AI

MDPI

Article

Enhancing *Tuta absoluta* Detection on Tomato Plants: Ensemble Techniques and Deep Learning

Nikolaos Giakoumoglou [1], Eleftheria-Maria Pechlivani [1,*], Nikolaos Frangakis [2] and Dimitrios Tzovaras [1]

[1] Centre for Research and Technology Hellas, Information Technologies Institute, 57001 Thermi, Greece; ngiakoumoglou@iti.gr (N.G.); dimitrios.tzovaras@iti.gr (D.T.)
[2] iKnowHow S.A., 15451 Athens, Greece; nfrangakis@iknowhow.com
* Correspondence: riapechl@iti.gr; Tel.: +30-231-125-7751

Abstract: Early detection and efficient management practices to control *Tuta absoluta* (Meyrick) infestation is crucial for safeguarding tomato production yield and minimizing economic losses. This study investigates the detection of *T. absoluta* infestation on tomato plants using object detection models combined with ensemble techniques. Additionally, this study highlights the importance of utilizing a dataset captured in real settings in open-field and greenhouse environments to address the complexity of real-life challenges in object detection of plant health scenarios. The effectiveness of deep-learning-based models, including Faster R-CNN and RetinaNet, was evaluated in terms of detecting *T. absoluta* damage. The initial model evaluations revealed diminishing performance levels across various model configurations, including different backbones and heads. To enhance detection predictions and improve mean Average Precision (mAP) scores, ensemble techniques were applied such as Non-Maximum Suppression (NMS), Soft Non-Maximum Suppression (Soft NMS), Non-Maximum Weighted (NMW), and Weighted Boxes Fusion (WBF). The outcomes shown that the WBF technique significantly improved the mAP scores, resulting in a 20% improvement from 0.58 (max mAP from individual models) to 0.70. The results of this study contribute to the field of agricultural pest detection by emphasizing the potential of deep learning and ensemble techniques in improving the accuracy and reliability of object detection models.

Keywords: deep learning; ensembles; Faster R-CNN; object detection; RetinaNet; tomato plants; *Tuta absoluta*; weighted boxes fusion

Citation: Giakoumoglou, N.; Pechlivani, E.-M.; Frangakis, N.; Tzovaras, D. Enhancing *Tuta absoluta* Detection on Tomato Plants: Ensemble Techniques and Deep Learning. *AI* **2023**, *4*, 996–1009. https://doi.org/10.3390/ai4040050

Academic Editor: Arslan Munir

Received: 7 October 2023
Revised: 2 November 2023
Accepted: 16 November 2023
Published: 20 November 2023

1. Introduction

The tomato plant (*Solanum lycopersicum* L.) plays a significant role in global agriculture, serving as a dietary staple and contributing significantly to the world's food systems, with an annual production exceeding 160 million tonnes globally [1]. Additionally, the tomato industry holds economic significance in various market segments such as canned goods, sauces, beverages, and cosmetics, with approximately a quarter of its production dedicated to processing, making it an essential commodity [2].

However, this thriving industry is facing a formidable adversary in the form of the tomato leaf miner, scientific name *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) and henceforth referred to as *T. absoluta* [3]. This pest, native to South America, has managed to proliferate far beyond its region of origin, causing considerable damage to tomato crops worldwide. The invasive pest was first detected in Europe in late 2006 [4] and has spread to North Africa and the Middle East at an exceptional speed. The severity of this infestation is such that it can cause production losses ranging from 80% to 100% [5]. This leads to a ripple effect of economic distress, affecting not just the agricultural sector but also the livelihoods of millions who depend on it [6].

The widespread presence of *T. absoluta* not only threatens the worldwide tomato supply but also discourages growers due to the rising costs of its integrated pest management.

Despite the availability of various pest management methods, such as chemical pesticides, pheromone traps, and the cultivation of resistant tomato varieties [7], effective control of *T. absoluta* spread remains elusive. Therefore, for the surveillance of *T. absoluta*, there is a need for digital-based technologies that leverage deep learning techniques, enabling fast, accurate, and non-invasive early detection of this pest [8].

Recently, the successful application of deep learning models in many fields of computer vision for early pest and disease detection in crops using convolutional neural network (CNN) has presented a unique opportunity for enabling efficient and rapid detection of *T. absoluta* infestation at early stages [9].

In this study, four ensemble techniques were benchmarked to evaluate their performance in improving the detections of *T. absoluta* infestations on tomato plants, as generated by deep learning object detection models. A dataset captured under real-life conditions in open fields and greenhouses was utilized, presenting various challenges in detecting *T. absoluta* damage. The outputs from multiple configurations of the employed Faster R-CNN and RetinaNet deep learning models, combined with ensemble techniques such as Non-Maximum Suppression (NMS), Soft Non-Maximum Suppression (Soft NMS), Non-Maximum Weighted (NMW), and Weighted Boxes Fusion (WBF) exhibited improved performance. In particular, the fusion of detection results using WBF, considering the detection scores, effectively improved the mean Average Precision (mAP) metric by 20%, elevating it from 0.58 (the highest observed mAP from individual deep learning models) to 0.70. This study demonstrates the superiority of the WBF ensemble technique in addressing the challenges associated with *T. absoluta* digital detection on tomato plants, highlighting its significance in enhancing the accuracy and reliability of deep learning models for agricultural applications. The innovation in this study lies in the combination of established object detection models with ensemble techniques, which reveals an improvement in *T. absoluta* detection accuracy under real-world agricultural settings. This approach not only establishes a robust method for pest detection but also extends the applicability of deep learning models in precision agriculture, aligning with the broader goal of integrated pest management.

The rest of the paper is structured as follows. Section 2 presents the related work in the field of tomato disease and pest identification and *T. absoluta* detection. Section 3 delineates the methodology employed in this study, encompassing the dataset, the object detection models, the ensemble techniques utilized, and the evaluation metrics adopted for detecting the impacts of *T. absoluta* on tomato plants. Section 4 exhibits the results. The study concludes in Section 5.

2. Related Work

In this section, the application of deep learning techniques for the identification of tomato diseases and pests, as well as the detection of *T. absoluta* effects, is reviewed. Significant progress has been made in agricultural applications by employing deep learning methods. In particular, the identification of diverse plant diseases and pests has yielded promising and impressive outcomes, as evidenced by recent studies [10–12].

2.1. Tomato Disease and Pest Identification

Tomato disease and pest identification have been actively investigated in the field of deep learning. Nevertheless, it is crucial to acknowledge that the nature of the tasks addressed in these investigations differs, with certain studies emphasizing classification while others prioritize object detection. It is worth noting that, in real-world applications, classification tends to present limitations and poses a comparatively simpler challenge compared to detection.

Wang et al. (2019) [13] used Faster R-CNN and Mask R-CNN in tomato disease detection. The models achieved a mAP of 88.53% and 99.64% for Faster R-CNN and Mask R-CNN, respectively. Wang et al. (2021) [14] used a YOLOv3-tiny network on a self-built tomato disease and pest dataset and dealt with light shadow, branch occlusion,

and overlapping leaves conditions. Yet, it should be noted that the objects depicted in the images were considerably prominent in size, which significantly facilitated the detection task by providing distinct and easily identifiable visual features. Fuentes et al. (2017) [15] used a deep-learning-based approach to detect diseases and pests in tomato plants using images captured in-place by camera devices with various resolutions. Using Faster R-CNN, R-FCN, and SSD, combined with various backbones such as VGG and ResNets, the authors demonstrated the performance of these models on images with diseases and pests.

Other studies primarily focused on the classification of tomato diseases rather than the detection of specific regions. Zaborowicz et al. (2017) [16] explored the application of computer image analysis and artificial neural modeling for assessing the quality of greenhouse tomatoes. Zaborowicz et al. (2013) [17] aimed to devise a neural network classification model for the automatic evaluation of greenhouse tomato quality, leveraging computer image analysis and artificial neural networks. Brahimi et al. (2017) [18] presented deep models, including AlexNet and GoogleNet, which were trained using a large dataset of 14,828 images to identify nine tomato diseases. Ferentinos (2018) [19] employed several deep models, namely AlexNet, GoogLeNet, and VGG, to recognize 58 diseases from a dataset consisting of 87,848 leaf images of various plants. Liang et al. (2019) [20] proposed a multitasking system based on the ResNet-50 architecture capable of diagnosing diseases, recognizing plant species, and estimating the severity of diseases. Rangarajan et al. (2018) [21] used two pre-trained deep learning models, VGG-16 and AlexNet, to classify six tomato diseases. Trivedi et al. (2021) [22] used CNNs to effectively define and classify tomato diseases with 98.49% accuracy.

In the field of tomato disease and pest identification, there is a distinguishable difference between classification and detection tasks, despite both utilizing deep learning models. Detection-focused studies, such as those by Wang et al. (2019) [13] and Fuentes et al. (2017) [15], are aimed at identifying and localizing specific affected areas within images, dealing with complexities like varied lighting and occlusions. Conversely, classification-centric research, exemplified by works by Brahimi et al. (2017) [18] and Ferentinos (2018) [19], is focused on categorizing images based on the type of disease or pest present, without localizing the affected areas. These differences underline the varied approaches and complexities within the broader scope of automated identification of plant diseases and pests.

2.2. Tuta Absoluta Identification

Tuta absoluta identification has been a subject of interest in the field of deep learning. However, while several studies have focused on the classification of *T. absoluta* effects, limited research has addressed the specific task of *T. absoluta* detection using object detection techniques. Notably, detection poses a more complex challenge compared to classification and carries greater relevance in real-world applications.

Among the studies that have explored *T. absoluta* detection, Loyani et al. (2021) [23] inoculated *T. absoluta* on randomly selected tomato plants and created a dataset of 5235 tomato images. The authors used semantic and instance segmentation models based on U-Net and Mask R-CNN to segment the effects of *T. absoluta* on tomato leaf images at the pixel level using field data. The Mask R-CNN achieved a mAP of 85.67%, while the U-Net model achieved an Intersection over Union (IoU) of 78.60% and Dice coefficient of 82.86%. Nonetheless, the large size of the depicted objects facilitated the detection task with distinct visual features. Georgantopoulos et al. (2023) [24] presented a dataset of multispectral images of tomato plants at various stages of infection with *Tuta absoluta* and Leveillula taurica. The authors applied a Faster-RCNN object detector for the localization and classification of lesions, achieving a 90% F1-score on binary classification, and 20.2% mAP on detection.

However, existing research on *T. absoluta* primarily focuses on its classification or the classification of the damage it causes. Rubanga et al. (2020) [25] determined the severity status of *T. absoluta*'s damage on tomato plants at an early stage of tomato growth. The authors pre-trained four CNN architectures to classify healthy and infested tomato leaves collected from real field experiments with an average accuracy of 87.2%. Mkonyi et al. (2020) [26] leveraged CNN network architectures for the identification of *T. absoluta* in

the tomato plant. The approach reinforces the classification of a dataset that consists of 2145 images with an accuracy of 91.9% on the tested images.

Research on *Tuta absoluta* identification via deep learning exhibits a division between classification and detection. Loyani et al. (2021) [23] and Georgantopoulos et al. (2023) [24] delve into object detection, focusing on identifying and localizing the impacts of *T. absoluta* on tomato plants, a task accentuated by its complexity and applicative relevance. Conversely, works by Rubanga et al. (2020) [25] and Mkonyi et al. (2020) [26] are anchored in classification, assessing the severity and identifying the presence of *T. absoluta* damage without specific localization. Despite the shared foundation in deep learning, a clear distinction emerges from the depth of analysis and practical implications each approach entails.

3. Methodology

The methodology section outlines the dataset utilized, details the object detection models implemented, and describes the evaluation metrics adopted to assess the effectiveness of these models in the detection of *T. absoluta* infestation on tomato plants. The model is tailored to run on a mobile platform aboard a fully autonomous mobile robot [27]. The collected data is processed within a robust decision support system [28] to ensure precise detection and evaluation of the infestations.

3.1. Dataset

The dataset used in this study consists of color images collected under real-life conditions in both fields and greenhouses depicting *T. absoluta* damage on tomato plants. These images were obtained by the EDEN Library (https://edenlibrary.ai (accessed on 1 March 2023)) in collaboration with its partners, ensuring the clear distinction of *T. absoluta* damage on tomato plant cultivation while preserving background information. The collection process involved farmers using mobile phones and other devices to capture images from various angles and under different lighting conditions. This collection style is consistent with typical precision farming and robotic tasks, and it provides a suitable basis for training deep learning models to detect *T. absoluta* infestations under diverse scenarios in real settings. The dataset's robustness is enhanced by the intentional inclusion of images captured under non-optimal illumination conditions and with variable resolution and color balance. Any images that did not meet the high standards set by experts were excluded from the dataset.

The dataset comprises a total of 659 images, with 396 reserved for training and 263 designated for validation. The average size of the training images is $(1964 \pm 1068) \times (1863 \pm 802)$ with a median of 1504×1504 (ranging from 960×720 to 4640×4000). The validation images have a mean size of $(1772 \pm 801) \times (1769 \pm 725)$ and a median of 1504×1504 (ranging from 1200×1504 to 4640×4000).

Each of the 659 images has been annotated with bounding boxes that enclose areas of *T. absoluta* damage. These annotations were executed by expert agronomists from the EDEN Library. All annotations underwent manual review and adjustment to ensure their consistency and accuracy. The training set contains 5443 annotations while the validation set includes 3267, totaling 8710 annotations. This equates to an average of 13.83 annotations per image across the one class. The annotation area for the training set has a mean size of 14,508 $\pm$ 31,806 pixels and a median of 5764, while the validation set's annotation area has a mean size of 12,234 $\pm$ 24,857 pixels and a median of 5304. Therefore, it can be deduced that the mean object, as indicated by the bounding boxes, captures approximately 0.39% (as far as training set is concerned). For a visual reference, an example image depicting *T. absoluta* damage on tomato leaves, annotated with bounding boxes, is presented in Figure 1. The aforementioned details highlight the inherent difficulty associated with addressing the object detection task within this dataset.

Figure 1. Sample featuring bounding box annotations that point out *T. absoluta* damage (class 0) on tomato leaves under various conditions and illuminations. Six different images are displayed, each containing annotations highlighting the damage. The images have been resized to a resolution of 1024×1024 pixels.

3.2. Object Detection Models

Two popular object detection models, Faster R-CNN [29] and RetinaNet [30], were utilized in this study to detect *T. absoluta* damage in tomato crops.

Faster R-CNN is a pivotal model in the domain of object detection that efficiently handles the task of object localization and classification in images. Its architecture is orchestrated in a manner that first employs a deep convolutional network as a backbone to extract a rich set of feature maps from the input image. Following the backbone, a Region Proposal Network (RPN) operates as the neck, generating a set of object proposals with associated objectness scores, significantly speeding up the detection process while maintaining a high level of accuracy. These proposals are then forwarded to the head of the model, where two sibling layers perform classification and bounding box regression to accurately identify and localize objects within the image.

For the Faster R-CNN models, configurations were established using ResNet-50 and ResNet-101 backbones [31] with Feature Pyramid Network (FPN), Convolutional Layer 4 (C4), and Dilated Convolutional Layer 5 (DC5) heads. The FPN head, paired with a ResNet+FPN backbone, was used with standard convolutional (conv) and fully connected (FC) heads for box prediction due to its optimal speed/accuracy tradeoff [32]. The C4 head employed a ResNet conv4 backbone with a conv5 head, following the original baseline configuration in the Faster R-CNN paper [29]. Lastly, the DC5 head utilized a ResNet conv5 backbone with dilations in conv5, as well as standard conv and FC heads for box prediction.

Meanwhile, RetinaNet is recognized for its effectiveness in detecting objects across a wide range of scales. Its architecture also comprises a backbone for feature extraction, an FPN neck for multi-scale feature aggregation, and a head with two sub-heads for classification and bounding box regression. Notably, RetinaNet introduces a novel loss function, the Focal Loss, designed to focus training on hard negative examples and thereby mitigate the foreground–background class imbalance.

RetinaNet configurations were set up with ResNet-50 and ResNet-101 backbones [31], complemented with an FPN head. The choice of these configurations was informed by their demonstrated efficiency in various object detection tasks, thus enabling a comprehensive evaluation of their performance in the detection of *T. absoluta* infestations.

The employment of different backbones and heads in these models serves to explore the speed–accuracy trade-off and the level of detail captured in the object detection task. For instance, a ResNet-101 backbone, with its deeper architecture, is expected to capture more intricate features compared to a ResNet-50 backbone, albeit at a higher computational cost. Similarly, the choice of head (FPN, C4, or DC5) influences the model's capability to handle varying object scales and complexities, which is crucial for accurately detecting *T. absoluta* damage in diverse real-world agricultural scenarios.

3.3. Ensemble Techniques

The process of combining predictions from the multiple object detection models was achieved using several ensemble techniques, namely Non-Maximum Suppression (NMS), Soft Non-Maximum Suppression (Soft NMS) [33], Non-Maximum Weighted (NMW) [34], and Weighted Boxes Fusion (WBF) [35].

NMS is a fundamental algorithm in object detection that removes duplicate detections by keeping only the bounding box with the highest confidence score and suppressing others that have a high IoU overlap with it. Soft NMS, an improvement over NMS, instead of discarding overlapping detections entirely, decreases their confidence scores based on the degree of overlap [33]. The NMW method uses a weighted combination of the overlapping detections to create a more accurate bounding box [34]. The weights are calculated based on the confidence scores of the detections, resulting in a bounding box that better represents the detections. The WBF method uses the detection score to obtain a weighted box [35]. This allows for a fused result that reflects all detection results according to the score, unlike existing methods that use only one result with the highest score when multiple models detect the same object.

All ensemble techniques were applied post the training of the models; hence, no additional training was required. The ensemble process was executed by aggregating the prediction outputs of the individual trained models. This strategy not only conserves computational resources by avoiding the need for further training but also enhances the robustness and accuracy of the detection process, making it a pragmatic approach for real-world agricultural applications [36].

3.4. Object Detection and Ensemble Workflow

Figure 2 illustrates the workflow employed for object detection and ensemble aggregation in this study. The process initiates with an input image which is fed into multiple object detection models. Each of these models processes the image independently to identify and localize *T. absoluta* infestations, producing images annotated with bounding boxes around detected instances of damage. The individual outputs from these models are then aggregated using an ensemble technique that merges the detection results, leveraging the strengths of each model to enhance accuracy and reliability. This ensemble approach effectively combines the diverse detection capabilities of the employed models, ensuring a more comprehensive representation of *T. absoluta* damage instances within the input image. The final output of this workflow is an image with aggregated bounding boxes around detected instances of *T. absoluta* damage, representing a consensus among the models employed.

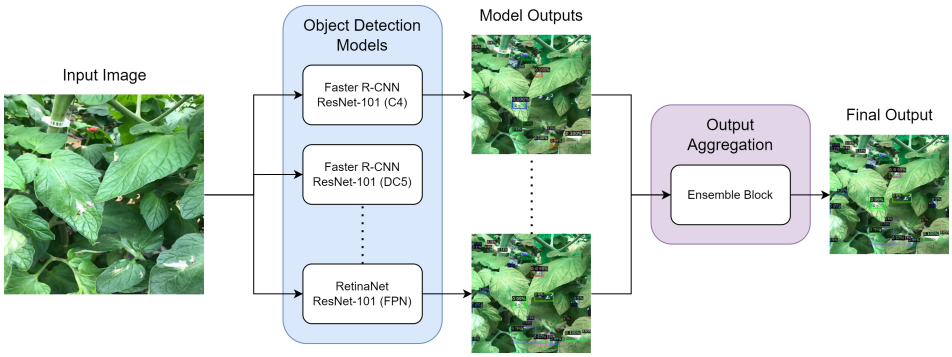

Figure 2. Object detection and ensemble workflow: the diagram delineates the process from input image, through individual model processing (Faster R-CNN, RetinaNet), to ensemble aggregation, producing the final output with aggregated bounding boxes around detected *T. absoluta* damage instances.

3.5. Object Detection Model Training

The object detection models were trained using a specific set of hyperparameters. Initially, each backbone was pre-trained on the COCO 2017 dataset [37]. The models were then trained on an image input size of 1024 × 1024 for a total of 20,000 iterations with a batch size of 2. The Stochastic Gradient Descent (SGD) optimization algorithm was used for model training. The learning rate was set at 0.001, with a momentum of 0.9 and a weight decay of 1×10^{-5}. In order to facilitate better learning, the learning rate was reduced by 0.1 at both the 80% and 90% completion stages of the total iterations. This strategy is known to help models converge more effectively towards the end of the training process. To improve model robustness and ensure that the models could generalize to new data, several data augmentation techniques were applied during training. These included vertical flipping, horizontal flipping, rotation, and Contrast Limited Adaptive Histogram Equalization (CLAHE) [38].

All models were trained using the Detectron2 framework [39] in PyTorch [40]. The training was executed on a PC equipped with an Intel Xeon CPU, an NVIDIA Tesla K80 GPU with 16GB of VRAM, and 12 GB of RAM.

3.6. Evaluation Metrics

The evaluation of the object detection models in this study was performed using the COCO detection metrics [37]. These metrics have been widely adopted in the field for their comprehensive evaluation of both localization and classification aspects of object detection. The primary metric of interest in this study was mAP_{50}, which represents the mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 across all classes (in this paper, the terms mAP and mAP_{50} are used interchangeably, indicating the same concept).

4. Results and Discussion

4.1. Initial Model Evaluations

The initial performance evaluation, as summarized in Table 1, revealed varying levels of effectiveness among the object detection models, taking both accuracy and inference time (in seconds per image) into consideration. An mAP_{50} score of 0.56 was achieved by the Faster R-CNN model with a ResNet-101 backbone and C4 head, with an inference time of 0.139 s per image. Its DC5 and FPN variants completed inferences in 0.086 and 0.051 s per image, respectively, and yielded mAP_{50} scores of 0.58 and 0.56.

Faster R-CNN with a ResNet-50 backbone and C4 head reached an mAP_{50} score of 0.58, taking 0.103 s per image for inference. The DC5 and FPN heads on the same backbone completed their inferences in 0.069 and 0.038 s per image, respectively, with mAP_{50} scores

of 0.58 and 0.55. Notably, the Faster R-CNN model with the ResNet-50 backbone and FPN head boasted the fastest inference time among all evaluated configurations.

With a ResNeXt-101 backbone and FPN head, the Faster R-CNN model displayed an mAP_{50} score of 0.56 and took 0.098 s per image for inference. RetinaNet models equipped with ResNet-101 and ResNet-50 backbones, both with FPN heads, showed mAP_{50} scores of 0.57. Their inference times were 0.054 and 0.041 s per image, respectively.

Table 1. Initial performance evaluation (mAP_{50} and inference time in seconds per image) of object detection models across various model configurations, including different backbones and heads.

Model	Backbone	Head	mAP_{50}	Inference Time
Faster R-CNN	ResNet-101	C4	0.56	0.139
Faster R-CNN	ResNet-101	DC5	0.58	0.086
Faster R-CNN	ResNet-101	FPN	0.56	0.051
Faster R-CNN	ResNet-50	C4	0.58	0.103
Faster R-CNN	ResNet-50	DC5	0.58	0.069
Faster R-CNN	ResNet-50	FPN	0.55	0.038
Faster R-CNN	ResNeXt-101	FPN	0.56	0.098
RetinaNet	ResNet-101	FPN	0.57	0.054
RetinaNet	ResNet-50	FPN	0.57	0.041

4.2. Qualitative Results

Detecting *T. absoluta* damage on tomato plants is a challenging task due to the inherent imperfections of individual detection models. When multiple object detection models are employed, it is common for certain instances of damage to be detected by one model while being missed by others (Figure 3). Such inconsistencies in detection results can significantly impact the overall accuracy and reliability of the detection process. To address this limitation, ensemble techniques offer a viable solution for improving *T. absoluta* detection. By combining the outputs of multiple models, ensemble methods can effectively encompass a wider range of *T. absoluta* damage instances, capitalizing on the strengths of different models and compensating for their respective shortcomings.

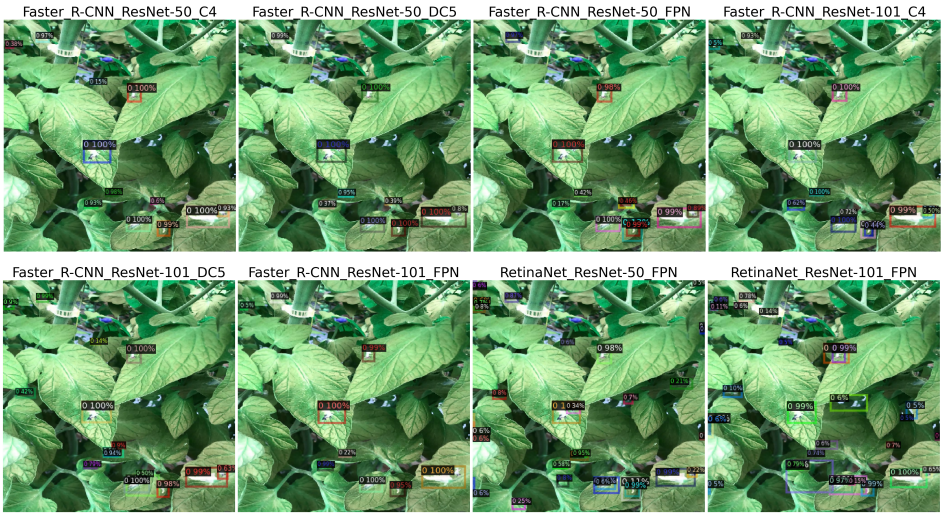

Figure 3. Example of *T. absoluta* damage instances identified by various models. The image represents the top left image of Figure 1, demonstrating varying bounding box predictions by each model. Each bounding box is annotated with its respective class, and class 0 signifies damage attributed to *T. absoluta*, accompanied by the predicted probability for each identified object.

Individual models may fail to accurately detect *T. absoluta* infestations due to various factors. The limitations often arise from the model's architecture, training data, and the inherent variability in real-world agricultural imagery [36]. Furthermore, the model's

ability to generalize might be hindered if the training data is not diverse enough to capture the range of conditions under which *T. absoluta* damage occurs [41]. Furthermore, the hyperparameters and the architecture of the model might not be optimal for the specific task of detecting *T. absoluta* infestations on tomato plants [36].

4.3. Ensemble Results

In an effort to improve the object detection performance, four ensemble techniques were benchmarked to assess their effectiveness in enhancing the detection of *T. absoluta* infestations on tomato plants. Different configurations of models were ensembled using NMS, Soft NMS, NMW, and WBF. NMS was configured with an IoU threshold of 0.1, a sigma of 0.01, and a box-skipping threshold of 0.01. For Soft NMS, only the box-skipping threshold of 0.01 was applied. NMW utilized an IoU threshold of 0.4 and a box-skipping threshold of 0.01. Lastly, the WBF method was implemented with an IoU threshold of 0.5 and a threshold of 0.01 for skipping a box.

The ensemble results, as depicted in Table 2, demonstrated the impact of different ensemble configurations on the mAP_{50} scores and the associated inference time. The total inference time for each ensemble configuration is calculated as $T_{\text{total}} = \sum_{i=1}^{N} T_i + T_{\text{ensemble}}$, where T_{total} is the total inference time, T_i is the inference time of the ith model, T_{ensemble} is the fixed additional latency due to the ensemble technique, and N is the total number of models in the ensemble.

A detailed examination of the ensemble techniques uncovered the specific average times associated with each. The NMS method recorded an average inference time of 1.9046×10^{-5} s. The Soft NMS recorded an average time of 4.2064×10^{-5} s. The NMW method emerged as faster with an average time of 3.5040×10^{-5} s. Impressively, the WBF method was the fastest, clocking an average time of 1.0002×10^{-5} s. However, these ensemble techniques contribute a slight increase to the total inference time of the processing.

The ensemble of all models resulted in mAP_{50} scores of 0.70, 0.66, 0.63, and 0.61 for WBF, NMW, Soft NMS, and NMS, respectively. When ensembling only Faster R-CNN models, the scores were slightly lower: 0.69 for WBF, 0.66 for NMW, 0.63 for Soft NMS, and 0.61 for NMS. Ensembling the RetinaNet models, the scores were 0.64, 0.63, 0.60, and 0.59 for WBF, NMW, Soft NMS, and NMS, respectively. For ensembles of Faster R-CNN models with a ResNet-101 backbone, the scores ranged from 0.64 for WBF to 0.59 for NMS. When only the Faster R-CNN models with a ResNet-50 backbone were ensembled, the scores increased to 0.67 for WBF and 0.60 for NMS. The ensemble of all Faster R-CNN models with a C4 head resulted in scores of from 0.63 for WBF to 0.59 for NMS. For the DC5 head, the scores were similar, ranging from 0.63 for WBF to 0.60 for NMS. Lastly, the ensemble of all Faster R-CNN models with an FPN head produced the lowest scores, ranging from 0.61 for WBF to 0.56 for NMS.

The success of ensemble models in this study is clearly illustrated in Figure 4, which demonstrates the superior performance of ensemble models in terms of mAP_{50} scores compared to individual models, despite the inference time. For a balance between accuracy and inference time, the ensemble of RetinaNet models lies on the top left of Figure 4, showcasing a favorable trade-off. However, for enhanced accuracy, ensembles involving all models or just Faster R-CNN models are preferred as they achieve higher mAP_{50} scores. In particular, the WBF ensemble stands out in achieving a higher mAP_{50} score, showcasing the advantage of aggregating detection results from different models.

These ensemble techniques, namely NMS, Soft NMS, NMW, and WBF, effectively merge the detection results from different models, leading to a more robust and accurate detection of *T. absoluta* damage. The aggregated knowledge from different models provides a broader understanding of the data, which is crucial for handling the variability and complexities inherent in real-world agricultural imagery [42]. Ensemble models often outperform single models as they benefit from the diversity among the models in the ensemble [43]. The diversity can arise from differences in the model architectures, training data, or the learning algorithms, which contribute to a more comprehensive and reliable

detection performance [36]. The fusion of detection results from diverse models, facilitated by ensemble techniques, forms a more complete representation of the data, thus enhancing the model's ability to accurately detect *T. absoluta* infestations across varying conditions encountered in agricultural settings.

Table 2. Ensemble results (mAP_{50}) using different configurations of models and ensemble techniques.

Ensemble Configuration	WBF	NMW	Soft NMS	NMS
All models	0.70	0.66	0.63	0.61
All Faster R-CNN	0.69	0.66	0.63	0.61
All RetinaNet	0.64	0.63	0.60	0.59
All Faster R-CNN ResNet-101	0.64	0.62	0.60	0.59
All Faster R-CNN ResNet-50	0.67	0.64	0.61	0.60
All Faster R-CNN C4	0.63	0.61	0.60	0.59
All Faster R-CNN DC5	0.63	0.62	0.60	0.60
All Faster R-CNN FPN	0.61	0.60	0.58	0.56

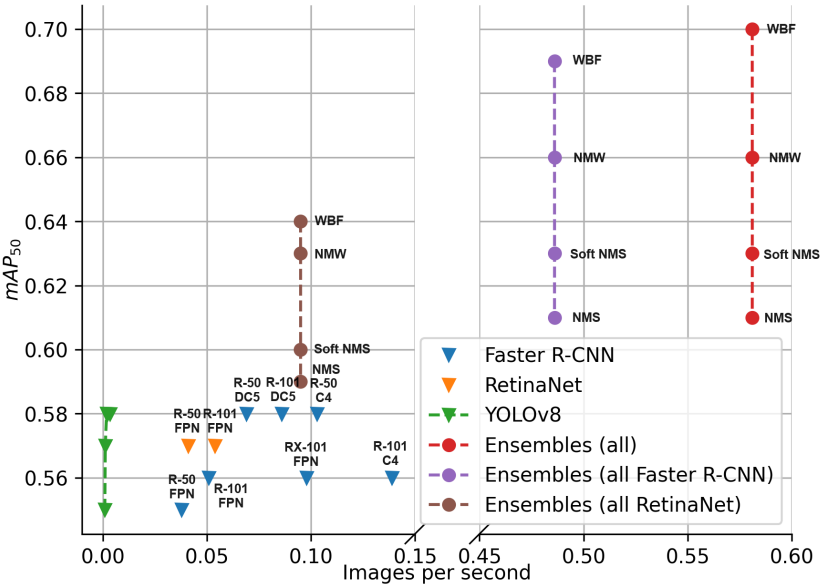

Figure 4. Comparison of mAP_{50} scores versus inference time (images per second) among individual and ensemble models. Higher mAP_{50} values indicate better model performance in accurately detecting *T. absoluta* damage.

4.4. Comparison with State-of-the-Art

The results generated in this study are significant, particularly when compared with those obtained by Loyani et al. (2021) [23] and Georgantopoulos et al. (2023) [24]. Loyani et al. managed to achieve a high mAP_{50} of 0.8567 using Mask R-CNN for detecting *T. absoluta* effects, which was facilitated by the large size of the depicted objects in their dataset. However, even though this study tackles a more challenging detection scenario, the ensemble methods, especially the WBF, managed to achieve an mAP_{50} score of 0.70, showcasing the robustness and effectiveness of the employed ensemble techniques.

In addition, Georgantopoulos et al. leveraged multispectral images, which provide richer spectral information compared to color images, yet their Faster-RCNN model only achieved a 0.202 mAP_{50} on detection. This stark contrast in performance, compared to the 0.70 mAP_{50} score achieved with the ensemble of all models in this study, underscores the significance of the ensemble techniques deployed. The refined configurations in the ensemble methods, such as combining models with different backbones and heads, demonstrated an ability to adjust to specific attributes of the data, thereby optimizing detection performance.

4.5. Benchmark Analysis with YOLOv8

For a more comprehensive benchmark analysis, additional state-of-the-art models were examined alongside the proposed ensemble techniques employed with Faster R-CNN and RetinaNet. Specifically, various versions of YOLOv8 [44], namely YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, were trained and evaluated. Every variant varied in network depth, denoted by a growing number of layers, and network width, demonstrated by an increasing number of filters in these layers. Each model was trained for 100 epochs, with a patience of 15, a batch size of 8 (4 in the case of YOLOv8x), and an image size of 1024×1024. The SGD optimizer was used with a learning rate of 0.01, momentum of 0.9, and a weight decay of 0.0005.

All models were trained using the Ultralytics framework [44] in PyTorch [40]. The training was executed on a PC equipped with an Intel Xeon CPU, an NVIDIA Tesla K80 GPU with 16 GB of VRAM, and 12 GB of RAM.

The results, shown in Table 3, reveal that the YOLOv8 models achieved comparable performance to the models in the initial model evaluations, albeit not surpassing them. This underlines the effectiveness of the ensemble approach proposed in this study, which has shown superior performance in detecting *T. absoluta* infestations on tomato plants, as shown in Figure 4. The ensemble techniques enhanced the mAP_{50} scores, showcasing the advantage of this study's methodology in the realm of agricultural pest detection.

Table 3. Performance evaluation (mAP_{50}, recall, and precision) of YOLOv8 models.

Model	Depth	Width	mAP_{50}	Recall	Precision
YOLOv8n	0.33	0.25	0.55	0.53	0.60
YOLOv8s	0.33	0.50	0.57	0.55	0.61
YOLOv8m	0.67	0.75	0.58	0.54	0.61
YOLOv8l	1.00	1.00	0.58	0.57	0.62
YOLOv8x	1.00	1.25	0.58	0.54	0.65

Using ensemble techniques with Faster R-CNN and RetinaNet models allows this study to draw from the strengths of each model, overcoming the limitations of each when used alone. Unlike using a single model like YOLOv8, the ensemble method combines the detection capabilities of different models, achieving more reliable and accurate detection output. The shared information between Faster R-CNN and RetinaNet models, facilitated by ensemble techniques, provides a better understanding of the data, leading to more accurate detection of *T. absoluta* infestations even in challenging real-world agricultural scenarios.

Moreover, the ensemble approach balances the distinct attributes of Faster R-CNN and RetinaNet, which, when combined, contribute to a higher mAP_{50} score compared to using a single model like YOLOv8. This pairing, arranged by ensemble techniques, results in a detection pipeline that performs better than the YOLOv8 models. The comparative performance of YOLOv8 models in this benchmark analysis further supports the chosen methodology of this study, affirming the better performance and usefulness of ensemble techniques in real-world agricultural pest detection tasks.

5. Conclusions

In conclusion, the detection of *T. absoluta* infestation on tomato plants was addressed in this study using object detection models combined with ensemble techniques. This approach contrasts with previous works by focusing on the challenges of detecting *T. absoluta* infestations in real-life conditions, including open-field and greenhouse environments, and specifically identifying and localizing *T. absoluta* damage on tomato plants.

Initial evaluations of the Faster R-CNN and RetinaNet models, involving different backbones and heads, suggested a need for a refined strategy to enhance performance. The application of ensemble techniques, namely NMS, Soft NMS, NMW, and WBF, successfully improved detection predictions, leading to enhanced mAP scores. In particular,

the employment of WBF led to a significant increment of 20% in the mAP score, from 0.58 to 0.70, highlighting the potential of ensemble methods to address the challenges associated with *T. absoluta* detection.

By enhancing the accuracy and reliability of *T. absoluta* detection using deep-learning-based object detection models combined with ensemble techniques, a contribution is made towards the development of more effective pest management strategies in tomato production. This study underscores the potential of these techniques in the field of agricultural pest detection and provides valuable insights for the development of advanced detection systems, aiding farmers in effectively managing *T. absoluta* infestations and minimizing economic losses in tomato production. Moreover, these technological strides resonate well with the objectives of integrated pest management, emphasizing a shift towards sustainable agriculture. This progression directly supports the Green Deal's ambitious aim of achieving a 50% reduction in pesticide use by 2030.

In future work, the refined models from this study could enable artificial-intelligence-enabled robotic traps to monitor and detect insects in real-time, integrating with decision support systems or mobile apps for immediate alerting to mitigate insect or moth attacks and spreads. Given that these traps capture images periodically, the enhanced accuracy of the models, even at the expense of minimal additional inference time, is a crucial attribute. Moreover, the models can be adapted for autonomous navigated mobile robots or unmanned aerial vehicles equipped for in-field detection and surveillance. Integrated with deep learning models, these robotic systems promise enhanced detection accuracy in precision agriculture, even in low fps modes, providing a sophisticated solution for real-time, on-site pest monitoring and management.

Author Contributions: Conceptualization, N.G. and E.-M.P.; methodology, N.G.; software, N.G.; validation, D.T.; formal analysis, E.-M.P.; investigation, E.-M.P. and N.G.; resources, E.-M.P.; data curation, N.F.; writing—original draft preparation, N.G. and E.-M.P.; writing—review and editing, N.F.; visualization, N.G.; supervision, D.T.; project administration, E.-M.P.; funding acquisition, E.-M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by "Horizon 2020 PestNu project, grant number 101037128". The APC was funded by the Centre for Research and Technology Hellas.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their purchase by the EDEN library.

Conflicts of Interest: The author Nikolaos Frangakis was employed by the company iKnowHow S.A. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

C4	Convolutional Layer 4
COCO	Common Objects in Context
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DC5	Dilated Convolutional Layer 5
FPN	Feature Pyramid Network
Faster R-CNN	Faster Region-Based Convolutional Neural Network
IoU	Intersection over Union

mAP	mean Average Precision
NMS	Non-Maximum Suppression
NMW	Non-Maximum Weighted
ResNet	Residual Network
SGD	Stochastic Gradient Descent
Soft NMS	Soft Non-Maximum Suppression
WBF	Weighted Boxes Fusion
YOLO	You Only Look Once

References

1. Zhang, S.; Griffiths, J.S.; Marchand, G.; Bernards, M.A.; Wang, A. Tomato brown rugose fruit virus: An emerging and rapidly spreading plant RNA virus that threatens tomato production worldwide. *Mol. Plant Pathol.* **2022**, *23*, 1262–1277. [CrossRef] [PubMed]
2. Tomato News. The Global Tomato Processing Industry. 2018. Available online: https://www.tomatonews.com/en/background_47.html (accessed on 1 May 2023).
3. Zekeya, N.; Chacha, M.; Ndakidemi, P.; Materu, C.; Chidege, M.; Mbega, E. Tomato Leafminer (*Tuta absoluta* Meyrick 1917): A Threat to Tomato Production in Africa. *J. Agric. Ecol. Res. Int.* **2017**, *10*, 1–10. [CrossRef]
4. Urbaneja, A.; Vercher, R.; Navarro, V.; Marí, F.G.; Porcuna, J.L. La polilla del tomate, *Tuta absoluta*. *Phytoma España* **2007**, *194*, 16–23.
5. Chidege, M.; Al-zaidi, S.; Hassan, N.; Julie, A.; Kaaya, E.; Mrogoro, S. First record of tomato leaf miner *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in Tanzania. *Agric. Food Secur.* **2016**, *5*, 17. [CrossRef]
6. Guimapi, R.Y.; Mohamed, S.A.; Okeyo, G.O.; Ndjomatchoua, F.T.; Ekesi, S.; Tonnang, H.E. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecol. Complex.* **2016**, *28*, 77–93. [CrossRef]
7. Guedes, R.N.C.; Picanço, M.C. The tomato boreri *Tuta absoluta*/iin South America: Pest status, management and insecticide resistance. *EPPO Bull.* **2012**, *42*, 211–216. [CrossRef]
8. Zahedi, S.R.; Zahedi, S.M. Role of Information and Communication Technologies in modern agriculture. *Int. J. Agric. Crop Sci.* **2012**, *4*, 1725–1728.
9. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [CrossRef]
10. Giakoumoglou, N.; Pechlivani, E.M.; Katsoulas, N.; Tzovaras, D. White Flies and Black Aphids Detection in Field Vegetable Crops using Deep Learning. In Proceedings of the 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS), Genova, Italy, 5–7 December 2022; Volume 5, pp. 1–6. [CrossRef]
11. Giakoumoglou, N.; Pechlivani, E.M.; Sakelliou, A.; Klaridopoulos, C.; Frangakis, N.; Tzovaras, D. Deep learning-based multi-spectral identification of grey mould. *Smart Agric. Technol.* **2023**, *4*, 100174. [CrossRef]
12. Giakoumoglou, N.; Pechlivani, E.M.; Tzovaras, D. Generate-Paste-Blend-Detect: Synthetic Dataset for Object Detection in the Agriculture Domain. *Smart Agric. Technol.* **2023**, *5*, 100258. [CrossRef]
13. Wang, Q.; Qi, F.; Sun, M.; Qu, J.; Xue, J. Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques. *Comput. Intell. Neurosci.* **2019**, *2019*, 9142753. [CrossRef] [PubMed]
14. Wang, X.; Liu, J.; Liu, G. Diseases Detection of Occlusion and Overlapping Tomato Leaves Based on Deep Learning. *Front. Plant Sci.* **2021**, *12*, 792244. [CrossRef]
15. Fuentes, A.; Yoon, S.; Kim, S.; Park, D. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors* **2017**, *17*, 2022. [CrossRef] [PubMed]
16. Zaborowicz, M.; Boniecki, P.; Koszela, K.; Przybylak, A.; Przybył, J. Application of neural image analysis in evaluating the quality of greenhouse tomatoes. *Sci. Hortic.* **2017**, *218*, 222–229. [CrossRef]
17. Zaborowicz, M.; Boniecki, P.; Koszela, K.; Przybył, J.; Mazur, R.; Kujawa, S.; Pilarski, K. Use of artificial neural networks in the identification and classification of tomatoes. In Proceedings of the SPIE Proceedings, Beijing, China, 21–22 April 2013. [CrossRef]
18. Brahimi, M.; Boukhalfa, K.; Moussaoui, A. Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. *Appl. Artif. Intell.* **2017**, *31*, 299–315. [CrossRef]
19. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]
20. Liang, Q.; Xiang, S.; Hu, Y.; Coppola, G.; Zhang, D.; Sun, W. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* **2019**, *157*, 518–529. [CrossRef]
21. Rangarajan, A.K.; Purushothaman, R.; Ramesh, A. Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Comput. Sci.* **2018**, *133*, 1040–1047. [CrossRef]
22. Trivedi, N.K.; Gautam, V.; Anand, A.; Aljahdali, H.M.; Villar, S.G.; Anand, D.; Goyal, N.; Kadry, S. Early Detection and Classification of Tomato Leaf Disease Using High-Performance Deep Neural Network. *Sensors* **2021**, *21*, 7987. [CrossRef]
23. Loyani, L.K.; Bradshaw, K.; Machuve, D. Segmentation of *Tuta absoluta*'s Damage on Tomato Plants: A Computer Vision Approach. *Appl. Artif. Intell.* **2021**, *35*, 1107–1127. [CrossRef]

24. Georgantopoulos, P.; Papadimitriou, D.; Constantinopoulos, C.; Manios, T.; Daliakopoulos, I.; Kosmopoulos, D. A Multispectral Dataset for the Detection of *Tuta absoluta* and Leveillula Taurica in Tomato Plants. *Smart Agric. Technol.* **2023**, *4*, 100146. [CrossRef]
25. Rubanga, D.P.; Loyani, L.K.; Richard, M.; Shimada, S. A Deep Learning Approach for Determining Effects of *Tuta absoluta* in Tomato Plants. *arXiv* **2020**, arXiv:2004.04023.
26. Mkonyi, L.; Rubanga, D.; Richard, M.; Zekeya, N.; Sawahiko, S.; Maiseli, B.; Machuve, D. Early identification of *Tuta absoluta* in tomato plants using deep learning. *Sci. Afr.* **2020**, *10*, e00590. [CrossRef]
27. Tsiakas, K.; Papadimitriou, A.; Pechlivani, E.M.; Giakoumis, D.; Frangakis, N.; Gasteratos, A.; Tzovaras, D. An Autonomous Navigation Framework for Holonomic Mobile Robots in Confined Agricultural Environments. *Robotics* **2023**, *12*, 146. [CrossRef]
28. Pechlivani, E.M.; Gkogkos, G.; Giakoumoglou, N.; Hadjigeorgiou, I.; Tzovaras, D. Towards Sustainable Farming: A Robust Decision Support System's Architecture for Agriculture 4.0. In Proceedings of the 2023 24th International Conference on Digital Signal Processing (DSP), Rhodes, Greece, 11–13 June 2023. [CrossRef]
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.
30. Lin, T.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
32. Lin, T.; Dollár, P.; Girshick, R.B.; He, K.; Hariharan, B.; Belongie, S.J. Feature Pyramid Networks for Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
33. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection With One Line of Code. *arXiv* **2017**, arXiv:cs.CV/1704.04503.
34. Zhou, H.; Li, Z.; Ning, C.; Tang, J. CAD: Scale Invariant Framework for Real-Time Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 760–768. [CrossRef]
35. Solovyev, R.; Wang, W.; Gabruseva, T. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image Vis. Comput.* **2021**, *107*, 104117. [CrossRef]
36. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ. Comput. Inf. Sci.* **2023**, *35*, 757–774. [CrossRef]
37. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2014**, arXiv:1405.0312.
38. Pizer, S.; Johnston, R.; Ericksen, J.; Yankaskas, B.; Muller, K. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In Proceedings of the 1990 First Conference on Visualization in Biomedical Computing, Atlanta, GA, USA, 22–25 May 1990. [CrossRef]
39. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 1 May 2023).
40. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
41. Gong, Z.; Zhong, P.; Hu, W. Diversity in Machine Learning. *IEEE Access* **2019**, *7*, 64323–64350. [CrossRef]
42. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]
43. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs* **2018**, *8*, e1249. [CrossRef]
44. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 1 May 2023).