Green Apple Detection Method Based on Optimized YOLOv5 Under Orchard Environment

Weike Zhang, Yanna Zhao, Yujie Guan, Ting Zhang, Qiaolian Liu, Weikuan Jia

***Abstract*—In orchards, detecting green apples can be challenging due to interference factors like similar fruit and background color, branch and leaf shading, and fruit overlap. To address this limitation, this paper presents a simple yet effective detection model based on improved YOLOv5, which can enhance the detection ability of green apples against a near-color background. Our contributions are twofold. Firstly, we added an attention mechanism to enhance the feature extraction network of the conventional YOLOv5. This modification focuses the network on green apple features and improves the detection performances on green apples. Secondly, we introduced the Focal Loss calculation method to the loss calculation of YOLOv5, to improve the model's results by controlling positive and negative sample weights as well as hard and easy to classify sample weights. Experimental results show that our model yields better results. While the base YOLOv5 model achieved an Average Precision (AP) of 86.3% and an Average Recall (AR) of 66.8% on the green apple dataset test, our improved YOLOv5 model reached an AP of 88.1% (a 1.8 percentage point improvement) and an AR of 69.1% (a 2.3 percentage point improvement). Our proposed model, therefore, significantly enhances detection efficiency.**

***Index Terms*—Green apple; Target Detection; YOLOv5 model; Attention Mechanism; Focal Loss**

I. Introduction

The development of science and technology has transformed the agricultural management mode, resulting in a significant increase in agricultural productivity through the incorporation of intelligent agricultural equipment. With one such example being the use of vision systems by intelligent agricultural equipment, which serve as the equipment's "eyes," and provide vital environmental perception information. In an agricultural or orchard environment, vision systems are capable of target fruit detection, which allows for crop maturity monitoring [1], yield predictions [2][3][4], automatic fruit picking [5][6][7], diagnosis of pest and disease spotting [8], and more. However, the complexity of the orchard environment adds numerous factors that can affect the accuracy of the vision system, including light and camera angles, background and fruit color similarity, shading of branches and leaves, as well as, overlapping of fruits leads to the possibility of missed and misidentified fruit, challenging the accurate and efficient detection of fruit.

Conventional machine learning algorithms have accumulated significant results in the field of target fruit detection. For instance, Li [9] proposed an improved spectral clustering algorithm to address the issue of overlapping fruits in natural environments. It employs a spectral clustering algorithm to segment the image first and then uses the randomized Hough transform to achieve fruit identification and localization. Seng [10] used nearest neighbor classification to obtain feature values that can fuse three features of color, shape, and size to detect fruit accurately. Song [11][12] proposed a method based on convex hull theory to handle identification and localization of obscured apple targets and an algorithm based on convex hull for segmentation and reconstruction of overlapping apple targets for branch occlusion and fruit overlap, respectively. Huang [13] proposed a new optimized method based on the framework of discriminative region feature integration (DRFI) algorithm that combines color, texture, and shape features of green fruits to achieve the detection of green fruits under the interference of various factors of natural conditions. However, the conventional target detection algorithm is greatly affected by the complex orchard environment, and it cannot meet the requirements of practical operation, leading to a bottleneck in the development of orchard target fruit recognition.

Deep learning technology has been increasingly applied to various fields, including facial detection [14], classification of MI signals in brain-computer interface [15], extraction of semantic maps of roads [16], and malaria parasite detection [17]. Naturally, Deep Learning has also found widespread application in fruit detection, bringing about significant improvements in the performance of the detection and the models' reliability. To provide technical support for intelligent fig planting management, Wu [18] proposed a fig fruit recognition method based on the YOLOv4 Deep Learning technique. In order to enable better agricultural tasks, Bargoti [19] devised a method for fruit detection by combining data augmentation techniques and Faster-RCNN. Zhao [20] proposed an apple localization method based on the YOLOv3 deep convolutional neural network, which

Manuscript received Jan. 2, 2023; revised May 25, 2023.
This work is supported by Natural Science Foundation of Shandong Province in China (ZR2022MF349, ZR2020MF076); New Twentieth Items of Universities in Jinan (2021GXRC049).
W.K. Zhang is postgraduate student of School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China (e-mail: 1984961276@qq.com)
Y.N. Zhao is Associate Professor of School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China (e-mail: yannazhao@outlook.com)
Y.J. Guan is postgraduate student of School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China (e-mail: 1399973427@qq.com)
T. Zhang is lecturer of School of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China (e-mail: zhangting0607@126.com)
Q. Liu is lecturer of School of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China (Corresponding author, phone: +86-632-3785947; fax: +86-632-3785947; e-mail: llgx1207@126.com)
W.K. Jia is Associate Professor of School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China (Corresponding author, phone: +86-531-86181755; fax: +86-531-86181750; e-mail: jwk_1982@163.com)

enables the picking robot to overcome various influencing factors to pick apples around the clock and to improve its efficiency and ability to identify fruits. To better apply the excellent performance of Deep Learning to fruit detection in orchards, Jia [21] devised a target fruit recognition algorithm based on Mask R-CNN, which combines ResNet and DenseNet as a feature extraction network, greatly improving the performance of fruit detection. For accurate counting of visible fruits in image sequences, Liu [22] combined depth segmentation, frame-to-frame tracking and 3D localization to devise a fruit counting method. Jia [23] proposed a fast and effective Foveabox detection model to improve the performance of fruit detection, achieving quick identification and localization of green apples. Despite the above-mentioned results, there are still significant challenges in detecting target fruit in orchards.

This paper introduces improvements to the YOLOv5 model, by incorporating ECA-Net and CBAM in its feature extraction. This mitigates the issues related to tree branch occlusion and fruit overlap, commonly occurring in the orchard's near-color background. These modifications help the model focus on fruit features during the feature extraction process. In addition, Focal Loss is employed in the loss calculation to assign weightage to positive and negative samples, as well as hard and easy classification samples. This enhances the model's performance.

The following chapters are structured as such: The second chapter outlines the acquisition and labeling process used for fruit images. In the third chapter, the network structure and loss function of our improved YOLOv5 model are presented. Finally, the fourth chapter provides a comparative analysis of our model against YOLOv5 and other Deep Learning-based models to demonstrate superior performance.

II. Production of Dataset

Target detection of fruits against a green background is challenging due to the similarity between the fruit color and the background color, making it difficult to distinguish the fruit features. To improve the model's accuracy in detecting target fruits against a green background, a dataset of high-quality images of green fruits was curated and used.

A. Image acquisition

To ensure accurate detection of green apples against a green background, this research focuses on using green apples as the target object. The images were captured from an apple production base located in Longwang Mountain, Fushan District, Yantai City, Shandong Province.

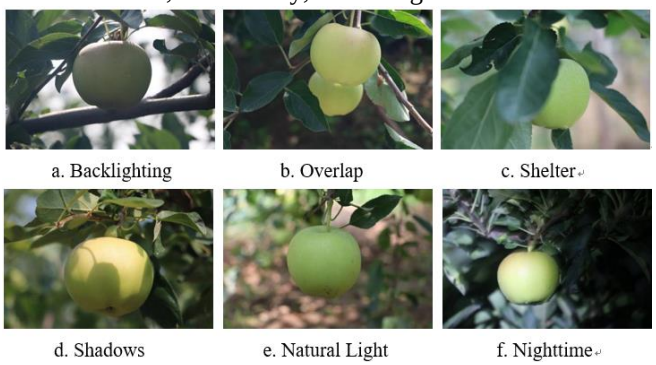


Fig. 1. Green apple in various cases.

Images were captured using a Canon EOS 80D DSLR camera equipped with a CMOS image sensor having a resolution of 6000×4000 pixels. To simplify the model, the images were compressed to 600×400 pixels and saved in .jpg format with 24-bit color depth.

Apple acquired 1361 images. There are 953 images for the training set and 408 images for the test set. These images were captured in various scenarios, such as backlighting, shading, overlapping, shadows, daytime and nighttime, as depicted in Figure 1.

B. Image annotation

LabelMe software was utilized in the experiment to annotate images of green apples. Initially, a marker point was used to trace the outline of the green fruit in each image, thus forming a boundary between the fruit and the background. This boundary aided in determining the position of the fruit within the picture. Subsequently, all essential information pertaining to the image and its annotation points is recorded in a corresponding json file. Finally, the collected data is converted into the COCO format dataset.

III. Optimized Yolo v5 Detection Algorithm

In real-world applications of orchards, fruit detection while in motion is crucial. Thus, a simple and efficient model is necessary. This study utilized the fifth generation of YOLOv5s for enhancement. YOLOv5s has the smallest depth and feature map width in the YOLOv5 family. Consequently, YOLOv5s is exceptionally rapid and appropriate for identifying fruits swiftly.

A. Basic Network

The YOLOv5 [24][25][26][27] network is composed of four main segments: the input section, backbone network, neck network, and head network. The input section includes Mosaic Data Augmentation, adaptive anchor box calculation, and adaptive image scaling. Meanwhile, the backbone network, which is formed by the CSPDarknet structure, is composed of Focus structure, Conv2D_BN_SiLU (CBS) structure, CSP structure, and SPP structure. These structures provide three effective feature layers, obtained through convolution modules, residual modules, and pooling operations, that are fed into the neck network. Essentially, the neck network enhances the effective feature layers and fuses them through the construction of the FPN. Finally, the head network adjusts the number of channels through a convolutional network and predicts the final results.

B. Improved YOLOv5 network

A higher level of independence of the fruit alongside clear edges will improve the accuracy of the model's detection capabilities. Nonetheless, orchard environments are highly intricate, and various obstructive factors can influence the fruit images captured. In particular, the similarity of fruit color and background, as well as branch and leaf shading, are environmental elements that contribute to the unclear edges in fruit images. Such factors can significantly impair the accuracy of the model's detection capabilities.

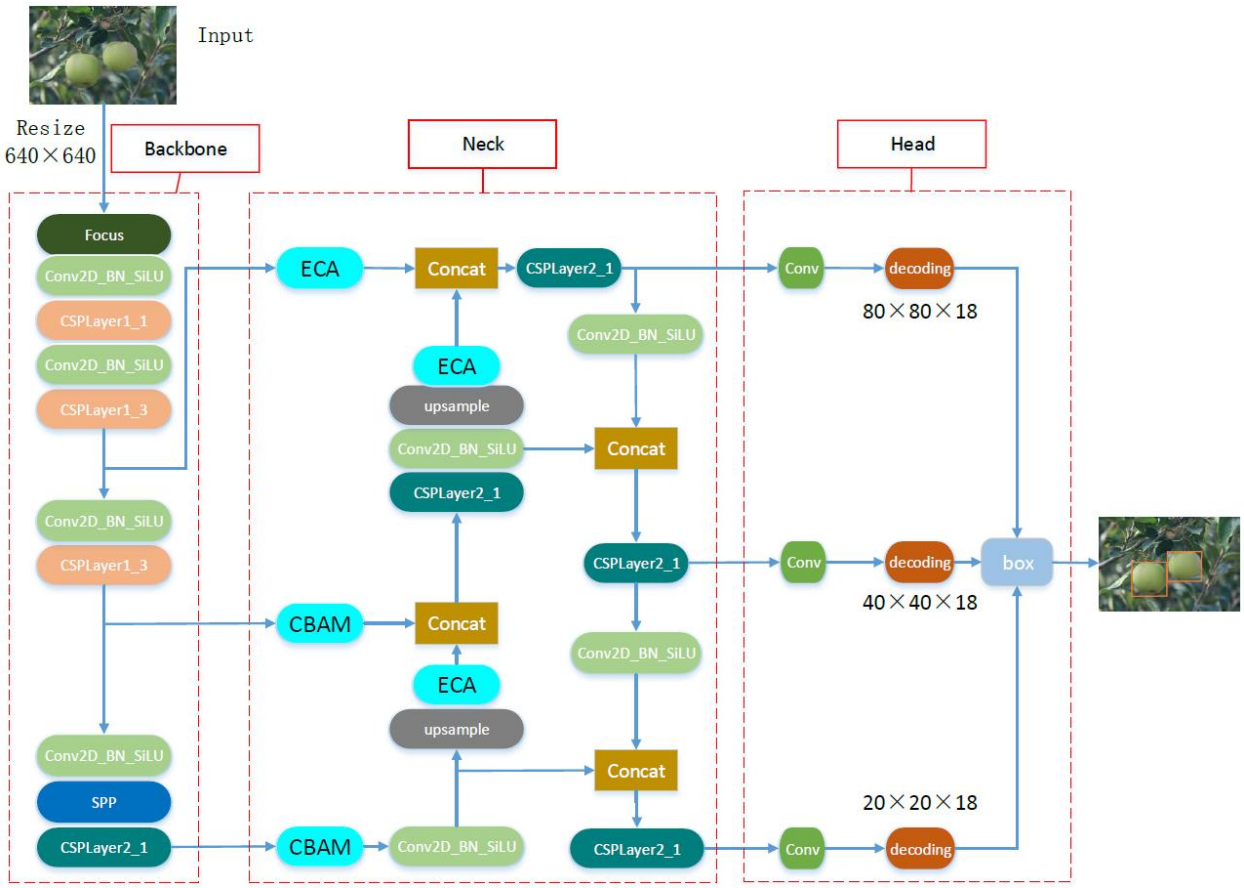


Fig. 2. Network structure of the improved YOLOv5 model.

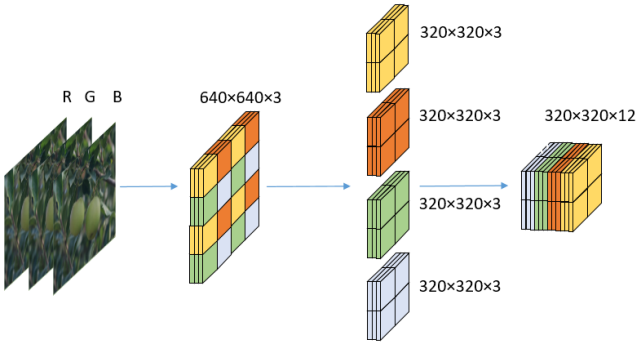


Fig. 3. Focus structure

Overall Network Structure

The simple structure of the YOLOv5 model detrimentally affects its detection performance in complex orchard environments. To address the unclear edges of fruit images, this paper proposes an improvement to the YOLOv5s model to increase its adaptability to detect green fruits in complex orchard environments. By incorporating attention mechanisms such as ECA-Net [28][29] and CBAM [30][31][32], the enhanced model focuses on fruit features, allowing for better fruit identification and classification in images. Figure 2 illustrates the structure of the improved model.

Specifically, the input fruit image is first resized to 640 × 640 × 3 after Mosaic Data Augmentation, and then to 320 × 320 × 12 after the Focus structure, as illustrated in Figure 3. In the Focus structure, a pixel value is taken every other pixel on the green apple image, then pixels of the same type are combined to form a feature layer, and finally these feature layers are stitched together. The end result is that the height and width of the fruit image is reduced by half and the number of channels is expanded to four times as many as before, achieving the effect of subsampling. This allows the height and width information of the fruit image to be fused to the channels without loss of feature information, making it easier to extract the fruit features.

After increasing the number of channels, feature extraction is carried out using convolutional network modules and residual network modules to obtain several feature layers. From these layers, the Neck selects three effective feature layers of varying scales for feature fusion. These three layers have sizes of 80×80×256, 40×40×512, and 20×20×1024, respectively. One of the feature layers, which has a size of 20×20×1024, is obtained through the SPP structure illustrated in Figure 4. The SPP structure downscales fruit features using a CBS module, passes them through three different max pooling layers, stitches the resulting scaled and pooled features, and then further adjusts the number of channels through another CBS. The SPP structure significantly improves the network's perceptual field, thereby enhancing fruit feature fusion.

ECA-Net

In this paper's feature fusion module of Neck, the ECA-Net attention mechanism is added after the 80×80×256 feature layer and upsampling to differentiate the background and fruit clearly and fuse the fruit features efficiently. The ECA-Net's cross-channel interactivity ensures that each fruit feature point has multiple features, improving the accuracy of the classifier and retractor. Figure 5 demonstrates the first step involves global average pooling (GAP) on the height and width of the fruit feature layer, followed by a 1D convolutional network that learns the channel information of the fruit feature with the same channel dimension. The

resulting value is considered a sigmoid value, and it becomes a weight for each channel, which multiplies with the initial fruit feature layer to obtain the fruit feature with channel attention.

ECA-Net's introduction not only enhances feature fusion but also adds minimal complexity, leading to improved fruit detection performance. Additionally, the model's detection speed is preserved, meeting the real-time demands of agricultural operations.

CBAM

To enhance the model's performance by providing more information on fruit features and better fusion of fruit features, this paper incorporates CBAM after two effective feature layers, 40×40×512 and 20×20×1024. CBAM comprises channel attention and spatial attention which improves attention to features in both channel and spatial dimensions, as illustrated in Figure 6. The layer with information on fruit features of size H×W×C enters the Channel Attention Module where global average pooling and global max pooling are performed on each fruit feature layer, creating two outputs that are processed by a shared, fully connected layer. The results are then added together to form a single output that is transformed into a sigmoid value as the weight of the Channel Attention. Next, the Spatial Attention Module performs max pooling and average pooling on the points of features of each fruit to obtain two outputs that are concatenated into a feature layer. A convolutional network adjusts the channel number to 1, from which the weight of Spatial Attention is obtained using a sigmoid function. Finally, CBAM multiplies the original feature layer with the weights of Channel Attention and the Spatial Attention to obtain a fruitful feature layer with both channel and spatial attention.

Introducing the CBAM attention mechanism can effectively help the network to focus on the significant features of the fruit while ignoring the similar background features. This results in the generation of a feature layer with a positive effect, thereby resolving the issue of unclear fruit image edges and improving the model's ability to detect fruits.

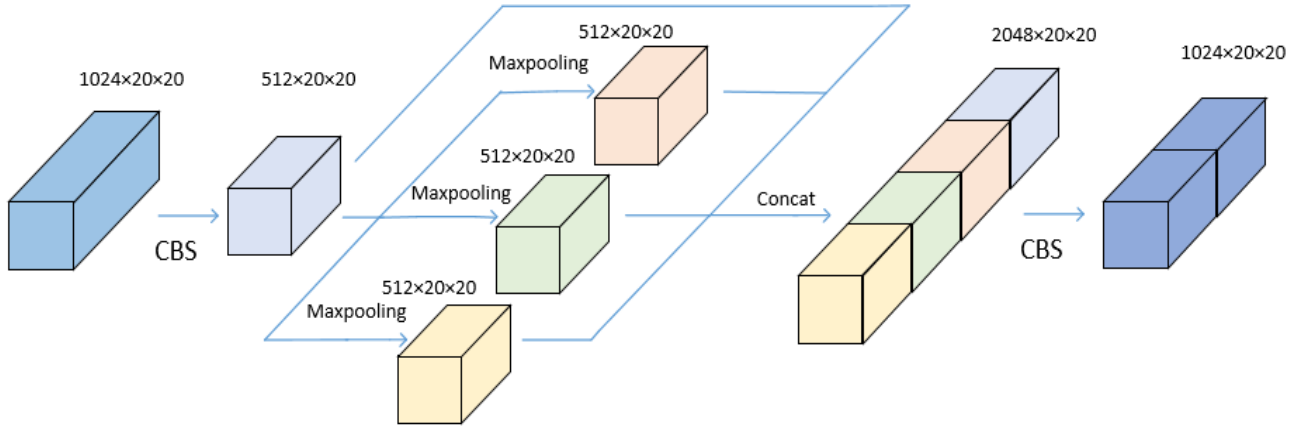


Fig. 4. SPP structure

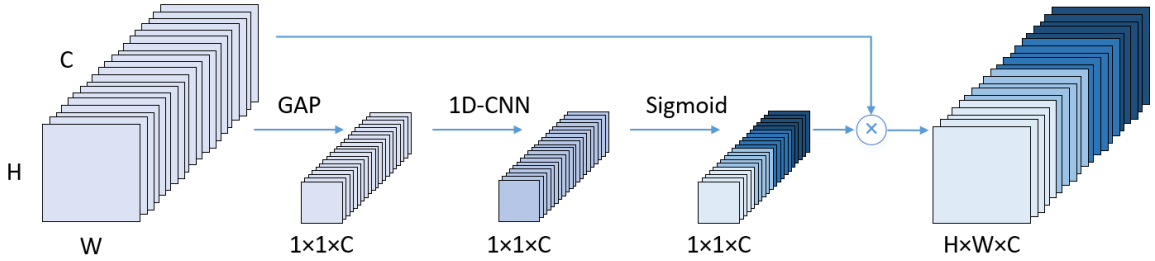


Fig.5. ECA-Net network structure

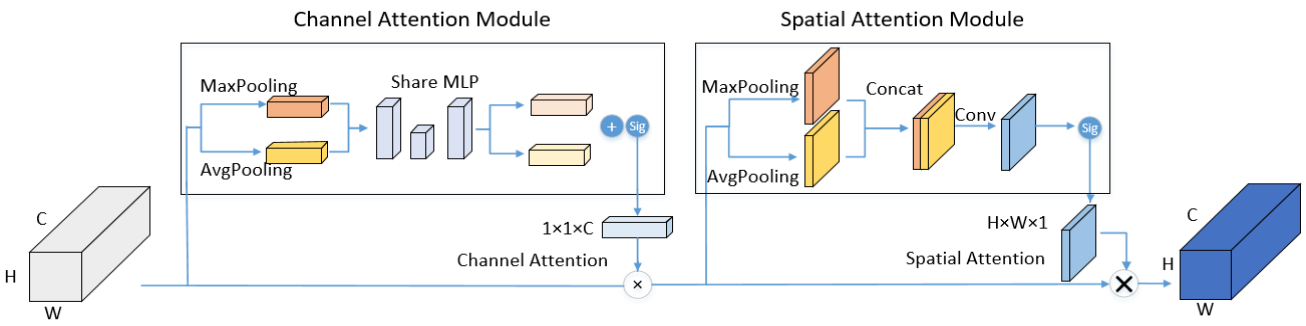


Fig. 6. CBAM network structure

C. Focal Loss function

Since the background environment color is similar to that of the fruit, it introduces bias in the Head prediction part of YOLOv5. Consequently, some non-fruits erroneously register as fruits. This condition generates an extensive number of false frames, which are negative samples that will not match the true frame. The majority of these false frames are easily identifiable samples. This situation in turn leads to the training process being focused on easily classified negative samples, thus providing little useful information to the training process. As a result, the model's training performance suffers. Additionally, the large number of easily classified negative samples can obscure the contribution of positive samples to loss calculation, leading to errors in the gradient update direction. The model thus fails to get useful fruit feature information, which ultimately leads to low fruit detection performance.

The loss function plays a crucial role in training the model. A well-suited loss function enhances the iterative optimization of the model training process, and the model achieves the best training effect by back propagation of the model gradient. This paper investigates the loss function of the improved YOLOv5 model to enhance its ability to detect fruitfulness.

YOLOv5's loss calculation formula, shown in Equation (1), comprises of three parts. The first part, loss_cls, calculates the classification loss using BCE loss but only for positive samples. The second part, loss_loc, uses GIoU loss to calculate the regression loss of positive samples only. Lastly, loss_conf calculates the target confidence loss of all samples using BCE loss.

$$\text{Loss} = \lambda_1 L_{cls} + \lambda_2 L_{loc} + \lambda_3 L_{conf} \quad (1)$$

where λ_1, λ_2, λ_3 are the equilibrium coefficients. To account for positive and negative samples in the calculation of the target confidence loss in the YOLOv5 loss function, this paper proposes the utilization of Focal Loss [33][34][35]. This is done to mitigate the adverse effects of negative samples that are easily classified on the model's detection performance by controlling the weights of positive and negative samples, as well as the weights of hard and easy samples. Equation (2) represents the formula of the Focal Loss function.

$$FL(p) = \begin{cases} -\alpha(1-p)^{\gamma}\log(p) & if\ y = 1 \\ -(1-\alpha)p^{\gamma}\log(1-p) & otherwise \end{cases} \quad (2)$$

y is the label value of the sample in the binary classification problem. The loss function is applied to positive samples when y equals 1 and to negative samples otherwise. The hyperparameter α is used to balance the contribution of positive and negative samples and is set to 0.25. The modulation factor γ is used to control the weights of easily and hardly classified samples. In this paper, γ is set to 2.2. By definition, p represents the probability that a sample is predicted as positive, and for samples with true negative labels, a smaller p signifies more accurate model prediction.

The introduction of the Focal loss effectively alleviates the influence of easily classified negative samples on the loss calculation and greatly enhances the contribution of a priori boxes that match real boxes to the model's predictions. This approach solves the problem of the imbalance of positive and negative samples in fruit detection. In addition, it greatly eases trouble in fruit detection caused by similar colors between fruit and the background.

IV. Experiments

This experiment aims to evaluate the enhanced detection capability of the improved YOLOv5 model in identifying green fruits against the green background. Initially, a comparison was made between the improved YOLOv5 model and the conventional YOLOv5 model to observe the effect of enhancements. Subsequently, the improved YOLOv5 model was pitted against other Deep Learning- based target detection models to draw comprehensive experimental conclusions.

A. Experimental environment

T The model operated on an Ubuntu 16.04 operating system and was equipped with an Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz processor, 64 GB of RAM, and 10 GB of graphics memory. CUDA 11.0, Python 3.7, and Pytorch 1.7.0 Deep Learning framework were utilized on a server to train the model.

To further optimize the model's detection capabilities, it was pre-trained with weights from the COCO dataset. The improved YOLOv5 model was then utilized to train the Green Fruit training dataset, and its performance was assessed by validation and evaluation on the Green Fruit validation dataset.

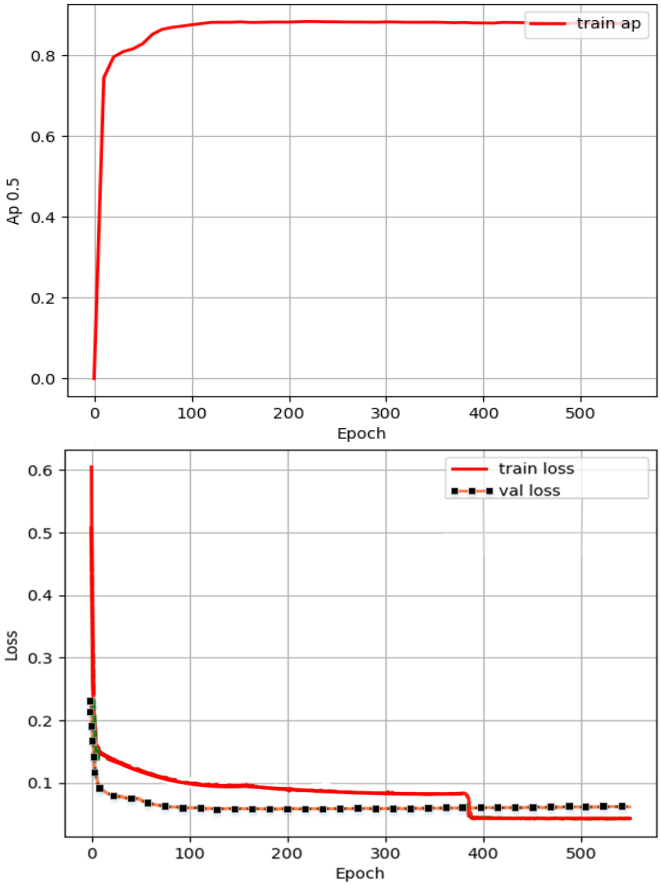


Fig.7.The curves of AP at IoU=0.50 and Loss

B. Train

In this study, we trained the model using the green apple dataset for 550 epochs and 52,800 iterations. To increase the dataset's diversity and the model's robustness, we applied Mosaic Data and Mixup Data Augmentation. The model used

label smoothing to prevent overfitting during the training, with a smoothing factor of 0.01. We used the SGD Optimizer with a momentum of 0.937 and a weight decay coefficient of 5e-4 to accelerate the model's convergence and avoid the issue of local optimum. We obtained training curve graphs for AP and Loss values, as illustrated in Figure 7. The convergence of both AP and Loss started from the 60th epoch, and by the 380th epoch, the training loss suddenly decreased while the alteration in the validation loss was not significant, implying an overfitting phenomenon. So to evaluate the model, we chose the appropriate training outcomes ranging from the 60th epoch to the 380th epoch.

C. Evaluation Criteria

This experiment employs Precision (P) and recall (R) as criteria to evaluate the efficacy of the model. Precision is measured using Equation (3), and Recall is measured using Equation (4). The model's merit is evaluated using various criteria, and this experiment focuses on Precision (P) and recall (R).

$$P = \frac{TP}{TP+FP} \tag{3}$$

$$R = \frac{TP}{TP+FN} \tag{4}$$

where TP stands for the number of true positive samples, FP stands for the number of false positive samples, and FN stands for the number of false negative samples. The Precision ratio refers to the proportion of true positive samples to all positive samples classified by the model's classifier; the Recall ratio refers to the proportion of true positive samples to all positive samples.

In order to conduct a thorough evaluation of the model, we utilize the average precision (AP) metrics at a specified IoU threshold and mAP. This approach is supported by the following equations:

$$AP_{IoU=i} = 1/101 \sum_{r \in R} p(r) \tag{5}$$

$$\text{mAP} = 1/10 \sum_{i \in I} AP_{IoU=i} \tag{6}$$

where i represents the IoU threshold and the letter I represents the set of IoU thresholds with a total of 10 values: [0.5, 0.55, 0.6, 0.65, ..., 0.99, 0.95]. r denotes Recall and R denotes the set of Recall, which consists of 101 values: [0, 0.01, 0.02, 0.03, ... 0.99, 1.0]. p(r) denotes the precision associated with the recall.

In addition, we use the following evaluation metrics.

Flops: The number of floating point operations per second, which measures the computational time complexity of the network model.

Params: The total number of parameters required in the training of the model, which measures the computational space complexity of the network model.

FPS: The number of frames per second, which means the number of images that can be processed in each second.

D. The detection effect of the improved model

A comparison of the proposed model with the conventional YOLOv5 model is showcased in this study to demonstrate the former's performance. The same experimental setup, dataset, and evaluation metrics are utilized to ensure fairness. Figure 8 presents a visual analysis of the detection outcomes. The first row of the images shows the detection results of the conventional YOLOv5 model, while the improved YOLOv5 model results are depicted in the second row. The triangular box in the first row indicates the misclassification of the background as the target fruit, whereas the rectangular box in the first row represents the failure to detect the target fruit.

Figure 8 illustrates that detecting independent green apples, there is no significant difference between the two models. However, the models perform differently when exposed to natural disturbances such as similar background and fruit color, overlapping fruit, branch, and leaf shading. The first column indicates that the improved YOLOv5 model does not, when compared to the conventional YOLOv5 model, incorrectly detect branches and leaves as apples. The second and third columns indicate that the improved YOLOv5 model performs better in identifying overlapping and obscured apples. To effectively showcase the model's superiority, we utilize AP at IoU = 0.50 and AR at maxDet = 100 as metrics in tandem with Flops, Params, and FPS.

The Table 1 presents the results of the comparison between the conventional YOLOv5 model and the improved YOLOv5 model.

TABLE I
PERFORMANCE COMPARISON OF THE IMPROVED YOLOV5 MODEL WITH THE ORIGINAL MODEL

Models	$AP_{IoU=0.50}$	$AR_{maxDet=100}$	Flops/G	Params/M	FPS
YOLOv5	86.30%	66.80%	**114.54**	46.63	**33.94**
Ours	**88.10%**	**69.10%**	114.56	46.63	32.29

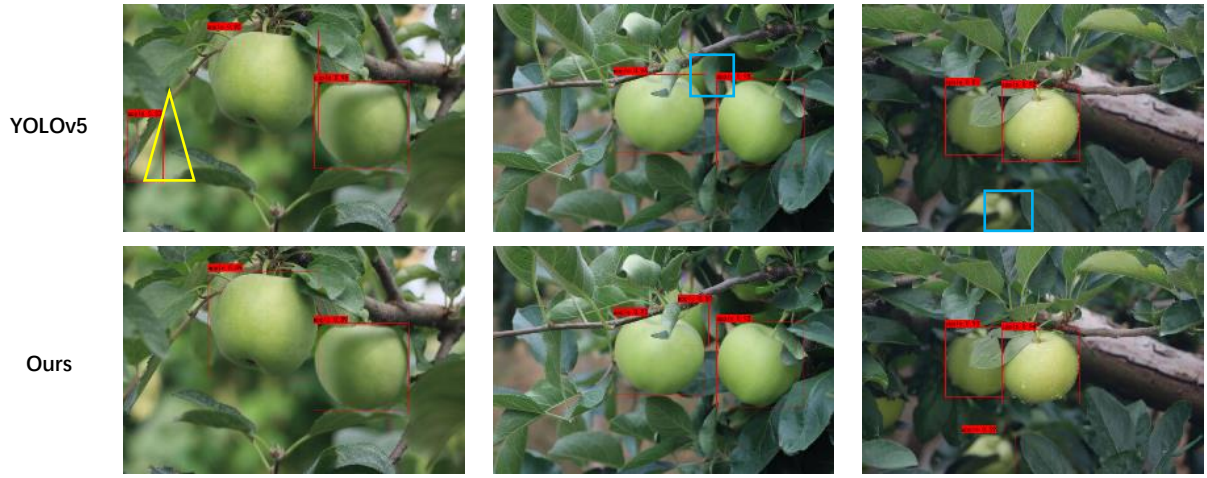
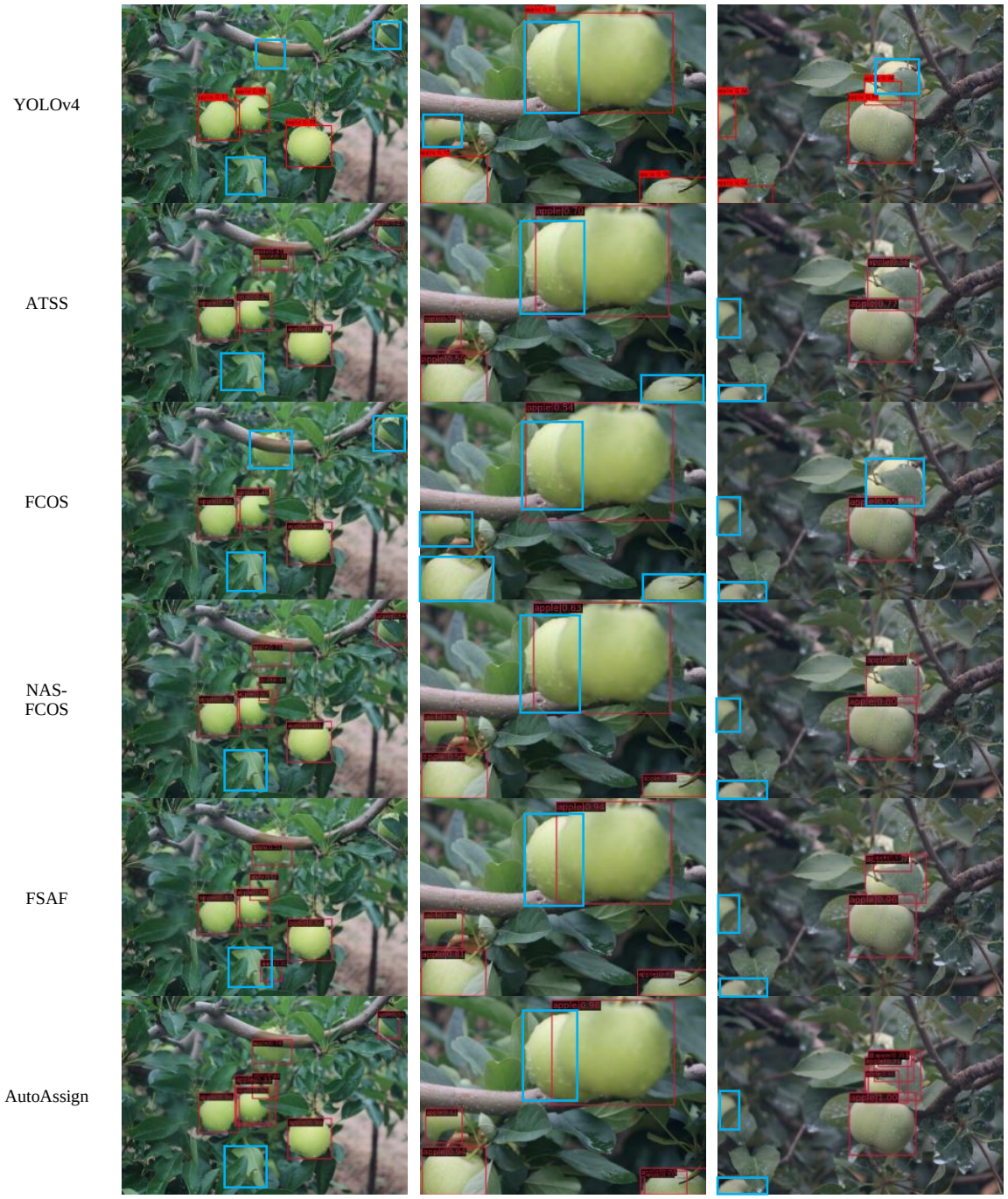


Fig. 8. Comparison of the detection of the model in this paper and the YOLOv5 model

Specifically, the improved YOLOv5 yielded an AP of 88.1%, exhibiting an increase of 1.8% in comparison to the conventional YOLOv5, which had an AP of 86.3%. Additionally, the improved YOLOv5 model increased the AR from 66.8% to 69.1%, showing a 2.3% enhancement. Meanwhile, there was no significant change in Params, and only a minimal increase of 0.02 G in Flops and decrease of 1.65 in Fps. These positive results indicate that the improved YOLOv5 model effectively enhances the detection accuracy with negligible reduction in detection speed. Therefore, it proves to be a more optimal solution for detecting green apples in orchards.

E. Comparison with other models

This study compares the detection results of the improved YOLOv5 with other state-of-the-art models on the green apple dataset. The green apple dataset was used to train YOLOv4, ATSS, FCOS, NAS-FCOS, FSAF, AutoAssign, Faster R-CNN, and Cascade R-CNN. Figure 9 presents the comparison of detection results of the models mentioned above, with YOLOv4, ATSS, FCOS, NAS-FCOS, FSAF, AutoAssign, Faster R-CNN, and Cascade R-CNN results listed consecutively from rows 1 to 8. In each row, rectangular boxes highlight green apples detected by the improved YOLOv5 model but missed by other models. And, triangular boxes indicate background or branches detected incorrectly as green apples by other model.



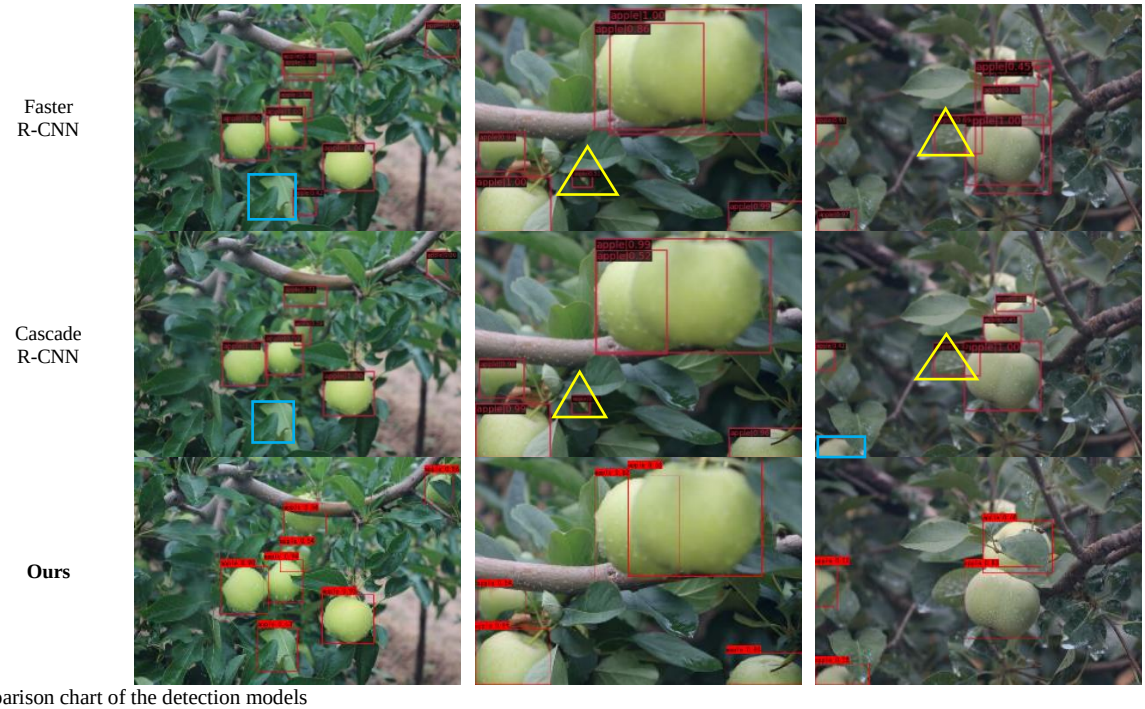


Fig. 9. Comparison chart of the detection models

TABLE II
PERFORMANCE OF THE IMPROVED YOLOV5 MODEL COMPARED WITH OTHER MODELS

Models	mAP	$AP_{IoU=0.50}$	$AR_{maxDet=100}$	Flops/G	Params/M	FPS
YOLOv4	56.4%	83.4%	67.8%	119.8	63.9	20.8
ATSS	54.4%	84.3%	63.1%	267.3	51.0	20.5
FCOS	47.2%	76.7%	58.3%	276.6	50.9	21.7
NAS-FCOS	58.3%	85.3%	66.6%	123.3	**38.1**	19.0
FSAF	60.7%	87.8%	67.5%	264.5	55.0	20.6
AutoAssign	57.7%	86.3%	67.0%	187.9	**35.9**	21.0
Faster R-CNN	58.1%	86.1%	65.4%	206.6	60.1	20.2
Cascade R-CNN	57.0%	83.1	62.6%	384.2	77.1	17.3
Ours	**61.2%**	**88.1%**	**69.1%**	**114.5**	46.6	**32.2**

The conditions in the orchard can significantly impede fruit detection due to various interferences. Figure 9 demonstrates that YOLOv4, ATSS, FCOS, NAS-FCOS, FSAF, AutoAssign, Faster R-CNN, and Cascade R-CNN are subject to fruit misses when detecting fruit that is obstructed by foliage and branches (as indicated by rectangular boxes). Due to overfitting, both Faster R-CNN and Cascade R-CNN can incorrectly identify branches and leaves as target fruit, resulting in triangular boxes. The final row depicts the improved YOLOv5, incorporating ECA-Net and CBAM, which enhances the model's sensitivity to fruit features and offers superior detection capability that can precisely detect the edges and areas of green apples even when masked by branches and leaves. In addition, the introduction of Focal Loss reduces the impact of easily classified negative samples on the model, resulting in the elimination of false positives samples due to the identification of other non-fruit targets as fruit.

To provide an equitable comparison of the models' performance, this research paper presents a comparative list of metrics (shown in Table 2) including mAP, AP at IoU=0.50, AR at maxDet=100, Flops, Params, and FPS.

According to Table 2, the improved YOLOv5 model generated 61.2% mAP, 88.1% AP for IoU=0.5, and 69.1% AR for maxDet=100, which exceed the metrics of the other models. Thus, the improved YOLOv5 model owns better detection capabilities without any misidentification or missing targets. The enhanced performance is attributed to two primary reasons. Firstly, the ECA-Net and CBAM made the fruit features more prominent, and secondly, Focal Loss reduced the impact of easily classified negative samples on the model gradient update direction. As a result of these improvements, this model achieved excellent accuracy and robustness.

The NAS-FCOS and AutoAssign models' parameters are slightly smaller than the 46.6M of the improved YOLOv5 model. Conversely, the other models' parameters are higher than the 46.6M of the improved model while all the Flops are more significant than the 114.5G of the improved model. This indicates that their computation is more extensive than the computation of the improved YOLOv5 model, and their FPS is lower than the 32.2 of the improved YOLOv5 model. The above outcomes demonstrate the superior performance of the improved YOLOv5 model in terms of speed and accuracy,

making it the most suitable smart device for real-time detection.

V. Conclusion

Detecting green fruits against the green background in a natural orchard environment is challenging due to factors such as color similarity between the fruit and background, occlusion caused by branches and leaves, and overlapping fruits. Thus, the detection model must have high accuracy. To address the aforementioned challenge, this paper proposes an improved YOLOv5 detection model. The proposed improvement includes two key modifications. First, we added attention mechanisms ECA-Net and CBAM to the feature extraction module of the base model. This modification enhances the model's focus on fruit features, thereby improving accuracy. Second, we added Focal loss to the loss calculation, which reduces the influence of negative samples on loss calculation. The results demonstrate that the improved model outperforms the original model. Evaluation against other models suggests the improved YOLOv5 model meets the accuracy requirement for identifying green fruits in complex orchard environments.

Real-time agriculture demands a model capable of achieving rapid target detection; however, the proposed model still requires improvements in terms of speed. Future work will aim to enhance the detection speed of the model while maintaining high detection accuracy. Specifically, the objective is to effectively apply the model to detect green fruits.

References

[1] Fengjun Chen, Xinwei Zhang, Xueyan Zhu, Zhiqiang Li, Jianhui Lin, "Detection of the olive fruit maturity based on improved EfficientDet," Transactions of the Chinese Society of Agricultural Engineering, vol. 38, no.13, pp158-166, 2022.

[2] Van Klompenburg Thomas, Kassahun Ayalew, Catal Cagatay, "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 177, 105709, 2020.

[3] Yinian Li, Shiwei Du, Min Yao, Yingwu Yi, Jianfeng Yang, Qishuo Ding, Ruiyin He, "Method for wheatear counting and yield predicting based on image of wheatear population in field," Transactions of the Chinese Society of Agricultural Engineering, vol. 34, no.21, pp185-194, 2018.

[4] Maldonado Walter, Barbosa José Carlos, "Automatic green fruit counting in orange trees using digital images," Computers and Electronics in Agriculture, vol. 127, pp572-581, 2016.

[5] Dandan Wang, Huaibo Song, Dongjian He, "Research advance on vision system of apple picking robot," Transactions of the Chinese Society of Agricultural Engineering, vol. 33, no.10, pp59-69, 2017.

[6] Weikuan Jia, Yan Zhang, Jian Lian, Yuanjie Zheng, Dean Zhao, Chengjiang Li, "Apple harvesting robot under information technology: A review," International Journal of Advanced Robotic Systems, vol. 17, no.3, 1729881420925310, 2020.

[7] Wei Ji, Dean Zhao, Fengyi Cheng, Bo Xu, Ying Zhang, Jingjing Wang, "Automatic recognition vision system guided for apple harvesting robot," Computers & Electrical Engineering, vol. 38, no.5, pp1186-1195, 2012.

[8] Ngugi Lawrence C, Abelwahab Moataz, Abo-Zahhad Mohammed, "Recent advances in image processing techniques for automated leaf pest and disease recognition–A review," Information Processing in Agriculture, vol. 8, no.1, pp27-51, 2021.

[9] Dahua Li, Hui Zhao, Xiao Yu, "Overlapping green apple recognition based on improved spectral clustering," Spectroscopy and Spectral Analysis, vol. 39, pp2974-2981, 2019.

[10] Seng Woo Chaw, Mirisaee Seyed Hadi, "A new method for fruits recognition system," 2009 International Conference on Electrical Engineering and Informatics, 2019 IEEE, 5-7 August, 2009, Bangi, Malaysia, pp130-134.

[11] Huaibo Song, Dongjian He, Jingpeng Pan, "Recognition and localization methods of occluded apples based on convex hull theory," Transactions of the Chinese Society of Agricultural Engineering, vol. 28, no.22, pp174-180, 2012.

[12] Huaibo Song, Chuandong Zhang, Jingpeng Pan, Xu Yin, Yibin Zhuang, "Segmentation and reconstruction of overlapped apple images based on convex hull," Transactions of the Chinese Society of Agricultural Engineering, vol. 29, no.3, pp163-168, 2013.

[13] Xiaoyu Huang, Guanglin Li, Chi Ma, Shihang Yang, "Green peach recognition based on improved discriminative regional feature integration algorithm in similar background," Transactions of the Chinese Society of Agricultural Engineering, vol. 34, no.23, pp142-148, 2018.

[14] Wan-Ting Chew, Siew-Chin Chong, Thian-Song Ong, and Lee-Ying Chong, "Facial Expression Recognition Via Enhanced Stress Convolution Neural Network for Stress Detection," IAENG International Journal of Computer Science, vol. 49, no.3, pp818-827, 2022.

[15] Xiaoyuan Dang, Guorui Liu, Xianlun Tang, Shifei Wang, Tianzhu Wang, and Mi Zou, "Motor Imagery EEG Recognition Based on Generative and Discriminative Adversarial Learning Framework and Hybrid Scale Convolutional Neural Network," IAENG International Journal of Applied Mathematics, vol. 52, no.4, pp946-954, 2022.

[16] Md. Abdul Alim Sheikh, Tanmoy Maity, and Alok Kole, "Deep Learning Approach using Patch-based Deep Belief Network for Road Extraction from Remote Sensing Imagery," IAENG International Journal of Applied Mathematics, vol. 52, no.4, pp760-775, 2022.

[17] Mousami Turuk, R Sreemathy, Sadhvika Kadiyala, Sakshi Kotecha, and Vaishnavi Kulkarni, "CNN Based Deep Learning Approach for Automatic Malaria Parasite Detection," IAENG International Journal of Computer Science, vol. 49, no.3, pp745-753, 2022.

[18] Yijing Wu, Yi Yang, Xue-fen, Jian Cui, Xinyun Li, "Fig fruit recognition method based on YOLO v4 deep learning," 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2021 ECTI-CON, 19-22 May, 2021, Chiang Mai, Thailand, pp303-306.

[19] Bargoti Suchet, Underwood James, "Deep fruit detection in orchards," IEEE International Conference on Robotics and Automation, 2017 ICRA, 29 May-3 June, 2017, Singapore, pp3626–3633.

[20] Dean Zha, Rendi Wu, Xiaoyang Liu, Yuyan Zhao, "Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background," Transactions of the Chinese Society of Agricultural Engineering, vol. 35, no.3, pp164-173, 2019.

[21] Weikuan Jia, Yuyu Tian, Rong Luo, Zhonghua Zhang, Jian Lian, Yuanjie Zheng, "Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot," Computers and Electronics in Agriculture, vol. 172, 105380, 2020.

[22] Xu Liu, Chen Steven W, Aditya Shreyas, Sivakumar Nivedha, Dcunha Sandeep, Qu Chao, Taylor Camillo J, Das Jnaneshwar, Kumar Vijay, "Robust fruit counting: Combining deep learning, tracking, and structure from motion," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018 IROS, 1-5 October, Madrid, Spain, pp1045-1052.

[23] Weikuan Jia, Zhifen Wang, Zhonghua Zhang, Xinbo Yang, Sujuan Hou, Yuanjie Zheng, "A fast and efficient green apple object detection model based on Foveabox," Journal of King Saud University-Computer and Information Sciences, vol. 34, no.8, pp5156-5169, 2022.

[24] Bin Yan, Pan Fan, Xiaoyan Lei, Zhijie Liu, Fuzeng Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," Remote Sensing, vol. 13, no.9, 1619, 2021.

[25] Upesh Nepal, Hossein Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," Sensors, vol. 22, no.2, 464, 2022.

[26] Xingkui Zhu, Shuchang Lyu, Xu Wang, Qi Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," 2021 IEEE/CVF International Conference on Computer Vision Workshops, 2021 ICCVW, 11-17 October, 2021, Montreal, BC, Canada, pp2778-2788.

[27] Qisong Song, Shaobo Li, Qiang Bai, Jing Yang, Xingxing Zhang, Zhiang Li, Zhongjing Duan, "Object Detection Method for Grasping Robot Based on Improved YOLOv5," Micromachines, vol. 12, no.11, 1273, 2021.

[28] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, Qinghua Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020 CVPR, 13-19 June, 2020, Seattle, WA, USA, pp11531-11539.

[29] Yuhao Qing, Wengyi Liu. "Hyperspectral image classification based on multi-scale residual network with attention mechanism," Remote Sensing, vol. 13, no.3, 335, 2021.

[30] Huan Zhang, Wenhua Cui, Tianwei Shi, Ye Tao, Jianfeng Zhang, "ATMLP: Attention and Time Series MLP for Fall Detection," IAENG

International Journal of Applied Mathematics, vol. 53, no.1, pp58-65, 2023.
[31] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, "Cbam: Convolutional block attention module," Proceedings of The European Conference on Computer Vision 2018, ECCV 2018, 8-14 September, 2018, Munich, Germany, pp3-19.
[32] Huixuan Fu, Guoqing Song, Yuchao Wang, "Improved YOLOv4 marine target detection combined with CBAM," Symmetry, vol. 13, no.4, 623, 2021.
[33] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, "Focal loss for dense object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no.2, pp318-327, 2020.
[34] Jishnu Mukhoti, Viveka Kulharia, Amartya Sanyal, Stuart Golodetz, Philip Torr, Puneet Dokania, "Calibrating deep neural networks using focal loss," Advances in Neural Information Processing Systems, vol. 33, pp15288-15299, 2020.
[35] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, Jian Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," Advances in Neural Information Processing Systems, vol. 33, pp21002-210112, 2020.