Supplementary Material

1 Data Preparation

Data preparation is the first step in the general pipeline for YOLO object detection training and, in this case, for weed detection exploiting UAV data specifically. Initially, a simple UAV capturing RGB images in the CP dataset has only image-level annotation. These annotations do not follow the format requirements set out by YOLO detector implementation [1], as YOLO detectors require bounding box labels on each image. Therefore, we labeled each image for weed recognition.

2 Annotation and Labeling Using Roboflow

Annotation is the process of physically drawing bounding boxes around a selected object (i.e., a weed). On the other hand, a label refers to any given name of the class allocated to the annotations that are stored as metadata for that class file. The manual process of image labeling and annotation is a tedious but critical step in supervised modeling. In order to train the model to identify every instance of an object in an image, it must have a method to grasp the location and position of an object in relation to other surrounding objects in an image where detection is possible. To properly annotate and label high-quality UAV-acquired images, an application with the ability to quickly label hundreds of weed objects in an image was needed. A labeling application called Roboflow was selected because it offers an appropriate labeling speed as well as convenient zoom and pan functions [2]. Specifically, Roboflow transforms each bounding box into five-digit patterns. The first digit represents the label of the class. The second and third digits refer to the x and y coordinates of the center position of the bounding box. The

fourth digit indicates the width, and the fifth digit represents the height of the bounding box.

Figure S1 provides an illustration of a bounding box annotation and labeling using the Roboflow application. The captured images were annotated by skilled operators who used the online tool "Roboflow" [2] to draw bounding boxes for each weed object in the images. This process was very time-consuming, as a couple of man-months of effort was needed to complete the task because each image was labeled manually. The weed object in the images was annotated by their ordinary names, such as "weed". The annotated dataset was saved as a .yaml file format, and weed experts also double-checked the results by visualizing the annotations to confirm the label and assure the dataset's quality. To our knowledge, this is the first chicory weed dataset related to chicory crop production.

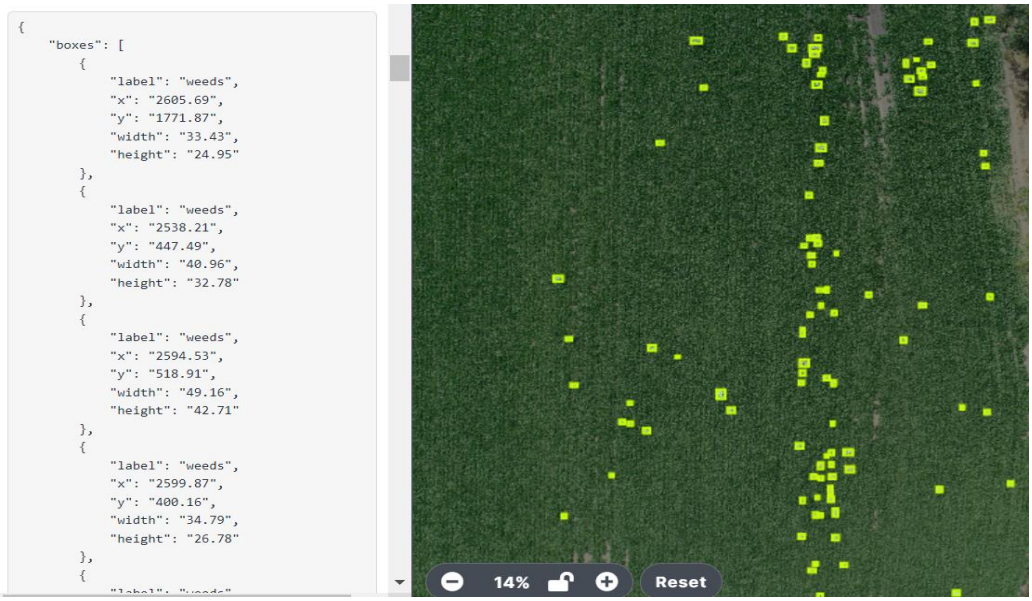

Figure S1: Overview of Roboflow manual labeling process. The left side of the figure shows a sample of the manual annotation process in the image with dimensions of 5472 × 3648 pixels, belonging to the Chicory Plant dataset proposed in this work. Yellow boxes are annotation boxes. The right side of the figure shows labels and information about the annotation boxes, such as the X and Y coordinates and the height and width.

Figure S2 shows one example of a 5472 × 3648 image from those used to create the dataset. By using a custom-written Python script, the truth boxes were created from the large image, and then it was cut into smaller patches to be processed by the neural network. Only the patches containing weeds were extracted and used to generate the CP dataset.

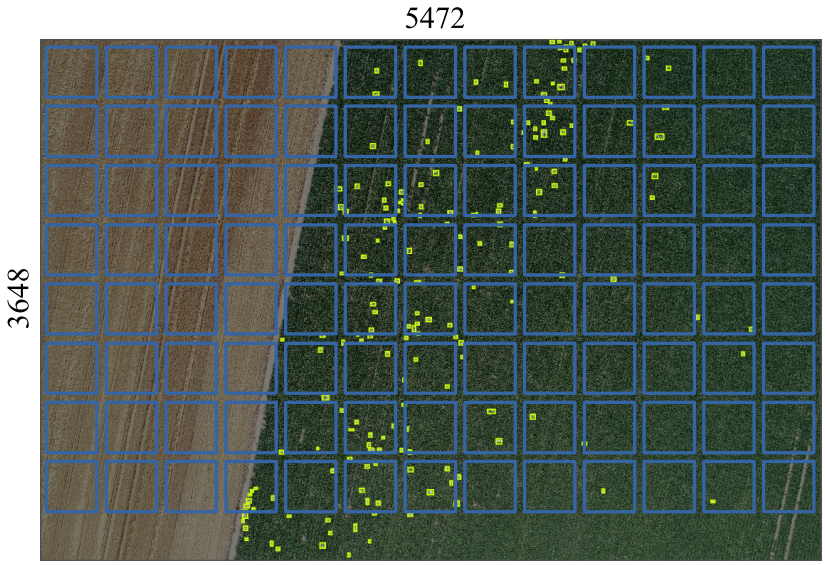

Figure S2: An example of an RGB image with dimensions of 5472 × 3648 pixels, belonging to the Chicory Plant dataset proposed in this work. The yellow boxes identify weeds tagged by an expert. The 13 × 8 cells in blue represent the mosaic used to extract patches 416 × 416 in size to be fed into the neural model. Only patches containing at least one bounding box are used by the neural model.

3 Data Augmentation

As we mentioned earlier, agricultural datasets are very scarce, it is difficult to generate new datasets, and annotations and labeling take a lot of time and human effort. However, for the efficient performance of DL detectors, a bigger and more suitable dataset is always considered a good option. For this purpose, we also used data augmentation in the proposed CP dataset. Data augmentation is a group of techniques that increase the quality and the

size of ML training datasets so that better DL approaches can be trained for the experiment process. Image augmentation entails modifying the training images to produce a synthetically larger dataset compared with the original dataset, and it may enhance the downstream efficiency of the model. Data augmentations involve flipping, adding noise, cropping, occluding portions of the image, rotating, and many other transformations [3]. We also applied Flip (horizontal or vertical), 90° Rotate (clockwise or counterclockwise), and mosaic data augmentation techniques in the CP dataset. Flip augmentation means turning the image horizontally or vertically according to the object in the image. The horizontal flipping of the columns and rows of pixels in an image and the process of vertically flipping all of the columns and rows of pixels in an image are considered horizontal and vertical flip augmentation, respectively. Mosaic augmentation [4] combines the four random images together and forms one image with a certain ratio. It combines those classes in the dataset that might not have been together in the training set. It is a unique augmentation that takes the original image and three additional random images whenever an image is loaded for training in the model. A 90° rotation means rotating the images in the dataset at a 90 degree angle using clockwise and counterclockwise directions.

We generated two versions of the CP dataset; the normal CP dataset and the augmented CP dataset. The final version of the normal CP dataset contained simple annotated chicory crop images without any augmentation techniques. On the one hand, for the augmented CP dataset, we used the Roboflow application and applied the above-mentioned image augmentation techniques to the images of the normal CP dataset. The LB dataset contained two classes named sugar beet and weed, while the CP dataset contained only one weed class for detection. Both datasets were randomly split into test, validation, and training sets with 20%, 10%, and 70% of all the images, respectively. For a fair comparison, we also split the data-augmented CP dataset into training, validation, and test sets with the same 70%, 20%, and 10% ratios, respectively. The images and annotations were cropped into two distinct sizes: 416 × 416 pixels and 640 × 640 pixels while maintaining the width/height ratio in order to test the techniques under various image resolutions.

References

[1] Chien, W. YOLOv7 repositry with all instruction. https://github.com/WongKinYiu/yolov7, 2022.

[2] Dwyer, J. Quickly Label Training Data and Export To Any Format. https://roboflow.com/annotate, 2020.

[3] Image augmentation in roboflow. https://docs.roboflow.com/imagetransformations/image-augmentation, December 2022.

[4] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934* **2020**.