Comparison of detection accuracy between model 4 and model 6.

TABLE 2. Comparison with different models.

	Method	Backbone	mAP(%)	FPS
One stage	RetinaNet500	ResNet101	77.43	**34.9**
	DSSD513	ResNet101	78.15	32.7
Two stage	RelationNet600	ResNet101	86.5	12
	SNIP800	DPN-98	89.79	8
Our method	**SSD+TM+RPN+GIoU512**	ResNet-101	**91.83**	28.4

The images (1-3) show the detection effect of SSD. Although some small targets are detected, there is still a missing detection. The images (4-6) show the detection effect of model 4, more small diseased position are detected and no missed detection occurred.

D. THE EFFECT OF GIOU ON MAP

From Fig. 10, it is clearly that the mAP is improved by optimizing the original loss function. Comparing model 3 (320*320) with model 5 in Data set B, the mAP increases from 85.65% to 88.79%, and the mAP also has improvement (1.76%) in Data set A. The best performance of our method is 91.83% (512*512) in model 6. The mAP is higher (1.23%) than that of model 4 (512*512). As can be seen from Fig. 10, the detection accuracy of diseased position has been improved. The main explanation is that GIoU is adopted to redefine the loss. The GIoU can accurately reflect the real detection situation compared with the traditional smooth-L1. The images (1-3) and the images (4-6) show the detection effects of model 4 and model 6 respectively. Adding GIoU into the basis of original model, the detection accuracy of diseased position is improved.

E. COMPARISON WITH OTHER MODELS

Based on the preprocessed data set B, Table 2 compares our model with other detection methods. Our method with Resnet-101 produces 91.83 mAP that is better than other detection models based on Resnet-101. If the input picture (i.e., 512*512) is further enlarged, a better detection effect may be obtained. Generally speaking, the one-stage detection method (e.g., RetinaNet, DSSD) still produces a relatively good FPS, but the detection accuracy is still worse than the two-stage method (e.g., RelationNet, SNIP). This is because the anchor generated by the one-stage detection method is only a logical structure, which only needs to be classified and regressed. The anchor generated by the two-stage detection will map to the area of feature map, and then re-enter the area to the full connection layer for classification and regression. Although our proposed method is slightly inferior to the one-stage detection method in FPS, it has greatly improved its FPS due to the sharing of features of the transmission module. As far as the disease data set we use, on the premise of ensuring the detection accuracy, improving the efficiency of the overall model will provide greater help to the whole production process of intelligent agriculture.

VI. CONCLUSION

In this paper, the convolutional neural network was applied to the detection of maize leaf blight. A promising detection performance in complex field was achieved, which could be attributed to the improvements that we had made based on SSD. In the proposed method, series of steps were amalgamated, including data preprocessing, feature fusion, feature sharing, disease detection. The main reason behind data preprocessing was to reduce the influence of high-intensity light on image identification and improve detection accuracy. In order to further improve the detection accuracy, feature fusion was utilized to produce the best possible results. In our proposed method, we also took into account the improvement of detection efficiency. The transmission module not only realized the feature fusion, but also transferred the relevant anchor information in the fine-tuning network to the detection modules, realizing the feature sharing between the modules, and improving the detection efficiency. Compared with the original SSD model, the mAP of new models was higher (from 71.80% to 91.83%) than the mAP of original SSD. The FPS of the new model also had certain improvement (from 24 to 28.4) and had reached the standard of real-time detection.

The new model of this study was useful for the detection of maize leaf blight in complex background. The disease detection model was efficient and accurate, which could replace the on-site identification of human experts. It could reduce the labor force and overcome the subjectivity of selecting features artificially. The model could be moved into the embedded system, which lays a theoretical foundation for the development of precise drug application and precise detection robot for maize leaf blight.

REFERENCES

[1] P.-Y. Tong, "Report on corn production in China," *Agric. Tech. Eqpt.*, vol. 9, pp. 12–18, Sep. 2011.

[2] C. DeChant, T. Wiesner-Hanks, S. Chen, E. L. Stewart, J. Yosinski, M. A. Gore, R. J. Nelson, and H. Lipson, "Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning," *Phytopathology*, vol. 107, no. 11, pp. 1426–1432, Nov. 2017.

[3] M.-Y. Zhu, H.-B. Yang, and Z.-W. Li, "Early detection and identification of rice sheath blight disease based on hyperspectral image and chlorophyll content," *Spectrosc. Spectr. Anal.*, vol. 39, pp. 1898–1904, Jun. 2019.

[4] C.-L. Chung, K.-J. Huang, S.-Y. Chen, M.-H. Lai, Y.-C. Chen, and Y.-F. Kuo, "Detecting Bakanae disease in rice seedlings by machine vision," *Comput. Electron. Agricult.*, vol. 121, pp. 404–411, Feb. 2016.

[5] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Comput. Electron. Agricult.*, vol. 153, pp. 69–81, Oct. 2018.

[6] Z. Lin, S. Mu, A. Shi, C. Pang, and X. Sun, "A novel method of maize leaf disease image identification based on a multichannel convolutional neural network," *Trans. ASABE*, vol. 61, no. 5, pp. 1461–1474, Oct. 2018.

[7] G. Zhou, W. Zhang, A. Chen, M. He, and X. Ma, "Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion," *IEEE Access*, vol. 7, pp. 143190–143206, 2019.

[8] Z. Libo, H. Tian, G. Chunyun, and M. Elhoseny, "Real-time detection of cole diseases and insect pests in wireless sensor networks," *J. Intell. Fuzzy Syst.*, vol. 37, no. 3, pp. 3513–3524, Oct. 2019.

[9] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.

[10] X. Bai, Z. Cao, L. Zhao, J. Zhang, C. Lv, C. Li, and J. Xie, "Rice heading stage automatic observation by multi-classifier cascade based rice spike detection method," *Agricult. Forest Meteorol.*, vol. 259, pp. 260–270, Sep. 2018.

[11] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.

[12] J. Ma, K. Du, L. Zhang, F. Zheng, J. Chu, and Z. Sun, "A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing," *Comput. Electron. Agricult.*, vol. 142, pp. 110–117, Nov. 2017.

[13] M. A. Khan, T. Akram, M. Sharif, M. Awais, K. Javed, H. Ali, and T. Saba, "CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features," *Comput. Electron. Agricult.*, vol. 155, pp. 220–236, Dec. 2018.

[14] Z. Lin, S. Mu, F. Huang, K. A. Mateen, M. Wang, W. Gao, and J. Jia, "A unified matrix-based convolutional neural network for fine-grained image classification of wheat leaf diseases," *IEEE Access*, vol. 7, pp. 11570–11590, 2019.

[15] S. Zhang, W. Huang, and C. Zhang, "Three-channel convolutional neural networks for vegetable leaf disease recognition," *Cognit. Syst. Res.*, vol. 53, pp. 31–41, Jan. 2019.

[16] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Comput. Electron. Agricult.*, vol. 154, pp. 18–24, Nov. 2018.

[17] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–11, Jun. 2016.

[18] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 21–37.

[20] Y. Shen, Y. Yin, C. Zhao, B. Li, J. Wang, G. Li, and Z. Zhang, "Image recognition method based on an improved convolutional neural network to detect impurities in wheat," *IEEE Access*, vol. 7, pp. 162206–162218, 2019.

[21] M. P. S. da Silva, M. S. M. Freitas, P. C. Santos, A. J. C. de Carvalho, and T. S. Jorge, "Capsicum var. Annuum under macronutrients and boron deficiencies: Leaf content and visual symptoms," *J. Plant Nutrition*, vol. 42, no. 5, pp. 417–427, Mar. 2019.

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[23] J. Wei, Q. Zhijie, X. Bo, and Z. Dean, "A nighttime image enhancement method based on Retinex and guided filter for object recognition of apple harvesting robot," *Int. J. Adv. Robot. Syst.*, vol. 15, no. 1, pp. 1–12, Jan. 2018.

[24] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[25] D. J. Jobson and G. A. Woodell, "Multiscale retinex for color rendition and dynamic range compression," *Proc. SPIE*, vol. 2847, pp. 183–191, Nov. 1996.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[27] X. Liu, D. Zhao, W. Jia, W. Ji, C. Ruan, and Y. Sun, "Cucumber fruits detection in greenhouses based on instance segmentation," *IEEE Access*, vol. 7, pp. 139635–139642, 2019.

[28] J. Sun, W.-J. Tan, and H.-P. Mao, "Identification of plant leaf diseases based on improved convolutional neural network," *Trans. Chin. Soc. Agric.*, vol. 33, pp. 151–162, Oct. 2017.

[29] N. Liu and J.-M. Kan, "Improved deep belief networks and multi-feature fusion for leaf identification," *Neurocomputing*, vol. 216, pp. 460–467, Dec. 2016.

[30] S. Bertrand, R. B. Ameur, G. Cerutti, D. Coquin, L. Valet, and L. Tougne, "Bark and leaf fusion systems to improve automatic tree species recognition," *Ecol. Informat.*, vol. 46, pp. 57–73, Jul. 2018.

[31] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 1999, pp. 1150–1157.

[32] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104846.

[33] M. Haggag, S. Abdelhay, A. Mecheter, S. Gowid, F. Musharavati, and S. Ghani, "An intelligent hybrid experimental-based deep learning algorithm for tomato-sorting controllers," *IEEE Access*, vol. 7, pp. 106890–106898, 2019.

[34] X. Ye and Q. Zhu, "Class-incremental learning based on feature extraction of CNN with optimized softmax and one-class classifiers," *IEEE Access*, vol. 7, pp. 42024–42031, 2019.

[35] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," *Signal Process.*, vol. 6, no. 2, pp. 113–133, Apr. 1984.

YU YANG is currently pursuing the master's degree with the School of Electrical and Information Engineering, Jiangsu University, China. His research interests include applications of deep learning in agriculture, mainly applications of convolution neural networks to the precise detection of crop diseases.

XIAOFEI HE is currently pursuing the master's degree in electronics and communication engineering with Jiangsu University, China. His research interests include applications of computer vision in agriculture, using deep learning method to study weed detection in complex background.

JUN SUN received the Ph.D. degree in mechanical electronics from Jiangsu University. He is currently a Professor and the Doctoral Director of the School of Electrical Information Engineering, Jiangsu University. He has published over 80 articles in his research related fields. His research focuses on applications of computer electronics in agriculture, including deep learning, hyperspectral technology, and nondestructive testing technology.

XIAOHONG WU born Hefei, China, in 1971. He is currently a Professor with Jiangsu University. He is mainly engaged in machine learning, pattern recognition, and spectral information processing. He is serving as a member of the Sixth Council of China Electronic Education Society.

• • •