AgriEngineering

MDPI

Article

Object Detection for Yellow Maturing Citrus Fruits from Constrained or Biased UAV Images: Performance Comparison of Various Versions of YOLO Models

Yuu Tanimoto [1,2,*], Zhen Zhang [1] and Shinichi Yoshida [1]

[1] Department of Engineering, Graduate School of Engineering, Kochi University of Technology, Kami 782-8502, Japan
[2] Kochi Agricultural Research Center Fruit Tree Experiment Station, Kochi 780-8064, Japan
* Correspondence: yuu_tanimoto@ken2.pref.kochi.lg.jp

Abstract: Citrus yield estimation using deep learning and unmanned aerial vehicles (UAVs) is an effective method that can potentially achieve high accuracy and labor savings. However, many citrus varieties with different fruit shapes and colors require varietal-specific fruit detection models, making it challenging to acquire a substantial number of images for each variety. Understanding the performance of models on constrained or biased image datasets is crucial for determining methods for improving model performance. In this study, we evaluated the accuracy of the You Only Look Once (YOLO) v8m, YOLOv9c, and YOLOv5mu models using constrained or biased image datasets to obtain fundamental knowledge for estimating the yield from UAV images of yellow maturing citrus (*Citrus junos*) trees. Our results demonstrate that the YOLOv5mu model performed better than the others based on the constrained 25-image datasets, achieving a higher average precision at an intersection over union of 0.50 (AP@50) (85.1%) than the YOLOv8m (80.3%) and YOLOv9c (81.6%) models in the training dataset. On the other hand, it was revealed that the performance improvement due to data augmentation was high for the YOLOv8m and YOLOv9c models. Moreover, the impact of the bias in the training dataset, such as the light condition and the coloring of the fruit, on the performance of the fruit detection model is demonstrated. These findings provide critical insights for selecting models based on the quantity and quality of the image data collected under actual field conditions.

Keywords: data augmentation; object detection; smart agriculture; yield estimation

Citation: Tanimoto, Y.; Zhang, Z.; Yoshida, S. Object Detection for Yellow Maturing Citrus Fruits from Constrained or Biased UAV Images: Performance Comparison of Various Versions of YOLO Models. *AgriEngineering* **2024**, *6*, 4308–4324. https://doi.org/10.3390/agriengineering6040243

Academic Editor: Simone Pascuzzi

Received: 4 October 2024
Revised: 5 November 2024
Accepted: 13 November 2024
Published: 15 November 2024

1. Introduction

Citrus fruits are essential crops for the global agricultural market. In 2021, citrus fruits were the second most produced worldwide, with a total production of 161.8 million tons. This extensive production spanned over 10.2 million hectares of agricultural land [1]. Global production of citrus fruits has increased by 50% over the past two decades, reflecting their growing importance in industry and people's livelihoods. As fruit production increases, the mechanization of cultivation also progresses. However, harvesting remains labor-intensive, especially in regions such as Japan. In this area, citrus crops are grown on sloping land where advanced machinery is limited in many cases [2], requiring a significant amount of hired labor in a short period. Therefore, accurate yield estimation before harvesting is crucial for reducing costs and optimizing labor hiring. Yield estimation is also necessary to maximize profits through strategic marketing and to further develop smart agriculture technology.

Studies have aimed to predict the number of fruits before harvesting using optimal sampling strategies for decades. For example, the United States Department of Agriculture [3] reported that sampling surveys of certain branches were sufficient to achieve a coefficient of variation within 10% when estimating fruit set early in the growing season

for early orange, Valencia orange, and grapefruit in blocks. Stout [4] observed that a survey of two trees per orchard and four revised frame counts per tree in Temple, tangerine, and tangelo varieties resulted in sampling errors regarding the average number of fruits per tree ranging from 14% to 32%. Sampling methods for estimating total orchard yield have also been developed for other fruit trees [5,6]. These sampling methods, created through the efforts of numerous researchers, are highly accurate in estimating yields; however, the sampling itself is labor-intensive, and more labor-saving methods are required.

Therefore, yield estimation using image processing has become the focus of recent studies. Yield estimation using image processing is an effective method with the potential to achieve both high accuracy and labor savings. In particular, advances in computer technology have promoted the development of object detection methods using Artificial Intelligence, particularly deep learning (DL). Recent studies on citrus fruit detection using DL were reported by Zhang et al. [7], Li et al. [8], Gremes et al. [9], Jing et al. [10], and Ang et al. [11]. These fruit detection methods demonstrate significant potential for reducing labor in sampling for yield estimation before harvest.

Simultaneously, progress in unmanned aerial vehicles (UAVs) has made it possible to acquire substantial amounts of image data inexpensively, reproducibly, and in a short timeframe. Consequently, several studies have been published that combine UAV images of fruit trees with DL. For instance, Apolo-Apolo et al. [12] with Navelina sweet orange, Novelero et al. [13] with mature coconut, Xiong et al. [14] with litchi, Wang et al. [15] with apple, and Arakawa et al. [16] with chestnut fruit (bur) have constructed fruit detection models. The combination of UAVs and DL has the potential to dramatically improve yield estimation methods.

In most of these studies, an object detection method called YOLO (You Only Look Once) has been used for fruit detection. Object detection methods can be classified into one- and two-stage detection models. A two-stage detection method is divided into two steps: the extraction of the region proposal and the identification of object classes within the region proposal. Faster R-CNN is the primary model in this category [17]. In general, two-stage object detection shows high detection performance but slow inference speed, limiting its use in actual scenarios. Thus, one-stage object detection, which offers both high accuracy and fast inference speeds, is used for fruit detection in actual orchards. A one-stage detection method is an approach that detects objects efficiently by predicting both the location and class of objects in a single time over the entire image. The one-stage object detection model proposed by Redmon et al. [18], called YOLO, is characterized by its speed. YOLO was released in 2016, and YOLOv2 [19], YOLOv3 [20], YOLOv4 [21], YOLOv5 [22], YOLOv6 [23], YOLOv7 [24], YOLOv8 [25], YOLOv9 [26], YOLOv10 [27], and YOLO11 [28] are currently available.

Achieving good performance in object detection models requires numerous labeled training images. However, in many cases, significant difficulties are encountered due to the scarcity of available images. Data augmentation (DA) is a key technique for generating a large number of images and conducting effective model training. DA includes geometric transformation, color transformation, blurring, and noise addition [29,30]. Additionally, advanced methods, such as CutMix [31], replace part of an image with a patch from another image, whereas Random Erasing [32] erases random regions in an image to enhance model robustness. To enhance dataset diversity, generating new images using generative adversarial networks (GANs) and incorporating them into training data is highly beneficial. This approach is particularly effective in scenarios in which data are scarce or when greater variation is required. GANs, proposed by Goodfellow et al. [33], consist of two networks: the generator aims to produce data closely resembling the training data to deceive the discriminator, which in turn distinguishes between the training and generated data. Through this adversarial process, the generator learns to generate data that aligns closely with the training distribution. Furthermore, Pix2pix [34], a type of conditional GAN, generates training data by mapping input images to the corresponding output images. These meth-

ods collectively address the challenges posed by limited data availability, enhance dataset diversity, and ultimately improve the performance of object detection models.

Considering the excellent compatibility between DL-based object detection and UAV images, further research on fruit yield estimation methods using these technologies is expected. However, many citrus fruit varieties with different shapes and colors require varietal-specific fruit detection models. Understanding the model performance on constrained or biased image datasets is crucial for determining a method for improving model performance. In certain cases, the training results for constrained or biased image datasets can be applied to a model for few-shot object detection.

Hence, the objective of this study was to evaluate the accuracy of the YOLOv8m, YOLOv9c, and YOLOv5mu models of the YOLO series with constrained or biased image datasets to obtain fundamental knowledge for estimating the yield of citrus trees based on UAV images. Although there have been many studies of object detection in citrus fruits with an orange color, there have not been many studies in fruits with a yellow color, such as lemons. Therefore, we conducted this study using *Citrus junos* Sieb. ex Tanaka, also known as yuzu, which is a yellow maturing citrus fruit (Figure 1). *Citrus junos* is a commonly cultivated citrus cultivar in Japan, Korea, and China [35–37]. This fruit has an attractive fragrance and is strongly acidic. In Japan, yuzu juice has traditionally been used as a substitute for vinegar and seasoning, instead of being consumed as fresh fruit.

Figure 1. Fruits of *Citrus junos* Sieb. ex Tanaka [38].

The remainder of this paper is organized as follows: Section 2 provides details on the dataset collection and describes the methodology used for detecting *Citrus junos* fruit. Section 3 presents and discusses the results. Finally, Section 4 summarizes the research and concludes this study.

2. Materials and Methods

2.1. Acquision of UAV Images

The UAV images were acquired using *Citrus junos* on trifoliate orange rootstocks in two orchards (Orchards 1 and 2) at the Kochi Agricultural Research Center Fruit Tree Experiment Station in Kochi City and two farmers' orchards (Orchards 5 and 7) in Kami City, Japan. The orchard number and location were consistent with those reported by Tanimoto and Yoshida [38]. The UAV images were collected between October and November from 2020 to 2023, during the yellow maturing period of the fruits. Figure 2 shows the locations of the experimental *Citrus junos* orchards, and Table 1 presents an overview of the experimental orchards.

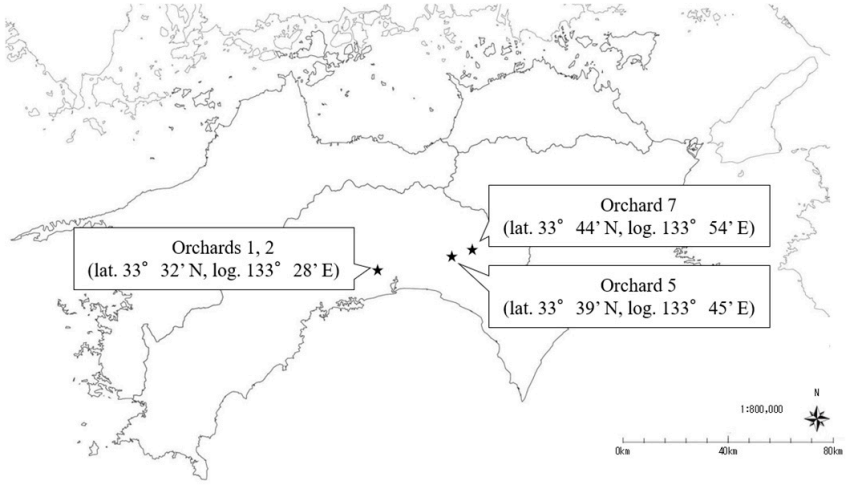


Figure 2. The locations of the experimental *Citrus junos* orchards. This study was conducted in two orchards (Orchards 1 and 2) at the Kochi Agricultural Research Center Fruit Tree Experiment Station in Kochi City and two farmers' orchards (Orchards 5 and 7) in Kami City, Japan. This figure was obtained from the report by Tanimoto and Yoshida [38].

Table 1. Overview of experimental *Citrus junos* orchards.

Train or Test	Orchard Number	Acquisition Date of UAV Images	Number of Trees	Weather Condition of Acquisition Date	Fruit Maturing of Acquisition Date	Harvesting Date
Train	1	2020/10/28	40	Cloudy	Completely Matured	2020/11/06
Train	1	2021/10/31	45	Cloudy	Completely Matured	2021/11/11
Test	1	2022/10/27	15	Cloudy	Completely Matured	2022/11/08
Train	2	2020/10/30	20	Sunny	Completely Matured	2020/11/04
Test	2	2021/11/05	15	Cloudy	Completely Matured	2021/11/12
Train	2	2022/10/27	20	Cloudy	Completely Matured	2022/11/02
Test	5	2023/10/18	15	Cloudy	Under-Matured	2023/11/07
Test	7	2023/10/25	10	Sunny	Under-Matured	2023/11/09

The weather conditions for acquiring UAV images for Orchard 2 in 2020 and Orchard 7 in 2023 were sunny, whereas images for the other orchards and other years were taken in cloudy conditions. Additionally, UAV images of Orchards 5 and 7 in 2023 were acquired when the fruit was under-matured, whereas images of the other orchards were captured when the fruit was completely matured. The DJI Mavic Mini UAV (Da-Jiang Innovations Science and Technology Co., Ltd., Shenzhen, China) was used to acquire images. The sensor was a 1/2.3-inch Complementary Metal-Oxide-Semiconductor camera with a field of view of 83°, a focal length of 24 mm (equivalent to 35 mm in full-frame format), a focal ratio of f/2.8, and focus from 1 m to infinity. To obtain comprehensive images for analysis, the average flight height was between 5 m and 9 m, and individual trees were photographed vertically. The image resolution was 4000 × 2250 pixels, and the files were in JPG format.

2.2. Dataset Construction

The training datasets included UAV images from Orchard 1 in 2020 and 2021 and Orchard 2 in 2020 and 2022 (Table 1). All images were cropped to 1200 × 1200 pixels from the center of *Citrus junos* trees to ensure consistency in the region of interest (Figure 3), and the files were in PNG format. To annotate fruits in training images, the open annotation software "Labelimg" was used, and annotations were implemented carefully and accurately by double-checking. In this study, three training datasets were constructed to compare the detection performance of the YOLOv8m, YOLOv9c, and YOLOv5mu models. Figure 4 illustrates a flowchart of the dataset preparation.

Figure 3. The training dataset UAV images. (**a**) A tree image from Orchard 1 in 2020. (**b**) A tree image from Orchard 1 in 2021. (**c**) A tree image from Orchard 2 in 2020. (**d**) A tree image from Orchard 2 in 2022.

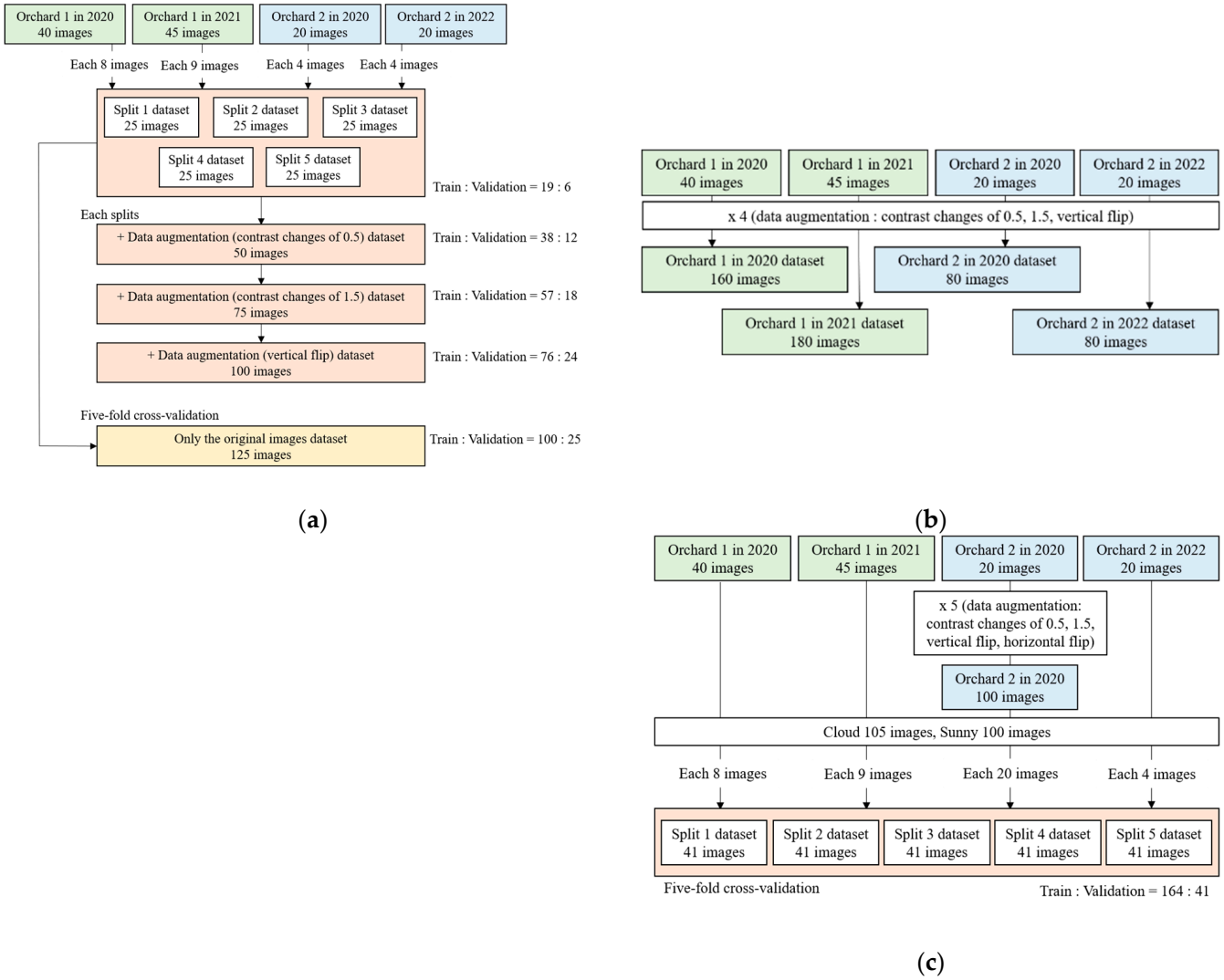


Figure 4. Flowchart of dataset preparation. (**a**) Constrained image dataset preparation process. (**b**) Biased image dataset preparation process. (**c**) Dataset preparation process ensuring approximately equal number of images collected on sunny and cloudy days.

For the first dataset, the data (images and annotations) from each orchard and year were randomly divided into five parts combined to generate five-fold datasets. Each dataset consisted of 25 images (25-image dataset). Based on each 25-image dataset, step-by-step

DA was performed to evaluate the performance of the fruit detection model using the constrained image datasets. This process included techniques of contrast changes of 0.5, contrast changes of 1.5, and vertical flips, applied sequentially (Figure 5). The datasets with these DA images added were named the 50-image dataset, the 75-image dataset, and the 100-image dataset. For a dataset containing only the original images (125-image dataset), *k*-fold cross-validation was performed to assess the validity of the model's performance. *k*-fold cross-validation is a reliable method for evaluating the performance of a model by splitting the data into *k* equally sized subsets. For each of the *k* folds, $k-1$ folds are used to train the model, and the remaining fold is used to test the model. This process was repeated *k* times, and a different fold was used as the test set each time. Finally, average performance metrics were calculated and evaluated. In this context, five-fold cross-validation was performed. The other datasets were trained using each of the five-fold datasets, and the averages were calculated for each dataset (Figure 4a). For the subsequent datasets, to evaluate the performance of the fruit detection model trained using data from each orchard and year, contrast change of 0.5, contrast change of 1.5, and vertically flipped DA images were added to the datasets from each orchard and each year. The data were randomly divided into five parts and five-fold datasets were constructed. For each orchard and year dataset, five-fold cross-validation was performed to confirm the reliability and robustness of the model (Figure 4b). For the third dataset, we prepared one with approximately an equal number of images collected on sunny and cloudy days. Twenty images for Orchard 2 in 2020 collected under sunny conditions were expanded to 100 images by means of DA, with contrast changes of 0.5, contrast changes of 1.5, vertical flips, and horizontal flips. This method is known as oversampling with DA [29]. The data from each orchard and year were randomly divided into 5 parts, which were combined to generate 5-fold datasets, each consisting of 41 images. Five-fold cross-validation was performed to confirm the reliability and robustness of the model (Figure 4c).

Figure 5. The data augmentation implemented for the model training. (**a**) The original image. (**b**) After a contrast change of 0.5. (**c**) After a contrast change of 1.5. (**d**) After a vertical flip.

Table 2 lists the number of images and average number of instances in each dataset. Based on the number of instances, the training/validation ratio was 77.5%:22.5% for the step-by-step DA dataset and 80.0%:20.0% for the other datasets. This ratio reflects the emphasis on the number of UAV images from each orchard (or year) included in each fold.

Based on the report by Yuan [39], the test dataset was prepared, and the performance of the training model was evaluated. In this study, images of different orchards or different years, separate from the training dataset, were selected for the test dataset. Specifically, UAV images and their annotations from Orchard 1 in 2022, Orchard 2 in 2021, and Orchards 5 and 7 in 2023 were used, as shown in Table 1. Each image was cropped to include only the tree region and then pasted onto a black image measuring 2250 × 2250 pixels, as illustrated in Figure 6. These test datasets minimized background complexity, containing only trees and fruits by removing distracting elements such as the ground, other plants, and shadows.

This approach allowed for a more focused evaluation of the model's fruit detection accuracy. The fruits in the test images were annotated using the LabelImg software.

Table 2. Number of images and the average number of instances in each dataset. (a) Constrained image datasets. (b) Biased image datasets. (c) Approximately equal number of images collected on sunny and cloudy days.

(a)

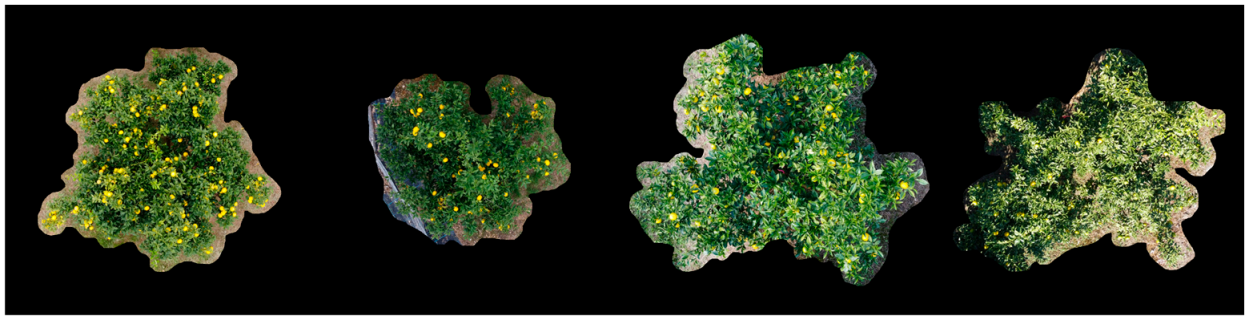
Datasets	Number of Images			Average Number of Instances		
	Train	Validation	Total	Train	Validation	Total
125-Image Dataset	100	25	125	9738	2434	12,172
25-Image Dataset	19	6	25	1887	547	2434
50-Image Dataset	38	12	50	3774	1095	4869
75-Image Dataset	57	18	75	5661	1642	7303
100-Image Dataset	76	24	100	7548	2190	9738

(b)

Datasets	Number of Images			Average Number of Instances		
	Train	Validation	Total	Train	Validation	Total
Orchard 1 in 2020 Dataset	128	32	160	10,819	2705	13,524
Orchard 1 in 2021 Dataset	144	36	180	20,000	5000	25,000
Orchard 2 in 2020 Dataset	64	16	80	3510	878	4388
Orchard 2 in 2022 Dataset	64	16	80	4621	1155	5776

(c)

Datasets	Number of Images			Average Number of Instances		
	Train	Validation	Total	Train	Validation	Total
Approximately Equal Number of Images on Sunny and Cloudy Days	164	41	205	13,248	3312	16,560



(**a**) (**b**) (**c**) (**d**)

Figure 6. The test datasets cropped to include only the tree region and then pasted onto a black background measuring 2250 × 2250 pixels. (**a**) A tree image from Orchard 1 in 2022. (**b**) A tree image from Orchard 2 in 2021. (**c**) A tree image from Orchard 5 in 2023. (**d**) A tree image from Orchard 7 in 2023.

2.3. Execution Environment of Deep Learning

All model training and testing processes were executed on Google Colaboratory (Colab), an online platform provided by Google (Google LLC, Mountain View, CA, USA), using a T4 GPU.

2.4. Overview of YOLOv5u, YOLOv8 and YOLOv9

To compare the fruit detection performance of the various models, each dataset was trained using the YOLOv8m, YOLOv9c, and YOLOv5mu models. Compared to other object detection models such as Faster R-CNN or SSD, YOLO has a faster inference speed and is

capable of real-time detection while it maintains its mean average precision (mAP) [40,41]. This means that it has the potential to be particularly useful in practical applications in the field, as it allows for the efficient detection of high-resolution images such as UAV images. For this reason, the YOLO series model was selected in this study.

YOLOv8 is a kind of YOLO series, released in 2023, and offers an optimal balance between accuracy and speed. For enhanced feature extraction and object detection performance, YOLOv8 introduces state-of-the-art backbone and neck architectures. The adoption of an anchor-free split Ultralytics head enables superior accuracy and efficient detection. YOLOv8 offers various series for different tasks, and certain modes, such as inference, validation, training, and export, accommodate users with different demands, thereby promoting the use of YOLO [25]. YOLOv9 is one of the new YOLO series, released in 2024. YOLOv9 adopts the programmable gradient information (PGI) concept and the advanced architecture of the generalized efficient layer aggregation network (GELAN). The integration of PGI enables the transfer of key information to deeper network layers and improves the accuracy of detection. The GELAN efficiently aggregates and utilizes information from each layer to improve model performance while reducing computational costs. Consequently, high performance can be achieved even with constrained resources. These technologies enhance the accuracy and adaptability of YOLOv9 [26]. YOLOv5u is a modified version of YOLOv5 designed to achieve higher accuracy and performance than the standard YOLOv5 model. It introduces the anchor-free split head employed in the YOLOv8 model, thereby improving the accuracy–speed trade-off in object detection [42]. In this study, YOLOv8m, YOLOv9c, and YOLOv5mu were selected because they have similar parameters for models trained on the Common Objects in Context (COCO) dataset. We deemed it appropriate to compare these YOLO series with constrained or biased image dataset scenarios (Table 2).

2.5. Model Training Execution

Based on the following hyperparameters, training was executed for the YOLOv8m, YOLOv9c, and YOLOv5mu models: 400 epochs, stopping early at 50 epochs, a batch size of 8, an image size of 640 × 640, the AdamW optimizer, a learning rate of 0.002, and a momentum of 0.9 (Table 3).

Table 3. Hyperparameters used for YOLOv8m, YOLOv9c, and YOLOv5mu model training.

Model	YOLOv8m	YOLOv9c	YOLOv5mu
Parameters (Millions) [1]	25.8	25.3	25.0
GFLOPs [1]	78.7	102.3	64.0
Epochs	400	400	400
Early Stopping	50	50	50
Batch Size	8	8	8
Image Size	640 × 640	640 × 640	640 × 640
Optimizer	AdamW	AdamW	AdamW
Learning Rate	0.002	0.002	0.002
Momentum	0.9	0.9	0.9

[1] The performance on the COCO dataset.

2.6. Evaluation Metrics

Several crucial metrics were used to evaluate model performance. These included the precision (*P*), recall (*R*), and F_1-score, which are crucial measures of the accuracy of a model in correctly identifying objects in an image. Additionally, the AP was used to measure the overall performance.

Precision measures the accuracy of positive predictions made by the model. It is defined as the ratio of true positive (*TP*) predictions to all positive predictions (both *TP*s

and false positives (*FP*s)). This metric indicates the ability of the model to identify positive instances correctly.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall measures the model's capability to identify relevant instances and is defined as the ratio of *TP* to the sum of actual positives (*TP*s and false negatives (*FN*s)). A high recall indicates that the model successfully captured most of the actual positive instances.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

The F_1-score is the harmonic mean of the precision and recall, providing a single measure that balances both metrics. The F_1-score is highest at one (perfect precision and recall) and lowest at zero.

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

AP evaluates the trade-off between precision and recall by calculating the area under the precision-recall curve. This metric summarizes the model's performance across all recall levels and provides a comprehensive perspective of its effectiveness; a higher *AP* indicates better overall performance in identifying *TP* while maintaining a lower *FP* rate. In this study, when the overlap between the correct bounding box and the predicted bounding box was 50% or more, it was defined as a correct prediction (IoU (intersection over union) = 50), and the AP at this time was calculated (AP@50).

$$Average\ Precision\ (AP) = \int_0^1 P(R)dR \quad (4)$$

2.7. Assessment of Constructed Models

All the model tests were conducted using the test datasets in Table 1 and Figure 6. The batch size was set to 8 and the image size to 1200 × 1200. Additionally, a confidence score of 0.25 was set as the minimum threshold for detections, and an IoU of 0.5 was set as the IoU threshold for detections during non-maximum suppression (Table 4).

Table 4. Hyperparameters used for YOLOv8m, YOLOv9c, and YOLOv5mu model tests.

Model	YOLOv8m	YOLOv9c	YOLOv5mu
Batch Size	8	8	8
Image Size	1200 × 1200	1200 × 1200	1200 × 1200
Confidence Score	0.25	0.25	0.25
IoU	0.50	0.50	0.50

3. Results and Discussion

3.1. Comparison of Training Models Using Constrained Images Datasets

To evaluate the performance of the three YOLO models trained on datasets with step-by-step image augmentation by means of DA, we used four metrics: precision, recall, F_1-score, and AP@50. Table 5 presents a comparison of the performances of the YOLO model trained on these different training dataset sizes. The models trained on the 125-image dataset containing only the original images yielded the highest values for all metrics across all training models. Among these models, the YOLOv9c model achieved the highest performance, with a precision of 87.7%, recall of 80.0%, F_1-score of 83.6%, and AP@50 of 89.0%. Among the models trained on the 25-image dataset that does not contain DA images, the YOLOv5mu model exhibited the highest values among the three training YOLO models, with a precision of 84.7%, recall of 74.6%, F_1-score of 79.3%, and AP@50 of 85.1%. The models trained on the 50-image dataset, which added contrast change of 0.5 DA images to the 25-image dataset, showed higher values for all evaluation metrics for the three YOLO models compared to the 25-image dataset. Among these, the increase in the

values of each evaluation metric for the YOLOv8m and YOLOv9c models compared to the YOLOv5mu model was higher than the increase from the models trained on the 25-image dataset. The models trained on the 75-image dataset, which added contrast change 1.5 DA images to the 50-image dataset, had the same or lower evaluation metric values for all three YOLO models, except for the recall of the YOLOv8m model. The models trained on the 100-image dataset, which added vertically flipped DA images to the 75-image dataset, had higher evaluation metric values for all three YOLO models except for the precision of the YOLOv8m model. The YOLOv9c model trained on the 100-image dataset had a precision of 86.7%, F_1-score of 81.7%, and AP@50 of 87.4%, which were higher than those for the YOLOv5mu and YOLOv8m models.

Table 5. Comparison of YOLO model performance trained on datasets with step-by-step image augmentation by means of DA.

Model	**Precision**				
	25-Image [1]	**50-Image**	**75-Image**	**100-Image**	**125-Image**
YOLOv8m	78.6% ± 11.0%	86.7% ± 2.3%	85.4% ± 2.0%	85.0% ± 1.6%	87.6% ± 1.4%
YOLOv9c	80.9% ± 17.0%	85.5% ± 3.7%	84.5% ± 2.9%	86.7% ± 1.6%	87.7% ± 1.3%
YOLOv5mu	84.7% ± 1.6%	85.3% ± 1.3%	85.3% ± 2.3%	85.9% ± 1.5%	86.7% ± 1.6%
Model	**Recall**				
	25-Image	**50-Image**	**75-Image**	**100-Image**	**125-Image**
YOLOv8m	72.3% ± 7.4%	73.4% ± 2.2%	73.6% ± 3.5%	76.0% ± 3.5%	78.1% ± 2.0%
YOLOv9c	73.5% ± 5.6%	75.9% ± 2.2%	75.3% ± 4.9%	77.4% ± 4.6%	80.0% ± 1.6%
YOLOv5mu	74.6% ± 4.3%	76.2% ± 4.8%	74.9% ± 3.4%	77.5% ± 5.0%	79.0% ± 0.9%
Model	**F_1-score**				
	25-Image	**50-Image**	**75-Image**	**100-Image**	**125-Image**
YOLOv8m	75.2% ± 8.7%	79.5% ± 2.1%	79.0% ± 2.4%	80.2% ± 2.3%	82.6% ± 1.5%
YOLOv9c	76.7% ± 11.3%	80.4% ± 2.5%	79.6% ± 3.1%	81.7% ± 2.8%	83.6% ± 1.0%
YOLOv5mu	79.3% ± 2.4%	80.4% ± 3.0%	79.7% ± 2.3%	81.4% ± 2.7%	82.7% ± 1.0%
Model	**AP@50**				
	25-Image	**50-Image**	**75-Image**	**100-Image**	**125-Image**
YOLOv8m	80.3% ± 9.3%	84.8% ± 2.3%	84.7% ± 3.1%	85.4% ± 2.4%	88.0% ± 1.1%
YOLOv9c	81.6% ± 12.7%	85.7% ± 2.1%	85.2% ± 2.9%	87.4% ± 2.9%	89.0% ± 0.9%
YOLOv5mu	85.1% ± 2.5%	85.9% ± 3.1%	85.2% ± 2.4%	86.7% ± 3.0%	88.0% ± 1.0%

[1] The values represent the mean ± standard deviation.

Based on these results, the generalization of these training models was tested using four test datasets from different orchards and years. Table 6 presents a comparison of the performance of the training models for the test datasets. For this test, AP@50 was selected as the evaluation metric. The results showed a similar trend for all test datasets. The model trained on the 125-image dataset achieved the highest AP@50 values among all the YOLO models. Among these models, the YOLOv9c model achieved the highest AP@50 values (88.7% for Orchard 1 in 2022, 87.6% for Orchard 2 in 2021, 81.5% for Orchard 5 in 2023, and 84.2% for Orchard 7 in 2023) among all orchards and years. The YOLOv5mu model trained on the 25-image dataset exhibited the highest AP@50 values (86.1% for Orchard 1 in 2022, 84.9% for Orchard 2 in 2021, 77.8% for Orchard 5 in 2023, and 79.5% for Orchard 7 in 2023) among the three YOLO models. The YOLOv8m model trained on the 50-image dataset achieved the highest AP@50 values (86.2% for Orchard 1 in 2022, 79.1% for Orchard 5 in 2023, and 81.3% for Orchard 7 in 2023) among the three YOLO models. The models trained on the 75-image dataset showed stagnation or a decrease in AP@50, except for the YOLOv8m model for Orchard 2 in 2021 and the YOlOv9c and YOLOv5mu models for Orchard 5 in 2023. However, all YOLO models trained on the 100-image dataset demonstrated a higher AP@50 compared to the models trained on the 75-image dataset.

Table 6. Comparison of performance of YOLO models trained on datasets with step-by-step image augmentation by means of DA in test datasets. (a) Results of Orchard 1 in 2022 test dataset. (b) Results of Orchard 2 in 2021 test dataset. (c) Results of Orchard 5 in 2023 test dataset. (d) Results of Orchard 7 in 2023 test dataset.

(a)

Model	AP@50				
	25-Image [1]	50-Image	75-Image	100-Image	125-Image
YOLOv8m	83.9% ± 4.7%	86.2% ± 1.0%	85.8% ± 0.6%	86.8% ± 0.4%	88.5% ± 0.5%
YOLOv9c	84.4% ± 4.9%	86.2% ± 1.0%	85.4% ± 1.5%	86.8% ± 0.8%	88.7% ± 0.3%
YOLOv5mu	86.1% ± 0.8%	86.0% ± 1.0%	85.8% ± 0.6%	87.3% ± 0.3%	88.2% ± 0.7%

(b)

Model	AP@50				
	25-Image	50-Image	75-Image	100-Image	125-Image
YOLOv8m	81.1% ± 8.6%	84.4% ± 0.8%	84.7% ± 1.4%	85.2% ± 0.7%	87.6% ± 0.7%
YOLOv9c	81.8% ± 7.3%	85.1% ± 1.3%	85.0% ± 1.9%	85.6% ± 1.0%	87.6% ± 0.3%
YOLOv5mu	84.9% ± 1.1%	85.1% ± 1.5%	84.5% ± 0.8%	85.8% ± 1.0%	87.4% ± 0.3%

(c)

Model	AP@50				
	25-Image	50-Image	75-Image	100-Image	125-Image
YOLOv8m	71.1% ± 11.6%	79.1% ± 2.4%	77.5% ± 2.4%	80.8% ± 2.4%	80.0% ± 1.0%
YOLOv9c	74.0% ± 7.5%	77.7% ± 1.2%	78.0% ± 3.1%	78.7% ± 1.6%	81.5% ± 1.3%
YOLOv5mu	77.8% ± 3.5%	78.0% ± 3.0%	79.7% ± 2.1%	80.2% ± 2.4%	80.9% ± 2.2%

(d)

Model	AP@50				
	25-Image	50-Image	75-Image	100-Image	125-Image
YOLOv8m	76.5% ± 9.1%	81.3% ± 1.6%	80.2% ± 1.7%	81.9% ± 1.4%	83.2% ± 1.5%
YOLOv9c	77.0% ± 6.1%	80.2% ± 1.8%	79.7% ± 2.3%	80.6% ± 0.8%	84.2% ± 1.2%
YOLOv5mu	79.5% ± 1.8%	80.4% ± 2.0%	80.3% ± 1.5%	81.4% ± 1.2%	82.2% ± 2.8%

[1] The values represent the mean ± standard deviation.

There are several studies on the effect of the type and combination of DA on the object detection performance of training models. Shijie et al. [43] reported the effects of different types and combinations of DA on image classification tasks. They confirmed that cropping, flipping, Wasserstein GAN, and rotation produced better performance improvements than the other augmentation methods, which were more pronounced for small datasets. They also reported that combinations of DA improved or worsened performance in certain cases. Alin et al. [44] compared various types of DA for drone object detection using the YOLOv5 model. Their results indicated that mosaic augmentation achieved the highest precision-recall value of 0.993 compared with the other augmentation types. In addition, Fu et al. [45] reported that the YOLOv5-AT model developed for the detection of green fruits showed a decrease in the mAP of the training model as the number of images in the training dataset decreased, whereas the detection performance of the model improved with the same number of images in the training dataset by incorporating DA images. In our study, the detection performance of the YOLOv8m and YOLOv9c models trained on the 50-image dataset improved significantly compared to the models trained on the 25-image dataset. On the other hand, the results for the models trained on the 50-image dataset and the models trained on the 75-image dataset showed that there was almost no change in detection performance for the same combination of DA methods (contrast change of 0.5 and 1.5). When vertically flipped images, representing a different type of DA method to contrast change, were added, the detection performance improved again for all YOLO models. Therefore, it was considered that combining different types of DA methods would also be

effective for improving the performance of fruit detection from UAV images of *Citrus junos*. In addition, it was revealed that the performance improvement due to DA was high for the YOLOv8m and YOLOv9c models. It was considered that lower detection performance of YOLOv8m and YOLOv9c in the 25-images dataset was due to overfitting caused by the too small amount of training data. Also, the enhancement in detection performance observed in the models was attributed to the increased training data generated through DA, which allowed the models to fully leverage their inherent feature extraction capabilities.

3.2. Comparison of Training Models Using Biased Images Datasets

We compared the performances of the three YOLO models using biased datasets constructed solely from the data for each year and for each orchard. Table 7 lists the model performance metrics based on these biased image datasets. Compared to the models trained on the 125-image dataset, all three YOLO models had equivalent or higher performance when trained on the Orchard 1 in 2020 dataset, while the performance of the models trained on the other datasets was equivalent or lower.

Table 7. Comparison of YOLO models' performance when trained on biased training datasets.

Model	Precision				
	125-Image [1]	**Orchard 1 in 2020**	**Orchard 1 in 2021**	**Orchard 2 in 2020**	**Orchard 2 in 2022**
YOLOv8m	87.6% ± 1.4%	87.7% ± 1.5%	86.1% ± 1.9%	85.1% ± 2.3%	83.5% ± 2.4%
YOLOv9c	87.7% ± 1.3%	89.2% ± 1.1%	87.4% ± 1.4%	85.9% ± 3.1%	88.0% ± 3.7%
YOLOv5mu	86.7% ± 1.6%	87.8% ± 1.5%	86.6% ± 1.5%	87.7% ± 3.5%	86.3% ± 5.3%
Model	**Recall**				
	125-Image	**Orchard 1 in 2020**	**Orchard 1 in 2021**	**Orchard 2 in 2020**	**Orchard 2 in 2022**
YOLOv8m	78.1% ± 2.0%	79.0% ± 1.6%	76.0% ± 2.5%	77.8% ± 2.8%	77.2% ± 5.2%
YOLOv9c	80.0% ± 1.6%	80.6% ± 1.6%	77.7% ± 3.5%	80.1% ± 3.0%	78.4% ± 3.3%
YOLOv5mu	79.0% ± 0.9%	79.8% ± 1.8%	77.3% ± 2.6%	78.5% ± 3.6%	77.8% ± 2.8%
Model	**F_1-score**				
	125-Image	**Orchard 1 in 2020**	**Orchard 1 in 2021**	**Orchard 2 in 2020**	**Orchard 2 in 2022**
YOLOv8m	82.6% ± 1.5%	83.1% ± 1.2%	81.9% ± 3.6%	81.3% ± 2.5%	80.1% ± 3.2%
YOLOv9c	83.6% ± 1.0%	84.7% ± 1.3%	82.2% ± 2.1%	82.9% ± 2.9%	82.9% ± 3.2%
YOLOv5mu	82.7% ± 1.0%	83.6% ± 1.1%	81.7% ± 2.0%	82.8% ± 3.1%	81.8% ± 3.2%
Model	**AP@50**				
	125-Image	**Orchard 1 in 2020**	**Orchard 1 in 2021**	**Orchard 2 in 2020**	**Orchard 2 in 2022**
YOLOv8m	88.0% ± 1.1%	88.2% ± 1.0%	86.6% ± 1.9%	86.1% ± 2.1%	85.9% ± 3.4%
YOLOv9c	89.0% ± 0.9%	89.2% ± 1.0%	87.7% ± 2.1%	87.5% ± 3.1%	87.6% ± 2.8%
YOLOv5mu	88.0% ± 1.0%	88.6% ± 1.5%	87.1% ± 2.1%	87.1% ± 2.9%	87.0% ± 3.7%

[1] The values represent the mean ± standard deviation.

The generalizability of these training models was assessed using four test datasets from various orchards and years. Table 8 presents the comparative performance of these models based on the test datasets, with AP@50 selected as the evaluation metric. Based on the test datasets (taken on cloudy days) for Orchard 1 in 2022 and Orchard 2 in 2021, the model trained on the Orchard 2 in 2020 dataset (taken on sunny days) had a lower AP@50 than the model trained on the other datasets. Based on the Orchard 5 in 2023 test dataset (including under-matured fruit) and the Orchard 7 in 2023 test dataset (including under-matured fruit), all YOLO models except the YOLv5mu model trained on the Orchard 2 in 2020 dataset had an AP@50 that was equal to or lower than the models trained on the 125-image dataset. There was no consistent trend in detection performance among the YOLO models.

Table 8. Comparison of performance of YOLO models trained on biased training datasets based on test datasets. (a) Results of Orchard 1 in 2022 test dataset. (b) Results of Orchard 2 in 2021 test dataset. (c) Results of Orchard 5 in 2023 test dataset. (d) Results of Orchard 7 in 2023 test dataset.

(a)

Model	AP@50				
	125-Image [1]	Orchard 1 in 2020	Orchard 1 in 2021	Orchard 2 in 2020	Orchard 2 in 2022
YOLOv8m	88.5% ± 0.5%	86.8% ± 0.9%	87.6% ± 0.6%	83.3% ± 0.9%	86.6% ± 0.4%
YOLOv9c	88.7% ± 0.3%	87.7% ± 0.8%	87.9% ± 0.5%	84.0% ± 1.2%	87.6% ± 0.6%
YOLOv5mu	88.2% ± 0.7%	86.5% ± 0.8%	87.9% ± 0.5%	84.3% ± 0.7%	87.6% ± 0.6%

(b)

Model	AP@50				
	125-Image	Orchard 1 in 2020	Orchard 1 in 2021	Orchard 2 in 2020	Orchard 2 in 2022
YOLOv8m	87.6% ± 0.7%	85.8% ± 0.9%	86.6% ± 0.4%	82.7% ± 0.5%	84.1% ± 1.0%
YOLOv9c	87.6% ± 0.3%	86.9% ± 1.0%	87.2% ± 0.7%	82.4% ± 2.0%	85.1% ± 0.8%
YOLOv5mu	87.4% ± 0.3%	85.7% ± 1.0%	87.3% ± 0.6%	83.4% ± 0.9%	85.4% ± 0.7%

(c)

Model	AP@50				
	125-Image	Orchard 1 in 2020	Orchard 1 in 2021	Orchard 2 in 2020	Orchard 2 in 2022
YOLOv8m	80.0% ± 1.0%	74.6% ± 1.8%	76.9% ± 1.3%	78.5% ± 2.5%	77.9% ± 1.5%
YOLOv9c	81.5% ± 1.3%	77.4% ± 1.4%	79.6% ± 1.6%	77.8% ± 2.2%	78.2% ± 2.5%
YOLOv5mu	80.9% ± 2.2%	75.5% ± 1.7%	81.0% ± 0.5%	80.8% ± 1.4%	80.1% ± 1.2%

(d)

Model	AP@50				
	125-Image	Orchard 1 in 2020	Orchard 1 in 2021	Orchard 2 in 2020	Orchard 2 in 2022
YOLOv8m	83.2% ± 1.5%	77.0% ± 1.4%	77.4% ± 1.4%	81.3% ± 0.5%	80.0% ± 1.7%
YOLOv9c	84.2% ± 1.2%	76.9% ± 2.1%	77.0% ± 2.0%	82.4% ± 0.8%	78.5% ± 2.2%
YOLOv5mu	82.2% ± 2.8%	75.3% ± 2.3%	79.5% ± 0.6%	83.6% ± 0.3%	78.6% ± 1.2%

[1] The values represent the mean ± standard deviation.

Mirhaji et al. [46] constructed orange fruit detection models for YOLOv2, YOLOv3, and YOLOv4 using images taken on cloudy and sunny days and at nighttime using 72 W LED lights. Xu et al. [47] reported that the detection performance of the fruit detection model for citrus fruits constructed using HPL-YOLOv4, which applies the lightweight feature extraction network GhostNet to YOLOv4, was higher than that of YOLOv3 and YOLOv4, even in cases where there was leaf and branch occlusion, light condition changes, and blurry images. We constructed the training model using a dataset for each orchard with biased light conditions (cloudy or sunny). Although there was a certain level of detection performance for all YOLO models, the detection performance tended to decrease when the light conditions of the training dataset and the test dataset did not match. In addition, the detection performance of all YOLO models was low for the detection dataset that included under-matured fruit, which was not included in the training dataset. These results show the impact of bias in the training dataset, such as the light conditions and the coloring of the fruit, on the performance of the fruit detection model.

3.3. Comparison of Models Trained Using Equal Number of Images Collected on Sunny and Cloudy Days

We compared the performances of the three YOLO models trained on datasets containing an equal number of images collected on sunny and cloudy days. Table 9 presents the model performance metrics for these training datasets. The YOLOv9c and YOLOv5mu models trained on an equal number of images showed higher precision, with increases of 1.0% and 0.9%, respectively, than the models trained on the 125-image dataset. Additionally,

the YOLOv8m model trained on an equal number of images exhibited a 1.0% higher recall than the model trained on the 125-image dataset. However, there were no differences in the F_1-score and AP@50 between the models trained on an equal number of images and those trained on the 125-image dataset.

Table 9. Comparison of performance of YOLO models trained on an equal number of images captured on sunny and cloudy days.

Model	Precision						Recall					
	125-Image [1]			Equal Images			125-Image			Equal Images		
YOLOv8m	87.6%	±	1.4%	86.9%	±	1.5%	78.1%	±	2.0%	79.1%	±	1.6%
YOLOv9c	87.7%	±	1.3%	88.7%	±	0.9%	80.0%	±	1.6%	79.7%	±	1.4%
YOLOv5mu	86.7%	±	1.6%	87.7%	±	2.1%	79.0%	±	0.9%	78.6%	±	0.8%
Model	F_1**-score**						**AP@50**					
	125-Image			**Equal Images**			**125-Image**			**Equal Images**		
YOLOv8m	82.6%	±	1.5%	82.8%	±	1.5%	88.0%	±	1.1%	88.1%	±	1.4%
YOLOv9c	83.6%	±	1.0%	84.0%	±	0.8%	89.0%	±	0.9%	89.1%	±	0.7%
YOLOv5mu	82.7%	±	1.0%	82.9%	±	0.8%	88.0%	±	1.0%	88.3%	±	0.6%

[1] The values represent the mean ± standard deviation.

Table 10 shows the comparison of the performance of YOLO models trained on an equal number of images collected on sunny and cloudy days in the test datasets. For Orchard 5 in the 2023 test dataset, the AP@50 values of the YOLOv8m and YOLOv5mu models trained on a dataset with an equal number of images taken on cloudy and sunny days were 1.2% and 1.6% higher than those of the models trained on the 125-image dataset, respectively. For Orchard 7 in the 2023 test dataset, the AP@50 of the YOLOv9c model trained on a dataset with an equal number of images taken on cloudy and sunny days was 0.8% lower than that of the model trained on the 125-image dataset, whereas that of the YOLOv5mu training model was 1.9% higher.

Table 10. Comparison of YOLO models' performance when trained on equal number of images captured on sunny and cloudy days in test datasets.

Model	Orchard 1 in 2022						Orchard 2 in 2021					
	125-Image [1]			Equal Images			125-Image			Equal Images		
YOLOv8m	88.5%	±	0.5%	87.9%	±	0.9%	87.6%	±	0.7%	87.1%	±	0.9%
YOLOv9c	88.7%	±	0.3%	88.3%	±	0.9%	87.6%	±	0.3%	87.4%	±	0.6%
YOLOv5mu	88.2%	±	0.7%	88.4%	±	0.3%	87.4%	±	0.3%	87.4%	±	0.4%
Model	**Orchard 5 in 2023**						**Orchard 7 in 2023**					
	125-Image			**Equal Images**			**125-Image**			**Equal Images**		
YOLOv8m	80.0%	±	1.0%	81.2%	±	1.5%	83.2%	±	1.5%	83.5%	±	1.8%
YOLOv9c	81.5%	±	1.3%	81.4%	±	1.2%	84.2%	±	1.2%	83.4%	±	1.3%
YOLOv5mu	80.9%	±	2.2%	82.5%	±	0.5%	82.2%	±	2.8%	84.1%	±	1.0%

[1] The values represent the mean ± standard deviation.

Buda et al. [48] reported that class imbalance has a negative impact on classification performance. In our results, even when there was a bias in the light conditions at the time of acquiring the training dataset, the detection performance of the three YOLO models was almost the same as when there was no bias in the training dataset. Based on these results, it was considered that, even if the light conditions of the images included in the training dataset were unbalanced, the effect on detection performance would be small.

4. Conclusions

In this study, we evaluated the performance of three YOLO models using UAV images of yellow maturing citrus fruits. The results indicated that YOLOv5mu exhibited superior detection performance for the constrained image dataset, whereas it was revealed that the performance improvement due to DA was high for the YOLOv8m and YOLOv9c models. Moreover, the impact of bias in the training dataset, such as the light conditions and the coloring of the fruit, on the performance of the fruit detection model was evaluated. In our results, it was considered that, even if the light conditions of the images included in the training dataset were unbalanced, the effect on detection performance would be small.

Our study employed simple DA methods, such as contrast changes and vertical flipping, to evaluate model performance. None of the evaluation metrics reached 90% with their highest value. However, these findings provide critical insights for selecting models based on the quantity and quality of the image data collected under actual field conditions. It is desirable to incorporate advanced augmentation techniques such as Pix2pix, which can introduce more diversity into the dataset and enhance model robustness in the case of actual model training.

Detecting yellow maturing fruits, such as *Citrus junos*, poses a greater challenge than detecting orange maturing fruits due to their visual similarity to yellow leaves and the ground surface. Limited research has been conducted on the detection of these specifically colored fruits. Furthermore, under-matured fruits are difficult to distinguish even with the naked eye because their color is similar to that of branches and leaves. Further improvement is required to resolve these issues. For instance, it is necessary to examine the differences in model performance under specific conditions (direct light or shadow). We believe that this study will contribute significantly to advancing research on the detection of fruits by providing valuable insights and recommendations for model selection and DA strategies.

Author Contributions: Conceptualization, Y.T., Z.Z. and S.Y.; methodology, Y.T.; validation, Y.T.; formal analysis, Y.T.; investigation, Y.T.; resources, Y.T.; data curation, Y.T.; writing—original draft preparation, Y.T.; writing—review and editing, Y.T., Z.Z. and S.Y.; visualization, Y.T.; supervision, S.Y.; project administration, S.Y.; funding acquisition, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a Cabinet Office grant in aid of "Evolution to Society 5.0 Agriculture Driven by IoP (Internet of Plants)", Japan.

Data Availability Statement: Data are contained within this article.

Acknowledgments: We acknowledge the valuable contributions of the staff at the Kochi Agricultural Research Center Fruit Tree Experiment Station.

Conflicts of Interest: The funders had no role in the study design; collection, analyses, or interpretation of data; writing of the manuscript; or decision to publish the results.

References

1. Gonzatto, M.P.; Santos, J.S. *Introductory Chapter: World Citrus Production and Research, Citrus Research—Horticultural and Human Health Aspects*; InTech Open: Rijeka, Croatia, 2023. Available online: https://www.intechopen.com/chapters/86388 (accessed on 2 November 2024).
2. Morinaga, K.; Sumikawa, O.; Kawamoto, O.; Yoshikawa, H.; Nakao, S.; Shimazaki, M.; Kusaba, S.; Hoshi, N. New Technologies and Systems for High Quality Citrus Fruit Production, Labor-Saving and Orchard Construction in Mountain Areas of Japan. *J. Mt. Sci.* **2005**, *2*, 59–67. [CrossRef]
3. United States Department of Agriculture. *Evaluation of Procedures for Estimating Citrus Fruit Yield*; United States Department of Agriculture: Washington, DC, USA, 1972. Available online: https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/Evaluation%20of%20Procedures%20for%20Estimating%20Citrus%20Fruit%20Yield.pdf (accessed on 2 November 2024).
4. Stout, R.G. Estimating Citrus Production by Use of Frame Count Survey. *J. Farm Econ.* **1962**, *44*, 1037–1049. [CrossRef]
5. United States Department of Agriculture. *Sampling for Objective Yields of Apples and Peaches*; United States Department of Agriculture: Washington, DC, USA, 1967. Available online: https://www.nass.usda.gov/Education_and_Outreach/

Reports,_Presentations_and_Conferences/Yield_Reports/Sampling%20for%20Objective%20Yields%20of%20Apples%20and%20Oranges.pdf (accessed on 2 November 2024).
6. Wulfsohn, D.; Aravena Zamora, F.; Potin, T.C.; Zamora, L.I.; García-Fiñana, M. Multilevel Systematic Sampling to Estimate Total Fruit Number for Yield Forecasts. *Precis. Agric.* **2012**, *13*, 256–275. [CrossRef]
7. Zhang, W.; Wang, J.; Liu, Y.; Chen, K.; Li, H.; Duan, Y.; Wu, W.; Shi, Y.; Guo, W. Deep-Learning-Based in-Field Citrus Fruit Detection and Tracking. *Hortic. Res.* **2022**, *9*, uhac003. [CrossRef] [PubMed]
8. Li, Y.; Gong, Z.; Zhou, Y.; He, Y.; Huang, R. Production Evaluation of Citrus Fruits Based on the YOLOv5 Compressed by Knowledge Distillation. In Proceedings of the 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Rio de Janeiro, Brazil, 24–26 May 2023; pp. 1938–1943.
9. Gremes, M.F.; Fermo, I.R.; Krummenauer, R.; Flores, F.C.; Andrade, C.M.G.; Lima, O.C.d.M. System of Counting Green Oranges Directly from Trees Using Artificial Intelligence. *AgriEngineering* **2023**, *5*, 1813–1831. [CrossRef]
10. Jing, J.; Zhai, M.; Dou, S.; Wang, L.; Lou, B.; Yan, J.; Yuan, S. Optimizing the YOLOv7-Tiny Model with Multiple Strategies for Citrus Fruit Yield Estimation in Complex Scenarios. *Agriculture* **2024**, *14*, 303. [CrossRef]
11. Gao, A.; Tian, Z.; Ma, W.; Song, Y.; Ren, L.; Feng, Y.; Qian, J.; Xu, L. Fruits Hidden by Green: An Improved YOLOV8n for Detection of Young Citrus in Lush Citrus Trees. *Front. Plant Sci.* **2024**, *15*, 1375118.
12. Apolo-Apolo, O.E.; Martínez-Guanter, J.; Egea, G.; Raja, P.; Pérez-Ruiz, M. Deep Learning Techniques for Estimation of the Yield and Size of Citrus Fruits Using a UAV. *Eur. J. Agron.* **2020**, *115*, 126030. [CrossRef]
13. Novelero, J.M.; Cruz, J.C.D. On-Tree Mature Coconut Fruit Detection Based on Deep Learning Using UAV Images. In Proceedings of the 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia, 16–18 June 2022; pp. 494–499.
14. Xiong, Z.; Wang, L.; Zhao, Y.; Lan, Y. Precision Detection of Dense Litchi Fruit in UAV Images Based on Improved YOLOv5 Model. *Remote Sens.* **2023**, *15*, 4017. [CrossRef]
15. Wang, H.; Feng, J.; Yin, H. Improved Method for Apple Fruit Target Detection Based on YOLOv5s. *Agriculture* **2023**, *13*, 2167. [CrossRef]
16. Arakawa, T.; Tanaka, T.S.T.; Kamio, S. Detection of On-Tree Chestnut Fruits Using Deep Learning and RGB Unmanned Aerial Vehicle Imagery for Estimation of Yield and Fruit Load. *Agron. J.* **2023**, *116*, 973–981. [CrossRef]
17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
20. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
21. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
22. Jocher, G.; Stoken, A.; Borovec, J.; Changyu, L.; Hogan, A.; Diaconu, L.; Rai, P. *Ultralytics/yolov5: v3.1—Bug Fixes and Performance Improvements, Version 3.1*; Zenodo: Geneva, Switzerland, 2020. Available online: https://zenodo.org/records/4154370 (accessed on 2 November 2024).
23. Li, C.; Li, L.; Geng, Y.; Jiang, H.; Cheng, M.; Zhang, B.; Ke, Z.; Xu, X.; Chu, X. YOLOv6 v3.0: A Full-Scale Reloading. *arXiv* **2023**, arXiv:2301.05586. [CrossRef]
24. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
25. Jocher, G.; Chaurasia, A.; Qiu, J. *YOLO by Ultralytics, version 8.0.0*; Ultralytics: Frederick, MD, USA, 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 2 November 2024).
26. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616. [CrossRef]
27. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458. [CrossRef]
28. Jocher, G.; Qiu, J. Ultralytics YOLO11. Available online: https://github.com/ultralytics/ultralytics (accessed on 2 November 2024).
29. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
30. Montserrat, D.M.; Lin, Q.; Allebach, J.; Delp, E.J. Training Object Detection and Recognition CNN Models Using Data Augmentation. *Electron. Imaging* **2017**, *29*, 27–36. [CrossRef]
31. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
32. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008.

33. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
34. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
35. Webber, H.J. *The Citrus Industry: History, World Distribution, Botany, and Varieties*, 2nd ed.; University of California Press: Berkeley, CA, USA, 1967; pp. 389–390.
36. Iwamasa, M. Citrus Cultivars in Japan. *HortScience* **1988**, *23*, 687–690. [CrossRef]
37. Lan-Phi, N.T.; Shimamura, T.; Ukeda, H.; Sawamura, M. Chemical and Aroma Profiles of Yuzu (*Citrus junos*) Peel Oils of Different Cultivars. *Food Chem.* **2009**, *115*, 1042–1047. [CrossRef]
38. Tanimoto, Y.; Yoshida, S. A Method of Constructing Models for Estimating Proportions of Citrus Fruit Size Grade Using Polynomial Regression. *Agronomy* **2024**, *14*, 174. [CrossRef]
39. Yuan, W. Accuracy Comparison of YOLOv7 and YOLOv4 Regarding Image Annotation Quality for Apple Flower Bud Classification. *AgriEngineering* **2023**, *5*, 413–424. [CrossRef]
40. Vilcapoma, P.; Parra Meléndez, D.; Fernández, A.; Vásconez, I.N.; Hillmann, N.C.; Gatica, G.; Vásconez, J.P. Comparison of Faster R-CNN, YOLO, and SSD for Third Molar Angle Detection in Dental Panoramic X-Rays. *Sensors* **2024**, *24*, 6053. [CrossRef]
41. Sarma, K.S.R.K.; Sasikala, C.; Surendra, K.; Erukala, S.; Aruna, S.L. A comparative study on faster R-CNN, YOLO and SSD object detection algorithms on HIDS system. *AIP Conf. Proc.* **2024**, *2971*, 060044.
42. Jocher, G. Ultralytics YOLOv5. 2020. Available online: https://github.com/ultralytics/yolov5 (accessed on 2 November 2024).
43. Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on Data Augmentation for Image Classification Based on Convolution Neural Networks. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170.
44. Alin, A.Y.; Yuana, K.A. Data Augmentation Method on Drone Object Detection with YOLOv5 Algorithm. In Proceedings of the 2023 Eighth International Conference on Informatics and Computing (ICIC), Manado, Indonesia, 8–9 December 2023; pp. 1–6.
45. Fu, X.; Zhao, S.; Wang, C.; Tang, X.; Tao, D.; Li, G.; Jiao, L.; Dong, D. Green Fruit Detection with a Small Dataset under a Similar Color Background Based on the Improved YOLOv5-AT. *Foods* **2024**, *13*, 1060. [CrossRef]
46. Mirhaji, H.; Soleymani, M.; Asakereh, A.; Mehdizadeh, S.A. Fruit Detection and Load Estimation of an Orange Orchard Using the YOLO Models Through Simple Approaches in Different Imaging and Illumination Conditions. *Comput. Electron. Agric.* **2021**, *191*, 106533. [CrossRef]
47. Xu, L.; Wang, Y.; Shi, X.; Tang, Z.; Chen, X.; Wang, Y.; Zou, Z.; Huang, P.; Liu, B.; Yang, N.; et al. Real-Time and Accurate Detection of Citrus in Complex Scenes Based on HPL-YOLOv4. *Comput. Electron. Agric.* **2023**, *205*, 107590. [CrossRef]
48. Buda, M.; Maki, A.; Mazurowski, M.A. A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef]