Improving the Efficiency of Image Recognition for Yuzu Fruit Counting Using Object Recognition Models

Takahiro Sugiyama and Shinichi Yoshida(✉)

Kochi University of Technology, Kochi 782-8502, Japan
275104w@gs.kochi-tech.ac.jp, yoshida.shinichi@kochi-tech.ac.jp

Abstract. Modern agriculture faces a labor shortage due to aging and a decrease in new farmers. Artificial intelligence (AI) and data utilization aim to improve productivity. Crop detection is one example, where object recognition models automate the process compared to manual detection relying on farmer experience. However, the challenge lies in training data requirements and variations in label assignment. This research investigates how different label assignment methods impact object recognition. We compare the labeling conditions using the YOLO and assess their effect on accuracy. Increasing target classes in test data helps maintain precision, while reducing recall. Detailed labeling improves average precision.

Keywords: image object detection · YOLO · agriculture · labeling

1 Introduction

In Japan, a problem on agriculture is a shortage and aging of agricultural workforce. According to the Ministry of Agriculture, Forestry, and Fisheries in Japan, the average age of basic agricultural workers was 67.1 years old with 1.757 million people in 2015. However, in 2022, the average age increased to 68.4 years old, and the number of workers decreased to 1.226 million. Furthermore, the number of new entrants decreased from 65,000 in 2015 to 52,000 in 2021. [1] To address these challenges, agriculture has adopted IT technologies, including IoT, AI, and robotics. Modern agriculture aims to improve productivity, reduce labor, and create an accessible environment for all through the use of IT technologies. An example is the IoP project in Kochi Prefecture, Japan, which manages cultivation information by comparing shipping and growth data, utilizes data for efficient farming, and employs image recognition technology for fruit yield estimation. This study focuses on the detection of crop types and varieties and prediction of shipping and yield using image recognition in agricultural technology. Traditional methods relied on the experience and intuition of agricultural experts or required significant labor from young individuals. However, AI-based image recognition enables accurate predictions without relying on human intuition or labor. Nonetheless, training AI models for crop recognition requires a

B. Xin et al. (Eds.): IWACIII 2023, CCIS 1932, pp. 156–166, 2024.
https://doi.org/10.1007/978-981-99-7593-8_15

large amount of image data collection, which poses challenges in terms of time and cost. Moreover, there is a lack of consensus on labeling methods, resulting in variations in how labels are assigned for different crops, and the impact of labeling methods on object detection accuracy remains unclear. Therefore, this research aims to investigate the influence of label assignment methods on the performance of image recognition models. By focusing on whether labels should be assigned only to fully visible objects or to all partially visible objects, the study compares different labeling conditions using object detection algorithms and examines the effect on object detection accuracy. Additionally, future research will target the early identification of green Yuzu (Japanese citrus) fruits for harvesting season predictions.

2 Related Works on Crop Detection Using Image Recognition AI

The detection and counting of crops using image recognition AI is a crucial challenge in the field of agriculture. Tasks such as crop detection, classification, and evaluation are time-consuming and difficult to perform efficiently manually. However, these tasks can be automated using image recognition AI, significantly contributing to improved productivity and labor efficiency by accurately counting the number of crops. In the literature, for example, a tomato harvesting robot was developed using image recognition algorithms and pattern recognition models to automatically detect and classify tomatoes [2]. Another study explored methods to improve detection accuracy by combining existing learning models for accurately recognizing and counting different types of grapes [3]. Inspired by these works, this study investigates the learning approaches that can enhance object detection accuracy in AI.

2.1 YOLO(You Look only Once)

YOLO (You Only Look Once) is a real-time object detection algorithm widely used to detect objects. The name "You Only Look Once" reflects its ability to classify objects and estimate their positions in an image or video by passing the image through a convolutional neural network (CNN) in a single forward pass. This allows for efficient and accurate identification of object names and their corresponding coordinates within the input image or video (YOLO Official Website).

Various applications have utilized YOLO, including lettuce detection in gardens [4] and tasks such as people counting, traffic monitoring, and intrusion detection in restricted areas [5].

One notable feature of YOLO (You Only Look Once) compared to other object detection models like Faster R-CNN is its significantly faster processing time, performing object detection approximately 6–7 times faster. Additionally, YOLO has the capability to make predictions for the entire image at once. A key component

Fig. 1. Example of object detection with YOLO

of YOLO's object detection process is the use of bounding boxes, which are rectangular shapes utilized to approximate the regions of objects within an image. Each bounding box is assigned coordinates, and the confidence score indicates the likelihood of an object's presence within that region [6].

Figure 1 illustrates an example of object detection using YOLO, where 8156 instances of the label "yuzu" were assigned. It demonstrates the confidence scores and corresponding bounding boxes for detecting yuzu fruits, showcasing the real-time detection capability. In this study, YOLOv5 [7] is employed as the image recognition model. Custom training data is created for each experimental condition, and the model is updated through training to investigate the impact of label assignment conditions on object detection accuracy.

We train pretrained YOLO model using label of Yuzu fruits. We employed LabelImg software for the annotation of Yuzu images (Fig. 2).

3 Experiment

This chapter describes the experiments conducted in the study.

3.1 Dataset for Object Detection

In this study, a total of 357 images depicting yuzu fruits were used. (Images without any yuzu fruit present were not included, and external factors such as

Fig. 2. Example of Labeling with LabelImg

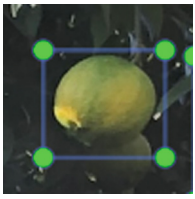

Fig. 3. yuzu

Fig. 4. half_yuzu

Fig. 5. unknown

location, date, weather, and lighting were not considered.) The images depicting yuzu fruits were labeled into three classes based on the author's subjective judgment: yuzu (images where yuzu fruits could be confidently identified), half_yuzu (images where yuzu fruits were partially obscured by leaves or other external factors, accounting for approximately 50% of the dataset), and unknown (images where it was difficult to determine if they contained yuzu fruits), as shown in Figs. 3, 4 and 5. The labeling process was performed manually by the author using LabelImg. As there may be individual differences in annotation, the author conducted the labeling process alone.

Next, the dataset was divided into 258 images for training, 116 images for validation, and 77 images for testing. The training dataset was used to update the model's weights during training. The validation dataset was used for tuning the hyperparameters, which are parameters set before the model's training to determine its performance. The hyperparameters were adjusted and the accuracy was evaluated using the validation dataset iteratively during training to find the best-performing hyperparameters. Finally, the testing dataset was used to evaluate the accuracy of the trained model. By using unseen data during training, it was possible to assess the model's ability to handle unknown data accurately.

Table 1. Number of each label

Label type	yuzu	half_yuzu	unknown
Number of labels	3516	1999	2641

Table 2. Conditions for correct labeling

Class of correct label before conversion	Class of correct label after conversion
Correct answer condition A: before conversion	correct answer condition A: after conversion
(0, 1, 2)	(0, x, x)
Correct answer condition B: before conversion	correct answer condition B: after conversion
(0, 1, 2)	(0, 0, x)
Correct answer condition C: before conversion	correct answer condition C: after conversion
(0, 1, 2)	(0, 0, 0)

The number of labels and the percentage of each label are shown in Table 1.

3.2 Label Requirements

The classes yuzu, half_yuzu, and unknown were categorized as (0, 1, 2), respectively, and the ground truth labels were classified as shown in Table 2. The intention behind this classification was as follows: Condition A represents images where yuzu fruits are confidently identified as the ground truth, Condition B includes images where yuzu fruits are partially obscured by leaves or other factors (approximately 50% of the dataset), and Condition C encompasses images where it is uncertain whether they contain yuzu fruits. This classification allowed for a comparison among these three ground truth conditions. Each condition was trained using the YOLOv5 framework.

The PR-curve and Average Precision (AP) were compared for all combinations of Condition A, Condition B, and Condition C in the training and testing labels, as shown in Table 3. In this evaluation, the IoU (Intersection over Union) threshold of 0.5 or higher was used. The training was conducted using YOLOv5, with a fixed batch size of 8 and 300 epochs.

4 Results and Discussion

4.1 PR-Curve for Each Combination for Each Condition

The PR-curves for each combination of test and training conditions are depicted in Figs. 6, 7 and 8. Figure 6 represents the PR-curves for the test condition A using the three training conditions, Fig. 7 shows the PR-curves for the test condition B, and Fig. 8 illustrates the PR-curves for the test condition C. By comparing these figures, we observe that increasing the number of classes considered as correct in the test data results in a reduced decline in Precision and a decrease in Recall.

Table 3. Combination of conditions for learning and testing

Conditions for correct labels for training	conditions for correct labels for testing
Correct answer condition A for study	A for test
Correct answer condition B for study	A for test
Correct answer condition C for study	A for test
Correct answer condition A for study	B for test
Correct answer condition B for study	B for test
Correct answer condition c for study	B for test
Correct answer condition A for study	C for test
Correct answer condition B for study	C for test
Correct answer condition C for study	C for test

This suggests that as the number of correct predictions increases in the test data, the probability of correct predictions also rises, leading to higher Precision values. Conversely, as the number of predicted labels increases relative to the number of correct labels, the possibility of false negatives occurring becomes more likely, resulting in a decrease in Recall. Notably, Fig. 8 exhibits Precision and Recall values closer to 1, indicating that more detailed labeling in both the test and training data yields more appropriate label assignments.

4.2 AP per Condition for Each Label

The comparison of AP values for each label condition is presented in Fig. 9. When the test condition is A and the training condition is C, the AP value is at its minimum, while both conditions being C result in the highest AP value. This suggests that when even partially visible objects are considered for detection, performing detailed labeling and considering all such instances as correct targets can lead to improved detection accuracy.

Furthermore, for the test condition of C, the AP values increase in the order of training conditions A, B, and C. However, for the test conditions of A or B, there is no consistent improvement in AP with respect to the order of training conditions A, B, and C. This discrepancy may be attributed to an increased number of predicted labels in locations other than the correct ones, which negatively impacts AP.

Moreover, it is observed that the AP is higher for the training condition B and the test condition A compared to both conditions being A. This indicates that inadequate training might have occurred for condition A due to the lower number of labels, whereas training for condition B approaches the number of correct labels in the test data. To address this issue, increasing the number of training images and the number of labels is considered a potential solution for improvement.

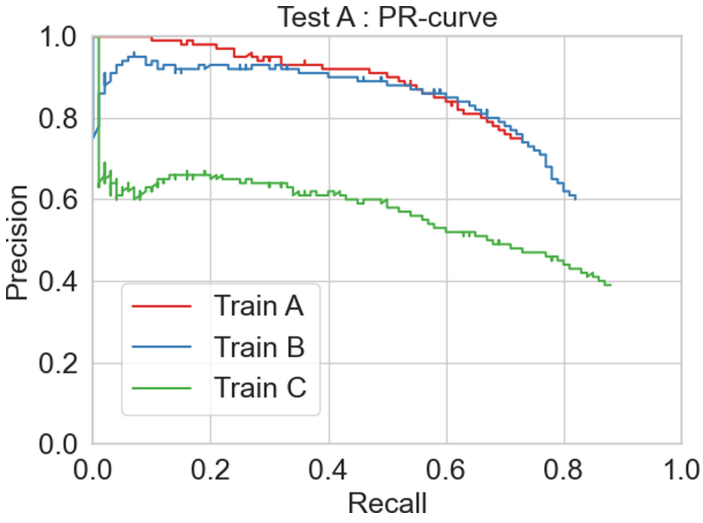

Fig. 6. PR-curve for correct answer condition A for testing

4.3 Detection Prediction Results After Learning Each Corrective Condition

After training for each condition, the predicted detection results for each condition are shown in Fig. 10 for condition A, Fig. 11 for condition B, and Fig. 12 for condition C. Upon comparison, it is observed that the number of predicted labels increases in the order of A, B, and C. Additionally, when comparing A and B, it is evident that while A correctly predicts the top-left yuzu, B fails to make a prediction. Despite increasing the number of classes considered as correct targets, this outcome may be attributed to the fusion of features from yuzu and half_yuzu during the feature extraction process in condition B, leading to the determination that it is not yuzu. In contrast, when training under condition C, there were no similar omissions in the predicted labels as observed in B, indicating a more appropriate learning process.

Based on the obtained results, compared with three labeling conditions, A, B, C, it is better to label all objects in the image regardless of the percentage of their presence (occlusion) in the label. Furthermore, in this study, labeling was performed using three classes corresponding to different percentages of yuzu presence. However, it is expected that further subdivision of classes and experimentation would lead to a stronger improvement trend in PR-curve and AP values.

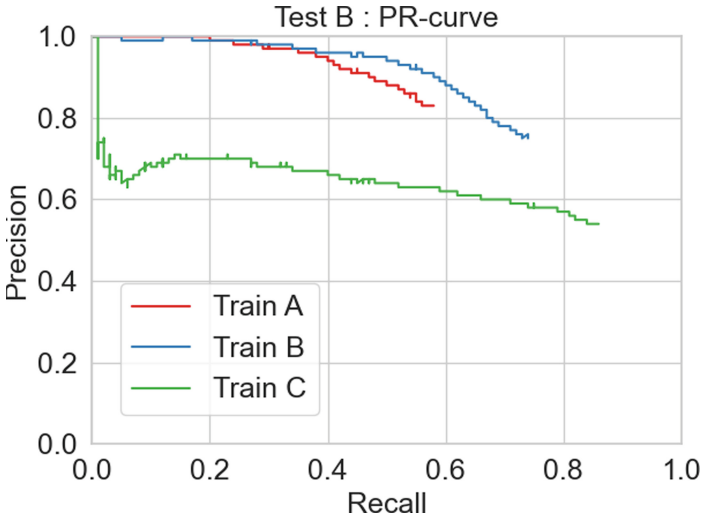

Fig. 7. PR-curve for correct answer condition B for testing

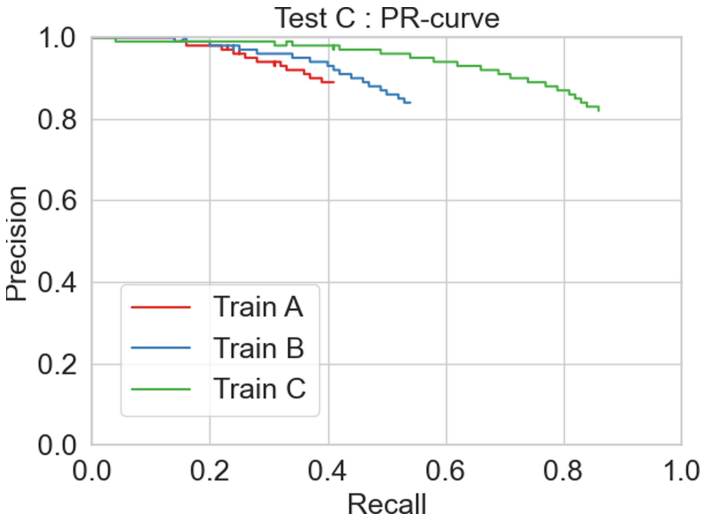

Fig. 8. PR-curve for correct answer condition C for testing

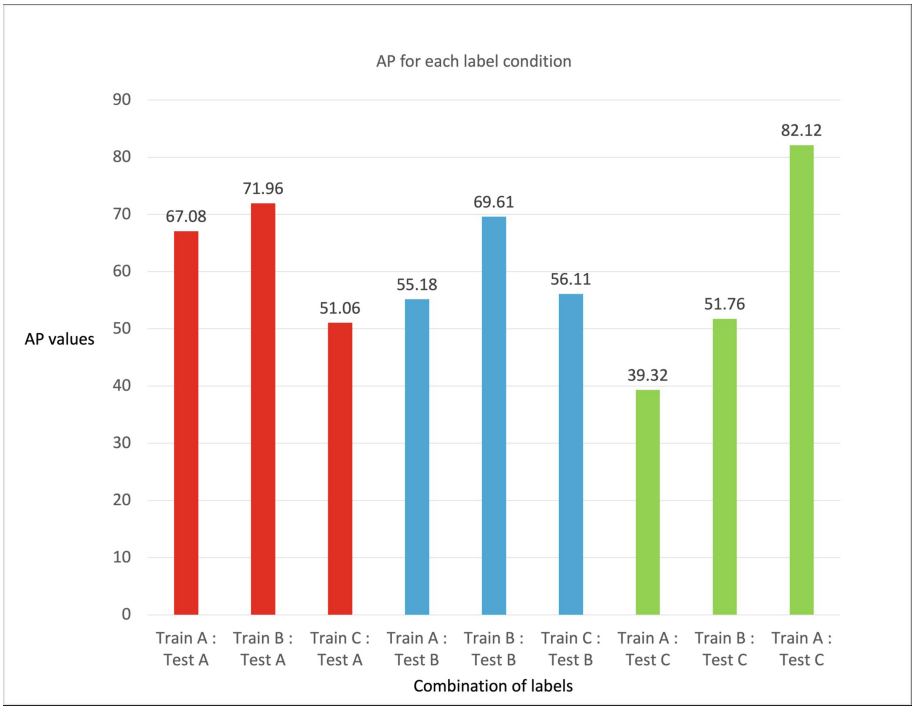

Fig. 9. AP per condition for each label

Fig. 10. Detection Prediction Results for Correct Condition A

Fig. 11. Detection Prediction Results for Correct Condition B

Fig. 12. Detection Prediction Results for Correct Condition C

5 Conclusion

In this study, the optimization of labeling was the main objective, focusing on the criteria for determining the presence of the target object in the image. Specifically, we investigated whether to label only the objects that are clearly visible or label all partially visible objects. Labels were assigned to three classes based

on the percentage of yuzu presence in the image. After training with YOLOv5, experiments were conducted to evaluate how the object detection accuracy varied across the nine combinations of test criteria and labeling. PR-curve, AP values, and changes in predicted images were used as evaluation metrics. Comparing the PR-curves, it was observed that increasing the number of correct answers in the test data led to higher precision values, as the probability of correct predictions increased. However, this also resulted in a decrease in recall values, as more predictions were made, leading to an increased probability of false negatives. Comparing AP values, it was concluded that labeling all target objects regardless of their presence percentage yielded better results. Furthermore, subdividing the classes into finer categories showed a strong improvement trend in PR-curves and AP values. When comparing the predictions after training with different test criteria, it was found that as the number of classes to be considered as correct answers increased, predictions that were previously correct disappeared, potentially due to the fusion of different features during feature extraction. Based on these findings, it was suggested that performing labeling corresponding to the predetermined percentage of object presence in the image across a large number of images would result in more appropriate labeling. Therefore, future research is needed to expand the number of data and the scope of the experiment, including further exploration of labeling methods and data expansion, in order to obtain more generic and reliable results. A future prospect for the first step is to concentrate on extracting only green yuzu for early detection at harvest time.

Acknowledgement. This work was supported by Cabinet Office grant in aid, the Advanced Next-Generation Greenhouse Horticulture by IoP (Internet of Plants), Japan.

References

1. Ministry of Agriculture, Forestry and Fisheries: Statistics on agricultural labor force (in Japanese) (2023). https://www.maff.go.jp/j/tokei/sihyo/data/08.html
2. Uegaki, S., Araki, H., Toshima, R., et al.: Tomato harvesting robot using AI for environment recognition. Panasonic Tech. J. **64**(1), 54–58 (2018)
3. Santos, T., Souza, L., et al.: Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association (2020). https://doi.org/10.1016/j.compag.2020.105247
4. Idemura, S., Senda, Y., et al.: Lettuce detection in the field using deep learning (in Japanese). In: The Society of Instrument and Control Engineers Chubu Branch Symposium 2017, Lecture No.PD-3 (2017)
5. Syunsaku, S., Toyota, M., et al.: Development of a Device to Assist the Visually Impaired Using YOLO. Kansai University (2022). https://wps.itc.kansai-u.ac.jp/acoust/wp-content/uploads/sites/190/2022/02/2021_b_shimizu.pdf
6. Redmon, J., Divvala, S., et al.: You look only look once: unified, real-time object detection (2016). https://arxiv.org/pdf/1506.02640v5
7. Jocher, G.: YOLOv5 (2021). https://github.com/ultralytics/yolov5
8. Tzutalin: Labelimg (2018). https://github.com/heartexlabs/labelImg