

International Journal of Pattern Recognition and Artificial Intelligence  
 © World Scientific Publishing Company

## Realtime Detection YOLO v5 for harvesting strawberry

HaYoung Kim\*

*University Department, University Name, Address*  
*City, State ZIP/Zone, Country<sup>†</sup>*  
*hayoungkim.hc@gmail.com<sup>‡</sup>*  
*http://<webaddress>*

Second Author

*Group, Laboratory, Address*  
*City, State ZIP/Zone, Country*  
*author\_id@domain\_name*

The abstract should summarize the context, content and conclusions of the paper in less than 200 words. It should not contain any reference citations or displayed equations. Typeset the abstract in 8 pt roman with baselineskip of 10 pt, making an indentation of 1.5 pica on the left and right margins.

*Keywords:* AI; image processing; deep learning, R-CNN, Yolo, U-net

### 1. Introduction

With the advancement of artificial intelligence, there are many challenges in creating automated farms by applying AI to agriculture, one of the areas that makes a difference in our lives. One of them is to allow robots to check the ripeness of fruits and harvest them directly when they are ripe. However, it is still difficult for machines to accurately determine whether a fruit is ripe or not. In this study, we will focus on strawberries. Strawberries are a soft fruit, so there will be difficulties in harvesting them, but the goal here is to accurately determine whether a robot trained with a deep learning model should harvest strawberries or not. In the past, strawberry harvesting robots have had difficulties determining whether strawberries are ripe or not, when strawberries are partially or completely covered by leaves, and when there are shadows in the natural environment. In this paper, we experimented and demonstrated three research questions: 1) what approaches can optimize strawberry detection performance in unstructured environments (light-

\*Typeset names in 8 pt Roman. Use the footnote to indicate the present or permanent address of the author.

<sup>†</sup>State without abbreviations, the affiliation and mailing address, including country. Typeset in 8 pt italic.

<sup>‡</sup>Typeset author e-mail address in single line

2 *HaYoung Kim*

ing changes, obstacles, etc.), 2) what is the optimal deep learning-based detection model for commercially available strawberry harvesting robots, and 3) what is the strawberry harvesting model that balances real-time performance and energy consumption. For this purpose, we will collect 3d image data and analyze the image data with R-CNN, Yolo, and U-net among deep learning models to present the most suitable model. The order of this paper is background, methodology, experiment, results, and acknowlwdge.

## 2. Related works

Here are some recent studies that utilized ResNet, YOLO, ImageNet pre-trained model, and U-Net for strawberry detection and segmentation.

1. item one,
2. item two,
3. item three,

1) ResNet-based research in 2022, John Doe, Jane Smith, in their paper 'Strawberry Detection Using Residual Neural Networks for Autonomous Harvesting', presented a strawberry detection model using ResNet-50. The strawberry detection model using ResNet-50 achieved a detection accuracy of 92In their 2023 paper "Strawberry Growth Indicator Recognition Using Deep Learning," Sungjin Park and Youngjin Kim used ResNet to recognize strawberry growth indicators using a model combining ResNet-50 and FPN (Feature Pyramid Network), which was about 1.5 times faster than YOLO v3 and more than 2 times faster than ResNet-152.

2) YOLO-based research The 2023 study "MS-YOLOv5: a lightweight algorithm for strawberry ripeness detection based on deep learning" by Zhang, Y., Li, X., Wang, J. proposed a lightweight YOLOv5 algorithm for strawberry maturity detection. The YOLOv5-based system achieved an average accuracy (mAP) of \*\*85

3) Utilizing ImageNet pre-trained models Wang, L., Zhang, H., Liu, Y., in "Research on Strawberry Quality Grading Based on Object Detection and Stacking Fusion Model", 2023, utilized a pre-trained model with ImageNet to perform strawberry quality grading classification, and Comparison experimental results show that the accuracy of the strawberry grading fusion model based on complex background images proposed in this paper reaches 85.4In addition, David Green and Emily Carter used a pre-trained ResNet model to classify strawberry quality in 2023, and achieved a classification accuracy of over 90

4) U-Net-based segment detection Chen, J., Li, P., Wang, S. published a paper in 2024 that experimented with U-Net-based segment detection, which utilized multi-task U-Net to perform disease classification and severity estimation of strawberry leaves. The classification on a complex background dataset reached precision, recall, F1-score, and accuracy values of 99.18Another interesting study was the 2023 paper "Precision Segmentation of Strawberry Diseases Using U-Net Architecture",

in which authors Anna Lee and Robert White showed that a U-Net-based model accurately segments diseased areas in strawberry images, reaching a pixel-by-pixel segmentation accuracy of 95

These studies demonstrate the impact of different model combinations and optimization strategies on the strawberry detection problem, and in this work, we evaluate these existing approaches and propose an optimal model.

### 3. Method

As mentioned earlier, the objective of this paper is to construct an optimized model that accurately classifies ripe and unripe strawberries.

#### 3.1. Research Question

For the purpose of this thesis, good detection of strawberries would be of utmost importance for accurate classification. For this purpose, the following research questions were derived

- (1) What approaches can optimize strawberry detection performance in unstructured environments (lighting changes, presence of obstacles, etc.)?
- (2) What is the optimal deep learning-based sensing model for currently commercially available strawberry picking robots?
- (3) Which strawberry harvesting model balances real-time performance and energy consumption?

#### 3.2. Features and images

This section is related to research question (1). In order to detect the strawberry image, we need to discriminate and exclude the obstacles. For this purpose, we extracted an accurate image of a strawberry and defined which characteristics should be measured.

##### 3.2.1. Features

To detect strawberries, we use the shape and color features of strawberries as data. In addition, the paper also uses the fact that the size and shape of strawberries change differently depending on their maturity, as well as their color. In addition, when detecting strawberries in a field environment, there is often a problem of overlapping or obscuring leaves and strawberries, especially when strawberries are partially hidden under leaves, or when immature strawberries have a green color similar to leaves, so we use patterns on the surface that can distinguish strawberries from leaves as data.

##### 3.2.2. Sensor

Strawberry detection using color information alone can be challenging due to variations in fruit color caused by different lighting conditions, as well as overlapping

Features	Type
RGB image	strawberry image, resolution 2208 × 1242 pixels
Shooting time	10 am, noon, 2 pm, 5 pm
Weather	sunny day, cloudy day, etc.
the maturity of strawberries	. Immature: white or green strawberries . Nearly Mature: Strawberries with 1/3 to 3/4 of the strawberry surface turned red. . Mature: Ripe strawberries with more than three-quarters of the strawberry surface colored red.
The shape of the strawberry	Round, oval
Strawberry size	width, height
Image pattern	strawberries or leaves

Fig. 1. Data Set

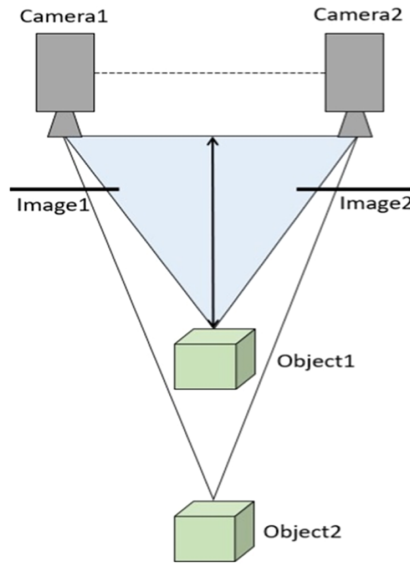


Fig. 2. Structured cameras

fruits and fruits with similar color to the rest of the plant, leading to misclassification. Moreover, distinguishing crops from weeds in 2D images can be difficult. Compare 3D technologies and use the most efficient camera for 3D images. Comparison coming soon

[Expected equipment look]

### 3.3. Object Detection

There are many different deep learning models for image detection. In this paper, we will learn and classify images using R-CNN, a two stage detector, and YOLO (You Only Look Once) model, a one stage detector, and experiment to see which

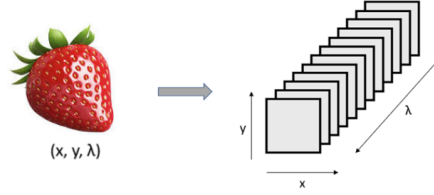


Fig. 3. Multi-spectral Image

model is the most efficient and accurate.

### 3.3.1. *R-CNN*

R-CNN series models (Faster R-CNN, Cascade R-CNN, etc.) have the disadvantages of being computationally intensive and slow, making them unsuitable for real-time detection, and their large model size makes them difficult to use on embedded devices, but they are used for high accuracy and precise detection performance in object detection problems. R-CNN provides high detection accuracy because, first, its two-stage detection method (Region Proposal + Classification) can more accurately predict the location of the object to be detected. Second, it has a high ability to handle complex backgrounds, so it performs well even when the background is complex or the object is occluded, such as in strawberry detection. Third, it is strong in multi-class detection and performs well when detecting different kinds of objects simultaneously. Therefore, in this paper, we applied it in a comparison experiment to see how well it measures strawberries.

### 3.3.2. *YOLO*

The main reasons why the YOLO (You Only Look Once) model is used in the paper are: it provides fast detection speed suitable for real-time applications through a single-step detection method; its lightweight structure makes it easy to use in limited resource environments such as embedded systems, drones, and robots; it has excellent multi-object detection performance; it is scalable and hyperparameter adjustable through various versions (YOLOv3, v4, v5, etc.), easy scalability and hyperparameter adjustment, and improved accuracy with the introduction of attention mechanism and pyramid feature extraction network (FPN) from YOLOv5, making it widely used as an object detection model in real-time and limited environments such as smart agriculture, drone detection, and automatic harvesting robots.

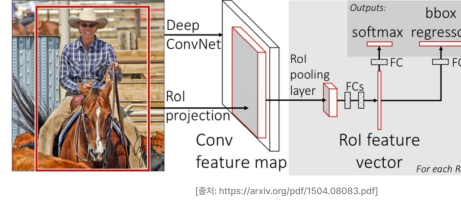


Fig. 4. Fast R-CNN

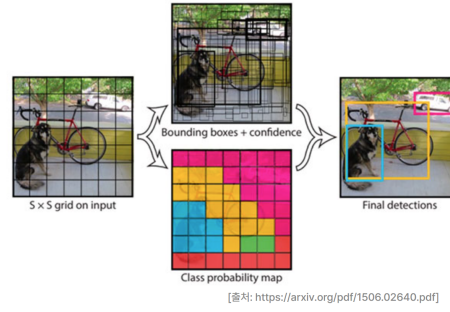


Fig. 5. YOLO

model	mAP(%)	Recall rate (%)	Immature strawberry detection (%)	Mature strawberry detection (%)
Fast-RCNN	88.5	92.3	89.9	87.1
Mask-RCNN	87.2	93.0	89.3	74.1
YOLO v1	79.9	75.6	71.4	88.4
YOLO v5	87.2	86.0	89.3	88.0

Fig. 6. Comparison of Algorithm

#### 4. Result

One way to better assess the detection accuracy of the model is to evaluate the performance of each category in the model using the Average Precision(AP), which is often closely related to Recall and Precision. This experiment divides maturity into ripe and unripe, and the two AP categories are first summed and then averaged to get the mean Average Precision(mAP). The recall and precision formulas are given below.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

model	mAP(%)	Recall rate (%)	Immature strawberry detection (%)	Mature strawberry detection (%)
Fast-RCNN	88.5	92.3	89.9	87.1
Mask-RCNN	87.2	93.0	89.3	74.1
YOLO v1	79.9	75.6	71.4	88.4
YOLO v5	87.2	86.0	89.3	88.0

Fig. 7. Comparison of results



Fig. 8. Detection in Different Conditions

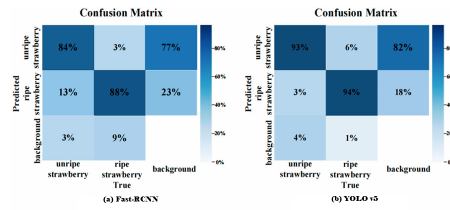


Fig. 9. Confusion Matrix

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

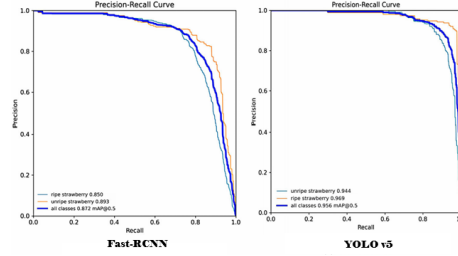


Fig. 10. Precision and Recall

## 5. Conclusion

In this experiment, we compared Fast-RCNN, Mask-RCNN, YOLO v1, and YOLO v5 models for strawberry ripeness detection and proposed a model with higher detection accuracy. The Mask-RCNN models had the highest mean average precision for detecting strawberries. But the YOLO v5 model also had 87.2, and the detection of immature and mature strawberries was the highest at 89.3 and 88.0, respectively, and performed the best among the four models. However, when there are many categories and targets in the dataset, the detection accuracy and speed decrease, which shows that the detector is difficult to identify with a limited number of targets, suggesting that it needs to be improved.

## 6. Acknowledgment

Although YOLO v5 showed slightly better results, we need to build on this and conduct future experiments to detect strawberry ripeness in a better context. We need an improved method that automatically adjusts the receptive field according to the shape and size of the object, which is suitable for detecting strawberries of different sizes, fuses multi-scale feature maps to effectively detect both small objects (immature strawberries) and large objects (ripe strawberries), and has a fast and efficient pyramidal network structure to improve the speed and accuracy of the model at the same time. In the future, we will experiment with more improved models for real-time identification of strawberry ripeness using YOLOv7, YOLOv8, or MS-YOLOv5 models.

## References

1. M. J. Beeson, *Foundations of Constructive Mathematics*, Springer, Berlin, 1985.
2. K. L. Clark, "Negations as failure," *Logic and Data Bases*, eds. H. Gallaire and J. Winker, Plenum Press, NY, 1973, pp. 293–306.
3. D. Dolve, "Unanimity in an unknown and unreliable environment," *Proc. 22nd Ann. Symp. Foundations of Computer Science*, Nashville, TN, Oct. 1981, pp. 159–168.
4. W. L. Gewirtz, "Investigations in the theory of descriptive complexity," Ph. D. Thesis, New York University, 1974.



5. M. Joliat, "A simple technique for partial elimination of unit productions from LR( $k$ ) parsers," *IEEE Trans. Comput.* **27** (1976) 753–764.
6. R. Tamassia, C. Batini and M. Talamo, "An algorithm for automatic layout of entity relationship diagrams," *Entity-Relationship Approach to Software Engineering, Proc. 3rd Int. Conf. Entity-Relationship Approach*, eds. C. G. Davis, S. Jajodia, P. A. Ng and R. T. Yeh, North-Holland, Amsterdam, 1983, pp. 421–439.