Received 21 October 2022, accepted 3 November 2022, date of publication 7 November 2022, date of current version 1 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3220234

RESEARCH ARTICLE

Real-Time Monitoring Method of Strawberry Fruit Growth State Based on YOLO Improved Model

QILIN AN[1,2], KAI WANG[2], ZHONGYANG LI[1], CHENGYUAN SONG[2], XIUYING TANG[1], AND JIAN SONG[2]

[1]College of Engineering, China Agricultural University, Beijing 100083, China
[2]School of Mechanical and Automation, Weifang University, Weifang 261061, China

Corresponding author: Jian Song (sjian11@163.com)

This work was supported by the Shandong Province Key Research and Development Program through the Public Welfare Category Project 2019GNC106144 and 2022CXGC010612.

ABSTRACT A key challenge in automated orchard management robots is the fast and accurate identification of crop growth conditions and maturity for subsequent operations such as automatic pollination, fertilization and picking. In particular, strawberry fruits have a short ripening period and the fruits are heavily overlapped and shaded by each other, which is time-consuming and ineffective based on traditional detection methods. Therefore, we designed and developed a strawberry growth detection algorithm, SDNet (Strawberry Detect Net). The algorithm is based on the YOLOX model and replaces the original CSP block in the backbone network with a self-designed feature extraction module C3HB block to improve the spatial interaction capability and monitoring accuracy of the detection algorithm; Then, the normalized attention module (NAM) is embedded in the neck to improve the detection accuracy and attention weight of small target fruits; and we use the latest SIOU objective loss function to improve the prediction accuracy of the detection model, which finally achieves the monitoring of strawberry fruits under five growth states. The experimental results show that the precision, accuracy, and recall of SDNet are 94.26%, 93.15%, and 90.72%, respectively, and the monitoring speed is 30.5 ms. It is 4.08%, 3.64 and 2.04% higher than the precision, accuracy, and recall of YOLOX, respectively, and there is no significant change in the model size. The research results can effectively solve the problem of low accuracy of strawberry fruit growth state monitoring under complex environments, and provide important technical reference for realizing unmanned farm and precision agriculture.

INDEX TERMS Fruit detection, digital agriculture, deep learning, real-time counting.

I. INTRODUCTION

Strawberries are one of the most popular fruits in the world, with worldwide production reaching more than 9 million tons in 2022 [1], [2], [3]. However, the detection of strawberry fruit at the fruit growth state in the orchard is usually based on farmers' experience and is highly subjective, which may lead to errors and affect the pollination, spraying, harvesting, and marketing of the fruit. Moreover, strawberries are highly perishable and the right maturity level for picking them often determines the quality and storage time of the fruit [4]. Therefore, effective real-time detection of the strawberry growth state not only provides an important reference for the quantity and quality loss of strawberry fruits during growth but also has important significance for automated orchard management.

Traditional detection methods apply color characteristics, surface texture characteristics, chemical composition, and odor to fruits and crops [5]. Surya et al. [6] evaluated the growth stages of bananas based on their color characteristics to determine the correct harvest time for banana growers. Sooyeon et al. [7] used gas chromatography/mass spectrometry (GC-MS) to compare plant volatile compounds of unripe and ripe fruits to distinguish the growth information

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

of the fruits. Campos et al. [8] and [9], used electronic tongues as a tool for monitoring grape ripeness, measuring its physicochemical parameters, acidity, pH, and other parameters for ripeness classification. Shao et al. [10] used a portable hyperspectral imager to determine the internal quality and ripeness of peaches to support field inspection and timely harvesting of peaches. Zhang et al. [11] developed a sensing intelligent manipulator to grade avocado ripeness. However, these methods require corresponding specialized equipment that is unsuitable for fruit inspection in a natural growth environment and are destructive, time-consuming, and labor-intensive.

With the rapid development of artificial intelligence in recent years, deep learning technology has also made great progress [12], [13], [14], [15]. Deep learning, in which machines mimic human activities such as seeing, hearing, and thinking to solve complex pattern recognition challenges, has been successfully applied to rapid crop and fruit detection [16]. Wilson et al. [17] determined the ripeness of The Cape gooseberry fruit by a combination of machine learning techniques and color spaces (RGB, HSV, and L∗a∗b∗). Wan et al. [18] proposed a combination of feature color values and back propagation neural network (BPNN) techniques for the detection of fresh tomato ripeness. Xu [19] use of the YOLOv3 detection algorithm for the detection of different growth stages and classes of tea buds classification. Harshana et al. [20] studied the identification detection method of strawberries in a greenhouse environment using deep convolutional neural networks with an average detection accuracy of 77.21%. Yu et al. [21] segmented and identified strawberries in an unstructured environment based on Mask-RCNN, which could only classify ripe and immature strawberries. Chen et al. [22] added the identification of strawberry flowers to the category of detected ripe and immature strawberries and estimated their yield based on the FAST-RCNN algorithm whose average accuracy of detection was 72%. Li et al. [23] used a combination of depth features and classifiers for strawberry appearance recognition, which was mainly applied to mechanized production lines, and the accuracy of the method was as high as 96.55%, but its single background was not suitable for detection under natural growth environment. Zhang [24] designed a lightweight real-time strawberry recognition device based on the YOLOv4-tiny algorithm, which is mainly applied to ripe strawberry harvest detection.

In summary, the research on fruit detection has been explored in depth, but the identification and detection of the fruit growth process are less and the strawberry target in the trial dataset is too single, and the model inference speed is too slow to meet the demand for real-time detection of multiple targets. In addition, strawberry fruits in the natural growth environment are densely distributed and have different postures, and the fruit leaves and fruits overlap each other to shade the growth. Different treatments are needed for strawberries at different growth stages, such as calcium phosphate fertilization, flower thinning, and hand pollination at the flowering stage. The green fruiting stage and white fruit stage need foliar fertilizer spraying and adding the nutrient solution to make the fruit ripen quickly. The ripening stage requires timely picking, which is necessary for the study of real-time detection of strawberry fruits growth state.

Therefore, in order to improve the real-time detection accuracy of strawberry fruits in different growth states under a natural environment, weaken the influence of overlapping occlusion of multiple targets on recognition results and enhance the real-time detection effect of the model. In this study, based on the homemade backbone feature extraction C3HB block, fused normalized attention module (NAM) and SIOU target frame loss function, the strawberry fruit detection model (SDNet) for different growth stages is proposed, and the possibility of the model for real-time detection of strawberry growth in the natural environment is realized. The overall process of building the SDNet model is shown in Fig. 1.

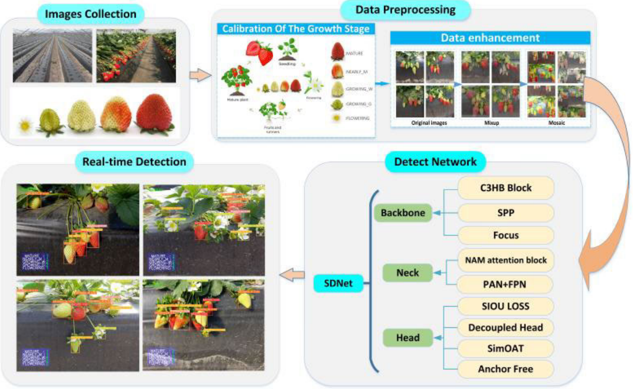


FIGURE 1. Flowchart for building a strawberry detection model (SDNet).

II. MATERIALS AND METHODS

A. DATASET CONSTRUCTION

1) IMAGE ACQUISITION

The fieldwork was conducted from November 2021 to April 2022 at the experimental station of the Academy of Agricultural Sciences in Weifang, Shandong Province (longitude 119.21866, dimension 36.819597). The strawberry varieties in the orchard are mainly 'Sui Zhu', 'Hong Yan', and 'Zhang Ji', which usually start to ripen in batches in autumn around November until the end of April of the following year. Fig. 2. shows the growth cycle of strawberries, from seedling, flowering, and fruiting to the final fruit maturity, which requires thinning, pollination, spraying, and picking at different fruit growth state.Therefore, based on the experience of strawberry growers and related information, we know that strawberry fruit growth states are shown on the right side of Fig. 2. and can be roughly divided into five states periods: flowering (FLOWING), green fruiting (GROWING_G), white ripening (GROWING_W), coloring (NEARLY_M) and mature (MATURE). The most important thing is to determine the ripeness of strawberries. The Chinese agricultural strawberry industry-standard NY/T 444-2001 states [25] that

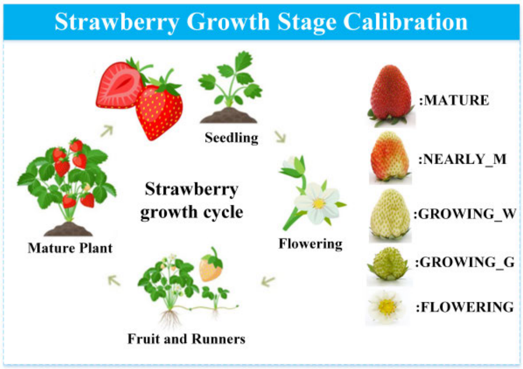


FIGURE 2. Strawberry growth stage calibration.

strawberries are ripe for picking and eating if their coloring rate is greater than 70%.

The strawberry dataset was photographed using a smartphone (Iphone11) and a camera with a resolution of 1920 × 1080, one shot per 3 strawberry plants at a distance of approximately 30-50 cm from the fruit. To ensure the diversity of the data set samples, a total of 5600 strawberries at different growth states were photographed under different weather and light conditions, respectively, and stored using JPG format. The obtained dataset was annotated using the annotation tool labelImag, and the smallest external rectangular box of strawberries was used as Ground truth for annotation to reduce the background pixels inside the real box. A corresponding XML-type annotation file was generated for each image and the dataset was adapted to Pascal VOC2007 format.

2) DATA ENHANCEMET

The deep learning network models of the YOLO family require a large number of datasets to reduce model overfitting or underfitting [26], [27], [28], [29], [30]. Therefore, the acquired samples need to be pre-processed to enhance the dataset before training the network model. The acquired Datasets are first enhanced by random combinations of panning, flipping, scaling, adding noise, rotating, etc. using methods from the OpenCV library to augment the sample Datasets images to 28,000. Next, the Datasets is then processed by the Mixup method, where two images are randomly mixed proportionally and their classified results are proportionally assigned. Finally, the blended Datasets are enhanced using the Mosaic method, and four images are randomly stitched together as in Fig. 3. to form a new image and its corresponding annotation information is also updated.

It is demonstrated that this greatly enriches the background of the detected objects and enhances the data computation of the images to improve the reliability of the network model training and the robustness of the detection.

B. OPTIMIZED YOLOX

In 2021 Megvii Technology Kuangwei Research Institute proposed the end-to-end deep learning detection algorithm YOLOX, which has faster inference speed and lower computational cost, and real-time detection. YOLOX network model fusing CSPDarknet53, Focus, and SPP structures as a backbone feature extraction network model. Although YOLOX has fast detection speed and real-time effect, it has low accuracy in detecting and recognizing strawberry growth conditions in the natural environment and poor recognition of overlapping occlusion of multiple targets. Therefore, this study improves the overall network model based on the YOLOX network model to make it possible to detect strawberry fruits at different growth stages in real-time with high accuracy.

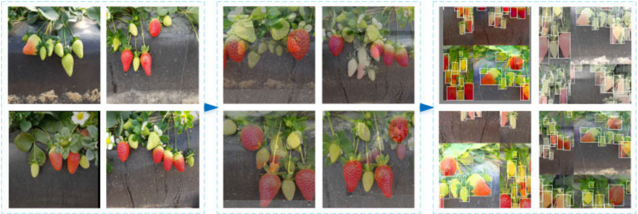
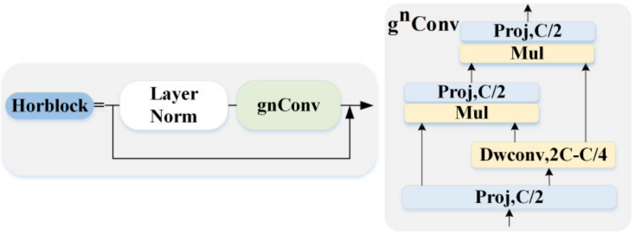

(a)Original images (b)Mixup (c)Mosaic

FIGURE 3. Strawberry Data Enhancement Process In Complex Environments.Where (a) is the original acquired image; (b) the effect after Mixup; (c)the effect after Mossic.

1) C3HB BLOCK

HorNet combines the advantages of Vision Transformer and CNN to propose (Recursive Gated Convolutions) [31]. It performs efficient, scalable, and panning higher-order spatial information interactions by the recursive design. It performs efficient, scalable, and panning higher-order spatial information interactions through recursive design. As represented in Fig. 4. is the structure of a gated convolution, which is composed using standard convolution, elemental multiplication, and linear projection, but with an adaptive mixing function similar to that of self-attentiveness. operation firstly adjusts the number of feature channels by two convolutional layers, and secondly divides the output features of the depth-separable convolution into multiple blocks, each of which interacts with the features of the previous block by further element-by-element multiplication to finally obtain the output features.



(a)framework of Horblock (b)g^nConv

FIGURE 4. Horblock and g^nConv basic building blocks. (a) shows the overall framework of Horblock;(b) shows the detailed implementation of g^2Conv.

Inspired by HorNet, this study designed the C3HB module to replace the backbone feature extraction module of the strawberry detection network model based on the YOLOX network model by fusing Horblock and C3 structure to enhance the feature extraction and spatial interaction capabilities of the detection model, as shown in Fig. 5. network structure that enhances the learning capability of CNNs and reduces computational bottlenecks and memory costs in a lightweight manner while maintaining the accuracy of the detection model.

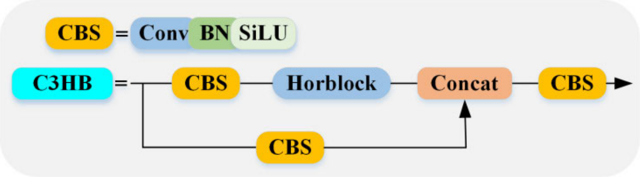


FIGURE 5. C3HB basic building blocks.

The C3HB module mainly consists of the CBS module, Horblock module, and Concat structure, The main flow of the C3HB module is shown in Fig. 5. The initial feature information input will be assigned to two branches. Firstly, the number of channels of the input feature information is halved and feature extraction is performed through the CBS module in branch one, and the other part of the feature information is passed through the Horblock module and CBS structure in branch two, and the output feature information of branches one and two are connected using the Concat operation depth. Finally, feature enhancement is performed by one more CBS structure, in which the output feature information size is the same as the input size of the C3HB module. The established C3HB module effectively improves the accuracy and higher-order spatial interaction capability of the strawberry detection model.

2) NAM ATTENTION MECHANISM

To enhance the accuracy of strawberry detection for small and medium target fruits, this study adds a normalization-based attention module (NAM) [32] to the network model. It redesigned the channel and spatial attention sub-modules based on CBAM, using the scaling factors of the weights on the two dimensions to enhance the effect of attention. The scaling factor in Batch Normalization is used in the channel attention module to reflect the weight of each channel and the importance of the channel, and the larger the scaling factor of Batch Normalization, the richer the channel variation and the more important its weight, as in Equation 1.

$$\mathrm{B_{out}} = \mathrm{BN}(\mathrm{B_{in}}) = \frac{\mu \mathrm{B_{in}} - \epsilon_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2}} + \beta \tag{1}$$

where $\epsilon_{\mathcal{B}}$ is the mean, $\sigma_{\mathcal{B}}^2$ is the variance, and μ and β are the trainable variation parameters (scale and displacement).

In the way of the channel attention module, for example, the BN scale factor, called Pixel Normalization (PN), is also applied in the spatial submodule, mainly to measure the importance of spatial dimensional features.PN focuses on the more informative pixels in the spatial sub-channel according to the scale factor and adjusts the associated weights to suppress the less significant pixels so that more attention is focused on the useful ones, as shown in Figure 6.

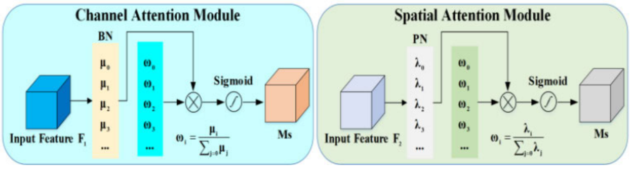


FIGURE 6. NAM attention block.

$$\mathrm{M_c} = \mathrm{sigmoid}(\mathrm{W}_{\mu}(\mathrm{BN(F1)})) \tag{2}$$

$$\mathrm{M_s} = \mathrm{sigmoid}(\mathrm{W}_{\lambda}(\mathrm{PN(F2)})) \tag{3}$$

where $\mathrm{M_c}$ denotes the output features for the channel attention submodule; μ denotes the scale factor on each channel dimension; channel weight $\mathrm{W}_{\gamma} = \gamma_{\mathrm{i}} \Big/ \sum_{\mathrm{j=0}} \gamma_{\mathrm{j}}$; $\mathrm{M_s}$ is the output feature of the spatial attention submodule; λ denotes the scale factor on each spatial dimension; and channel weight $\mathrm{W}_{\lambda} = \lambda_{\mathrm{i}} \Big/ \sum_{\mathrm{j=0}} \lambda_{\mathrm{j}}$.

3) LOSS FUNCTION

The main role of the loss function is to measure the accuracy of the expected effect of the detection model. Generally, a minimum of three loss functions need to be defined: object loss, classification loss, and bounding box loss. The bounding box loss often affects the detection accuracy and convergence speed of the network model. The bounding box loss used in the YOLOX network model is the original IOU, which mainly relies on factors such as the distance, overlap area, and aspect ratio between the prediction frame and the real frame to measure the difference between them, however, the matching of the angular orientation between the prediction frame and the real frame is not taken into account. Therefore, we embed the SIOU loss function [33] in the network model, which introduces the vector angle between the desired regressions in the calculation of the boundary regression loss function, and redefines the penalty metric of the boundary regression loss function. This makes the predicted target frame more stable and the network model detects strawberry fruits closer to the size and location of the real target frame. The SIOU loss function plays an important role in the strawberry detection network mainly through the linear combination of four components: Angle cost, Distance cost, Shape cost, and IoU cost.

$$\mathrm{Loss} = \mathrm{L_{SIOU}} + \mathrm{L_{object}} + \mathrm{L_{classification}} \tag{4}$$

The overall loss evaluation parameter of SDNet is shown in Eq.4, which is used to evaluate the degree of inconsistency of the prediction results with the ground truth, and it consists of three components: bounding box loss (SIOU), object loss,

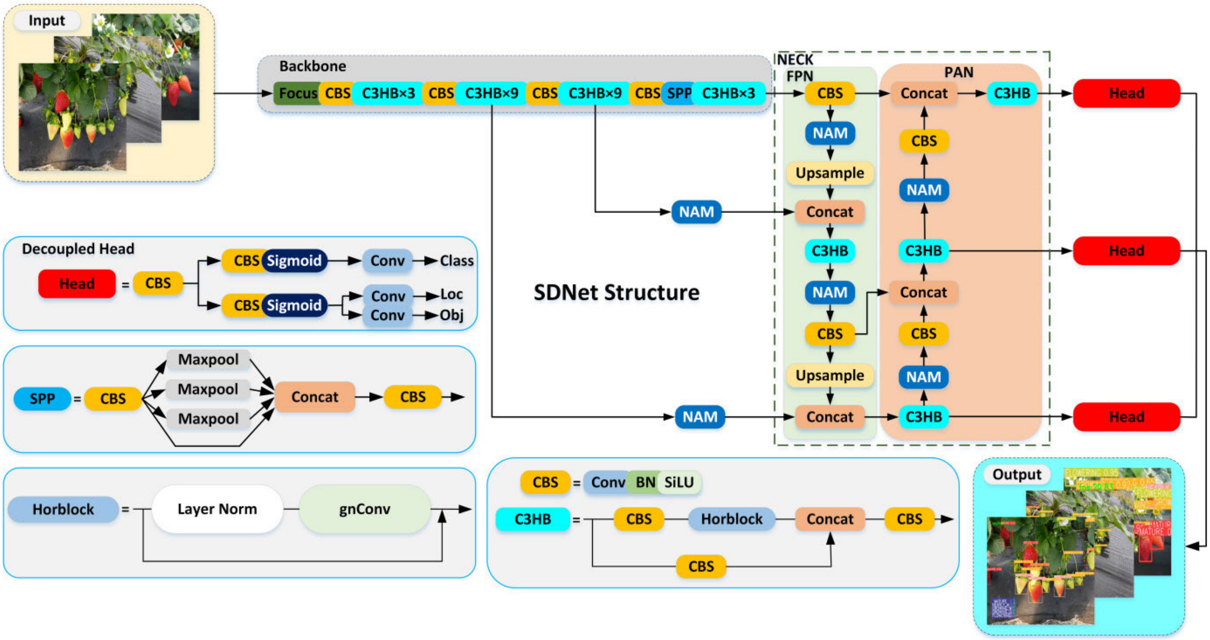


FIGURE 7. The architecture of the propose SDNet.

and classification loss, which affect the rectangular box, confidence level, and classification probability, respectively.

C. SDNET WORKING PROCESS

The structure of the SDNet network model is shown in Fig. 7.: it is mainly composed of the detection model backbone composed of C3HB, NECK with embedded NAM attention mechanism, and Decoupled Head. The backbone part retains the FOCUS model for slicing operation of input images and the SPP structure for spatial pyramid pooling operation in the original YOLOX network model and embeds 3, 9, 9, and 3 layers of C3HB feature extraction modules in turn. The NECK part mainly consists of FPN and PAN structures, which up and down-sample the input feature information to enhance the feature extraction capability. Moreover, the NAM and C3HB modules based on the normalized attention module are added to the NECK part to enhance the SDNet network model for small target strawberry feature information enhancement. Finally, the feature information obtained from the SDNet model is passed into the Decoupled Head section to obtain information on the growth condition of strawberries.

III. TRAINING AND DISCUSSION

In this section, we evaluate SDNet and compare the relevant parameters with the current mainstream models. We used five quantitative methods to evaluate the model, containing average precision AP, recall, F1, Precision, and mAP. The calculations are as follows:

$$\mathrm{AP} = \int_0^1 (\mathrm{pre} \times \mathrm{recall})\mathrm{drec} \tag{5}$$

$$\mathrm{recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \times 100\% \tag{6}$$

$$\mathrm{precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \times 100\% \tag{7}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{pre} \times \mathrm{recall}}{\mathrm{pre} \times \mathrm{recall}} = \frac{2\mathrm{TP}}{2\mathrm{TP} \times \mathrm{FP} \times \mathrm{FN}} \times 100\% \tag{8}$$

$$\mathrm{mAP} = \frac{1}{\mathrm{c}} \sum\nolimits_{\mathrm{i}=1}^{\mathrm{c}} \mathrm{AP_i} \tag{9}$$

where TP denotes the number of correctly classified positive samples, FP denotes the number of misclassified positive samples, and FN denotes the number of misclassified negative samples.

A. MODEL TRAINING

1) HYPERPARAMETER SETTING

We proposed strawberry detection network model SDNet was trained and tested on Ubuntu (18.04.4) and Windows 10 system workstations, and Table 1 shows the specific settings of the training and testing devices. Since the weights of the network model trained from scratch are too random and less effective for feature extraction, we first pre-trained it in the initial training phase using the COCO dataset and trained the obtained weights on a homemade strawberry dataset by transfer learning, which can improve the efficiency and robustness of the model training. The hyperparameters are trained using SGD optimizer with an initial learning rate of 0.001, momentum factor designed to be 0.9, batch size can be set to 16, and epoch for migration learning training is set to 100.

2) MODEL TRAINING

We first redraw the P-R curves for each epoch during the training of the SDNet algorithm, as shown in Fig. 8. the AP, the integrated area under the curve, is calculated

TABLE 1. Training and testing equipment settings.

Equipment	Training equipment	Test equipment
System	window 10	Ubuntu18.4
Language	Python	Python
framework	Pytorch1.7.1	Pytorch1.7.1
GPU	NVIDIA RTX 3090	NVIDIA GeForce RTX 2060
VRAM	32G	6G
CPU	Intel(R) Xeon(R) Gold 6234 @3.30GHz	Intel(R) Core(TM) i7-1165G7 @2.80GHz
RAM	192GB	32GB

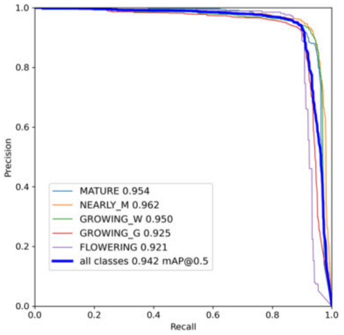


FIGURE 8. P-R curves of SDNet.

in the legend. The larger the AP, the better the detection performance for its phase. Where precision and recall are mutually constrained, if precision increases, recall decreases, so the detection algorithm needs to find a balance between the two. Usually, the intersection IOU threshold and confidence threshold are the two basic metrics of deep learning algorithms.

We also plot the confusion matrix of SDNet results for the test set with an IOU threshold of 0.5 and a confidence threshold of 0.5. As shown in Fig. 9., the confusion matrix aggregates the data in the dataset in matrix form according to the true category and the category judgment predicted by the classification model. The rows in the graph represent the predicted categories, the columns represent the actual categories, and the data on the diagonal lines represent the proportion of categories correctly classified. It was shown that it is the most difficult to detect fruits in the green and flowering stage, which are easy to be missed and misidentified because they are tiny targets and more similar to the background. The strawberries in the coloring stage are easily misidentified as a ripe and white ripening stage.

B. ALGORITHMIC OPTIMIZATION RESULTS

1) COMPARISON OF DIFFERENT ALGORITHMS

We performed validation on a homemade strawberry dataset and compared SDNet with various mainstream detection models, which include YOLOv4, YOLOv5, Swin Transformer, and YOLOX. The results of the detection algorithm are shown in Table 2, and the training results are compared in Figure 10.

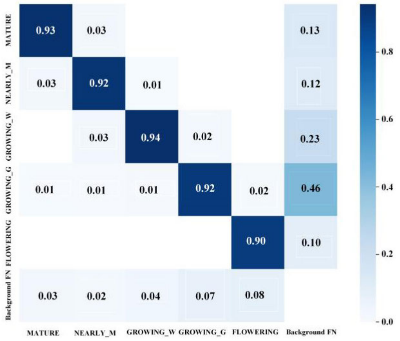


FIGURE 9. Confusion matrix for SDNet.

TABLE 2. Test results of different detection algorithm models.

Method	F1(%)	Precision (%)	Recall (%)	mAP (%)	Paramet ers(%)
YOLOv4	79.82	78.91	80.75	83.37	50.0
YOLOv5	86.29	85.9	86.7	89.19	47.1
STr	88.86	89.42	88.32	90.78	174.4
YOLOX-l	88.35	89.51	88.68	90.18	54.2
SDNet	91.91	93.15	90.72	94.26	54.6

The proposed SDNet has the best detection among all the algorithm models with the average detection accuracy, F1, precision, and recall of 94.26%, 91.91%, 93.15%, and 90.76%.SDNet shows a significant improvement in accuracy compared to the original YOLOX network model, where the average accuracy mAP, recall, and F1 values are improved by 4.08%, 2.04%, and 3.56%, respectively. And compared to YOLOv4, YOLOv5 and STr (Swin Transformer) algorithm models its average detection accuracy is improved by 10.89%, 5.07%, and 3.48%, which is the best among all current advanced detection models.

In terms of model size, the SDNet detection model has 54.6 MB of parameters, which is slightly larger than the original YOLOX network model. The Swin Transformer model, which has a higher average detection accuracy, has a higher number of parameters of 174.4 MB, which is three times higher than the SDNet model. Compared to YOLOv4 and YOLOv5, the number of parameters is smaller but the detection accuracy is lower and the recognition error is larger than for strawberry growth detection. the comprehensive analysis of the data in the table shows that although SDNet does not have much advantage in terms of model parameter size, it has the best detection accuracy and effect and is more suitable for monitoring the growth of strawberries.

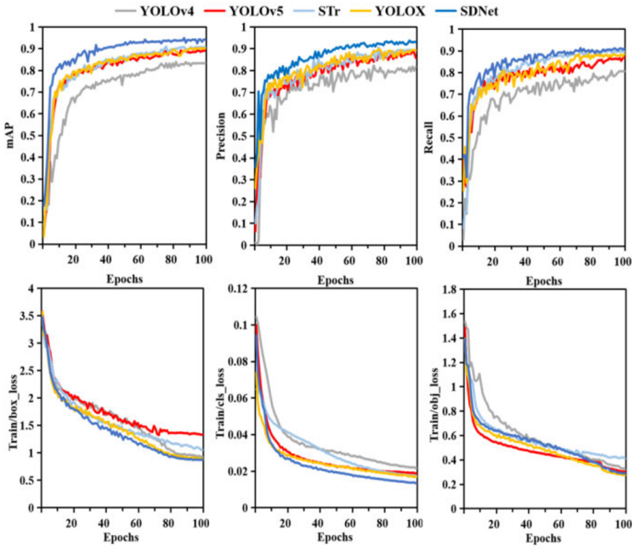


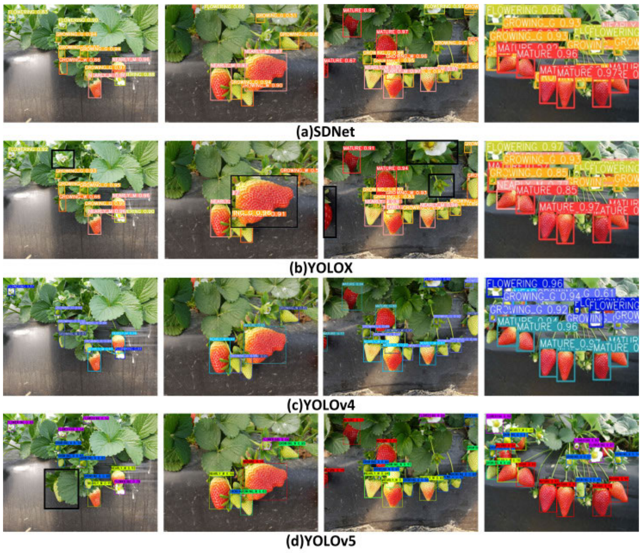
FIGURE 10. Training results of different algorithms. the mAP is the average accuracy of the algorithm. precision is the accuracy of the algorithm. Train's loss is the loss of the training set, which consists of target box loss (box_loss), object loss (obj_loss), and classification loss (cls_loss).

TABLE 3. Comparison of AP for strawberry detection at different stages.

Growth Stage	YOLOv4	YOLOv5	STr	YOLOX	SDNet
MATURE	86.83	92.2	93.81	90.67	95.41
NEARLY_M	83.55	93.22	92.27	92.14	96.26
GROWING_W	85.12	90.21	89.93	89.35	95.58
GROWING_G	86.68	85.74	89.67	89.91	92.12
FLOWERING	74.67	84.58	88.25	88.87	92.52

2) COMPARISON OF PRECISION PER CATEGORY

The detection results of strawberries in different growth states of fruits directly reflect the ability of the models to extract and discriminate the target features, so the detection results of individual state periods of the five models were further compared, as shown in Table 3. For the maturity and coloring stage phases, the detection accuracy of "MATURE" and "NEARLY_M" was 95.41% and 96.26%, respectively. This is an improvement of 4.74% and 4.12% over the original YOLOX, proving that the algorithm can identify coloring and ripe fruits more accurately. The detection accuracy in the "GROWING_G" state was the lowest, 92.12% because the target fruit size of green-fruited strawberries was smaller, providing less information and harder to distinguish from the background. In summary, SDNet has improved the accuracy for each state period and can accurately detect the growth condition of strawberries.


(a)SDNet

(b)YOLOX

(c)YOLOv4

(d)YOLOv5

FIGURE 11. Comparison of the effect of different detection algorithms:(a)SDNet; (b) YOLOX-l; (c) YOLOv4; (d) YOLOv5; where the black boxes put most of the cases for partial misses and misdetections.

3) COMPARISON OF DETECTION RESULTS

The detection results of different algorithmic models are shown in Fig. 11., and four very representative detection results are selected for a brief analysis. To quickly analyze the model detection results, labels with "MATURE", "NEARLY_M", "GROWING_W", "GROWING_G" and "FLOWERING" were used with different colors to frame different growth periods, such as ripening, coloring, white ripening, green fruit, and flowering. The detection results of different algorithmic models are shown in Fig. 11., and four very representative detection results are selected for a brief analysis. As shown in Fig. 11.(a), the proposed SDNet detection model applies to the detection of strawberry growth stages under shading conditions, detecting with high confidence the fruits in the white fruit stage and the flowers in the strawberry flowering stage, where more than half of the fruits are shaded by the leaves. It can also be applied to the detection of heterogeneous strawberry fruits, small target fruits, and fruits with different multiple fruit positions. In Fig. 11.(b,c,d), the black boxes are selected and enlarged as the results caused by missed and false detection by the corresponding detection algorithm, which seriously affects the effectiveness of the detection model and is very detrimental to the subsequent operations such as autonomous strawberry picking, spraying, and fertilization. Although fewer fruits were missed in Fig. 11.(c, d), their overall confidence level was low and prone to false detection. As can be seen from the detection effect plots, the SDNet model effectively solves the problems of low confidence and low recognition rate of obscured and small target fruits in the existing models for fruit detection at different growth stages of strawberries.

4) ABLATION EXPERIMENTS

The above results show that the SDNet designed in this study achieved the most accurate detection results in detecting strawberries at different growth stages, which illustrates the effectiveness of SDNet.To further validate the effects of hybrid Mosaic and Mixup data enhancement methods, C3HB feature extraction module, normalized attention mechanism NAM and SIOU loss function modules on strawberry fruit detection at different growth stages, ablation experiments were designed in this research, and the results are shown in Table 4.

TABLE 4. Ablation experiments.

Model	Precision (%)	Recall (%)	mAP (%)	FPS
Base (YOLOX)	89.51	88.68	90.18	26.6
+DH	89.87	88.75	90.80	26.1
+C3HB	91.29	90.81	92.95	29.7
+NAM	91.30	91.02	93.51	29.8
+SIOU (SDNet)	93.15	90.72	94.26	30.5

The experimental results using the hybrid data augmentation method resulted in only 0.07%, 0.62%, and 0.36% improvement in the recall, mAP, and precision values for each growth length state of the model, which is a certain enhancement to the performance of the model. The new C3HB module has been added to increase the global feature extraction and spatial interaction capability of the model, resulting in further improvements of about 2.15%, 1.42%, and 2.06% in the parameters of mAP, precision, and recall. And because Horblock uses hole convolution also makes the model's real-time detection better, with FPS increasing from 26.1 to 29.7. The addition of the NAM module and SIOU ultimately improved the model mAP to 94.26% and the frames per second transferred FPS to approximately 30.5fps on the test device RTX 2060 (6G VRAM). The specific details of the comparison are shown in Fig.12.

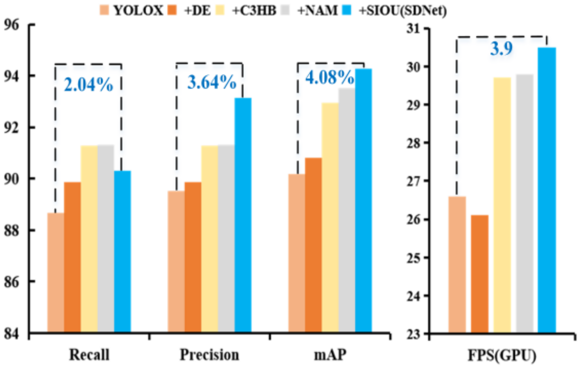


FIGURE 12. Ablation experiment comparison.

C. FEATURE MAP ANALYSIS

We visualize the feature maps of the SDNet detection process to get a more intuitive sense of the strengths and weaknesses of the detection model. The visualization processing of feature maps refers to the feature maps output by each intermediate module (including modules such as a convolutional layer, C3HB module, and NAM) after model processing for a given input image display.

As shown in Fig. 13, four sets of feature maps extracted from strawberry maps in different natural environments were randomly selected to analyze the strengths and weaknesses of the model. It can be seen that in the shallow feature map, there is not only the feature information of strawberry fruit but also the cluttered background and texture information. As the network goes deeper and deeper, the information extracted from the feature map is continuously abstracted and blurred to retain and enhance the strawberry fruit information with high-dimensional deep semantic feature information. After the model feature extraction, only strawberry feature information of different growth stages is highlighted in the feature map, which effectively shows the effectiveness and accuracy of the detection model. The SDNet detection model increases the perceptual domain of the feature map in the detection process and refines the feature information of strawberries at different growth stages, making the detection model more accurate and effective.

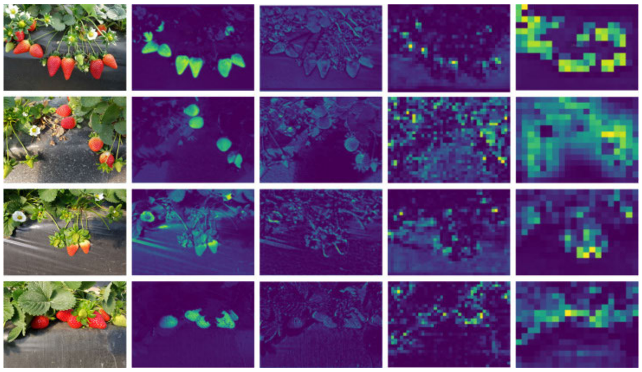


FIGURE 13. Strawberry detection feature map.

D. REAL-TIME DETECTION

In the detection of strawberry growth stages, only the location information and confidence degree of strawberry were visualized, while the quantity information was ignored, and the fruit quantity of each growth stage could not be obtained intuitively. Therefore, in this study, in order to display the fruit quantity information of different growth stages of strawberry more accurately, the fruit identification quantity of different growth stages was visualized by autonomous programming and the quantity information was directly output in real time combined with videos and pictures. As shown in Fig. 14., the SDNet algorithm was performed to experiment the real-time detection effect. The top left corner of the figure shows the real-time frames per second FPS, while the bottom left corner

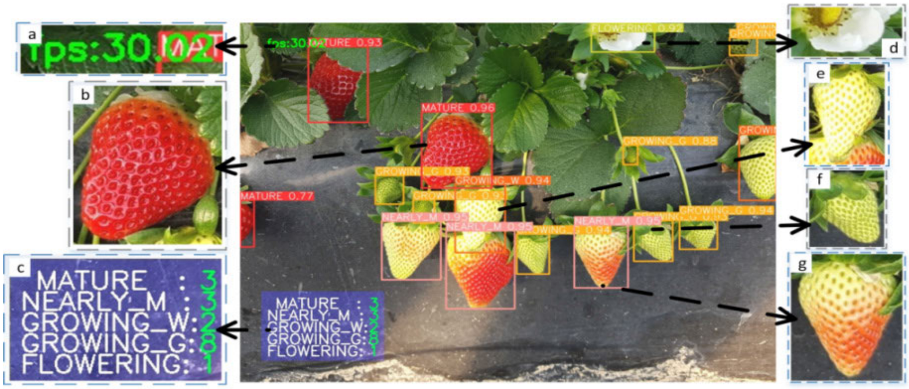


FIGURE 14. Real-time detection effect map.

visualizes the fruit quantity information of strawberries at different growth stages in the real-time detection screen. The monitoring method can be used with a simple environmental configuration on different devices.

IV. CONCLUSION

In this research, we designed and developed a method to detect the growth state of strawberry fruits based on the SDNet algorithm, and finally realized the real-time monitoring of strawberry fruits in five different states. The experimental results are as follows.

(1) This study tested the improvement of different modules on the performance of the YOLOX algorithm and verified the effectiveness of the improved model. For example, the CSP block is replaced by a self-designed C3HB block in the backbone part; the normalized attention module (NAM) is introduced in the neck and the latest SIOU loss function is embedded. The mAP, precision, and recall are improved by 4.08%, 3.64%, and 2.04%, respectively. And compared with five popular detection algorithms, SDNet has the best detection results. SDNet also meets the requirement of real-time monitoring.

(2) Deep learning-based strawberry growth state monitoring is server-dependent and still has limitations. Future research will focus on operations such as picking, spraying, and fertilizing relative to each detected growth state in order to build out a complete automated orchard management robot that can contribute to subsequent unmanned farms and precision agriculture.

REFERENCES

[1] P. Constante, A. Gordon, O. Chang, E. Pruna, F. Acuna, and I. Escobar, "Artificial vision techniques to optimize Strawberry's industrial classification," *IEEE Latin Amer. Trans.*, vol. 14, no. 6, pp. 2576–2581, Jun. 2016, doi: 10.1109/TLA.2016.7555221.

[2] Y. Ge, Y. Xiong, G. L. Tenorio, and P. J. From, "Fruit localization and environment perception for strawberry harvesting robots," *IEEE Access*, vol. 7, pp. 147642–147652, 2019, doi: 10.1109/ACCESS.2019.2946369.

[3] T. Ilyas, A. Khan, M. Umraiz, Y. Jeong, and H. Kim, "Multiscale context aggregation for strawberry fruit recognition and disease phenotyping," *IEEE Access*, vol. 9, pp. 124491–124504, 2021, doi: 10.1109/ACCESS.2021.3110978.

[4] Y. Yu, K. Zhang, H. Liu, L. Yang, and D. Zhang, "Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot," *IEEE Access*, vol. 8, pp. 116556–116568, 2020, doi: 10.1109/ACCESS.2020.3003034.

[5] A. Subeesh, S. Bhole, K. Singh, N. S. Chandel, Y. A. Rajwade, K. V. R. Rao, S. P. Kumar, and D. Jat, "Deep convolutional neural network models for weed detection in polyhouse grown bell peppers," *Artif. Intell. Agricult.*, vol. 6, pp. 47–54, Jan. 2022.

[6] D. S. Prabha and J. S. Kumar, "Assessment of banana fruit maturity by image processing technique," *J. Food Sci. Technol.*, vol. 52, no. 3, pp. 1316–1327, Mar. 2015, doi: 10.1007/s13197-013-1188-3.

[7] S. Lim, J. G. Lee, and E. J. Lee, "Comparison of fruit quality and GC–MS-based metabolite profiling of kiwifruit 'Jecy green': Natural and exogenous ethylene-induced ripening," *Food Chem.*, vol. 234, pp. 81–92, Nov. 2017, doi: 10.1016/j.foodchem.2017.04.163.

[8] I. Campos, R. Bataller, R. Armero, J. M. Gandia, J. Soto, R. Martínez-Máñez, and L. Gil-Sánchez, "Monitoring grape ripeness using a voltammetric electronic tongue," *Food Res. Int.*, vol. 54, no. 2, pp. 1369–1375, Dec. 2013, doi: 10.1016/j.foodres.2013.10.011.

[9] M. Larrain, A. R. Guesalaga, and E. Agosin, "A multipurpose portable instrument for determining ripeness in wine grapes using NIR spectroscopy," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 294–302, Feb. 2008, doi: 10.1109/TIM.2007.910098.

[10] Y. Shao, Y. Wang, and G. Xuan, "In-field and non-invasive determination of internal quality and ripeness stages of feicheng peach using a portable hyperspectral imager," *Biosyst. Eng.*, vol. 212, pp. 115–125, Dec. 2021, doi: 10.1016/j.biosystemseng.2021.10.004.

[11] J. Zhang, X. Wang, J. Xia, S. Xing, and X. Zhang, "Flexible sensing enabled intelligent manipulator system (FSIMS) for avocados (persea Americana Mill) ripeness grading," *J. Cleaner Prod.*, vol. 363, Aug. 2022, Art. no. 132599, doi: 10.1016/j.jclepro.2022.132599.

[12] G. Ayalew, Q. U. Zaman, A. W. Schumann, D. C. Percival, and Y. K. Chang, "An investigation into the potential of Gabor wavelet features for scene classification in wild blueberry fields," *Artif. Intell. Agricult.*, vol. 5, pp. 72–81, Jan. 2021, doi: 10.1016/j.aiia.2021.03.001.

[13] S. O. Cisneros, J. M. R. Varela, M. A. R. Acosta, J. R. Dominguez, and P. M. Villalobos, "Pollen grains classification with a deep learning system GPU-trained," *IEEE Latin Amer. Trans.*, vol. 20, no. 1, pp. 22–31, Jan. 2022, doi: 10.1109/TLA.2022.9662170.

[14] D. Elavarasan and P. M. D. Vincent, "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications," *IEEE Access*, vol. 8, pp. 86886–86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

[15] P. K. Kashyap, S. Kumar, A. Jaiswal, M. Prasad, and A. H. Gandomi, "Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17479–17491, Aug. 2021, doi: 10.1109/JSEN.2021.3069266.

[16] S. Shorewala, A. Ashfaque, R. Sidharth, and U. Verma, ''Weed density and distribution estimation for precision agriculture using semi-supervised learning,'' *IEEE Access*, vol. 9, pp. 27971–27986, 2021, doi: 10.1109/ACCESS.2021.3057912.

[17] W. Castro, J. Oblitas, M. De-la-Torre, C. Cotrina, K. Bazán, and H. Avila-George, ''Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces,'' *IEEE Access*, vol. 7, pp. 27389–27400, 2019, doi: 10.1109/ACCESS.2019.2898223.

[18] P. Wan, A. Toudeshki, H. Tan, and R. Ehsani, ''A methodology for fresh tomato maturity detection using computer vision,'' *Comput. Electron. Agricult.*, vol. 146, pp. 43–50, Mar. 2018, doi: 10.1016/j.compag.2018.01.011.

[19] W. Xu, L. Zhao, J. Li, S. Shang, X. Ding, and T. Wang, ''Detection and classification of tea buds based on deep learning,'' *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106547, doi: 10.1016/j.compag.2021.106547.

[20] H. Habaragamuwa, Y. Ogawa, T. Suzuki, T. Shiigi, M. Ono, and N. Kondo, ''Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network,'' *Eng. Agricult., Environ. Food*, vol. 11, no. 3, pp. 127–138, Jul. 2018, doi: 10.1016/j.eaef.2018.03.001.

[21] 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, doi: 10.1016/j.compag.2019.06.001.

[22] Y. Chen, W. S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang, and Y. He, ''Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages,'' *Remote Sens.*, vol. 11, no. 13, p. 1584, Jul. 2019, doi: 10.3390/rs11131584.

[23] H. Zheng, G. Wang, and X. Li, ''Identifying strawberry appearance quality by vision transformers and support vector machine,'' *J. Food Process Eng.*, vol. 45, Oct. 2022, Art. no. e14132, doi: 10.1111/jfpe.13982.

[24] Y. Zhang, J. Yu, Y. Chen, W. Yang, W. Zhang, and Y. He, ''Real-time strawberry detection using deep neural networks on embedded system (RTSD-Net): An edge AI application,'' *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106586, doi: 10.1016/j.compag.2021.106586.

[25] *China Agricultural Strawberry Industry Standard NY/T 444-2001, Compilation of Chinese Agricultural Standards. Fruits and Vegetables Volume (Next Volume)*, China Standard Publishing House, Beijing, China, 2004.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[27] J. Redmon and A. Farhadi, ''YOLO9000: Better, faster, stronger,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[28] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, *arXiv:1804.02767*.

[29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, *arXiv:2004.10934*.

[30] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, ''YOLOX: Exceeding Yolo series in 2021,'' 2021, *arXiv:2107.08430*.

[31] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S.-N. Lim, and J. Lu, ''HorNet: Efficient high-order spatial interactions with recursive gated convolutions,'' 2022, *arXiv:2207.14284*.

[32] Y. Liu, Z. Shao, Y. Teng, and N. Hoffmann, ''NAM: Normalization-based attention module,'' 2021, *arXiv:2111.12419*.

[33] Z. Gevorgyan, ''SIoU loss: More powerful learning for bounding box regression,'' 2022, *arXiv:2205.12740*.

QILIN AN is currently pursuing the master's degree with the College of Engineering, China Agricultural University. He is working on robotics research and industrial automation design. His research interests include applications of artificial intelligence in agriculture, mobile robot target localization, autonomous navigation, and deep learning-based machine.

KAI WANG received the Ph.D. degree in engineering from Nanjing Agricultural University, in 2021. He is working on mobile robot navigation and environment perception technology, robotic arm target detection, and grasping technology. His research interests include agricultural robotics, agricultural navigation, environment perception, and SLAM.

ZHONGYANG LI is currently pursuing the master's degree with the College of Engineering, China Agricultural University. He is working on the research and design of agricultural robots. His research interests include 3D target reconstruction, autonomous robot navigation, path planning, and SLAM.

CHENGYUAN SONG is currently pursuing the master's degree in engineering with Weifang University, Weifang, China. Since, he has been involved in many non-standard machinery and equipment development, industrial control, and robot control projects. His current research interests include industrial control, machine vision, and image processing.

XIUYING TANG received the Ph.D. degree. He is currently pursuing the doctoral degree. He is a Professor with the College of Engineering, China Agricultural University, and the Supervisor of master's student. He has presided over three national research projects and participated in six national projects. He has published more than 40 research articles, including nearly 30 SCI/EI articles. He has declared more than 20 invention patents for related research. His research interests include agricultural non-destructive testing technology, agricultural robotics, and optomechanical integration technology.

JIAN SONG received the M.S. and Ph.D. degrees in engineering from China Agricultural University, Beijing, China, in 2003 and 2006, respectively. He is currently a Professor with the School of Mechatronics and Vehicle Engineering, Weifang University, Weifang, China. His current research interests include robotics and intelligent agricultural equipment, machine vision and image processing, and machine learning.