

## Article

# TobSet: A New Tobacco Crop and Weeds Image Dataset and Its Utilization for Vision-Based Spraying by Agricultural Robots

Muhammad Shahab Alam <sup>1,\*</sup>, Mansoor Alam <sup>2</sup>, Muhammad Tufail <sup>2,3,\*</sup>, Muhammad Umer Khan <sup>4</sup>, Ahmet Güneş <sup>1</sup>, Bashir Salah <sup>5,\*</sup>, Fazal E. Nasir <sup>2</sup>, Waqas Saleem <sup>6</sup> and Muhammad Tahir Khan <sup>2,3</sup>

<sup>1</sup> Defense Technologies Institute, Gebze Technical University, Gebze 41400, Turkey; ahmet.gunes@gtu.edu.tr

<sup>2</sup> Advanced Robotics and Automation Laboratory, National Center of Robotics and Automation (NCRA), Peshawar 25000, Pakistan; mansooralam129047@gmail.com (M.A.); fazalnasir.uet@gmail.com (F.E.N.); tahir@uetpeshawar.edu.pk (M.T.K.)

<sup>3</sup> Department of Mechatronics Engineering, University of Engineering & Technology, Peshawar 25000, Pakistan

<sup>4</sup> Department of Mechatronics Engineering, Atilim University, Ankara 06830, Turkey; umer.khan@atilim.edu.tr

<sup>5</sup> Department of Industrial Engineering, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia

<sup>6</sup> Department of Mechanical and Manufacturing Engineering, Institute of Technology, F91 YW50 Sligo, Ireland; saleem.waqas@itsligo.ie

\* Correspondence: shahab@gtu.edu.tr (M.S.A.); tufail@uetpeshawar.edu.pk (M.T.); bsalah@ksu.edu.sa (B.S.)



**Citation:** Alam, M.S.; Alam, M.; Tufail, M.; Khan, M.U.; Güneş, A.; Salah, B.; Nasir, F.E.; Saleem, W.; Khan, M.T. TobSet: A New Tobacco Crop and Weeds Image Dataset and Its Utilization for Vision-Based Spraying by Agricultural Robots. *Appl. Sci.* **2022**, *12*, 1308. <https://doi.org/10.3390/app12031308>

Academic Editors: Anselme Muzirafuti and Dimitrios S. Paraforos

Received: 12 October 2021

Accepted: 22 December 2021

Published: 26 January 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/>).

**Abstract:** Selective agrochemical spraying is a highly intricate task in precision agriculture. It requires spraying equipment to distinguish between crop (plants) and weeds and perform spray operations in real-time accordingly. The study presented in this paper entails the development of two convolutional neural networks (CNNs)-based vision frameworks, i.e., Faster R-CNN and YOLOv5, for the detection and classification of tobacco crops/weeds in real time. An essential requirement for CNN is to pre-train it well on a large dataset to distinguish crops from weeds, lately the same trained network can be utilized in real fields. We present an open access image dataset (TobSet) of tobacco plants and weeds acquired from local fields at different growth stages and varying lighting conditions. The TobSet comprises 7000 images of tobacco plants and 1000 images of weeds and bare soil, taken manually with digital cameras periodically over two months. Both vision frameworks are trained and then tested using this dataset. The Faster R-CNN-based vision framework manifested supremacy over the YOLOv5-based vision framework in terms of accuracy and robustness, whereas the YOLOv5-based vision framework demonstrated faster inference. Experimental evaluation of the system is performed in tobacco fields via a four-wheeled mobile robot sprayer controlled using a computer equipped with NVIDIA GTX 1650 GPU. The results demonstrate that Faster R-CNN and YOLOv5-based vision systems can analyze plants at 10 and 16 frames per second (fps) with a classification accuracy of 98% and 94%, respectively. Moreover, the precise smart application of pesticides with the proposed system offered a 52% reduction in pesticide usage by spotting the targets only, i.e., tobacco plants.

**Keywords:** precision agriculture; selective spraying; vision-based crop and weed detection; convolutional neural networks; Faster R-CNN; YOLOv5

## 1. Introduction

Tobacco is grown in more than 120 countries around the world, covering millions of hectares of land. In Pakistan, it is regarded as an important crop as it generates substantial revenue. According to an estimate, in rural areas of the country, 80k–90k tonnes of Flue-Cured Virginia (*Nicotiana Tabacum*) is produced annually [1]. In addition to being a profitable crop, it is important to highlight that tobacco's leaf is highly susceptible to pests and pathogens, and the crops demand meticulous effort and care in order to protect them from seasonal insects, as shown in Figure 1. Local farmers rely upon the use of conventional agrochemical spray methods for combating these pests and pathogens. Pesticides are applied to tobacco plants usually five to six times in one season (over three months),

which makes it a highly pesticide-dependent crop. Two methods are commonly used for pesticide spraying: manual knapsack spraying in which human labor carries the equipment and performs spray on every plant and broadcast spraying via a tractor-mounted sprayer in which the entire field is sprayed indiscriminately. Both methods are imprecise and, therefore, cause serious damage to farmers' health (due to first-hand/direct exposure) and to the environment (due to overdosing) [2–6]. Despite its hazards, agrochemical spraying is still common in practice as it is a viable and economical means to protect tobacco crop from pests and pathogens [7]. The solution, therefore, lies not in eliminating the use of agrochemicals but instead in optimizing their application by embracing advanced techniques and methodologies.



**Figure 1.** Weeds and tobacco leaf infestation due to pests.

Artificial Intelligence is rapidly bringing a substantial paradigm shift in the agriculture sector. Endowing agricultural spraying systems the cognitive ability of understanding, learning, and responding to different crop conditions greatly improves spraying operations. Precision spraying methods combine techniques from emerging disciplines such as artificial intelligence, robotics, and computer vision, which provides a spraying system the ability to identify plants (crop) and weeds and apply precise doses only on the desired targets [8–13].

Over the last decade, numerous promising attempts have been made by researchers for the development of intelligent spraying systems for different crops [14–23]. Surprisingly, not much work is found in the literature on vision-based site-specific spraying systems for crops. The vision-based system tends to deal with numerous variations, such as varying leaf sizes at different growth stages, varying light intensities, different soil textures, varying leaf colors due to different water levels, high weed densities, and crop plant occlusion by weeds, etc.

Existing methods for vision-based plant/weed detection and precision spraying are mostly based on traditional machine learning-based techniques [24–32]. Although high accuracies have been achieved with these techniques, the hand-crafted features formulation and generation of a decision function over the extracted features make them less robust. Therefore, they are certainly not a preferred choice for tobacco plant and weed detection (keeping in view the factors of variations and complexities involved in tobacco fields) due to poor generalization capabilities. Over the past few years, deep learning-based computer vision algorithms have demonstrated their ability to perform well on complex problems from training examples [33–41]. CNNs are the main architecture of these computer vision algorithms. Deep learning algorithms learn the features and decision functions in an end-to-end fashion. Lopez-Martin et al. [42] proposed a classifier known as gaNet-C for type-of-traffic forecast problem. An additive network model, gaNet, has the capability to forecast k-steps beforehand by utilizing time-series of last computed values for each node. The proposed model demonstrates good performance on two detection forecast problems.

The advantages that deep learning algorithms offer, such as feature learning capabilities, high accuracy, and better performance in intricate problems, make them best suited for complex tasks such as detecting tobacco plants under several variations in outdoor fields. Several studies have been reported with respect to deep learning-based plant and weed detection [43–49]. The latest research on plant and weed detection mainly utilizes computer vision [50–56]. For instance, Costa et al. [57] used deep learning for finding defects in

tomatoes by applying Deep ResNet classifiers. According to their finding, ResNet50 with fine-tuned layers was reported as the best model that achieved an average precision of 94.6% and a recall of 86.6%. Moreover, it was observed that fine-tuning outperformed feature extraction process. Santos Ferreira et al. [58] detected weeds in soybean crops using ConVNets and SVM classifiers. ConVNets was able to achieve higher accuracy of more than 97% in weed detection. Yu et al. [59] used deep learning algorithms for detecting multiple weed species in Bermuda grass. The study reported that VGGNet performed well with an F1-score of over 0.95 than compared to GoogleNet. Moreover, F1-scores of over 0.99 were reported for detecting weeds via DetectNet. The authors, based on attained results, concluded the effectiveness of deep convolutional neural networks in the weed detection problem. In another study, Sharpe et al. [60] evaluated three CNNs—DetectNet, VGGNet, and GoogLeNet—for the detection of weeds in strawberry fields. It was observed that the image classification DetectNet model produced the best results for image-based remote sensing of weeds. Le et al. [61] used Faster R-CNN for the detection of weeds in Barley crops using several feature extractors. In the study, mean Average Precision (mAP) with Inception-ResNet-V2 was found better than the mAP for other networks. Moreover, an inference time of 0.38 s per image was also reported. Quan et al. in [62] presented an improved version of the Faster R-CNN vision system for the identification of maize seedlings in tough field environments. The images were taken with a camera at an angle ranging from 0 to 90 degrees. The results reported detection accuracy of 97.71%. In the study performed by [63], the authors reported F1-scores of 88%, 94%, and 94% for SVM, YOLOv3, and Mask R-CNN for detecting weeds in lettuce crops, respectively. The work reported by Wu et al. [64] used YOLOv4-based vision system for detecting apple flowers. The model based on CSPDarkNet-53 framework was simplified with a channel pruning algorithm for detecting the target object in real time. They reported achieving a mAP of 97.31% at a detection speed of 72.33 fps.

Despite the impressive accomplishments in deep learning-based object detection, the performance of these algorithms has yet to be evaluated in the realm of tobacco plants and weeds detection; for instance, the use of region-based methods such as Faster R-CNN or one stage detectors such as YOLOv5. Moreover, published reports also lack experimental validation in actual field environments. The aim of this study includes the replacement of conventional broadcast spraying methods in tobacco fields with a site-specific (drop-on-demand) spraying system. The proposed method detects and classifies tobacco plants and weed automatically, determines their position, i.e., their location in the crop rows, and finally performs agrochemical spray on the detected targets.

This paper focuses on automatic vision-based tobacco plant detection that is considered a vital part of the precision spraying system. The basic frameworks of two off-the-shelf deep-learning algorithms—Faster R-CNN and YOLOv5—are employed for detection and classification models. The robustness and ability of the models are enhanced by fine-tuning detection of tobacco plants in challenging field conditions. Both detection models are tested on a vision-guided mobile robot platform in real tobacco fields. A comparative study is also carried out between both frameworks in terms of robustness, accuracy, and inference/computational speed. The Faster R-CNN-based vision-based model demonstrated higher accuracy but lower real-time detection speed, whereas the YOLOv5-based model produced slightly lower accuracy but higher real-time detection speed. Therefore, YOLOv5-based vision model, based on its performance, is considered best suited for real-time tobacco plant and weed detection. The main contributions of this study are summarized as follows:

1. Development and deployment of a vision-based robotic spraying system for replacement of the conventional broadcast spraying methods with a site-specific selective spraying technique that can detect tobacco plants and weed and classify them in a real time;
2. Building a tobacco image dataset (TobSet) that comprises labeled images of tobacco crop and weed. The dataset is collected under challenging real in-field conditions to

train and evaluate the latest state-of-the-art deep learning algorithms. TobSet is an open-source dataset and is publicly available at <https://github.com/mshahabalam/TobSet> (accessed on 11 October 2021).

The rest of the paper is organized as follows: Section 2 covers the description about the image dataset and Section 3 briefly explains the materials and methods employed in this study. The workings of Faster R-CNN and YOLOv5 algorithms are discussed in Section 4. The hardware setup for the implementation is explained in Section 5. Evaluation of the proposed approaches is carried out in Section 6 along with discussion and comparative analysis, and a brief concluding remarks are provided in Section 7.

## 2. Data Description

Due to the unavailability of any image dataset of tobacco plants, we developed an extensive image dataset, TobSet, from the actual fields in Swabi, Khyber Pakhtunkhwa, Pakistan ( $34^{\circ}09'07.3''$  N  $72^{\circ}21'36.2''$  E). The main objective of building this dataset is to provide real-field data for training and evaluating the performance of state-of-the-art algorithms for tobacco crop and weed detection. TobSet comprises (a) 7000 images of tobacco plants and (b) 1000 images of bare soil and weeds (that grow up in tobacco fields), with a resolution of  $640 \times 480$ . The images are captured using a 13-megapixels color digital camera possessing a CMOS-image sensor (IMX258 Exmor RS by Sony, Japan), 28 mm focal length,  $65.4^{\circ}$  horizontal FOV, and  $51.4^{\circ}$  vertical FOV. A comprehensive dataset was built over a period of 2 months, i.e., starting from the first week of tobacco seedling transplantation from seedbeds to the time when plants gain an approximate height of 1.25 m. All images in the dataset were captured manually by human scouts in the months of June and July 2020. No artificial shading and sources of lightning were used while collecting the images. During image acquisition, the camera's height was adjusted between 1 and 1.5 m. In order to maintain diversity in the dataset, all images in TobSet are captured under several factors of variations: different growth stages, different day timings, varying lighting and weather conditions (i.e., on normal, bright sunny, and cloudy days), and visual occlusions of crop leaves by weeds, etc. The existing literature on vision-based detection of crops and weeds lacks experimental validation on hard real-world datasets such as TobSet. Some sample images from the publicly available TobSet are presented in Figure 2. After data acquisition, the main step involved in crop/weed detection is the annotation of images for ground truth data. All images in the TobSet are manually labeled with the LabelImg tool.



**Figure 2.** Illustration of factors of variation in the actual tobacco fields.

TobSet is publicly available and offers multi-faceted utilities:

1. It comprises labelled images of tobacco plants and weeds that can be utilized by computer scientists for performance evaluation of their developed computer vision algorithms;
2. Scientists working on agricultural robotics can use it to train their robots for variable rate-spray applications, plant or weed detection, and detection of plant diseases;
3. It can also be used by agriculturists and researchers for studying various aspects of tobacco plant growth, weed management, yield enhancement, leaf diseases, and pest prevention, etc.

### 3. Materials and Methods

For targeted agrochemical spray, the application equipment must have the following capabilities: (a) discriminating the crop plants from weeds, (b) determining the robot's location in the field, and (c) applying agrochemicals on the targeted plants, i.e., crop or weeds. Considering these aspects, our developed agrochemical spraying robot has three main systems: a vision-based crop or weed identification system, a robot navigation system, and an actuation system for spraying on targeted plants. This paper is focused only on the predominant sensing modalities of developed spraying robot that enables it to identify crop plants and weeds, i.e., a vision-based detection framework.

Due to the nature of the application, i.e., harsh or challenging tobacco field conditions, the vision system essentially must be robust in order to process data and generate accurate results in real-time. Due to excellent performance, deep-learning algorithms are currently state-of-the-art for computer vision applications. This is attributed to the availability of large-sized labeled data, and deeply layered architectures. However, due to increasing depth, the algorithms are computationally very expensive, especially for resource-limited portable machines. The study presented herein aims to develop a deep-learning-based vision framework with low inference cost, thereby it can be used in real-time detection and classification of tobacco crops and weeds. In order to achieve this, two state-of-the-art CNN algorithms, i.e., Faster R-CNN and YOLOv5, are utilized for implementation.

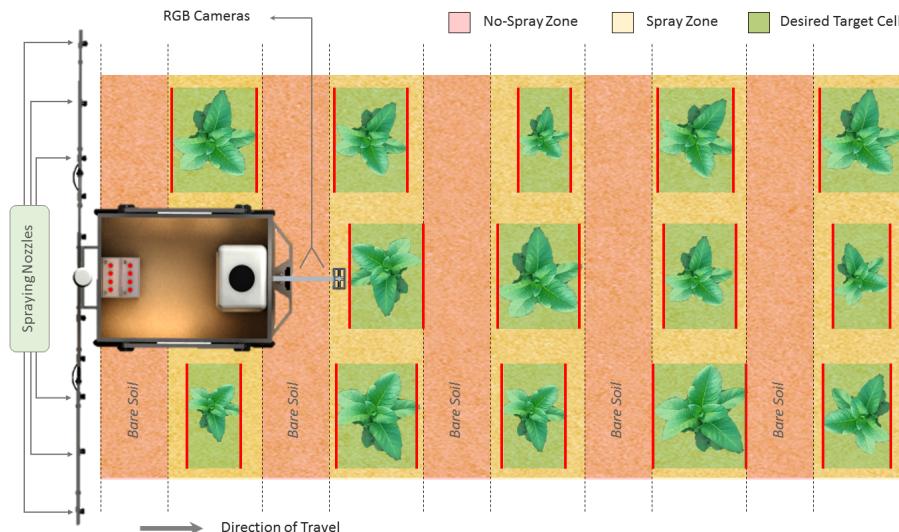
Pesticide application on the tobacco plants begins immediately after the first week when the seedlings are transplanted from the seedbed into the fields and continues periodically until their maturity. As shown in Figure 3, inter-row spacings of approximately 1 m and intra-row spacings of approximately 0.75 m were kept between any two consecutive plants. Therefore, indiscriminate broadcast application of pesticides on the complete tobacco field, particularly at earlier growth stages when the plants' canopy sizes are very small, results in off-the-target pesticide spray on bare soil spots. This unnecessary pesticide application on bare soil or weed patches engenders polluting the environment and leaching of toxic pesticides into the ground.



**Figure 3.** Inter-row and intra-row plant spacing in tobacco fields.

Moreover, all crop plants across a tobacco field do not necessarily grow homogeneously due to variation in seedling health, size of the plant at the time of transplantation, and water

and nutrients variability across the field. Due to these reasons, intra-row and inter-row spacing varies across the entire field according to plant leaf sizes. Our system proposes dividing the camera's field of view into grids. In each grid, the deep-learning-based detector detects plants and assigns a cell to each plant based on its coverage such that it apprehends plant canopies. Since our spray application module comprises flat fan nozzles, the lateral length of the grid is set according to the swath size of each corresponding nozzle. Furthermore, the vertical size of the cell is adjusted based on the detected plant's canopy, as shown in Figure 4, by the green boxes.



**Figure 4.** Desired pinpointed spray zones.

Two separate vision systems are employed on the robot. One vision system is for the detection and localization of the tobacco crops and weeds, whereas the other vision system helps with crop row structure detection for guiding the robot along the crop rows. As stated earlier, this paper focuses only on the vision system for crop and weed detection. The tobacco crop or weed detection and spraying processes are performed in the following sequence: (a) acquiring an image with the camera via image grabber; (b) sending the acquired image to the NVIDIA GPU for processing; (c) detection of crop plants and weeds; (d) determining the location of the plant and size of its attributed grid cell based on the plant's coverage; (e) sending the required control signal for spray via USB port to the embedded controller; and (f) actuation of the corresponding nozzles upon reaching the target plant.

#### 4. CNN-Based Detection and Classification Frameworks

The primary objective of this research is to enable an agriculture sprayer robot to identify tobacco plants and weeds in real time using an onboard vision system. Two different deep-learning algorithms are utilized in the detection of tobacco plants and weeds, i.e., Faster R-CNN and YOLOv5. Despite some differences in the overall frameworks of Faster R-CNN and YOLO, both rely upon CNNs as their core working tool. Faster R-CNN processes the entire image using CNN and then divides it for several region proposals in two steps, whereas YOLO splits the image into grid cells and processes it through CNN in one step.

##### 4.1. Faster R-CNN

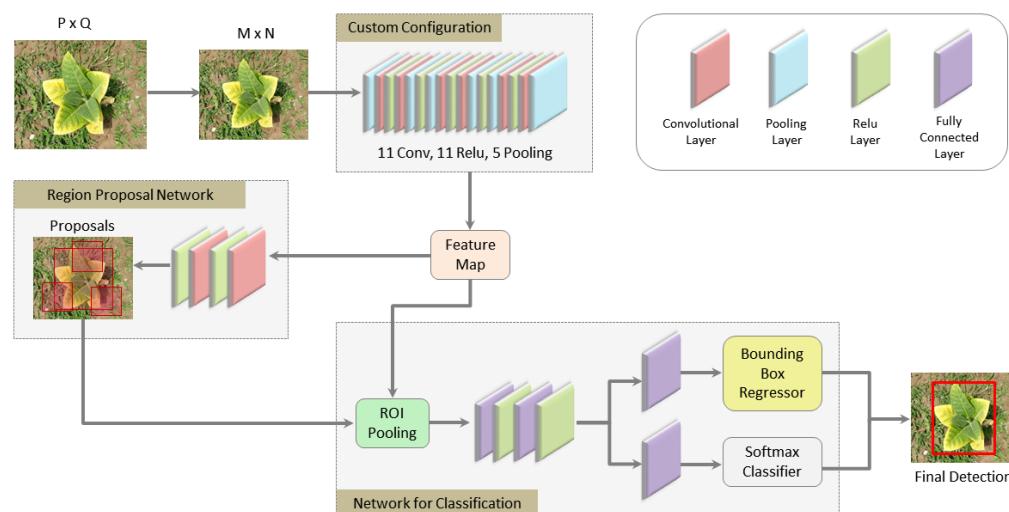
Faster R-CNN, proposed by Ren et al. [65], is a combination of Fast R-CNN and region proposal network (RPN). The aim behind the introduction of Faster R-CNN was to make the detection process less time consuming and more accurate. Primarily, its structure comprises feature extraction, region proposals, and bounding box regression.

The submodules involved in the algorithm for our tobacco crop and weed detection are explained in the following subsections.

#### 4.1.1. Convolutional Layers

Being a CNN-based detection approach, we use the basic *convolutional*, *relu*, and *pooling* layers for extracting feature maps from tobacco and weeds images. Rather than using the models of Simonyan and Zisserman [66] or Zeiler and Fergus [67], we customized the architecture of the model. The in-depth structure of our model comprised eleven *Conv* layers, eleven *relu* layers, and five *pooling* layers. In each *Conv* layer, the *kernel* size is set to 3 and *padding* and *stride* are set to 1, whereas in the *pooling* layers, the *kernel* size is set to 2, *padding* is set to 0, and *stride* is set to 2. The detection and classification pipeline of the Faster R-CNN-based detection model is shown in Figure 5.

All the convolutions are expanded in the *Conv* layers using *padding* size of 1 to transform the original input image size to  $(M + 2) \times (N + 2)$ , and then a *kernel* of size  $(3 \times 3)$  is applied to obtain an output image of  $(M \times N)$ , i.e.,  $(640 \times 480)$ . This helped the input and output matrix sizes to remain unchanged in the *Conv* layers. Moreover, the *pooling* layer, *kernel*, and the *stride* sizes are set to 2 in the *Conv* layers. Thus, every  $(640 \times 480)$  matrix that goes past the *pooling* layer is converted to  $(640/2) \times (480/2)$ . In all of the *Conv* layers, the input and output sizes of the *Conv* and *relu* layers are kept the same. However, the *pooling* layer forces the output length and width to be 1/2 of the input. Next, a matrix with a size of  $(640 \times 480)$  is switched to  $(640/16) \times (480/16)$  by the *Conv* layers; hence, the feature map produced by *Conv* layers can be associated with the original image. The feature maps are fed to the subsequent RPN and fully connected layers.



**Figure 5.** Faster R-CNN-based tobacco crop detection framework.

#### 4.1.2. Region Proposal Networks

The RPN network being small is slid over the feature map for generating regional proposals. RPN classifies the corresponding regions and regresses bounding box locations, simultaneously. To find out that whether the anchors belonged to the foreground or background, we used *softmax* in this layer. Furthermore, the anchors are adjusted with the bounding box regression in order to obtain precise proposals. The classic approach generates a very time-consuming detection framework. Therefore, instead of the traditional sliding window and selective search approaches, the RPN method is used directly for generating detection frames. This served as a plus point of the Faster R-CNN method as compared to classical detection methods in improving the detection frame generation speed to some extent [65].

#### 4.1.3. ROI Pooling

In the ROI *pooling* layer, the region proposals are collected and split into smaller windows. Next, feature maps are extracted from these regions, which are further sent to the subsequent *fullyconnected* layer for determining the target class in this layer. Moreover, our ROI *pooling* layer comprises two inputs:

1. Original feature maps;
2. RPN output proposal boxes of different sizes.

In traditional CNNs such as AlexNet, VGG, etc., the size of the input image essentially should be constant, and the output of the network should also be a fixed-size vector or matrix when the network is trained. Therefore, a remedy is proposed for variable input image sizes: (a) parts of images are cropped, and (b) the images are warped to the desired size. Despite adopting these approaches, either the structure of the entire image is altered after the images are cropped or the shape information of the original image is altered when the images are warped. Similarly to the proposal's generation approach of RPN's bounding box regression on foreground anchors, the image properties achieved in this manner has dissimilar shapes and sizes. To cater with this complexity, ROI pooling is utilized. Since it corresponds to the  $(640 \times 480)$  scale, the spatial scale parameter is first used for mapping it back to  $(640/16) \times (480/16)$ -sized feature maps. Next, horizontal ( $pooled_w$ ) and vertical ( $pooled_h$ ) division of each property is performed. Finally, *maxpooling* is applied to each property. This approach ensured an output of the same size and fixed length.

#### 4.1.4. Classification

Pseudo feature maps are used to compute the proposal's class and, simultaneously, the final position of the detection frame is acquired by the bounding boxes. Since the network deals with  $P \times Q$  input size images, they are first scaled down to a constant size of  $(M \times N)$ , i.e.,  $(640 \times 480)$ , and passed onto the network. The convolution layers contains 11 *Conv* layers, 11 *relu* layers, and 5 *pooling* layers. The RPN network employs  $3 \times 3$  convolution and then generates foreground or background anchors and the associated bounding box regression offsets. Then, proposals are calculated and ROI pooling is performed, which computes the feature maps and sends them to the subsequent fully connected *softmax* network for classification. The classification section uses the acquired property feature maps for calculating the specific category (i.e., tobacco plants and weeds) that each property belongs to via the *fullyconnected* layer and *softmax*.

Finally the probability for the class is computed, and bounding box regression is once more used for obtaining the position offset for each proposal. The classification section of the proposed network is highlighted by the shaded region of Figure 5. After obtaining the  $7 \times 7 = 49$  sized features, feature maps from ROI pooling, and then sending them to the succeeding network, the following two steps were performed:

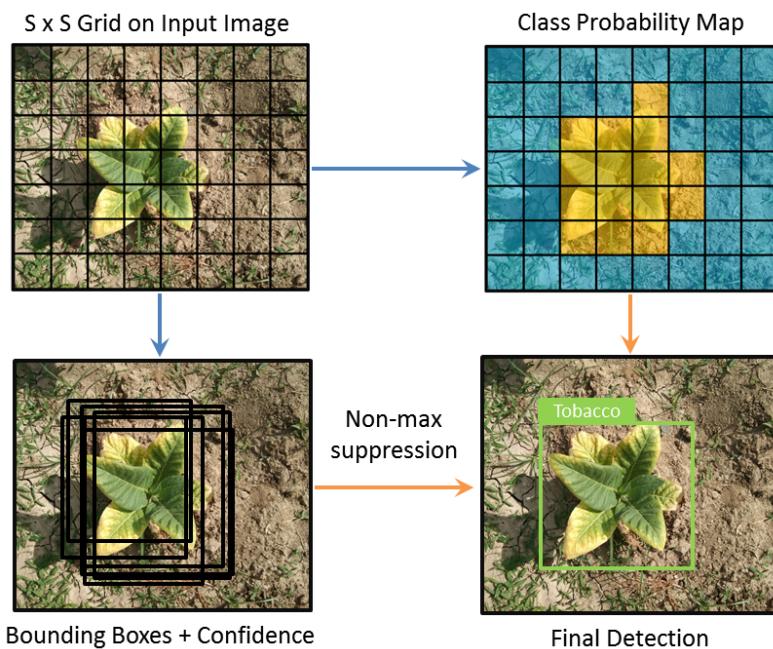
1. Classification of proposals by *fullyconnected* layer and *softmax*;
2. Bounding box regression on the proposals for acquiring more accurate rectangular boxes.

#### 4.2. You Only Look Once (YOLO)

YOLO is a fast one-stage object detection model that was developed by Redmon et al. [68] in 2015. YOLO as compared to Faster R-CNN is less error-prone to background errors in images as it observes the larger context. The main trait that dignifies YOLO from other similar networks is its capability to detect objects (with bounding boxes) and calculate class probabilities in a single step, i.e., detection and class predictions are performed simultaneously after a single evaluation of the input image. Training is performed on complete images, and the performance of the detection is optimized directly. YOLO, unlike the region proposal and sliding window-based methods, processes the complete image during training and testing phases, which enables it to translate class-specific information and its outlook implicitly.

There are three main elements involved in the YOLO network: (a) backbone, (b) neck, and (c) head. The backbone comprises CNNs that serve the purpose of aggregations and image feature formation from several image granularities. The neck is composed of a series of layers used for mixing and combining the extracted features and subsequently transmitting them to the prediction layer. Finally, the head is used for the features prediction, bounding box creation, and class prediction.

The algorithm works by first splitting the input image into a grid of  $S \times S$  and then predicting  $B$  bounding boxes for each grid cell, as shown in Figure 6. Every bounding box in the grid cell is assigned a confidence score to denote the probability of an object's existence inside the defined box.



**Figure 6.** YOLO detection pipeline.

Grid cells are accountable for detecting objects if their centers fall inside a grid cell. If the center of the bounding box (of the same object) is predicted to fall in multiple grid cells, a non-max suppression eliminates redundant bounding boxes and retains the one possessing the highest probability. Each bounding box has four associated predictions that include the  $(x, y)$  coordinates of the center of the box, width  $w$ , height  $h$ , and confidence  $C$ . Confidence  $C$  can be formulated as follows:

$$C = Pr(Class_i) * IOU_{pred}^{truth} \quad (1)$$

where  $IOU$  is the intersection over union, i.e., the overlapped area between predicted and ground truth bounding boxes. The  $IOU$  value of 1 represents a perfect prediction of the bounding box relative to ground truth.

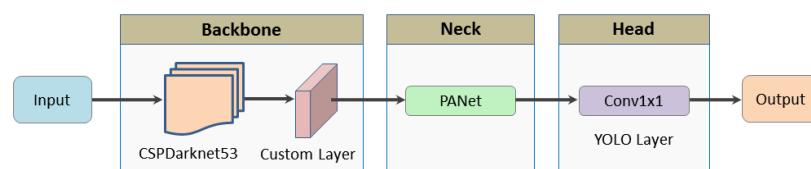
Bounding boxes and conditional class probabilities for each grid cell are computed at the same time. Conditional class probabilities and bounding box confidence predictions during the test phase are multiplied to obtain confidence scores of a particular class of each box as follows.

$$Pr(Class_i|Object) * Pr(Object) * IOU_{pred}^{truth} = Pr(Class_i) * IOU_{pred}^{truth} \quad (2)$$

#### Network Architecture

The baseline architecture of YOLOv5 is very similar to YOLOv4, primarily comprising a Backbone, Neck, and Head. The backbone of YOLOv5 can be ResNet-50, VGG16, ResNeXt-

101, EfficientNet-B3, or CSPDarkNet-53. We used the CSPDarkNet-53 neural network as our model's backbone, which encompasses cross-stage partial connections, and it is considered as the most optimal model [57]. CSPDarknet-53 has 53 convolutional layers, and it originates from DenseNet architecture. DenseNet network uses the preceding input, and, prior to stepping into dense layers, it concatenates the previous input with the current one [69]. The robustness of our YOLOv5-based vision framework greatly improved with the CSP application approach, i.e., by applying the CSP1\_x to the backbone and CSP2\_x to the neck. First, data were fed as input to CSPDarkNet-53 for extracting features. For improving feature extraction from different growth stages of tobacco plants, an additional layer was inserted into the model's backbone, which helped to improve the mAP. Next, the extracted features were fed to PANet (Path Aggregation Network) for fusing features. Finally, output results, i.e., class, score, etc., of detection were provided by the YOLO layer. Our model's head part used an anchor-free one-stage object detector YOLO. The modified YOLOv5 architecture used in this study is illustrated in Figure 7.



**Figure 7.** Modified YOLOv5 detection pipeline.

## 5. Experimental Evaluation

This section deals with the experimental setup that we used for conducting in-field experiments, the dataset used for training both deep learning-based vision models, and the infield real-time results obtained with our vision models.

### Hardware Setup

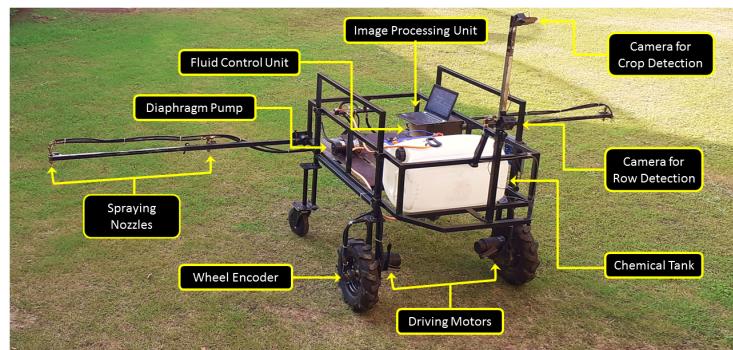
The proposed frameworks are implemented in the tobacco fields with a four-wheeled mobile robot platform. The robot has a track width and wheelbase of 1 and 1.3 m, respectively. In order to protect tobacco plants from the robot, the ground clearance of the platform was carefully chosen as 0.9 m. Moreover, the height of the robot's platform can be adjusted anywhere between 0.9 and 0.4 m depending on different crops.

In order to keep robot design and control simple, a differential drive scheme was chosen with two driving wheels (front) and two passive wheels (rear). The robot is equipped with two DC motors connected to motor controllers for steering and driving the robot along the straight crop rows. Two separate RGB cameras are mounted on the robot: One is used for the crop row detection (for navigation), and the other is used for crop/weed detection (for spraying). The camera for row structure detection is mounted at the front with its face towards the ground and a horizon at an angle of 35° with the horizontal axis, covering three rows simultaneously. The camera for crop and weed detection is mounted at the front of the robot and oriented facing downwards to the ground at a fixed distance of 1.8 m from nozzles on the boom. The distance between the crop and weed detection camera and the boom is kept at the maximum in order to provide the desired time delay between detection and position estimation of the crop plant and the spray application process on every corresponding grid cell.

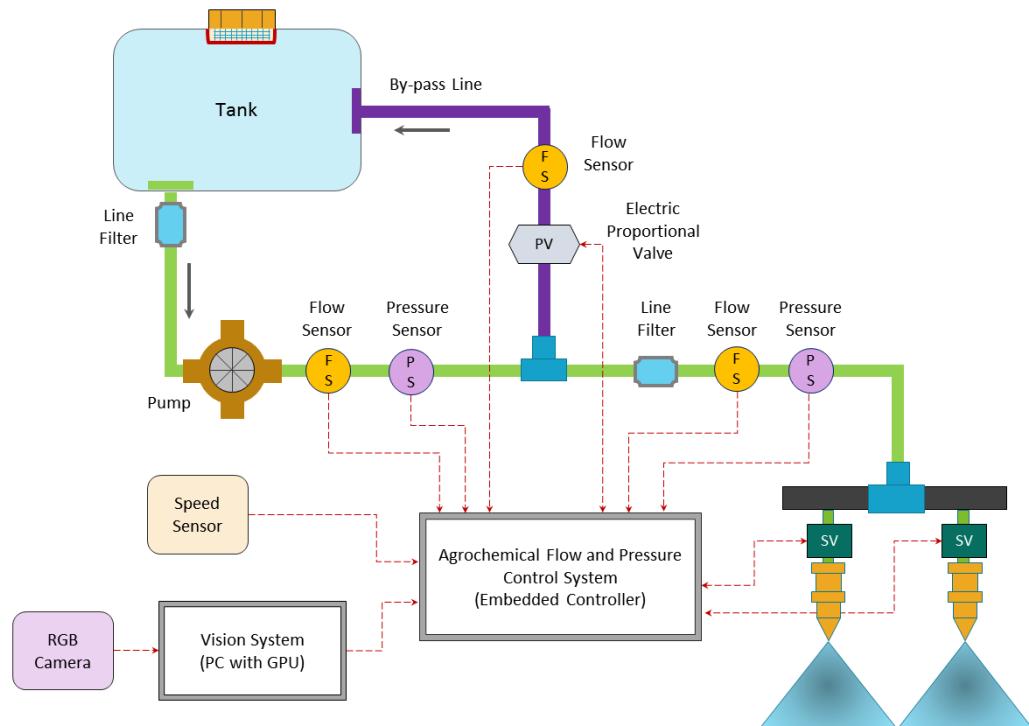
The vision-based detection system is coupled with spraying equipment and other sensing modules, thereby making a complete precision agricultural robotic spraying system. A 12 V DC diaphragm pump is used to pressurize the fluid system. An electronic pressure regulated valve maintains a constant line pressure when different nozzles on the boom are switched ON and OFF based on feedback from the vision system and other sensing modules. The outflow line from the pump is divided into a bypass line that diverts excess flow back to the tank and a boom line onto which the nozzles are mounted. Two rotary incremental encoders (with resolutions of 1000 pulses per revolution) connected to the

embedded controller are mounted on the front wheels' axles to measure the rotation (and thereby speed) of the wheels. The incremental encoders and a GPS module facilitates the robot in determining its position and heading direction for navigation. Moreover, the optical data acquired via cameras are synchronized to the robot's position through incremental encoders and GPS module.

The robot used ROS (Robot Operating System) as the middleware software framework. The cameras were connected to a computer possessing an Intel Core E5-1620, a 3.50 GHz processor, 32 GB RAM, and an 8 GB NVIDIA GTX 1650Ti GPU for processing the images. Moreover, Microsoft Visual Studio and Python were used for program development. The developed agricultural robot sprayer and its overall functional block diagram is shown in Figures 8 and 9, respectively.



**Figure 8.** Developed prototype of the agricultural robotic sprayer.



**Figure 9.** Block diagram illustrating the developed vision and fluid flow control systems.

## 6. Results and Discussions

In order to validate and demonstrate the effectiveness of both vision-based frameworks for tobacco crop/weed detection and classification, the models are trained and tested on real-field tobacco images from TobSet. The dataset consists of 8000 images; 7000 images are of tobacco plants, and the rest of the images are of weeds. Images from both classes are divided with a 70 to 30 ratio into training and testing sets. The training set comprised

a total of 5600 images (4900 tobacco and 700 weeds), whereas the testing set comprised 2400 images (2100 tobacco and 300 weeds).

In the implementation phase, the models are trained using down-sampled images (with a resolution of  $640 \times 480$ ). A learning rate is initialized as 0.0002 for the training. Google's TensorFlow API is utilized for implementation purposes. Batch sizes of 1 and 10 k epochs are used for training the models. Table 1 lists the hyper-parameters and their corresponding losses (against the epochs) for both models. It can be observed from Table 1 that for obtaining better results with Faster R-CNN-based vision model, the learning rate is kept the same, whereas the other hyper-parameters did change. With an increase in the number of epochs, total loss is reduced. The confusion matrices for Faster R-CNN and YOLOv5-based models, given in Tables 2 and 3 respectively, are used for computing the evaluation measures listed in Table 4.

**Table 1.** Hyper-parameters for Faster R-CNN and YOLOv5.

S. No.	Learning Rate	Epoch	Loss for Faster R-CNN	Loss for YOLOv5
1	0.0002	2 k	0.046	0.124
2	0.0002	4 k	0.029	0.081
3	0.0002	6 k	0.028	0.066
4	0.0002	8 k	0.025	0.058
5	0.0002	10 k	0.017	0.049

After training the models with the given training set, performance evaluation of both models is conducted on the testing data from TobSet. The accuracy results obtained by using the Faster R-CNN-based vision model show its supremacy over YOLOv5. A total of 635 predictions were produced on unseen test images for each model. Detection results for both models are presented in Figures 10 and 11. The YOLOv5-based model did not perform well on some test samples, as illustrated in Figure 12.

**Table 2.** Confusion matrix for Faster R-CNN-based model.

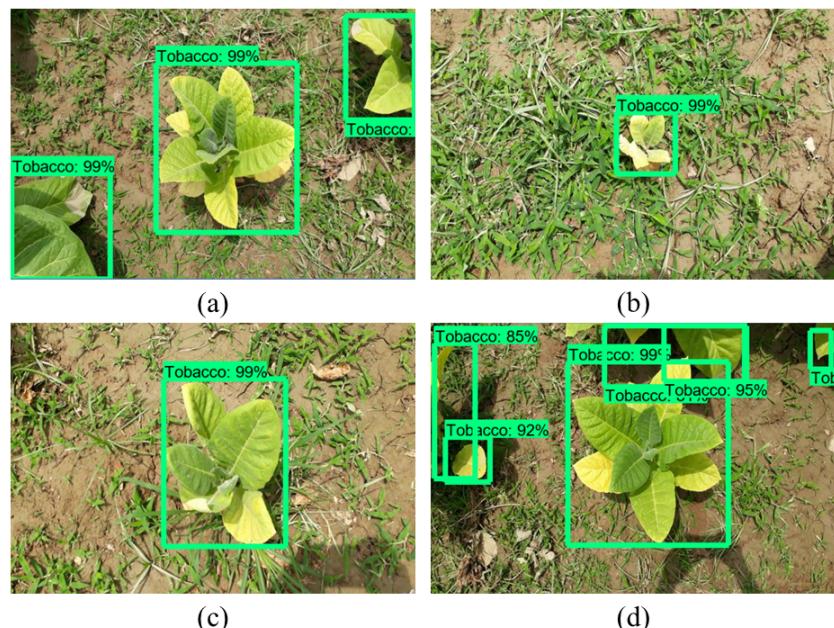
True Class	Predicted Class	
	Tobacco	Weeds
	Tobacco	454
Weeds	7	168
	98.48%	96.55%

**Table 3.** Confusion matrix for YOLOv5-based model.

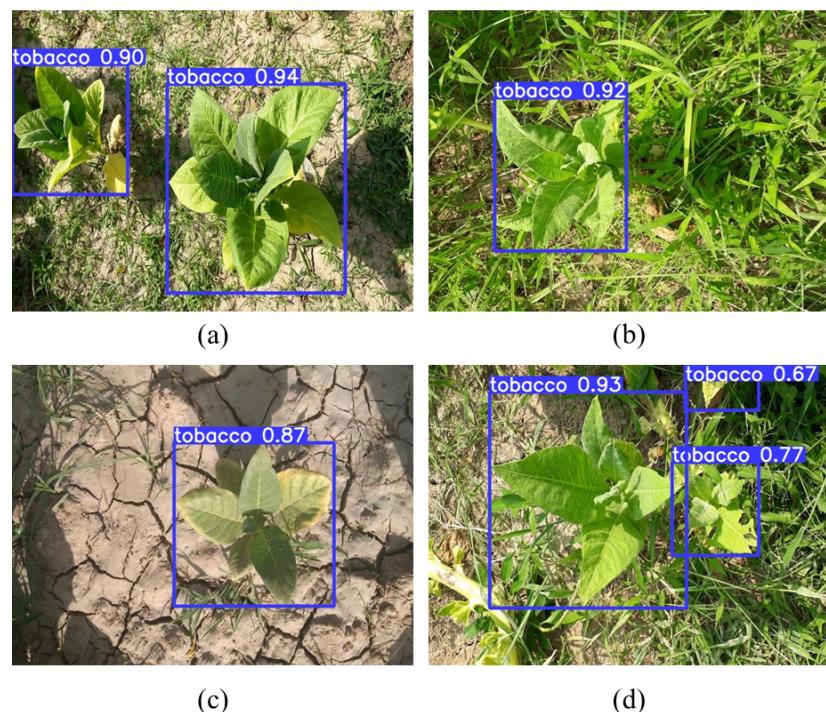
True Class	Predicted Class	
	Tobacco	Weeds
	Tobacco	437
Weeds	24	161
	94.79%	92.52%

**Table 4.** Evaluation measures for Faster R-CNN and YOLOv5.

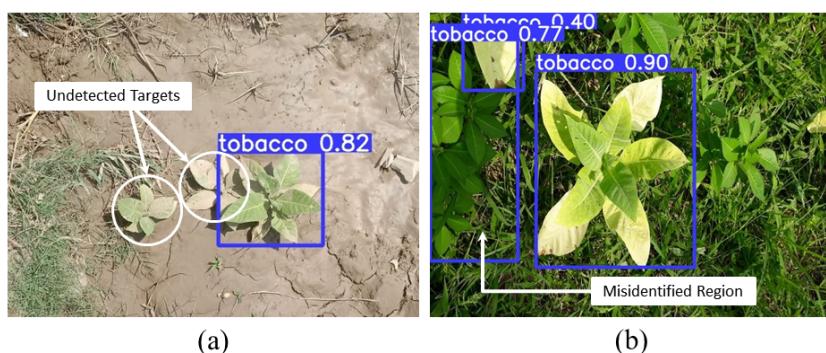
S.No.	Evaluation Measure	Faster R-CNN	YOLOv5
1	Precision	0.9829	0.9481
2	Recall	0.9863	0.9732
3	F1-Score	0.9859	0.9576
4	Accuracy	0.9834	0.9445



**Figure 10.** Faster R-CNN detection results of tobacco from testing data in varying scenarios: (a) high intra-row plant distance. (b) high weed density. (c) low weed density. (d) low intra-row plant distance.



**Figure 11.** YOLOv5 detection results of tobacco from testing data in varying scenarios: (a) high intra-row plant distance. (b) high weed density. (c) low weed density. (d) low intra-row plant distance.

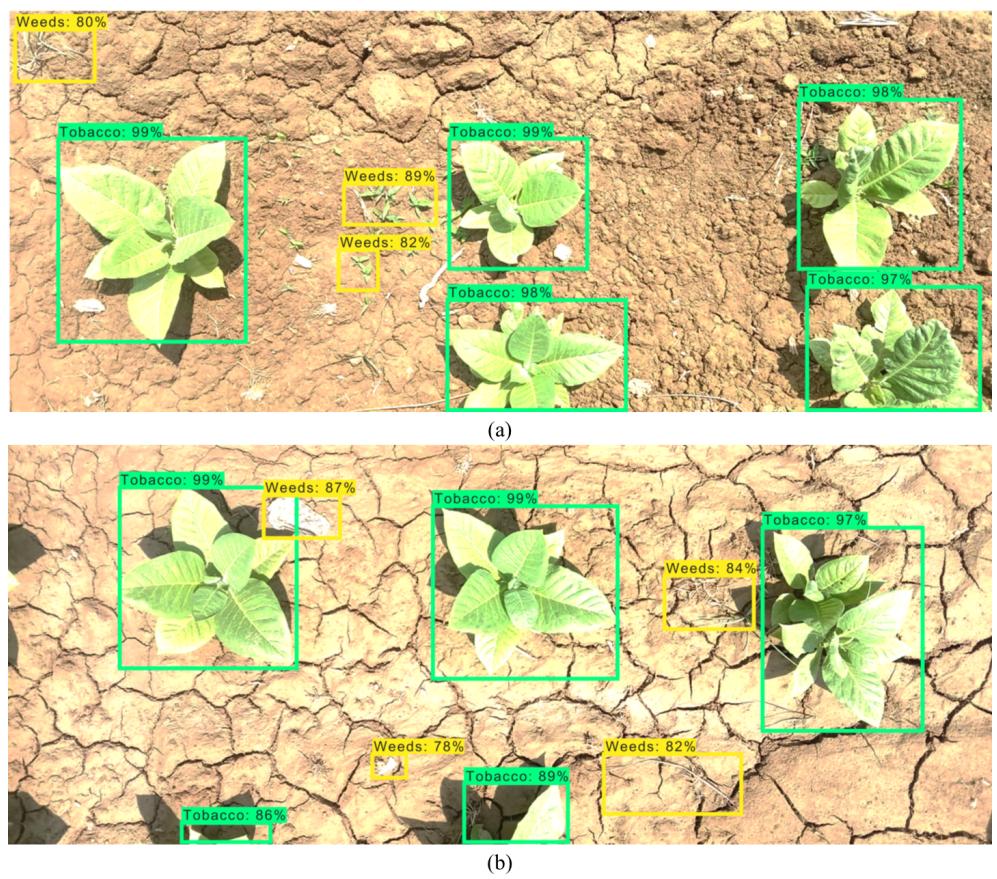


**Figure 12.** YOLOv5 detection results with (a) undetected targets, and (b) misidentified regions.

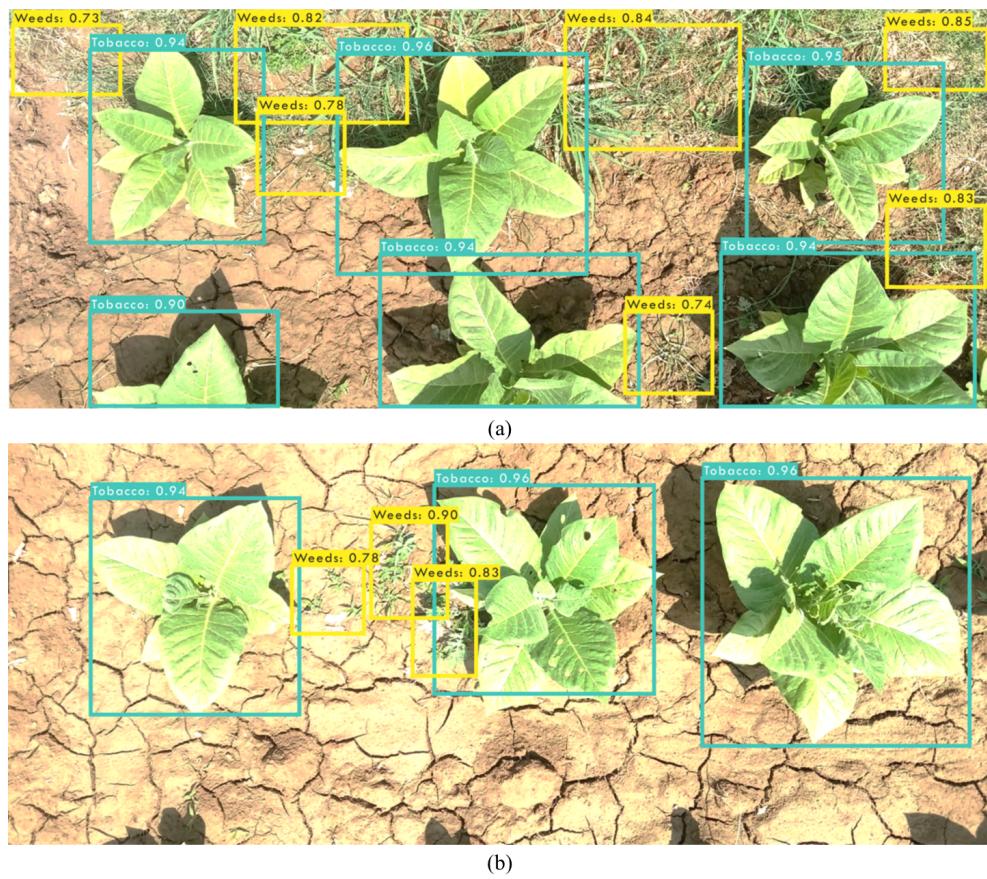
#### Real-Time Inference

The proposed vision models are evaluated in real tobacco fields on a mobile robot spraying platform. For obtaining higher inference in real-time, NVIDIA's optimized library for faster deep-learning inference, i.e., NVIDIA TensorRT, was used. The modified Faster R-CNN and YOLOv5-based vision models identified tobacco plants at 10 and 16 fps, and with classification accuracies of 98% and 94%, respectively, at a robot speed of approximately 3 km/h. The modified YOLOv5-based model can process images at a higher frame rate compared to the Faster R-CNN model, thus making it a better choice for real-time deployment on a spraying robot. Real-time detection results for both models are presented in Figures 13 and 14.

Table 5 presents each model's inference results in real time. YOLOv5 outperformed the Faster R-CNN model in terms of inference speed.



**Figure 13.** Real-time Faster R-CNN detection of tobacco crop and weeds in scenarios with (a) low intra-row plant distance, and (b) high intra-row plant distance.



**Figure 14.** Real-time YOLOv5 detection of tobacco crop and weeds in scenarios with (a) low intra-row plant distance, and (b) high intra-row plant distance.

**Table 5.** Inference results of the models in real-time.

Model	Faster R-CNN	YOLOv5
Inference time (ms)	$98.5 \pm 5$	$62 \pm 5$
Frames per second (fps)	10.15	16.12
mAP	0.94	0.91

## 7. Conclusions

Intelligent precision agriculture robot sprayers for agrochemical application must be robust enough to distinguish crops from weeds to perform targeted spraying to reduce the usage of agrochemicals. In this paper, two different CNN-based approaches, namely, Faster R-CNN and YOLOv5, are explored in order to develop a vision-based framework for the detection and classification of tobacco crop and weeds in the actual fields. Both frameworks are first trained and then tested on a self-developed tobacco plants and weeds dataset, TobSet. The dataset comprises 7000 images of tobacco plants and 1000 images of bare soil and weeds taken manually with digital cameras periodically over 2 months. The Faster R-CNN-based vision framework demonstrated higher accuracy and robustness, whereas the YOLOv5-based vision framework demonstrated lower inference time. Experimental implementation is conducted in the tobacco fields with a four-wheeled mobile robot sprayer with a computer possessing a GPU. Classification accuracies of 98% and 94% and frame rates of 10 and 16 fps were recorded for Faster R-CNN and YOLOv5-based models, respectively. Moreover, the precise smart application of pesticides with the proposed system offered 52% reduction in pesticide usage by pinpointing the targets, i.e., tobacco plants.

Faster R-CNN produces higher accuracy but lower fps on computers (especially without GPUs); high computational cost of training makes it challenging for real-time

applications. TobSet demonstrated true assessment of the deep-learning algorithms as it comprises real field images with challenging scenarios possessing different factors of variation, such as dense weed patches, lightening variation, color similarity with weeds, color variation of tobacco plant at different growth stages, and varying growth stages. The classification results of both approaches in real time were slightly lower than the prediction results obtained on the dataset due to higher sunlight intensities. Intended future studies include real-time tobacco plant segmentation for finding canopy size and the desired application flowrate of spray for each tobacco plant.

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

**Funding:** This study received funding from King Saud University, Saudi Arabia, through researcher's supporting project number (RSP-2021/145). The APCs were funded by King Saud University, Saudi Arabia, through researcher's supporting project number (RSP-2021/145).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset used in this study is publicly available on GitHub at <https://github.com/mshahabalam/TobSet> (accessed on 11 October 2021).

**Acknowledgments:** The authors extend their appreciation to King Saud University, Saudi Arabia, for funding this study through researcher's supporting project number (RSP-2021/145).

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

## References

1. Iqbal, J.; Rauf, A. Tobacco Revenue and Political Economy of Khyber Pakhtunkhwa. *FWU J. Soc. Sci.* **2021**, *15*, 11–25.
2. Wang, G.; Lan, Y.; Yuan, H.; Qi, H.; Chen, P.; Ouyang, F.; Han, Y. Comparison of spray deposition, control efficacy on wheat aphids and working efficiency in the wheat field of the unmanned aerial vehicle with boom sprayer and two conventional knapsack sprayers. *Appl. Sci.* **2019**, *9*, 218. [[CrossRef](#)]
3. Liu, W.; Wu, C.; She, D. Effect of spraying direction on the exposure to handlers with hand-pumped knapsack sprayer in maize field. *Ecotoxicol. Environ. Saf.* **2019**, *170*, 107–111. [[CrossRef](#)] [[PubMed](#)]
4. Hughes, E.A.; Flores, A.P.; Ramos, L.M.; Zalts, A.; Glass, C.R.; Montserrat, J.M. Potential dermal exposure to deltamethrin and risk assessment for manual sprayers: Influence of crop type. *Sci. Total Environ.* **2008**, *391*, 34–40. [[CrossRef](#)]
5. Ellis, M.B.; Lane, A.; O'Sullivan, C.; Miller, P.; Glass, C. Bystander exposure to pesticide spray drift: new data for model development and validation. *Biosyst. Eng.* **2010**, *107*, 162–168. [[CrossRef](#)]
6. Kim, K.D.; Lee, H.S.; Hwang, S.J.; Lee, Y.J.; Nam, J.S.; Shin, B.S. Analysis of Spray Characteristics of Tractor-mounted Boom Sprayer for Precise Spraying. *J. Biosyst. Eng.* **2017**, *42*, 258–264.
7. Matthews, G. *Pesticide Application Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
8. Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H.; Shah, M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artif. Intell. Agric.* **2020**, *4*, 58–73. [[CrossRef](#)]
9. Mavridou, E.; Vrochidou, E.; Papakostas, G.A.; Pachidis, T.; Kaburlasos, V.G. Machine vision systems in precision agriculture for crop farming. *J. Imaging* **2019**, *5*, 89. [[CrossRef](#)]
10. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2020**, *7*, 1–19. [[CrossRef](#)]
11. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
12. Osman, Y.; Dennis, R.; Elgazzar, K. Yield Estimation and Visualization Solution for Precision Agriculture. *Sensors* **2021**, *21*, 6657. [[CrossRef](#)]
13. Bechar, A.; Vigneault, C. Agricultural robots for field operations: Concepts and components. *Biosyst. Eng.* **2016**, *149*, 94–111. [[CrossRef](#)]
14. Berenstein, R.; Edan, Y. Automatic adjustable spraying device for site-specific agricultural application. *IEEE Trans. Autom. Sci. Eng.* **2017**, *15*, 641–650. [[CrossRef](#)]

15. Arakeri, M.P.; Kumar, B.V.; Barsaiya, S.; Sairam, H. Computer vision based robotic weed control system for precision agriculture. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Udupi, India, 13–16 September 2017; pp. 1201–1205.
16. Gázquez, J.A.; Castellano, N.N.; Manzano-Agüero, F. Intelligent low cost telecontrol system for agricultural vehicles in harmful environments. *J. Clean. Prod.* **2016**, *113*, 204–215. [[CrossRef](#)]
17. Faiçal, B.S.; Freitas, H.; Gomes, P.H.; Mano, L.Y.; Pessin, G.; de Carvalho, A.C.; Krishnamachari, B.; Ueyama, J. An adaptive approach for UAV-based pesticide spraying in dynamic environments. *Comput. Electron. Agric.* **2017**, *138*, 210–223. [[CrossRef](#)]
18. Zhu, H.; Lan, Y.; Wu, W.; Hoffmann, W.C.; Huang, Y.; Xue, X.; Liang, J.; Fritz, B. Development of a PWM precision spraying controller for unmanned aerial vehicles. *J. Bionic Eng.* **2010**, *7*, 276–283. [[CrossRef](#)]
19. Yang, Y.; Hannula, S.P. Development of precision spray forming for rapid tooling. *Mater. Sci. Eng. A* **2008**, *477*, 63–68. [[CrossRef](#)]
20. Tellaeche, A.; BurgosArtizzi, X.P.; Pajares, G.; Ribeiro, A.; Fernández-Quintanilla, C. A new vision-based approach to differential spraying in precision agriculture. *Comput. Electron. Agric.* **2008**, *60*, 144–155. [[CrossRef](#)]
21. Tewari, V.; Pareek, C.; Lal, G.; Dhruw, L.; Singh, N. Image processing based real-time variable-rate chemical spraying system for disease control in paddy crop. *Artif. Intell. Agric.* **2020**, *4*, 21–30. [[CrossRef](#)]
22. Rincón, V.J.; Grella, M.; Marucco, P.; Alcatrão, L.E.; Sanchez-Hermosilla, J.; Balsari, P. Spray performance assessment of a remote-controlled vehicle prototype for pesticide application in greenhouse tomato crops. *Sci. Total Environ.* **2020**, *726*, 138509. [[CrossRef](#)]
23. Gil, E.; Llorens, J.; Llop, J.; Fàbregas, X.; Escolà, A.; Rosell-Polo, J. Variable rate sprayer. Part 2—Vineyard prototype: Design, implementation, and validation. *Comput. Electron. Agric.* **2013**, *95*, 136–150. [[CrossRef](#)]
24. Alam, M.; Alam, M.S.; Roman, M.; Tufail, M.; Khan, M.U.; Khan, M.T. Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture. In Proceedings of the International Conference on Electrical and Electronics Engineering, Antalya, Turkey, 14–16 April 2020; pp. 273–280.
25. Tufail, M.; Iqbal, J.; Tiwana, M.I.; Alam, M.S.; Khan, Z.A.; Khan, M.T. Identification of Tobacco Crop Based on Machine Learning for a Precision Agricultural Sprayer. *IEEE Access* **2021**, *9*, 23814–23825. [[CrossRef](#)]
26. Garcia-Ruiz, F.J.; Wulfsohn, D.; Rasmussen, J. Sugar beet (*Beta vulgaris* L.) and thistle (*Cirsium arvensis* L.) discrimination based on field spectral data. *Biosyst. Eng.* **2015**, *139*, 1–15. [[CrossRef](#)]
27. Li, X.; Chen, Z. Weed identification based on shape features and ant colony optimization algorithm. In Proceedings of the International Conference on Computer Application and System Modeling, Taiyuan, China, 22–24 October 2010; Volume 1, p. V1-384.
28. Burgos-Artizzi, X.P.; Ribeiro, A.; Guijarro, M.; Pajares, G. Real-time image processing for crop/weed discrimination in maize fields. *Comput. Electron. Agric.* **2011**, *75*, 337–346. [[CrossRef](#)]
29. Cheng, B.; Matson, E.T. A feature-based machine learning agent for automatic rice and weed discrimination. In *International Conference on Artificial Intelligence and Soft Computing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 517–527.
30. Guru, D.; Mallikarjuna, P.; Manjunath, S.; Shenoi, M. Machine vision based classification of tobacco leaves for automatic harvesting. *Intell. Autom. Soft Comput.* **2012**, *18*, 581–590. [[CrossRef](#)]
31. Haug, S.; Michaels, A.; Biber, P.; Ostermann, J. Plant classification system for crop/weed discrimination without segmentation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, 24–26 March 2014; pp. 1142–1149.
32. Rumpf, T.; Römer, C.; Weis, M.; Sökefeld, M.; Gerhards, R.; Plümer, L. Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*. *Comput. Electron. Agric.* **2012**, *80*, 89–96. [[CrossRef](#)]
33. Ouyang, W.; Zeng, X.; Wang, X.; Qiu, S.; Luo, P.; Tian, Y.; Li, H.; Yang, S.; Wang, Z.; Li, H.; et al. DeepID-Net: Object detection with deformable part based convolutional neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1320–1334. [[CrossRef](#)]
34. Diba, A.; Sharma, V.; Pazandeh, A.; Pirsavash, H.; Van Gool, L. Weakly supervised cascaded convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 914–922.
35. Toshev, A.; Szegedy, C. Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1653–1660.
36. Chen, X.; Yuille, A. Articulated pose estimation by a graphical model with image dependent pairwise relations. *arXiv* **2014**, arXiv:1407.3399.
37. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
38. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
39. Lin, L.; Wang, K.; Zuo, W.; Wang, M.; Luo, J.; Zhang, L. A deep structured model with radius-margin bound for 3d human activity recognition. *Int. J. Comput. Vis.* **2016**, *118*, 256–273. [[CrossRef](#)]
40. Cao, S.; Nevatia, R. Exploring deep learning based solutions in fine grained activity recognition in the wild. In Proceedings of the 23rd International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 384–389.
41. Doulamis, N. Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. *Multimed. Tools. Appl.* **2018**, *77*, 9651–9689. [[CrossRef](#)]

42. Lopez-Martin, M.; Carro, B.; Sanchez-Esguevillas, A. IoT type-of-traffic forecasting method based on gradient boosting neural networks. *Future Gener. Comput. Syst.* **2020**, *105*, 331–345. [[CrossRef](#)]
43. 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](#)]
44. Yu, J.; Schumann, A.W.; Cao, Z.; Sharpe, S.M.; Boyd, N.S. Weed detection in perennial ryegrass with deep learning convolutional neural network. *Front. Plant Sci.* **2019**, *10*, 1422. [[CrossRef](#)] [[PubMed](#)]
45. Asad, M.H.; Bais, A. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Inf. Process. Agric.* **2019**, *7*, 535–545. [[CrossRef](#)]
46. Umamaheswari, S.; Arjun, R.; Meganathan, D. Weed detection in farm crops using parallel image processing. In Proceedings of the Conference on Information and Communication Technology, Jabalpur, India, 26–28 October 2018; pp. 1–4.
47. Bah, M.D.; Dericquebourg, E.; Hafiane, A.; Canals, R. Deep learning based classification system for identifying weeds using high-resolution UAV imagery. In *Science and Information Conference*; Springer: Cham, Switzerland, 2018; pp. 176–187.
48. Forero, M.G.; Herrera-Rivera, S.; Ávila-Navarro, J.; Franco, C.A.; Rasmussen, J.; Nielsen, J. Color classification methods for perennial weed detection in cereal crops. In *Iberoamerican Congress on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 117–123.
49. Wang, A.; Zhang, W.; Wei, X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* **2019**, *158*, 226–240. [[CrossRef](#)]
50. Hu, K.; Wang, Z.; Coleman, G.; Bender, A.; Yao, T.; Zeng, S.; Song, D.; Schumann, A.; Walsh, M. Deep Learning Techniques for In-Crop Weed Identification: A Review. *arXiv* **2021**, arXiv:2103.14872.
51. Loey, M.; ElSawy, A.; Afify, M. Deep learning in plant diseases detection for agricultural crops: A survey. *Int. J. Serv. Sci. Manag. Eng. Tech.* **2020**, *11*, 41–58. [[CrossRef](#)]
52. Weng, Y.; Zeng, R.; Wu, C.; Wang, M.; Wang, X.; Liu, Y. A survey on deep-learning-based plant phenotype research in agriculture. *Sci. Sin. Vitae* **2019**, *49*, 698–716. [[CrossRef](#)]
53. Bu, F.; Gharajeh, M.S. Intelligent and vision-based fire detection systems: A survey. *Image Vis. Comput.* **2019**, *91*, 103803. [[CrossRef](#)]
54. Chouhan, S.S.; Singh, U.P.; Jain, S. Applications of computer vision in plant pathology: A survey. *Arch. Comput. Methods Eng.* **2020**, *27*, 611–632. [[CrossRef](#)]
55. Bonadies, S.; Gadsden, S.A. An overview of autonomous crop row navigation strategies for unmanned ground vehicles. *Eng. Agric. Environ. Food* **2019**, *12*, 24–31. [[CrossRef](#)]
56. Tripathi, M.K.; Makedar, D.D. A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey. *Inf. Process. Agric.* **2020**, *7*, 183–203. [[CrossRef](#)]
57. da Costa, A.Z.; Figueroa, H.E.; Fracarolli, J.A. Computer vision based detection of external defects on tomatoes using deep learning. *Biosyst. Eng.* **2020**, *190*, 131–144. [[CrossRef](#)]
58. dos Santos Ferreira, A.; Freitas, D.M.; da Silva, G.G.; Pistori, H.; Folhes, M.T. Weed detection in soybean crops using ConvNets. *Comput. Electron. Agric.* **2017**, *143*, 314–324. [[CrossRef](#)]
59. Yu, J.; Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. Deep learning for image-based weed detection in turfgrass. *Eur. J. Agron.* **2019**, *104*, 78–84. [[CrossRef](#)]
60. Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. Detection of Carolina geranium (*Geranium carolinianum*) growing in competition with strawberry using convolutional neural networks. *Weed Sci.* **2019**, *67*, 239–245. [[CrossRef](#)]
61. Le, V.N.T.; Truong, G.; Alameh, K. Detecting weeds from crops under complex field environments based on Faster RCNN. In Proceedings of the IEEE Eighth International Conference on Communications and Electronics, Phu Quoc Island, Vietnam, 13–15 January 2021; pp. 350–355.
62. Quan, L.; Feng, H.; Lv, Y.; Wang, Q.; Zhang, C.; Liu, J.; Yuan, Z. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R-CNN. *Biosyst. Eng.* **2019**, *184*, 1–23. [[CrossRef](#)]
63. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodríguez, L. A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering* **2020**, *2*, 471–488. [[CrossRef](#)]
64. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [[CrossRef](#)]
65. 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](#)]
66. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
67. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 818–833.
68. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
69. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.