IJPAS ISSN: 2149-0910 | INTERNATIONAL JOURNAL PURE AND APPLIED SCIENCES

Vol. 10(1);162-173 (2024)
DOI: 10.29132/ijpas.1448068

Research Article

Evaluation of YOLOv8 Model Series with HPO for Object Detection in Complex Agriculture Domains

Jale BEKTAŞ

Received: 06.03.2024
Accepted: 09.05.2024

Mersin University, Faculty of Engineering, Department of Computer Engineering, Mersin, Türkiye; jale@mersin.edu.tr

Abstract: In recent years, many studies have been conducted in-depth investigating YOLO Models for object detection in the field of agriculture. For this reason, this study focused on four datasets containing different agricultural scenarios, and 20 different trainings were carried out with the objectives of understanding the detection capabilities of YOLOv8 and HPO (optimization of hyperparameters). While Weed/Crop and Pineapple datasets reached the most accurate measurements with YOLOv8n in mAP score of 0.8507 and 0.9466 respectively, the prominent model for Grapes and Pear datasets was YOLOv8l in mAP score of 0.6510 and 0.9641. This situation shows that multiple-species or in different developmental stages of a single species object YOLO training highlights YOLOv8n, while only object detection extracting from background scenario naturally highlights YOLOv8l Model.

Keywords: YOLOv8 state-of-the-art networks; hyperparameter optimization; agricultural images; object detection

Karmaşık Tarım Senaryoları Üzerinde Nesne Tespiti için HPO ile YOLOv8 Model Serisinin Değerlendirilmesi

Özet: Son yıllarda, tarım alanında nesne tespitine yönelik YOLO modellerini derinlemesine inceleyen birçok çalışma yapılmıştır. Bu nedenle bu çalışmada farklı tarımsal senaryolar içeren dört veri seti üzerine odaklanılmış ve YOLOv8 üzerinde HPO'nun (hiper parametrelerin düzenlenmesi) tespit yeteneklerinin anlaşılması amacıyla 20 farklı eğitim gerçekleştirilmiştir. Weed/Crop ve Pineapples veri setleri sırasıyla 0.8507 ve 0.9466 mAP skorunda YOLOv8n ile optimal ölçümlere ulaşırken, Grapes ve Pears veri setleri için öne çıkan model 0.6510 ve 0.9461 mAP skorunda YOLOv8l olmuştur. Bu durum birden fazla türün veya tek bir türün farklı gelişim aşamalarındaki nesneler üzerinde YOLO eğitiminin YOLOv8n'i öne çıkardığını, yalnızca arka plan senaryosundan elde edilen nesne algılama görevinde ise YOLOv8l modelini doğal olarak öne çıkardığını göstermiştir.

Anahtar Kelimeler: YOLOv8 state-of-the-art networkler; hiper parametre optimizasyonu; zirai görüntüler; nesne tespiti

1. Introduction

Deep learning-based models that provide state-of-the-art performance are very attractive in the domain of object detection problems [1-2]. These models combine the following merits: momentum with stable, fast, and real-time object detection in the recognition of objects with image properties in agricultural environments. Among the models developed for the solution of object recognition problems, YOLO (You Only Look Once) stands out in real-time object detection with its effective overall average precision (mAP) values [3]. On the other hand, images are complex due to the diversity of details in agricultural application data and YOLO methods can produce solutions that can more accurately express the data sets improving the evaluation performances [4] and optimizing the hyperparameters. Moreover, rectangular bounding boxes are the basis for object detectors that determine the region of each object sample more precisely than the traditional rule-based image processing techniques one of which is known as pixel-wise [5-6].

2. Literature Review

Nowadays, intelligent applications [7] in the domain of agriculture have become increasingly popular because taking management measures in various agricultural activities requires accurate adjustments according to the specific conditions of each agricultural process. As the number of objects and the interaction of objects increases in images containing real production data, the complexity of the problem increases exponentially. Especially when images with agricultural content are evaluated, these images may consist of fruits growing on trees, mature/unmatured fruit characteristics, multi-featured fruit groups in a bunch, complex agricultural field images containing different tree species, or complex structures requiring crop/weed detection. Most of the time, uncertainty in plant characteristics, subjectivity in expert evaluation, and lack of object labeling can lead to decreased object detector performances [8]. Unlike region-based detectors based on two stages phenomena such as the CNN (R-CNN) series, YOLO, all variations of which are referred to as one-stage detectors in computer vision, have demonstrated remarkable performance by combining region detection and object classification in simple structures to achieve speeds in higher performance. In the literature, prediction studies on agricultural images in recent years can be classified as object detection prediction by considering these two main detector groups:

- Convolutional Neural Network (CNN) based methods incorporate fine-tuned networks [9] for the detection of specific objects assisting some techniques for object visualization from VGG16 feature maps.
- Region-based Convolutional Neural Network (R-CNN) methods, although time-consuming a region-based strategy for object detection first determines the possible areas [10] of the object and then operate sequentially independent CNNs (Convolutional Neural Networks) in these areas. Despite this technique having significant performance, and two additional processes, it increases the number of operations on the image and results in a low FPS (Frames per second).
- Faster Region-based Convolutional Neural Network (Faster R-CNN) [11] based methods employ a CNN network and obtain fixed-size features from the topmost feature map. These methods detect objects by designing an effective region proposals algorithm that can solve detection efficiency problems.
- Mask R-CNN has previously been used in agricultural applications. Adding a mask prediction branch to the Faster R-CNN and Mask R-CNN is capable of revealing objects and inferring boundaries precisely [12].
- YOLOv7x and its other variants have recently been successfully applied in a variety of real-world applications, including object-to-object detection in images from X-ray-captured agricultural fields [13]. Studies have been conducted to measure speed and performance among many YOLO variants models such as the YOLOv5, YOLOv6, YOLOv7, and YOLOv7x [4].
- Architectures based on YOLOV8 variants are built on the capability of previous YOLO versions, adding new specific features to gradually increase performance [14]. Thanks to its ease of use and speed in real-time applications at different scales, YOLOv8's popularity has increased in a wide range of object detection and tracking, sample segmentation, and image classification processes.

Traditional image processing methods require some filtering processes to reveal the distinctive features in the image. Although researchers need to produce extra solutions suitable for many experimental processes and data in the complex parameter tuning stages, the model is prone to overfitting and the decisions it makes after training can be misleading. On the other hand, images are complex due to the diversity of details in agricultural application data, and deep learning as an object detection method can produce solutions that can more accurately express the dataset [15]. Some studies use the Faster R-CNN based on the ResNeXt-101 which is developed to extract prominent features to improve the detection capability of Faster R-CNN [16] or use R–CNN with VGG19 processed for weed identification [17]. The R-CNN series (including R-CNN, Fast R-CNN, Faster R-CNN, Mask-RCNN [18], and YOLO series (including YOLO-v1-v5) are mostly used in the agricultural field [19-20]. Wang et al. experiment with YOLOv8 and the other is the two-stage model Mask R-CNN that achieves precision rates exceeding 90% [21]. Therefore, it can be said that compared to traditional methods, to solve problems including complex agricultural datasets one-stage detectors are getting more popular thanks to their flexibility.

2.1. Paper Contribution

Architectures based on YOLOV8 variants were investigated based on images comprising complex agricultural environments. The prominent dynamics of the investigation can be summarized as follows:

- Using four different problem domains included for training with the newest versions of YOLOv8n, YOLOv8s, YOLOv8l, YOLOv8m, and YOLOv8x for feature representation.
- Investigating the YOLOV8-based versions supported by synthetic data augmentation strategy to gradually increase performance.
- Deeply understanding the YOLOv8 detection capability and practices in different agricultural scenarios. The objective for optimizing hyperparameters is that they provide a general measurement value for object detection in agricultural images containing different types of scenarios.

3. Data Materials

The network was trained and tested on Benchmark image datasets. While the Grapes, Pineapple and Pear datasets contain consequently 76, 1128, and 721 different-grown fruit images, the Weed/Crop dataset contains 644 weed plants at early growth and growth stages. Four datasets obtained from different sources included images from the field in different seasons and under different light and illumination conditions. Randomly selected sample images are depicted in Fig.1.

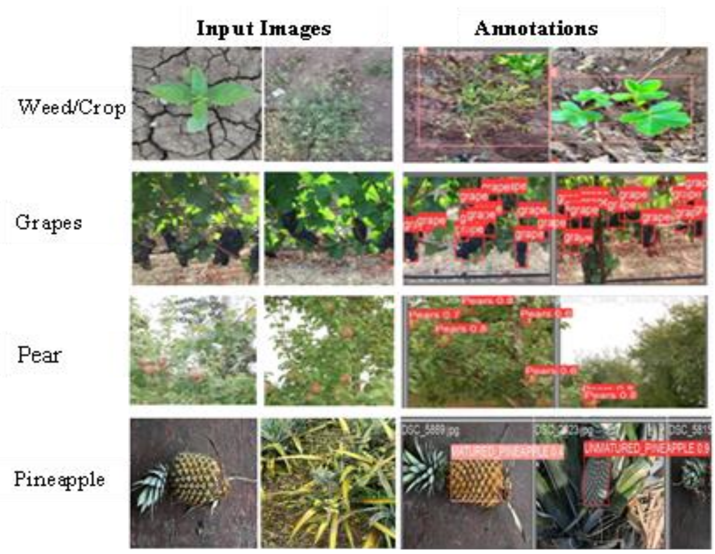

Figure 1. Sample input images and annotations randomly selected.

4. Methods

4.1. Overview of YOLOv8

YOLOv8 which has been maintained by Ultralytics Group is shared on GitHub environment. This open-source work, which includes different variants, is structured with features that can perform many tasks such as segmentation, classification, and object detection. [14]. YOLOv8 network structure is given in Fig.2. YOLOv8 comprises five versions, namely n based on 3.2Mparams, s based on 11.2Mparams, m based on 25.9Mparams, based on 43.7Mparams and x based on 68.2Mparams, with a robust network architecture developed over different models. Moreover, YOLOv8 incorporates three significant structures, including:

- The backbone is the main block of the network and comprises the C2f module that supports richer gradient flows, improving the feature extraction capacity of the backbone network.
- The Neck integrates the backbone and head and comprises SPPF and New CSP-PAN structures.
- The head is responsible for the production of final decisions to optimize the loss calculation process that is the basic structure of [21] Distribution Focal Loss.

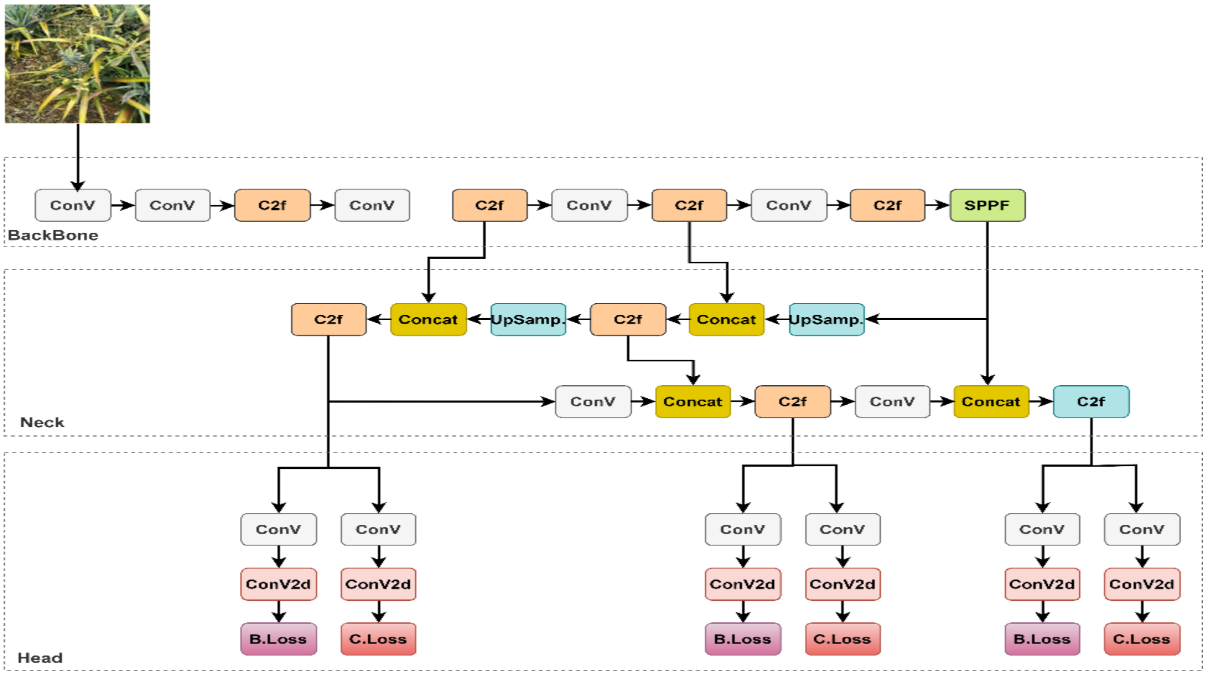

Figure 2. Schematic summarized diagram for YOLOv8 with the BackBone, Neck, and Head blocks.

4.2. Hyperparameter Evaluation

Hyperparameter training is a time-consuming process, especially for deep learning networks. It can take decades of GPU processing to finish the entire process, as even training a single neural network to converge takes almost a day [22]. It is common to use any hyperparameter optimization toolkit, which bridges network training and hyperparameter tuning. Scikit-Optimize is the expanded library specified by many more frameworks. Several machine learning methods including probabilistic approaches, searching, and some other optimization algorithms. Common aspects of all these methods include their application to search algorithms, support for deep learning training frameworks, and application programs for experiments.

4.2.1. Mosaic Augmentation

The Mosaic Augmentation method is a method that combines four randomly selected images from the training samples into a single image and creates unique diversity in object detection by using scales in various variations. YOLOv8 uses mosaic augmentation to speed up the training process.

5. Design of Experiment

When evaluating the object detector's F1 score, Precision (P) in Eq. (4.1) comprises True Positive s (TP), False Positives (FP), Recall (R) in Eq. (4.2) comprises True Positives (TP) and False Negatives (FN), average precision mAP@0.5, and finally average precision mAP@0.5:0.95 are considered the most common metrics in every scenario [19] Basic definitions are obtained from the confusion matrix as follows: According to these metrics:

$$Precision(P) = \frac{TP}{TP+FP} \tag{4.1}$$

$$Recall(R) = \frac{TP}{TP+FN} \tag{4.2}$$

In this study, the number of weeds, the bunch of grapes, the pears, and the pineapple per image was recorded as a ground truth value. TP comprises the true positives which mean weeds with a bounding box for the Weed/Crop dataset, the bunch of grapes for the Grapes dataset, the pears for the Pears dataset, and the matured pineapples for the Pineapple dataset; FP corresponds to the false positives without weeds, without matured pineapples and the background for the rest of the datasets. FN indicates false negatives in case the target weed and matured pineapples are not detected. The IoU is calculated according to the difference between the bounding box that is obtained by the model and the ground truth. The trained model obtains a confidence score for each object separately, which it performs according to the YOLOV8 algorithm, and provides a TP using the bounding box coordinates. Precision recall curve area determines the average precision (AP) in Eq. (4.4).

$$F1\ score = 2x\frac{P\ x\ R}{P+R} \tag{4.3}$$

$$Average\ Precision(AP) = \int_0^1 P(R)dR \tag{4.4}$$

When discussing the outputs of the model, the average of the AP is used on a class-by-class basis and an average precision (mAP) value is obtained. The mAP value is more sensitive in measuring different precision values obtained from the recall function than the AP value measurements. In this study, consequently, three metrics were produced: F1 score in Eq. (4.3) and precision values using two different thresholds: mAP@0.5 and mAP@[0.5:0.95].

6. Experimental Results

6.1. Training Environments and Hyperparameter Settings

A 5-fold cross-validation Monte-Carlo procedure was employed partitioning the data randomly into training, validation, and test sets for model performance assessment using CoLaboratory (Google LLC, Mountain View, CA, USA) environment which has a speed option given as 280 FPS on an NVIDIA A100 TensorCore GPU. For the hyperparameter adjustment [23], the pixel size of the images of the input network was set to 640X640. The Stochastic Gradient Descent method was chosen as the model optimizer which is one of the most important hyperparameters. The weight decay value was given as 0.0005. The batch size was chosen as 16. The epoch number is fixed as 50 for all of the model training periods. The initial learning rate was chosen as 0.01 integrating the momentum value to 0.937. To boost the training process during mosaic augmentation the last ten epochs during YOLOv8 training, augmentation is frozen for image HSV (hue, saturation, and value) parameters were given consequently: h value was set to 0.015, s was set to 0.7, and v was set to 0.4. Translate and scale values were given as 0.1, flip left-right was set to 0.5, and finally, the mosaic parameter was set to 1.0.

6.2. Accuracy Assessment on the Validation Datasets

YOLOv8 Models were assessed in various ways across the Weed/Crop, Grapes, Pears, and Pine-apple datasets and showed significant performance on validation processes. Performances of the five models on four datasets are resumed and depicted in Fig.3.

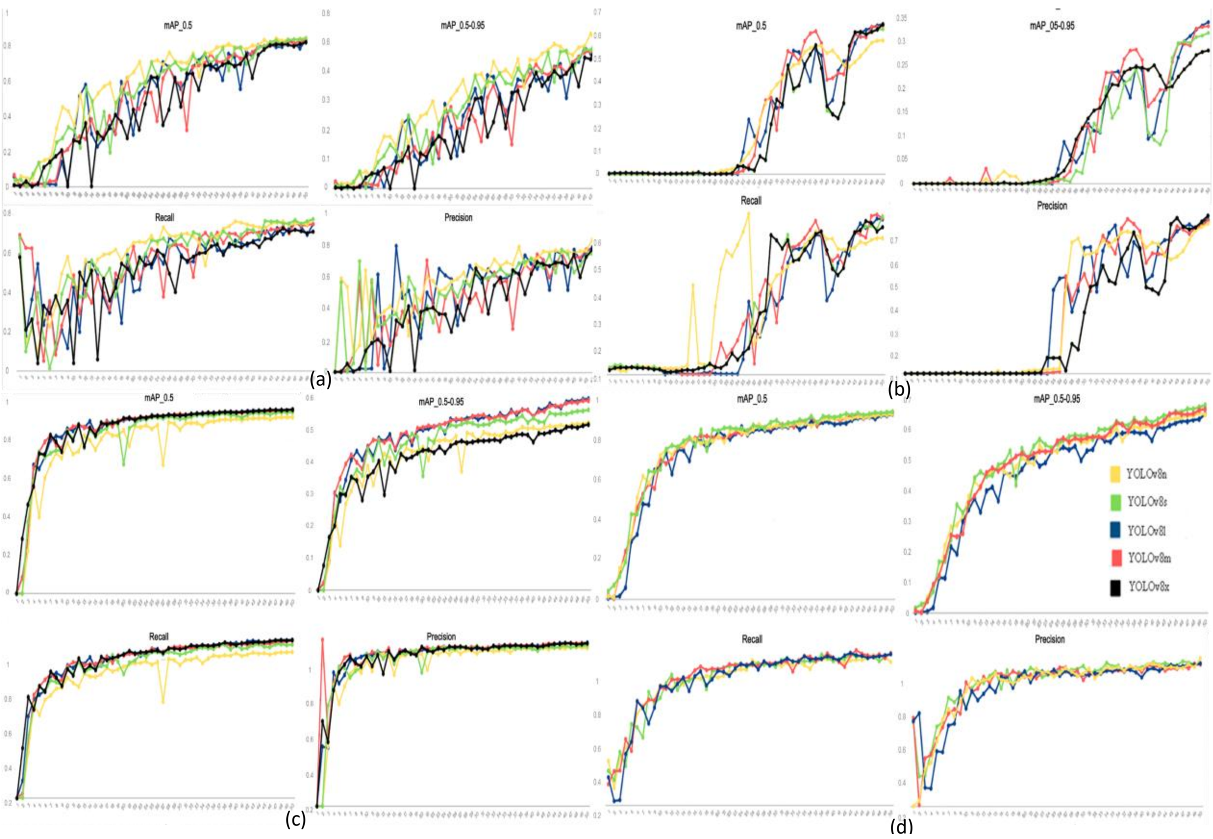

Figure 3. Different model scales of precision, recall, mAP@ 0.5, and mAP@0.5-0.95 for 50 epochs. (a)Weed/Crop scales, (b) Grapes dataset scales, (c) Pear dataset scales, (d) Pineapple dataset scales

Performance results and specifically F1 measures of best YOLO model versions for the five studied YOLO versions (YOLOv8n, YOLOv8s, YOLOv8l, YOLOV8m, and YOLOv8x), and best Model results were highlighted in bold. After 50 epochs during the training of the five studied Models on the four datasets, the results revealed that the performance decreased drastically with the existing Grapes dataset results. Among YOLOv8 model measurements, YOLOv8n obtained the highest P with 0.8949 on the Pear dataset, while YOLOv8n obtained the best R with 0.8568, mAP@0.5 with 0.9466 and mAP@0.5:0.95 with 0.6815 on the Pineapple dataset which makes YOLOv8n the most appropriate Model for the Pineapple dataset. Among YOLOv8 model measurements, YOLOv8s obtained the best P with 0.9094 and mAP@0.5 with 0.9509 on the Pear dataset, while YOLOv8s obtained the best R with 0.9059 and mAP@0.5:0.95 (0.6498) on the Pineapple dataset. YOLOv8l obtained the best P with 0.9266, the best R with 0.9155, and mAP@0.5 with 0.9509 on the Pear dataset. YOLOv8m obtained the best P (0.9196), the best R with 0.9110, and mAP@0.5 with 0.9613 on the Pear dataset. Among YOLOv8 model measurements, YOLOv8x gained the best P with 0.9232 and mAP@0.5 with 0.9631 on the Pear dataset, while YOLOv8x obtained the best R with 0.9883 and mAP@0.5:0.95 with 0.6476 on the Pineapple dataset

Among the five studied YOLO versions, YOLOv8n superimposed the Weed/Crop dataset for all the metrics resulting in P with 0.8341, R with 0.7725, mAP@0.5 with 0.8507, mAP@0.5:0.95 with 0.5609, and F1 with 0.8015. Among the five studied YOLO versions, YOLOv8n superimposed the Pineapple dataset for all the metrics resulting in P with 0.8885, R with 0.8568, mAP@0.5 with 0.9466, mAP@0.5:0.95 with 0.6815, and F1 with 0.8810. YOLOv8l surpassed the other four studied YOLO versions for the Grapes and Pear datasets. YOLOv8l trials resulted in P with 0.6639, R with 0.5787, mAP@0.5 with 0.6510, mAP@0.5:0.95 with 0.3433, and F1 with 0.6186 for the Grapes dataset. Similar to these results, YOLOv8l obtained P with 0.9266, R with 0.9155, mAP@0.5 with 0.9641, mAP@0.5:0.95 with 0.5975, and F1 with 0.9210 for the Pear dataset that is the best experiment results overall the studied datasets.

Table 1. *Performance and F1 scores of YOLOv8 model measurements for the five studied YOLO versions (YOLOv8n. YOLOv8s. YOLOv8l. YOLOV8m. and YOLOv8x).*

Dataset	Model	P	R	mAP@0.5	mAP@0.5:0.95	F1
Weed/Crop	*YOLOv8n*	0.8341	0.7725	0.8507	0.5609	0.8015
	YOLOv8s	0.8209	0.7713	0.8423	0.5365	0.7959
	YOLOv8l	0.8386	0.7464	0.8369	0.5168	0.7898
	YOLOv8m	0.8208	0.7472	0.8294	0.5136	0.7822
	YOLOv8x	0.8272	0.7085	0.8239	0.5133	0.7632
Grapes	YOLOv8n	0.6465	0.5091	0.5888	0.2809	0.5696
	YOLOv8s	0.6807	0.5499	0.6296	0.3206	0.6083
	YOLOv8l	0.6639	0.5787	0.6510	0.3433	0.6186
	YOLOv8m	0.6591	0.5829	0.6439	0.3346	0.6183
	YOLOv8x	0.6806	0.5499	0.6296	0.3206	0.6083

Pear	YOLOv8n	0.8949	0.8448	0.9238	0.5237	0.8691
	YOLOv8s	0.9094	0.8882	0.9509	0.5630	0.8986
	YOLOv8l	0.9266	0.9155	0.9641	0.5975	0.9210
	YOLOv8m	0.9196	0.9110	0.9613	0.5914	0.9152
	YOLOv8x	0.9232	0.9137	0.9631	0.6031	0.9184
Pineapple	*YOLOv8n*	0.8885	0.8568	0.9466	0.6815	0.8810
	YOLOv8s	0.8575	0.9059	0.9350	0.6498	0.8753
	YOLOv8l	0.8579	0.9041	0.9333	0.6519	0.8803
	YOLOv8m	0.8519	0.9107	0.9427	0.6685	0.8803
	YOLOv8x	0.8486	0.9883	0.9252	0.6476	0.9131

It can be observed from Table 1 that YOLOv8n is the prior Model for the Weed/Crop and Pineapple datasets, while YOLOv8l is the prior Model for the Grapes and Pear datasets. Additionally, multiple-species distinctions or in different developmental stages of a single-species object YOLO training were supported by Table 2 with YOLOv7 variants. It has been observed that the experiments presented in Table are lower in proportion than the results obtained on the same dataset with YOLOv8n. Comparisons made for Yolov8 variants over different object detection scenarios in the same domain are given in Table 3.

Table.2. *Performance comparisons of YOLOv7 variants on the Weed/Crop dataset*

Model	P	R	F1	mAP@0.5
YOLOv7	0.74	0.73	0.74	77.26
YOLOv7x	0.81	0.75	0.81	82.90

Weed/Crop detection training was also evaluated with YOLOv7 and YOLOv7x.TP comprises the true positives which mean weeds with a bounding box for the Weed/Crop dataset, the matured pineapples for the Pineapple dataset, while TP comprises the bunch of grapes for the Grapes dataset, and the pears for the Pears dataset. FP corresponds to the false positives without weeds for the Weed/Crop dataset, without matured pineapples for the Pineapple dataset, and the background for the rest of the datasets. These two scenarios were observed on four datasets, and based on the objectives of the datasets in object detection (Weed/Crop and Pineapple detect two separate object types, Grapes and Pear datasets distinguish between objects and ground), it was observed that they were naturally grouped in pairs and were successful in two separate YOLOv8 Models. As shown in Fig.4., the highest sensitivity detection was obtained because of the morphology of the pear fruit and tree pattern on all YOLO training models.

Weed/Crop Dataset training with YOLOv8n

Grapes Dataset training with YOLOv8n

Pear Dataset training with YOLOv8l

Pineapple Dataset training with YOLOv8n

Figure 4. *Sample detections on the YOLOV8 versions with the most accurate models on four datasets.*

Table.3. YOLOv8 object detection studies in the agriculture domain

Models	Dataset Domain	P	R	mAP@0.5
[12] YOLOv8n	Rapaseed Pod Detection	0.98	0.99	0.99
[14] YOLOv8x	Pest-Infected Leaf Region	0.75	0.65	0.67
[25] YOLOv8m	Insect detection on PlantDoc dataset	0.752	0.549	0.627
[26] YOLOv8s	Strawberry ripeness detection	0.90	0.79	0.93

7. Discussion

When identifying crops in agricultural environments, generalization problems often occur due to structural differences in the crops' size changes over time. In this study, four different data sets were analyzed with YOLO-v8, in which different variants were used to detect the temporal change of the crop, distinguish it from different objects, or separate it from other components in the environment. Additionally, synthetic augmentation of the dataset in lower fidelity, slightly distorted but realistic images by manual marking (especially in images containing objects with transitions to each other) helped to reduce overfitting and improved the generalization ability in object detection, as can be seen from the F1 values in Table 1. While Weed/Crop and Pineapple datasets reached the most sensitive and accurate measurements with YOLOv8n in F1-score and mAP@0.5 measurements, the prominent model for Grapes and Pear datasets was YOLOv8l. The lowest sensitivity detection was obtained for the Grapes dataset that remained under an F1-score of 0.62. Relatively low results compared to other experiments on the grape dataset may depend on several factors including different tree species, leaf morphologies, weather, and season conditions affecting the grape bunch view in the images [24]. Multiple species or in different developmental stages of a single species object YOLO training achieved with mAP score of 0.8507 for Weed/Crop, with mAP score of 0.9466 for matured/unmatured pineapple detection. Although satisfactory outputs are obtained from investigations on datasets, efforts still need to be made to improve the overall accuracy in complex problems such as grapes bunch detection.

8. Conclusion

YOLOv8's popularity has increased in a wide range of object detection and tracking, sample segmentation, and image classification processes because of its model practices and speed in real-time applications in different environments. Therefore, in this study, a total of 20 trainings were carried out by applying similar hyperparameters on YOLOv8 models to enable object detection in 4 different datasets. The Mosaic image augmentation technique was used to increase image diversity. Experiments have produced surpassing results on two models of YOLOv8. While Weed/Crop and Pineapple datasets reached the most sensitive and accurate measurements with YOLOv8n in F1-score and mAP@0.5 measurements, the prominent model for Grapes and Pear datasets was YOLOv8l. This situation shows that multiple-species or in different developmental stages of a single species object YOLO training highlights YOLOv8n, while only object detection extracting from background scenario naturally highlights YOLOv8l Model.

This research proposes to deeply understand the YOLOv8 detection capability and practices in different agricultural scenarios by using YOLOV8 models on four datasets. The objectives that are based on regularizations of the hyperparameters measure the closeness between the estimated and experimental records for different object detection. These regulations have concluded the study by highlighting two models, namely YOLOv8n and YOLOV8l. Although the results obtained from experiments on different datasets are satisfactory, efforts still need to be made to improve the overall accuracy in complex problems such as grapes bunch detection.

Future Work

Since it is difficult to accurately label the entire area of a single object in a manual labeling process, hybrid use of detection methods with a clustering method can improve performance. As future works, YOLOv8 models can be compound together with a density-based clustering method.

Conflict of interest

The Author reports no conflict of interest relevant to this article.

Research and publication ethics statement

The author declares that this study complies with research and publication ethics.

References

[1] Jin, X., Sun, Y., Che, J., Bagavathiannan, M., Yu, J., & Chen, Y. (2022). A novel deep learning-based method for detection of weeds in vegetables. *Pest Management Science*, 78(5), 1861-1869.

[2] Andaç, İ. M. A. K., Doğan, G., ŞENGÜR, A., & Ergen, B. (2023). Asma Yaprağı Türünün Sınıflandırılması için Doğal ve Sentetik Verilerden Derin Öznitelikler Çıkarma, Birleştirme ve Seçmeye Dayalı Yeni Bir Yöntem. *International Journal of Pure and Applied Sciences*, 9(1), 46-55.

[3] Dahirou, Z., & Zheng, M. (2021). Motion Detection and Object Detection: Yolo (You Only Look Once). *In 2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, Guiyang, China, IEEE, 250-257.

[4] Strand, J. F. (2000). Some agrometeorological aspects of pest and disease management for the 21st century. *Agricultural and Forest Meteorology*, 103(1-2), 73-82.

[5] Dominguez-Sanchez, A., Cazorla, M., & Orts-Escolano, S. (2018). A new dataset and per-for-mance evaluation of a region-based cnn for urban object detection. *Electronics*, 7(11), 301.

[6] Joseph, E. C., Bamisile, O., Ugochi, N., Zhen, Q., Ilakoze, N., & Ijeoma, C. (2021). Systematic Advancement of Yolo Object Detector For Real-Time Detection of Objects. *In 2021 18th In-ter-national Computer Conference on Wavelet Active Media Technology and Information Pro-cessing (ICCWAMTIP)*, Chengdu, China, IEEE, 279-284.

[7] Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., ... & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications, and recent trends. *Knowledge-Based Systems*, 194, 105596.

[8] Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS journal of photogrammetry and remote sensing*, 117, 11-28.

[9] Menikdiwela, M., Nguyen, C., Li, H., & Shaw, M. (2017). CNN-based small object detection and visualization with feature activation mapping. *In 2017 international conference on image and vision computing*, New Zealand (IVCNZ), IEEE, 1-5.

[10] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 580-587.

[11] Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detec-tion system using deep neural networks. *Sensors*, 16(8), 1222.

[12] Wang, N., Liu, H., Li, Y., Zhou, W., & Ding, M. (2023). Segmentation and Phenotype Cal-culation of Rapeseed Pods Based on YOLO v8 and Mask R-Convolution Neural Networks. *Plants*, 12(18), 3328.

[13] Liu, K., Tang, H., He, S., Yu, Q., Xiong, Y., & Wang, N.(2021). Performance validation of YOLO variants for object detection. *In Proceedings of the 2021 International Conference on Bioinfor-matics and intelligent computing*, New York, NY, United States,239-243.

[14] Zhu, R., Hao, F., & Ma, D. (2023). Research on Polygon Pest-Infected Leaf Region Detec-tion Based on YOLOv8. Agriculture, 13(12), 2253.

[15] Zhang, K., Wu, Q., & Chen, Y. (2021). Detecting soybean leaf disease from synthetic image using multi-feature fusion faster R-CNN. *Computers and Electronics in Agriculture*, 183.

[16] Mu, Y., Feng, R., Ni, R., Li, J., Luo, T., Liu, T., ... & Hu, T. (2022). A Faster R-CNN-Based Model for the Identification of Weed Seedling. *Agronomy*, 12(11).

[17] Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., & Yuan, Z. (2019). Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN. *Biosystems Engineering*, 184.

[18] Jabir, B., Moutaouakil, K. E., & Falih, N. (2023). Developing an Efficient System with Mask R-CNN for Agricultural Applications. Agris on-line Papers in Economics and Informatics, 15(1).

[19] Dang, F., Chen, D., Lu, Y., & Li, Z. (2023). YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. *Computers and Electron-ics in Agriculture*, 205.

[20] Gallo, I., Rehman, A. U., Dehkordi, R. H., Landro, N., La Grassa, R., & Boschetti, M. (2023). Deep Object Detection of Crop Weeds: Performance of YOLOv7 on a Real Case Dataset from UAV Images. *Remote Sensing*, 15(2).

[21] Wang, F., Fu, X., Duan, W., Wang, B., & L, H. (2023). Visual Detection of Lost Ear Tags in Breeding Pigs in a Production Environment Using the Enhanced Cascade Mask R-CNN. *Agriculture*, 13(10), 2011.

[22] Altun, S., & Talu, M. F. (2021). Derin sinir ağları için hiperparametre metodlarının ve kitlerinin incelenmesi. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 12(2), 187-199.

[23] Wang, C.-Y.,Bochkovskiy, A., Liao, H.-Y.M.(2022). 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*, 7464-7475.

[24] Sozzi, M., Cantalamessa, S., Cogato, A., Kayad, A., & Marinello, F. (2022). Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algo-rithms. *Agronomy*, 12(2), 319.

[25] Giakoumoglou, N., Pechlivani, E. M., & Tzovaras, D. (2023). Generate-Paste-Blend-Detect: Synthetic dataset for object detection in the agriculture domain. Smart Agricultural Technology, 5, 100258.

[26] Yang, S., Wang, W., Gao, S., & Deng, Z. (2023). Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. Computers and Electronics in Agriculture, 215, 108360.