Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 6, No. 1, 2023

Fruit Freshness Detection Based on YOLOv8 and SE attention Mechanism

Zihe Wei*, **Miaosen Chang, Yipeng Zhong**

University of Science and Technology Liaoning, Anshan 114051, China
* Corresponding author

Abstract: Fruit is a crucial component of daily diets, emphasizing the importance of ensuring its freshness. Our proposed method utilizes YOLOv8 and SE attention mechanism for detecting the freshness of fruits. Our method balances accuracy and real-time performance while detecting the freshness of different types of fruits. The proposed method achieves an average accuracy of 87.8% and a maximum accuracy of 95.0% for detecting the freshness level of a single fruit category. Experimental results confirm the effectiveness of our proposed method in accurately detecting and localizing the freshness level of fruits, demonstrating its potential for wide application and promotion in the fruit industry.

Keywords: Fruit freshness detection, YOLOv8, SE attention mechanism.

1. Introduction

Fruit is an indispensable part of People's Daily diet, but in the process of storage and transportation of fruit, due to the influence of various factors, such as temperature, humidity, oxygen and carbon dioxide, fruit freshness is often affected.

Many researchers have come up with a variety of methods to detect fruit freshness, such as those[1][2]based on multiple sensors such as infrared imaging, smell, sound and vision. However, these traditional methods often have some problems, such as high cost, poor accuracy, not easy to popularize.

The fruit freshness detection method based on computer vision and deep learning technology has many advantages, such as high accuracy, low cost, easy implementation and popularization. Among them, the fruit freshness detection method using YOLOv8 and SE attention mechanism can realize real-time detection and positioning of multiple fruit categories, with high accuracy and practicability.

2. Introduction of YOLOv8

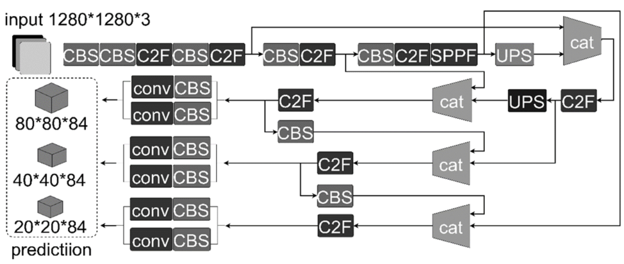

Figure 1. The network structure of YOLOv8

YOLOv8 is a cutting-edge model that builds upon the design principles of YOLOv5 and YOLOv7 ELAN, as illustrated in Figure 1. This model retains the basic framework of YOLOv5 while introducing new features and improvements to enhance performance and flexibility. YOLOv8 includes a new backbone network architecture, an Anchor-Free detection head, and a new loss function. It offers models of different sizes ranging from N/S/M/L/X scales, which are adjusted based on scaling coefficients.

The backbone network and Neck sections are based on the design philosophy of YOLOv7 ELAN, with adjustments to improve model performance.

YOLOv8 has undergone significant changes in the Head section, using the decoupled head structure to separate the classification and detection heads, and changing the detection head from Anchor-Based to Anchor-Free. The Loss calculation uses the TaskAlignedAssigner positive sample allocation strategy and the Distribution Focal Loss.

In summary, YOLOv8 is an advanced model that builds upon the success of previous YOLO models while introducing new features and improvements. It offers different size models and incorporates various design changes to improve performance and flexibility.

3. Introduction of SE Attention Mechanism

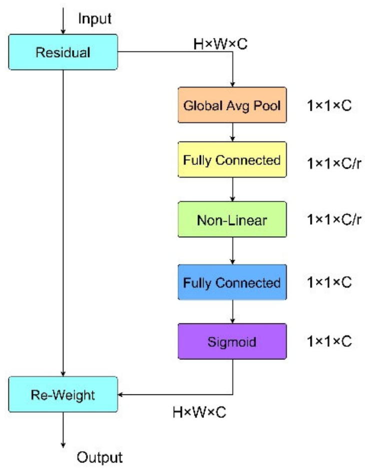

Figure 2. Structure of SE attention

As shown in Figure 2, The SE (squeeze-and-excitation)[3]attention mechanism is a channel-wise attention mechanism that improves the interdependencies between channels in a feature map.

It takes the feature map as input and performs global average pooling to reduce the spatial features to 1 x 1. Then, it establishes channel connections through two fully connected layers with a nonlinear activation function.

The sigmoid activation function is used to obtain normalized weights, which are multiplied with the original feature map in a channel-wise manner to complete the channel attention process.

The first fully connected layer reduces the dimensionality of the feature values, which reduces computation and parameters, and the second fully connected layer restores the original channel number, establishing correlations between channels.

4. Experiment and Analysis

The dataset used in our study is named "fruit_bad_dataset" and contains a total of 401 images. The dataset consists of four classes: "apple" with 95 images, "bad apple" with 69 images, "banana" with 94 images, and "bad banana" with 60 images. Additionally, 25 images depict a mix of apples and bananas, and 58 images depict a mix of bad apples and bad bananas. Examples of each category are shown in Figure 3.

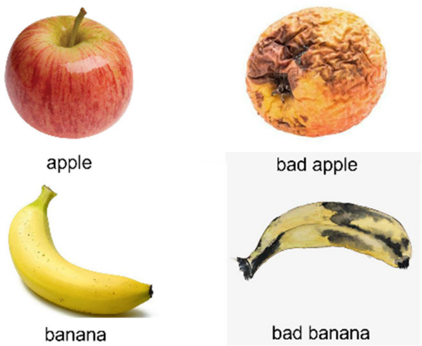

Figure 3. Example graphs for each category

The images in the dataset were selected to represent common fruits that are prone to decay, spoilage, and damage. The "bad" classes were included to simulate real-world scenarios where fruits may be unfit for consumption due to various reasons such as insect infestation, physical damage, or microbial contamination. The mixed classes were added to further challenge the classification task by introducing combinations of different fruit types and their associated defects.

In our study, we utilized the YOLOv8 model to train on the fruit_bad_dataset. We employed a stochastic gradient descent algorithm with a batch size of 8 and used SGD as the optimizer. The learning rate was decayed to 0.01 after 100 iterations, and the training process continued for a total of 1 hour on an RTX3060 device.

5. Experimental Application and Results

As shown in Figure 4,The confusion matrix shows how the classifier behaves on the test set.

In the apple category, 95% of samples were correctly classified and the misjudgment rate was 5%. In the bad banana category, 94 percent of the sample was correctly classified, with a misjudgment rate of 6 percent. In the banana category, 85% of the samples were correctly classified and the misjudgment rate was 15%. In the bad apple category, 77 percent of the samples were correctly classified, with a misjudgment rate of 23 percent.

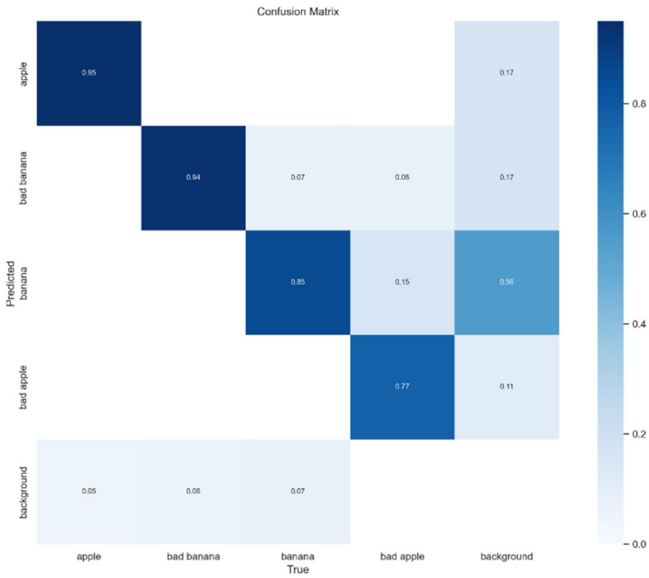

Figure 4. Confusion matrix result

From Figure 5, we can observe that the trained model is able to predict and mark all possible targets in the pictures, instead of just distinguishing one category for each picture. Remarkably, the model maintains a good classification effect even when the targets are partially or fully blocked.

It's worth noting that the model's performance is impressive given that each image is processed in approximately 20 milliseconds, which demonstrates its efficiency and speed in real-time applications.

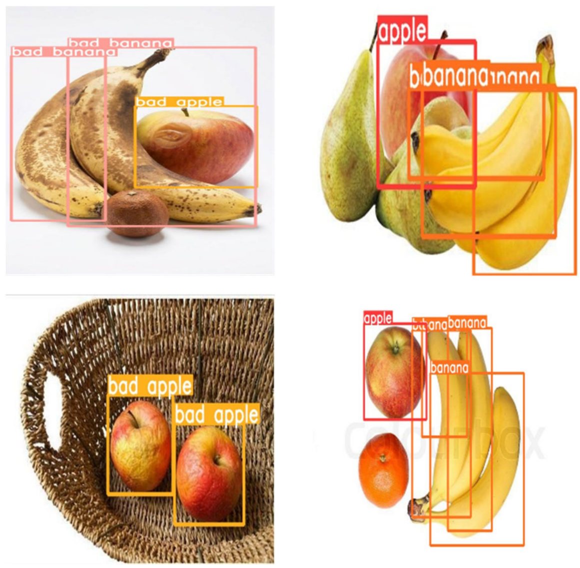

Figure 5. Classification prediction result

6. Conclusion

In conclusion, the trained model demonstrated impressive performance in predicting and marking all possible targets in pictures, while maintaining good classification accuracy even when targets were partially or fully blocked. The model's efficiency and speed, with each image being processed in approximately 20 milliseconds, make it suitable for real-time applications. Overall, the model's capabilities and speed make it a promising tool for a variety of image recognition tasks.

References

[1] Chen Ling, Cheng Guodong, Xu Jianhang, Hu Zhiming, Jia Youdong. Fruit freshness detection based on hyperspectral technology feasibility study [J]. Journal of electronic test, 2021 (24): 41-43. DOI: 10.16520 / j.carol carroll nki. 1000-8519.2021.24.014.

[2] LIU Y G. Research on rapid recognition technology of apple freshness feature based on gas sensor array optimization [D]. North central university, 2021. DOI: 10.27470 /, dc nki. GHBGC. 2021.000587.

[3] Xiao Pengcheng, Xu Wenguang, Zhang Yan, Zhu Liguang, Zhu Rong, Xu Yunfeng. Scrap classification rating method based on SE attention mechanism [J]. Journal of engineering science, 2023, (8) : 1342-1352. The DOI: 10.13374 / j.i ssn2095-9389.2022.06.10.002.

[4] Yu Ying, Yang Tingting, Yang Boxiong. Performance evaluation and Python implementation of confusion matrix Classification [J]. Modern Computers,2021(20):70-73+79.