Research Article

Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense

Yunong Tian,1,2 Guodong Yang,1,2 Zhe Wang,1,2 En Li1,2 and Zize Liang1,2

1Institute of Automation, Chinese Academy of Sciences, The State Key Laboratory of Management and Control for Complex Systems, 95 Zhongguancun East Road, Beijing 100190, China
2University of Chinese Academy of Sciences, No. 19(A) Yuquan Road, Beijing 100049, China

Correspondence should be addressed to En Li; en.li@ia.ac.cn

Received 24 December 2018; Accepted 7 March 2019; Published 8 April 2019

Guest Editor: José A.Padilla-Medina

Copyright © 2019 Yunong Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Plant disease is one of the primary causes of crop yield reduction. With the development of computer vision and deep learning technology, autonomous detection of plant surface lesion images collected by optical sensors has become an important research direction for timely crop disease diagnosis. In this paper, an anthracnose lesion detection method based on deep learning is proposed. Firstly, for the problem of insufficient image data caused by the random occurrence of apple diseases, in addition to traditional image augmentation techniques, Cycle-Consistent Adversarial Network (CycleGAN) deep learning model is used in this paper to accomplish data augmentation. These methods effectively enrich the diversity of training data and provide a solid foundation for training the detection model. In this paper, on the basis of image data augmentation, densely connected neural network (DenseNet) is utilized to optimize feature layers of the YOLO-V3 model which have lower resolution. DenseNet greatly improves the utilization of features in the neural network and enhances the detection result of the YOLO-V3 model. It is verified in experiments that the improved model exceeds Faster R-CNN with VGG16 NET, the original YOLO-V3 model, and other three state-of-the-art networks in detection performance, and it can realize real-time detection. The proposed method can be well applied to the detection of anthracnose lesions on apple surfaces in orchards.

1. Introduction

Nowadays, in fruit agricultural production, most of the farming works rely on the manual labor of fruit planters. A great quantity of simple and repetitive labors not only consumes time and energy and increases production costs but also brings more uncertainties to agricultural production. Fruit farmers may make mistakes in the judgment of farming due to the lack of knowledge and experience. The wrong judgment will directly lead to the wrong implementation of farming, which could have a serious impact on crop yield. With the continuous progress of precision agricultural technology [1], sensors have become the prime sources of crop information. As one of the major components of sensor information, image data plays a significant role in obtaining crop growth status and judging crop health states [2]. With the development of vision sensors, automation and intelligence of agricultural production have been promoted, and various image processing approaches have been applied in agricultural production [3, 4].

Plant disease is one of the primary reasons of crop yield reduction. Timely detection of plant diseases and accurate judgment of disease types are decisive preconditions to take corresponding control measures and diminish agricultural economic losses. However, due to the planting area and crown height of fruit trees, it is difficult to discover the lesions on branches, leaves, flowers, and fruits in time. In addition, the knowledge and experience of fruit planters differ greatly, which may lead to misinterpretation of disease types and erroneous measurement. Inappropriate farming decisions result in the waste of labors and resources, and plant diseases cannot be well controlled. Normally, diseases alter the surface of plant branches, leaves, and fruits, which is in sharp contrast with healthy plants. Therefore, it is feasible to use optical sensors to collect images of plants and fruits and then use the computer vision technology to process the images, so as to
evaluate health states of plants and the types of plant diseases [5]. Li et al. used background removal and defect segmentation methods to detect apple surface lesions [6]. This method can achieve desirable results on pipeline but has larger errors in the orchard environment because of the complex background and variable illumination. Xiao-bo et al. used a camera system to detect apple surface lesions on pipeline [7]. This method adopted multithreshold processing to obtain the region of interest (ROI) of apple surface lesions but could not classify the types of lesions.

While moving ahead with machine learning technology, the traditional image processing methods are gradually replaced by methods such as neural network and support vector machine (SVM). Ebrahim et al. used SVM to detect insects on flowers [8]. Arribas et al. adopted General Softmax Perceptron neural network to train the extracted image features and then used the features to classify sunflower leaves [9]. In our previous work, a BP neural network improved by genetic algorithm was applied to realize multithreshold image processing. The region of green apple in the image was segmented, and the lesion area was extracted by subsequent SVM method. The diseased apple was further identified [10]. This method has been well applied in the images collected in orchards but could not achieve real-time image processing.

With the rapid growth of computing and storage capacities of the GPU processor, deep learning technology gains more popularity. Deep learning achieves desirable performance in computer vision since it takes the advantage of mass amount of data and does not need to extract the image feature manually. In agriculture, deep learning techniques have also been widely applied in crop detection [12–14] and classification [15], pest and disease identification [16] and diagnosis [17], and so on. In the classification of plant diseases, Tan et al. used convolutional neural networks (CNN) to identify and diagnose the surface lesions of fruits [17]. Mohanty et al. used AlexNet, GoogLeNet, and some other deep learning models to classify 26 plant diseases. The highest classification accuracy reached 99.35% [18]. Ferentinos compared the classification performance of 5 deep learning methods such as AlexNet and VGG on 58 different plant diseases, and the highest accuracy reached 99.53% [19]. Fuentes et al. used the improved Faster RCNN model to detect nine different pests and diseases and achieved good results [20]. However, because Faster R-CNN [21] and other R-CNN models include two procedures of region proposal generation and classification, the model cannot realize real-time detection.

YOLO [22–24] unifies object classification and detection into a regression process. It does not adopt the procedure of region proposal but directly utilizes regression to detect the target. Therefore, it effectively accelerates the detection process. The latest version YOLO-V3 [24] not only owns desirable precision and high speed but also has a fantastic performance on small target detection. So it is a splendid reference for apple lesion detection. Nevertheless, YOLO-V3 model has not been widely applied in plant lesion detection.

Generally, due to the existence of downsampling, convolution, and other operations, the size of the feature map in deep neural network architecture gradually reduces with the increase of layers, which brings about the loss of features in the process of network propagation and weakens the utilization of features. YOLO networks share the same issues. To address this, densely connected neural network (DenseNet) [25] proposed by Huang et al. connects all feature layers together. In DenseNet architecture, the input of each layer contains the feature maps of its previous layers, and the output of each layer is connected to each of its succeeding layers. These layers are grouped together through depth concatenation. The application of DenseNet enhances feature propagation. It effectively alleviates the issue of vanishing gradient and boosts the performance of deep learning models. In our previous work [26], DenseNet was utilized to substitute the low resolution feature layers in YOLO network. The detection performance was therefore improved. However, the purpose of the task was to detect apples in orchards and the image data were relatively convenient to obtain, so the techniques based on image generation were not adopted in the image augmentation process.

In our research, an improved YOLO-V3 deep learning model is employed to detect anthracnose lesions on apple surfaces. Because the incidence of apple disease is relatively random, it is very difficult to acquire a large number of specific disease images. To overcome this deficiency, CycleGAN [27] deep learning method is adopted to expand the datasets in this paper. CycleGAN can learn the features of one type of images and transplant them to another. By this means, features of apple lesion images can be extracted and transplanted to healthy apple images. Hence, the dataset of diseased apple images is enlarged.

The rest of the paper is organized as follows. In Section 2, the methods of image data preprocessing are introduced, including image collection, image augmentation, and image dataset production. In Section 3, the YOLO-V3-dense model is described in detail. In Section 4, the experimental setup and experimental results are discussed to demonstrate the effectiveness of the method proposed in this paper. Finally, in Section 5, the main contents of our research are summarized.

2. Image Preprocessing

2.1. Original Image Data Acquisition. As shown in Figure 1, the original images used in this paper contain two types: healthy apple images and diseased apple images. In this paper, anthracnose apple images are taken as the diseased apple image samples. In the early stage of apple anthracnose, the light brown water-soaked small round spots appear on the fruit surface and rapidly expand. When the diameter of the lesion reaches 10-20 mm, many small particles are formed on the lesion surface and then become black, slightly in concentric ring pattern. In moist environment, the lesions will gradually become dark brown. Because apple diseases occur randomly, it is very difficult to acquire a large number of images of specific diseases. In this paper, the diseased apple images are collected in two ways: orchard field collection and online collection. The healthy apple images are relatively easy to collect. In this paper, 500 healthy apple images and 140
anthracnose apple images are collected as datasets for training CycleGAN model.

2.2. Image Data Augmentation. In this paper, traditional image augmentation techniques and the CycleGAN deep learning method are used to preprocess the diseased apple images and expand the image dataset.

2.2.1. Image Augmentation Using Traditional Methods. In our research, three traditional image augmentation techniques, color, angle, and brightness transformation, are used to process the original images.

(1) Color Transformation. In this paper, the Gray World algorithm [28] is used to augment image data in color aspect. Human vision system has color constancy and can acquire the invariant color features of the object from changing illumination environment and imaging conditions, but optical sensors such as imaging equipment do not have such adjusting capacities. Gray World algorithm is usually adopted to reduce the effect of illumination on color rendering. In this paper, 140 images processed by Gray World algorithm are added to the training dataset.

(2) Brightness Transformation. In this research, the brightness of training set images is transformed as follows:

\[ L' = kL \]

where \( L \) is the brightness value of the original image, \( L' \) is the brightness value of the processed image, and \( k \) is randomly selected in the range of 0.8 to 1.2. Then the 140 new results are saved to the training set. This procedure can simulate well the imaging situation under different illumination intensities and increase the adaptability of the neural network model to different illumination intensities.

(3) Angle Transformation. In order to simulate more image viewing angles, four operations, including rotating the picture 90 degrees, 180 degrees, and 270 degrees and mirroring, are randomly selected to process the original images. Then the 140 transformed images are appended to the training image set. The angle transformation of training set images can enhance the robustness of the trained model to different imaging angles.

2.2.2. Image Augmentation Using CycleGAN Deep Learning Method. Generative adversarial net (GAN) [29] deep learning model can learn the features of a type of data and generate similar data. GAN consists of a generative model and a discriminative model. The purpose of the generative model is to generate samples that are increasingly similar to the true samples by a generator (\( G \)). By capturing the distribution of the true sample data, the generator can produce a sample similar to the true training data with a noise \( z \) which obeys a certain distribution such as uniform distribution, Gaussian distribution. The discriminator (\( D \)) is a binary classifier used to estimate the probability that a sample belongs to training data rather than generated data. If the sample belongs to the true training data, the discriminator will output a high probability, otherwise a low probability.

In the process of training, the generator and the discriminator alternately optimize their networks until Nash equilibrium is reached. At this time, the generator can create samples with the same distribution as the true data, and the discriminator can recognize the generated samples with the accuracy of 50%. The final generative model is used to generate new required data.
CycleGAN [27] used in this paper can transform one type of images into another by extracting and transferring image features. This will be very helpful for transferring the features of diseased apple images to healthy apple images so as to transform healthy apple images into diseased apple images. The principle of CycleGAN can be summarized as transforming one type of data into another, that is, there are two sample spaces of $X$ and $Y$; the CycleGAN model is expected to transform samples in $X$ into samples in $Y$ by a generator.

In order to avoid loss invalidation caused by $G$ mapping all images in $X$ to the same image in $Y$, the “cycle consistency loss” is proposed. The cycle consistency loss method assumes that a mapping function $F$ can transform the image $y$ in $Y$ into image $F(y)$ in $X$. CycleGAN algorithm simultaneously learns $G$ and $F$. It is expected to achieve $G(F(y)) = y$ and $F(G(x)) = x$, that is, to establish a one-to-one mapping relationship between $X$ and $Y$.

At the same time, a discriminator $D_Y$ is introduced for $F$. The loss function of CycleGAN model is as follows:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, X, Y) + \lambda L_{cyc}(FmG),$$

(2)

where $L_{GAN}(G, D_Y, X, Y)$ is the loss of $G$ and $D_Y$, $L_{GAN}(F, D_X, X, Y)$ is the loss of $F$ and $D_X$, and $L_{cyc}(F, G)$ is the cycle consistency loss. The expectation of CycleGAN model is as follows:

$$G^*, F^* = \arg \min_{G,F} \max_{D_X,D_Y} L(G, F, D_X, D_Y).$$

(3)

In this paper, 140 images of anthracnose apples are used as training set $B$ and 500 healthy apple images as training set $A$. The images in $A$ are transformed into images in $B$ by the CycleGAN model. As shown in Figure 2, the healthy apple images are transformed into anthracnose apple images.

2.3. Image Labelling and Dataset Production. For the purpose of better comparing the performance of different detection models, the final training set images are made into Pascal VOC format. In the process of training set production, we first readjust the training set images to Pascal VOC scales. Then the resized images are numbered uniformly. After numbering, images are manually annotated, including bounding box marking and category partition. In this paper, lesions which are too small or extremely unclear are not labeled to prevent these samples from decreasing the detection performance of neural networks.

3. Target Detection Algorithm

3.1. YOLO-V3 Algorithm. The development of YOLO algorithm goes through the YOLO [22], YOLO-V2 [23], and YOLO-V3 [24]. The series of R-CNN need region proposal to detect the RoI in the image. Compared with R-CNN, YOLO uses the regression algorithm to solve the problem of target detection. YOLO generates the coordinates of bounding boxes and the probability of each category directly through regression operation, which remarkably promotes the speed of object detection.

At the training stage, the YOLO model first divides the input image into $7 \times 7$ grids. A grid takes charge of the target detection in case that the object ground truth is contained in it. $B$ bounds and their confidence scores are predicted by the grid, as well as $C$ conditional class probabilities. Confidence scores represent the precision of the predicted bounds when the grid contains targets. When various bounds contain the same object, the nonmaximum suppression algorithm is adopted to decide the most outstanding one with the highest confidence score.

The loss function of YOLO model is composed of three parts: coordinate prediction error, intersection over union (IoU) error, and classification error. The loss is defined as follows:

$$Loss = Error_{coord} + Error_{IoU} + Error_{cls}. $$

(4)

The coordinate prediction error $Error_{coord}$ is defined as follows:

$$Error_{coord} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij} \left[ (x_i - x_{\lambda_i})^2 + (y_i - y_{\lambda_i})^2 \right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij} \left[ (w_i - w_{\lambda_i})^2 + (h_i - h_{\lambda_i})^2 \right].$$

(5)
where $\lambda_{\text{coord}}$ is a coefficient to regulate the coordinate prediction error. $\lambda_{\text{noobj}}$ equals 5 in this paper. $S$ equals 7 and $B$ equals 9. $I_{ij}^{\text{obj}}$ indicates that the true value of the probability of the object in grid $i$, $(x_i, y_i, w_i, h_i)$ are predicted bounding box parameters of center coordinates and box size. $(x_i, y_i, w_i, h_i)$ are actual parameters.

The IoU error is defined as follows:

\[
\text{Error}_{\text{IoU}} = \sum_{i=0}^{S} \sum_{j=0}^{B} I_{ij}^{\text{obj}} (C_i - \tilde{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S} \sum_{j=0}^{B} I_{ij}^{\text{obj}} (C_i - \tilde{C}_i)^2, \tag{6}
\]

where $\lambda_{\text{noobj}}$ is a coefficient to adjust the IoU error. $\lambda_{\text{noobj}}$ equals 0.5 in our research. $\tilde{C}_i$ is the predicted parameter of the confidence score and $C_i$ is the actual one.

The error of classification is as follows:

\[
\text{Error}_{\text{cls}} = \sum_{i=0}^{S} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \sum_{c=0}^{C} (p_i(c) - \tilde{p}_i(c))^2, \tag{7}
\]

where $p_i(c)$ indicates that the true value of the probability of the object in grid $i$ belonging to class $C$. $\tilde{p}_i(c)$ is the predicted value.

Although YOLO largely raises the speed of target detection using regression method, it still presents a considerable margin of error. To address the defect, YOLO-V3 brings anchor box and K-means clustering algorithms in its network to generate appropriate prior bounds. YOLO-V3 renovates the design of the network structure on the basis of YOLO-V2 and utilizes convolution in the output layer. To further improve the detection result, YOLO-V3 also adopts batch normalization, high resolution classifier, and so on. Multi-scale prediction is also adopted to detect the final target in YOLO-V3 model. Therefore, YOLO-V3 achieves more desirable results in small target detection. This provides a theoretical basis for the detection of apple lesions in orchards.

3.2. YOLO-V3 Model Improved by DenseNet. In the training process of neural networks, the operations of convolution and downsampling reduce the size of feature layers, resulting in feature loss. To exploit the features more effectively and minimize the loss of them, DenseNet uses the method of feedforward to connect each layer to every other layer. Therefore, the layer $l$ of DenseNet receives all the features of the preceding layers as inputs as follows:

\[
x_i = H_l[x_0, x_1, \ldots, x_{l-1}], \tag{8}
\]

where $[x_0, x_1, \ldots, x_{l-1}]$ is the spliced feature map of layers $x_0, x_1, \ldots, x_{l-1}$. $H_l$ is the transfer function of the spliced feature map.

As shown in Figure 3, this paper improves the original Darknet-53 architecture. DenseNet is adopted to substitute the feature layers which have smaller scales. When features are transmitted to these layers, they will be reused by multiple feature layers in DenseNet, thus lessening the loss.

4. Experiments and Discussion

4.1. Evaluation Indicators. In this paper, detection models are checked out on the NVIDIA Tesla V100 server. The
The following indicators are used to evaluate the performance of detection models.

4.1.1. Precision-Recall \((P - R)\) Curve and \(F_1\) Score. Precision \((P)\), recall \((R)\), and \(F_1\) score are essential indicators for evaluating the performance of object detection models. The precision represents the prediction accuracy, representing the percentage of the actual targets in all the predictions. Recall indicates the percentage of the actual targets detected in all actual targets. The higher the recall, the more actual targets are detected.

The \(P - R\) curve can be acquired by combining precision and recall into a two-dimensional coordinate system. In this paper, \(F_1\) score is also used to evaluate the performance of the detection models as follows:

\[
F_1 = \frac{(k+1) \times P \times R}{k \times k \times (P+R)},
\]

where \(k\) indicates the relative importance of precision to recall. Recall has a greater impact on \(F_1\) score when \(k > 1\). When \(k < 1\), precision has a greater impact. In this paper, \(k\) equals 1.

4.1.2. IoU. IoU is a parameter for evaluating the prediction precision of the bounding boxes. IoU verifies the detection performance by assessing the overlap ratio between predicted bounds and actual bounds.

4.1.3. Loss Value. Loss value is an essential standard to estimate the convergence result of a neural network model in training stage. In this paper, by comparing the difference of loss between the improved YOLO-V3 model and the unmodified YOLO-V3 model in training stages, the advantages of the improved model can be brought out.

4.1.4. Average Detection Time. In this paper, we compare the average detection time and analyze the real-time performance of several models.

4.2. Results of Image Augmentation. Due to the random occurrence of apple diseases, it is arduous to collect large quantities of a particular type of lesion images. In order to meet the demand of image quantity for training deep neural networks, three traditional data augmentation methods, including color, brightness, and angle transformation, are implemented. The image augmentation performance is shown as Figure 4.

In order to further expand the image dataset, CycleGAN deep learning method is utilized to generate diseased apple images. The initial parameters of CycleGAN model are shown in Table 1. The generated images are shown in Figure 5.

From the above experimental results, it can be seen that CycleGAN can learn the features of anthracnose apple and healthy apple images well through training and can generate anthracnose lesions on the surface of healthy apple images. This allows to generate more new images, making up for the lack of image data due to the low incidence of apple diseases.

<table>
<thead>
<tr>
<th>Size of input images</th>
<th>Batch</th>
<th>Momentum</th>
<th>Initial learning rate</th>
<th>(\lambda)</th>
<th>Training steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>256 × 256</td>
<td>1</td>
<td>0.5</td>
<td>(2e^{-4})</td>
<td>10</td>
<td>150000</td>
</tr>
</tbody>
</table>
diseases. The application of the CycleGAN method produces lesions of different shapes, sizes, and quantities, which greatly enriches the diversity of the image dataset. In order to compare with other traditional image augmentation techniques, CycleGAN is applied in this paper to transform 140 healthy apple images into anthracnose apple images, so as to maintain the same number of images as other methods. The composition of the final training set is shown in Table 2.

### Table 2: The training set after data augmentation.

<table>
<thead>
<tr>
<th>Initial images</th>
<th>Color transformation</th>
<th>Brightness transformation</th>
<th>Angle transformation</th>
<th>CycleGAN generation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>140</td>
<td>700</td>
</tr>
</tbody>
</table>

4.3. Detection Performance of YOLOV3-Dense Model. In our research, DenseNet is adopted to improve the feature usage efficiency. The detailed parameters and structure of the proposed model are shown in Figure 6. Firstly, in the course of model training, the input images are rescaled into $512 \times 512$ pixels. Then the downsampling layers with the lowest and second-lowest resolution are replaced by DenseNet.

In our research, the structure $BN - ReLU - Conv(1 \times 1) - BN - ReLU - Conv(3 \times 3)$ which is composed of batch normalization (BN), rectified linear units (ReLU), and convolution (Conv) is adopted as $H_1$ to realize the nonlinear calculation of $[x_0, x_1, \ldots, x_{l-1}]$. In feature layers with the resolution of $32 \times 32$, $x_i$ is composed of 64 subfeature layers. $H_1$ carries out $BN - ReLU - Conv(1 \times 1)$ nonlinear calculation on $x_{i0}$, and then $BN - ReLU - Conv(3 \times 3)$ operation is employed. $H_3$ also performs the above operations on the spliced feature map $[x_{i0}, x_{i1}]$. The rest operations are done in the same manner. Finally, the spliced feature map $[x_{i0}, x_{i1}, x_{i2}, x_{i3}]$ with the size of $32 \times 32 \times 512$ continues to propagate forwards. In feature layers with the resolution of $16 \times 16$, $x_i$ is composed of 128 subfeature layers. The above feature layer splicing and feature propagation steps are also carried out, and the feature layers are finally spliced into $16 \times 16 \times 1024$ size for forward propagation. Finally, the improved detection model predicts bounds in three different resolution and classifies the categories of the bounding boxes, so as to realize lesion detection.
The initialization parameters of the proposed model are shown in Table 3.

So as to increase the model performance, the input size of the image is adjusted to $512 \times 512$ pixels in this paper. Considering the insufficient memory of the server, the batch size is set to 8. In order to better analyze the training process of the model, training steps are set to 70000 in this paper. After setting up the initialization parameters, the YOLOV3-dense model is trained. When model training is completed, 50 test images of $500 \times 500$ pixels are used to conduct a series of experiments to verify the performance of the algorithm. To better observe the bounding boxes, we use “1” to substitute the label of “anthracnose.” The lesion detection results of the proposed model are shown in Figure 7.

As can be seen from Figure 7, the YOLOV3-dense model proposed in our research can detect the anthracnose lesions on apple surface under the condition of complex environment and illumination, and it can also achieve good results for the detection of small lesions.

4.4. Comparison of Different Detection Models. To better verify the superiority of the improved method, we compare YOLOV3-dense with Faster R-CNN and the original YOLO-V3. These models are evaluated from several aspects such as loss curves, $P$-$R$ curves, $F_1$ scores, detection time, IoU, and the image test results. We also compare the detection accuracy of the proposed model with three state-of-the-art models [19].

In the training stages, the loss curve of the YOLOV3-dense model converges rapidly before 2000 steps. The convergence speed slows down after 2000 steps and is essentially stopped after 40000 steps. Finally, the loss of the YOLO-V3-dense model is about 0.29, around 0.43 lower than the loss of the YOLO-V3. It can be perceived that the YOLOV3-dense model has higher utilization of image features than the YOLO-V3 model.

In this paper, the proposed model is compared with the other two latest detection models on the image test set. The
Figure 7: Detection performance of YOLOV3-dense model. (a, c) Images of anthracnose apples. (b, d) Images of the detection results.

Figure 8: Comparison of loss curves between YOLOV3-dense and YOLO-V3.
In order to obtain better detection results, color, brightness, and angle transformation and CycleGAN deep learning are adopted to expand the training image dataset. To evaluate the influence of the augmentation techniques on YOLOV3-dense model, the control variate technique is adopted to get rid of one data augmentation approach every time and get the indicators in the absence of this method, as shown in Table 7.

From the above experimental results, removing the images generated by CycleGAN has the greatest impact on the detection performance of the model. This indicates that the image data generated by CycleGAN plays a crucial part in enriching the diversity of training dataset. Compared with the traditional methods of color, brightness, and angle transformation, CycleGAN generates lesions with new backgrounds, textures, and shapes, which is of great help to enhance the robustness of the detection model. In particular, for the detection of the lesion images collected in the orchard environment, because the orchard environment is complex and the distribution and shape of lesions are variable, the diseased apple images generated by CycleGAN deep learning method can make up for the deficiency of insufficient training set images, so that the detection model can achieve a better consequence.

To better assess the performance of the improved model, we compare the detection results of anthracnose lesions in original images versus the images generated by transformation. In this paper, another 150 images are generated through the methods of color, brightness, and angle transformation by the 50 images in the test set. Meanwhile, 50 diseased apple images generated by the CycleGAN deep learning method are randomly selected for testing. The test results are shown in Table 8.

It can be seen from Table 8 that the detection results of the test images generated by color transformation are better than those of the original images, while the detection results of the test images converted by brightness transformation are slightly worse. This demonstrates that color transformation effectively eliminates the influence of illumination and reserves the original color characteristics, which alleviate the difficulty of detection. While the brightness transformation can simulate different illumination intensity and has certain influence on the detection performance. Compared with the above two methods, the influence of angle transformation on detection results is relatively slight. The YOLOV3-dense model in this paper can also achieve desirable lesion detection results for the diseased apple images generated by CycleGAN.

4.6. Influence of the Training Set Image Number on Detection Results. In order to evaluate the influence of training set image number on YOLOV3-dense model, 10, 50, 100, 200, 400, 500, 600, and 700 images are randomly selected from the 700 training set images to make up different training datasets. The YOLOV3-dense model is trained on these datasets, and the \( P-R \) curves, \( F_1 \) scores and IoU of the trained models are shown as Figure 11 and Table 9.

From the above experimental results, with the increase of training set size, the detection performance of the proposed model is gradually enhanced. When the number of training set images exceeds 400, the rate of model performance

\[ \text{Figure 9: } P-R \text{ curves of three detection models.} \]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Model} & \text{Faster R-CNN with VGG16 NET} & \text{YOLO-V3} & \text{YOLOV3-dense} \\
\hline
\text{F1 score} & 0.765 & 0.797 & 0.816 \\
\text{IoU} & 0.863 & 0.903 & 0.917 \\
\text{Average detection time (s)} & 0.289 & 0.027 & 0.032 \\
\hline
\end{array}
\]

\textit{Table 4: Detection indicators of several models.}

\( P-R \) curves of several models are shown in Figure 9; other detection indicators are shown in Table 4.

The detection performance of the three models is shown in Figure 10. In Figure 10, 19 lesions are manually labeled, 10 lesions are detected by Faster R-CNN model, 16 lesions are detected by YOLO-V3 model, and 17 lesions are detected by YOLOV3-dense model. Detection results of 50 test images are shown in Table 5.

Ferentinos [19] compared the classification performance of several state-of-the-art deep learning methods such as AlexNet and VGG. In this paper, these methods are also used to identify anthracnose lesions. The detection accuracy of these models is shown in Table 6.

It can be observed from the improved \( F_1 \) score and IoU that the improved model in this paper enhances the detection result compared with the unmodified YOLO-V3. The improved model obviously precedes the Faster R-CNN with VGG16 NET. The average speed of YOLOV3-dense detection is 31 FPS, and it is capable to achieve real-time detection. As shown in Table 6, compared with other three models, the proposed model has the highest detection accuracy.

4.5. Impact of Data Augmentation Methods on Detection Results. In order to obtain better detection results, color, brightness, and angle transformation and CycleGAN deep learning are adopted to expand the training image dataset.
enhancement slows down with the further expansion of the training set.

### 5. Conclusion

In this paper, an anthracnose lesion detection method is proposed based on deep learning. Firstly, in view of the problem of insufficient image dataset due to the random occurrence of apple diseases, CycleGAN deep learning method is adopted to extract the features of healthy apples and anthracnose apples and to produce anthracnose lesions on the surface of healthy apple images. Compared with the traditional image augmentation methods, this method greatly enriches the diversity of training dataset and provides plentiful data for model training. Based on the data augmentation, DenseNet is adopted in our research to substitute the lower resolution layers of the YOLO-V3 model, which further exploits the features in the neural network and enhances the detection performance of the model. Compared with the Faster R-CNN, the average detection time of the YOLOV3-dense model is greatly shortened, and real-time detection can be realized.

![Figure 10: Detection results of three detection models.](image)

(a) Anthracnose apple images. (b) Detection results of Faster R-CNN with VGG16 NET. (c) Detection results of YOLO-V3. (d) Detection results of YOLOV3-dense.

<table>
<thead>
<tr>
<th>Model</th>
<th>Manual annotation</th>
<th>Faster R-CNN with VGG16 NET</th>
<th>YOLO-V3</th>
<th>YOLOV3-dense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of detected lesions</td>
<td>219</td>
<td>168</td>
<td>185</td>
<td>194</td>
</tr>
<tr>
<td>Number of missed lesions</td>
<td>0</td>
<td>51</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>Number of wrong lesions</td>
<td>0</td>
<td>17</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNetOWTBN</td>
<td>95.19%</td>
</tr>
<tr>
<td>VGG</td>
<td>95.54%</td>
</tr>
<tr>
<td>GoogleNet</td>
<td>93.84%</td>
</tr>
<tr>
<td>YOLOV3-dense</td>
<td>95.57%</td>
</tr>
</tbody>
</table>
Moreover, the proposed model outperforms three other state-of-the-art models in detection accuracy.

In the future works, we will transfer the method mentioned in this article to more fruit surface lesion detection problems to further test the performance of the proposed model. And on the basis of identifying the type of lesions, the diagnosis of the disease stages will be carried out, which will provide a guidance for taking corresponding disease prevention and control measures.

### Data Availability

Since the data used in this paper was acquired by self-collection, the dataset is being further improved, so the dataset is not available for the time being.

### Conflicts of Interest

The authors claim no conflicts of interest.
Acknowledgments

This project was supported by the National Key Research and Development Program of China (Grant no. 2017YFD0701401).

References


