agronomy

MDPI

Article

Weed Detection by Faster RCNN Model: An Enhanced Anchor Box Approach

Muhammad Hammad Saleem [1], Johan Potgieter [2] and Khalid Mahmood Arif [1,*]

[1] Department of Mechanical and Electrical Engineering, School of Food and Advanced Technology, Massey University, Auckland 0632, New Zealand; H.Saleem@massey.ac.nz
[2] Massey AgriFood Digital Lab, Massey University, Palmerston North 4472, New Zealand; J.Potgieter@massey.ac.nz
* Correspondence: K.Arif@massey.ac.nz

Abstract: To apply weed control treatments effectively, the weeds must be accurately detected. Deep learning (DL) has been quite successful in performing the weed identification task. However, various aspects of the DL have not been explored in previous studies. This research aimed to achieve a high average precision (AP) of eight classes of weeds and a negative (non-weed) class, using the DeepWeeds dataset. In this regard, a DL-based two-step methodology has been proposed. This article is the second stage of the research, while the first stage has already been published. The former phase presented a weed detection pipeline and consisted of the evaluation of various neural networks, image resizers, and weight optimization techniques. Although a significant improvement in the mean average precision (mAP) was attained. However, the Chinee apple weed did not reach a high average precision. This result provided a solid ground for the next stage of the study. Hence, this paper presents an in-depth analysis of the Faster Region-based Convolutional Neural Network (RCNN) with ResNet-101, the best-obtained model in the past step. The architectural details of the Faster RCNN model have been thoroughly studied to investigate each class of weeds. It was empirically found that the generation of anchor boxes affects the training and testing performance of the Faster RCNN model. An enhancement to the anchor box scales and aspect ratios has been attempted by various combinations. The final results, with the addition of 64 × 64 scale size, and aspect ratio of 1:3 and 3:1, produced the best classification and localization of all classes of weeds and a negative class. An enhancement of 24.95% AP was obtained in Chinee apple weed. Furthermore, the mAP was improved by 2.58%. The robustness of the approach has been shown by the stratified k-fold cross-validation technique and testing on an external dataset.

Keywords: deep learning; convolutional neural network; weed detection; optimization algorithms; transfer learning

check for updates

Citation: Saleem, M.H.; Potgieter, J.; Arif, K.M. Weed Detection by Faster RCNN Model: An Enhanced Anchor Box Approach. *Agronomy* **2022**, *12*, 1580. https://doi.org/10.3390/agronomy12071580

Academic Editor: Wen-Hao Su

Received: 22 May 2022
Accepted: 27 June 2022
Published: 29 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Conventional weed control methods are generally cost-ineffective and produce adverse effects on the environment [1]. Hence, there is a need for an automatic weed control system that can reduce human and machinery efforts. Before deploying a weed management system, a precise identification/detection of weeds is a mandatory task. In this regard, computer vision plays a vital role in conjunction with artificial intelligence. Deep learning (DL), a class of machine learning (ML), has been a promising tool for performing various real-life object detection tasks, including agricultural operations, such as identification of plant diseases [2], agricultural land cover classification [3], fruit recognition [4], plant recognition [5], and many others [6].

Recently, weed recognition has also been focused on by the research community. Several DL-based methods have been introduced to detect various weed classes. A recent review paper presents a summary of the characteristics along with the advantages and

disadvantages of traditional ML-based techniques, deep learning algorithms, and available datasets and weeding equipment [7]. At the beginning of research into DL-based weed identification, most of the studies were focused on the weed classification task by DL-based feature extractors. For instance, one study presents a combined convolutional neural network (CNN) with a support vector machine (SVM) for the recognition of weeds [8]. Similarly, other research has proposed a graph-based CNN with ResNet-101 for weed classification [9]. Hu et al. [10] also presented a graph-based DL model to classify weeds. Then, the research community started to focus on both classification and localization tasks using DL-based object detection methods. For example, a research article used two DL models including EfficientDet and a version of You Only Look Once (YOLO-v5) to detect monocot/dicot weeds [11]. For that research, in-field images were collected that showed the usefulness of the state-of-the-art deep learning models for real/complex environments. On the other hand, loss plots should also be presented to show the training performance of the models. Furthermore, some of the older DL architectures could also be tested to show the significance of the selected model. Various DL classification models such as GoogLeNet, VGG, and DetectNet were used to detect broadleaf weeds [12]. The research evaluated various feature extractors, but training profiles that could be useful to understand the training performance of the DL models were not presented. A research showed the importance of the YOLO-v3 architecture for weed identification, and the performance of two commonly used DL frameworks was evaluated, including TensorFlow and Keras [13]. The trained model successfully detects weeds, but the selection of YOLO-v3 was not justified clearly. Hence, a comparative analysis with other object detection methods could further support the selection of the YOLO-v3 model. A recent article addressed an important task of weed identification in the seedling growth stage using DL architectures [14]. That study not only addressed an important agricultural problem but also evaluated different DL classification and detection models along with various considerations of image sizes. However, training performance should also be analyzed. Gao et al. [15] proposed a reduced model of YOLO for the recognition of weeds and proved the usefulness of the proposed approach by comparing it with former versions of YOLO. That research provided inference time for the proposed methods, which is one of the main practical aspects of the implementation of such research. The CenterNet model was used to detect weeds in [16].

The research community also evaluated transfer learning for weed recognition. The DL models were used in research for classification purposes along with conventional/traditional ML techniques to detect several classes of weeds [17]. A study presented the performance of 35 DL models through transfer learning for the identification of weeds in the real environment [18]. Suh et al. [19] presented a transfer learning-based approach to classifying weeds using pre-trained weights on ImageNet. One of the key features of that research was the evaluation of DL models on the images collected at different periods. It shows the strength of DL for weed detection. Although the transfer learning methods produced significant results, the main research gap was that the proposed approaches have not been tested on an external dataset with the same weed/crop classes. This would show the robustness of the methodologies in different field environments. Few of the studies analyzed the significance of DL model for Unmanned Aerial Vehicles (UAVs). A study investigated images of UAVs to recognize weeds using a ResNet-based model [20]. Research has been conducted on the identification of weeds using the ResNet model [21]. Ukaegbu et al. [22] used a CNN model on a quadcopter to detect broadleaf and grassweeds and to evaluate herbicide spraying. A UAV was used to detect weeds using Faster RCNN and SSD models [23]. An article presented an improved version of the Faster RCNN model for the image collected by UAV for weed identification [24]. The performance of the proposed method was compared with other DL and ML methods. However, the specifications of the anchor boxes were not analyzed in detail, which is a major research gap. Few robotic platforms have also been proposed for weed detection by DL. For example, the YOLO-v3 model with a DarkNet-53 feature extractor was trained and tested on a mobile platform to identify crops and weeds in a practical field, and a chemical spray was applied

to the detected weeds [25]. Kounalakis et al. [26] designed a prototype robotic system for weed detection using a CNN model. Furthermore, few studies have been conducted to perform weed segmentation [27–29], but it is not the focus of this research.

From the literature presented earlier, it can be summarized that most of the previous studies relied on either proposing the backbone models/feature extractors or leveraging the state-of-the-art DL object detection methods for weed identification purposes. However, to the best of our knowledge, none of the previous approaches has provided a systematic way to analyze the robustness of deep learning by exploiting various aspects of image resizing and optimization methods for the recognition of weeds. Moreover, a detailed analysis of a well-known object detector, Faster Region-based Convolutional Neural Network (RCNN), has not been performed at its architectural level to successfully detect weeds. Therefore, a DL-based approach has been proposed for weed identification, divided into two stages. The first stage has already been published in another journal. It presented a weed detection framework and obtained the best DL architecture along with the selection of the most optimum image resizer, interpolator, weights initializer, and DL optimizer (in the absence and presence of batch normalization) [30]. However, this paper is dedicated to the second stage of the research. The proposed method is based on our observations during the initial stage of the study.

It was found during the former phase of the study that all weed classes attained a high (more than 90%) average precision (AP), except for the Chinee Apple class from the selected dataset called DeepWeeds [31]. The Faster RCNN ResNet-101 model was found to be the most suitable method trained with the RMSProp optimizer and aspect ratio resizer technique with area interpolation. A high AP of seven weed classes showed that the ResNet-101 performed well in extracting the unique features of the weed classes. However, the distinct features of the Chinee apple weed were not well extracted. Few of the test images belonging to the Chinee apple were detected with a high confidence score. However, most of the images of the Chinee apple were undetected by the trained model. Hence, there was a need to further investigate the Faster RCNN model to detect the remaining images of the Chinee apple, along with maintaining the high AP of the other weed classes (and a negative/non-weed class).

The main contributions of this research are:

1. Obtained a 24.95% higher AP for Chinee apple and maintained the performance of all other weed classes and a negative class with AP > 89%;
2. Investigated another robust way to improve the weed detection task by deep learning;
3. Evaluated the significance of various anchor box scales and aspect ratios for weed identification that can be replicated for other agricultural applications;
4. Achieved an improvement of 2.58% in mean average precision compared to the former phase of the research;
5. Shown the robustness of the approach by the stratified k-fold cross-validation technique and testing on an externally generated dataset;
6. Made all data including the final weights of the optimized Faster RCNN model publicly available to reuse as a transfer learning for other weeds-related datasets.

The rest of the paper is organized as follows: Section 2 describes the dataset, training, specifications, and steps of the enhanced anchor boxes. Section 3 presents the summary of the first phase of the research and results and discussion of the proposed approach by graphical/training plots, and detection outcomes. Section 4 concludes the research along with future directions.

2. Materials and Methods

This section elaborates on the selected dataset, specifications related to the deep learning setup, and methodology for the enhancement in the anchor boxes.

The overall flow of both stages of the research is presented in Figure 1.

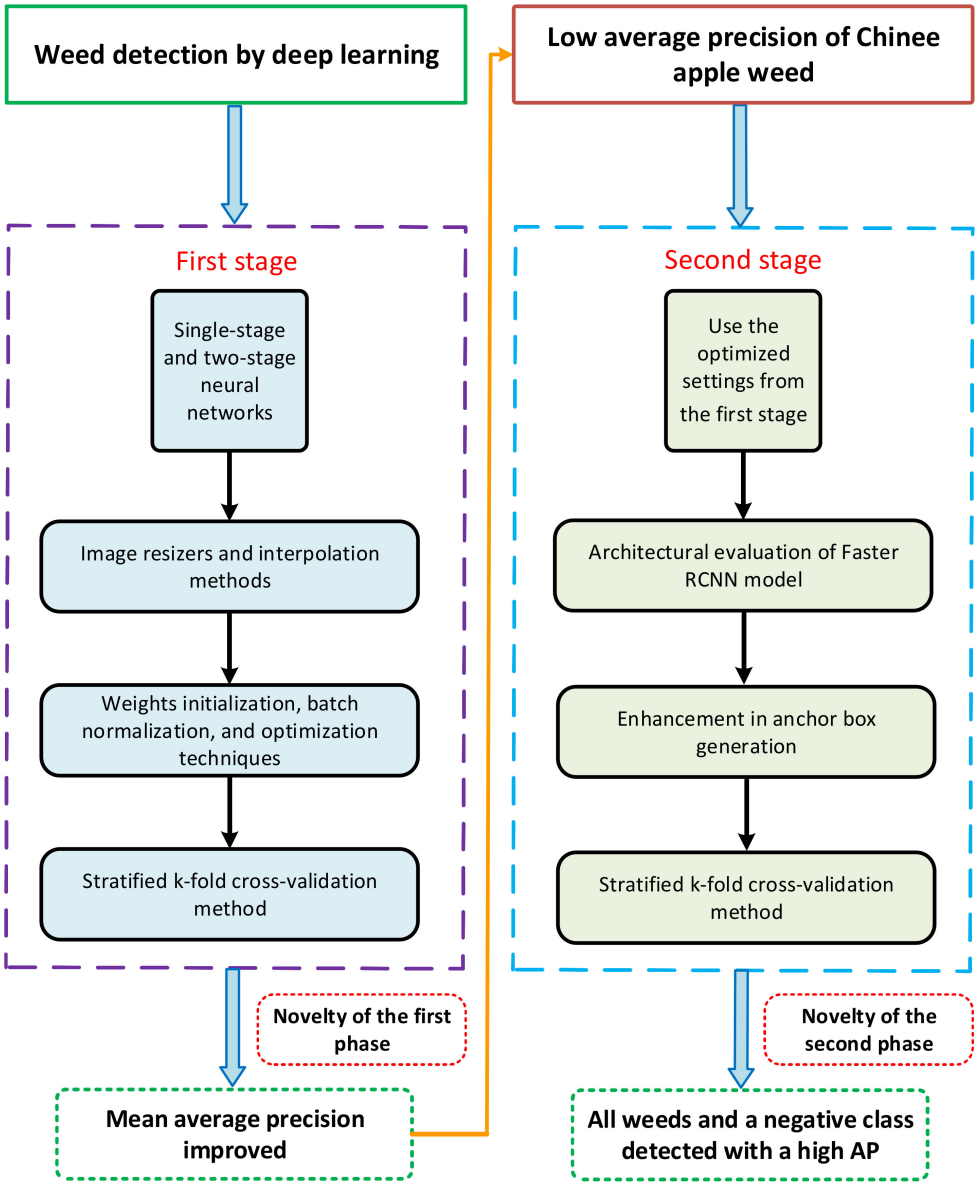


Figure 1. The overall methodology of the two-step DL-based weed identification.

2.1. Dataset Specifications

The DeepWeeds dataset [31] was used throughout this research. This dataset contains images of eight weed classes including a negative/non-weed class, collected in Northern Australia. The reasons for the dataset selection were the diverse nature and consideration of various properties of the real field environment, including actual background, inconsistent lighting, occlusion, etc. Hence, the high detection precision of each class in such a dynamic dataset would show the effectiveness/robustness of the deep learning-based method for weed identification purpose. The dataset contains 17,509 images; it was divided into three sub-datasets for training (70%), validation (20%), and testing (10%). The class names were shortened to visualize the detected results more clearly, such as Chinee apple was annotated as C_App, Lantana was annotated as Lntna, Prickly acacia was annotated by P_acacia, Parthenium with P_nium, Rubber vine with R_vine, Siam weed with S_weed, Parkinsonia with P_sonia, Snakeweed with Snk_wd, and Negative with Ngtv. For the annotations of the dataset images, the XML files were obtained using an open-source tool named LabelImg. These XML files were later converted into CSV and then TF records [32].

2.2. Deep Learning Specifications

From the first step of the research [30], the Faster RCNN model was found to be the best model. It was trained through TensorFlow Object Detection API 1. The transfer learning technique was applied using the weights of the Common Objects in Context (COCO) dataset [33]. Furthermore, all training was conducted using NVIDIA GeForce GTX 1080 Ti GPU card. The most suitable batch size for the Faster RCNN model was equal to 2 [30]. Moreover, the hyperparameters used in the DL optimization method (learning rate, epsilon, momentum, and decay) were selected using the random search technique [34]. According to the methodology presented for the first phase of the research, different DL optimizers were used including Stochastic Gradient Descent (SGD) with momentum, Root Mean Square Propagation (RMSProp), and Adaptive Moment Estimation (Adam). RMSProp attained the best results in terms of mean average precision with the hyperparameters: learning rate = 3×10^{-4}, momentum = 0.9, rho = 0.9, and epsilon = 1.0.

The performance of the DL model was evaluated in terms of average precision (AP) for each class obtained by the 11-point interpolation method [35] and defined as AP at unique recall levels. This method first evaluates the precision at different recall levels and then an interpolated precision is evaluated by taking the maximum/highest precision for a specific recall level. Finally, the average of the AP values for each class is calculated and the mean average precision (mAP) is evaluated. More details can be seen in [35].

Validation of the final results was performed using the stratified k-fold cross-validation method. This technique was adopted due to the class imbalance problem of the DeepWeeds dataset as the negative class has a considerably higher number of images than all weed classes. This method maintains the class distribution in each fold [30]. Furthermore, an external dataset has been generated by random internet search to test the approach in different environments.

2.3. Selection of the Faster RCNN ResNet-101

The first phase of the research evaluated various DL models to select the best-suited architecture for weed detection. In this regard, models including YOLO-v4, Single-Shot MultiBox Detector (SSD) with Inception-v2, MobileNet, and ResNet-50, EfficientDet, CenterNet ResNet-50, Region-based Fully Convolutional Networks (RFCN) ResNet-101, Faster Region-based Convolutional Neural Networks (RCNN) with feature extractors such as Inception-v2, ResNet-50, and ResNet-101, were trained and tested on DeepWeeds dataset [30]. The Faster RCNN ResNet-101 attained superior performance due to its highest mean average precision of 87.64% [30]. Therefore, performance optimization was attempted for the Faster RCNN model in the previous stage and further enhancement has been presented in this article.

2.4. Methodology of the Enhanced Anchor Box Approach

The comprehensive analysis presented in the first stage of the research provided a solid ground for investigating the performance of Faster RCNN in more detail. Careful observation of the detected images of the Chinee apple led to the conclusion that some of the test images were identified with a high confidence score, but most of them remain undetected. To cope with this problem, the main architecture of the Faster RCNN has been thoroughly investigated. The generation of the anchor boxes is one of the main characteristics of the Faster RCNN model. Therefore, an in-depth analysis of anchor box scales and aspect ratios has been performed to improve the AP of the Chinee apple. Furthermore, the AP of all other classes was aimed to be maintained.

2.4.1. Major Novelty of the Original Faster RCNN Model

The weakness of the former version of the Faster RCNN model was its slow speed due to the use of the selective search (SS) method to generate the region proposals [36]. It was addressed through the Region Proposal Network (RPN) in the Faster RCNN model. This version of RCNN takes input images to a feature extractor through a convolutional

neural network model (CNN). Then, the feature map from the convolutional layer is fed to the RPN that generates region proposals in the form of a sliding window. Then, a similar structure comes to its previous model (Fast RCNN), including Region of Interest (ROI) pooling, a classifier, and a bounding box regressor. The main concept and characteristics of the original Faster RCNN are summarized as follows:

- A new network named Region Proposal Network (RPN) was introduced that generates proposals with different scales and aspect ratios.
- In contrast to the Fast RCNN model, the region proposals can be modified according to the specific application.
- The Faster RCNN model combines the RPN and Fast RCNN model; the same convolutional layers are shared between the parts of the model. Hence, no additional time is required to generate the proposals.
- The concept of an anchor box was developed, which is a reference box of specific size and aspect ratio. While training RPN, the training images were passed through the sliding windows using various anchor boxes. The dimensions of the anchor boxes were specified in terms of scale size and aspect ratio and placed in the center of the sliding window. The candidate boxes are then obtained using anchor boxes that operate for the regression. Hence, the anchor box contributes to identifying and localizing the objects with varying dimensions/coordinates.
- The generation of the anchor box was one of the key elements behind the success of the Faster RCNN model. These boxes help solve multiclass classification problems, detect objects of variable size in the dataset, and identify overlapping objects. These reference boxes are placed at various points in the image.

2.4.2. Steps to Obtain Enhanced Anchor Boxes

As discussed in the previous subsection, the generation of several anchor boxes plays a vital role in the detection of target objects. Various aspect ratios and scales together generate anchor boxes, and these boxes make a sliding window that passes through the images.

While annotating the training images from the DeepWeeds dataset, it was empirically observed that the weed classes vary in terms of their bounding box coordinates. Therefore, an enhancement in the anchor box has been attempted through this research to provide a better weed detection outcome. This idea led to a range of new experiments with Faster RCNN to successfully detect and classify all classes of weeds, including the Chinee apple. A detailed explanation of the enhanced anchor box approach is shown in Figure 2.

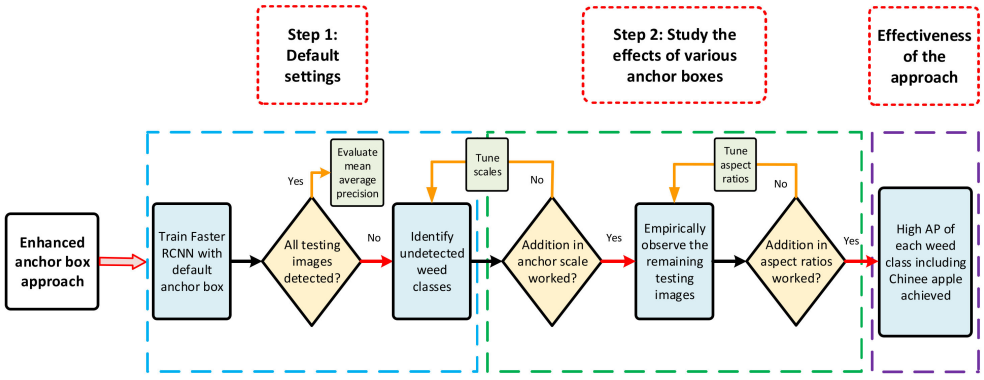


Figure 2. An explanation of the enhanced anchor box approach.

The overall flow of this approach is explained as under:

- First, Faster RCNN was trained on the DeepWeeds dataset with the default scale size and aspect ratio of the anchor boxes according to the original Faster RCNN model. In

this regard, anchor boxes with 128 × 128, 256 × 256, and 512 × 512 scales combined with aspect ratios of 1:1, 2:1, 1:2, and only 1:1 were considered.

- Then, an empirical observation was made to observe whether all testing images are detected or not. In the case of finding undetected images, an enhancement was attempted in the anchor boxes according to the features/characteristics of the weed classes.
- Primarily, the anchor boxes were enhanced by adding one scale. If the insertion of a scale fits the anchor boxes to detect the weed classes, those scales were fixed for the rest of the analysis. Otherwise, more scales were added to obtain the optimum combination. The addition of the scales was performed in the form of twice/half of the previous scale size. For example, the default sizes were 128 × 128, 256 × 256, and 512 × 512; then, a 64 × 64 scale size was inserted. The effects of adding a scale were evaluated in terms of training loss, mAP, and AP of individual classes.
- If the mAP was improved significantly, again, the remaining testing images were carefully observed to check whether all classes were detected or not.
- Then, aspect ratios were modified in two further stages. First, aspect ratios with a small gap, such as a ratio of 1:4 between the default anchor boxes, were added. If the model produces unsuccessful results, reciprocal aspect ratios were considered. The detected images with their confidence scores were compared with the results obtained in the previous step of the analysis.
- In this way, the empirical adjustment of the scale size and aspect ratio contributed to getting the successful detection of weeds. It was further observed that the proposed method not only obtained better true positive results, but also reduced the region proposal network (RPN), classification, and localization errors.

3. Results and Discussion

3.1. Highlights of the First Stage of the Research

3.1.1. Analyzed the Performance of Several DL-Based Object Detectors

First, a comprehensive analysis of various single- and two-stage neural networks was performed in terms of percentage training loss and mean average precision. The Faster RCNN architecture achieved the highest mean average precision. It was trained with various DL backbone/classification models such as Inception-v2, ResNet-50, and ResNet-101. ResNet-101 was found to be the most suitable model that successfully extracted distinct features of seven weed classes and a negative class. The mean average precision was found to be 87.64%, and most of the weed classes were successfully detected, except for the negative class that achieved the lowest average precision (AP) of 62.35% [30].

3.1.2. Studied the Effects of Image Resizing and Interpolation Techniques

The second step was studying the effects of image resizing techniques on the deep learning models. Two image resizers were evaluated, including aspect ratio and fixed-shape resizers. Moreover, image interpolation methods were also used with both resizing techniques, including bilinear, bicubic, area, and nearest neighbor. The aspect ratio resizer with the area interpolation method was found to be the most suitable technique. This step contributed to improving the AP of the negative class to 96.61% and the mAP was also improved to 91.55%. On the other hand, this step of the work degraded the performance of the Chinee apple class to 75.78%.

3.1.3. Performance Optimization by Weight Initializers, Batch Normalization, and DL Optimizers

Finally, the weights of the Faster RCNN model were optimized. This step was divided into three stages. First, various parameters of weight initialization were investigated. Truncated normal, scaling variance, and random normal initialization techniques were used. The truncated normal and scaling variance initializers performed well, and the random normal initializer did not produce satisfactory results. The most appropriate

parameters of the truncated normal were found to be 0.01 standard deviation and zero mean; the parametric values of scaling variance were 1.0 scaling factor with the normal distribution and considered an average of the input and output weight tensor as the mode of operation.

Then, two well-known DL optimizers including RMSProp [37] and Adam [38] were analyzed in the presence and absence of batch normalization to further enhance the performance of the Faster RCNN model. It was found that the RMSProp optimizer (without batch normalization) attained the highest mAP of 93.44%. It was also observed that all classes achieved more than 90% AP, except the Chinee apple, which was degraded with an AP of 68.62%. Therefore, it was concluded that future research should attempt a higher AP of Chinee apple along with sustaining the AP of all other weed classes. Additionally, the outcomes of the former stage were validated by a stratified k-fold cross-validation technique.

Each step of the first phase of the research contributed to an improvement in mean average precision. A summary of the results obtained through the former stage of the research is presented in Figure 3.

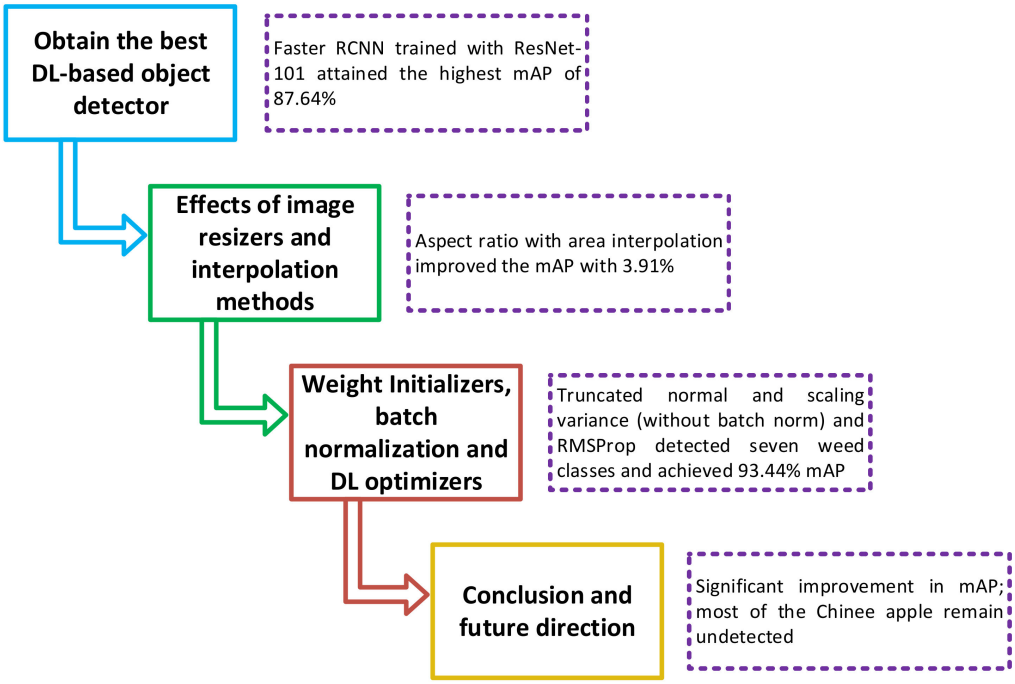


Figure 3. Highlights of the first step of the research.

3.2. Performance of the Second Phase of the Research: Enhanced Anchor Box Approach

This section presents the effects of the enhanced anchor box approach by using various combinations of scale size and aspect ratios, on the detection of weed classes and a negative class. The training performance has been analyzed by the graphical plots. The testing performance has been shown by the average precision (AP) of each class along with the mean average precision (mAP) (Table 1) along with the detection outcomes of the weed classes, to show the effectiveness of the proposed methodology.

Table 1. Effects of anchor scale sizes and aspect ratios on the average precision of each class.

Anchor Box Scale Sizes	Anchor Box Aspect Ratios	Average Precision (%)									mAP (%)	Training Time (h)
		Chinee Apple	Lantana	Prickly Acacia	Parthenium	Parkinsonia	Rubber vine	Siam Weed	Snake Weed	Negative		
$\{128^2, 256^2, 512^2\}$	{1:2, 1:1, 2:1}	66.39	99.83	81.23	100	99.95	98.13	87.84	24.56	49.3	78.58	10
$\{128^2, 256^2, 512^2\}$	1:1	56.11	60.12	92.81	38.11	98.33	99.84	98.94	23.62	94.9	73.63	10
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 2:1}	68.62	99.78	90.09	100	99.89	99.76	98.71	93.22	90.9	**93.44**	10
$\{32^2, 64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 2:1}	71.60	85.54	72.12	100	93.75	75.62	79.41	78.39	86.18	82.51	9.72
$\{16^2, 32^2, 64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 2:1}	70.39	87.38	76.11	100	94.88	72.55	86.83	79.38	85.50	83.67	9.61
$\{64^2, 128^2, 256^2, 512^2\}$	{1:4, 1:2, 1:1, 2:1}	81.65	100	65.97	100	100	100	98.95	83.92	58.95	87.71	10
$\{64^2, 128^2, 256^2, 512^2\}$	{1:4, 1:2, 1:1, 1.5:1, 2:1}	71.38	100	53.49	100	99.37	99.76	40.79	19.75	92.34	75.21	10.72
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 1.5:1, 2:1}	98.16	96.02	88.15	99.76	99.78	96.19	98.94	88.23	86.05	**94.58**	10.17
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1.25, 1:1, 1.5:1, 2:1}	80.94	100	48.03	100	98.91	95.67	11.82	26.35	55.51	68.58	10.22
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 1.25:1, 1.5:1, 2:1}	99.64	92.63	34.44	100	99.72	99.61	99.92	30.48	79.08	81.72	10.11
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 1.5:1, 1.75:1, 2:1}	98.58	99.11	94.13	68.79	97.74	100	94.97	51.55	94.21	88.78	10.61
$\{64^2, 128^2, 256^2, 512^2\}$	{1:2, 1:1, 1.75:1, 2:1}	100	97.25	58.4	97.7	100	97.72	37.36	72.92	64.21	80.61	10
$\{64^2, 128^2, 256^2, 512^2\}$	{1:3, 1:1, 3:1}	93.57	96.14	89.93	99.5	99.84	99.6	98.98	89.59	97.06	**96.02**	10.17
$\{64^2, 128^2, 256^2, 512^2\}$	{1:4, 1:1, 4:1}	98.64	99.72	95.16	99.56	97.79	99.08	80.86	28	89.11	87.56	10.67
$\{64^2, 128^2, 256^2, 512^2\}$	{1:3, 1:2, 1:1, 1.5:1, 2:1, 3:1}	92.63	93.8	71.02	98.1	100	99.84	99.56	92.56	86.83	92.70	10.5

16^2: 16×16, 32^2: 32×32, 64^2: 64×64, 128^2: 128×128, 256^2: 256×256, 512^2: 512×512.

3.2.1. Default Anchor Box Scale

During the first phase of the research, the Faster RCNN model was trained with the default settings. According to this, the model was trained with anchor box scales 64 × 64, 128 × 128, 256 × 256, and 512 × 512 with aspect ratios 1:2, 1:1, and 2:1. For this step of the research, the effects of various scales and aspect ratios have been analyzed step by step.

The authors of the original Faster RCNN architecture used several combinations of aspect ratio and scale size explained in [36]. The scale size of 128 × 128, 256 × 256, and 512 × 512 with the aspect ratios of 1:2, 1:1, 2:1, and with only 1:1, attained the highest mean average precision for that research. Similarly, both default combinations of scales and aspect ratios were applied for this study. The aspect ratio of 1:2, 1:1, and 2:1 achieved a better mean average precision, as shown in Table 1. However, the classes including Chinee apple, Snakeweed, and negative classes did not reach high average precision and were confused with other classes such as Lantana. A pictorial representation of the anchor box according to the default settings is presented in Figure 4. It can be observed that the area of the sliding window is determined by the scale size and aspect ratio. Consequently, in the case of a modification in scale and ratio, the area of the anchor box will be changed. Therefore, an important reason for an unsatisfactory result by the default settings is an unacceptable anchor box dimension that fits the characteristics/bounding box coordinates of the testing images. Hence, some of the images belonging to the testing dataset could not be detected. This reduced the mAP of the Faster RCNN model. An example of some of the classes detected by the default settings is presented in Figure 5.

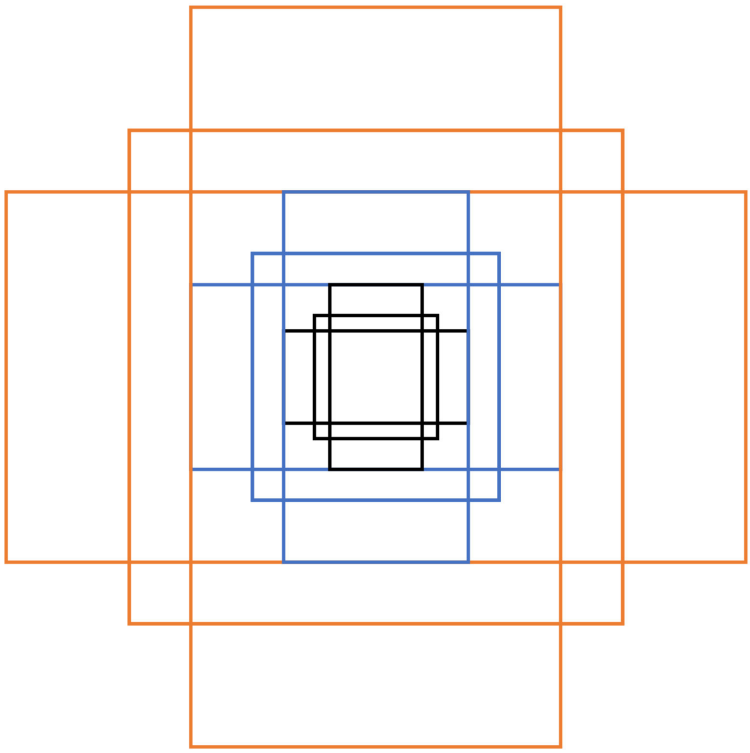


Figure 4. Anchor box with the default settings.

3.2.2. Enhancement in Anchor Box Scales

Next, smaller scale sizes were added to obtain an enhancement in the anchor boxes. In this regard, a 64 × 64-sized window was added, and its effects were evaluated. It was found that the inclusion of the 64 × 64 scale significantly improved the mAP with a huge margin of 14.86%. From the training plots presented in Figure 6a–d, both region proposal network (RPN) losses, including localization loss (R_loc_loss) and objectness loss (R_obj_loss), improved with a small difference of 0.014% and 0.104%, respectively.

Furthermore, losses including box classifier localization loss (Loc_class) and classification loss (Class_loss) were also reduced by 0.44% and 0.38%, respectively, compared to those obtained during training with default scale sizes, as shown in Figure 6e–h. It can be observed from Figure 6 that the Faster RCNN model did not suffer from overfitting. The loss plots were converged and settled down to the final value. Additionally, there was no such abrupt increase in the loss after achieving a steady-state condition. The anchor boxes after the addition of the 64 × 64 scale are presented in Figure 7. It can be seen that this modification in the anchor box should be useful for detecting and localizing weeds having small coordinates.

Then, the performance of the testing dataset was evaluated. The detection results for the classes including Prickly acacia, Siam weed, Snakeweed, and negatives attained a higher AP compared to the results obtained by the previous/default anchor boxes as shown in Table 1. It implies that the images belonging to these classes had coordinate ensembles on the scale of 64 × 64, resulting in a significantly higher AP. However, only Chinee apple weed was found to have low AP. There were two reasons for the insufficient results of the Chinee apple; one was confusion with other classes, including Lantana and Negative, and the second was that some images could not be detected. Examples of the four truly detected classes along with a false-positive result for the Chinee apple are presented in Figure 8.

Later, smaller scale sizes were also used to observe their effects on the Faster RCNN. For example, 32 × 32 and 16 × 16 scales were added; these anchor box scales achieved a lower mAP due to a low AP of the classes including Lantana, Prickly acacia, rubber vine, Siam weed, and Snakeweed (as shown in Table 1). However, the training time was slightly reduced with small scale sizes as compared to the default anchor box scales, as shown in Table 1. In summary, due to a substantial improvement in the performance of the Faster RCNN model after the insertion of the 64 × 64 scale, the scale sizes of 64 × 64, 128 × 128, 256 × 256, and 512 × 512 were fixed for the rest of the analysis.

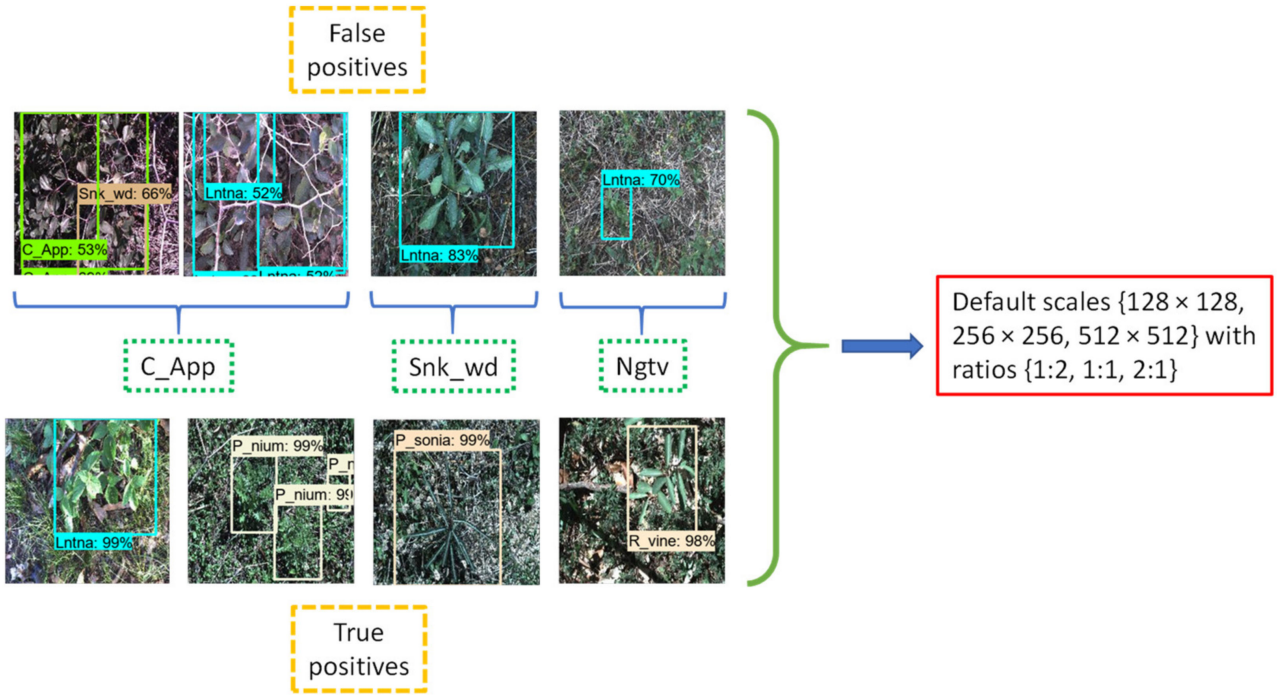


Figure 5. True-positive and false-positive results using the default scale size and aspect ratio.

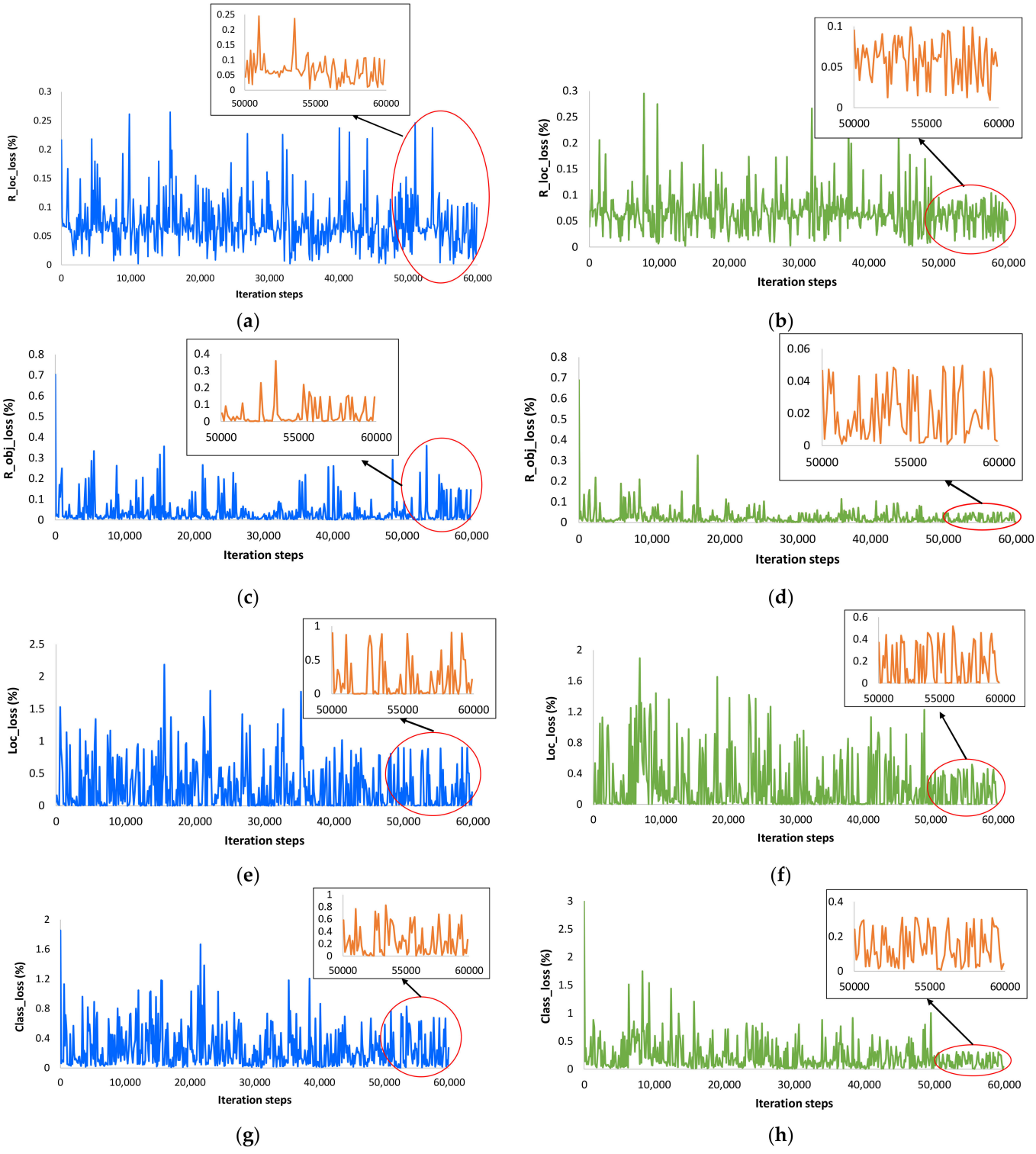


Figure 6. Training loss plots of the Faster RCNN model: (**a**) RPN localization loss with default settings; (**b**) RPN localization loss after adding 64 × 64-scale-size window; (**c**) RPN objectness loss with default settings; (**d**) RPN objectness loss after adding 64 × 64-scale-size window; (**e**) box classifier localization loss with default settings; (**f**) box classifier localization loss after adding 64 × 64-scale-size window; (**g**) box classifier classification loss with default settings; (**h**) box classifier classification loss after adding 64 × 64-scale-size window.

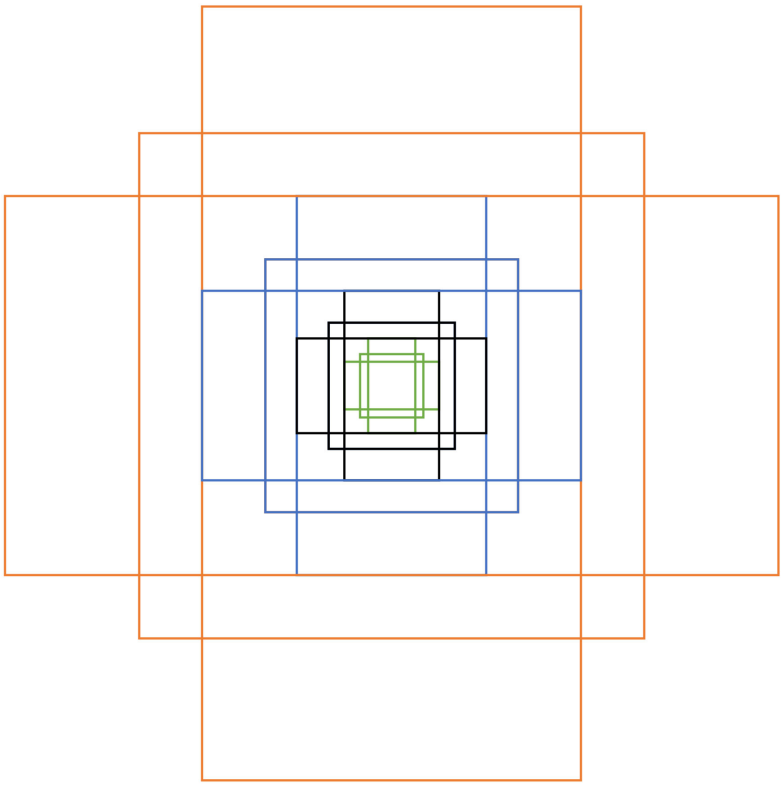


Figure 7. Enhanced anchor box after the addition of a 64 × 64 scale (green box).

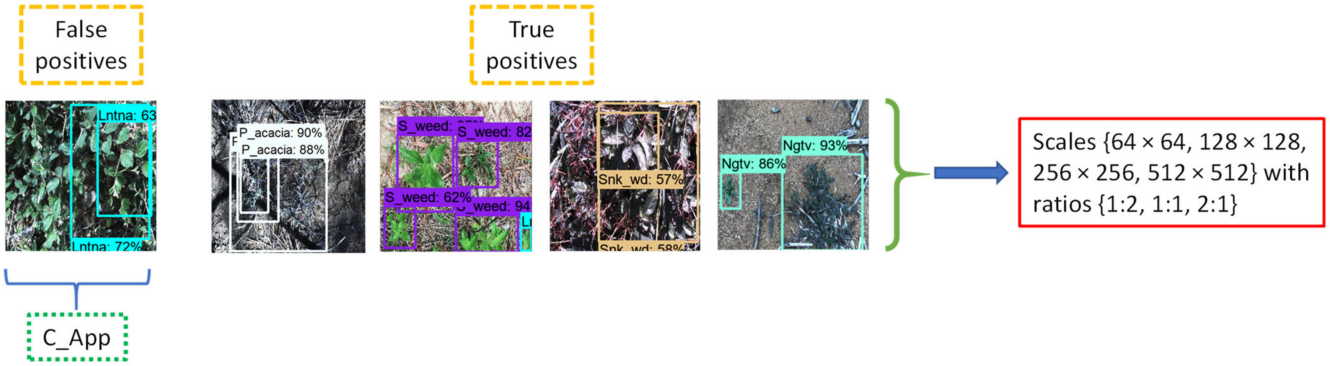
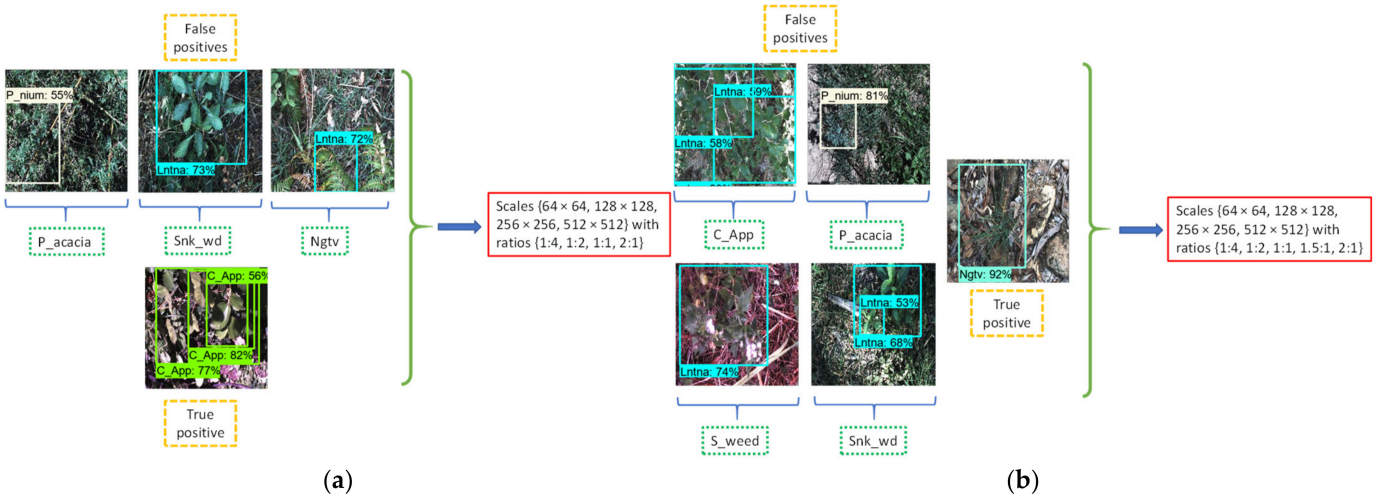


Figure 8. True positive results of Prickly acacia, Siam weed, Snakeweed, and negative class; False positive result of Chinee apple after the addition of a 64 × 64 scale.

3.2.3. Effects of Different Aspect Ratios

a. Gradual Enhancement in Aspect Ratios

The next step was to study the effects of the aspect ratio to obtain the optimum anchor boxes. First, a smaller aspect ratio of 1:4 was added to the default values. These aspect ratios marginally improved the average precision of the Chinee apple, but few classes, including Prickly acacia, Snakeweed, and negative, were degraded, as shown in Figure 9a. This was due to the mismatch of the required anchor boxes with the modified sizes. The Faster RCNN perceived (with the configurations presented in this step) that the testing images of Snakeweed and negative belonged to the Lantana. For a similar reason, Prickly acacia was confused with Parthenium.



(**a**) (**b**)

Figure 9. True and false-positive results after addition to default aspect ratios: (**a**) Results after 1:4 aspect ratio; (**b**) Results after 1:4 and 1.5:1 aspect ratios.

Then, a larger (than the original 1:1) aspect ratio was added to observe its effects on the model's performance. Therefore, an aspect ratio of 1.5:1 was added to 1:4 and default ratios, but the resulting anchor boxes were found to be not suitable for the classes of Chinee apple, Siam weed, Prickly acacia, and Snakeweed. The reason was that the addition of 1.5:1 and 1:4 aspect ratios made an anchor box that observed these classes as Lantana and Parthenium, as presented in Figure 9b. Therefore, it was proved that the simultaneous addition of a smaller and larger aspect ratio was inappropriate for the detection of weed classes.

Later, only the inclusion of an aspect ratio of 1.5:1 was examined with the default (1:2, 1:1, and 2:1). It was noticed that an addition of a 1.5:1 aspect ratio was feasible to obtain better weed detection outcomes. The Chinee apple was detected with a higher average precision of 98.16% along with a higher mAP. The successful detection results of the Chinee apple show that its images required all aspect ratios with a difference of 1:2. Moreover, four weed classes maintained their AP including Parthenium, Parkinsonia, Prickly acacia, and Siam weed. The other four classes, such as Lantana, Rubber vine, Snakeweed, and negative, achieved a lower AP. These results were acceptable since mAP was also slightly improved by 1.14% along with a significant improvement in AP of the Chinee apple. The resulting anchor boxes are presented in Figure 10.

Furthermore, it was also observed that only the box classifier localization loss (Loc_loss) improved with a margin of 0.08% compared to the loss that occurred during the previous stage of the method; the rest of the losses did not show an improvement. This small reduction in Loc_loss produced a significant improvement in the AP of the Chinee apple class. A few examples of the successful outcomes of all weed classes are presented in Figure 11.

The anchor box specifications after the addition of the 1.5:1 ratio improved the weed detection results. It shows that the required anchor box sizes can be obtained by gradual tuning of the aspect ratio. Furthermore, the correct finding of the anchor box is subjected to the empirical observations on a gradual change in anchor box scales and aspect ratios. Hence, certain intervals between anchor box enhancements were considered to identify the weeds correctly.

Few other combinations of the aspect ratio were also studied. For example, 1:1.25, 1.25:1 and 1.75:1 were added with an aspect ratio of 1.5:1 and default (1:2, 1:1, and 2:1). The aspect ratios of 1:1.25 and 1:25:1 gave a lower mAP (as shown in Table 1). For reference to their false-positive results, Figure 12a,b present the detection results. However, an aspect ratio of 1.75:1 (combined with 1.5:1 and default aspect ratios) was reasonable to detect the classes of weeds, except for the Parthenium and Snakeweed, which were confused with Prickly acacia and Chinee apple, respectively, as shown in Figure 12c. Finally, the effects of

a 1.75:1 aspect ratio with default ratios were studied. The classes including Prickly acacia, Siam weed, Snakeweed, and negative classes were seen to not detect well (Figure 12d). The training time with the addition of aspect ratios was marginally increased, as presented in Table 1.

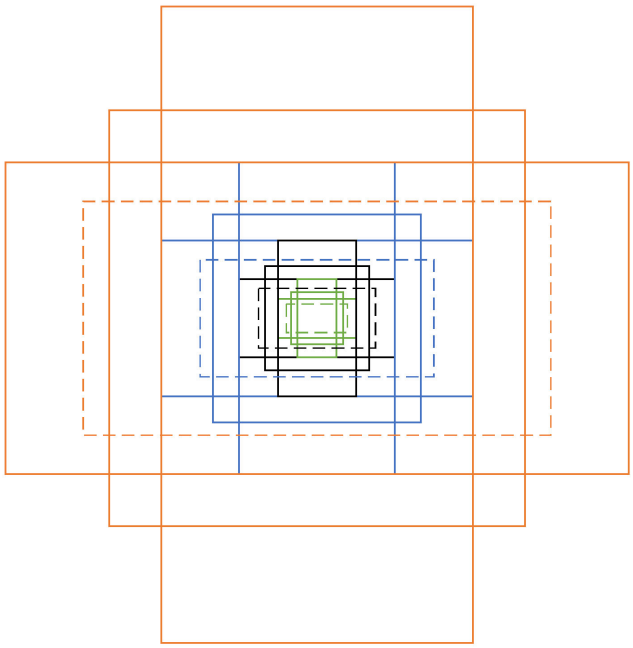


Figure 10. Resultant anchor box after adding 1.5 aspect ratio (dotted line boxes).

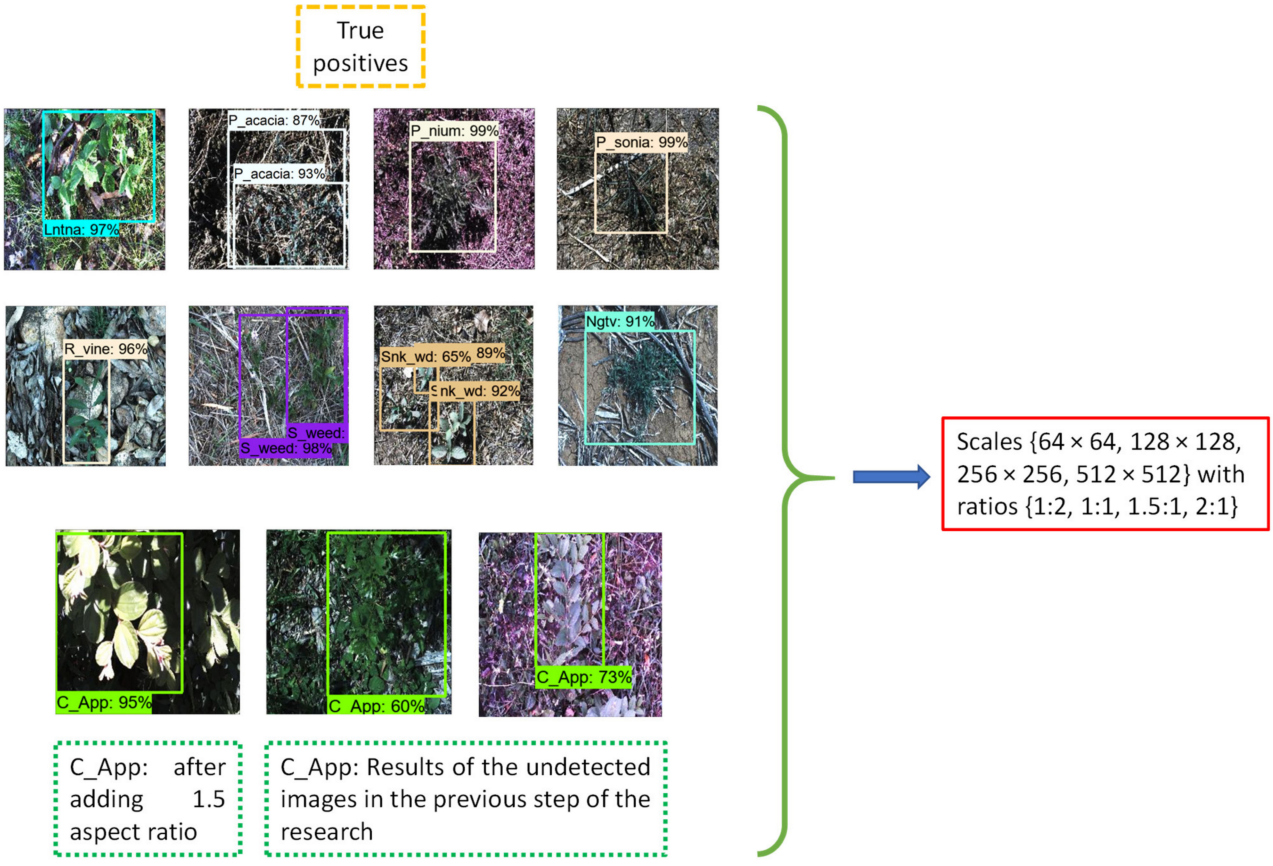


Figure 11. Detection results for all classes after the addition of a 1.5 aspect ratio.

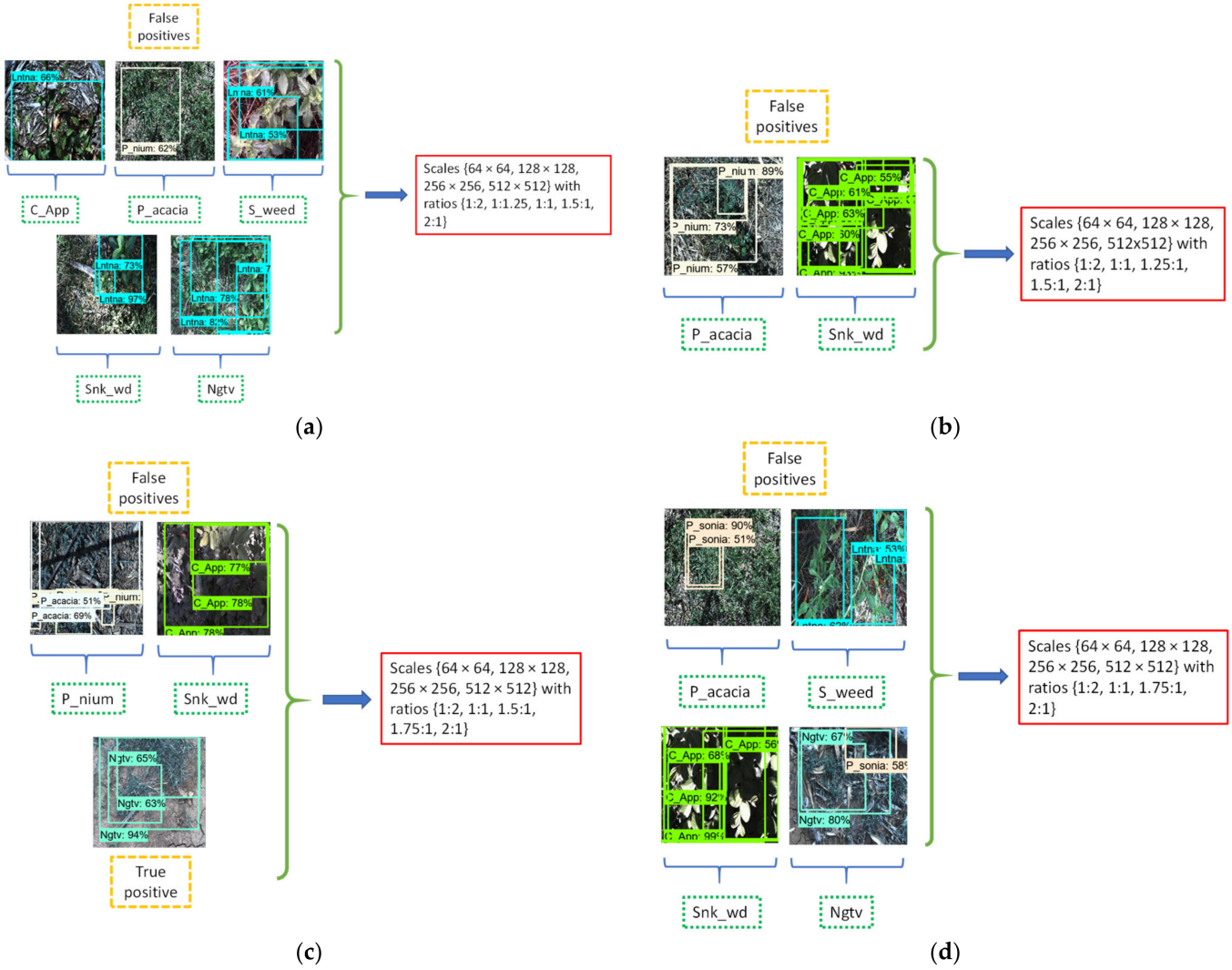


Figure 12. Detection results after the addition of various aspect ratios to the default ratios: (**a**) presents false positives with aspect ratios of 1:1.25 and 1.5:1; (**b**) presents false positives with aspect ratios of 1.25: 1 and 1.5:1; (**c**) presents true and false positives with aspect ratios of 1.5:1 and 1.75:1; (**d**) presents false positives with an aspect ratio of 1.75:1.

b. Reciprocal Enhancement in Aspect Ratios

After the addition of a 1.5:1 aspect ratio, the AP of the Chinee apple improved significantly. However, the "Negative" class was marginally degraded. From a practical perspective, the non-weed/negative class should also be detected and localized accurately. It is very useful for a site-specific weed management system when there is always a need for discrimination between a weed and a non-weed class to apply the herbicide spray precisely. Moreover, one of the main objectives of this research was to maintain the high AP of all classes. Therefore, an attempt was made to improve the AP of the negative class. In this regard, the anchor boxes were enhanced by reciprocal aspect ratios.

First, the effects of the aspect ratio of 1:3 and 3:1 (Figure 13) on the enhanced anchor scales were studied. Most training losses were reduced, including box classifier classification and localization losses, and RPN objectness loss with a margin of approximately 0.12%, 0.11%, and 0.02%, respectively, compared to the loss obtained with a gradual addition in an aspect ratio of 1.5:1, as shown in Figure 14. The training plots show that the model did not suffer from commonly occurring problems such as overfitting, as the loss value reached its steady point and there was no sudden change after obtaining the convergence. These training results were also validated by the testing outcomes; the negative class

achieved a significantly higher AP of 97.06% along with maintaining the AP of the classes that included Parthenium, Parkinsonia, Rubber vine, and Siam weed. The Chinee apple again attained a high AP of 93.57%. Few classes including Lantana, Prickly acacia, and Snakeweed were degraded with a bearable margin of 1–4% average precision as shown in Table 1. However, the mAP also improved to 96.02%, which was 2.58% better than the results achieved in the previous stage of the research. A few examples of all detected classes are represented in Figure 15. Furthermore, few other reciprocal combinations of aspect ratios were tested, which did not provide any significant improvements. Only ratios of 1:4 and 4:1 attained considerable outcomes, but Snakeweed was detected with only 28% AP as shown in Table 1. Finally, an anchor box with a combination of the two best results was also generated, considering the aspect ratios 1:3, 1.5:1, and 3:1, with the default ratios of 1:2, 1:1, and 2:1. The results of all weed classes with these aspect ratios were satisfactory except for prickly acacia. Therefore, the aspect ratios of 1:3 and 3:1 were found the most optimum solution to detect all the weed classes and a negative class.

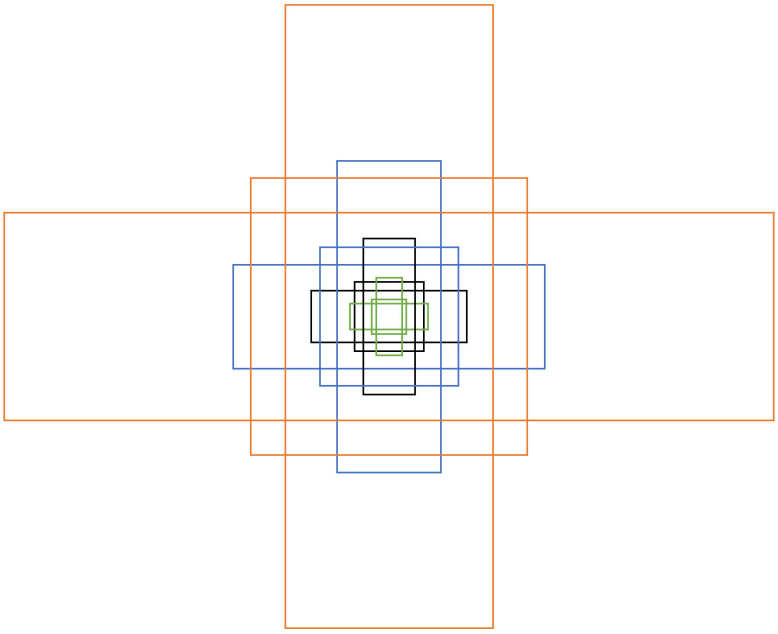


Figure 13. Enhanced anchor box after aspect ratios of 1:3 and 3:1.

The proposed anchor box method for the Faster RCNN model significantly improved the detection of weed classes, specially the Chinee apple, which was undetected/unsuccessful in the previous stage of the research. In terms of computation time, the Faster RCNN model with the default anchors required a slightly lower detection time of around 0.829 s/image compared to the final settings (with 64 × 64 scale size and 1:3, 3:1 aspect ratio) taking 0.840 s/image. However, it should be noticed that the modified Faster RCNN needed 12 anchor boxes compared to 9 anchor boxes with the default scale size and aspect ratio. Therefore, a small difference in the detection time after increasing some anchor boxes does not show any negative aspect of the proposed approach. The overall significance of the proposed method has been shown in Table 2, which compares the mean average precision with the latest DL models. It can be observed that the Faster RCNN ResNet-101 model attained the highest mAP as compared to all other DL architectures. Furthermore, it can also be seen that the performance of the model was improved by 5.8% mAP through the optimized version of the model as presented in the previous phase of the research [30]. However, the performance of the model has been further enhanced by the proposed approach in this article (enhanced anchor box approach) with a margin of 2.58% mAP. So, the overall improvement in the model's outcome was 8.38% in mAP as compared to the default settings, which shows the contribution of this research.

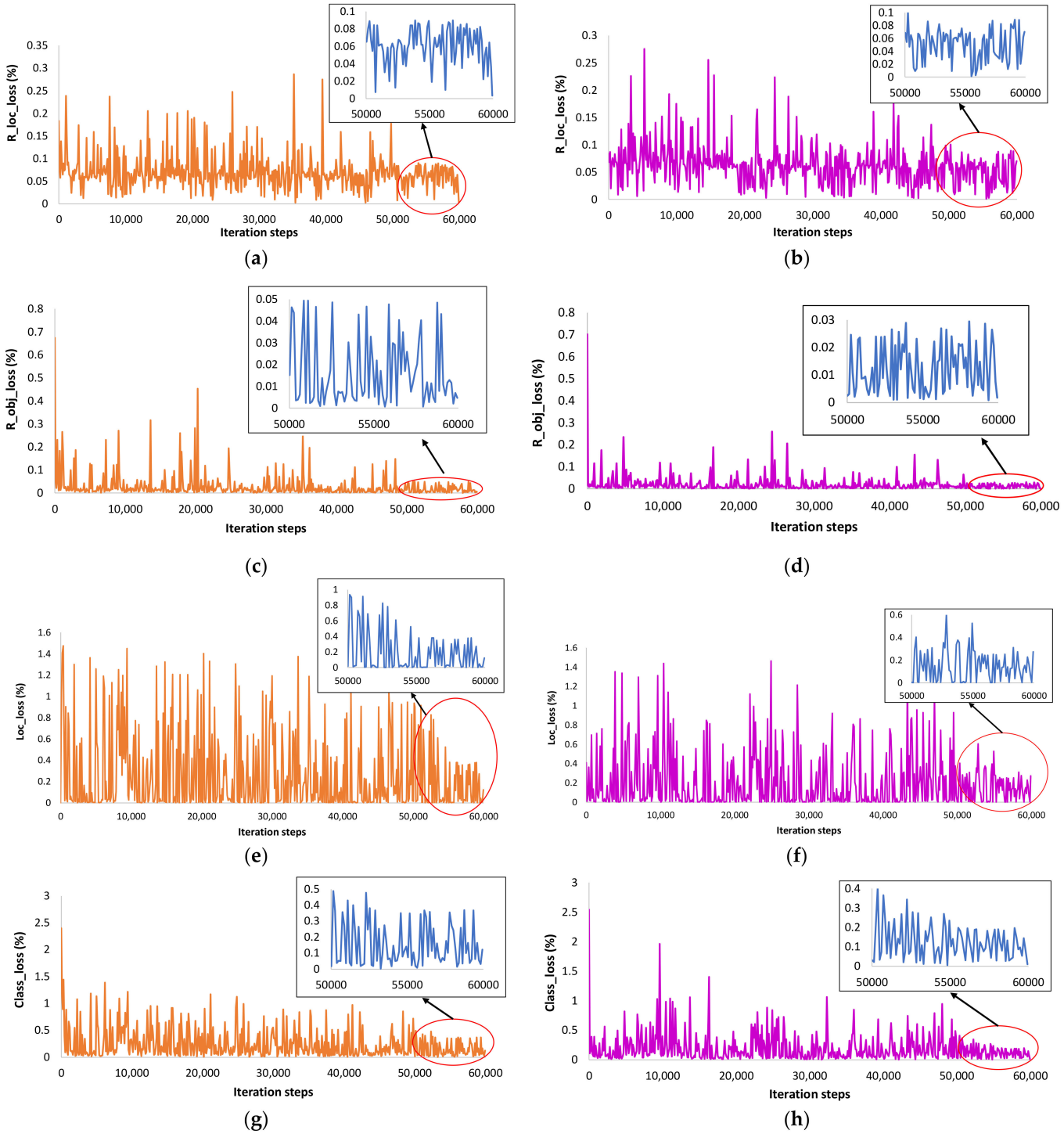


Figure 14. Training plots after adding different aspect ratios to the default: (**a**) presents the RPN localization loss with 1.5:1; (**b**) presents the RPN localization loss with 1:3 and 3:1; (**c**) presents the RPN objectness loss with 1.5:1; (**d**) presents the RPN objectness loss with 1:3 and 3:1; (**e**) presents the box classifier localization loss with 1.5:1; (**f**) presents the box classifier localization loss with 1:3 and 3:1; (**g**) presents the box classifier classification with 1.5:1; (**h**) presents the box classifier classification loss with 1:3 and 3:1.

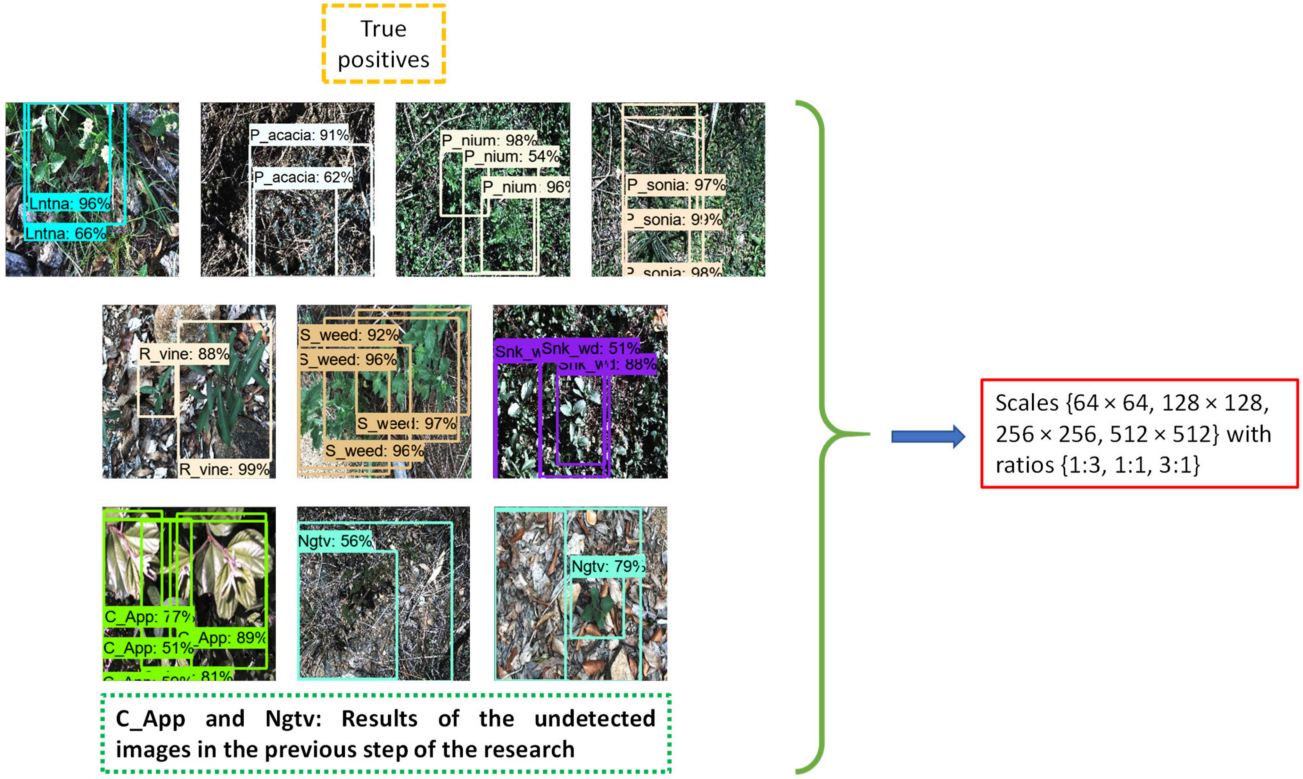


Figure 15. True-positive results with aspect ratios 1:3 and 3:1.

Table 2. Performance of the proposed method compared to different DL architectures.

DL Models with Respective Feature Extractors	Mean Average Precision (%)
YOLO-v4 CSPDarknet-53	79.68
SSD Inception-v2	48.12
SSD MobileNet	34.99
SSD ResNet-50 (RetinaNet)	21.69
EfficientDet EfficientNet	36.59
CenterNet ResNet-50	27.36
RFCN ResNet-101	55.06
Faster RCNN Inception-v2	73.59
Faster RCNN ResNet-50	87.23
Faster RCNN ResNet-101	87.64
Faster RCNN ResNet-101 (optimized model)	93.44
Faster RCNN ResNet-101 (optimized model + enhanced anchor box approach)	**96.02**

3.2.4. Validation of the Approach

Similar to the first stage of research [30], the effectiveness of this phase of the work has also been validated using a stratified five-fold cross-validation method. The final mAP obtained through the enhanced anchor box was validated in four other folds of the dataset (presented in our repository https://github.com/kmarif/DL-Weed-Identification, accessed on 20 May 2022). The presented analysis was performed in the first fold (fold1). A slight

difference in the mAP was obtained with the margin of 0.07 to 0.38% and achieved 95.95%, 96.19, 96.40%, and 95.80% by fold2, fold3, fold4, and fold5, respectively. Furthermore, the difference in AP of the Chinee apple was also not significant in all folds, with 94.27, 92.66, 92.34, and 94.95% of AP with the final enhanced anchor boxes.

To further show the robustness of the proposed research, a small external testing dataset has been generated by random google search and tested using the weights obtained by the final proposed model. The mAP is 95.83% with each class attaining AP > 90%. Figure 16 shows a sample of each detected weed and a non-weed class.

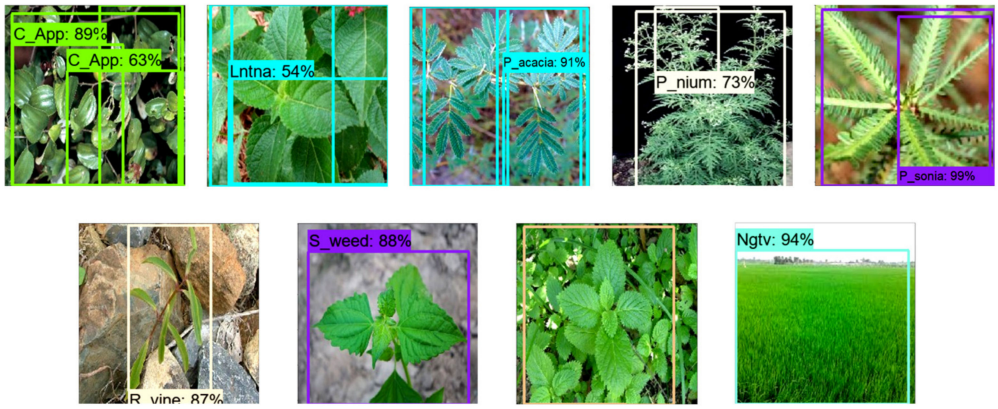


Figure 16. A sample of each class by externally generated dataset.

4. Conclusions

The architectural details of the Faster RCNN model provided a solid motivation to propose an enhanced anchor box approach to successfully identify and localize weeds. The effects of various anchor box scale sizes and aspect ratios on the training and testing performance of the model have been presented. The addition of a 64 × 64 scale size and replacement of the default aspect ratio with 1:3 and 3:1 attained the optimum results. A significant improvement of 24.95% in the average precision of the Chinee apple weed was achieved that was not successfully detected in the previous stage of the study. Furthermore, the mean average precision was 2.58% better compared to the first step of the research. In addition to that, the AP of the remaining weed classes was also maintained, which proved the effectiveness of the method. Moreover, the AP of the negative class was also improved by 6.16%. Furthermore, the successful detection results were obtained without compromising the detection time, which shows the practicality of the work. Moreover, the robustness of the work has been presented by the stratified k-fold cross-validation method and external testing dataset with a small difference of 0.38% and 0.19% in mAP, respectively. This study has provided another way to visualize and analyze deep learning-based object detectors for further development in agricultural tasks.

Although the proposed methodology has been validated by two techniques, a similar approach should also be tested on other weed classes/datasets. In the future, segmentation-based DL models can be explored to perform pixel-wise detection of weeds. Moreover, the proposed methodology for anchor box enhancement has the potential to be applied to other relevant datasets. The final weights of the proposed model could be reused for real-time weed detection in field trials using a portable system. It would be beneficial for the precise application of herbicide sprays on weeds.

Author Contributions: Conceptualization, methodology, investigation, visualization, M.H.S. and K.M.A.; software, validation, writing—original draft preparation, M.H.S.; writing—review and editing, M.H.S. and K.M.A.; project administration, supervision, funding acquisition, J.P. and K.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Business, Innovation, and Employment (MBIE), New Zealand, Science for Technological Innovation (SfTI) National Science Challenge.

Data Availability Statement: The data presented in this study are made publicly available at https://github.com/kmarif/DL-Weed-Identification, (accessed on 20 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hasan, A.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [CrossRef]
2. Liu, J.; Wang, X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* **2021**, *17*, 1–18. [CrossRef] [PubMed]
3. Saleem, M.H.; Potgieter, J.; Arif, K.M. Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precis. Agric.* **2021**, *22*, 2053–2091. [CrossRef]
4. Wan, S.; Goudos, S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* **2020**, *168*, 107036. [CrossRef]
5. Quiroz, I.A.; Alférez, G.H. Image recognition of Legacy blueberries in a Chilean smart farm through deep learning. *Comput. Electron. Agric.* **2020**, *168*, 105044. [CrossRef]
6. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
7. Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y. Review of Weed Detection Methods Based on Computer Vision. *Sensors* **2021**, *21*, 3647. [CrossRef]
8. Tao, T.; Wei, X. A hybrid CNN–SVM classifier for weed recognition in winter rape field. *Plant Methods* **2022**, *18*, 1–12. [CrossRef]
9. Jiang, H.; Zhang, C.; Qiao, Y.; Zhang, Z.; Zhang, W.; Song, C. CNN feature based graph convolutional network for weed and crop recognition in smart farming. *Comput. Electron. Agric.* **2020**, *174*, 105450. [CrossRef]
10. Hu, K.; Coleman, G.; Zeng, S.; Wang, Z.; Walsh, M. Graph weeds net: A graph-based deep learning method for weed recognition. *Comput. Electron. Agric.* **2020**, *174*, 105520. [CrossRef]
11. Teimouri, N.; Jørgensen, R.N.; Green, O. Novel Assessment of Region-Based CNNs for Detecting Monocot/Dicot Weeds in Dense Field Environments. *Agronomy* **2022**, *12*, 1167. [CrossRef]
12. Yu, J.; Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. *Pest Manag. Sci.* **2019**, *75*, 2211–2218. [CrossRef]
13. Ahmad, A.; Saraswat, D.; Aggarwal, V.; Etienne, A.; Hancock, B. Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. *Comput. Electron. Agric.* **2021**, *184*, 106081. [CrossRef]
14. Zhuang, J.; Li, X.; Bagavathiannan, M.; Jin, X.; Yang, J.; Meng, W.; Li, T.; Li, L.; Wang, Y.; Chen, Y. Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Manag. Sci.* **2021**, *78*, 521–529. [CrossRef]
15. Gao, J.; French, A.P.; Pound, M.P.; He, Y.; Pridmore, T.P.; Pieters, J.G. Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields. *Plant Methods* **2020**, *16*, 1–12. [CrossRef]
16. Jin, X.; Che, J.; Chen, Y. Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access* **2021**, *9*, 10940–10950. [CrossRef]
17. Espejo-Garcia, B.; Mylonas, N.; Athanasakos, L.; Fountas, S.; Vasilakoglou, I. Towards weeds identification assistance through transfer learning. *Comput. Electron. Agric.* **2020**, *171*, 105306. [CrossRef]
18. Chen, D.; Lu, Y.; Li, Z.; Young, S. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput. Electron. Agric.* **2022**, *198*, 107091. [CrossRef]
19. Suh, H.K.; Ijsselmuiden, J.; Hofstee, J.W.; van Henten, E.J. Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosyst. Eng.* **2018**, *174*, 50–65. [CrossRef]
20. Bah, M.D.; Hafiane, A.; Canals, R. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sens.* **2018**, *10*, 1690. [CrossRef]
21. de Camargo, T.; Schirrmann, M.; Landwehr, N.; Dammer, K.-H.; Pflanz, M. Optimized Deep Learning Model as a Basis for Fast UAV Mapping of Weed Species in Winter Wheat Crops. *Remote Sens.* **2021**, *13*, 1704. [CrossRef]
22. Ukaegbu, U.F.; Tartibu, L.K.; Okwu, M.O.; Olayode, I.O. Development of a Light-Weight Unmanned Aerial Vehicle for Precision Agriculture. *Sensors* **2021**, *21*, 4417. [CrossRef]
23. Veeranampalayam Sivakumar, A.N.; Li, J.; Scott, S.; Psota, E.; Jhala, A.J.; Luck, J.D.; Shi, Y. Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery. *Remote Sens.* **2020**, *12*, 2136. [CrossRef]
24. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Anwar, S. Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precis. Agric.* **2021**, *22*, 1711–1727. [CrossRef]
25. Ruigrok, T.; van Henten, E.; Booij, J.; van Boheemen, K.; Kootstra, G. Application-specific evaluation of a weed-detection algorithm for plant-specific spraying. *Sensors* **2020**, *20*, 7262. [CrossRef]
26. Kounalakis, T.; Triantafyllidis, G.A.; Nalpantidis, L. Deep learning-based visual recognition of rumex for robotic precision farming. *Comput. Electron. Agric.* **2019**, *165*, 104973. [CrossRef]
27. Khan, A.; Ilyas, T.; Umraiz, M.; Mannan, Z.I.; Kim, H. Ced-net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics* **2020**, *9*, 1602. [CrossRef]

28. Wang, A.; Xu, Y.; Wei, X.; Cui, B. Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access* **2020**, *8*, 81724–81734. [CrossRef]
29. Zou, K.; Chen, X.; Wang, Y.; Zhang, C.; Zhang, F. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. *Comput. Electron. Agric.* **2021**, *187*, 106242. [CrossRef]
30. Saleem, M.H.; Velayudhan, K.K.; Potgieter, J.; Arif, K.M. Weed Identification by Single-Stage and Two-Stage Neural Networks: A Study on the Impact of Image Resizers and Weights Optimization Algorithms. *Front. Plant Sci.* **2022**, *13*, 920. [CrossRef]
31. Olsen, A.; Konovalov, D.A.; Philippa, B.; Ridd, P.; Wood, J.C.; Johns, J.; Banks, W.; Girgenti, B.; Kenny, O.; Whinney, J. DeepWeeds: A multiclass weed species image dataset for deep learning. *Sci. Rep.* **2019**, *9*, 2058. [CrossRef] [PubMed]
32. Saleem, M.H.; Khanchi, S.; Potgieter, J.; Arif, K.M. Image-based plant disease identification by deep learning meta-architectures. *Plants* **2020**, *9*, 1451. [CrossRef] [PubMed]
33. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014, Proceedings of the 13th European conference, Zurich, Switzerland, 6–12 September 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 740–755.
34. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
35. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
36. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
37. Hinton, G.; Srivastava, N.; Swersky, K. Neural networks for machine learning. *Coursera Video Lect.* **2012**, *264*, 2146–2153.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: https://arxiv.org/abs/1412.6980 (accessed on 20 May 2022).