Turkish Journal of Science & Technology
18(2), 519-533, 2023

Research Paper
https://doi.org/10.55525/tjst.1342555

# Enhancing Strawberry Harvesting Efficiency through Yolo-v7 Object Detection Assessment

**Mehmet NERGİZ**[1*]
[1] Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Dicle Üniversitesi, Diyarbakır, Türkiye
[*1] mnergiz@dicle.edu.tr

**(Geliş/Received: 13/08/2023; Kabul/Accepted: 31/08/2023)**

**Abstract:** Strawberry fruits which are rich in vitamin A and carotenoids offer benefits for maintaining healthy epithelial tissues and promoting maturity and growth. The intensive cultivation and swift maturation of strawberries make them susceptible to premature harvesting, leading to spoilage and financial losses for farmers. This underscores the need for an automated detection method to monitor strawberry development and accurately identify growth phases of fruits. To address this challenge, a dataset called Strawberry-DS, comprising 247 images captured in a greenhouse at the Agricultural Research Center in Giza, Egypt, is utilized in this research. The images of the dataset encompass various viewpoints, including top and angled perspectives, and illustrate six distinct growth phases: "green", “red”, "white", "turning", "early-turning" and "late-turning". This study employs the Yolo-v7 approach for object detection, enabling the recognition and classification of strawberries in different growth phases. The achieved mAP@.5 values for the growth phases are as follows: 0.37 for "green," 0.335 for "white," 0.505 for "early-turning," 1.0 for "turning," 0.337 for "late-turning," and 0.804 for "red". The comprehensive performance outcomes across all classes are as follows: precision at 0.792, recall at 0.575, mAP@.5 at 0.558, and mAP@.5:.95 at 0.46. Notably, these results show the efficacy of the proposed research, both in terms of performance evaluation and visual assessment, even when dealing with distracting scenarios involving imbalanced label distributions and unclear labeling of developmental phases of the fruits. This research article yields advantages such as achieving reasonable and reliable identification of strawberries, even when operating in real-time scenarios which also leads to a decrease in expenses associated with human labor.

**Key words:** Strawberry, Yolo-v7, Object detection, Agriculture, Deep learning.

## Yolo-v7 Nesne Tespiti ile Çilek Hasat Verimliliğinin Artırılması

**Öz:** A vitamini ve karotenoidler açısından zengin olan çilek meyveleri, sağlıklı epitel dokularını korur ve büyümeyi destekleyici faydalar sunar. Çileklerin yoğun ekimi ve hızlı olgunlaşması, bu meyveyi erken hasada duyarlı hale getirerek, çiftçiler için çürük hasat elde etmeye ve mali kayıplara yol açar. Bu durum, çilek gelişimini izlemek ve meyvelerin büyüme aşamalarını doğru bir şekilde belirlemek için otomatik bir algılama yöntemine olan ihtiyacı arttırmaktadır. Bu zorluğun üstesinden gelmek için, bu araştırmada Mısır'ın Giza kentindeki Tarımsal Araştırma Merkezi'ndeki bir serada çekilen 247 görüntüden oluşan Strawberry-DS adlı bir veri seti kullanılmıştır. Veri kümesinin görüntüleri, üstten ve açılı perspektifler dâhil olmak üzere çeşitli bakış açılarını kapsayacak şekilde altı farklı büyüme aşamasını içermektedir: "yeşil", "kırmızı", "beyaz", "dönüşüm", "erken-dönüşüm" ve "geç-dönüşüm". Bu çalışma, farklı büyüme evrelerindeki çileklerin tanınmasını ve sınıflandırılmasını tespit etmek için Yolo-v7 nesne tespiti yöntemini kullanmaktadır. Büyüme aşamaları için elde edilen mAP@.5 değerleri şu şekildedir: "yeşil" için 0,37, "beyaz" için 0,335, "erken-dönüşüm" için 0,505, "dönüşüm" için 1,0, "geç-dönüşüm" için 0,337 ve "kırmızı" için 0,804. Tüm sınıflardaki kapsamlı performans sonuçları ise şu şekildedir: 0,792'de kesinlik, 0,575'te hatırlama, 0,558'de mAP@.5 ve 0,46'da mAP@.5:.95. Özellikle, bu sonuçlar, dengesiz etiket dağılımları ve meyvelerin gelişim evrelerinin etiketlerinin net olmaması gibi etiketleri de içeren bir veri seti ile eğitilip test edilmesine rağmen, hem performans değerlendirmesi hem de görsel değerlendirme açısından önerilen araştırmanın etkinliğini göstermektedir. Bu araştırma makalesi, gerçek zamanlı senaryolarda çalışırken bile çileklerin makul ve güvenilir bir şekilde tespit edilmesi gibi avantajlar sağlamakta ve bu da işçilik maliyetlerinde azalmayı sağlamaktadır.

**Anahtar kelimeler:** Çilek, Yolo-v7, Nesne tanıma, Tarım, Derin öğrenme.

## 1. Introduction

The perennial plant, scientifically named Fragaria x Ananassa, belongs to the Rosaceae family and is recognized as a herbaceous perennial within the genus strawberry [1]. The wild strawberry has its beginnings in Europe, Asia and America whereas the origin of the contemporary planted strawberry, known for its larger fruits, can be traced back to France [1]. Abundant in carotenoids and vitamin A, strawberries offer advantages for preserving robust epithelial tissues and stimulating growth and maturation [1-2]. The considerable amount of dietary fiber found in strawberries could potentially play a positive role in aiding the digestive process within the gastrointestinal system, as well as in the prevention of both acne and colon cancers [1]. Due to their dense cultivation and rapid maturation, premature harvesting of strawberries can readily result in the spoilage of the fruit, causing financial setbacks for farmers [1]. Currently, the predominant method for gathering strawberries is manual

* Corresponding author: mnergiz@dicle.edu.tr. ORCID Number of authors: [1] 0000-0002-0867-5518

labor, imposing significant strain on this process due to elevated labor expenses, demanding physical effort, and limited employee productivity [1,3]. Because of these factors, overseeing the development of strawberries proves to be a challenging endeavor, and the manual collection of mature strawberries is a monotonous and time-intensive undertaking [1]. More and more, in recent times, there have been reports of a decline in the number of agricultural laborers due to aging, particularly highlighted during the COVID-19 pandemic, even extending to the cultivation of fruit crops [4-5]. Thus, cultivating strawberries in open fields demands a substantial amount of human workforce, a task that is becoming progressively challenging to enlist manpower for [6].

All these facts about the strawberry cultivation necessitate the creation of an automated detection technique to oversee the progress of strawberries and accomplish precise recognition of matured fruit. Computer vision is a field in which machine learning techniques is commonly applied in medical and agricultural images [7-11]. Presently, this research field functions as a primary instrument for detecting agricultural commodities and has found extensive application in tasks such as identifying maturity levels, remotely monitoring crops, predicting yields, facilitating harvesting robots, and aiding in the selection of suitable plant varieties [1, 12-19] . Nonetheless, the intricate and ever-changing natural surroundings still exert particular impacts on fruit identification. Instances of these include the obscuring caused by leaves, fruit clustering and disturbances from plant arrangement, and fluctuations in lighting. These are prevalent aspects that impact the precision of fruit recognition [1].

There are two primary classifications for object detection model architectures such as one-step and dual-step [1]. One-step object detection algorithms identify targets by capturing features just once, primarily encompassing the SSD technique and the YOLO models [1]. Dual-step object detection algorithms necessitate the initial creation of potential regions before proceeding to employ a convolutional neural network (CNN) for target detection. This category primarily involves the SPPNet and the R-CNN series of algorithms [1]. Historically, in contrast to the dual-step object detection models, the single-step object detection models exhibit superior real-time capabilities while compromising somewhat on accuracy [1]. Nonetheless, due to the ongoing enhancement and advancement of the YOLO algorithm, its precise and effective detection capabilities have garnered substantial attention and practical implementation [1]. The YOLO series algorithms exhibit remarkable versatility and resilience, enabling them to adjust to object detection assignments amid intricate scenarios, encompassing various sizes, orientations, and obstructions. This adaptability is highly valuable for their application in real-world agricultural contexts [1].

### 1.1. Related works:

Li et al. presented YOLOv5-ASFF, an enhanced real-time deep learning model for detecting strawberries in multiple stages derived from the enhanced YOLOv5 architecture. By integrating the adaptive spatial feature fusion (ASFF) component to the YOLOv5, Li et al. indicated that the model dynamically acquires the combined spatial weights of strawberry activation maps across different scales. This approach targeted to capture comprehensive image feature data related to strawberries more effectively. To validate the capabilities of YOLOv5-ASFF, Li et al. curated an extensive strawberry image dataset that encompasses diverse complex scenarios, including instances of leaf shading, overlapping fruits, and densely clustered fruits [1].

Lemsalu et al. created a real-time application for identifying strawberries and their peduncles using the YOLOv5 model, implemented on an edge device. This system effectively discerned both mature and immature strawberries along with their respective peduncles for automated harvesting purposes. Furthermore, Lemsalu et al. compiled a dataset of strawberries, annotated it, and utilized it to train their model [6].

Zhang et al. introduced YOLOv5s-Straw, a specialized model based on YOLOv5. They adapted the base model by substituting the C3 component in the foundational network with the C2f component, a change that enhanced the flow of feature gradients. Additionally, Zhang et al. integrated the Spatial Pyramid Pooling Fast and the Cross Stage Partial Net into the concluding layer of the YOLOv5s backbone network [4]. This integration aimed to bolster the model's capacity to generalize across the strawberry dataset examined in their research [4].

Lawal introduced YOLOStrawberry, a model version built upon the adapted YOLOv5 structure, and conducted comparisons against other YOLO lightweight variants. In the architecture of YOLOStrawberry, the backbone network incorporates elements like SPPF, Conv_Maxpool, ResNet, Shuffle_Block and SElayer. The neck network employed FPN, while the achievement ratio of strawberry detection was enhanced through the application of the CIoU loss function [20].

Mao et al. presented the RTFD model, an acronym for real-time fruit detection, a slight model tailored for edge CPU gadgets aimed for detecting the fruit and vegetables. Using the PicoDet-S model as a backbone, RTFD refined the architecture, loss and activation functions to elevate the live detection performance specifically for edge CPU gadgets [21].

Mejia et al., introduced an independent rover system designed to identify strawberries and gauge their ripeness within a genuine agricultural field. This was achieved by implementing an image processing pipeline that makes use of visual data from a stereo camera. Additionally, they established a comprehensive strawberry map that

imparts crucial insights to farmers regarding the fruit's development stage, condition, and potential yield. This map was especially valuable for navigating ridge planting environments, characterized by uneven landscapes, confined spaces, and challenging backgrounds [22].

Ren et al., introduced a mobile robotics framework encompassing essential components such as a stereoscopic camera, a robotic manipulator with 6 degrees of freedom and a gripper situated on an independent mobile base. Through the utilization of the Yolo-v4-tiny algorithm, they acquired information concerning the position and maturity level of individual fruits. A software-based open loop algorithm for fruit positioning and manipulation was also developed, enabling the accomplishment of tasks spanning fruit identification, pinpointing, gripping, release and placement [23].

Within this research, the Yolo-v7 approach for object detection is employed to identify strawberries across their various developmental phases and categorize them according to their growth stages. The findings of this proposed research showcase satisfactory outcomes, both from the perspective of performance evaluations and visual assessments, even when confronting intricate circumstances involving infrequent labels and ambiguous developmental phases of the fruits.

Some of the advantageous aspects of this research article on localization and phase classification of strawberries using the YOLO-V7 method are as follows [1]:

- Automated detection and monitoring
- Precise and accurate object detection even in real time
- Adaptability for complex images like obstructions, different orientations, colors and sizes of strawberries
- Reduction of labor costs

## 2. Material

"Strawberry-DS" dataset which is used in this study is a collection of annotated images featuring strawberries at various phases of development. Renowned for their distinct flavor and nutritional value, strawberries (Fragaria X Ananassa) are globally cultivated fruits utilized fresh or processed. With substantial economic importance and export potential, assessing strawberry characteristics during growth stages is pivotal for cultivar selection and yield estimation. Traditionally, growth phase assessment relies on time-intensive visual inspection, prompting the creation of this dataset. Comprising 247 high-resolution RGB images, Strawberry-DS captures strawberries at different developmental stages, annotated by hand through the Roboflow tool. Annotations are provided in YOLO format for reference to the region of interest [24].

Captured using a Sony Xperia Z2 LTE-A D6503 smartphone camera with a 20.7 MP CMOS sensor, the dataset images encompass fully visible strawberries and those partially hidden by foliage or other fruits. Collected from a greenhouse at the Agricultural Research Center in Giza, Egypt, the images span top and various angled views. Strawberry-DS.zip houses 247 .jpg images, depicting six growth phases: "green", “red”, "white", "turning", "early-turning" and "late-turning". The general turning phases signify color transitions, with “early-turning” displaying around 10% and 30% red color, “turning” featuring around 30% and 60% red color and “late-turning” showing around 60% and 90% red color [24].

The dataset's significance extends to advanced agriculture automation, offering support for developing robotic harvesting systems that increase yield and reduce production costs. It plays a vital role in constructing robust machine vision models for decision-making in during, before and after harvesting strawberry operations [24]. Furthermore, Strawberry-DS enables smartphone and drone-based monitoring, yield prediction, and accurate evaluation of strawberry harvesting times, providing essential aid to farmers and advancing the agricultural sector. Some sampled images for each class depicting the gradual phase changes of the growth of the strawberries are shared in Figure 1. The image counts and label numbers per class are shared for train, validation and test datasets in Table 1.

## 3. Methods

The real-time identification of objects has become a pivotal element in a wide array of applications, encompassing diverse domains like self-driving vehicles, robotics, video monitoring, and augmented reality [25].

**Table 1**. The images numbers and label numbers per class for train, validation and test datasets

| | Train | | Validation | | Test | | Total | |
|---|---|---|---|---|---|---|---|---|
| **Class** | **Images** | **Labels** | **Images** | **Labels** | **Images** | **Labels** | **Images** | **Labels** |
| all | 172 | 733 | 49 | 225 | 26 | 104 | 247 | 1062 |
| white | 172 | 168 | 49 | 54 | 26 | 35 | 247 | 257 |
| green | 172 | 308 | 49 | 104 | 26 | 43 | 247 | 455 |
| early-turning | 172 | 19 | 49 | 7 | 26 | 2 | 247 | 28 |
| turning | 172 | 23 | 49 | 10 | 26 | 2 | 247 | 35 |
| late-turning | 172 | 37 | 49 | 14 | 26 | 3 | 247 | 54 |
| red | 172 | 178 | 49 | 36 | 26 | 19 | 247 | 233 |

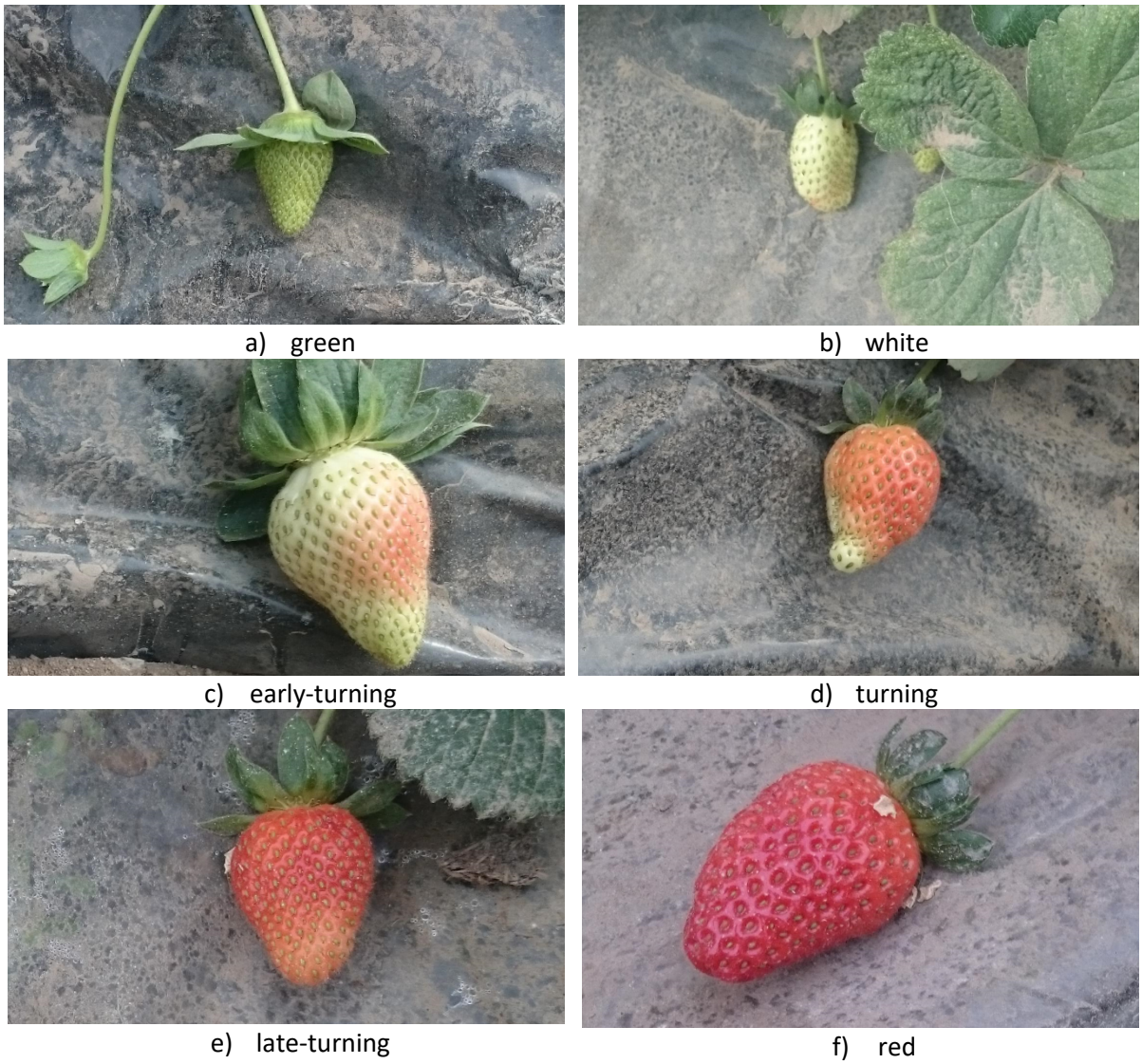



**Figure 1**. Sampled images for each class of Strawberry-DS dataset [24]

Amid the numerous algorithms designed for object detection, the YOLO (You Only Look Once) framework has gained prominence due to its exceptional equilibrium between swiftness and precision. This characteristic empowers swift and dependable object recognition within images [25]. Redmon and his colleagues released the initial YOLO paper during the CVPR conference of 2016 [25-26]. For the very first time, it introduced a real-time comprehensive technique for identifying objects. The acronym YOLO signifies its unique ability to perform object detection in just one iteration on the model graph, distinct from earlier methods. Since its introduction, the YOLO

lineage has undergone several iterations, with each iteration improving upon the earlier versions to rectify constraints and elevate efficiency, as illustrated in Figure 2 [25].

The real-time object detection process of YOLO has proven indispensable in self-driving vehicle systems, facilitating swift recognition and monitoring of diverse entities like cars, pedestrians, bicycles, and impediments. Such abilities have found utilization across a multitude of domains, encompassing tasks such as discerning actions within video sequences to facilitate surveillance, analyzing sports activities, and supporting interactions between humans and computers. Within the realm of medicine, YOLO has found application in the identification of cancer, delineating skin areas, and recognizing pills. This application has resulted in heightened precision of diagnoses and streamlined procedures for medical treatment. In the realm of remote sensing, YOLO has been enlisted for the purpose of recognizing and categorizing objects within satellite and aerial images. Yolo has also contributed to tasks such as charting land utilization, urban blueprinting, and overseeing ecological conditions. YOLO models find implementation in the agricultural sector to identify and categorize crops, pests, and diseases. This aids in the adoption of precision agriculture strategies and the automation of farming operations [25].

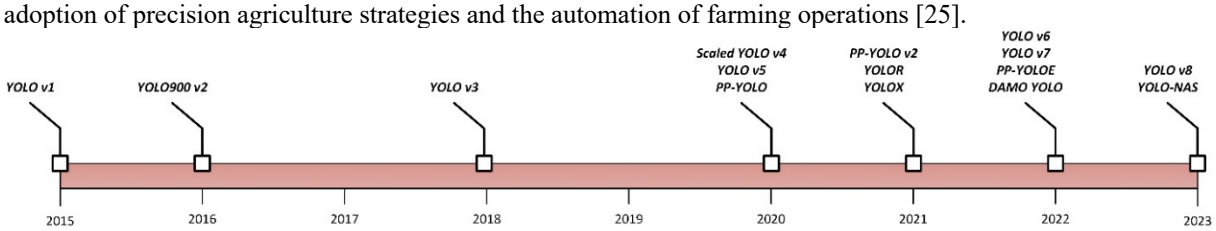



**Figure 2.** The emergence of lineage of YOLO models throughout the history [25]

The YOLO-v1 model integrates the stages of object detection, enabling the concurrent identification of all bounding boxes at a single scan. To achieve this, YOLO-v1 partitions the image as an S by S rectangles, anticipating B bounding boxes for a particular category and estimating the confidence associated with C distinct classes for each grid segment. YOLO-v1 employs a loss function, also used by the other YOLO models, that comprised sum of three sum-squared error components: one for the accuracy of bounding box coordinates (localization loss), another for object presence or absence confidence (confidence loss), and the third for the precision of category predictions (classification loss) [25].

Since the introduction of the YOLO-v3 model, the structure of YOLO models is delineated into three units: the “backbone”, “neck” and “head”. The “backbone” units are tasked with deriving valuable features from the input image and are generally constituted by a CNN that undergoes training in a comprehensive image classification assignment, like that of ImageNet. The intermediary unit, termed as the "neck," acts as a bridge connecting the “backbone” with the final unit “head”. Its role involves consolidating and honing the features extracted by the “backbone”, often emphasizing the refinement of spatial and semantic data across varying dimensions. Constituting the concluding unit of an object detection system, the "head" assumes the duty of generating forecasts using the features derived from the “backbone” and “neck” units. Generally, it comprises one or more subnetworks tailored to specific tasks, encompassing localization, classification and in more recent developments, duties such as pose estimation and instance segmentation.

The authors behind YOLO-v4 and YOLOR also introduced YOLO-v7 in July 2022. In YOLO-v7, some sort of alterations are introduced to the architecture along with a range of enhancements aimed at boosting accuracy. These changes are implemented through a collection of beneficial optimizations that improved precision while maintaining the inference speed, albeit impacting solely the duration of the training phase [25,27]. Yolo-v7 proposes new modifications in its architecture known as Extended Efficient Layer Aggregation Network (E-ELAN) and Model Scaling for Concatenation-Based Models.

ELAN serves as a technique that enhances the efficiency of deep models by managing the gradient flow across the shortest and longest pathways, thus facilitating more effective learning and convergence [25,28]. YOLO-v7 introduces the concept of E-ELAN, a mechanism designed to function effectively across models containing an unrestricted number of stacked computational blocks. E-ELAN enhances network learning without disrupting the original gradient trajectory by amalgamating features from various clusters through a process of shuffling and merging cardinality. In YOLO-v7, a novel approach for scaling models based on concatenation is also introduced. This method involves proportionally adjusting both the block's depth and width, ensuring the model's optimal structure is preserved [25,27].

The other proposed changes of Yolo-v7 model are as follows:

- The identity link within reparametrized convolutions (RepConv) is eliminated and referred to as RepConvN [25,27].
- The auxiliary head is set to receive coarse label assignment while the primary head is set to obtain precise label assignment [25,27].

- During the inference phase, the convolutional layer's bias and weight are adjusted to incorporate the mean and variance from batch normalization [25,27].

In this study, the Strawberry-DS image dataset is benchmarked by using the Yolo-v7 technique, employing its standard configuration with the sole exception being the adjustment of the default image dimensions from 640x640 to 960x960. This alteration is implemented to improve the detection potential of the model for smaller strawberries [29]. The training phase entails the execution of 600 epochs, a process executed on a workstation equipped with dual Nvidia RTX A4000 16GB GPUs, an Intel i7-11700F CPU operating at 3.6 GHz and 64 GB of RAM.

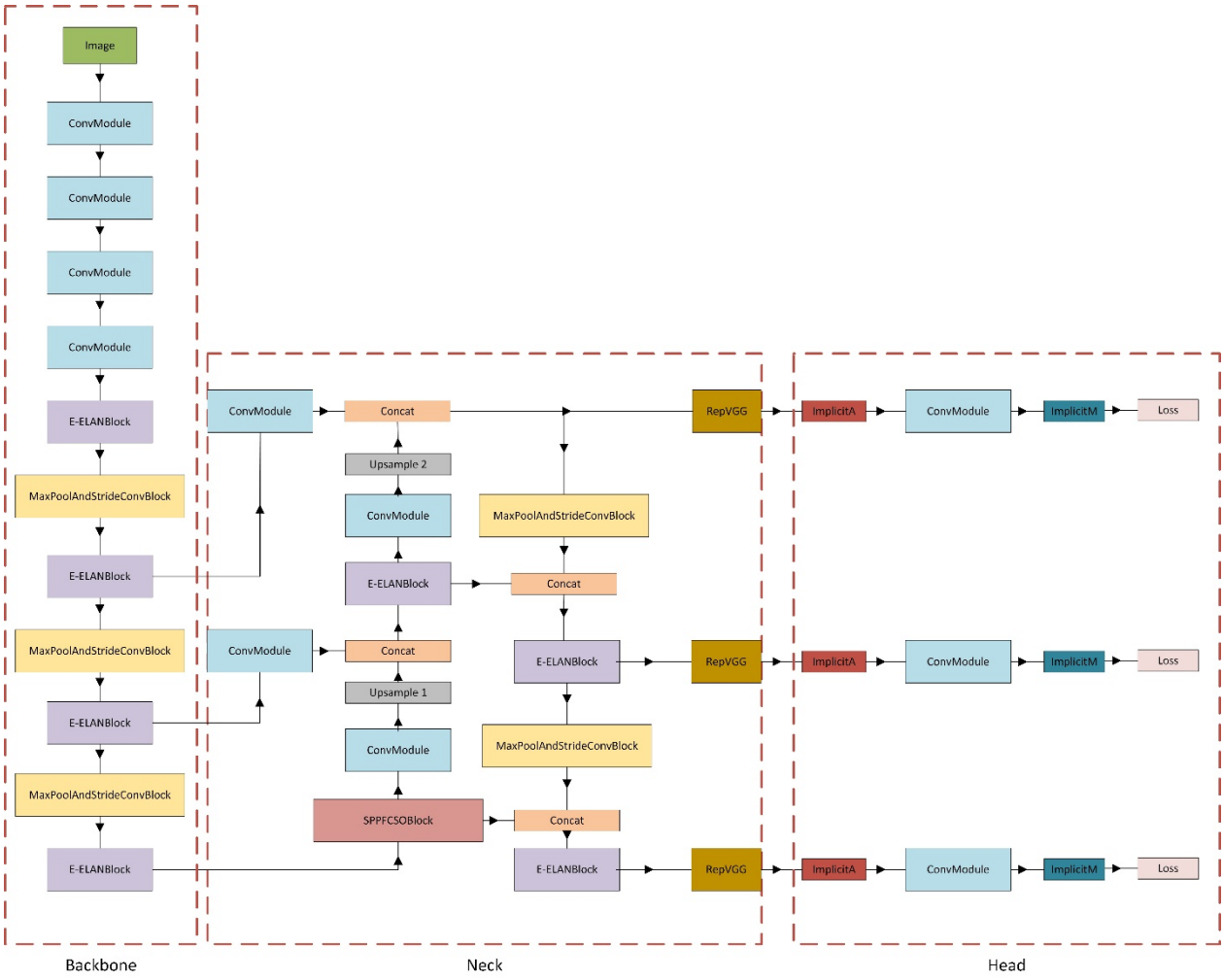



**Figure 3.** The main architecture of YOLO-v7 model [25,27]

## 4. Results & Discussion

In this section, the outcomes of the proposed study's training, validation, and testing phases are given. The outcomes of the validation phase, encompassing all categories, are outlined in Table 2. A total of 49 validation images have been examined during training, and 225 labels have been assigned to these images. Among these labels, the "green" class stands as the most prevalent, while the "early-turning" class is the least common. Equations 1, 2, and 3 provide the formulations for intersection over union (IoU), recall and precision metrics respectively. The average precision at an IoU value of 0.5 (AP@.5) is computed by calculating the area under the precision x recall curve as illustrated in Figure 5 (d) and equation 4 in which "p(r)" signifies the precision value linked to the indicated recall point on the horizontal axis. The term "mAP@.5" denotes the mean average precision for all classes at an IoU of 0.5. Furthermore, "mAP@.5:.95" signifies the average mean average precision along the IoU range of [0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95]. Higher IoU values tend to reduce false positives and

amplify false negatives, leading to heightened precision and diminished recall. This IoU impact strongly molds the precision x recall curve and governs mAP@IoU values as a secondary effect [30].

$$IoU = \frac{area(ground\ truth \cap prediction)}{area(ground\ truth \cup prediction)} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

$$AP@IoU = \int_0^1 p(r)dr \quad (4)$$

Table 2 reveals a noteworthy observation: the class "red," despite having fewer labels, attains the highest mAP@.5:.95 metric value. This outcome aligns with expectations, given that red strawberries exhibit a larger, reddish appearance, rendering them distinguishable against the backdrop of greenish leaves. In contrast, the class "late-turning" records the lowest mAP@.5: metric, owing to the lack of labeled ground truths and its resemblance to the “red” class which eventually poses a challenge the model to learn this subtle gradual color change.

In Figure 4, the confusion matrix, representing all validation dataset classes, unveils a distinct pattern. The dark blue diagonal values signify true positives, exposing a noteworthy revelation: the "late-turning" class records the smallest true positive value at 0.38, with 0.54 of all labeled "late-turning" ground truths being erroneously classified as background by the model. In comparison, the model overlooks none of the "early-turning" labels even the least number of labels are provided for this class.

**Table 2.** Performance results of validation dataset for each class

| Class | Images | Labels | Precision | Recall | mAP@.5 | mAP@.5:.95 |
|---|---|---|---|---|---|---|
| all | 49 | 225 | 0.689 | 0.671 | 0.66 | 0.526 |
| white | 49 | 54 | 0.7 | 0.648 | 0.583 | 0.441 |
| green | 49 | 104 | 0.707 | 0.719 | 0.726 | 0.473 |
| early-turning | 49 | 7 | 0.738 | 0.714 | 0.786 | 0.67 |
| turning | 49 | 10 | 0.541 | 0.6 | 0.507 | 0.423 |
| late-turning | 49 | 14 | 0.606 | 0.429 | 0.461 | 0.394 |
| red | 49 | 36 | 0.839 | 0.917 | 0.896 | 0.754 |

Figure 5 illustrates the recall, precision, precision x recall and F1 score curves relating to the validation dataset. Notably, the confidence level represents the model's degree of certainty in its predictions, distinct from the concept of IoU. When predictions coincide with the same ground truth label, the prediction boasting the highest confidence receives the true positive label, while the remaining predictions are classified as false positives. Elevating the confidence threshold results in heightened precision but reduced recall. This pattern is evident in Figure 5 a) and b), where adjustments in precision and recall values align as anticipated along the confidence level axis. In Figure 5 b), while the precision curves for all “red” class converge faster than the rest of the classes. The "late-turning" class's recall curve experiences an earlier decline compared to the delayed reduction in the "red" class's recall curve. The F1 score and precision x recall curves partially mirror the trajectory of the recall curve, as they integrate the precision and recall metrics as multipliers and axes, respectively.

Figure 6 presents a comprehensive view of the loss and performance metrics spanning epochs for both training and validation outcomes. In the leftmost column, there exists the localization loss for the predicted boxes [25]. Moving to the center column, the loss associated with the objectness of these predicted boxes is depicted, often referred to as confidence loss [25]. The third column portrays the classification loss. Notably, across the training and validation sets, all three losses exhibit a consistent decrease over epochs, barring the validation objectness loss, which exhibits a sign of overfitting by a spike around the 200th epoch. As we analyze the trend along the epoch axis, a discernible pattern emerges: the recall, precision, mAP@0.5, and mAP@0.5:0.95 metrics exhibit a steady rise before plateauing. This observation underscores the progressive improvement and eventual stabilization of these metrics over the course of training and validation epochs.

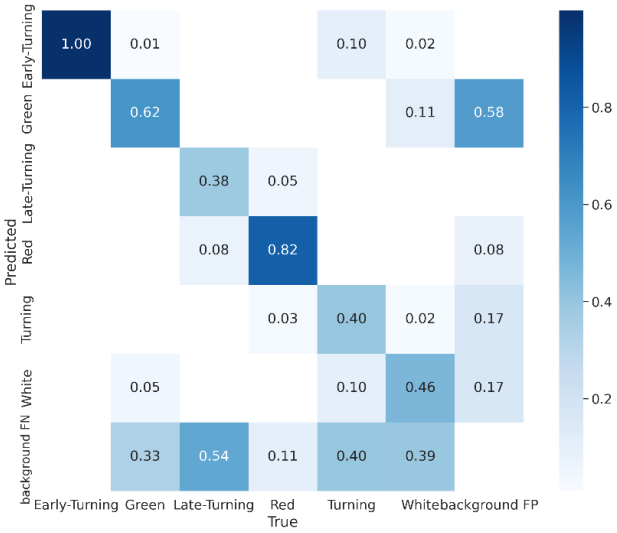
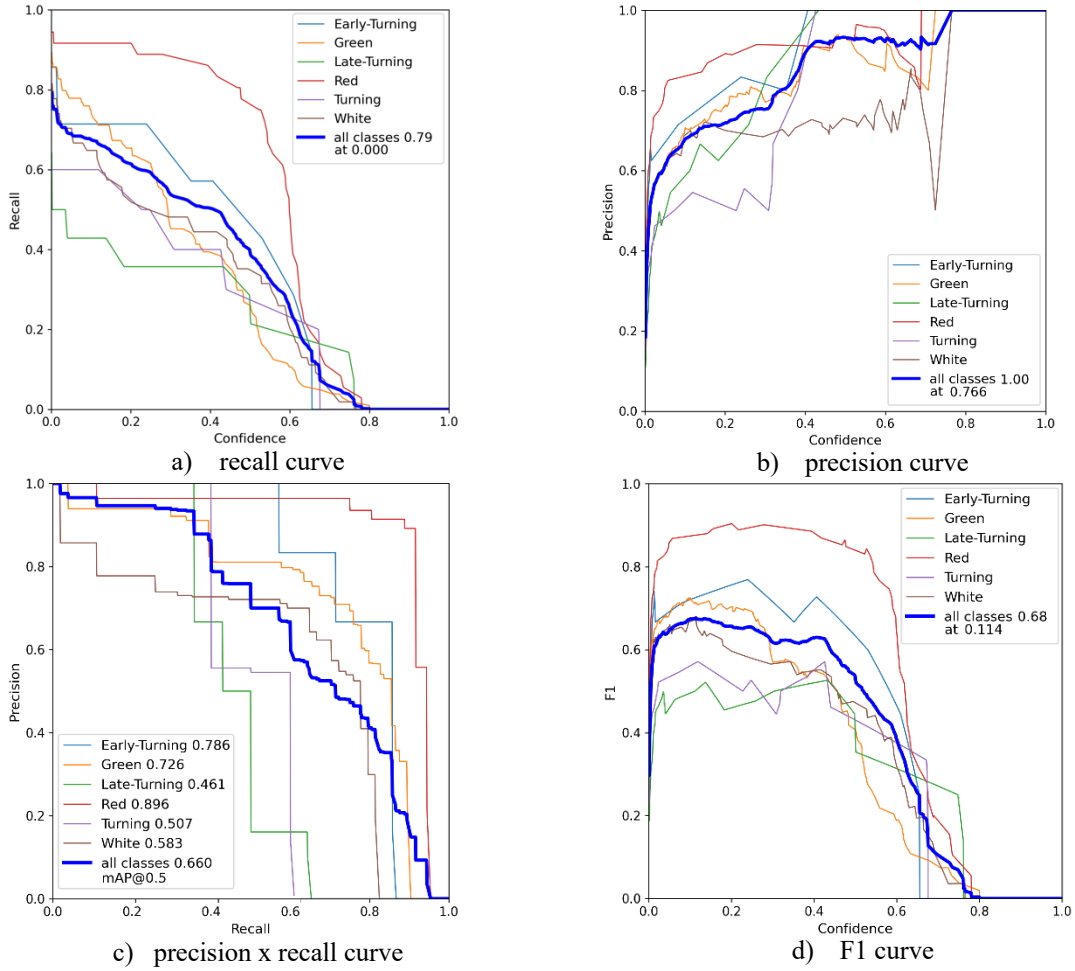



**Figure 4.** The confusion matrix of validation dataset




a) recall curve

b) precision curve

c) precision x recall curve

d) F1 curve

**Figure 5.** The recall, precision, precision x recall and F1 score curves of validation dataset

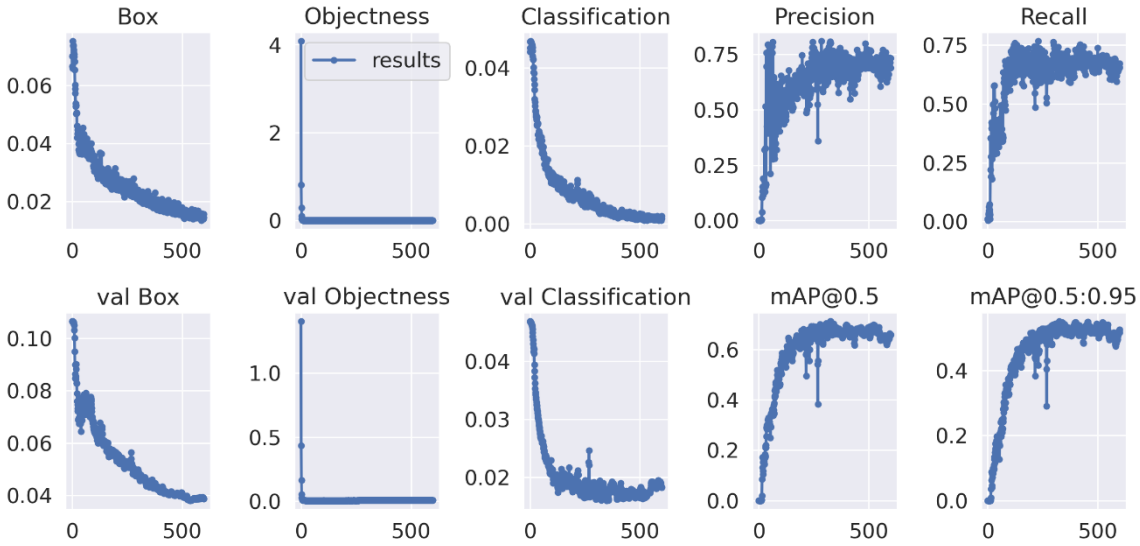



**Figure 6.** The performance results along the epoch number for training and validation datasets

A comprehensive breakdown of performance metrics across all classes pertaining to the test dataset are shared in Table 3. Notably akin to the validation performance outcomes within Table 2, both the "turning" and "red" categories exhibit the most impressive mAP@0.5 metric values, in contrast to the rest of the classes. The highest performance is obtained by “turning” class because of having only two labels in the test dataset and both of these cases are detected successfully. A higher number of labels of “turning” class would yield a more realistic result. Since the “white” and ”green” classes have more than 40 labels and their recall values are not high like “red” and “turning” classes, their mAP@0.5 values are obtained as 0.335 and 0.37 even if their precision values are around seventies. The validation and test results are obtained by a confidence level threshold of 0.25 [29].

The original test dataset have only 2 labels for “early-turning” class and the trained model is observed as not to detect them. In order to get a more balanced result, 2 images having “early-turning” class from the validation dataset is copied to the test data set and thus the total number of the labels are increased to 4. The relative decline of performance of “early-turning” class on the validation dataset with respect to test dataset is because of the noisy structure of the whole dataset especially between the transitionary classes like “early-turning”, ”turning”, ”late-turning”. This sort of noisy datasets are likely to be encountered in the machine learning data ecosystem. The noisy datasets are occasionally handled by the generalization capacity of the state of the art models. However, if some classes are represented only by a few noisy images than this issue should be focused and dataset should be monitored for these classes. Thus, in this study, 2 images from the validation dataset is copied to the test dataset to balance the “early-turning” class performance and obtain a more realistic result.

Another intriguing observation lies in the "turning" class, where the true positive count reaches to 1, and the instances erroneously labeled as background are minimized as 0 derived from the insights emerging from the test dataset's confusion matrix, thoughtfully presented in Figure 7. In tandem, Figure 8 provides a detailed visual representation of recall, precision, precision x recall and F1 score curves corresponding to the test dataset outcomes.

**Table 3.** Performance results of test dataset for each class

| Class | Images | Labels | Precision | Recall | mAP@.5 | mAP@.5:.95 |
|---|---|---|---|---|---|---|
| all | 28 | 119 | 0.792 | 0.575 | 0.558 | 0.46 |
| white | 28 | 42 | 0.631 | 0.381 | 0.335 | 0.27 |
| green | 28 | 47 | 0.769 | 0.426 | 0.37 | 0.228 |
| early-turning | 28 | 4 | 1 | 0.5 | 0.505 | 0.455 |
| turning | 28 | 2 | 1 | 1 | 1 | 0.85 |
| late-turning | 28 | 3 | 0.5 | 0.333 | 0.337 | 0.303 |
| red | 28 | 21 | 0.85 | 0.81 | 0.804 | 0.653 |

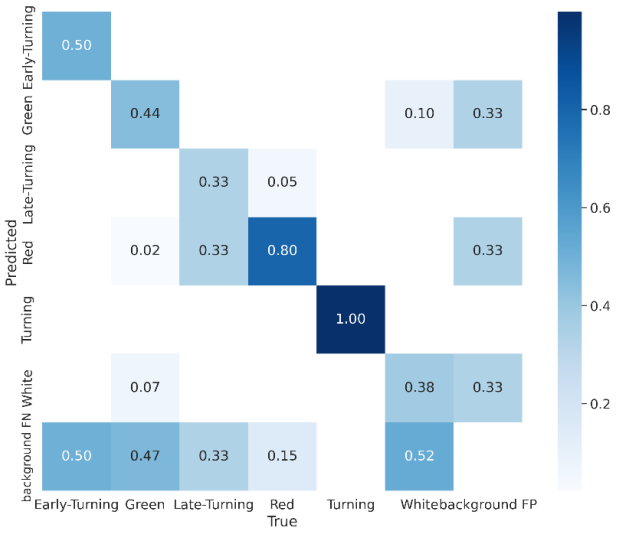



**Figure 7.** The confusion matrix of test dataset

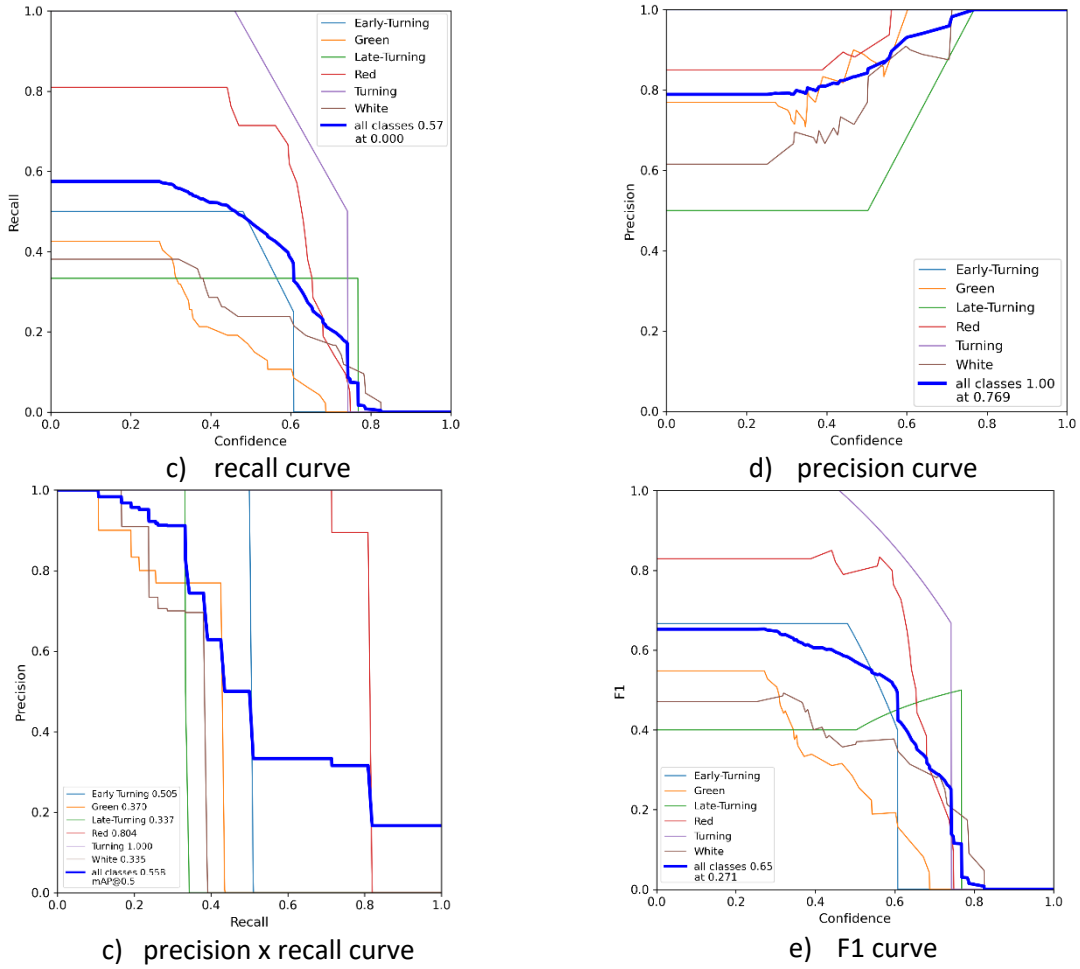



**Figure 8.** The recall, precision, Precision x Recall and F1 score curves of test dataset

In certain instances, the same images exhibit varying annotations, reflecting the complexities of accurately categorizing strawberries during these intermediary stages. This underscores the need for continual refinement in

both dataset annotation and model development to enhance the robustness and reliability of our detection and classification system. Some instances within the dataset exhibit inconsistent annotations, where identical plants have been labeled differently as shown in Figure 9. This inconsistency introduces challenges to the training and testing processes, as the model must tolerate conflicting labels to establish accurate patterns for classification.

In our study, the sampled images consistently demonstrated impressive detection results for strawberries as shown in Figure 10. For these sampled images, the Yolo-v7 algorithm efficiently identified and localized strawberries within the images, accurately outlining their boundaries with minimal false positives. This level of precision showcases the effectiveness of our approach in robustly identifying and distinguishing strawberries from their background, underscoring its potential for real-world applications in agriculture and food processing.

In our strawberry dataset, it's also important to acknowledge the presence of labeling errors that occasionally affect the accuracy of classification performance and localization boxes. Despite these errors, our model has demonstrated a remarkable ability to still detect many strawberries correctly, often classifying these correctly detected labels as false positives due to the misaligned localization coordinates as shown in Figure 11. Furthermore, the dataset's transitionary classes, such as "early-turning," "turning," and "late-turning," present challenges due to ambiguous labels as shown in Figure 11 e) and f).

The impact of extending the training process to 600 epochs in comparison to 100 epochs becomes evident when examining the background false negative values in their respective confusion matrices, as outlined in Table 4. This comparison provides insights into how the model's ability increases to correctly identify the strawberries from the background, particularly those initially missed, evolves as the training duration increases. By analyzing these values, we can gain a deeper understanding of the model's enhanced capacity to distinguish between strawberries and the background, thus highlighting the significance of prolonged training for improved detection accuracy.


a) Ground truth error of validation dataset

b) Ground truth error of validation dataset

**Figure 9.** Different labeling for identical plants in the Strawberry-DS dataset.

**Table 4.** Comparison of background false negative values for 100 and 600 epochs training

| epochs | early-turning | green | late-turning | red | turning | white |
|---|---|---|---|---|---|---|
| 100 | 1.0 | 0.9 | 1.0 | 0.08 | 1.0 | 0.8 |
| 600 | 0.5 | 0.47 | 0.33 | 0.15 | 0.0 | 0.52 |

### 4.1. Limitations and future work

The potential limitations of this research can be listed as follows:

- Data diversity and generalization capability of model: The trained model by Strawberry-DS dataset might struggle to generalize to entirely new and unforeseen conditions, underrepresented classes and different weather as well as lighting conditions.

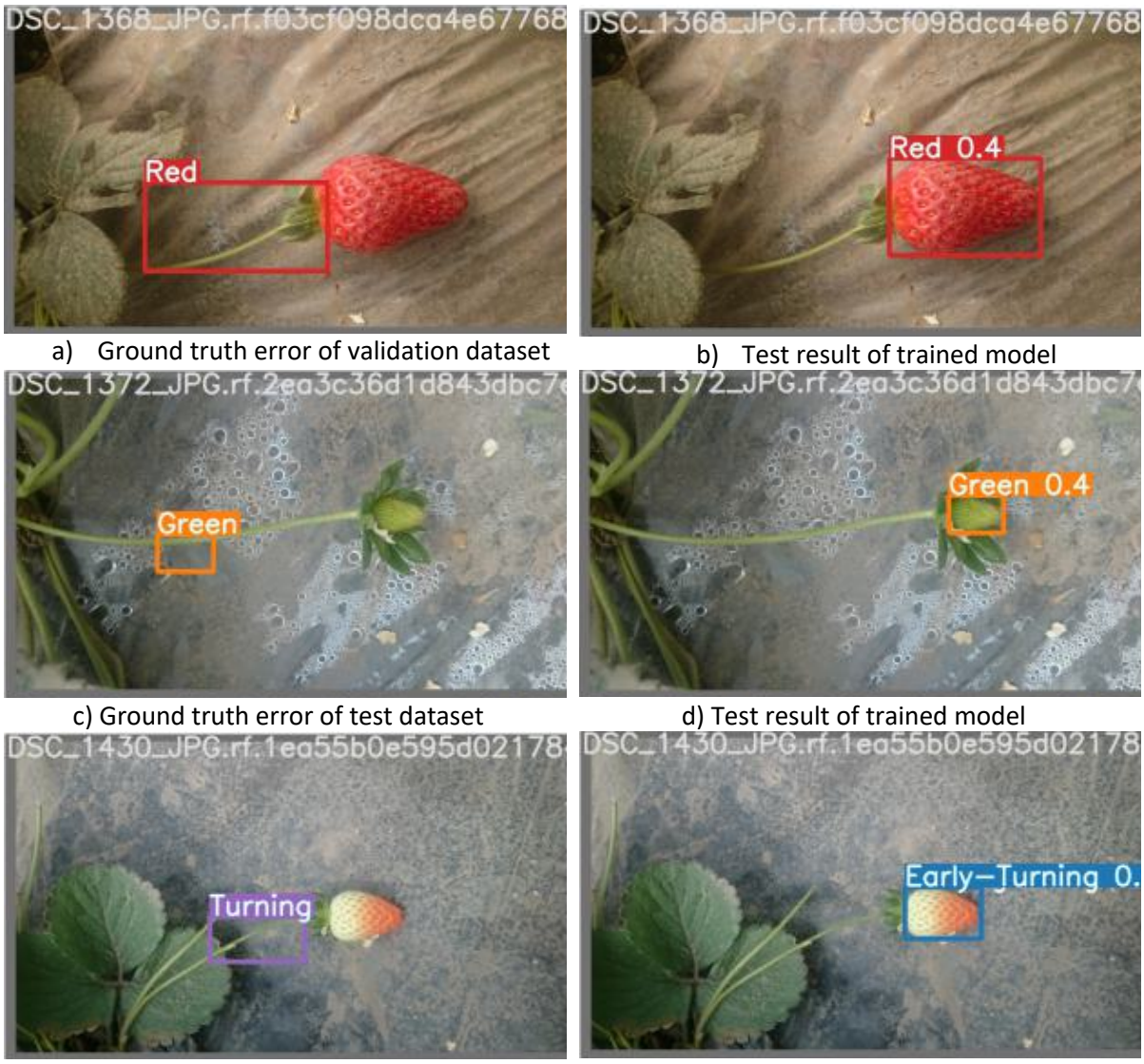

**Figure 10.** Some sampled images having successful detection results.

- Ambiguity in developmental phases: The performance of the model may be affected when confronted with ambiguous or transitional phases which do not fit neatly into predefined categories.
- Real-time performance and computational resource constraints: In case of deploying the trained model to the environments like edge devices, there may occur some extra limitations in terms of processing power, memory, and energy consumption, which could affect the feasibility of deployment.

The potential improvements of this research can be outlined in the subsequent manner:

- Improved data collection and annotation: Future research can include more other diverse and comprehensive datasets, including extreme scenarios that may not have been previously considered.
- Adapting to the edge devices: The further research direction can focus on optimizing the YOLO-v7 model specifically for edge devices by using techniques like model compression and quantization etc.

- Collaboration of human and AI: In order to tackle the ambiguous phases arises, human input feedback loops and active learning based approaches can be employed to continuously improve model performance.




a) Ground truth error of validation dataset
b) Test result of trained model
c) Ground truth error of test dataset
d) Test result of trained model
e) Ground truth error of validation dataset
f) Test result of trained model

**Figure 11.** Some sampled localization based labeling errors in the Strawberry-DS dataset.

## 5. Conclusion

In this proposed study, an automated object detection mechanism is benchmarked to precisely localize and monitor strawberry growth phases. The six classes representing the growth phases of 247 images of Strawberry-DS dataset is localized and classified via the Yolo-v7 model. The classified images are constitute of diverse capturing perspectives and angles as well as imbalanced and noisy labels which stands as an obstacle for an optimal training and testing tasks. The obtained mAP@.5 values for the distinct stages of ripeness stand as follows: 0.37, 0.804, 0.335, 1.0, 0.505 and 0.337 respectively for "green", “red”, "white", "turning", "early-turning" and "late-turning". The holistic performance metrics across all classes exhibit 0.792, 0.575, 0.558 and 0.46 respectively for precision, recall, mAP@.5 and mAP@.5:.95. Employing the Yolo-v7 methodology for object detection, the study effectively enables the identification of strawberries across their varying growth phases and the classification of these phases even with a relatively small imaging dataset.

## Data Availability

The public Strawberry-DS dataset is publicly available and can be reached from https://data.mendeley.com/datasets/z6dtfdpzz8/1

## References


[1] Li Y, Xue J, Zhang M, Yin J, Liu Y, Qiao X, Zheng D, Li Z. YOLOv5-ASFF: A Multistage Strawberry Detection Algorithm Based on Improved YOLOv5. Agronomy 2023; 13(7): 1901.

[2] Baby B, Antony P, Vijayan R. Antioxidant and anticancer properties of berries. Crit. Rev. Food Sci. Nutr 2018; 58(15): 2491–2507.

[3] Zhou C, Hu J, Xu Z, Yue J, Ye H, Yang G. A Novel Greenhouse-Based System for the Detection and Plumpness Assessment of Strawberry Using an Improved Deep Learning Technique. Front. Plant Sci. 2020; 11, 559: 1–13.

[4] He Z, Khana SR, Zhang X, Karkee M, Zhang Q. Real-time Strawberry Detection Based on Improved YOLOv5s Architecture for Robotic Harvesting in open-field environment. *arxiv.org* 2023; [Online]. Available: http://arxiv.org/abs/2308.03998.

[5] Charlton D, Castillo M. Potential Impacts of a Pandemic on the US Farm Labor Market. Appl. Econ. Perspect. Policy 2021; 43(1): 39–57

[6] Lemsalu M, Bloch V, Backman J, Pastell M. Real-Time CNN-based Computer Vision System for Open-Field Strawberry Harvesting Robot. IFAC-PapersOnLine 2022; 55(32): 24–29.

[7] Baygin M, Tuncer T, Dogan S. New pyramidal hybrid textural and deep features based automatic skin cancer classification model: Ensemble DarkNet and textural feature extractor. arxiv.org 2022; [Online]. Available: http://arxiv.org/abs/2203.15090.

[8] Yaman O, Tuncer T. Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images. Biomed. Signal Process. Control 2022; 73:103428.

[9] Baygin M, Yaman O, Barua PD, Dogan S, Tuncer T, Acharya UR. Exemplar Darknet19 feature generation technique for automated kidney stone detection with coronal CT images. Artif. Intell. Med. 2022; 127:102274.

[10] Yaman O, Tuncer T. Bitkilerdeki Yaprak Hastalığı Tespiti için Derin Özellik Çıkarma ve Makine Öğrenmesi Yöntemi. Fırat Üniversitesi Mühendislik Bilim. Derg. 2022; 34(1): 123–132.

[11] Fırat H. Sıkma - Uyarma Artık Ağı kullanılarak Beyaz Kan Hücrelerinin Sınıflandırılması Classification of White Blood Cells using the Squeeze- Excitation Residual Network. Bilişim Teknolojileri Dergisi 2023; 16(3):189–205.

[12] Li S, Zhang S, Xue J, Sun H. Lightweight target detection for the field flat jujube based on improved YOLOv5. Comput. Electron. Agric. 2022; 202:107391.

[13] Qiao Y, Guo Y, He D. Cattle body detection based on YOLOv5-ASFF for precision livestock farming. Comput. Electron. Agric. 2022; 204:107579.

[14] Koirala A, Walsh KB, Wang Z, McCarthy C. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO. Precis. Agric. 2019; 20(6): 1107–1135.

[15] Ji W, Pan Y, Xu B, Wang J. A Real-Time Apple Targets Detection Method for Picking Robot Based on ShufflenetV2-YOLOX. Agric. 2022; 12(6): 1–23.

[16] Lawal OM. Development of tomato detection model for robotic platform using deep learning. Multimed. Tools Appl. 2021; 80(17): 26751–26772.

[17] Montoya-Cavero LE, Torres RDL, Espinosa AG, Cabello JAE. Vision systems for harvesting robots: Produce detection and localization. Comput. Electron. Agric. 2022; 192: 106562.

[18] Lawal OM. YOLOMuskmelon: Quest for fruit detection speed and accuracy using deep learning. IEEE Access 2021; 9:15221–15227.

[19] Fu X, Li A, Meng Z, Yin X, Zhang C, Zhang W,Qi L. A Dynamic Detection Method for Phenotyping Pods in a Soybean Population Based on an Improved YOLO-v5 Network. Agronomy 2022; 12(12): 3209.

[20] Lawal OM. Study on strawberry fruit detection using lightweight algorithm. Multimed. Tools Appl. 2023; [Online]. Available: https://doi.org/10.1007/s11042-023-16034-0.

[21] Mao DH, Sun H, Li XB, Yu XD, Wu JW, Zhang QC. Real-time fruit detection using deep neural networks on CPU (RTFD): An edge AI application. Comput. Electron. Agric. 2023; 204:107517.

[22] Mejia G, Oca AM, Flores G. Strawberry localization in a ridge planting with an autonomous rover. Eng. Appl. Artif. Intell. 2022; 119:105810.

[23] Ren G, Wu T, Lin T, Yang L, Chowdhary G, Ting KC, Ying Y. Mobile robotics platform for strawberry sensing and harvesting within precision indoor farming systems. J. F. Robot. 2023; [Online]. Available: https://doi.org/10.1002/rob.22207.

[24] Elhariri E, El-Bendary N, Saleh SM. Strawberry-DS: Dataset of annotated strawberry fruits images with various developmental stages. Data Br. 2023; 48:109165.

[25] Terven J, Cordova-Esparza D. A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond. arxiv.org 2023; 1–33. [Online]. Available: http://arxiv.org/abs/2304.00501.

[26] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit; 26 June-1 July 2016; Las Vegas, NV, USA: 779–788.

[27] Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arxiv.org 2022; 1–15. [Online]. Available: http://arxiv.org/abs/2207.02696.

[28] Wang CY, Liao HYM, Yeh IH. Designing Network Design Strategies Through Gradient Path Analysis. arxiv.org 2022; [Online]. Available: http://arxiv.org/abs/2211.04800.
[29] "GitHub - WongKinYiu/yolov7: Implementation of paper - YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors". [Online]. Available: https://github.com/WongKinYiu/yolov7 (accessed Aug. 13, 2023).
[30] Padilla R, Passos WL, Dias TLB, Netto SL, Da Silva EAB. A comparative analysis of object detection metrics with a companion open-source toolkit. Electron. 2021; 10(3): 1–28.