

# YOLO-Banana: An Effective Grading Method for Banana Appearance Quality

Dianhui Mao, Xuesen Wang, Yiming Liu, Denghui Zhang, Jianwei Wu<sup>✉</sup>, Junhua Chen<sup>✉</sup>

**Abstract:** The increasing trend towards independent fruit packaging demands a high appearance quality of individually packed fruits. In this paper, we propose an improved YOLOv5-based model, YOLO-Banana, to effectively grade banana appearance quality based on the number of banana defect points. Due to the minor and dense defects on the surface of bananas, existing detection algorithms have poor detection results and high missing rates. To address this, we propose a density-based spatial clustering of applications with noise (DBSCAN) and K-means fusion clustering method that utilizes refined anchor points to obtain better initial anchor values, thereby enhancing the network's recognition accuracy. Moreover, the optimized progressive aggregated network (PANet) enables better multi-level feature fusion. Additionally, the non-maximum suppression function is replaced with a weighted non-maximum suppression (weighted NMS) function based on distance intersection over union (DIoU). Experimental results show that the model's accuracy is improved by 2.3% compared to the original YOLOv5 network model, thereby effectively grading the banana appearance quality.

**Keywords:** YOLOv5; banana appearance grading; clustering algorithm; weighted non-maximum suppression (weighted NMS); progressive aggregated network (PANet)

## 1 Introduction

In recent years, individually packaged fruits are more popular among consumers than directly sold fruits. As people's standard of living increases, the quality of fruit is becoming more and more demanding, and individually packaged fruit looks more advanced and reduces the possi-

bility of contamination during transportation and sales, making the fruit safe. Taking bananas as an example, GOOD FARMER Company provided a single package of good quality after screening out defective bananas, which improved the quality and storability of bananas and allowed consumers to buy multiple bananas as needed, thus avoiding waste and arousing consumers' desire to buy. However, relying on the manual sorting of intact bananas is time-consuming and labor-intensive, so the use of machines for sorting bananas is of great importance and has a promising market [1].

The realization of banana surface defect detection is the premise of banana sorting by machine. Recently, the use of equipment to detect, identify and screen banana surface defects has become a hot research topic at home and abroad [2]. With the development and application of computer vision, various image processing technologies based on machine learning are used for fruit defect recognition, proving the fea-

---

Manuscript received Jan. 12, 2023; revised Feb. 19, 2023; accepted Feb. 20, 2023. The associate editor coordinating the review of this manuscript was Dr. Dong An. This work was supported by the Beijing Science Foundation (No. 9232005), the Beijing Municipal Philosophy and Social Science Foundation of China (No. 19GLB036), and the Beijing Science and Technology Project (No. Z221100005822014).

Dianhui Mao, Xuesen Wang, Yiming Liu and Denghui Zhang are with School of Computer, Beijing Technology and Business University, Beijing 100048, China.

Jianwei Wu is with Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China.

Jianwei Wu is with Beijing PAIDE Science and Technology Development Co. Ltd., Beijing 100097, China.

Junhua Chen is with Institute of Standardization Theory and Strategy, China National Institute of Standardization, Beijing 100088, China.

✉ Corresponding author. Email: wujw@nrcita.org.cn, chenjunh@cnis.ac.cn

DOI: [10.15918/j.jbit1004-0579.2023.004](https://doi.org/10.15918/j.jbit1004-0579.2023.004)

sibility of machine sorting fruits [3]. However, traditional machine learning is limited by the need to extract features manually through high-precision equipment, resulting in incomplete extraction of essential elements. As a result, developing specific and efficient feature-matching algorithms that can be easily transplanted or generated is difficult. Furthermore, the low accuracy of the model in detecting minor and dense targets poses a challenge to its application in the industry.

The development of convolutional neural networks effectively overcomes the drawbacks above. These networks enable computers to learn from data and extract features from a deep network without relying on an artificial setting. Deep learning-based target detection networks fall into two categories: two-stage detection algorithms are represented by faster-region convolutional neural networks (faster-RCNN) [4], while one-stage detection algorithms are represented by single shot multibox detector (SSD) [5] or you only look once (YOLO) [6-8]. Applying a single-step detection algorithm is more extensive because of the high real-time requirements of the detection tasks in the actual scenario [9-13]. However, compared to existing studies, YOLO still has excellent potential in fruit detection because few studies use it for the fine-grained detection of fruits. Thus, this paper focuses on identifying subtle defects on the surface of bananas using YOLO, aiming to address the low model accuracy.

This paper presents a model, YOLO-Banana, for identifying banana surface defects. The model aims to detect targets of different sizes by utilizing an input image scale of 320 and feature layers of 8, 16, and 32. The density-based spatial clustering of applications with noise (DBSCAN) [14] and K-means [15] fusion clustering is introduced to recalculate the anchor frame to get a better initial anchor value. Meanwhile, an optimized progressive aggregated network (PANet) will bring more feature fusion to improve the network to detect minor defects on

the banana surface. The experiment aims to assess the quality of bananas based on the number of regions with defective spots on the fruit's surface. Consequently, identifying these bad spots poses a significant challenge. To mitigate the impact of missing defective region prediction frames on quality assessment, the model uses the distance intersection over union (DIoU)-based weighted non-maximum suppression (NMS) function [16] in place of the standard NMS function.

In summary, this model provides a robust solution for detecting banana surface defects and assessing fruit quality, potentially improving quality control processes in the agriculture industry.

## 2 Materials and Methods

### 2.1 Image Preprocessing

In this study, an image acquisition system constructed a dataset of banana defects, as shown in Fig. 1. The object used in the experiment was an Ecuadorian banana. The image acquisition system consists of a camera facing the banana at about 50 cm to simulate the actual scene of a single banana on an intelligent machine sorting conveyor belt in a factory. The camera will record the natural decay process of a single banana to capture the evolution of the banana itself from defective point to decay.

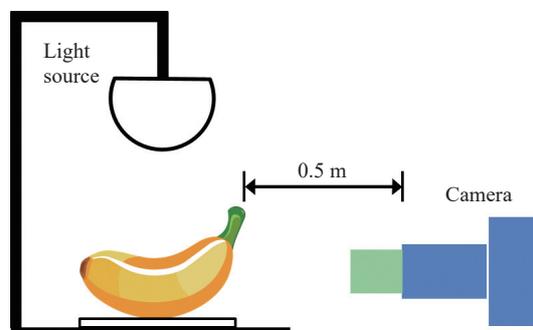


Fig. 1 Image acquisition system

### 2.2 Image Annotation and Dataset Production

In this study, the captured video was processed frame by frame, and 250 images were extracted

throughout the banana decay process according to a fixed number of frames. These images were then expanded to 2 000 using data enhancement methods, resulting in the experiment’s dataset. The image data was reduced to increase the network’s processing speed. The image size was adjusted to  $480 \times 480$  in a 1:1 ratio and  $640 \times 480$  in a 4:3 ratio while maintaining the original image’s aspect ratio.

In this study, Labellmg was used to label the banana defect images and generate XML-type annotations. Each object was centered within the label box during the labeling process. The dataset was divided into a training set of 1 400 images, a validation set of 400 images, and a test set of 200 images, in a ratio of 7:2:1. To test the robustness of the model, 80 additional images of a single banana were crawled from the network

and added to the test set.

The visualization results in Fig. 2 provide an overview of the banana defect dataset used in the study. Specifically, Fig. 2 shows the analysis and visualization results of the data set. The banana defect data set in Fig. 2(a) includes 1 188 bananas, 12 304 defect points, and 3 068 rotten parts. In Fig. 2(b),  $X$  and  $Y$  refer to the position of the center point. The darker the color is, the more concentrated the center point of the target box is. Fig. 2(b) width and height, respectively, represent the width and height of the object in the picture. From Fig. 2(b) and Fig. 2(c), the distribution of things in the data set is not very uniform. The proportion of small objects is dense, the proportion of medium objects is relatively small, and it conforms with actual scenarios.

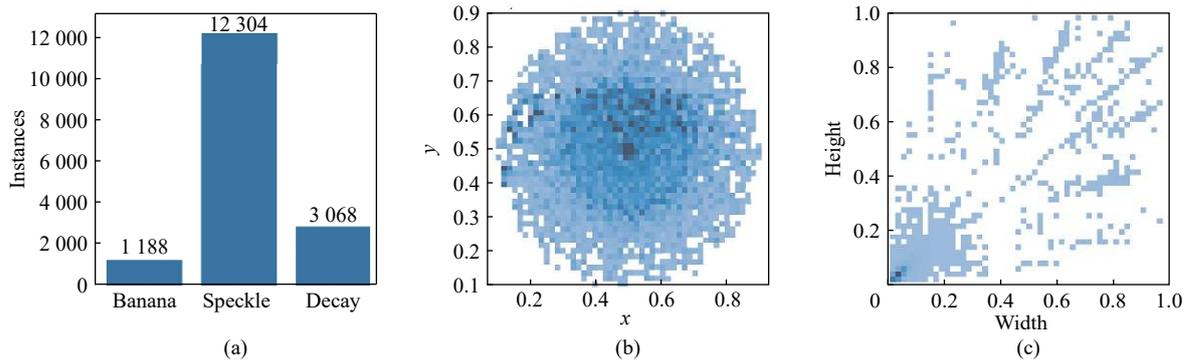


Fig. 2 Data analysis: (a) category distribution; (b) object center position distribution; (c) size distribution

### 3 Improved YOLOv5 Algorithm

The detection of banana defect points requires more sophisticated recognition and positioning capabilities. Therefore, we propose an improved YOLO-Banana model to improve the target detection accuracy by introducing the refined, improved anchor, PANet and NMS.

#### 3.1 Refined Anchor

An appropriate anchor frame generation algorithm can be used to obtain better initial anchor values to achieve more accurate positioning of banana defect points. DBSCAN is a density-based clustering algorithm that determines the proximity of sample distribution to identify

whether samples of the same category are closely connected. In contrast, the K-means clustering algorithm uses distance as the evaluation index of similarity. The smaller the distance between two objects, the higher their similarity. Therefore, using the combination of DBSCAN and K-means clustering algorithm to re-cluster the box of the training dataset will achieve better results.

Therefore, we use the DBSCAN algorithm to preprocess the frames in the data set to eliminate the isolated points away from the density center and then use the K-means algorithm to analyze the preprocessed data to re-determine  $N$  cluster centers. Because the K-means clustering algorithm will not produce errors due to individ-

ual outliers when calculating the cluster center point after removing outliers, thus improving the accuracy of the results. In the experiment, we use a genetic algorithm to carry out 1000 mutations on  $n$  cluster centers, with a mutation rate of 9:1, and take the value of anchor\_fitness as the evaluation standard. The anchor after mutation value is the anchor ( $kg$ ), and the before is the anchor ( $k$ ). Suppose  $\text{anchor}(kg) > \text{anchor}(k)$ ; the anchor( $kg$ ) is used; otherwise, anchor( $k$ ) is used.

In the clustering process, the distance formula of the K-means clustering algorithm adopts the method of 1-IoU to calculate the distance. The specific distance formula is defined as follows

$$\begin{cases} D(B, C) = 1 - \text{IoU}(B, C) \\ \text{IoU} = \frac{B \cap A}{B \cup A} \end{cases} \quad (1)$$

In Eq. (1) above,  $A$  represents the size of the cluster box,  $B$  represents the sample box size,  $C$  represents the cluster center generated by the K-means clustering algorithm, and IoU represents the intersection ratio of the sample box size and the cluster box size. If the IoU between sample box  $B$  and cluster center  $C$  is more giant, the distance 1-IoU will be smaller, and the distance will be closer.

### 3.2 Improved PANet

As illustrated in Fig. 3 (a), conventional object detectors utilize the backbone network to extract and predict features based on the feature pyramid network (FPN) hierarchy [17]. YOLOv3 also

follows this paradigm but replaces the original FPN with the progressive aggregated network (PANet) [18], as depicted in Fig. 3 (b). The PANet incorporates a bottom-up path aggregation network based on the FPN to fuse deep and shallow feature maps at different levels and enhance target detection accuracy. However, in practice, the contribution of each node to the output features is not equal due to the other input resolutions. In particular, a node with only one input edge and no feature fusion has a relatively minor contribution to the feature network.

The paper proposes an optimized PANet, as shown in Fig. 3 (c), to address this feature-level imbalance issue. The optimized PANet weakens the node weight of feature fusion with only one input by emphasizing the importance of intermediate bidirectional output nodes. Specifically, edges are added from the original input to output nodes, and horizontal connections are introduced between feature layers at the same scale to increase the feature function weight of the coating, thereby achieving higher-level feature fusion. Fig. 3 illustrates the structure of the improved YOLOv5 network model, while Fig. 4 provides an overview of the overall structure of the proposed model.

### 3.3 Weighted NMS Based on DIoU

The YOLOv5 [19] model is a variant of the YOLO algorithm, known for its fast detection speed and capability to return the position coordinates and category of the target box in the output layer. The YOLO algorithm divides the

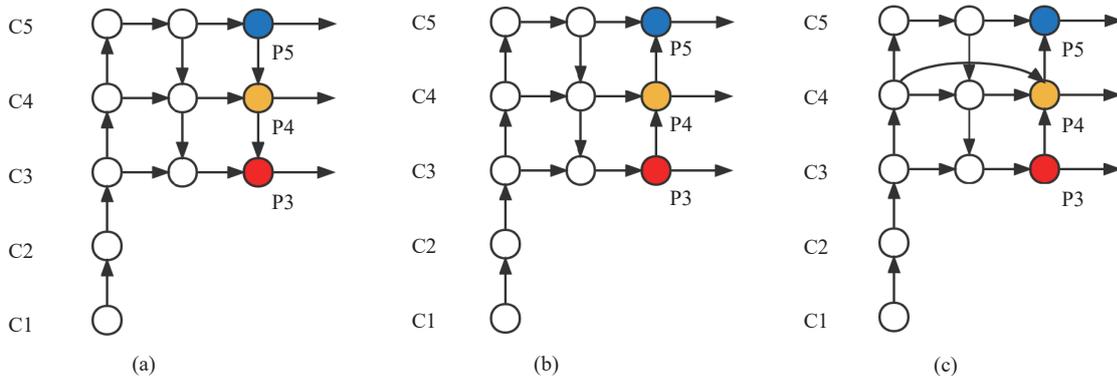


Fig. 3 Fusion diagram: (a) FPN; (b) PANet; (c) Craded PANet

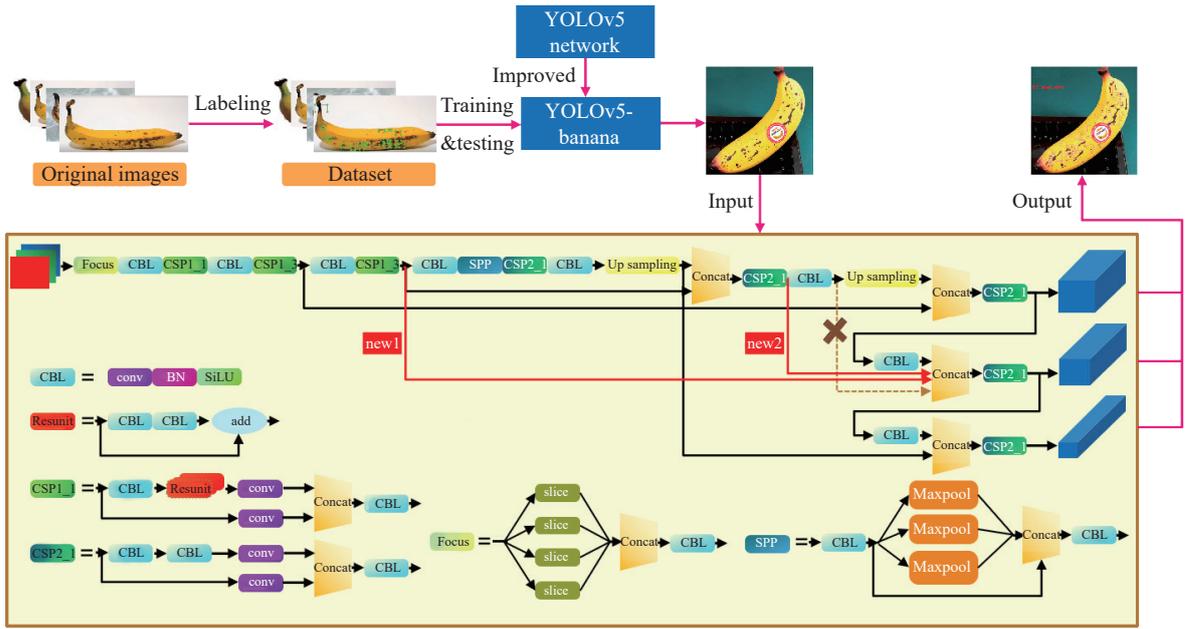


Fig. 4 Improved structure of the YOLOv5 network

input image into a  $7 \times 7$  grid, containing the target center responsible for predicting that target. Each grid predicts two target frames, returning the position coordinates of that target frame and the prediction confidence value. A confidence

threshold is set to filter out the target boxes with low confidence. The retained packages are then subjected to NMS (non-maximum suppression) to obtain the final prediction results, as illustrated in Fig. 5.

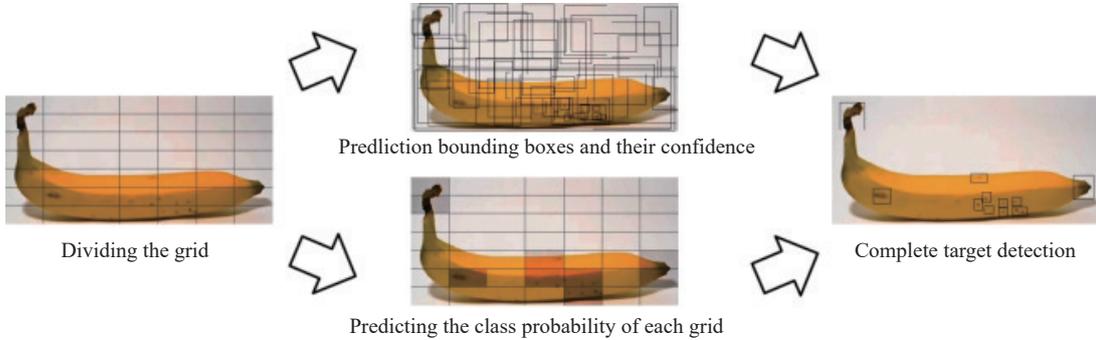


Fig. 5 YOLO object detection model

However, traditional NMS suffers from several problems, depicted in Fig. 6. In this figure, green represents the actual box, red represents the prediction box and black represents the overlapping part. In Fig. 7, there are three overlapping modes with the same IoU value, which highlights that generalized intersection over union (GIoU) [20] cannot accurately reflect the overlapping degree between the prediction box and the actual box.

This paper uses the DIoU [21] bounding box regression loss to address this limitation. As illus-

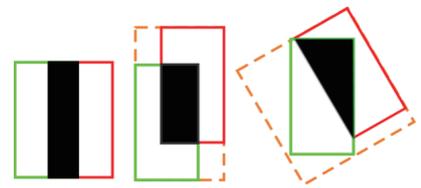


Fig. 6 Three cases with the same IoU value

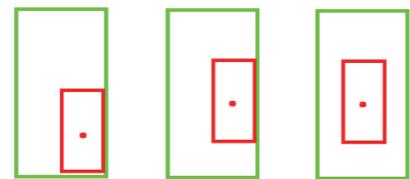


Fig. 7 Cases: GIoU loss is the same as IoU loss

trated in Fig. 8, DIoU considers the overlap area and the center point distance between the two boxes. The NMS based on DIoU can be defined using following equations for the prediction box  $N$  with a high score.

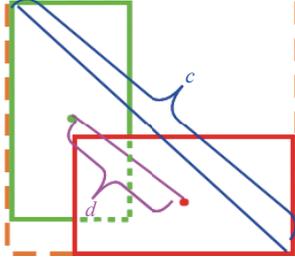


Fig. 8 DIoU calculation diagram

$$k_i = \begin{cases} k_i, \text{IoU} - R_{\text{IoU}}(N, B_i) < \varepsilon \\ 0, \text{IoU} - R_{\text{DIoU}}(N, B) > \varepsilon \end{cases} \quad (2)$$

$$\text{IoU} = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (3)$$

$$R_{\text{DIoU}} = \frac{\rho^2(b, b^{gt})}{c^2} \quad (4)$$

$$N = \frac{\sum_i \omega_i B_i}{\sum_i \omega_i} \quad (5)$$

$$B_i \in \{B | \text{DIoU}(N, B) \geq \text{thresh}\} \cup (N) \quad (6)$$

$$\omega_i = S_{\text{DIoU}}(N, B_i) \quad (7)$$

Intersection over union (IoU) is a widely used evaluation metric in object detection that measures the overlap between the predicted bounding box and the ground truth bounding box.  $\rho^2(b, b^{gt})$  is the Euclidean distance between the center point of the predicted bounding box and the center point of the ground truth bounding box.  $c^2$ , the diagonal length of the smallest closed rectangular box formed between the predicted bounding box and the ground truth bounding box. The threshold  $\varepsilon$  is a manually set parameter used in NMS to filter out the predicted boxes with low confidence scores.  $B$  is the anchor value corresponding to each box, and  $K$  is the classification score for different object categories.

However, traditional NMS only selects the

highest score box in each iteration and may miss some objects in dense scenes, especially for prediction boxes with more than two central points located on different objects. This paper introduces weighted NMS based on the DIoU loss function to address this issue. The DIoU considers the overlap area and the center point distance between two boxes. Weighted NMS does not directly eliminate the boxes with the same category but rather weights and sums all prediction boxes with confidence scores higher than the threshold to obtain a new coordinate box. Eq. (7) represents the weighted weight and the product of the score and DIoU. By replacing traditional NMS with weighted NMS, the YOLOv5 model achieves improved performance in dense scenes with multiple objects.

## 4 Experiment Analysis

### 4.1 Evaluation Metrics

The experiments in this study were conducted on a CentOS operating system, using an Intel(R) Xeon(R) CPU E5-2678 v3 with a clock speed of 2.50 GHz, an NVIDIA TITAN XP GPU, and 12 GB of memory. The software stack used for the experiments included CUDA version 10.2, cuDNN version 7.6.5, and PyTorch as the deep learning framework. The dataset was converted into the VOC format and was divided into training and test sets in an 8:2 ratio. The initial training was performed for 200 epochs, using a batch size of 12 and an input image size of  $320 \times 320$ .

The performance evaluation of the model was based on precision, recall, AP (average precision), mAP (mean average precision), and detection speed. In this study, a true positive case was defined as an IoU (Intersection over Union) greater than or equal to 0.5, while a false positive case was defined as an IoU of less than 0.5. The calculations for precision, recall, and mAP is shown as

$$\text{Precision} = P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recal} = R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (10)$$

$$\text{mAP} = \frac{\text{AP}_{\text{banana}} + \text{AP}_{\text{speckle}} + \text{AP}_{\text{decay}}}{3} \quad (11)$$

where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. The AP in Eq. (10) represents the area under the precision-recall curve. For multi-category detection, the mAP is often used as a performance metric, calculated as the mean of the APs for each category, as shown in Eq. (11).

## 4.2 Results and Analysis

### 4.2.1 Comparison of Different Anchors

The DBSCAN algorithm is used as a preprocessing step to eliminate isolated points far from the density center and select nine clustering centers after re-clustering. Tab. 1 compares anchor difference and detection results with the banana defect dataset, demonstrating that the improved anchor leads to better results.

Tab. 1 Experiment results 1

| Method   | Recall | Precision |
|----------|--------|-----------|
| Original | 81.8%  | 87.3%     |
| Refined  | 82.1%  | 88.5%     |

### 4.2.2 Ablation Experiments

The performance changes caused by the change

in network structure are gradually verified through the comparison of ablation experiments. Model changes can be understood through the name of the experiment.

The experimental process and experimental test results are shown in Tab. 2. Based on the experimental results, it can be observed that the performance of model B is not optimal when compared to model A. The reason behind this could be attributed to dense spots in the banana region during the recession, leading to the generation of a large number of prediction boxes containing another. This situation causes overlap between prediction boxes, making it challenging for the model to detect and localize the defect regions accurately. While the weighted NMS can potentially address this issue, the limitations of GIoU hinder its effectiveness in these experiments. The introduction of an optimized PANet in experiment C yields significant improvements in various performance metrics, albeit with a slight increase in model parameters and a decrease in processing speed. In experiment D, the introduction of DIoU addresses the limitations of GIoU and improves the issue of overlapping prediction boxes to some extent. While the improvement of experimental indicators is not substantial, experiment D performs better than experiment C in actual detection scenarios.

Tab. 2 Experiment results 2

| Label | Refinement                  | mAP@0.5       | mAP@0.5:0.95 | Recall        | Precision     | Size (MB) | Speed (f·s <sup>-1</sup> ) |
|-------|-----------------------------|---------------|--------------|---------------|---------------|-----------|----------------------------|
| A     | YOLOv5s                     | 88.1%         | 60.7%        | 82.1%         | 88.5%         | 14.3      | 55.56                      |
| B     | A+Weighted NMS              | 87.4%         | 62.4%        | 82.3%         | 85.8%         | 14.3      | 55.87                      |
| C     | B+Optimized PANet           | 90.2%         | 63%          | 85.8%         | 88.7%         | 15.4      | 52.08                      |
| D     | C+DIoU <sub>(YOLO-Ba)</sub> | 90.4% (+2.3%) | 63.7% (+3%)  | 85.0% (+2.9%) | 89.9% (+1.4%) | 15.4      | 51.54                      |

Fig. 9 provides a visual comparison of the improved and baseline models' detection results for an NMS threshold of 0.6. In Fig. 9(a) and Fig. 9(b), the scenario involves a high density of banana defect points. The detected defect regions in Fig. 9(b) have a high degree of overlap, with severe occlusions and omissions and an omission

of the banana logo. In contrast, the detection of the defect regions in Fig. 9(a) is more precise and does not count the banana logo as a defect region. In Fig. 9(c) and Fig. 9(d), the most oversized prediction box in Fig. 9(d) already contains the other two smaller detection boxes. In counting defective areas, Fig. 9(c) is more accu-

rate. Since banana decay tends to progress from local to global levels, the “decay” part of the dataset is labeled with medium-sized boxes to distinguish the degree of decay of the banana by the number of medium-sized boxes. In Fig. 9(e) and Fig. 9(f), the area of overlapping boxes in Fig. 9(e) is better than in Fig. 9(f). Finally, in comparing Fig. 9(g) and Fig. 9(h), it can be seen that the improved model is more sensitive to detecting defective spots on the surface of the banana.

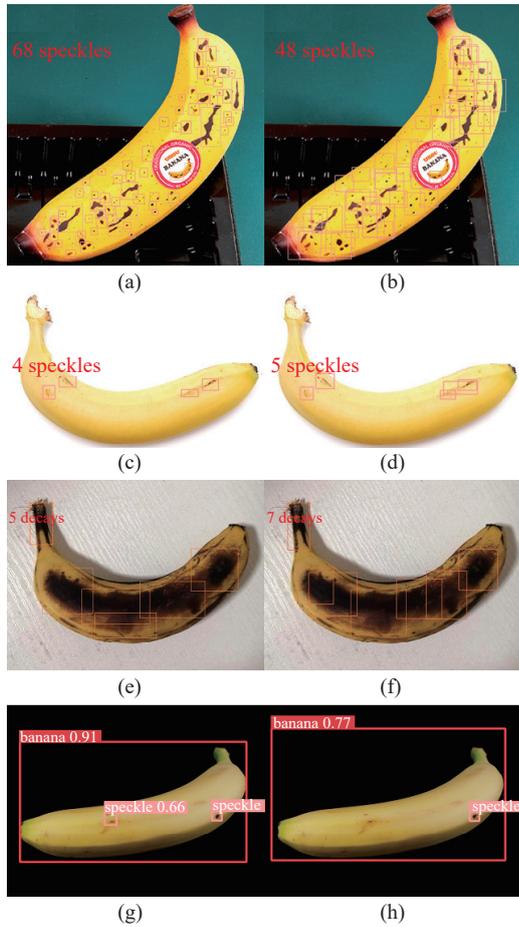


Fig. 9 YOLO-Banana (left) and baseline (right) detection effect

Fig. 10 compares the detection performance between YOLO-Banana and baseline models on banana clusters. Notably, Fig. 10(b) and Fig. 10(d) show the detection results obtained by YOLO-Banana. Although our training dataset only has a single banana, we still want to explore the detection effect of the model on the whole banana. The figure shows that the improved model can detect more defect spots on the banana surface. Furthermore, the improved model performs better detecting small fuzzy objects with extremely dense distribution than the baseline model. However, there are still some tiny objects that remain undetected. In Fig. 10(c) and Fig. 10(d), the model erroneously identifies the black background between banana stems as defect areas or rotten areas and the black area of banana roots as defect areas. The main reason for such misidentification could be that the defects formed during banana decay are typically dark in color. Additionally, our preprocessing steps involve retaining only a single banana body and do not label the banana stem and root, which results in weak robustness for complex scene detection.

To conclude, the proposed YOLO-Banana model significantly improves the detection of defective areas in banana fruit. Adopting a weighted NMS based on the DIoU loss function enhances the localization accuracy of the model. Moreover, optimizing the PANet promotes the model to learn the defective regions more effectively, achieving higher-level feature fusion. Consequently, the YOLO-Banana model outperforms the baseline model and achieves better

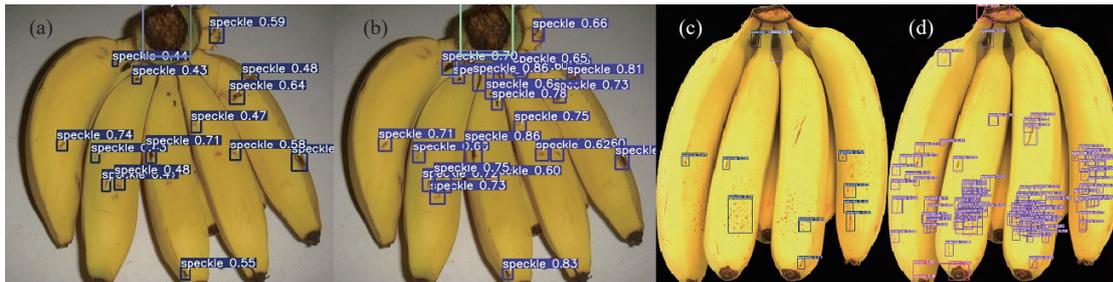


Fig. 10 Comparison diagram of banana cluster detection

detection results, making it a promising solution for automated banana fruit defect detection.

**4.2.3 Evaluation and Analysis of Banana Appearance Quality**

Packaging fruits separately has several advantages, such as increased safety and improved quality. The YOLO-Banana model enables the counting of defective areas on the surface of individual bananas, which helps in distinguishing the appearance of bananas based on the number of “speckles” and “decay”. The accuracy of the model has been established through a series of tests. The judgment criteria for banana appearance quality are presented in Tab. 3. Fig. 11 compares the judgment effect of banana appear-

ance quality. The image on the left has more than ten defective spots and areas of decay and is therefore judged to be of poor appearance. The image on the right only detects two defects, so it is considered a better-looking banana. In fact, for the independent packaging of single bananas, only the ones with good appearance quality will be selected, and neither the not good nor the bad will be considered.

**Tab. 3 Appearance quality judgment standard**

| Numbers of “speckles”                     | Numbers of “decay”                   | Appearance quality |
|---|--------------------------------------|--------------------|
| $0 \leq \text{areas of speckles} \leq 10$ | None                                 | Good               |
| $10 < \text{areas of speckles} \leq 25$   | $0 \leq \text{reas of decay} \leq 2$ | Not good           |
| $25 < \text{areas of speckles}$           | $2 < \text{areas of decay}$          | Bad                |



Fig. 11 Judgment effect of banana appearance quality

**5 Conclusion**

In this study, a model called YOLO-Banana is proposed for identifying surface defect areas on bananas to determine the grading of their appearance quality based on the number of defects. The model achieves high accuracy in identifying defects on the surface of bananas. It includes several improvements, such as recalculating the size of the anchor boxes and using an optimized PANet for multi-level feature fusion, which enhances the model’s ability to identify densely speckled and decaying areas. Additionally, a weighted non-maximum suppression (NMS) function based on DIoU reduces the missed detection rate due to dense defects.

The increasing trend of individual fruit pack-

aging has made the classification of single banana appearance quality more critical for producers. The proposed model can classify the quality of a single banana’s appearance through defective areas, ensuring that only good-quality bananas are packed individually. However, the limitation of the dataset only allows the model to identify black defect areas on the banana surface. Future research can explore how to classify the appearance quality grade of bananas more precisely according to the area of detected defect areas and according to national standards.

**References:**

[1] R. Zhang, X. Li, L. Zhu, M. Zhong, and Y. Gao, “Target detection of banana string and fruit stalk based on YOLOv3 deep learning network, ” in

- 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, pp. 346-349, 2021.
- [2] L. Fu, J. Duan, X. Zou, J. Lin, L. Zhao, J. Li, and Z. Yang, "Fast and accurate detection of banana fruits in complex background orchards," *IEEE Access*, vol. 8, pp. 196835-196846, 2020.
- [3] D. Mao, H. Sun, B. Li, X. Yu, J. Wu, and Q. Zhang, "Real-time fruit detection using deep neural networks on CPU (RTFD): An edge AI application," *Computers and Electronics in Agriculture*, vol. 204, pp. 107517, 2023.
- [4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, pp. 21-37, 2016.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779-788, 2016.
- [7] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6517-6525, 2017.
- [8] A. Bochkovskiy, C. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint, arXiv: 2004.10934, 2020, doi: 10.48550/arXiv.2004.10934.
- [9] Y. Xue, N. Huang, S. Tu, L. Mao, A. Yang, X. Zhu, X. Yang, and P. Chen, "Immature mango detection based on improved YOLOv2," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 7, pp. 173-179, 2018.
- [10] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417-426, 2019.
- [11] L. Fu, Y. Feng, J. Wu, Z. Liu, F. Gao, Y. Majeed, A. Al-Mallahi, Q. Zhang, R. Li, and Y. Cui, "Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model," *Precision Agriculture*, vol. 22, pp. 754-776, 2021.
- [12] D. Wu, S. Lv, J. Mei, and H. Song, "Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments," *Computers and Electronics in Agriculture*, vol. 178, pp. 105742, 2020.
- [13] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sensing*, vol. 13, no. 9, pp. 1619, 2021.
- [14] M. Z. Fauzi and A. Abdullah, "Clustering of public opinion on natural disasters in indonesia using DBSCAN and K-Medoids algorithms," *Journal of Physics: Conference Series*, vol. 1783, no. 1, pp. 012016, 2021.
- [15] A. Vattani, "K-means requires exponentially many iterations even in the plane," in *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry (SCG '09)*, Association for Computing Machinery, New York, NY, USA, pp. 324-332, 2009.
- [16] S. Roman, W. Wang, and T. Gabruseva, "Weighted boxes fusion: Ensembling boxes for object detection models," *Image and Vision Computing*, vol. 107, pp. 104117, 2021.
- [17] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936-944, 2017.
- [18] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8759-8768, 2018.
- [19] M. L. Mekhalfi, C. Nicolò, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif, and E. A. Maghayreh, "Contrasting YOLOv5, transformer, and efficientdet detectors for crop circle detection in desert," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, no. 3003205, pp. 1-5, 2022.
- [20] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 658-666, 2019.

- [21] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IOU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993-13000, 2020.



**Dianhui Mao** was born in 1979, Ph. D., professor. He received the degree from the Huazhong University of Science and Technology. He is currently a Professor at the School of Computer, Beijing Technology and Business University. He has a wealth of experience in the application studies of block-chain and AI. He has received the Leading Talent of Enterprise Innovation and Entrepreneurship Award in Jiangsu, China. He is also a member of the Think Tank Committee, All-China Federation of Industry and Commerce (ACFIC).



**Xuesen Wang** graduated with a Bachelor of Engineering degree from the School of Computer and Information Engineering, Beijing Technology and Business University, in 2019. He is currently a graduate student studying Computer Technology at the same university. His primary research interests include object detection and entity resolution.



**Yiming Liu** graduated from the School of Information Engineering of Hebei University of Geosciences with a bachelor's degree in engineering. Currently, she is a graduate student in the School of Computer Science and Technology at Beijing Technology and Business University. Her

main interests include natural language processing, information extraction, and knowledge graph construction.



**Denghui Zhang** received a B.S. degree in software engineering from the Jiangsu Ocean University of Computer and Engineering, in 2018. He is currently a graduate research candidate with the Computer Technology of Beijing Technology and Business University. His main interests include image processing, target detection, and fine-tuning of the large language model.



**Jianwei Wu** was born in Xianghe, Hebei in 1980, and is a professor-level senior engineer. He serves as the Chairman and General Manager of Beijing PAIDE Science and Technology Development Co., Ltd., and the Director of the Key Laboratory of Agricultural Internet of Things System Integration under the Ministry of Agriculture and Rural Affairs. He has been selected for talent programs such as Beijing's Outstanding Youth, Science and Technology Nova, and Excellent Young Engineer. His main research interests are smart orchards and digital villages.



**Junhua Chen** is affiliated with the Institute of Standardization Theory and Strategy at the China National Institute of Standardization, and holds a Doctorate in Law from Wuhan University. He has long been involved in the formulation and research of standards across various industries, with his main focus on standardization strategy, standardization law, and intellectual property.