Deep Learning-Based Object Detection System for Identifying Weeds Using UAS Imagery

Aaron Etienne 1, Aanis Ahmad 2, Varun Aggarwal 2 and Dharmendra Saraswat 1,*

1 Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA; aetienne@purdue.edu
2 Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; ahmad31@purdue.edu (A.A.); aggarw82@purdue.edu (V.A.)
* Correspondence: saraswat@purdue.edu; Tel.: +1-(765)-494-5013

Abstract: Current methods of broadcast herbicide application cause a negative environmental and economic impact. Computer vision methods, specifically those related to object detection, have been reported to aid in site-specific weed management procedures for targeted herbicide application within a field. However, a major challenge to developing a weed detection system is the requirement for a properly annotated database to differentiate between weeds and crops under field conditions. This research involved creating an annotated database of 374 red, green, and blue (RGB) color images organized into monocot and dicot weed classes. The images were acquired from corn and soybean research plots located in north-central Indiana using an unmanned aerial system (UAS) flown at 30 and 10 m heights above ground level (AGL). A total of 25,560 individual weed instances were manually annotated. The annotated database consisted of four different subsets (Training Image Sets 1–4) to train the You Only Look Once version 3 (YOLOv3) deep learning model for five separate experiments. The best results were observed with Training Image Set 4, consisting of images acquired at 10 m AGL. For monocot and dicot weeds, respectively, an average precision (AP) score of 91.48% and 86.13% was observed at a 25% IoU threshold (AP @ T = 0.25), as well as 63.37% and 45.13% at a 50% IoU threshold (AP @ T = 0.5). This research has demonstrated a need to develop large, annotated weed databases to evaluate deep learning models for weed identification under field conditions. It also affirms the findings of other limited research studies utilizing object detection for weed identification under field conditions.

Keywords: weed detection; artificial intelligence; machine learning; deep learning; object detection; UAS

1. Introduction

Weed infestations have been globally reported to cause yield losses in all field crops. In 2018, noxious weed infestation alone contributed to 30% of total yield loss worldwide [1]. In North America alone, weed infestations were reported to cause a 40-billion-dollar (USD) loss in harvest profit in the 2018 growing season [2]. Chemical weed control via the use of herbicides is a crucial component to crop health and yield. Broadcast application, the current standard in agriculture, involves the uniform distribution of herbicide over the entire field, regardless if there are weeds present or not. This practice has negative environmental implications and is financially detrimental to farming operations [3]. The ability to detect, identify, and control weed growth in the early stages of plant development is necessary for crop development. In the early crop production season, an effective management strategy helps prevent weed infestation from spreading to other field areas. Early-season site-specific weed management (ESSWM) is achievable by implementing this strategy on a plant-by-plant level [4]. In the practice of ESSWM, an automatic weed detection strategy can be utilized to spray only where a weed is present in-field. Advances in computer vision techniques have generated researchers’ interest in developing automated systems capable of accurately identifying weeds.
Various computer vision techniques have been used across different engineering disciplines. Computer vision is commonly used in the healthcare industry to evaluate different diseases [5], lesions [6,7], and detect cancer [8]. It was observed that the YOLO object-detection model performed the best for breast cancer detection [8]. In addition, computer vision is used for autonomous vehicles to develop self-driving cars [9], ground robots [10], and unmanned aerial systems [11]. Security applications such as facial recognition [12], pedestrian avoidance [13], and obstacle avoidance [14] also rely on computer vision. Although computer vision is commonly used, its recent implementation in precision agriculture applications has shown promising results for the detection of different stresses within crop fields, such as weeds [15], diseases [16], pests [17], nutrient deficiencies [18], etc. In addition, it has been used for fruit counting [19], crop height detection [20], automation [21], and assessment of fruit and vegetable quality [22]. Data from different sensors are utilized to implement computer vision techniques in agriculture. Stereo camera sensors have also been used for computer vision applications [10,14,20]. Hyperspectral and multispectral sensors are commonly used for weed identification to obtain detailed information and pick up multiple different channels [23]. Although research has been conducted and solutions have been developed using a range of sensors, red, green, and blue (RGB) sensors are the most popular, cost less, are easy to use, and are readily available [23,24].

Before the popularity of deep learning-based computer vision, traditional image processing and machine learning algorithms were commonly used by the research community. Computer vision systems using image processing were developed to discriminate between crop rows and weeds in real-time [25] and weed identification using multispectral and RGB imagery [26]. Machine learning was also recently used to identify weeds in corn using hyperspectral imagery [1] and in rice using stereo computer vision [27]. However, as traditional image processing and machine learning algorithms relied on manual feature extraction [28], the algorithms were less generalizable [29] and prone to bias [30]. Therefore, training deep learning models gained popularity as they rely on convolutional neural networks capable of automatically extracting important features from images [31]. Deep learning was recently used for weed identification in corn using the You Only Look Once (YOLOv3) algorithm [15].

Although promising results have been reported for weed identification, the development of deep learning models capable of accurately identifying weeds from UAS are limited. UAS mounted with hyperspectral and multispectral [32] sensors were used for weed identification. However, as RGB sensors are cost-effective, machine learning was recently used by mounting RGB sensors on UAS for acquiring images at 30, 60, and 90 m altitude, respectively, for weed identification [33]. Machine learning-based support vector machines (SVM), along with the YOLOv3 and Mask RCNN deep learning models, were used for weed identification using multispectral imagery acquired using a UAS at an altitude of 2 m [30]. YOLOv3 was also used to identify weeds in winter wheat using UAS-acquired imagery at 2 m altitude [34].

Flying a UAS at a low altitude allows obtaining higher spatial resolution imagery than manned aircraft or satellites [35]. A UAS also provides a high temporal resolution to track physical and biological changes in a field over time [36]. UAS-based imagery was implemented to train a DNN for weed detection [37], resulting in high testing accuracy. Similarly, UAS-based multispectral imagery was successfully used to develop a crop/weed segmentation and mapping framework on a whole field basis [32].

Despite a few successful research outcomes reported previously, weed detection has proven difficult within tilled and no-till row-crop fields. These fields present a complex and challenging environment for a computer vision application. Weeds and crops have similar spectral characteristics and share physical similarities early in the growing season. Soil conditions may also vary heavily within a small area, and the presence of stalks and debris in no-till or minimal-till fields can cause problems with false detection. In addition, differing weather conditions can affect how well a weed detection system can discriminate weeds from non-weeds. In recent years, various deep learning methods have been used
Remote Sens. 2021, 13, 5182

for UAS-based weed detection applications, as deep learning relies on convolutional neural networks capable of learning essential features from an imagery dataset. A fully convolutional network (FCN) was proposed for semantic labeling and weed mapping based on UAS imagery [38]. Classification accuracy of 96% has been reported between weed and rice crops using an eight-layer, custom FCN.

Besides classification, detection is another popular deep learning approach utilizing deep convolutional neural networks (DCNNs). Unlike classification, detection can locate and identify multiple objects within images. Therefore, detection is a superior technique for agricultural applications such as a UAS-based SSWM, as the detection model can be trained to identify and locate multiple weeds within imagery [39]. As the breadth and depth of machine learning techniques in agriculture continue to grow, DCNNs prove among the best-performing approaches for detection tasks in remote sensing applications [40].

Faster region-based convolutional neural networks (R-CNNs) [41] and single-shot detectors (SSD) [42] are standard detection models that have been utilized for weed identification. Different pre-trained networks were used with the Faster R-CNN detection model to identify and locate maize seedlings and weeds acquired using a ground robot under three different field conditions [43]. An earlier version of YOLO, YOLOv2, was reported with an F1 score of 97.31% to compare the detection models’ performance. Faster R-CNN was again used along with the SSD detection model to identify weeds in UAS-acquired imagery [39]. Flights were conducted using a DJI Matrice 600 UAS at an altitude of 20 m to acquire images of five weed species in soybean at the V6 growth stage. Although similar results were reported for the two models, it was concluded that Faster R-CNN yielded better confidence scores for identification, which is consistent with the comparisons conducted [44]. Recently, Faster R-CNN was also used to identify two weeds in barley [45]. Although detection is becoming increasingly popular due to the ability to identify and locate objects accurately, faster models and algorithms are required for SSWM.

The You Only Look Once (YOLO) real-time object detection network is a DCNN that has become popular among detection tasks but has limited applications reported in agricultural remote sensing. YOLO predicts multiple bounding boxes and their corresponding class probabilities. It can thus process entire images at once and inherently encode contextual class information and their related appearances [46]. In a study by Li et al. [47], YOLOv3, Faster R-CNN, and SSD detected agricultural greenhouses from satellite imagery. YOLOv3 reported the best detection results. YOLOv3 was compared with other detection models, including Faster R-CNN and SSD, for identifying weeds in lettuce plots [44]. YOLOv3 was also used for identifying and locating multiple vegetation and weed instances within acquired imagery from 1.5 m above the ground [48]. A ground-based spraying system, developed using a customized YOLOv3 network trained on a fusion of UAS-based imagery and imagery from a ground-based agricultural robot, resulted in a weed detection accuracy of 91% [49].

The challenge with using a DCNN, such as YOLO, to train on a weed-detection task using UAS-based imagery is the presence of multiple weed species of varying shapes and sizes within a single image. Image pre-processing techniques such as resizing and segmentation have been applied to make individual weed examples clearer [50]. To the best of our knowledge, limited research studies have been conducted that utilize the YOLO network architecture for weed detection in a production row-crop field, using UAS-based RGB sensor-acquired color imagery. Furthermore, no current studies have reported weed detection results on UAS imagery acquired at different heights above ground level (AGL). The overall goal of this study was to evaluate the performance of YOLOv3 for detecting dicot (broadleaf) and monocot (grass type) weeds, during early crop growth stages, in corn and soybean fields. The study’s specific objectives were (1) compile a database of UAS-acquired images using an RGB sensor for providing critical training data for YOLOv3 models; (2) annotate the RGB image dataset of weeds obtained from multi-location corn and soybean test plots into monocot and dicot classes in YOLO format, and (3) evaluate
the performance of YOLOv3 for identifying monocot and dicot weeds in corn and soybean fields at different UAS heights above ground level.

The remainder of the manuscript is organized as follows: explaining the dataset, preprocessing, evaluation metrics, and models are presented in the Materials and Methods section (Section 2). The results are then presented in Section 3, followed by the discussion in Section 4. Finally, the study is concluded in Section 5.

2. Materials and Methods

This research utilized a DJI Matrice 600 Pro hexacopter UAS for data collection. A 3-axis, programmable gimbal was installed to carry the RGB sensor. A FLIR Duo Pro R (FLIR Systems Inc., Wilsonville, OR, USA), RGB sensor was operated to collect imagery in corn and soybean fields at the four different research sites in Indiana, USA. The RGB camera on the Flir Duo R Pro had a sensor size of 10.9 × 8.7 mm, a field of view (FoV) of 56° × 45°, a spatial resolution of 1.5 cm at 30 m AGL and 0.5 cm at 10 m AGL, a focal length of 8 mm, a pixel size of 1.85 micrometers (µm), and a pixel resolution of 4000 × 3000. Data were collected throughout 2018 and the first month of the 2019 growing season. This study’s experimental research plots were chosen at different locations within Indiana for their weed density and diversification throughout the growing season. Research locations used were the Pinney Purdue Agricultural Center (PPAC) (41.442587, −86.922809), Davis Purdue Agricultural Center (DPAC) (40.262426, −85.151217), Throckmorton Purdue Agricultural Center (TPAC) (40.292592, −86.909956), and The Agronomy Center for Research and Education (ACRE) (40.470843, −86.995994).

Cost-effective yet accurate Ground control points (GCPs) were created from 1 × 1 checkerboard-patterned linoleum squares. The GCPs were placed on a heavy ceramic tile and used to mark each experimental plot’s boundary. At the beginning of the 2018 and 2019 growing seasons, each GCP was surveyed with a Trimble V™ real-time kinematic (RTK) base station and receiver [22] to identify their latitude and longitude coordinates. Waypoints were created and stored so that GCPs did not have to stay in the field year-round yet could be placed in identical locations for missions at each plot. As GCP measurements fell within a 2–3-cm accuracy of the GCP’s physical field locations, they proved to be an invaluable tool in improving the quality of image stitching techniques. While stitched imagery was not utilized for analysis in this study, it proved important in identifying heavy weed infestation areas throughout the growing season. All data collection missions occurred within 30 minutes of solar noon. This meant that UAS flights were performed, and imagery was collected when the sun was at its highest point in the sky. Solar position, time of day, and weather conditions such as fog, clouds, and haze can all affect the spectral composition of sunlight. Constricting the temporal resolution of flights to solar noon was an important step to reducing the atmospheric scattering of radiation. This practice is referred to as radiometric calibration. To ensure temporal accuracy across different data acquisition dates, sites, and altitudes, a Spectralon™ panel with a known reflectance value was utilized before and after each flight. This was done to calibrate RGB sensors. Spectral targets of black, white, red, gray, and green were also utilized to check the wavelength range. To measure the radiance of these targets and the crop and soil background, a SpectraVista™ mass-spectrometer was used. For the post-processing stage of image analysis, data gained from these procedures were downloaded and converted to reflectance values.

DroneDeploy™ mobile flight mission planning software was utilized for data collection missions. DroneDeploy™ is “a flight automation software for unmanned aerial systems that allows users to set a predetermined flight path, speed, and percentage value of side and front overlap” [51]. Side and front overlap values were set at 75%. This allowed each consecutive image captured to overlap 75% with the previous flight line (beside) and the previous image (in front of).

At PPAC, the most commonly found weeds within the research plot were dicot weeds, including giant ragweed (Ambrosia trifida), velvetleaf (Abutilon theophrasti), common lambsquarters (Chenopodium album), and redroot pigweed (Amaranthus retroflexus). Monocot
weeds at PPAC were predominately giant foxtail (*Setaria faberi*). At DPAC, the most common weeds within the research plot were tall waterhemp (*Amaranthus tuberculatus*), velvetleaf, and prickly sida (*Sida spinosa*). At TPAC, the most commonly occurring weeds found in the research plots were giant ragweed and redroot pigweed (*Amaranthus retroflexus*).

Monocot weeds commonly found in the research plots at TPAC were green foxtail (*Setaria viridis*) and panicum (*Poaceae virgatum*). At ACRE, common dicot weeds included giant ragweed. For the 2018 growing season at TPAC, the soybean research plot was planted on 27 May 2018. Planting began on the TPAC corn plots and the PPAC and DPAC corn and soybean research plots on 27 April, 28 May, and 17 May. A timeline of the UAS data acquisition dates over the 2018 growing season is shown in Figure 1.

![Figure 1. Data acquisition dates for 2018 and 2019. Each date also denotes where UAS flights were performed to collect data.](image)

Data collection was also done during the first month of the 2019 growing season at ACRE and TPAC. Due to the high amount of rain received during the Spring of 2019, planting at these locations did not occur until late May. Planting at the TPAC and ACRE
corn plots began on 24 May and 28 May, respectively. The timeline of the 2019 data collection is shown in Figure 1.

Research plot fields were rotated between corn and soybeans between the 2018 and 2019 growing seasons at TPAC. Therefore, the location of the study area moved down six research plots (to the west within the research field) for the 2019 planting season, due to a change in experimental layout and planting. Data collection at ACRE was done to increase the number of images for the 2019 growing season. Late planting and the research conclusion dictated the removal of DPAC and PPAC from the flights conducted in 2019, as neither location had planted plots when data collection was undertaken. The first location was flown on 22 May 2019. This was done during pre-plant (PP). The second plot location, flown 4 June 2019, had corn at the first vegetative (V1) growth stage. A further explanation of corn and soybean growth stages is given in Figure 2, where growth stages from emergence (VE) to fifth vegetative state (V5) are visualized. VE refers to the stage where the corn plant breaks through the soil surface.

| Corn Growth Stages in Image Sets | Soybean Growth Stages in Image Sets |
|----------------------------------|-------------------------------------|
| VE     | V1 | V3 | V5 | VE | V1 | V3 | V5 |

Figure 2. Corn and soybean growth stages (adapted from [52,53]).

2.1. Network Architecture—YOLOv3

This research utilized an unmodified YOLOv3 network for weed detection from UAS images [54]. In total, YOLOv3 consists of 106 layers. The initial 53 layers are called the network’s backbone, while the latter 53 are called the detection layers. An illustration of the YOLOv3 architecture is shown in Figure 3.

Figure 3. YOLOv3 network architecture—detection at three different scales are combined for the final detection (Figure adapted from [55]).
The function of the first 53 layers is to extract features from the input images. These layers are based on an open-source neural network framework called Darknet [54]. Hence, together, these layers are called Darknet53. The extracted features are used for detection at three different scales. These detections are then combined to get the final detection. The scale of plants in UAS images varies with changes in flight height and changes in the plant’s growth stage. Hence the multiscale detection of YOLOv3 is beneficial, especially for identifying monocot and dicot plants in UAS images.

2.2. Labeling and Annotation of the Weeds Dataset

Images chosen for manual annotation were acquired during the 2018 (April through September) and early 2019 (May and June) growing season for the corn and soybean crops. Annotation refers to the physical creation of bounding boxes around a selected object. In contrast, a label refers to the name of the class (for example—monocot or dicot in this study) assigned to the annotation that is saved as metadata for the class file. Manual image annotation and labeling are tedious yet crucial steps in supervised learning. To train a network to detect each instance of an object, it must have a way to learn how the object looks and its position in relation to other objects within an image where detection is warranted. To accurately annotate and label the high-resolution UAS-based imagery collected, an image-editing tool with enough speed to label hundreds of weeds in an image and enhanced zoom features that focus on a particular region and clarity of the weeds present therein was required. After meticulous research and testing were performed, a labeling tool was finally found, LabelImg, which provided adequate labeling speed and user-friendly zoom and pan features [56]. LabelImg converts each bounding box to five numbers. The first number corresponds to the class label. The second and third numbers correspond to the x- and y-coordinates of the center of the bounding box. Finally, the fourth and fifth numbers correspond to the width and height of the bounding box. An example of bounding box annotation using LabelImg is shown in Figure 4 below.

![Figure 4. Labelling manual labeling process and output.](image)

Due to the amount of time required to complete manual annotation, speed and ease of use become increasingly important in the overall dataset creation workflow. A labeled RGB imagery dataset was created to train the weed detection model. RGB cameras were chosen as they are more affordable and more commonly used by farmers than other UAS-based sensors, such as multispectral or thermal ones [57].

2.3. Methods for Choosing Imagery for Manual Annotation

A total of 25 flights, each lasting approximately 24.5 min, were carried out in the 2018 and early 2019 growing seasons to acquire an average of 1113 RGB images per flight. While 27,825 RGB images were acquired in 2018 and 2019 combined, only 374 of these images
were annotated for network training dataset creation. The reason for selecting a smaller percentage of total imagery data is because only early season and pre-plant imagery were chosen for manual annotation. Individual weed examples during later crop growth stages could not be used for annotation due to the difficulty of finding individual weeds within the imagery to be labeled. All of the later growth stage imagery had areas of heavy weed infestation, where weeds were clumped together. These images might be useful for other researchers and future network training, but the main focus of the reported research was on the early season, individual weed identification. A flowchart of the manually annotated training image set creation is shown in Figure 5. Each set will be denoted as “Training Image Set #” to explain the training set creation methods better.

![Training Image Set Creation Flowchart](image)

**Figure 5.** Training image set creation flowchart. All training image sets are trained with YOLOv3.

During the formation of Training Image Set 1, UAS-based weed imagery was acquired during the early growth stages of corn and soybeans in 2018 at 30 m AGL. For the creation of Training Image Set 2, 2019 pre-plant weed imagery was acquired at 10 m AGL and added to Training Image Set 1. Images of newly emerged to early-stage corn plots in 2019 were acquired at 10 m AGL for Training Image Set 3 to compare the differences between object detection performance on a multi-crop image set, such as Training Image Sets 1 and 2, to an image set from a single crop plot. There were no monocot weeds in this imagery, as no monocot weeds were present in this early growth stage imagery. This observation has been denoted as N/A in Table 1. While configuring Training Image Set 4, early growth-stage corn plot imagery was acquired at 10 m AGL. No imagery was acquired from soybean plots during this time. Within these images, corn had fully emerged, and it was easier to decipher between weeds and crops. The criteria for identifying emerged corn included a longer average leaf length, larger average leaf diameter, a succinct spacing of corn plants based on planting pattern, and other physical traits, such as leaf structure [58]. A weed-free image showing only corn was added to the training set for every monocot and dicot labeled image. A corresponding empty text file was also needed to inform the YOLO network that there were no labeled objects. This step ensured that there were as many negative samples [59], i.e., images containing only non-desirable objects, as there are numbers of labeled images (positive samples).

| Front-Ends | # of Nodes | Cores per Node | Memory per Node | GPUs per Node |
|------------|------------|----------------|-----------------|---------------|
| GPU        | 2          | 20             | 96 GB           | 1 NVIDIA Tesla P100 |

**Table 1.** Description of the Gilbreth hardware specifications.
2.4. Hardware and Software Setup for Deep Learning Network Training

An Alienware R3 laptop computer with a 2.8 GHz Core i7-7700HQ processor with 32 gigabytes (GB) of RAM was used to train the YOLOv3 network. A 6 GB NVIDIA GTX 1060 graphical processing unit (GPU) was installed on the laptop to enable deep learning network training. The Ubuntu operating system, version 18.04, was installed on the laptop. Ubuntu was disk partitioned to 200 GB, as it was set up to be dual booted alongside the Windows operating system already available in the computer. Access to Purdue University’s community supercomputer cluster allowed for network training at a larger scale. The Gilbreth cluster utilized for this training is “optimized for communities running GPU intensive applications such as machine learning.” The Gilbreth cluster is comprised of “Dell compute nodes with Intel Xeon processors and NVIDIA Tesla GPUs” [60]. A detailed overview of the compute nodes and GPU specifications can be found in Table 1 above.

2.5. Constraints in Training Image Set Creation

For the creation of Training Image Set 1, 100 UAS acquired RGB images from the 2018 early season corn and soybean fields were selected for dataset creation. These raw images were of size 4000 × 3000 pixels. Furthermore, images were acquired at 30 m AGL in a two-acre soybean plot surrounded by corn. The soybeans in these images were in the V1–V2 growth stage. The surrounding corn plots planted prior to the soybean plot were in the V3–V4 growth stage. Images were annotated using bounding boxes, and present weeds were labeled as either monocot or dicot. Before the bounding box annotations were created and labeled, images were resized to 416 × 416 pixels during preprocessing. A 1:1 aspect ratio was used per YOLOv3 network constraints. A total of 8638 weeds were annotated and labelled for an average of 86 weeds per image. Each weed was categorized as either monocot or dicot. Aside from allowing for quicker annotation, these two categories were chosen to provide a broader representation of weed types present in the field. Furthermore, the abundance of either weed type will require a different herbicide application to implement an effective weed management strategy. For example, certain herbicides, such as 2,4-D and Dicamba, are used to target dicot weeds. While they can be applied to monocot crops such as corn without damage, dicot crops such as soybean need to be tolerant to avoid harm [61]. The YOLOv3 network reads a bounding box by the pixel location of its four corners within an image, allowing for the bounding box’s width and height to be normalized by the width and height of the resized image. Specifically, the x-center, y-center, width, and height are float values relative to the width and height of an image. The values of these indices were normalized between 0.0 and 1.0. In addition to resizing the images, anchor boxes were calculated based on the manually labeled bounding boxes, using the calc_anchors bash command. This command, built into the Darknet repository [60], was used in this study. Anchor boxes were calculated through a k-means clustering technique, shown in Equations (1)–(5) below:

\[
b_x = \sigma(t_x) + C_x \\
b_y = \sigma(t_y) + C_y \\
b_w = P_w \cdot e^{bw} \\
b_h = P_h \cdot e^{bh} \\
Pr(\text{object}) \times \text{IOU}(b, \text{object}) = \sigma(t_0)
\]

where, \(b_x, b_y, b_w, \) and \(b_h\) are the x coordinate, y coordinate, width, and height of the predicted bounding box, respectively; \(t_x, t_y, t_w, \) and \(t_h\) are the bounding box coordinate predictions made by YOLO; \(C_x\) and \(C_y\) is the top left corner of the grid cell that the anchor lies within; and \(P_w\) and \(P_h\) are the width and height of the anchor [62]. \(C_x, C_y, P_w,\) and \(P_h\) were normalized by the width and height of the image in which the anchor was
being predicted. Finally, $\sigma(t_0)$ outputs the box confidence score. $Pr$ is the probability of a predicted bounding box containing an object, $Pr(Object)$.

Training Image Set 2 added an additional 108 raw images of size $4000 \times 3000$ pixels of the 2019 pre-plant weed imagery. This imagery was taken at 10 m AGL. Training Image Set 2, therefore, comprised of imagery acquired at the height of 30 m and 10 m AGL. After the images were resized to $416 \times 416$ pixels, they were added to Training Image Set 1, making a total of 208 images in Training Image Set 2. After resizing, bounding box annotations were created for these images and labeled as either monocot or dicot. An additional 4077 monocot and dicot weeds were annotated and labeled, resulting in a total of 12,715 weed annotations. This gave an average of 61 weeds per image. The identified weeds were small dicots (less than five centimeters (cm) in diameter). Anchors were again calculated to ensure accuracy with the updated data, and the learning rate was modified from 0.01 to 0.001 in the configuration file. During this time, the training intersection over union (IoU) of the annotated vs. predicted bounding boxes was calculated to predict areas where a specific class of weed was present. The network calculated the predicted bounding boxes based on Equation (1). IoU is the ratio between the area of overlap and the area of union concerning the predicted (through network training) versus ground-truth (manually annotated) bounding boxes [63]. An illustration of the IoU is shown in Figure 6 below.

![Figure 6. Illustration of the IoU on the dicot dataset images.](image)

Computing the IoU scores for each detection allows a threshold, $T$, to be set for converting each calculated IoU score into a specific category. IoU scores above a given threshold are labeled positive predictions, while those below are labeled false predictions. Predictions can be further classified into true positive (TP), false positive (FP), and false negative (FN). A TP occurs when the object is present, and the model detects it, with an IoU score above the specified threshold. There are two scenarios in which an FP can occur. The first happens when an object is present, but the IoU score falls below the set threshold. The second occurs when the model detects an object that is not present. A FN happens when the object is present, but the model does not detect it. In essence, the ground-truth object goes undetected [62]. Precision is the probability of predicted bounding boxes matching the ground-truth boxes. This is detailed in Equation (6).

$$\text{Precision} = \frac{TP}{TP + FP}$$

where:
TP = true positives, FP = false positives, and precision = the true object detection overall detected boxes. Precision scores range from 0 to 1, with high precision implying that most detected objects match the ground-truth objects. For example, a precision score of 0.7 means a 70% match. Recall measures the probability of the ground-truth objects being correctly detected. This is shown in Equation (7).

\[
\text{Recall} = \frac{TP}{TP + FN}
\]  

(7)

where FN = false negatives and recall = the true object detection overall from the ground-truth boxes. The recall is also referred to as the sensitivity and ranges from 0 to 1 [62]. After obtaining the precision and recall, the average precision (AP) and mean average precision (mAP) were calculated for each class and image set, as shown in Equations (8) and (9), respectively.

\[
\text{AP} = \frac{1}{M} \sum_{m=1}^{M} \text{Precision}(m) \triangle \text{Recall}(m)
\]  

(8)

The mAP was calculated by dividing the sum of the APs for all classes by the total number of objects to be detected [55].

\[
\text{mAP} = \frac{\sum_{n=1}^{N} \text{AP}(n)}{N}
\]  

(9)

For the creation of Training Image Set 3, the imagery was collected during the 2019 early season data collection missions. Altitude was set at 10 m AGL for these flights after poor object detection performance was observed for Training Image Set 1 and Training Image Set 2, consisting of images acquired at 30 m AGL and a combination of 30 m and 10 m AGL, respectively. This set comprises 100 raw images of size 4000 × 3000 pixels. As corn was the only crop planted during these data collection dates, corn plots were chosen to make up this dataset. Nearly all weeds present in these fields were dicot. Therefore, all bounding box annotations for the images were labeled as a dicot. For this training, a number of changes were made to the yolov3_weeds.cfg file. Image size was increased to 512 × 512, from 416 × 416 in the previous image set, to test whether resizing to a higher resolution would improve training accuracy. A total of 7795 dicot weeds were annotated and labeled in this set for an average of 78 weeds per image. It was found that image size could be increased by changing the network hyperparameters in the configuration file. Specifically, adjusting the batch and subdivision values to reduce the likelihood of a memory allocation error was experimented with. The learning rate was also modified from 0.001 to 0.0001. The network also recalculated Anchors using the calc_anchors command, which performs Equation (1). Before creating Training Image Set 4, a number of changes had to be made to increase the average precision, lower the false-positive rate, and introduce imagery with monocot weeds present. These steps were undertaken to test the network performance with two-class detection.

Training Image Set 4 contained 166 RGB images, acquired at 10 m AGL, from the V1–V4 growth stages corn fields. Additional raw images of size 4000 × 3000 pixels were collected from the 2019 early season corn fields of the V3–V4 growth stages. These were split into 20 smaller tiles instead of resizing the original image into a smaller pixel resolution. Each weed was manually annotated with bounding boxes and labeled either monocot or dicot. These image tiles were then resized from 894 × 670 to 800 × 800 pixels, matching the configured network size. Weeds within these image tiles were easier to identify and annotate. Creating bounding box annotations and labeling them took less time compared to original, non-split images. The number of weeds per image was significantly decreased as smaller sections of the original image were used. It is recommended that image size be as similar to the network parameter as possible while training [64]. Sixty-six tiles from the split images, resized to 800 × 800 pixels, were chosen to be labelled and annotated. These were added to the 100 previously resized images used in Training Image Set 3. An
additional 5147 monocot and dicot weed instances were annotated and labelled for a total of 12,945 weeds in Training Image Set 4. This also gave an average of 78 weeds per image for network training.

Access to Purdue University’s Gilbreth supercomputer cluster, optimized for GPU-intensive applications [60], allowed for network training at a larger scale. Before the 2019 early season corn training set, Training Image Set 4, could be trained on the Gilbreth cluster, several changes needed to be made in the network configuration file, yolov3.cfg. The additional network training of 20,000 iterations for Training Image Set 4 on the Gilbreth cluster has been defined as Training Image Set 4+. These changes included setting the batch size to 64, subdivisions to 16, angle to 30, and random to 1. The angle hyperparameter allows for random rotation of an image over the training iterations by a specified degree, i.e., 30 degrees. This image augmentation has improved network performance and reduced loss [65]. The random hyperparameter has also been reported to augment images by resizing them every few batches [65]. For network training, the 50% detection threshold (T = 0.5) was used for Image Sets 1, 2, 3, and 4. At this value, only objects that are identified with a confidence value greater than 50% are considered [54]. For Image Set 4+, the 50% threshold was again used along with the 25% detection threshold (T = 0.25). All other parameters remained the same in Training Image Set 4+ as in Training Image Set 4.

3. Results

3.1. Annotated Weed Database Creation Using UAS Imagery

A total of 374 images across the image sets were used for annotation. The specifications for creating Training Image Sets 1–4 are shown in Table 2 below.

| Training Image Set | #Images | T-V Split | Average Diameter [b] | Average Height | Crop Growth Stage |
|--------------------|---------|-----------|----------------------|---------------|------------------|
|                    |         |           | Monocot | Dicot | Monocot | Dicot | Monocot | Dicot | Monocot | Dicot |
| 1                  | 100     | 80–20%    | 16.51 cm | 12.7 cm | 17.78 cm | 8.89 cm | V1–V4 | V1–V4 |
| 2                  | 208     | 80–20%    | 13.97 cm | 11.87 cm | 15.24 cm | 7.89 cm | PP [c]–V4 | PP [c]–V4 |
| 3                  | 100     | 80–20%    | N/A     | 3.81 cm | N/A     | 2.54 cm | N/A | V1–V2 |
| 4/4+               | 166     | 80–20%    | 10.16 cm | 7.62 cm | 11.43 cm | 5.08 cm | V1–V4 | V1–V4 |

[a] T-V Split = training–validation split. [b] Average Diameter = average diameter at the weed flower end to end. [c] PP = pre-plant.

3.2. Training Image Set 1 (30 Meters AGL)

After 20,000 iterations, network training on Training Image Set 1 resulted in an average loss of 4.7%. At an IoU threshold of T = 0.5, the AP of the monocot and dicot weed detection was a meager 9.88% and 3.63%, respectively (AP @ T = 0.5). AP @ T = 0.25 was 25% across all classes. Single-class AP @ T = 0.25 was not available at the time of this training due to memory constraints of the laptop GPU used. A major contributor to this poor performance was the high number of false positives for both monocot and dicot classes. The AP is given for individual classes, while mAP is given for all classes in the dataset. The network training results for Training Image Set 1 are shown in Table 3.

| Value         | Monocot | Dicot |
|---------------|---------|-------|
| AP            | 9.88%   | 3.63% |
| True Positives| 188     | 103   |
| False Positives| 224    | 499   |

mAP @ T = 0.5: 7.29%

This network training showed a higher false-positive rate for dicots than monocots. Poor spatial accuracy was a major contributing factor to this poor result. Testing results with the final network training weights for this image set are shown in Figure 7.
Figure 7. Monocot and dicot detection results from the YOLOv3 network training on Image Set 1.

Initial image collection at 30 m AGL caused difficulty in object labeling and contributed to poor network training performance. The mean average precision (mAP) and loss scores over each training iteration are shown in Figure 8 below.

Figure 8. Map and loss scores over each training iteration from YOLOv3 network training on Image Set 1.
3.3. Training Image Set 2 (30 and 10 Meters AGL)

The results of this training continued to be poor, with a high average loss of 7.77%. The AP @ T = 0.5 was 9.32% and 4.22% for detecting monocot and dicot weeds, respectively. The AP @ T = 0.25 was 22 percent across all classes. Again, many false positives in both classes were a major factor in poor network performance. As this Training Image Set consisted of the images acquired at 30 m AGL from Training Image Set 1 and at 10 m AGL from the additional pre-plant imagery, the object detection model was unable to train well, resulting in poor performance. The network training results for Training Image Set 2 are shown in Table 4 below.

Table 4. Network training results for Training Image Set 2.

| Value            | Monocot | Dicot  |
|------------------|---------|--------|
| AP               | 9.32%   | 4.22%  |
| True Positives   | 188     | 133    |
| False Positives  | 101     | 199    |
| mAP @ T = 0.5    | 6.76%   |        |

While false positives were greatly reduced from the previous training set results, the mAP @ T = 0.5 similarly performed poorly. An example of monocot detection is shown in Figure 9.

Figure 9. Monocot weed detection test set result from the YOLOv3 training on Image Set 2.

Due to the physical differences in size and shape of the monocot weeds to the V1-stage corn (also a monocot), this class AP was significantly better than the dicot weeds, which looked more similar to the early season corn.
3.4. Training Image Set 3 (10 Meters AGL)

Performance measures of the network training results improved slightly with an AP @ T = 0.5 of 20.42%. The average loss of this training was also lower at 3.9%. The AP @ T = 0.25 across all classes (only dicot in this training) was 32%. There were no monocot weeds present in this training set. The network training results for Training Image Set 3 are shown in Table 5 below.

Table 5. Network training results for Training Image Set 3.

| Value              | Monocot | Dicot  |
|--------------------|---------|--------|
| AP @ T = 0.5       | N/A     | 20.42% |
| AP @ T = 0.25      | N/A     | 32%    |
| True Positives     | N/A     | 15     |
| False Positives    |         | 44     |
| mAP @ T = 0.5      |         | 10.21% |

Limiting training to a single class and using images acquired at 10 m AGL improved the dicot AP performance and significantly reduced the number of false positives. Testing results with the final network training weights for this image set are shown in Figure 10 below.

Figure 10 illustrates the improvement in AP of dicot detection but shows poor performance in detecting the intra-row and very small weeds. All the weeds in this image were verified from the ground-truthing as a dicot.

3.5. Training Image Set 4 (10 Meters AGL)

Average loss was substantially lower during this network training, at 0.522%. The AP @ T = 0.5 for monocots and dicots was 58.31% and 31.13%, respectively. The AP @ T = 0.25 increased to 80%, across all classes. The network training results on Image Set 4 are shown in Table 6 below.
Table 6. Network training results for Training Image Set 4.

| Value                  | Monocot   | Dicot     |
|------------------------|-----------|-----------|
| AP                     | 58.31%    | 31.13%    |
| True Positives         | 166       | 410       |
| False Positives        | 38        | 99        |
| mAP @ T = 0.5          | 44.72%    |           |

As this image dataset did not mix corn and soybean plot imagery and only focused on early-season corn, it was found that monocot and dicot weeds could be labeled within the same images and still produce a significant improvement in class AP @ T = 0.5 and mAP @ T = 0.5. This also resulted in a significantly increased number of true positives and a significantly reduced number of false positives for each class. Furthermore, as all the images were acquired at 10 m AGL and were resized to a higher resolution of 800 × 800 pixels, the object detection model was able to identify the weed with greater AP and mAP.

3.6. Training Image Set 4+ (10 Meters AGL)

Training Image Set 4+ was trained on the same images collected for Training Image Set 4. The final weight file (20,000 iterations) from Training Image Set 4, when trained over 20,000 iterations on the Purdue Gilbreth supercomputer cluster, resulted in a slightly higher average loss at 0.98%. However, this final round of training produced an AP @ T = 0.5 of 65.37% and 45.13% for monocots and dicots, respectively. In this final training, AP @ T = 0.25 was measured for each class due to access to increased memory on the Gilbreth supercomputer cluster. At T = 0.25, the monocot AP improved to 91.48%, while the dicot AP increased to 86.13%. The results for Training Image Set 4 are shown in Table 7 below.

Table 7. Continued network training results for Training Image Set 4 (Set 4+).

| Value                  | Monocot   | Dicot     |
|------------------------|-----------|-----------|
| AP @ T = 0.25          | 91.48%    | 86.13%    |
| AP @ T = 0.5           | 65.37%    | 45.13%    |
| True Positives @ T = 0.5| 219       | 515       |
| False Positives @ T = 0.5| 43        | 268       |
| mAP @ T = 0.25 [a]     | 88.81%    |           |
| mAP @ T = 0.5 [b]      | 54.25%    |           |

* Mean average precision at an IoU threshold of 0.25.  
*b Mean average precision at an IoU threshold of 0.5.

It was observed that performing an additional 20,000 iterations of network training, for a total of 40,000 iterations, improved the class AP @ T = 0.5 by nearly 10%. While the number of true positives significantly increased for each class, the number of false positives also increased as more objects were detected.

Dicot class prediction average precision on a test set image from the final YOLO weights is shown in Figure 11.

The results in Figure 11 illustrate improvements in intra-row dicot weed detection but still show that very small, more soil-colored weeds still have difficulty getting detected. Network training results on Training Image Set 4 demonstrated that the addition of negative samples greatly improved the average precision of detection and reduced the number of detected false positives for each class. A summary of the results from network training on each Training Image Set is shown in Table 8 below.
Table 8. Summary of the results from the network training.

| Training Image Set | Iterations | Monocot Value | Dicot Value |
|--------------------|------------|---------------|-------------|
| 1                  | 20,000     | AP @ T = 0.5  | 9.88%       | 3.63%       |
| 2                  | 20,000     | AP @ T = 0.5  | 9.32%       | 4.2%        |
| 3                  | 20,000     | AP @ T = 0.5  | N/A%        | 10.21%      |
| 4                  | 40,000     | AP @ T = 0.5  | 58.31%      | 31.13%      |
| 4+                 | 40,000     | AP @ T = 0.25 | 65.37%      | 45.13%      |

Table 8 shows a significant improvement in class AP @ T = 0.5 from Training Image Set 3 onward. It was found that transfer learning from the 20,000-iteration weight file of Training Image Set 4 allowed for a greater increase in class AP at the T = 0.5 and T = 0.25 thresholds. It was decided that the nearly 30% increase in class AP @ T = 0.25 warranted an increase in false positives for each class. These findings are consistent with the initial YOLOv3 network training [54].

4. Discussion

Deep learning is quickly becoming popular among the research community for precision agriculture applications, including weed management. Identifying the weeds within a field can help farmers implement more efficient management practices. Image classification using DCNN has been gaining the attention of researchers for identifying weeds. However, as image classification can only identify single objects, its use on UAS acquired images at high altitudes can prove challenging to accurately identify multiple weeds within images. Nevertheless, the identification of weeds is of vital importance as locating the weeds within fields is necessary to ensure effective SSWM. Therefore, object detection was used in this study to identify and locate multiple weeds within UAS-acquired images. By identifying and locating the weeds, regions of the field consisting of weeds can be targeted by farm-
ers for herbicide application, resulting in reduced costs, reduced negative environmental impacts, and improved crop quality.

This study used deep learning-based object detection to identify and locate weeds. Although object detection may be implemented using multiple different DCNN-based algorithms, YOLOv3 was selected due to the fewer parameters needed for faster training, inference speed, and high detection scores for fast and accurate weed identification and localization. This network was particularly helpful in detecting multiple instances of weeds, especially when the neighboring emergent crops were of similar color and size. For example, in Training Image Set 4+, Figure 11 showed that the network could detect over 40 instances of dicot weeds in a single image. Moreover, YOLOv3 has fewer parameters than its predecessors, which keeps the model size at around 220 MB. Compared to other object-detection networks, such as Faster R-CNN, whose model size is over 500 MB [41], the smaller size translates into faster inference and the ability to run on resource-limited computing devices, such as Raspberry Pi or NVIDIA Nano.

In summary, this research resulted in four training image sets of early-season monocot and dicot weeds in corn and soybean fields acquired at heights of 30 m and 10 m AGL using a UAS. The impact of height on UAS imagery was evident as images acquired at lower heights resulted in more accurate results. The models trained for each image set were compared using different evaluation metrics. However, AP was used as the primary evaluation metric based on the practices present in the current literature. Through improvements in image set creation and training methods, a significant improvement was made to the object detection average precision at each successive stage. It was found that training time was significantly faster on the Gilbreth cluster than locally, using a GPU-enabled laptop. For example, a training time over 20,000 iterations took 108 h to complete on the laptop for Training Image Set 4. This same dataset with the same specifications took 62 h to complete for 20,000 epochs on the Gilbreth cluster, defined as Training Image Set 4+.

After training the YOLOv3 model for each training image set, the results were compared and reported using AP and mAP scores. The mAP scores for the training image set results are specific to the 50% and 25% IoU thresholds (T = 0.5 and T = 0.25), as denoted in each respective table. For Training Image Set 1, images were acquired at 30 m AGL. This leads to a lower spatial resolution, making it harder for the network to detect monocot and dicot weeds. This was observed when low true positives, high false positives, low AP, and low mAP were reported. Hence, a mix of images acquired at 10 m and 30 m AGL was used for Training Image Set 2 to evaluate whether adding higher resolution images acquired from 10 m AGL helps make the network more robust. As is apparent from the results, a significant increase was not observed—the mAP score decreased by ~0.5%. Training Image Set 3 was then acquired from a height of 10 m and was used exclusively to detect the dicot weeds in the corn field, leading to an increase in the AP score. However, the increase was not significant as only a single weed type was identified, which resulted in identifying all plants, and ultimately a high FP rate was observed. Having learned that images collected from 30 m AGL does not provide enough spatial information for the network to train adequately, only images collected from a height of 10 m AGL were used in the final image set, Training Image Set 4. The network was trained for another 20,000 iterations on the same image set, leading to an increased network performance on par with the literature. The final training for Training Image Set 4+ led to the highest network performance with an AP @ 0.5 of 65.37% and 45.13% and AP @ 0.25 of 91.48% and 86.13% for monocots and dicots, respectively. This resulted in a mAP @ 0.5 of 55.25% and mAP @ 0.25 of 88.81%. This was close to the expected performance of YOLOv3, where for the MS COCO dataset, a mAP @ T = 0.5 score of 57.9% was reported [46]. Furthermore, network performance increased when the images were resized to a higher 800 × 800 pixels for Training Image Sets 4 and 4+. The performance was comparable to recent weed studies, which utilized object detection for weed identification. Recently, the authors in [45] identified weeds with total mAP scores of 22.7%, 25.1%, 17.1%, and 28.9% for Faster RCNN with ResNet50, ResNet101,
InceptionV2, and InceptionResNetV2 feature extractors, respectively. Additionally, mAP @ $T = 0.5$ of 48.6%, 51.5%, 40.9%, and 55.5% were reported.

Recently, a new object-detection network belonging to the YOLO family was released. The network is called YOLOv4 [64]. YOLOv4 uses the Cross Stage Partial (CSP) Network in addition to Darknet53 for feature extraction [66]. The primary advantage of CSP is a reduction in computations by almost 20% while being more accurate on the ImageNet dataset when used as a feature extractor. The evaluation of Yolov4 for identifying and locating weeds in UAS images will be explored in future studies.

5. Conclusions

This study has demonstrated that training manually annotated and labeled monocot and dicot class image sets on the YOLOv3 object detection network can yield promising results towards automating weed detection in a row-crop field. This study has also demonstrated the feasibility of utilizing high-resolution, UAS-based RGB imagery to train an object detection network to an acceptable accuracy. A dataset of 374 annotated images, 27,825 raw RGB images, and the final training weight files from each training has been made publicly available on GitHub. It can be concluded that an annotated image set should be created for specific crops in which detection is to be performed. This was shown in Training Image Sets 3 and 4. Compared to the first two sets that mixed corn and soybean crops, the average precision improved significantly when single crop imagery was used. For example, unique image sets should be created for corn and soybeans separately. This will reduce the number of false positives from the resulting network training. It was also found that as many negative samples (non-labeled images) should be added to an annotated training set as positive samples (labeled images). The negative samples should not contain any objects that are desired for detection. By acquiring UAS imagery at the height of 10 m AGL, a significant improvement in YOLOv3 object-detection performance was observed for the creation of the training image sets for monocot and dicot weed detection, with an improved AP at both the $T = 0.5$ and $T = 0.25$ levels. After obtaining these promising results, we conclude that deep learning-based object detection is helpful for identifying weeds with field images using UAS for SSWM, especially early in the growing season.

Future research will first involve increasing the annotation and classes for object detection to reduce the false-positive rate. The trained models will then be moved towards a real-time, UAS-based weed detection system. This can be achieved by installing a microcomputer, such as a Raspberry Pi or NVIDIA Nano to the UAS, interfacing with the camera collecting the imagery, and installing a lighter memory version of YOLO, such as TinyYOLO, to the microcomputer. The final trained weights (the results of completed network training) can be loaded onto the microcomputer, and a bash script can achieve real-time video detection testing.

The future scope of this research will also include determining the effect of spatial and radiometric resolution on the accuracy of weed identification and localization [3]. A systematic study could be conducted that determines the effect of UAS flight height on the accuracy of weed detection. It would help develop a recommendation on the optimal flight height for weed detection in practical applications. Moreover, a study could be conducted to establish the generalization accuracy of the YOLO network for weed detection over changing weather conditions in the field, as it would reduce the need for radiometric calibration [67].

Author Contributions: Conceptualization, D.S. and A.E.; methodology, A.E.; validation, A.E., A.A., and V.A.; resources, D.S. and A.E.; writing—original draft preparation, A.E.; writing—review and editing, D.S., A.A. and V.A.; supervision, D.S.; project administration, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2018 Indiana Corn Marketing Council Gary Lamie Graduate Assistantship. Fund number 650100000, Purdue University Ross Fellowship, grant number 401423, Wabash Heartland Innovation Network (WHIN) grant number 18024S89 and USDA-NIFA Hatch Fund#1012501.
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data used in this study has been added to a publicly available Github repository aetienne93/UAS-based-Weed-Detection. This repository can be found at https://github.com/aetienne93/UAS-based-Weed-Detection.

Acknowledgments: We would like to acknowledge Bryan Young and his graduate students in Botany and Plant Pathology at Purdue University for providing the research plots used for this study. Their agronomic knowledge and help in setting up the experiment trials were vital in the success of this study. We would also like to thank Ben Hancock for his technical expertise in helping set up and debug the YOLO deep learning network used in this study.

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

References
1. Gao, J.; Nuyttens, D.; Lootens, P.; He, Y.; Pieters, J.G. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. Biosyst. Eng. 2018, 170, 39–50. [CrossRef]
2. Soltani, N.; Dille, J.A.; Burke, I.C.; Everman, W.J.; VanGessel, M.J.; Davis, V.M.; Sikkema, P.H. Potential corn yield losses from weeds in North America. Weed Technol. 2016, 32, 342–346. [CrossRef]
3. Lingenfelter, D.D.; Hartwig, N.L. Introduction to Weeds and Herbicides. Available online: https://extension.psu.edu/introduction-to-weeds-and-herbicides (accessed on 6 October 2021).
4. Pérez-Ortiz, M.; Peña, J.M.; Gutiérrez, P.A.; Torres-Sánchez, J.; Hervás-Martínez, C.; López-Granados, F. A Semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Appl. Soft Comput. J. 2015, 37, 533–544. [CrossRef]
5. Zhang, X.; Wang, S.; Liu, J.; Tao, C. Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC Med. Inform. Decis. Mak. 2018, 18, 69–76. [CrossRef]
6. Yan, K.; Wang, X.; Lu, L.; Summers, R.M. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging 2018, 5, 36501. [CrossRef] [PubMed]
7. Maglogiannis, I.; Doukas, C.N. Overview of advanced computer vision systems for skin lesions characterization. IEEE Trans. Inf. Technol. Biomed. 2009, 13, 721–733. [CrossRef] [PubMed]
8. Hamed, G.; Marey, M.A.E.R.; Amin, S.E.S.; Tolba, M.F. Deep learning in breast cancer detection and classification. Adv. Intell. Syst. Comput. 2020, 1153, 322–333. [CrossRef]
9. Agarwal, N.; Chiang, C.W.; Sharma, A. A study on computer vision techniques for self-driving cars. Lect. Notes Electr. Eng. 2018, 542, 629–634. [CrossRef]
10. Khan, M.; Hassan, S.; Ahmed, S.I.; Iqbal, J. Stereovision-based real-time obstacle detection scheme for unmanned ground vehicle with steering wheel drive mechanism. In Proceedings of the 2017 International Conference on Communication, Computing and Digital Systems, C-CODE, Islamabad, Pakistan, 9 March 2017; pp. 380–385. [CrossRef]
11. Al-Kaff, A.; Martín, D.; García, F.; de la Escalera, A.; María Armingol, J. Survey of computer vision algorithms and applications for unmanned aerial vehicles. Expert Syst. Appl. 2018, 92, 447–463. [CrossRef]
12. Kaur, P.; Krishan, K.; Sharma, S.K.; Kanchan, T. Facial-recognition algorithms: A literature review. Med. Sci. Law 2020, 60, 131–139. [CrossRef]
13. Zhao, C.; Chen, B. Real-time pedestrian detection based on improved YOLO model. In Proceedings of the 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC, Hangzhou, China, 25 August 2019; Volume 2, pp. 25–28. [CrossRef]
14. Heng, L.; Meier, L.; Tanskanen, P.; Fraundorfer, F.; Pollefeys, M. Autonomous obstacle avoidance and maneuvering on a vision-guided mav using on-board processing. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 12 May 2011; pp. 2472–2477. [CrossRef]
15. Ahmad, A.; Saraswat, D.; Aggarwal, V.; Etienne, A.; Hancock, B. Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. Comput. Electron. Agric. 2021, 184, 106811. [CrossRef]
16. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. Front. Plant Sci. 2016, 7, 1419. [CrossRef]
17. Liu, L.; Wang, R.; Xie, C.; Yang, P.; Wang, F.; Sudirman, S.; Liu, W. PestNet: An end-to-end deep learning approach for large-scale multi-class pest detection and classification. IEEE Access 2019, 7, 45301–45312. [CrossRef]
18. Wulandhari, L.A.; Gunawan, A.A.S.; Qurania, A.; Harsani, P.; Tarawan, T.F.; Hermawan, R.F. Plant nutrient deficiency detection using deep convolutional neural network. ICIC Express Lett. 2019, 13, 971–977.
19. Luo, L.; Liu, W.; Lu, Q.; Wang, J.; Wen, W.; Yan, D.; Tang, Y.; Gasteratos, A. Grape berry detection and size measurement based on edge image processing and geometric morphology. Machines 2021, 9, 233. [CrossRef]
20. Shrestha, D.; Steward, B.; Kaspar, T.; Robert, P. Others determination of early stage corn plant height using stereo vision. In Proceedings of the International Conference on Precision Agriculture Abstracts and Proceedings, Minneapolis, MN, USA, 15 July 2002.

21. 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]

22. Bhargava, A.; Bansal, A. Fruits and vegetables quality evaluation using computer vision: A review. *J. King Saud Univ. Comput. Inf. Sci.* **2021**, *33*, 243–257. [CrossRef]

23. Mohidem, N.A.; Norasma, N.; Ya, C.; Shukor Juraimi, A.; Fazilah, W.; Ilahi, F.; Huzuafah, M.; Roslim, M.; Sulaiman, N.; Saberioon, M.; et al. How can unmanned aerial vehicles be used for detecting weeds in agricultural fields? *Agriculture* **2021**, *11*, 1004. [CrossRef]

24. Hassanein, M.; El-Sheimy, N. An efficient weed detection procedure using low-cost uav imagery system for precision agriculture applications. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, XLII-1, 181–187. [CrossRef]

25. Burgos-Artizzu, 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]

26. Swain, K.C.; Nørremark, M.; Jørgensen, R.N.; Midtbø, H.S.; Green, O. Weed identification using an Automated Active Shape Matching (AASM) technique. *Biosyst. Eng.* **2011**, *110*, 450–457. [CrossRef]

27. Dadashzadeh, M.; Abbaspour-Gilandeh, Y.; Mesri-Gundoshmian, T.; Sabzi, S.; Hernández-Hernández, J.L.; Hernández-Hernández, M.; Ignacio Arribas, J. Weed classification for site-specific weed management using an automated stereo computer-vision machine-learning system in rice fields. *Plants* **2020**, *9*, 599. [CrossRef] [PubMed]

28. Bakshhipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [CrossRef]

29. 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]

30. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodriguez, L.; Co, A.P. A deep learning approach for weed detection in lettuce crops using multispectral imagery. *AgriEngineering* **2020**, *2*, 471–488. [CrossRef]

31. Olsen, A.; Konovalov, D.A.; Phillipa, B.; Ridd, P.; Wood, J.C.; Johns, J.; Banks, W.; Giglenti, B.; Kenny, O.; Whinney, J.; et al. DeepWeeds: A multiclass weed species image dataset for deep learning. *Sci. Rep.* **2019**, *9*, 2058. [CrossRef]

32. Sa, I.; Popovic, M.; Khanna, R.; Chen, Z.; Lottes, P.; Liebisch, F.; Nieto, J.; Stachniss, C.; Walter, A.; Siegwart, R. WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sens.* **2018**, *10*, 1423. [CrossRef]

33. Perez-Ortiz, M.; Gutierrez, P.A.; Pena, J.M.; Torres-Sanchez, J.; Lopez-Granados, F.; Hervas-Martinez, C. Machine learning paradigms for weed mapping via unmanned aerial vehicles. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence, Athens, Greece, 6–9 December 2016. [CrossRef]

34. Zhang, R.; Wang, C.; Hu, X.; Liu, Y.; Chen, S. Weed location and recognition based on UAV imaging and deep learning. *Int. J. Precis. Agric. Aviat.* **2020**, *3*, 23–29. [CrossRef]

35. Milioto, A.; Lottes, P.; Stachniss, C. Photogrammetry. In *Remote Sensing*. Springer, Cham, Switzerland, 2017; pp. 41–48. [CrossRef]

36. Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Madrigal, V.P.; Mallinis, G.; Dor, E.B.; Helman, D.; Estes, L.; Ciriaolo, G.; et al. On the use of unmanned aerial systems for environmental monitoring. *Remote Sens.* **2018**, *10*, 641. [CrossRef]

37. dos Santos Ferreira, A.; Matte Freitas, D.; Gonçalves da Silva, G.; Pistori, H.; Theophilo Folhes, M. Weed detection in soybean crops using convnets. *Comput. Electron. Agric.* **2017**, *143*, 314–324. [CrossRef]

38. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Zhang, L. A fully convolutional network for weed mapping of Unmanned Aerial Vehicle (UAV) imagery. *PLoS ONE* **2018**, *13*, e0196302. [CrossRef]

39. Sivakumar, A.N.V.; Li, J.; Scott, S.; Psota, E.; Jhala, A.J.; Luck, J.D.; Shi, Y. Comparison of object detection and patch-based classification deep learning models on mid-to-late-season weed detection in UAV imagery. *Remote Sens.* **2020**, *12*, 2136. [CrossRef]

40. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]

41. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]

42. Liu, W.; Angleuel, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Amsterdam, The Netherlands, 11–14 October 2016.

43. 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]

44. Espinoza, M.; Le, C.Z.; Raheja, A.; Bhandari, S. Weed identification and removal using machine learning techniques and unmanned ground vehicles. In *Proceedings of the SPIE 9866, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, Baltimore, MD, USA*, 17 May 2016.
45. Le, V.N.T.; Truong, G.; Alameh, K. Detecting weeds from crops under complex field environments based on faster RCNN. In Proceedings of the ICCE 2020—2020 IEEE 8th International Conference on Communications and Electronics, Phu Quoc Island, Vietnam, 13 January 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021; pp. 350–355.

46. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

47. Li, M.; Zhang, Z.; Lei, L.; Wang, X.; Guo, X. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. Sensors 2020, 20, 4938. [CrossRef] [PubMed]

48. Sharpe, S.M.; Schumann, A.W.; Yu, J.; Boyd, N.S. Vegetation detection and discrimination within vegetable plasticulture row-middle using a convolutional neural network. Precis. Agric. 2020, 21, 264–277. [CrossRef]

49. Partel, V.; Charan Kakarla, S.; Ampatzidis, Y. Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. Comput. Electron. Agric. 2019, 157, 339–350. [CrossRef]

50. Ham, S.; Oh, Y.; Choi, K.; Lee, I. Semantic segmentation and unregistered building detection from uav images using a deconvolutional network. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2018, XLII-2, 419–424. [CrossRef]

51. Etienne, A.; Saraswat, D. Machine learning approaches to automate weed detection by UAV based sensors. In Proceedings of the SPIE 11008, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, Baltimore, MD, USA, 14 May 2019. [CrossRef]

52. Novus Ag Growth Stages of Corn & Soybeans. Available online: https://www.novusag.com/2019/06/growth-stages-of-corn-soybeans/ (accessed on 17 October 2021).

53. Pioneer Staging Corn Growth. Available online: https://www.pioneer.com/us/agronomy/staging_corn_growth.html (accessed on 23 August 2021).

54. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. arXiv 2018, arXiv:1804.02767.

55. Kathuria, A. What’s new in YOLO v3? A Review of the YOLO v3 Object Detection. Available online: https://towardsdatascience.com/yolo-v3-object-detection-53b7d3bf66 (accessed on 29 November 2021).

56. Lin, T. LabellImg: Graphical Image Annotation Tool. Available online: https://github.com/tzutalin/labelImg (accessed on 29 November 2021).

57. Margaritoff, M. Drones in Agriculture: How UAVs Make Farming More Efficient. Available online: https://www.thedrive.com/tech/18456/drones-in-agriculture-how-uavs-make-farming-more-efficient (accessed on 8 August 2021).

58. Thomison, P. How to Identify Emergence Issues in Corn. Available online: https://cfaes.osu.edu/news/articles/how-identify-emergence-issues-in-corn (accessed on 12 August 2021).

59. Bochkovskiy, A. How to Train (to Detect Your Custom Objects). Available online: https://github.com/AlexeyAB/darknet#how-to-train-to-detect-your-custom-objects (accessed on 15 May 2019).

60. Information Technology at Purdue University Overview of Gilbreth Community Cluster. Available online: https://www.rcac.purdue.edu/compute/gilbreth (accessed on 10 March 2019).

61. National Pesticide Information Center Dicamba General Fact Sheet. Available online: http://npic.orst.edu/factsheets/dicamba_gen.pdf (accessed on 28 February 2021).

62. Aidouni, M. Evaluating Object Detection Models: Guide to Performance Metrics. Available online: https://manalelaidouni.github.io/Evaluating-Object-Detection-Models-Guide-to-Performance-Metrics.html (accessed on 12 November 2021).

63. Hui, J. Real-Time Object Detection with YOLO, YOLOv2 and Now YOLOv3. Available online: https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-2881b93e2088 (accessed on 12 October 2021).

64. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. arXiv 2020, arXiv:2004.10934.

65. Bochkovskiy, A. Yolo Mark Labelling Tool. Available online: https://github.com/AlexeyAB/Yolo_mark (accessed on 18 November 2018).

66. Wang, C.Y.; Mark Liao, H.Y.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A New backbone that can enhance learning capability of CNN. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 16–18 June 2020; pp. 1571–1580. [CrossRef]

67. Esposito, M.; Crimaldi, M.; Cirillo, V.; Sarghini, F.; Maggio, A. Drone and sensor technology for sustainable weed management: A review. Chem. Biol. Technol. Agric. 2021, 8, 18. [CrossRef]