Düzce University Journal of Science&Technology, 10 (2022) 1589-1602

Düzce University
Journal of Science & and Technology

Research Article

Strawberry Ripeness Assessment Via Camouflage-Based Data Augmentation for Automated Strawberry Picking Robot

Ferhat SADAK [a,*]

[a] *Department of Mechanical Engineering, Bartin University, Bartin, 74100, Turkey*
** Corresponding author's email address: fsadak@bartin.edu.tr*
DOI: 10.29130/dubited.1075572

ABSTRACT

Vision-based strawberry picking and placing is one of the main objectives for strawberry harvesting robots to complete visual servoing procedures accurately. Occlusion is the main challenge in strawberry ripeness detection for agriculture robots. In this study, strawberry ripeness detection was proposed using a camouflage-based data augmentation strategy to simulate the natural environment of strawberry harvesting conditions. Yolov4, Yolov4 tiny and Yolov4 scaled, and their traditional data augmentation and camouflage-based data augmentation derivatives were used to find out the effect of camouflage-based augmentation technique in overcoming the occlusion issue. Then the results were mainly evaluated based on mean Intersection over Union (IoU), F-1 score, average precision (AP) for ripe and unripe strawberries and frame per second (fps). Yolov4 tiny with camouflage-based data augmentation technique has demonstrated superior performance in detecting ripe and unripe strawberries with 84% IoU accuracy ~99% AP for ripe and unripe strawberries at an average of 206-fps, satisfying the agriculture strawberry harvesting robot operation need. The performance of the suggested technique was then tested successfully using a dataset termed the challenge dataset in this study to demonstrate its performance in a complex and occluded strawberry harvesting environment. Camouflage-based data augmentation technique helps to increase the detection procedure of ripe and unripe strawberries toward autonomous strawberry harvesting robot.

Keywords: *deep learning, Yolov4, data augmentation, strawberry ripeness detection, harvesting robot*

Otonom Çilek Toplama Robotu İçin Kamuflaj Tabanlı Veri Artırma Yoluyla Çilek Olgunluğu Değerlendirmesi

ÖZET

Görüye dayalı çilek toplama ve yerleştirme, çilek hasat robotlarının görsel servo prosedürlerini doğru bir şekilde tamamlaması için ana hedeflerden biridir. Tarım robotları için çilek olgunluğunun saptanmasındaki ana zorluk oklüzyondur. Bu çalışmada, çilek hasat koşullarının doğal ortamını simüle etmek için kamuflaj tabanlı bir veri artırma stratejisi kullanılarak çilek olgunluğu tespiti önerilmiştir. Kamuflaj tabanlı veri artırma tekniğinin oklüzyon sorununun üstesinden gelmedeki etkisini bulmak için Yolov4, Yolov4 tiny ve Yolov4 scaled ve bunların geleneksel veri artırma ve kamuflaj tabanlı veri artırma türevleri kullanılmıştır. Daha sonra sonuçlar esas olarak ortalama kesişim değeri (IoU), F-1 skoru, ortalama hassaslık (AP), ve saniyedeki kare sayısı (fps) temel alınarak değerlendirilmiştir. Kamuflaj tabanlı veri artırma tekniğine sahip Yolov4 tiny, ortalama 206 fps'de olgun ve olgunlaşmamış çilekler için %84 IoU doğruluğu ~%99 AP ile olgun ve olgunlaşmamış çilekleri tespit etmede üstün performans göstererek, tarım çilek hasat robotu operasyon ihtiyacını karşılamıştır. Önerilen tekniğin performansını çilek hasat ortamında göstermek için bu çalışma kapsamında belirlenen bir veri seti kullanılarak başarılı bir şekilde test edilmiştir. Kamuflaj tabanlı veri artırma tekniği, otonom çilek hasat robotları için olgun ve olgunlaşmamış çileğin tespit prosedürünü artırmaya yardımcı olmuştur.

Received: 18/02/2022, Revised: 14/03/2022, Accepted: 20/04/2022

Keywords: *Derin öğrenme, Yolov4, Veri artırımı, çilek olgunluğu tespiti, hasat robotu*

I. INTRODUCTION

Deep learning's rapid advancement has brought considerable convenience to our lives, and academics in the area are demanding to contribute to its development. Agricultural robot technology is also fast-evolving, and deep learning offers a wide range of possibilities for effective fruit detection [1]. Strawberry harvesting is the most time-consuming and labor-intensive part of the production process. Because manual harvesting does not meet the basic needs of the strawberry industry due to low production rates and a labour shortage, which has a negative impact on the industry's long-term viability and sustainability. Researchers have recently developed prototypes of commercial strawberry harvesting robots [2]–[4]. Target recognition and precise position detection (localization) of strawberry harvesting robots are difficult due to the complexity of the farm's natural environment and the unstructured features of the fruits. The natural factors include the intensity of the natural light, overlap of few fruits and occlusion of leaves. Therefore new solutions are necessary to overcome these challenges for developing autonomous strawberry harvesting robots [5].

Fruit ripeness detection has been investigated for various fruits to aid fully autonomous harvesting agriculture robots in making decisions about whether to pick the fruit. For instance, the colour and size values retrieved from RGB images, including 40 images for each of three stages, acquired using a computer vision system were used to determine banana maturity [6]. A total of 925 Cape gooseberry (physalis peruviana) fruit samples were gathered and manually graded into one of seven different maturity grades approaches to classify the ripeness of Cape gooseberry using three colour spaces (RGB, HSV, and L*a*b*) and the combination of four machine learning techniques [7].

To classify the different ripeness stages of tomatoes based on colour features, researchers employed an automated multi-class classification technique that included Principal Components Analysis (PCA), Support Vector Machines (SVMs), and Linear Discriminant Analysis (LDA) algorithms [8]. The Fuzzy Rule-Based Classification method (FRBCS) was used in a study to categorize the different ripeness stages of tomatoes, with six classes representing the six different phases of tomato ripeness [9]. The suggested FRBCS achieved a tomato ripeness classification accuracy of 94.29%. Based on the improving Otsu adaptive threshold algorithm and a novel feature in OHTA color space, an autonomous extraction method of strawberry maturity was examined in a complicated agricultural setting to assist fruit picking robot with a 95% extraction accuracy [10].

Deep learning (DL) approaches are currently widely recognized as the most promising technology for a wide range of computer vision applications, including medical robotics [11] and agricultural robotics [12], because they produce a superior performance to traditional machine learning techniques [13]. Faster-RCNN was used by Fanfang Gao et al. to detect apples, with an average precision of 0.909, 0.899, 0.858, and 0.848 for non-occluded, leaf-occluded, branch/wire-occluded, and fruit-occluded fruit, respectively [14]. The average precision in the fruit occluded scenario dropped to 6.1 % compared to the non-occluded example, demonstrating the importance of occlusion and its negative consequences in agricultural harvesting robotics. The Faster R-CNN was also used to detect apples, mangoes, and almonds, and the average accuracy results show that using this method can improve the accuracy of fruit recognition if data augmentation techniques such as flip and scale augmentation are used, as opposed to transferring the weight between orchards [15]. However, the processing time for Faster-RCNN is not sufficient in real-time operation that can be used in agriculture robotics.

You Only Look Once (YOLO) [16] and its improved versions are currently in high demand in various research fields, including agriculture robotics. YOLO's concept is to solve target detection as a regression task. Different research teams, for example, proposed apple detection based on an improved Yolov3 model [17], tomato detection based on a modified Yolov3 [18], and MangoYOLO based on Yolov3 and Yolov2(tiny) [19]. According to the Yolov4 model, occlusion performance is more accurate. As a result, more attention must be paid to the Yolov4 model [20] and its derivations to detect mature and immature strawberries for robotic harvesting tasks accurately. For instance, a deep convolutional neural network was used to distinguish between mature and immature strawberries in a greenhouse. It was discovered that the mature class had a problem detecting occluded strawberries. In contrast, the immature class was confused by the backdrop colour [21]. Hence, the development of YOLO architecture as a feature extractor aside, attention is also given to the data augmentation techniques so that the model can be robustly used in the agriculture field.

To overcome these challenges above, while using the superior performance of deep learning techniques, the development of new data augmentation techniques were shown to be useful [22]–[24]. Kamilaris et al. reviewed 40 research papers dealing with the use of deep learning in the agriculture field and showed that 37% of the reviewed paper utilized data augmentation techniques [25]. This is because increasing the number of training images improves the overall learning procedure and performance and for generalization purposes, which is critical when developing a robust detection and classification algorithm for agricultural robots. However, the developmental stages of deep learning algorithms mostly tried to benefit from classical data augmentation techniques in agriculture. This includes various rotation, scaling and intensity of the object in [26], image mirroring and rotation in 90-degree increments [27], flip, scale, flip-scale and in [15], image cropping, resized by a factor of 0.75 in [28]. These data augmentation techniques are commonly used in various research fields, such as medical robotics, and were not developed solely to serve agriculture robot autonomy. Furthermore, none of these studies developed data augmentation techniques that considered the nature of the agriculture field, such as occlusion.

This paper proposes a camouflage-based data augmentation technique to be employed as a robust ripeness detection of a strawberry as a part of the autonomy of the strawberry harvesting robot. The technique is evaluated against traditional data augmentation techniques using Yolov4, Yolov4 tiny and Yolov4-scaled framework, and this technique can be generalized for any type of fruit to detect its ripeness stage. Later, experimental validations are presented to illustrate the generalization of the proposed technique.

II. MATERIALS AND METHOD

A. DATASET PREPARATION AND TRAINING

112 images which consist of 443 ripe strawberries and 306 unripe strawberries, were created from a strawberry harvesting field in Turkey with the size of 416 *416 pixels. The dataset was prepared under a natural environment to simulate the real-world condition for strawberry harvesting. 80% of the data was used for training procedures, while 20% was used for testing purposes to evaluate the developed model. Given the size of our dataset, this ratio is sufficient to demonstrate the performance of the trained neural network and its generalization. The dataset was annotated using the graphical image annotation tool LabelImg, developed in [29]. A farmer with more than 30 years of experience in strawberry planting from Turkey's Bartin province performed the annotation of ripe and unripe strawberries. When classified as ripe or unripe, the annotation essentially relies on human perception. The bounding box was carefully drawn to the strawberry's outer margin based on the farmer's instructions.

The training hyper-parameters were configured as follows: the number of max epochs was set to 6000, the learning rate was set to 0.001, the momentum was set to 0.949, and batch and subdivision were set to 64 and 16, respectively. The training procedure for all neural networks was carried out on Google Colaboratory. Google Colaboratory is a free Jupyter notebook environment that requires no installation and runs entirely in the cloud was used throughout our study. NVIDIA Tesla K80 GPU and 12 GB RAM were used with Google Colaboratory in all training procedures to make a reasonable comparison. Figure 1 depicts a strawberry sample dataset as well as its annotation.

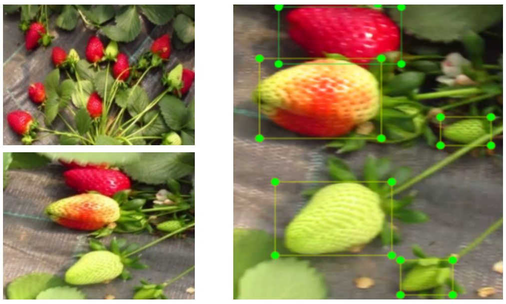
(a) (b)

Figure 1. ***(a)*** *Sample dataset of mature and immature strawberry* ***(b)*** *Sample annotation of mature and immature strawberry by LabelImg.*

B. YOLOV4

The Yolov4 model, a state-of-the-art real-time detection model, is an improved version of the Yolov3. The fundamental difference between the network structures of Yolov3 and Yolov4 is that in Yolov3, the DarketNet53 network is replaced with CSP DarketNet53, and the CSP DarketNet53 network, which is based on DenseNet, is employed as the backbone network in Yolov4. Additionally, the use of universal features such as Weighted-Residual-Connections (WRC), Cross-Stage-Partial-Connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT), and Mish-activation aids Yolov4 in achieving outstanding outcomes by increasing the learning capability of the Convolutional Neural Network (CNN). In comparison to Yolov3, Yolov4 is 12% faster and 10% more accurate.

The Yolo network approaches the detection problem as a regression problem, generating boundary coordinates and probability estimates for each class. If the center of the observed object falls within the artificially specified place, the network performs target detection using the trained Yolov4 loss functions, which include the bounding box location loss (L_{CIoU}), confidence loss($L_{confidence}$), and classification loss (L_{class}) [20].

$$Loss = L_{CIoU} + L_{confidence} + L_{class} \tag{1}$$

The CIoU loss, in comparison to the IoU loss, incorporates two new concepts: central point distance and aspect ratio measurement between ground truth and predicted bounding box, and uses these three metrics to improve the quality of the predicted bounding box.

$$CIOU = IOU - \frac{\rho^2(b,b^{gt})}{c^2} - \propto \upsilon \tag{2}$$

where $\rho^2(b, b^{gt})$ is the Euclidean distance between the prediction and real frames' center points, IOU defines the rate of overlap between the predicted and ground truth bounding box, and c signifies the diagonal distance of the smallest area that can contain both the prediction and real frames.

$$\propto = \frac{\upsilon}{1-IOU+\upsilon} \tag{3}$$

Here, $\propto$ is a positive trade-off parameter,

$$\upsilon = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{4}$$

and υ measures the consistency of the aspect ratio.
Where h is the height of the bounding box and w is the width of the bounding box.

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \propto \upsilon \quad (5)$$

C. YOLOV4 TINY

The compressed version of Yolov4 is Yolov4 tiny. It is based on Yolov4 and is intended to simplify the network structure and decrease parameters. This makes it possible to use it on mobile and embedded devices. Yolov4 tiny has two Yolo heads compared Yolov4 model, which has three heads. Furthermore, Yolov4 tiny was trained from 29 pre-trained convolutional layers, whereas Yolov4 was trained from 137 pre-trained convolutional layers, which directly impacts training and detection time. Instead of the CSPDarknet53 network used in the Yolov4 technique, the Yolov4-tiny approach employs the CSPDarknet53-tiny network as the backbone network. Instead of using the ResBlock module in the residual network, the CSPDarknet53- a small network employs the CSPBlock module in the cross-stage partial network. As a result, the gradient flow can propagate along two different network channels, increasing the gradient information's correlation difference.

The FPS (Frames Per Second) of Yolov4 tiny is almost eight times faster than Yolov4. The accuracy of Yolov4 tiny is two-thirds that of Yolov4 when tested on the MS COCO dataset. As a result, the Yolov4 model that should be used is determined by the application considering the model's accuracy and detection time. It should be noted that highly accurate and fast models are in high demand in the agricultural robotics industry.

D. YOLOV4 SCALED

Scaled Yolov4 is a set of neural networks built on top of the Yolov4 network, which has been enhanced and scaled without using any pre-trained model [30]. The Yolov4-scaled model employs a scaling strategy that alters the network's depth, width, resolution, and structure, such as image size, number of layers and channels while optimizing the model performance and inference speed. ResNet, ResNeXt, and the traditional Darknet backbone, as well as a few CSP-ized CNN backbones, are all considered in their network. Applying the concepts given out in the Cross-Stage Partial Networks is what it means to CSP-ize. The CSP is a new approach to designing CNNs that reduces computation by 50% for various CNN networks. In Yolov4 scaled, network architecture was improved by optimizing the backbone and Cross-stage-partial (CSP) connections and Mish activations. In contrast to Yolov4, where just one neural network was trained for all resolutions, a separate neural network is trained for each resolution of the network in Yolov4 scaled.

E. PERFORMANCE EVALUATION

Yolov4, Yolov4 tiny and Yolov4 algorithms were evaluated based on precision, recall, F-1 score, mAP@0.5, AP and Average IoU.

Precision is defined as the purity of positive detections relative to the ground truth, indicating how many of the predicted objects had a matching ground truth annotation.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall is described as the completeness of the positive predictions compared to the ground truth, answering what proportion of actual positives was correctly identified.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

The harmonic mean of the model's precision and recall is used to calculate the F-score.

$$F_1\ score = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

Intersection over Union is a metric used to evaluate the accuracy of an object detector on a given dataset. It is the rate at which the ground truth and prediction output overlap. Equation 9 shows the calculation of the IoU which *gt* refers to ground truth while *po* stands for prediction output.

$$IoU = \frac{Bgt \cap Bpo}{Bgt \cup Bpo} \quad (9)$$

The visual demonstration of the IoU calculation on an unripe strawberry as an example is shown in Figure 2.

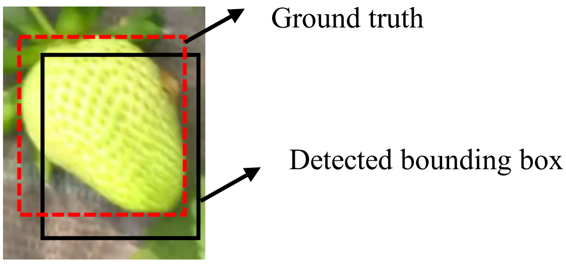

Figure 2. *Intersection over Union (IoU) calculation.*

Average precision (AP) is an evaluation metric for measuring the accuracy of the object detector. It is the measure of the area under the precision-recall curve. Finally, mAP@0.5 (AP with IoU = 0.50) is calculated as the average of the AP values for each class (ripe and unripe strawberry in our case).

F. DATA AUGMENTATION

In this study, eight traditional data augmentation techniques were applied: angle, saturation, exposure, hue, horizontal and vertical flip, Gaussian noise, and blur. 112 images were augmented for each augmentation procedure. The values and number of images set during the augmentation procedure in the Python environment are shown in Table 1.

Table 1 *The values and number of images set during the traditional augmentation procedure.*

	Angle [-30 30]	saturation =1.5	exposure = 1.5	hue=.1	horizontal flip=0.5	vertical flip=0.5	gaussian noise=1	blur=1
Number of augmented images	112	112	112	112	112	112	112	112

G. CAMOUFLAGE-BASED DATA AUGMENTATION TECHNIQUE

The camouflage-based data augmentation approach is a simple technique used in this study to describe the location of distinct strawberry leaves in five various shapes and sizes. 50 images of ripe and unripe strawberries were used to generate artificial data, then used to generate data from five distinct strawberry leaves. The leaf was chosen randomly to illustrate the most common occlusion scenario that deep learning systems may face. The primary idea behind this method is to expand the training dataset when the developed deep learning-based algorithms fail to recognize strawberries in complex scenarios. This is because typical data augmentation techniques such as scaling, and rotation are ineffective for detecting strawberries since they do not replicate the strawberries' natural state. The demonstration of the artificial data produced using five different strawberry leaf is shown in Figure 3.

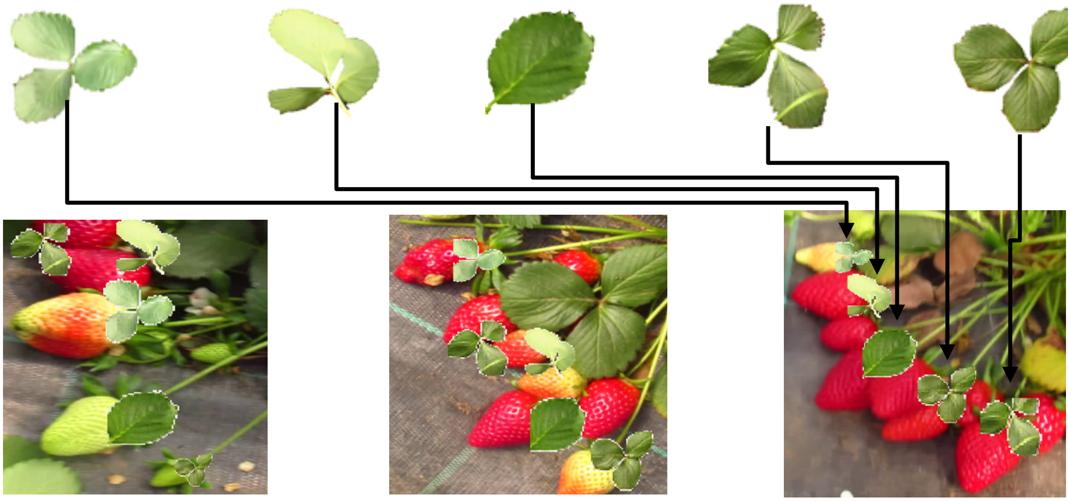

Figure 3. *Illustration of five different strawberry leaf and the produced artificial data.*

III. RESULTS

The results were evaluated based on precision, recall, F-1 score, AP-ripe strawberry, AP-unripe strawberry, mAP@0.5, Average IoU and frame per second (fps). Average IoU and fps are critical evaluation metrics since the overlap rate between ground truth and the detected bounding box for localization, and computational cost of detection for a real-time operation is a prerequisite for strawberry harvesting robots. The results were then revealed for Yolov4, Yolov4 tiny, Yolov4 scaled and Yolov4 with traditional data augmentation, Yolov4 tiny with traditional data augmentation, Yolov4 scaled with traditional data augmentation, Yolov4 with camouflage-based data augmentation, Yolov4 tiny with camouflage-based data augmentation and Yolov4 scaled with camouflage-based data augmentation. In this way, the usefulness of traditional data augmentation techniques and camouflage-based data augmentation techniques can be shown on state-of-the-art detection techniques, which are Yolov4, Yolov4 tiny and Yolov4 scaled in this study. Table 2 demonstrates the performance of deep learning models with different data augmentation approaches for detecting strawberry ripeness.

Table 2 *Performance evaluation of various deep models for strawberry ripeness detection.*

Parameters	Yolov4	Yolov4-tiny	Yolov4 -scaled	Yolov4 + Traditional data augmentation	Yolov4 tiny + Traditional data augmentation	Yolov4 scaled + Traditional data augmentation	Yolov4 + Camouflage-based data augmentation	Yolov4 tiny + Camouflage-based data augmentation	Yolov4 scaled+ Camouflage-based data augmentation
Precision	0.80	0.79	0.79	0.78	0.76	0.83	0.93	0.95	0.92
Recall	0.87	0.90	0.84	0.83	0.91	0.89	1.00	0.98	0.95
F-1 score	0.83	0.84	0.81	0.81	0.83	0.86	0.96	0.96	0.95
AP- ripe strawberry	0.9203	0.8989	0.8622	0.8859	0.9226	0.8896	0.9996	0.9849	0.9992
AP- unripe strawberry	0.8482	0.8215	0.7708	0.8768	0.8338	0.7944	1.00	0.9987	0.9987
mAP@0.5	0.8843	0.8602	0.8164	0.8813	0.8782	0.8420	0.9997	0.9918	0.9990
Average IoU	0.6311	0.6194	0.6333	0.6052	0.5980	0.6784	0.8067	0.8434	0.7863
Frame per second (fps)	30.45	206.61	33.19	30.64	205.88	33.25	30.62	206.73	33.45

On the left side of Figure 4, mAP@0.5, F-1 score, and Average IoU are evaluated and compared, and on the right side of Figure 4, AP for ripe and unripe strawberry is assessed and compared, as shown Figure 4.

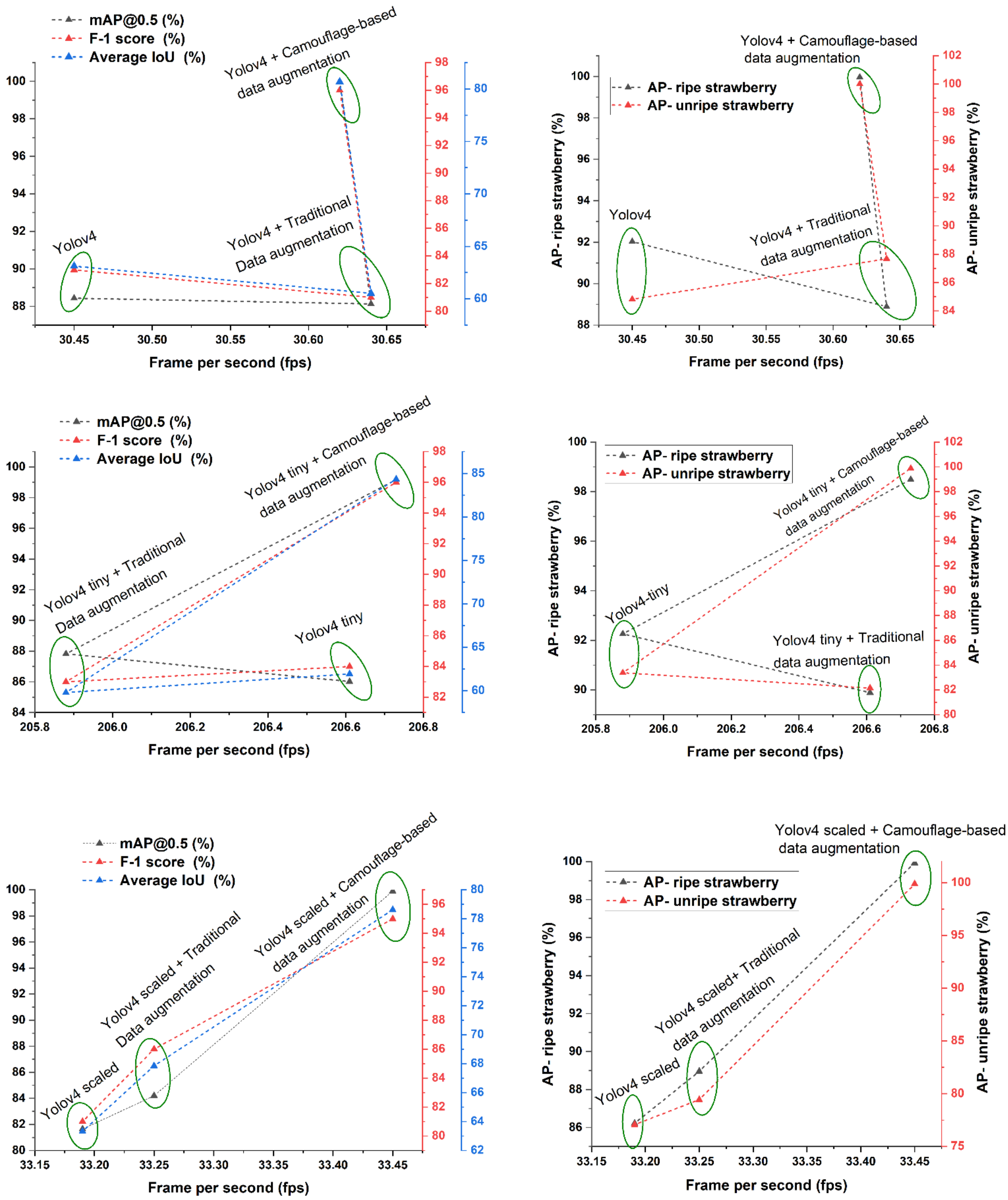

Figure 4. *Performance evaluation of the model based-on mAP@0.5, F-1 score, Average IoU and AP for ripe and unripe strawberry ripeness detection.*

In the following sections, the results will be examined, and a case study will be conducted to verify the effectiveness of the proposed augmentation approach in strawberry ripeness assessment task. Then the results will be compared with the literature to show its superior performance.

IV. DISCUSSION

As depicted from Table 2, F1 score for Yolov4 was calculated as 0.83, and its score was slightly decreased to 0.81 when the traditional data augmentation technique was used. On the other hand, its F1-score was increased from 0.83 to 0.96, representing a 13% rise. The AP for ripe strawberries was reduced by ~3%, and AP for unripe strawberries were increased by ~3%. Therefore, the mAP@0.5 was not changed between Yolov4 and Yolov4 using the traditional data augmentation technique. On the other hand, Yolov4 with a camouflage-based data augmentation approach has shown promising results for AP for ripe and unripe strawberry, 99 % and 100%, respectively. Compared to Yolov4 and Yolov4 with traditional data augmentation techniques, approximately 17% and 19% increase were obtained in AP for ripe and unripe strawberries. One of the most important phenomena resulted in the IoU metric. While IoU for Yolov4 and Yolov4 with traditional data augmentation technique resulted in 63%, its value has increased by ~17% and reached ~80% when the camouflage-based data augmentation is used along with Yolov4. Hence, the camouflage-based data augmentation strategy enhances the accuracy and detector's capacity to localize. No significant changes were observed for the fps parameter, which reached ~30 fps for Yolov4 and its augmentation derivatives.

F-1 score for Yolov4 tiny and Yolov4 with traditional data augmentation technique has only shown a 1% difference in favor of Yolov4 tiny. On the other hand, the F-1 score has reached ~83% to 96% when Yolov4 is used with the camouflage-based data augmentation technique. AP for ripe and unripe strawberries was also increased ~3% and ~1% when Yolov4 tiny is used traditional data augmentation. However, AP for ripe and unripe strawberry was increased to 9% and 17%, respectively, while almost reached 100% AP for both ripe and unripe strawberry classification when camouflage-based data augmentation is used. IoU score for Yolov4 tiny and Yolov4 with traditional data augmentation technique almost obtained 60%, while this score has increased to ~84%, which is nearly 24% difference compared to Yolov4 tiny and its traditional data augmentation derivative. This is a crucial stage in developing an autonomous agriculture robot since the precision of localization directly impacts picking and placing the strawberry crop by the robot's end effector. The fps score has nearly remained constant for all, reaching 206 fps for Yolov4 tiny and its augmented versions. Yolov4 and its data augmentation derivatives have over seven times greater fps score. This is due to the Yolov4 tiny's lighter neural network architecture.

F-1 score for Yolov4 scaled was obtained 81% and reached to 86% when it is used with data augmentation technique. On the other hand, its score has up to 14 % and reached a F-1 score of 95% when the camouflage-based data augmentation is used, demonstrating the effectiveness of this approach and its suitability within agriculture applications. AP for ripe and unripe strawberry was also increased ~2% for Yolov4 scaled with traditional data augmentation compared to Yolov4 scaled. However, AP for ripe and unripe strawberries was increased by 13% and almost reached to 100% AP for both ripe and unripe strawberry classification when Yolov4 is used with camouflage-based data augmentation technique. IoU score was increased from ~63% to ~67% when Yolov4 scaled is used with traditional data augmentation technique, while 78% IoU score was obtained when Yolov4 scaled is used with camouflage-based data augmentation technique. Fps score has recorded almost the same for all, which got to ~33 fps for Yolov4 scaled and its augmentation derivatives. This is approximately 3% higher than Yolov4 and its data augmentation derivatives and almost seven times less than Yolov4 tiny and its data augmentation derivatives.

Following the demonstration of the effectiveness of the camouflage-based data augmentation technique within each method, all models are evaluated and compared to each other based on mAP@0.5, F-1 score, and Average IoU on the left side of Figure 4, and AP for ripe and unripe strawberry on the right side of Figure 4. This is to determine the most promising ripeness detection neural network model and the data augmentation technique used. Figure 4 leads to the following conclusion.

- Although Yolov4 and Yolov4 scaled and their data augmentation derivatives demonstrate ~30 fps and 33 fps, respectively, Yolov4 tiny and its data augmentation derivatives illustrate almost seven times faster processing speed of ~206 fps. However, since all models are meeting the real-time operation to automate a strawberry harvesting robot, the fps score for all models included in their data augmented versions are all in an acceptable range.

- Yolov4 and Yolov4 tiny have shown similar behaviour when used with traditional data augmentation techniques. At the same time, YoloV4 scaled with traditional data augmentation technique were slightly superior to Yolov4 scaled in terms of mAP@0.5, F-1 score and Average IoU.
- AP for ripe and unripe strawberry detection by Yolov4, Yolov4 tiny and Yolov4 scaled, and their traditional data augmentation techniques were similar to each other with no notable differences. On the other hand, Yolov4, Yolov4 tiny and Yolov4 scaled models used with camouflage-based data augmentation technique were superior to other models used in AP for ripe and unripe strawberry detection.
- Overall, while differentiating between ripe and unripe strawberries, the camouflage-based data augmentation strategy enhanced the Yolov4, Yolov4 tiny, and Yolov4 scaled models in all evaluation criteria. Yolov4 tiny with camouflage-based data augmentation technique outperformed other models, achieving 84 % IoU and 99 % AP for ripe and unripe strawberries with a ~206 frames per second processing speed. These results are enough for use with the agricultural strawberry harvesting robot. Yolov4 tiny with camouflage-based augmentation approach can also be employed to the mobile phone due to its fast-processing speed.

Therefore, based on the analysis of results, Yolov4 tiny with camouflage-based data augmentation technique was selected to conduct further investigation and comparison with the literature.

A. A CASE STUDY ON CHALLANGE DATASET

To demonstrate the efficiency of the Yolov4 tiny model when combined with the camouflage-based data augmentation technique for strawberry ripeness identification, a challenge dataset was constructed. This collection contains 57 strawberry objects, 36 of which are ripe and 21 of which are unripe. The dataset was carefully designed to handle the occlusion and complicated scene of a strawberry harvesting field where an autonomous strawberry harvesting robot operates. Only one out of twenty-one unripe strawberries could not be detected, whilst all ripe strawberries were accurately detected. Table 3 demonstrates the detected ripe and unripe strawberries for each image used in the experiment.

Table 3 *Experimental results of correctly detected ripe and unripe strawberries.*

Image number	Number of ripe strawberries	Number of unripe strawberries	Number of correctly detected ripe strawberry	Number of correctly detected unripe strawberry
0	5	0	5	0
1	1	2	1	2
2	3	0	3	0
3	1	1	1	1
4	2	1	2	1
5	3	1	3	0
6	3	1	3	1
7	8	1	8	1
8	1	4	1	4
9	3	3	3	3
10	1	5	1	5
11	1	1	1	1
12	4	1	4	1

As demonstrated in Figure 5, the mean accuracy for ripe strawberries ranges from 92% to 100%, while the mean accuracy for unripe strawberries ranges from 51% to 100%. It should be noted that only images 6 and 11 showed less than 70% mean accuracy for unripe strawberries since the unripe strawberries in the image are hardly visible due to occlusion or are not entirely in the frame. However, the proposed method can still appropriately detect and classify strawberry ripeness.

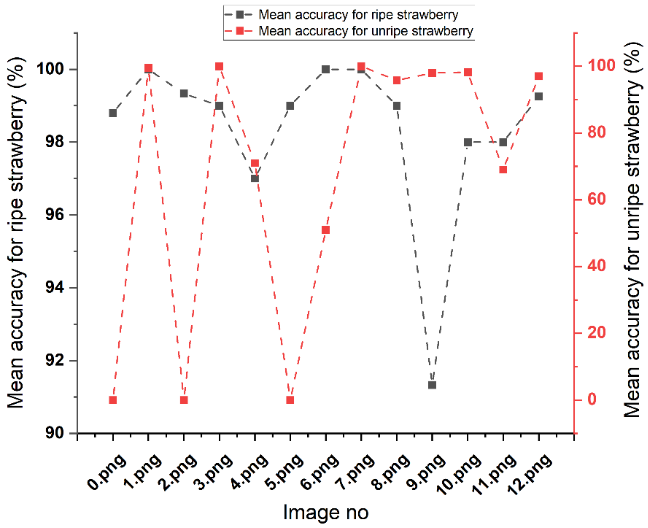

Figure 5 *Mean accuracy for ripe and unripe strawberry for each image.*

Figure 6 is depicted to demonstrate the effectiveness of the proposed approach. Figure 6 shows the promising results of the Yolov4 tiny with camouflage-based data augmentation technique and shows the complex scene of the strawberry harvesting circumstance. For instance, in Figure 6a, the unripe strawberry on the left side of the image is partially visible, while the leaf occludes the ripe strawberry on the right side. The method used in this study is not affected by the complex scene and occlusion by the leaf and successfully detects and classifies the strawberry. Occlusion can occur because of the strawberry itself. Strawberry can also obscure other strawberries, resulting in misdetection or misclassification. Figure 6f depicts eight ripe strawberries and one unripe strawberry, with the ripe strawberries stacked on top of each other and the unripe strawberry partially visible. The proposed approach again successfully detected and classified the ripe and unripe strawberries. The rest of the images in Figure 6 are also mostly occluded, or the strawberries are partially visible. Under this complex environment, the proposed approach can detect the ripeness of strawberries as a part of the visual-servoing of an autonomous strawberry harvesting robot.

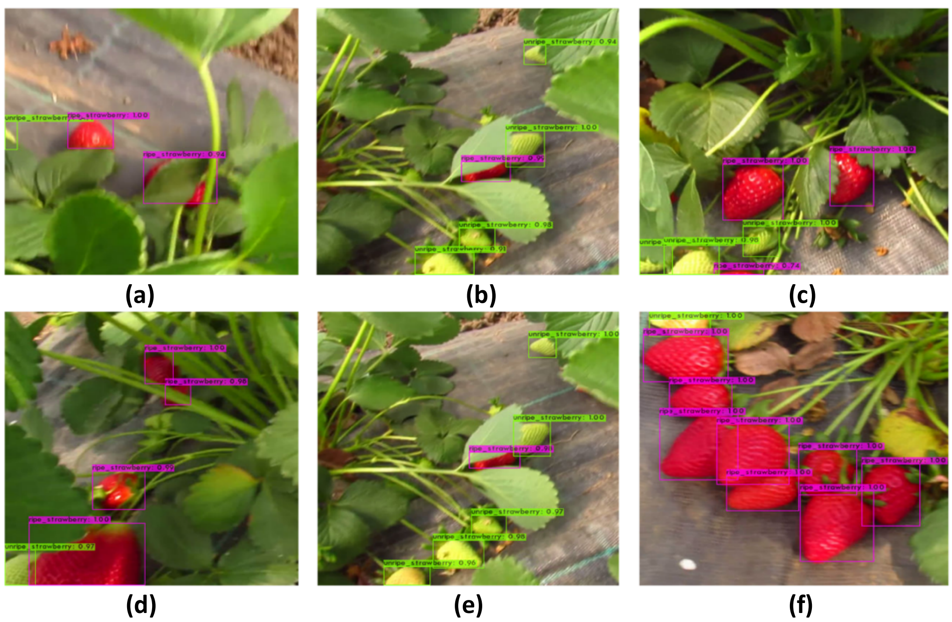

Figure 6 *Example of detection results of ripe and unripe strawberries from the challenge dataset.*

B. COMPARISON OF OUR WORK WITH LITERATURE

To the best of our knowledge, there is no publicly available benchmark to test strawberry ripeness detection. However, a few similar publications enable us to compare the proposed method in this study to the literature. Shao et al. [31] used portable hyperspectral imaging to determine strawberry maturity with a 96.7 % accuracy. However, the proposed method does not operate in real-time, and occlusion, the primary problem in strawberry ripeness assessment, was not prioritized. Habaragamuwa, on the other hand, provided a strawberry mature and immature detection framework utilizing a deep convolutional neural network, which is the first application of strawberry ripeness detection to be published in 2018 [21]. Occlusion is identified as the key issue and is excluded from the labeling process because it has reduced the AP by 50%. Our study has mostly focused on overcoming the occlusion issue to assist autonomous strawberry harvesting robots in picking the ripe strawberry while avoiding picking the unripe strawberry, regardless of whether the strawberries are occluded or not. This is consistent with the agricultural strawberry harvesting robot's real-time and natural operation conditions. This study has also shown that sufficient attention should be given to the data processing stage compared to the neural network architecture itself. The challenge dataset will be made accessible upon request so that researchers can put their approaches to the test in an obstructed strawberry harvesting scenario and compare the results provided in this study.

V. CONCLUSION

This study proposes a strawberry ripeness detection system through camouflage-based data augmentation technique. After the evaluation of the Yolov4, Yolov4 tiny and Yolov4 scaled, and their traditional data augmentation and camouflage-based data augmentation derivatives, Yolo tiny with camouflage-based data augmentation technique have established superior performance in detection of ripe and unripe strawberry with 84% IoU accuracy, ~99% AP for ripe and unripe strawberries at an average of 206-fps. Then, the performance of the proposed technique was tested on a challenging dataset and showed approximately above 92% to 100% mean accuracy for ripe strawberries. At the same time, the mean accuracy for unripe strawberries varies from ~51% to 100% due to occlusion or being partially in the frame for images 6 and 11. Real-time strawberry ripeness detection system in a complex and occluded strawberry harvesting environment was proposed to assist autonomous strawberry harvesting robots in completing visual servoing operations successfully. The proposed technique is not limited to strawberry harvesting robots but also can be used as a guide to be implementing it in other types of fruits. Future work will include an automated camouflage-based data augmentation framework in Python language specifically developed to assist for various fruit harvesting robotic applications that can be implemented automatically to the feature extractor used.

Funding: The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Conflicts of Interest: The author has no relevant financial or non-financial interests to disclose.

VI. REFERENCES

[1] N. Zhu *et al.*, "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *Int. J. Agric. Biol. Eng.*, vol. 11, no. 4, pp. 32–44, 2018, doi: 10.25165/j.ijabe.20181104.4475.

[2] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," *J. F. Robot.*, vol. 37, no. 2, pp. 202–224, 2020, doi: 10.1002/rob.21889.

[3] A. De Preter, J. Anthonis, and J. De Baerdemaeker, "Development of a Robot for Harvesting

Strawberries," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 14–19, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.08.054.

[4] Q. Feng, J. Chen, M. Zhang, and X. Wang, "Design and Test of Harvesting Robot for Table-top Cultivated Strawberry," in *2019 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*, 2019, pp. 80–85, doi: 10.1109/WRC-SARA.2019.8931922.

[5] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan, "Harvesting robots for high-value crops: State-of-the-art review and challenges ahead," *J. F. Robot.*, vol. 31, no. 6, pp. 888–911, 2014.

[6] D. Surya Prabha and J. Satheesh Kumar, "Assessment of banana fruit maturity by image processing technique," *J. Food Sci. Technol.*, vol. 52, no. 3, pp. 1316–1327, 2015, doi: 10.1007/s13197-013-1188-3.

[7] W. Castro, J. Oblitas, M. De-La-Torre, C. Cotrina, K. Bazán, and H. Avila-George, "Classification of Cape Gooseberry Fruit According to its Level of Ripeness Using Machine Learning Techniques and Different Color Spaces," *IEEE Access*, vol. 7, pp. 27389–27400, 2019, doi: 10.1109/ACCESS.2019.2898223.

[8] N. El-Bendary, E. El Hariri, A. E. Hassanien, and A. Badr, "Using machine learning techniques for evaluating tomato ripeness," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 1892–1905, 2015, doi: https://doi.org/10.1016/j.eswa.2014.09.057.

[9] N. Goel and P. Sehgal, "Fuzzy classification of pre-harvest tomatoes for ripeness estimation – An approach based on automatic rule learning using decision tree," *Appl. Soft Comput.*, vol. 36, pp. 45–56, 2015, doi: https://doi.org/10.1016/j.asoc.2015.07.009.

[10] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik (Stuttg).*, vol. 125, no. 19, pp. 5684–5689, 2014, doi: https://doi.org/10.1016/j.ijleo.2014.07.001.

[11] F. Sadak, M. Saadat, and A. M. Hajiyavand, "Real-Time Deep Learning-Based Image Recognition for Applications in Automated Positioning and Injection of Biological Cells," *Comput. Biol. Med.*, p. 103976, 2020, doi: 10.1016/j.compbiomed.2020.103976.

[12] Y. Onishi, T. Yoshida, H. Kurita, T. Fukao, H. Arihara, and A. Iwai, "An automated fruit harvesting robot by using deep learning," *Robomech J.*, vol. 6, no. 1, pp. 1–8, 2019, doi: 10.1186/s40648-019-0141-2.

[13] M. H. Saleem, J. Potgieter, and K. M. Arif, "Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments," *Precis. Agric.*, pp. 1–39, 2021, doi: 10.1007/s11119-021-09806-x.

[14] F. Gao *et al.*, "Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN," *Comput. Electron. Agric.*, vol. 176, p. 105634, 2020, doi: https://doi.org/10.1016/j.compag.2020.105634.

[15] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3626–3633, doi: 10.1109/ICRA.2017.7989417.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[17] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agric.*, vol. 157, pp. 417–426, 2019, doi: https://doi.org/10.1016/j.compag.2019.01.012.

[18] M. O. Lawal, "Tomato detection based on modified YOLOv3 framework," *Sci. Rep.*, vol. 11, no. 1, p. 1447, 2021, doi: 10.1038/s41598-021-81216-5.

[19] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO,'" *Precis. Agric.*, vol. 20, no. 6, pp. 1107–1135, 2019, doi: 10.1007/s11119-019-09642-0.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[21] H. Habaragamuwa, Y. Ogawa, T. Suzuki, T. Shiigi, M. Ono, and N. Kondo, "Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network," *Eng. Agric. Environ. Food*, vol. 11, no. 3, pp. 127–138, 2018, doi: https://doi.org/10.1016/j.eaef.2018.03.001.

[22] P. Pawara, E. Okafor, L. Schomaker, and M. Wiering, "Data Augmentation for Plant Classification," in *European Conference on Computer Vision*, 2017, pp. 615–626.

[23] A. Hernández-García and P. König, "Further Advantages of Data Augmentation on Convolutional Neural Networks BT - Artificial Neural Networks and Machine Learning – ICANN 2018," in *International Conference on Artificial Neural Networks*, 2018, pp. 95–103.

[24] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[25] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70–90, 2018, doi: https://doi.org/10.1016/j.compag.2018.02.016.

[26] K. A. Steen, P. Christiansen, H. Karstoft, and R. N. Jørgensen, "Using Deep Learning to Challenge Safety Standard for Highly Autonomous Machines in Agriculture," *Journal of Imaging* , vol. 2, no. 1. 2016, doi: 10.3390/jimaging2010006.

[27] M. Dyrmann, H. Karstoft, and H. S. Midtiby, "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, vol. 151, pp. 72–80, 2016, doi: https://doi.org/10.1016/j.biosystemseng.2016.08.024.

[28] P. Christiansen, L. N. Nielsen, K. A. Steen, R. N. Jørgensen, and H. Karstoft, "DeepAnomaly: Combining Background Subtraction and Deep Learning for Detecting Obstacles and Anomalies in an Agricultural Field," *Sensors* , vol. 16, no. 11. 2016, doi: 10.3390/s16111904.

[29] Tzutalin, "LabelImg." 2015, [Online]. Available: https://github.com/tzutalin/labelImg.

[30] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," in *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13029–13038, [Online]. Available: http://arxiv.org/abs/2011.08036.

[31] Y. Shao *et al.*, "Assessment of strawberry ripeness using hyperspectral imaging," *Anal. Lett.*, vol. 54, no. 10, pp. 1547–1560, 2020.