AgriEngineering

Article

Accuracy Comparison of YOLOv7 and YOLOv4 Regarding Image Annotation Quality for Apple Flower Bud Classification

Wenan Yuan

Independent Researcher, Burr Ridge, IL 60527, USA; wenan.yuan@huskers.unl.edu

Abstract: Object detection is one of the most promising research topics currently, whose application in agriculture, however, can be challenged by the difficulty of annotating complex and crowded scenes. This study presents a brief performance assessment of YOLOv7, the state-of-the-art object detector, in comparison to YOLOv4 for apple flower bud classification using datasets with artificially manipulated image annotation qualities from 100% to 5%. Seven YOLOv7 models were developed and compared to corresponding YOLOv4 models in terms of average precisions (APs) of four apple flower bud growth stages and mean APs (mAPs). Based on the same test dataset, YOLOv7 outperformed YOLOv4 for all growth stages at all training image annotation quality levels. A 0.80 mAP was achieved by YOLOv7 with 100% training image annotation quality, meanwhile a 0.63 mAP was achieved with only 5% training image annotation quality. YOLOv7 improved YOLOv4 APs by 1.52% to 166.48% and mAPs by 3.43% to 53.45%, depending on the apple flower bud growth stage and training image annotation quality. Fewer training instances were required by YOLOv7 than YOLOv4 to achieve the same levels of classification accuracies. The most YOLOv7 AP increase was observed in the training instance number range of roughly 0 to 2000. It was concluded that YOLOv7 is undoubtedly a superior apple flower bud classifier than YOLOv4, especially when training image annotation quality is suboptimal.

Keywords: agriculture; AP; growth stage; mAP; object detection; training instance number

check for updates

Citation: Yuan, W. Accuracy Comparison of YOLOv7 and YOLOv4 Regarding Image Annotation Quality for Apple Flower Bud Classification. *AgriEngineering* **2023**, *5*, 413–424. https://doi.org/10.3390/agriengineering5010027

Academic Editor: Murali Krishna Gumma

Received: 16 November 2022
Revised: 13 February 2023
Accepted: 14 February 2023
Published: 20 February 2023

1. Introduction

Object detection is the computer vision task of identifying and locating target object instances in digital images. Since the invention of AlexNet for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [1], convolutional neural networks (CNNs) have attracted much research attention and become the dominant mechanism for analyzing image spatial patterns in modern object detectors. CNNs are a type of feedforward neural network that are able to learn deep image features through sliding kernels performing element-wise product at multiple convolutional layers along with pooling layers and fully connected layers [2]. Object detectors usually highlight individual object instances using rectangular bounding boxes, while a further extension of object detectors, instance segmenters, mark individual object instances more precisely using pixel-wise masks [3,4]. Such functions of CNN-based algorithms allow the types, counts, and densities of objects in images to be conveniently estimated [5,6], which can be rather challenging to traditional rule-based image processing techniques.

With the rapid advancement of sensing technologies in recent years, optical instruments such as cameras are being extensively implemented in a broad spectrum of agricultural applications [7], from fruit grading, counting, and yield estimation to weed, insect, and disease detection [8]. As a result, algorithms such as object detectors that are able to extract valuable high-level information from complex image data have significant implications for agricultural research and farming [9,10]. Being widely known and adopted in computer vision systems [11,12], each generation of the you look only once (YOLO) series arguably represented the state of the art of one-stage object detectors at the time of releasing. In contrast to two-stage detectors such as the region-based CNN (R-CNN) series, one-stage detectors combine region detection and object classification in straightforward architectures to achieve higher inference speeds [13], and much success has been achieved over the past decade by such algorithms [14]. Since the creation of the original YOLO detector by Redmon et al. [15] in 2015, researchers have continuously contributed to the expansion of the YOLO family, which now includes YOLO [15], YOLOv2 [16], YOLOv3 [17], YOLOv4 [18], YOLOv5 [19], PP-YOLO [20], Scaled-YOLOv4 [21], PP-YOLOv2 [22], YOLOR [23], YOLOX [24], YOLOv6 [25], YOLOv7 [26], etc. As one of the newer members to the YOLO family, YOLOv7 was reported to outperform its predecessors in terms of both inference time and accuracy [26]. YOLOv7′s architecture features a proposed extended efficient layer aggregation network (E-ELAN) backbone for continuously enhanced image feature learning through a expand, shuffle, and merge cardinality manner, a compound model scaling approach to optimize network architecture search (NAS) for computing device fitting, a planned re-parameterized convolution module for detector performance improvement without inference-time cost, and finally an auxiliary head for model training assistance and a lead head for final classification output [26].

As modern object detectors are generally based on supervised learning, the quality of ground truth, or manual image annotations of target objects, is immensely important [27]. Existing research has reported that improved training image annotation quality can increase object detector performance and alter the difficulty of object detection tasks [28]. While image annotation preparation for large datasets is always time-consuming and tedious, it is not necessarily difficult for individual images containing only a few target objects. However, in the context of agriculture, crops such as fruit trees can have complex structures and densely distributed organs such as stems, leaves, buds, flowers, and fruits that are challenging to identify in images even for humans. When such objects need to be annotated, significant manpower will be required for timely training dataset preparation. Yet, oftentimes, the ambiguity in plant characteristics, the subjectivity in human judgement, the inconsistency in annotation style, and the incompleteness in object labelling can lead to low-quality training image annotations, and thus, reduced object detector performances [29].

Evaluating object detectors regarding image annotation quality with a focus on agriculture is a niche and underexplored research topic. For example, the Microsoft COCO dataset, one of the gold standard benchmarks widely employed by researchers to assess state-of-the-art computer vision models, does not have many agriculture-related object classes and contains on average only 7.7 object instances per image [30]. Consequently, it is generally unknown how a newly invented object detector would perform against complicated agricultural datasets, which may contain hundreds of object instances per image with imperfect image annotations. Moreover, current object detectors are being updated at an increasingly higher frequency than they have ever been, yet it is unclear whether the algorithmic advancements are meaningful to agricultural research. For example, there was only a two-month interval between the releases of YOLOv6 [25] and YOLOv7 [26] in 2022. Meanwhile, YOLOv4 [18], which was released in 2020 and reported to underperform YOLOv7, is still being actively adopted in various disciplines and has been cited more than 4000 times in 2022 alone according to Google Scholar. Constant, timely evaluations of state-of-the-art object detectors using a dedicated agricultural benchmark can provide insights to researchers on whether a newly developed algorithm has only a marginal or substantial performance improvement compared to the existing ones, and whether it is worthwhile to upgrade outdated detectors for more accurate plant target location and classification.

In the preceding works to the current study, YOLOv4 was utilized to classify apple flower bud growth stages [29] and examined in detail regarding its capacity to resist test image distortion and low training image annotation quality in terms of completeness [31]. It was discovered that training instance number was a critical factor that affected YOLOv4 accuracy, whose minimum number for optimal classification results was in the rough range of 3000 to 4000 for individual object classes. Under the same total training instance number, YOLOv4 models trained with fewer images but higher image annotation qualities outperformed those trained with more images but lower image annotation qualities, indicating the importance of training image annotation quality over training dataset size. On the basis of the previous works, the current study focused on, practically, how a state-of-the-art object detector compares to and has evolved from an antiquated predecessor in terms of apple flower bud detection accuracy and ideal training image annotation workload. Using YOLOv7 as a representation of state-of-the-art object detectors and YOLOv4 as a baseline reference, the objectives of the study included: (1) investigating the accuracy improvement of YOLOv7 at various training image annotation quality levels; (2) determining the optimal training instance number of YOLOv7.

2. Materials and Methods

2.1. Apple Flower Bud Image Dataset

The image dataset used in [31] was employed in the current study. A drone-based red–green–blue (RGB) camera with a 90° pitch and a 1920 × 1080 resolution was used to collect the images over an apple orchard (40°42′28.5″ N, 77°57′15.7″ W) on five dates. The four-row orchard had two apple varieties, including Jonagold and Daybreak Fuji, with 16 trees per row and two rows per variety. Mainly four apple flower bud growth stages were captured by the images collected on the first four dates from April to May, 2020, namely tight cluster, pink, bloom, and petal fall. On the last date in September, 2020, images of apple tree canopies and ground containing no flower buds were collected, serving as "negative samples" during model training. Although it has been concluded previously that negative samples do not affect YOLOv4 performance [31], negative samples were still utilized in the current study so that YOLOv7 and YOLOv4 models were developed using identical datasets. The image dataset in total consisted of 3060 images, with 450 images from each of the first four dates and 1260 from the last date. A tiny portion of the dataset contained motion blur due to either drone or tree movement during data collection. However, most apple flower buds could still be clearly identified in such images. The image annotations were prepared by four trained human annotators and the author and double-checked by the author to ensure annotation completeness, correctness, and style consistency. The annotation rules were defined in [29] for ambiguous and difficult annotation scenarios.

2.2. YOLOv7 Model Development

Following a commonly adopted data split rule in machine learning research [32,33], the images from each of the first four dates were split into 70%, 20%, and 10% segments for model training, model validation during training, and model independent test respectively. All negative samples from the last date were only used for model training. The training, validation, and test datasets totally contained 2520, 360, and 180 images respectively. To simulate low image annotation quality, training image annotations were artificially manipulated by randomly removing a percentage of annotations from each image. Seven training datasets were created accordingly, which had identical images but different levels of image annotation qualities, including 100%, 90%, 70%, 50%, 20%, 10%, and 5%. The quality levels were chosen to be consistent with the ones adopted in the previous study [31] for direct and convenient comparison between YOLOv4 and YOLOv7 models. Figure 1 shows a flowchart of the dataset preparation process, and examples of various training image annotation quality levels can be found in Figure 2.

Using the seven training datasets and the same validation and test datasets, seven YOLOv7 models were developed using CoLaboratory (Google LLC, Mountain View, CA, USA). As mentioned above, previous YOLOv4 models were also trained, validated, and tested using the identical datasets [31]. A few customized YOLOv7 model hyperparameters included a batch size of 32, an epoch number of 3000, which was much more than necessary to ensure sufficient model training, and an image size of 480, which was consistent with previous YOLOv4 model development [31]. All other hyperparameters were kept to be the default values as provided by the authors of YOLOv7 [26]. Model performances were evaluated based on average precisions (APs) and mean APs (mAPs) at 50% intersection over union (IoU) [34]. The APs and mAPs of YOLOv7 and YOLOv4 models were compared based on relative change (RC) in terms of percentage.

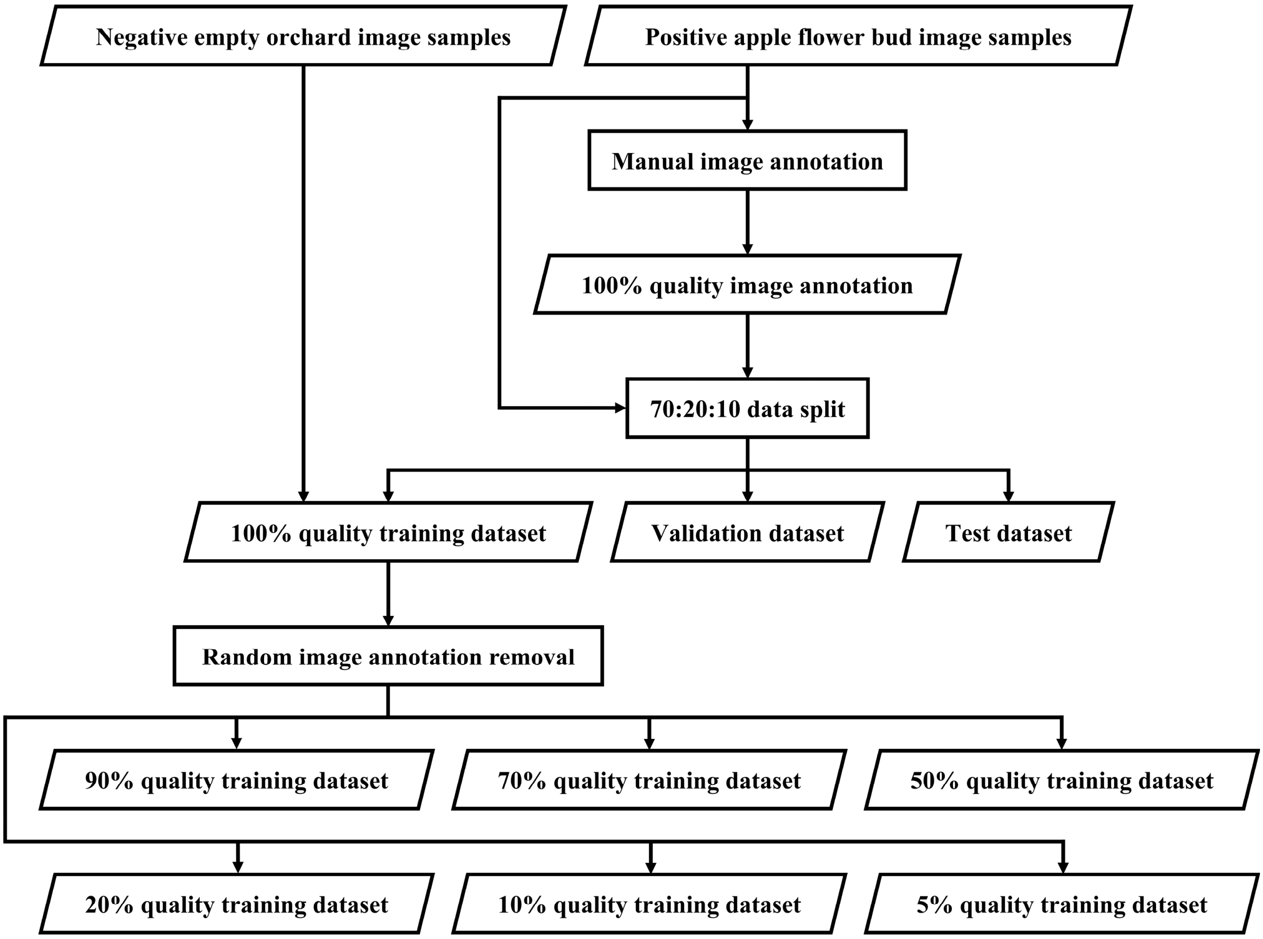
Negative empty orchard image samples
Positive apple flower bud image samples
Manual image annotation
100% quality image annotation
70:20:10 data split
100% quality training dataset
Validation dataset
Test dataset
Random image annotation removal
90% quality training dataset
70% quality training dataset
50% quality training dataset
20% quality training dataset
10% quality training dataset
5% quality training dataset

Figure 1. Flowchart of the preparation process for the seven training datasets with different image annotation qualities, one common validation dataset, and one common test dataset.

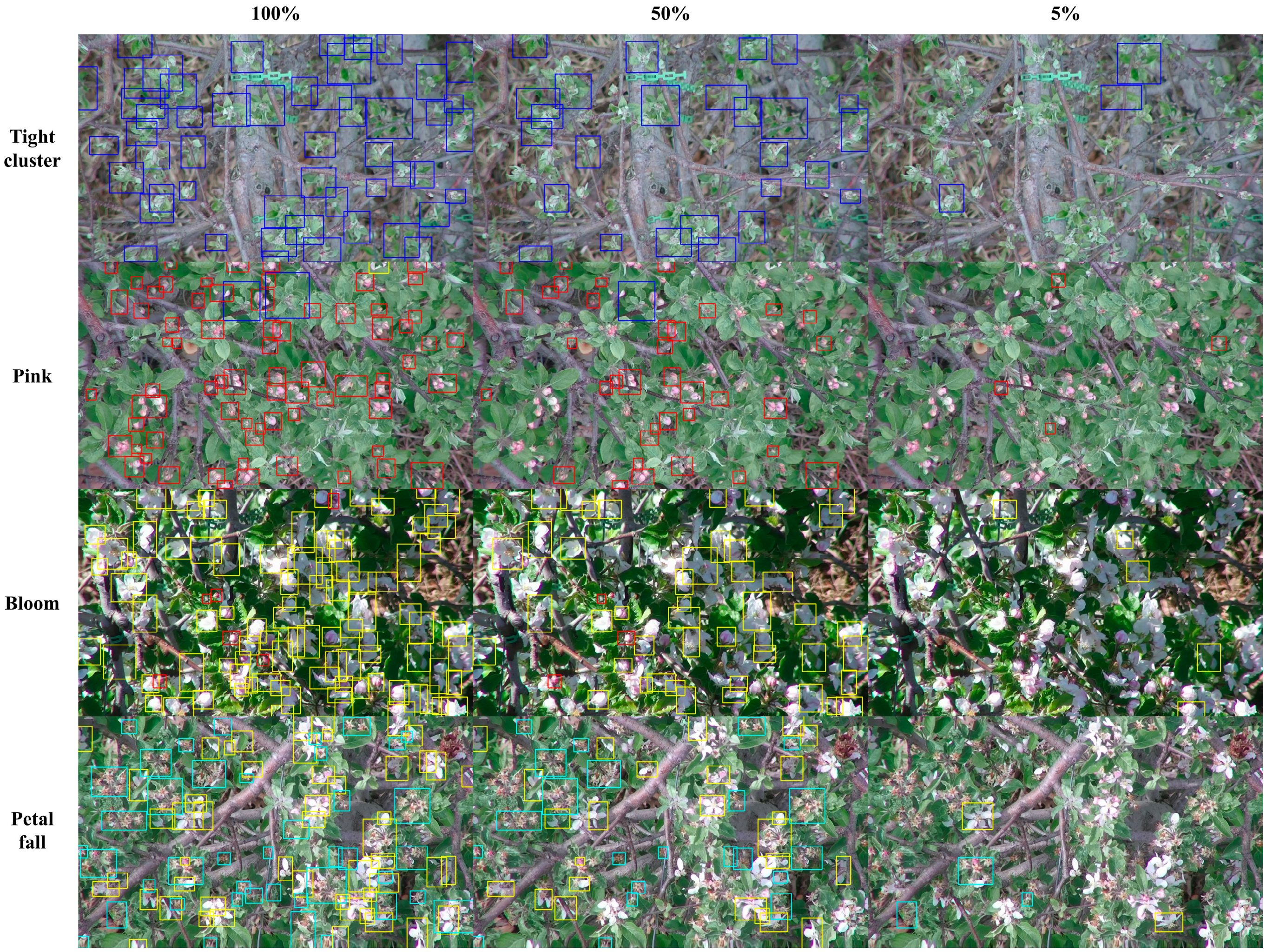

Figure 2. Sample training images containing apple flower buds at four growth stages with different levels of image annotation qualities.

3. Results and Discussion

3.1. Training and Validation Datasets

For model evaluation purposes, training and validation datasets are not as objective and appropriate as test datasets. Models can be overfitted to training datasets to achieve high accuracies, while validation datasets can be biased towards superior model performance results since validation images are also exposed to models during training. For this reason, YOLOv7 performances on the training and validation datasets are not discussed in detail. Nonetheless, the information is provided in this section for reference.

Table 1 shows the APs for individual apple flower bud growth stages, as well as mAPs of the YOLOv4 and YOLOv7 models trained with different image annotation qualities according to the training datasets. The APs of both YOLOv4 and YOLOv7 decreased for all growth stages as training image annotation quality decreased, which was the result of both lower model performance and lower ground truth quality. In other words, failing to recognize a flower bud that is correctly annotated and successfully recognizing a flower bud that is not annotated can both lead to reduced classification accuracies. However, as is shown below, training image annotation quality was the major reason for the extremely low AP and mAP values in Table 1, as the same models were able to achieve much better results for the validation and test datasets. Interestingly, when training image annotation quality was low, YOLOv7 was able to achieve much higher APs or follow “false” ground truths much better than YOLOv4. Using the most extreme case as an example, the AP of YOLOv7 for petal fall at 5% training image annotation quality was 609% higher than that of YOLOv4. Yet, such an AP improvement was not because YOLOv7 learned random patterns unique to the training datasets or YOLOv7 was overfitted to the training datasets, since YOLOv7 still outperformed YOLOv4 for the validation and test datasets as shown below. Additionally, YOLOv7 had greater potential to improve the weaker YOLOv4 APs and mAPs, or less potential to improve the stronger YOLOv4 APs and mAPs. For example, when training image annotation quality ranged from 100% to 70%, where YOLOv4 also had its best performances, YOLOv7 either had minimal performance improvements or performance decreases compared to YOLOv4. YOLO7 was able to improve the APs of petal fall the most and tight cluster the least, which were also in general the least and most accurate growth stages for YOLOv4 respectively.

Table 2 shows the APs and mAPs of YOLOv4 and YOLOv7 according to the validation dataset. Similarly, YOLOv4 and YOLOv7 generally had lower APs and mAPs as training image annotation quality decreased, which indicated poorer model performances. For tight cluster, pink, bloom, and petal fall, YOLOv7 achieved 0.75 to 0.92, 0.60 to 0.77, 0.72 to 0.87, and 0.45 to 0.64 AP ranges respectively. No extremely low AP and mAP values were observed for either YOLOv4 or YOLOv7, as the image annotation quality of the validation dataset was 100%. YOLOv7 consistently outperformed YOLOv4 for all growth stages across all training image annotation quality levels. Again, as YOLOv4 APs and mAPs decreased, the RCs between YOLOv7 and YOLOv4 generally increased. YOLOv7 demonstrated significantly more robustness against low ground truth quality than YOLOv4 by improving the APs from 1.23% to 184.45% and mAPs from 2.36% to 55.48%, although the improvement rates were lower than what were achieved for the training datasets. As is shown in the next section, the results obtained from the validation dataset were very similar to the test dataset results.

Table 1. YOLOv4 and YOLOv7 model performance comparison based on the training datasets.

Training Image Annotation Quality			100%	90%	70%	50%	20%	10%	5%
YOLOv4	AP	Tight cluster	0.939	0.866	0.692	0.525	0.166	0.085	0.043
		Pink	0.855	0.776	0.621	0.455	0.128	0.067	0.030
		Bloom	0.923	0.844	0.671	0.501	0.161	0.079	0.041
		Petal fall	0.790	0.713	0.565	0.422	0.085	0.037	0.017
	mAP		0.877	0.800	0.637	0.476	0.135	0.067	0.032
YOLOv7	AP	Tight cluster	0.946	0.858	0.699	0.552	0.252	0.114	0.106
		Pink	0.850	0.758	0.614	0.486	0.207	0.102	0.078
		Bloom	0.914	0.830	0.672	0.509	0.243	0.112	0.090
		Petal fall	0.773	0.659	0.569	0.447	0.193	0.087	0.117
	mAP		0.871	0.776	0.638	0.498	0.224	0.104	0.098
RC (%)	AP	Tight cluster	0.745	−0.924	0.953	5.203	51.716	34.118	146.512
		Pink	−0.585	−2.370	−1.111	6.790	61.215	52.012	164.865
		Bloom	−0.954	−1.635	0.194	1.698	50.932	41.058	122.963
		Petal fall	−2.090	−7.535	0.708	6.050	127.059	133.602	609.091
	mAP		−0.639	−2.964	0.110	4.732	65.803	54.762	202.160

Table 2. YOLOv4 and YOLOv7 model performance comparison based on the validation dataset.

Training Image Annotation Quality			**100%**	**90%**	**70%**	**50%**	**20%**	**10%**	**5%**
YOLOv4	AP	Tight cluster	0.885	0.875	0.850	0.831	0.727	0.611	0.497
		Pink	0.753	0.748	0.726	0.695	0.554	0.500	0.374
		Bloom	0.854	0.846	0.831	0.797	0.734	0.669	0.586
		Petal fall	0.626	0.622	0.594	0.547	0.362	0.291	0.158
	mAP		0.780	0.773	0.750	0.718	0.594	0.518	0.404
YOLOv7	AP	Tight cluster	0.915	0.912	0.892	0.879	0.862	0.813	0.748
		Pink	0.773	0.766	0.735	0.732	0.699	0.651	0.597
		Bloom	0.866	0.867	0.851	0.836	0.817	0.773	0.716
		Petal fall	0.638	0.635	0.647	0.592	0.559	0.475	0.450
	mAP		0.798	0.795	0.781	0.760	0.734	0.678	0.628
RC (%)	AP	Tight cluster	3.343	4.193	5.003	5.751	18.537	33.061	50.473
		Pink	2.629	2.420	1.226	5.324	26.287	30.148	59.540
		Bloom	1.405	2.482	2.456	4.920	11.278	15.477	22.122
		Petal fall	1.998	2.156	8.959	8.148	54.377	63.230	184.450
	mAP		2.360	2.886	4.133	5.909	23.527	30.913	55.484

3.2. Test Dataset

Table 3 shows the performance comparison between YOLOv4 and YOLOv7 according to the test dataset, which is in general agreement with the observations from the training and validation datasets. YOLOv7 achieved AP ranges of 0.73 to 0.90 for tight cluster, 0.60 to 0.79 for pink, 0.70 to 0.87 for bloom, and 0.47 to 0.66 for petal fall. Overall, YOLOv7 showed exceptional robustness against poor training image annotation quality, achieving a 0.80 mAP at 100% annotation quality, a 0.76 mAP at 50% annotation quality, and a 0.63 mAP at 5% annotation quality. By missing annotating 95% of all apple flower buds in the training images, YOLOv7 performance only experienced a 0.17 mAP decrease for the test images.

Again, both YOLOv4 and YOLOv7 generally performed worse when training image annotation quality decreased, which was due to both the reduced training instance numbers and lowered ground truth quality for the models to learn from [31]. With no exception, YOLOv7 outperformed YOLOv4 for every single growth stage at every single training image annotation quality level. When training image annotation quality was equal to or above 50%, relatively similar results were obtained by YOLOv4 and YOLOv7, with a 1.52% to 12.46% AP improvement range and a 3.16% to 6.52% mAP improvement range. When training image annotation quality was below 50%, YOLOv7 had significantly better performances than YOLOv4, with a 12.68% to 166.48% AP improvement range and a 23.65% to 53.45% mAP improvement range.

There was a general positive correlation between the performances of YOLOv4 and YOLOv7, whose accuracy changes were mostly consistent as growth stage and training image annotation quality varied. There was also a general negative correlation between the APs and mAPs of YOLOv4 and the corresponding RC values. That is, the lower the YOLOv4 classification accuracies were, the more YOLOv7 was able to improve upon, as discussed previously. For example, as training image annotation quality decreased, YOLOv4 performance also generally decreased, while the RC values between YOLOv7 and YOLOv4 generally increased. Tight cluster and bloom were the two most accurate growth stages for YOLOv4, which also showed the least improvements by YOLOv7. Petal fall and pink were the first and second least accurate growth stages for YOLOv4 respectively, yet they also had the highest RC values especially when training image annotation quality was low.

Table 3. YOLOv4 and YOLOv7 model performance comparison based on the test dataset.

Training Image Annotation Quality			100%	90%	70%	50%	20%	10%	5%
YOLOv4	AP	Tight cluster	0.886	0.872	0.852	0.827	0.710	0.643	0.507
		Pink	0.727	0.729	0.722	0.691	0.542	0.493	0.364
		Bloom	0.852	0.847	0.833	0.800	0.731	0.664	0.582
		Petal fall	0.625	0.626	0.590	0.529	0.391	0.325	0.176
	mAP		0.773	0.769	0.749	0.712	0.594	0.531	0.407
YOLOv7	AP	Tight cluster	0.904	0.899	0.880	0.865	0.836	0.783	0.734
		Pink	0.789	0.769	0.742	0.738	0.694	0.660	0.598
		Bloom	0.868	0.868	0.852	0.832	0.824	0.773	0.701
		Petal fall	0.634	0.655	0.617	0.595	0.580	0.472	0.469
	mAP		0.799	0.798	0.773	0.758	0.734	0.672	0.625
RC (%)	AP	Tight cluster	1.997	3.096	3.250	4.608	17.829	21.754	44.688
		Pink	8.543	5.444	2.813	6.848	27.950	33.874	64.150
		Bloom	1.842	2.443	2.281	4.013	12.676	16.468	20.509
		Petal fall	1.521	4.616	4.576	12.455	48.338	45.231	166.477
	mAP		3.430	3.812	3.163	6.521	23.652	26.506	53.450

A visualized comparison between YOLOv4 and YOLOv7 can be found in Figure 3, where the least accurate YOLOv4 and YOLOv7 models, trained with 5% image annotation quality, were used to detect apple flower buds on randomly selected test images using a 3% confidence threshold. Expectedly, as total training instance number decreases, overall model prediction confidence would also decrease, hence a low confidence threshold needed to be used to show the detection results. In the images, YOLOv7 was able to detect substantially more apple flower buds than YOLOv4 for all four growth stages. Particularly, YOLOv4 detected only a few pinks and no petal falls in the images with over 3% confidence, in contrast to YOLOv7′s ground truth-like detection results. This performance difference is also reflected in Table 3, where YOLOv7 improved YOLOv4 APs of pink and petal fall at 5% training image annotation quality by 64.15% and 166.48% respectively.

3.3. Optimal Training Instance Number

Figure 4 shows the relationship between training instance number and AP of the seven YOLOv4 models and seven YOLOv7 models at individual growth stages based on the test dataset. General power laws existed for both YOLOv4 and YOLOv7, meaning the more annotated training instances there were, the better model performances would be, and the slower model performance improvement would be. Although YOLOv7 did not improve the relatively large APs of YOLOv4 in the 0.6 to 0.9 range by a large margin, it was able to significantly improve the ones under 0.4 to above 0.45. Generally, unlike YOLOv4, where a clear increasing trend of APs could be observed in the training instance number range of 0 to 10,000, APs of YOLOv7 no longer increased at a substantial rate when training instance number was larger than 5000. The largest AP improvement rate for YOLOv7 can be observed in the training instance number range of 0 to 2000 approximately, implying YOLOv7 requires considerably fewer training instances to learn from than YOLOv4 to achieve the same level of model performance. This result could be an indication that modern object detectors progress mainly in their ability to extract and learn patterns from limited data. When abundant training data are available, however, newer detectors might not have a significant advantage over older detectors. Regardless, the improvements achieved by YOLOv7 implies that difficult annotation scenarios in agricultural contexts are much less of a challenge to modern state-of-the-art object detectors, even when training instance number is insufficient. Utilizing techniques such as pseudo labeling [35], it is promising that future object detectors can be robustly developed involving only little human image annotation effort.

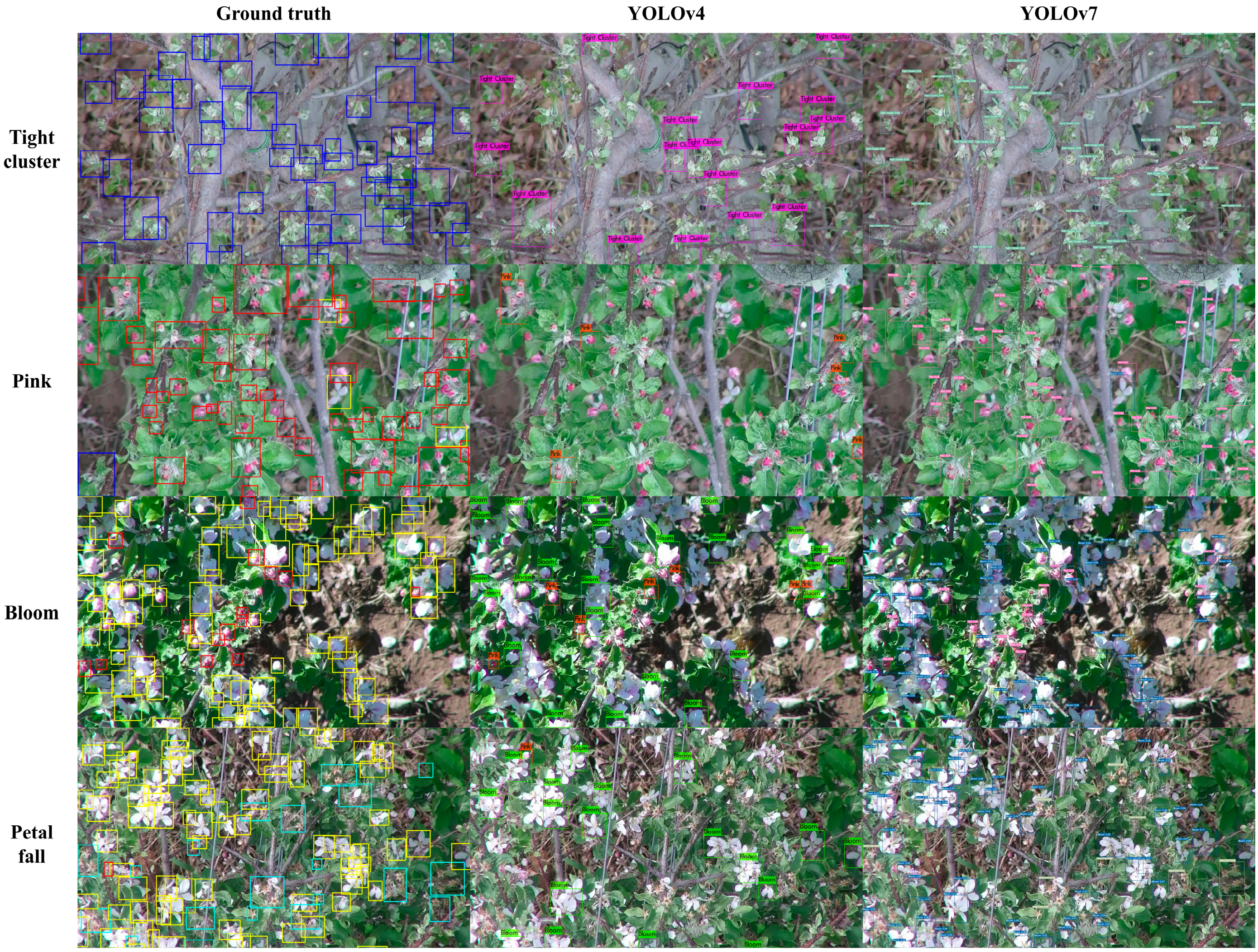

Figure 3. Apple flower bud detection results on sample test images based on a 3% confidence threshold using YOLOv4 and YOLOv7 models trained with 5% image annotation quality.

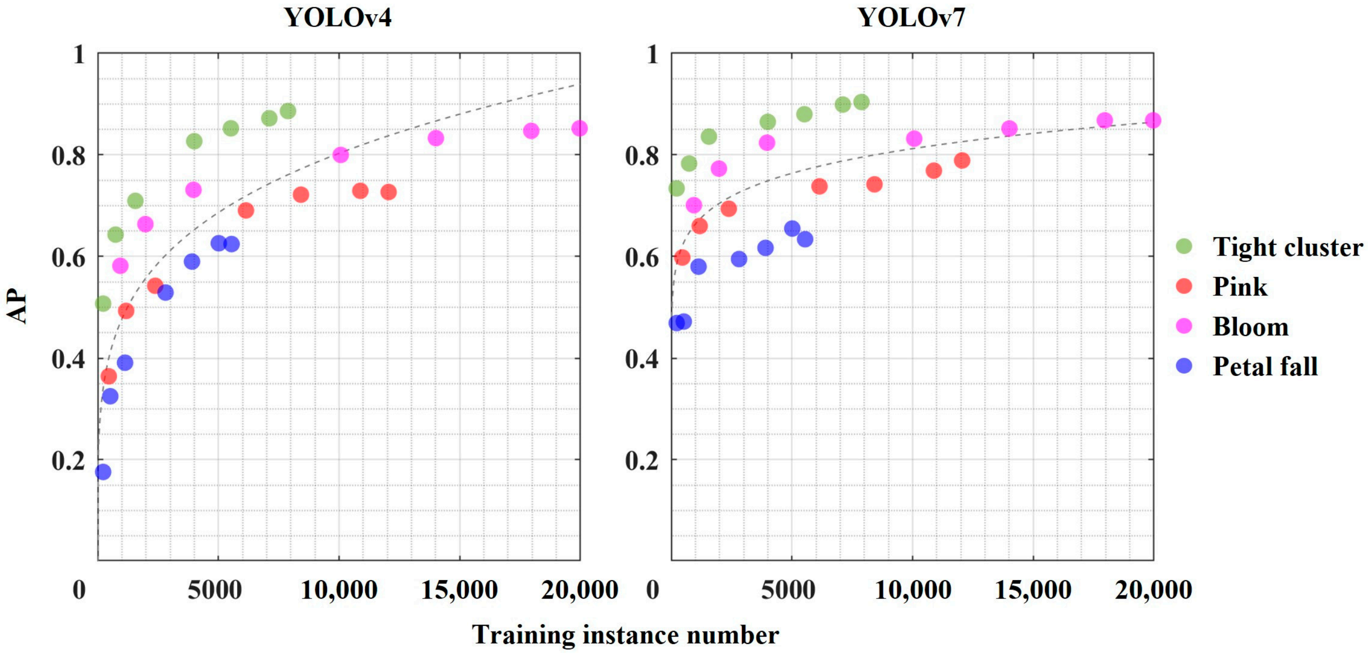

Figure 4. Relationships between training instance numbers of individual apple flower bud growth stages and model average precisions (APs) of YOLOv4 and YOLOv7 with fitted two-term power series trend lines.

4. Conclusions

Per the examination of the study, YOLOv7 is a conclusively superior object detector than YOLOv4 in terms of apple flower bud classification, which is most likely true for alternative classification tasks. Although YOLOv7 and YOLOv4 generally showed similar behaviors towards different training datasets, YOLOv7 demonstrated significantly stronger robustness against low training image annotation quality and required fewer training instances than YOLOv4 to achieve similar accuracies. However, only marginal performance improvements were observed between YOLOv7 and YOLOv4 when training instance number was sufficient, indicating the utility of antiquated but well-trained object detection models such as YOLOv4. Annotating at least 1000 to 2000 training instances for individual object classes is recommended to ensure optimal YOLOv7 performance for complicated agricultural datasets.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
2. Liu, Y.; Pu, H.; Sun, D.W. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends Food Sci. Technol.* **2021**, *113*, 193–204. [CrossRef]
3. Hafiz, A.M.; Bhat, G.M. A survey on instance segmentation: State of the art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [CrossRef]
4. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection with Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]
5. Sindagi, V.A.; Patel, V.M. A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognit. Lett.* **2018**, *107*, 3–16. [CrossRef]

6. Cholakkal, H.; Sun, G.; Shahbaz Khan, F.; Shao, L. Object counting and instance segmentation with image-level supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June; pp. 12389–12397.
7. Yeong, T.J.; Jern, K.P.; Yao, L.K.; Hannan, M.A.; Hoon, S.T.G. Applications of photonics in agriculture sector: A review. *Molecules* **2019**, *24*, 2025.
8. Mavridou, E.; Vrochidou, E.; Papakostas, G.A.; Pachidis, T.; Kaburlasos, V.G. Machine vision systems in precision agriculture for crop farming. *J. Imaging* **2019**, *5*, 89. [CrossRef]
9. Zhang, Q.; Liu, Y.; Gong, C.; Chen, Y.; Yu, H. Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors* **2020**, *20*, 1520. [CrossRef]
10. Li, G.; Huang, Y.; Chen, Z.; Chesser, G.D.; Purswell, J.L.; Linhoss, J.; Zhao, Y. Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review. *Sensors* **2021**, *21*, 1492. [CrossRef]
11. Taverriti, G.; Lombini, S.; Seidenari, L.; Bertini, M.; Del Bimbo, A. Real-Time Wearable Computer Vision System for Improved Museum Experience. In Proceedings of the MM '16: Proceedings of the 24th ACM international conference on Multimedia, Santa Barbara, CA, USA, 23–27 October 2016; pp. 703–704.
12. Chen, C.; Lu, J.; Zhou, M.; Yi, J.; Liao, M.; Gao, Z. A YOLOv3-based computer vision system for identification of tea buds and the picking point. *Comput. Electron. Agric.* **2022**, *198*, 107116. [CrossRef]
13. Soviany, P.; Ionescu, R.T. Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. In Proceedings of the 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 20–23 September 2018; pp. 209–214.
14. Fan, J.; Huo, T.; Li, X. A review of one-stage detection algorithms in autonomous driving. In Proceedings of the 2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI), Hangzhou, China, 18–20 December 2020; pp. 210–214.
15. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
16. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
17. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
19. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J. Ultralytics/yolov5: V6.2—YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations. *GitHub* **2022**.
20. Long, X.; Deng, K.; Wang, G.; Zhang, Y.; Dang, Q.; Gao, Y.; Shen, H.; Ren, J.; Han, S.; Ding, E.; et al. PP-YOLO: An Effective and Efficient Implementation of Object Detector. *arXiv* **2020**, arXiv:2007.12099.
21. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.M. Scaled-YOLOv4: Scaling Cross Stage Partial Network. *arXiv* **2020**, arXiv:2011.08036.
22. Huang, X.; Wang, X.; Lv, W.; Bai, X.; Long, X.; Deng, K.; Dang, Q.; Han, S.; Liu, Q.; Hu, X.; et al. PP-YOLOv2: A Practical Object Detector. *arXiv* **2021**, arXiv:2104.10419.
23. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. You Only Learn One Representation: Unified Network for Multiple Tasks. *arXiv* **2021**, arXiv:2105.04206.
24. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
25. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
26. 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. *arXiv* **2022**, arXiv:2207.02696.
27. Xu, M.; Bai, Y.; Ghanem, B. Missing Labels in Object Detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; pp. 1–10.
28. Ma, J.; Ushiku, Y.; Sagara, M. The Effect of Improving Annotation Quality on Object Detection Datasets: A Preliminary Study. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, New Orleans, LA, USA, 19–20 June 2022; pp. 4850–4859.
29. Yuan, W.; Choi, D. UAV-Based Heating Requirement Determination for Frost Management in Apple Orchard. *Remote Sens.* **2021**, *13*, 273. [CrossRef]
30. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common objects in context. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
31. Yuan, W.; Choi, D.; Bolkas, D.; Heinemann, P.H.; He, L. Sensitivity Examination of YOLOv4 Regarding Test Image Distortion and Training Dataset Attribute for Apple Flower Bud Classification. *Int. J. Remote Sens.* **2022**, *43*, 3106–3130. [CrossRef]
32. Böselt, L.; Thürlemann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641–2658. [CrossRef] [PubMed]

33. Nowell, D.; Nowell, P.W. A machine learning approach to the prediction of fretting fatigue life. *Tribol. Int.* **2020**, *141*, 105913. [CrossRef]
34. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
35. Lee, D.-H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the ICML 2013 Workshop on Challenges in Representation Learning, Atlanta, GA, USA, 21 June 2013; pp. 1–6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.