

# WeedVision: Multi-Stage Growth and Classification of Weeds using DETR and RetinaNet for Precision Agriculture

Taminul Islam

*School of Computing*

*Southern Illinois University*

Carbondale, USA

taminul.islam@siu.edu

Toqi Tahamid Sarker

*School of Computing*

*Southern Illinois University*

Carbondale, USA

toqitahamid.sarker@siu.edu

Khaled R Ahmed

*School of Computing*

*Southern Illinois University*

Carbondale, USA

khaled.ahmed@siu.edu

Cristiana Bernardi Rankrake

*Department of Plant Soils and Agricultural Systems*

*Southern Illinois University*

Carbondale, USA

cris.rankrake@siu.edu

Karla Gage

*School of Agricultural Sciences / School of Biological Sciences*

*Southern Illinois University*

Carbondale, USA

kgage@siu.edu

**Abstract**—Weed management remains a critical challenge in agriculture, where weeds compete with crops for essential resources, leading to significant yield losses. Accurate detection of weeds at various growth stages is crucial for effective management yet challenging for farmers, as it requires identifying different species at multiple growth phases. This research addresses these challenges by utilizing advanced object detection models—specifically, the Detection Transformer (DETR) with a ResNet-50 backbone and RetinaNet with a ResNeXt-101 backbone—to identify and classify 16 weed species of economic concern across 174 classes, spanning their 11-week growth stages from seedling to maturity. A robust dataset comprising 203,567 images was developed, meticulously labeled by species and growth stage. The models were rigorously trained and evaluated, with RetinaNet demonstrating superior performance, achieving a mean Average Precision (mAP) of 0.907 on the training set and 0.904 on the test set, compared to DETR’s mAP of 0.854 and 0.840, respectively. RetinaNet also outperformed DETR in recall and inference speed of 7.28 FPS, making it more suitable for real-time applications. Both models showed improved accuracy as plants matured. This research provides crucial insights for developing precise, sustainable, and automated weed management strategies, paving the way for real-time species-specific detection systems and advancing AI-assisted agriculture through continued innovation in model development and early detection accuracy.

**Index Terms**—Object Detection, Weed Management, DETR, Weed Growth Classification, Weed Detection

## I. INTRODUCTION

In the vast agricultural landscape of the USA, weed management remains a critical challenge for farmers and agronomists. The diverse climates and fertile soils ideal for crop production also create favorable conditions for a wide variety of weed species [1]. These unwanted plants compete with crops for essential resources such as water, nutrients, and sunlight, potentially leading to significant yield losses and economic

setbacks for farmers. Traditional weed control methods often rely on broad-spectrum herbicides or mechanically or labor-intensive removal [2]. However, these approaches can be environmentally harmful, economically inefficient, and increasingly ineffective due to the development of herbicide-resistant weed populations [3]. As such, there is a growing need for more precise, sustainable, and automated weed management strategies.

Recent advancements in computer vision and deep learning have shown promise in addressing this agricultural challenge. Object detection and classification techniques applied to weed identification offer the potential for highly accurate, real-time weed management solutions [4]. However, several research gaps persist in this domain, such as (a) limited datasets: most existing studies rely on small datasets or images captured at specific growth stages, failing to capture the dynamic nature of weed development, and (b) lack of diversity: many datasets focus on a limited number of weed species, not reflecting the full range of weeds farmers encounter in real-world scenarios.

The scope of this work addresses these gaps by focusing on 16 weed species of greatest economic concern found commonly across multiple geographies in USA agriculture, tracking their growth from the seedling stage through 11 weeks of development. We created a robust, diverse dataset and implemented advanced object detection models to improve the accuracy and efficiency of weed identification and classification.

Our research makes several key contributions to the field:

- Creation of a unique dataset comprising 203,567 images, capturing the full growth cycle of 16 of the most common and troublesome weed species in USA agriculture.
- Meticulous labeling of the dataset, categorized by species and growth stage (week-wise), providing a comprehen-

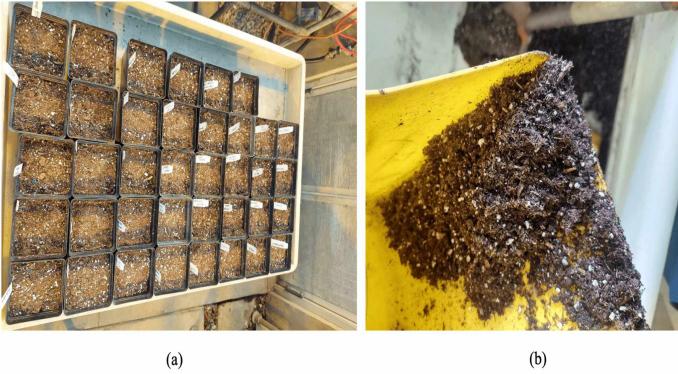


Fig. 1. Soil preparation and labeling for planting weed seeds in pots inside the greenhouse. (a) shows the prepared pots with soil and pot stakes, (b) displays the close-up of the soil mix used for planting.

sive resource for weed identification research.

- Implementing the Detection Transformer (DETR) [5] and RetinaNet [6], adapting these state-of-the-art object detection architectures for weed identification.
- Comprehensive comparison of model results, culminating in evidence-based recommendations for farmers on the most effective model for weed detection in real-world scenarios.

This research utilizes DETR and RetinaNet due to their state-of-the-art performance in object detection tasks. DETR introduces a state-of-the-art transformer-based approach, offering end-to-end object detection with potential benefits in handling complex scenes and object relationships. RetinaNet, known for its efficiency and accuracy, employs a focal loss function to address class imbalance issues common in detection tasks. By implementing and comparing these two advanced models, the study aims to evaluate their effectiveness in the specific context of weed detection and classification. This research not only contributes to the growing body of work on AI-assisted agriculture but also provides practical insights for farmers and beyond. By developing more accurate and efficient weed detection systems, we pave the way for precision agriculture techniques that can significantly reduce herbicide use, lower production costs, and minimize environmental impact.

In the following sections, this paper presents related work, followed by a comprehensive outline of the data collection and pre-processing techniques employed. The methodology section describes the steps taken in this research. Subsequently, the models section introduces the implementation and evaluation of DETR and RetinaNet for detecting and classifying 16 weed species at various growth stages. The results section showcases the performance metrics of these models. In conclusion, it summarizes these research findings for the 16 growth stage detection and classification with actionable recommendations for farmers based on the study's outcomes.

## II. RELATED WORK

Recent advancements in deep learning and computer vision have revolutionized weed detection and classification

in precision agriculture. Researchers have developed various approaches to address the challenges associated with accurate and efficient weed identification in diverse crop environments. Object detection models have shown promising results in weed identification. Hasan et al. (2024) [7] created a dataset of 5,997 images featuring corn and four weed species, demonstrating that YOLOv7 achieved the highest mean average precision (mAP) of 88.50%, further improved to 89.93% with data augmentation. Wang et al. (2024) [8] proposed the CSCW-YOLOv7 model for weed detection in wheat fields, achieving superior precision (97.7%), recall (98%), and mAP (94.4%). Transfer learning has proven effective for weed species detection. Shackleton et al. (2024) [9] evaluated seven pre-trained CNN models for rangeland weed detection, with EfficientNetV2B1 achieving the highest accuracy of 94.2%. Ahmad et al. (2021) [10] employed various models for image classification and object detection in corn and soybean systems, with VGG16 achieving 98.9% accuracy and YOLOv3 reaching 54.3% mAP. Traditional machine learning algorithms have also been applied to weed detection. Islam et al. (2021) [11] compared Random Forest, Support Vector Machine, and k-Nearest Neighbors for weed detection in chilli pepper fields, with RF and SVM achieving 96% and 94% accuracy, respectively. Semantic segmentation approaches have shown promise. Khan et al. (2020) [12] introduced CED-Net, outperforming traditional models like U-Net and SegNet. Arun et al. (2020) [13] developed a Reduced U-Net architecture, achieving 95.34% segmentation accuracy on the CWFID dataset. Autonomous weeding applications have benefited from deep learning. Adhikari et al. (2019) [14] proposed ESNet for autonomous weeding in rice fields, utilizing semantic graphics for data annotation. Teimouri et al. (2018) [15] developed a method to classify weeds into nine growth stages, achieving a maximum accuracy of 78% for *Polygonum* spp. Ensemble learning frameworks have been introduced to improve detection under varied field conditions. Asad et al. (2023) [16] proposed an approach using diverse models in a teacher-student configuration, significantly outperforming single semantic segmentation models. Moldvai et al. (2024) [17] explored weed detection using multiple features and classifiers, achieving a 94.56% recall rate with limited data.

Despite these advancements, several limitations persist in existing research. These include the need for larger and more diverse datasets [18] [19], class imbalance issues [7], and computational complexity [16]. Most studies focus on a limited number of weed species [17] [10] and growth stages, which may not fully represent real-world agricultural settings. Our research addresses these limitations by creating a comprehensive dataset of 203,567 images featuring 16 common and troublesome weeds in USA agriculture, capturing their full 11-week growth cycle. We implement and adapt state-of-the-art object detection models, DETR and RetinaNet, for weed growth identification. Through a comprehensive comparison of model results, we provide practical insights for weed management in precision agriculture. This work distinguishes itself by



Fig. 2. Greenhouse environment with lighting, temperature, and watering setup.

focusing on a large-scale, diverse dataset, considering multiple weed species and growth stages, and offering practical recommendations for farmers on the most effective model for weed detection in real-world scenarios. By addressing the limitations of previous studies, our research contributes significantly to the field of weed identification and management in precision agriculture.

### III. DATA DESCRIPTION AND PREPROCECCING

In this research, we conducted a study on 16 weed species at the SIU Horticulture Research Center greenhouse. We began by preparing soil for seed planting, as shown in Figure 1(b). Potting soil (Pro-Mix ® BX) was placed into 32 square pots (10.7 cm x 10.7 cm x 9 cm), each labeled by species with white pot stakes. Two seeds from each species were planted per pot. Environmental conditions in the greenhouse, including temperature and lighting, were carefully controlled. Plants were watered as needed and fertilized with all-purpose 20-20-20 nutrient solution every 3 days. Figure 2 provides an overview of the greenhouse environment. We monitored the growth stages of each plant on a weekly basis, capturing images from the first week until week 11. Image capture ceased when the weeds entered their flowering stage, which marked the final growth phase in our study. We have captured our images by using an iPhone 15 Pro Max. Table I provides a comprehensive overview of our study, detailing the weed species codes, their corresponding scientific and common names, and the number of frames captured for each species on a weekly basis.

Among the 16 species of weeds studied, SORHA did not emerge in weeks 1 and 2. Consequently, the research encompassed a total of 174 classes. The full dataset initially comprised 2,494,476 frames. After a thorough review process to remove substandard images, 203,567 images were ultimately selected for training. Figure 3 presents sample images of four weed species at different growth stages. For ABUTH, images

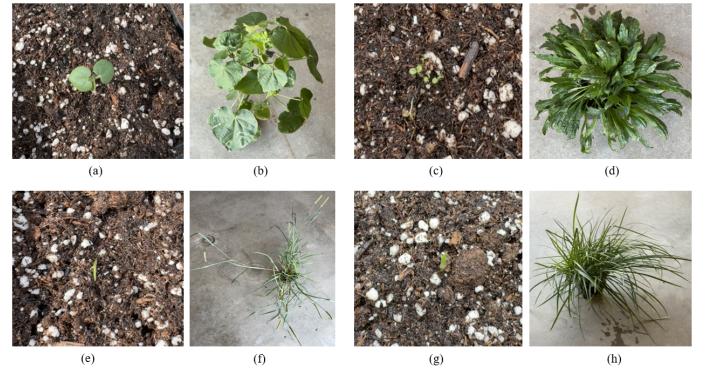


Fig. 3. Growth stages example of four weed species. (a,b) ABUTH in week 1 and 11; (c,d) ERICA in week 1 and 11; (e,f) SETFA in week 1 and 11; (g,h) CYPES in week 1 and 11. Images show progression from seedling emergence to mature plants across different species.

from week 1 (a) and week 11 (b) are shown. Similarly, ERICA is represented by its week 1 (c) and week 11 (d) images. SETFA is depicted in its first week (e) and eleventh week (f) of growth. Lastly, CYPES is illustrated in its initial (g) and final (h) weeks of the study period. Notably, while several species produced flowers in their final growth stages, others did not, reflecting natural growth processes and photoperiod sensitivities.

#### A. Data Preprocessing and Augmentation

Our preprocessing pipeline begins with image normalization, a fundamental step that standardizes the input data. Each image is meticulously scaled to a 0-1 range by dividing all pixel values by 255.0. This normalization process is crucial as it ensures consistency across the dataset and aligns with the input requirements of neural networks, facilitating more efficient and effective training [23]. Following normalization, we perform a color space conversion, transforming the images from the standard RGB (Red, Green, Blue) color space to the HSV (Hue, Saturation, Value) color space. The HSV color space allows us to more precisely isolate plant areas from the background, enhancing the accuracy of subsequent processing steps.

The next step in our pipeline is green area detection. We employ carefully calibrated thresholds for the HSV channels to create a mask that highlights potential plant regions. Specifically, we use hue values ranging from 25/360 to 160/360, a minimum saturation value of 0.20. These thresholds have been empirically determined to effectively isolate green regions corresponding to plant matter while minimizing false positives from non-plant green objects. We apply morphological operations [24] to refine the green mask and improve the continuity of detected plant areas. The refined green areas are then subjected to connected component analysis, which identifies and labels distinct regions within the image. This step is crucial for differentiating individual plants or plant clusters, allowing for more precise analysis and annotation. Fig 4 shows the process of the data augmentation.

TABLE I  
OVERVIEW OF WEED SPECIES OF ECONOMIC CONCERN, CORRESPONDING CODES, AND WEEKLY FRAME COUNTS CAPTURED FOR EACH SPECIES ACROSS 11 WEEKS IN THE GREENHOUSE

Species Code [20]	Scientific Name [21]	Common Name [22]	Family	Total Frames	Number of frames/week										
					W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_10	W_11
ABUTH	<i>Abutilon theophrasti</i> Medik.	Velvetleaf	Malvaceae	14754	1084	2451	1212	1819	1414	981	677	1164	1084	1500	1368
AMAPA	<i>Amaranthus palmeri</i> S. Watson.	Palmer Amaranth	Amaranthaceae	17525	1441	1408	2110	2014	2441	1290	923	1478	1393	1667	1360
AMARE	<i>Amaranthus retroflexus</i> L.	Redroot Pigweed	Amaranthaceae	15380	1017	1363	2110	1923	1884	1150	736	1237	1082	1596	1282
AMATU	<i>Amaranthus tuberculatus</i> (Moq.) Sauer.	Water Hemp	Amaranthaceae	14852	1325	1459	1565	1664	1942	837	730	969	1638	1573	1150
AMBEL	<i>Ambrosia artemisiifolia</i> L.	Common Ragweed	Asteraceae	17427	1022	2215	1846	1739	2162	1093	1066	1432	1092	2045	1715
CHEAL	<i>Chenopodium album</i> L.	Common Lambsquarters	Chenopodiaceae	8015	1108	954	1416	661	1056	305	418	641	453	429	574
CYPES	<i>Cyperus esculentus</i> L.	Yellow Nutsedge	Cyperaceae	14275	909	1512	1032	1499	2273	978	1224	1391	1182	1170	1105
DIGSA	<i>Digitaria sanguinalis</i> (L.) Scop.	Large Crabgrass	Poaceae	16962	732	1312	2411	2596	1649	1335	1166	1261	1120	1628	1692
ECHCG	<i>Echinocloa crus-galli</i> (L.) P. Beauv.	Barnyard Grass	Poaceae	16564	1349	2067	2029	1426	2221	1240	929	1280	1371	1332	1320
ERICA	<i>Erigeron canadensis</i> L.	Horse Weed	Asteraceae	15134	930	2183	1691	1542	2715	1189	609	742	915	1217	1401
PANDI	<i>Panicum dichotomiflorum</i> Michx.	Full Panicum	Poaceae	15182	1198	1400	2143	1296	1979	952	887	1350	1425	1034	1518
SETFA	<i>Setaria faberii</i> Herrm.	Gaint Foxtail	Poaceae	14655	1614	1195	2083	1348	1944	1091	715	1466	843	1342	994
SETPU	<i>Setaria pumila</i> (Poir.) Roem.	Yellow Foxtail	Poaceae	15211	887	1390	1732	1654	2040	1093	747	1361	1325	1348	1634
SIDSP	<i>Sida spinosa</i> L.	Prickly Sida	Malvaceae	14452	1035	1782	1583	1259	2142	1373	804	1059	1186	1303	926
SORHA	<i>Sorghum halepense</i> (L.) Pers.	Johnson Grass	Poaceae	10958	0	0	1444	1268	1395	945	749	1215	1328	1116	1498
SORVU	<i>Sorghum bicolor</i> (L.) Moench.	Shatter Cane	Poaceae	9573	945	1340	1959	832	1065	525	279	748	714	592	574

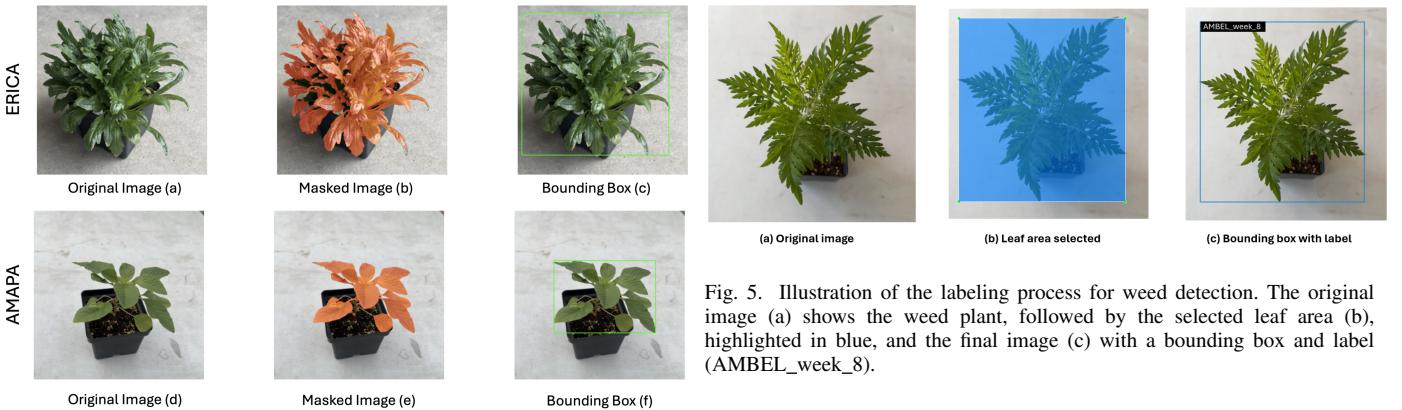


Fig. 4. Data Augmentation process with original image, masked image, and bounding box, respectively, for ERICA (a,b,c) and AMAPA (d,e,f).

### B. Data Labeling

Our labeling process creates comprehensive annotations for detected plants, including bounding box coordinates and detailed Pascal VOC XML annotations. We use Python libraries like Pillow [25], NumPy, and scikit-image for image processing. To ensure accuracy, we implemented a rigorous quality control process, manually refining annotations using LabelImg software [26] when necessary. Our labeling convention includes both species code and week number, enhancing the dataset's utility for tracking plant development and species-specific analysis. This meticulous approach results in a high-quality dataset with precise annotations and consistent formatting, suitable for various plant analysis tasks and growth stage tracking.

Figure 5 illustrates this process, presenting a side-by-side comparison of an original image and its corresponding labeled version, which we refer to as the ground truth.

### IV. METHODOLOGY

After annotating the dataset, we split the dataset into training, validation, and test sets. We used 184,719 images (~80%) to train our object detection models and 23,090 images (~10%) to validate the model during training time. The rest of the 23,090 images (~10%) are held out to test

the trained model's performance. In this study, we employed two advanced deep-learning models for weed detection and classification: RetinaNet with a ResNeXt-101 backbone and Detection Transformer (DETR) with a ResNet-50 backbone. These models were tasked with classifying weed species and their respective growth stages (in weeks), while simultaneously localizing them within the images via bounding box predictions. We configured and trained these models using PyTorch and mmDetection on an NVIDIA RTX 3090 GPU.

#### A. Detection Transformer with ResNet-50

The Detection Transformer (DETR) model is an end-to-end object detection architecture that combines a convolutional backbone with a transformer encoder-decoder [27]. This approach effectively addresses the complexities of identifying weeds in agricultural images. The backbone of our model ResNet-50 is a convolutional neural network, pre-trained on ImageNet (open-mmlab://resnet50). This 50-layer network, organized into four stages, serves as a powerful feature extractor. We utilize the output from the final stage (out\_indices=(3,)) and freeze the initial stages during training to preserve pre-learned features. The backbone's output can be represented as:

$$F_{\text{resnet}} = \text{ResNet50}(I) \quad (1)$$

where  $I$  is the input image. A Channel Mapper follows the backbone, transforming ResNet-50's 2048-channel output into a 256-channel feature map suitable for the transformer. This

dimensionality reduction is achieved through a 1x1 convolution:

$$F_{\text{neck}} = \text{Conv1x1}(F_{\text{resnet}}) \quad (2)$$

The core of our DETR model is the transformer module, comprising a 6-layer encoder and decoder. Each encoder layer incorporates a self-attention mechanism with 8 heads, followed by a feed-forward network (FFN) with ReLU activation. The model's bounding box head processes the decoder's output to predict class labels and bounding boxes. We employ cross-entropy loss for classification and a combination of L1 and Generalized IoU losses for bounding box regression. The overall loss function [28] is defined as:

$$L = \alpha \cdot L_{\text{cls}} + \beta \cdot L_{\text{bbox}} + \gamma \cdot L_{\text{iou}} \quad (3)$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are weight coefficients.  $L_{\text{cls}}$  represents the classification loss, which in this case is the cross-entropy loss.  $L_{\text{bbox}}$  represents the bounding box regression loss, which is a combination of L1 loss and Generalized IoU loss, and  $L_{\text{iou}}$  represents the IoU loss, which is specifically aimed at improving the localization accuracy by penalizing the model based on the intersection over union between the predicted and ground truth bounding boxes. During training, we utilize the Hungarian algorithm [29] for bipartite matching, ensuring a one-to-one correspondence between predicted and ground-truth boxes. This approach optimizes the model's ability to accurately locate and classify weeds within agricultural images. By integrating the robust feature extraction capabilities of ResNet-50 with the DETR architecture's powerful attention mechanisms, our model achieves good performance in weed detection with 174 classes.

### B. RetinaNet with ResNeXt-101

RetinaNet is a single-stage object detection model designed to address the extreme foreground-background class imbalance encountered during training [30]. The architecture comprises three main components: a backbone network for feature extraction, a neck (FPN) for generating multi-scale feature maps, and a detection head for predicting bounding boxes and class probabilities. We utilized ResNeXt-101 as the backbone, a variant of the ResNet architecture that employs grouped convolutions for improved efficiency and performance. The ResNeXt-101 backbone consists of 101 layers organized into four stages, with 32 groups and a base width of 4 channels per group. We initialized the backbone with weights pretrained on ImageNet (`open-mmlab://resnext101_32x4d`) to leverage transfer learning. Batch normalization is applied after each convolutional layer to stabilize the learning process.

The Feature Pyramid Network (FPN) enhances the backbone's feature maps by combining high-level semantic features with low-level detailed features, enabling the detection of objects at various scales. The FPN generates multiple feature maps of different resolutions, which are then fed into the detection head. The detection head of RetinaNet comprises two subnetworks: a classification subnetwork for predicting object presence probabilities and a regression subnetwork for predicting bounding box coordinates. Each subnetwork

consists of four convolutional layers, followed by a final convolutional layer that produces the desired outputs. To handle class imbalance, we employed the focal loss function [31] for training the classification subnetwork:

$$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (4)$$

where  $p_t$  is the predicted probability,  $\alpha_t$  is a balancing factor, and  $\gamma$  is the focusing parameter.

We trained our model using an epoch-based training loop with the AdamW optimizer (learning rate  $lr = 0.0001$ , weight decay  $wd = 0.0001$ ). The learning rate schedule incorporated a linear warmup over the first 1000 iterations. We trained for 12 epochs with a batch size of 16, employing automatic learning rate scaling to accommodate potential batch size changes.

### C. Evaluation Metrics

To assess the performance of our weed detection models, we employ a comprehensive set of metrics that capture both the accuracy and robustness of the detections. Our primary metrics are Average Precision (AP), Average Recall (AR), and Mean Average Precision (mAP) evaluated across various Intersection over Union (IoU) thresholds.

AP provides a single-value summary of the precision-recall curve, effectively balancing the trade-off between precision and recall. Precision (P) is defined as the ratio of true positive detections to the sum of true positive and false positive detections:  $P = \frac{TP}{TP+FP}$  and Recall (R) is the ratio of true positive detections to the sum of true positive and false negative detections:  $(R) = \frac{TP}{TP+FN}$ .

In this research, a true positive is a detected bounding box that correctly identifies a weed species and has an IoU above a specified threshold (e.g., 0.50) with the ground truth bounding box. A false positive is a detection that either does not sufficiently overlap with any ground truth box or incorrectly identifies the weed species. A false negative occurs when a ground truth weed instance is not detected by the model. AP [32] is calculated by integrating the precision over the recall range and it can be defined as:

$$\text{AP} = \int_0^1 P(R) dR \quad (5)$$

AR [33] measures the model's ability to detect all relevant objects. It is computed as the average of maximum recalls at specified IoU thresholds:

$$\text{AR} = \frac{1}{N} \sum_{i=1}^N R_{\max}(IoU_i) \quad (6)$$

mAP is the mean of AP values across different classes and is a common metric for evaluating object detection models. It provides a balanced measure of precision and recall across various IoU thresholds. It can be defined as:

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^C \text{AP}_c \quad (7)$$

where  $\text{AP}_c$  is the Average Precision for class  $c$ , and  $C$  is the total number of classes. We evaluate these metrics at

various IoU thresholds. This multi-faceted evaluation approach allows us to comprehensively analyze our models' capabilities in detecting and classifying weeds across various scenarios, providing insights into their precision, recall, and overall detection performance.

## V. EXPERIMENTAL EVALUATION

The evaluation encompasses both training and test datasets, with a detailed analysis across 16 weed species. We employ various metrics, including AP, AR at different IoU thresholds and detection limits, as well as mAP and mean average recall (mAR). Additionally, we compare the inference speed of both models to provide a holistic view of their capabilities.

Table II compares DETR and RetinaNet performance on training and test sets, highlighting key metrics. RetinaNet consistently outperforms DETR across all presented metrics. In terms of mean Average Precision (mAP), RetinaNet achieves superior scores of 0.907 and 0.904 on training and test sets respectively, compared to DETR's 0.854 and 0.840. This trend continues in mean Average Recall (mAR), where RetinaNet approaches near-perfect scores with 0.997 (training) and 0.989 (test), while DETR achieves 0.941 and 0.936. Notably, RetinaNet's inference speed is significantly faster, operating at 7.28 Frames Per Second (FPS), more than twice the speed of DETR's 3.49 FPS. This substantial difference in processing speed, combined with RetinaNet's superior accuracy metrics, suggests it may be the more efficient choice for real-time or high-volume weed detection tasks.

Table III delves deeper, breaking down performance across all individual weed species. This table shows the average value of all 11 weeks results for 16 species. This view reveals nuances in each model's capabilities. RetinaNet demonstrates more consistent performance across species, with less variation in mAP scores. In contrast, DETR's performance fluctuates more widely, excelling with some species like AMBEL (mAP 0.817) and SIDSP (mAP 0.771), while struggling with others such as CHEAL (mAP 0.503) and SORHA (mAP 0.527). RetinaNet shines particularly bright with species like AMATA (mAP 0.832) and AMAPA (mAP 0.877), though it faces challenges with ECHCG (mAP 0.566). Across all species, RetinaNet consistently achieves higher recall, often nearing or reaching 1.0, while DETR's recall, though generally high, shows more variability. Both models exhibit the expected decline in mAP as the IoU threshold increases from 0.5 to 0.75, but RetinaNet maintains higher scores more consistently throughout this range.

We have selected four species for presenting their growth-wise experimental evaluation in this paper: Palmer amaranth (AMAPA), waterhemp (AMATA), giant foxtail (SETFA), and velvetleaf (ABUTH). These species are considered "driver weeds" or weeds that drive management decisions in USA agriculture due to their aggressive growth habits, herbicide resistance, and significant impact on crop yields [34]. AMAPA and AMATA are particularly notorious for their rapid growth and resistance to multiple herbicide modes of action, making

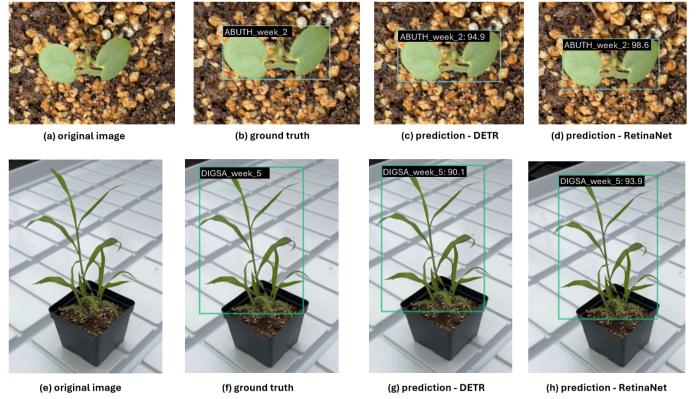


Fig. 6. Comparison of object detection results for ABUTH and DIGSA using DETR and RetinaNet models. Row 1 displays predictions for ABUTH, and Row 2 displays predictions for DIGSA, with ground truth and model confidence scores indicated for each detection.

them difficult to control and highly competitive with crops.

Tables IV, V, VI, and VII present comprehensive performance comparisons between DETR and RetinaNet across four weed species (SETFA, AMAPA, ABUTH, and AMATA) over 11 weeks. Both models demonstrated high performance across various metrics, including mAP, mAP<sub>50</sub>, mAP<sub>75</sub>, and Recall. RetinaNet generally outperformed DETR, showing more consistent and often higher scores across most species and weeks. For instance, RetinaNet achieved peak mAP scores of 0.843 for SETFA, 0.902 for AMAPA, 0.924 for ABUTH, and 0.968 for AMATA. DETR's highest mAP scores were comparable, reaching 0.859 for SETFA, 0.912 for AMAPA, 0.924 for ABUTH, and 0.905 for AMATA. Both models frequently achieved perfect scores of 1.000 in mAP<sub>50</sub> and Recall metrics across various weeks and species, indicating excellent detection accuracy at lower IoU thresholds and high object detection rates.

However, both models exhibited some performance fluctuations, particularly in the early weeks. DETR often struggled more in the initial weeks, with notably low mAP scores such as 0.355 for SETFA in Week 1, 0.096 for AMAPA in Week 1, and 0.001 for AMATA in Week 1. RetinaNet generally showed more stability, with its lowest mAP scores being higher than DETR's in most cases. For example, RetinaNet's lowest mAP for SETFA was 0.555 in Week 4, for AMAPA it was 0.481 in Week 1, and for AMATA it was 0.529 in Week 2. These early-week challenges could be attributed to factors such as less robust feature extraction, difficulty in detecting small objects, or lower-quality images in the initial stages of plant growth. Despite these early challenges, both models demonstrated significant improvement over time, with peak performances often occurring in later weeks (Weeks 8-11). This trend suggests that as plants matured and image quality potentially improved, both DETR and RetinaNet were able to more accurately detect and classify the weed species.

Figure 6 shows the prediction result of DETR and RetinaNet model. The top row focuses on ABUTH, where the first image (a) shows the original plant without any annotations, followed

TABLE II  
PERFORMANCE COMPARISON OF DETR AND RETINANET ON TRAINING AND TEST SETS

Model	mAP		mAR		FPS
	Train	Test	Train	Test	
DETR	0.854	0.840	0.941	0.936	3.49
RetinaNet	<b>0.907</b>	0.904	<b>0.997</b>	0.989	<b>7.28</b>

TABLE III  
PERFORMANCE COMPARISON OF DETR AND RETINANET ACROSS WEED SPECIES

Species Code	DETR				RetinaNet			
	Average mAP	Average mAP_50	Average mAP_75	Average Recall	Average mAP	Average mAP_50	Average mAP_75	Average Recall
ABUTH	0.683	0.907	0.719	0.973	0.720	0.924	0.779	0.993
AMAPA	0.617	0.835	0.672	0.975	0.877	0.985	0.939	0.994
AMARE	0.575	0.807	0.598	0.957	0.617	0.941	0.684	0.987
AMATA	0.536	0.721	0.565	0.869	0.832	0.977	0.905	0.997
AMBEL	0.817	0.978	0.898	0.993	0.663	0.926	0.740	0.994
CHEAL	0.503	0.846	0.502	0.962	0.871	0.993	0.957	0.997
CYPES	0.643	0.861	0.680	0.986	0.781	0.971	0.853	0.995
DIGSA	0.578	0.864	0.594	0.995	0.664	0.878	0.753	0.976
ECHCG	0.655	0.899	0.715	0.986	0.566	0.814	0.612	0.950
ERICA	0.718	0.918	0.752	0.977	0.678	0.918	0.749	0.992
PANDI	0.670	0.929	0.723	0.979	0.724	0.934	0.799	0.993
SETFA	0.680	0.903	0.756	0.990	0.785	0.967	0.854	0.993
SETPU	0.597	0.852	0.652	0.973	0.794	0.949	0.858	0.993
SIDSP	0.771	0.980	0.826	0.993	0.739	0.954	0.832	0.991
SORVU	0.582	0.791	0.624	0.871	0.713	0.925	0.789	0.995
SORHA	0.527	0.715	0.544	0.892	0.693	0.858	0.780	0.894

TABLE IV  
PERFORMANCE COMPARISON OF DETR AND RETINANET FOR SETFA

Class Name	DETR				RetinaNet			
	mAP	mAP_50	mAP_75	Recall	mAP	mAP_50	mAP_75	Recall
SETFA_Week_1	0.355	0.605	0.348	0.986	0.671	0.870	0.767	0.980
SETFA_Week_2	0.400	0.763	0.414	1.000	0.623	0.801	0.738	0.989
SETFA_Week_3	0.740	0.999	0.887	0.929	0.755	0.991	0.830	1.000
SETFA_Week_4	0.607	0.860	0.717	0.974	0.555	0.764	0.611	1.000
SETFA_Week_5	0.741	0.964	0.868	1.000	0.657	0.899	0.708	0.995
SETFA_Week_6	0.658	0.859	0.669	1.000	0.648	0.936	0.682	1.000
SETFA_Week_7	0.825	0.974	0.860	1.000	0.822	0.980	0.795	1.000
SETFA_Week_8	0.743	1.000	0.818	1.000	0.822	1.000	0.943	1.000
SETFA_Week_9	0.856	0.955	0.949	1.000	0.843	0.983	0.932	1.000
SETFA_Week_10	0.696	0.956	0.802	0.986	0.643	0.956	0.738	0.986
SETFA_Week_11	0.859	1.000	0.988	1.000	0.808	1.000	0.938	1.000

TABLE VII  
PERFORMANCE COMPARISON OF DETR AND RETINANET FOR AMATA

Class Name	DETR				RetinaNet			
	mAP	mAP_50	mAP_75	Recall	mAP	mAP_50	mAP_75	Recall
AMATA_Week_1	0.001	0.003	0.000	0.982	0.641	0.981	0.742	0.992
AMATA_Week_2	0.004	0.021	0.000	1.000	0.529	0.923	0.525	0.966
AMATA_Week_3	0.157	0.397	0.076	0.391	0.747	0.998	0.934	1.000
AMATA_Week_4	0.484	0.910	0.486	0.397	0.763	0.985	0.838	1.000
AMATA_Week_5	0.544	0.974	0.541	0.839	0.738	0.961	0.822	0.994
AMATA_Week_6	0.763	0.960	0.878	0.970	0.923	0.994	0.972	1.000
AMATA_Week_7	0.905	1.000	0.977	0.995	0.968	1.000	0.974	1.000
AMATA_Week_8	0.756	0.913	0.808	1.000	0.889	0.979	0.954	0.990
AMATA_Week_9	0.881	0.960	0.952	1.000	0.926	0.990	0.972	1.000
AMATA_Week_10	0.520	0.797	0.529	0.989	0.625	0.927	0.670	1.000
AMATA_Week_11	0.882	0.993	0.965	1.000	0.849	0.998	0.933	1.000

by the (b) ground truth with a labeled bounding box indicating "ABUTH week 2." The subsequent images display predictions by (c) DETR and (d) RetinaNet models, each bounding box labeled with the species name, the corresponding week, and the model's confidence score, with RetinaNet showing a slightly higher score (98.6) compared to DETR (94.9). The bottom row repeats this structure for DIGSA, showing the (e) original image, (f) the ground truth ("DIGSA week 5"), and the predictions from (g) DETR and (h) RetinaNet. For DIGSA, the confidence scores are close, with DETR predicting 90.1 and RetinaNet predicting 93.9, both models accurately detecting the plant but with varying degrees of confidence.

## VI. CONCLUSION

This research marks a pivotal advancement in precision agriculture by demonstrating the effectiveness of AI models, particularly RetinaNet, in weed detection and classification across various growth stages and species. Our study, conducted on a comprehensive dataset of 203,567 images spanning 16 weed species over 11 weeks, reveals RetinaNet's superior performance with mAP scores of 0.907 and 0.904 on training and test sets, and an inference speed of 7.28 FPS, significantly outpacing DETR's 0.854 and 0.840 mAP scores and 3.49

TABLE VI  
PERFORMANCE COMPARISON OF DETR AND RETINANET FOR ABUTH

Class Name	DETR				RetinaNet			
	mAP	mAP_50	mAP_75	Recall	mAP	mAP_50	mAP_75	Recall
ABUTH_Week_1	0.418	0.723	0.471	0.994	0.605	0.899	0.689	1.000
ABUTH_Week_2	0.576	0.988	0.530	1.000	0.829	0.990	0.952	1.000
ABUTH_Week_3	0.356	0.697	0.346	1.000	0.790	0.996	0.899	1.000
ABUTH_Week_4	0.408	0.771	0.396	0.996	0.725	0.973	0.844	0.995
ABUTH_Week_5	0.445	0.923	0.377	0.871	0.730	0.974	0.789	1.000
ABUTH_Week_6	0.850	1.000	1.000	0.886	0.924	0.970	0.970	0.972
ABUTH_Week_7	0.885	0.932	0.931	0.993	0.966	1.000	1.000	1.000
ABUTH_Week_8	0.856	1.000	0.982	1.000	0.911	1.000	1.000	1.000
ABUTH_Week_9	0.912	0.977	0.949	0.975	0.876	0.978	0.920	1.000
ABUTH_Week_10	0.880	0.967	0.923	1.000	0.868	0.971	0.893	1.000
ABUTH_Week_11	0.924	1.000	1.000	0.989	0.924	1.000	1.000	1.000

FPS speed. Both models exhibit improved accuracy with plant maturation, yet challenges persist during the early growth stages (weeks 1-2) due to poor differentiation between emerging plants and soil. These findings underscore the practical implications for weed management, with RetinaNet recommended for real-time applications due to its accuracy and speed. To integrate these models into existing agricultural practices, farmers should implement mobile-based applications for in-field weed detection using RetinaNet, calibrate the model for specific weed species with their growth stages prevalent in their region, and combine AI-driven detection with GPS-guided precision spraying systems. Despite the controlled greenhouse setting and early-stage detection challenges, this study lays the groundwork for future research aimed at enhancing detection accuracy through custom transformer models and expanding the dataset to include real field conditions. These AI-driven innovations hold the promise of revolutionizing weed management by enabling species-specific, growth-stage-aware detection, potentially reducing herbicide use, cutting costs, and minimizing environmental impact. By following these integration guidelines, farmers can leverage AI models to optimize their weed management strategies, leading to more sustainable and efficient agricultural practices.

## REFERENCES

- [1] A. Monteiro and S. Santos, "Sustainable approach to weed management: The role of precision weed management," *Agronomy*, vol. 12, no. 1, p. 118, 2022.
- [2] Z. Ren, D. J. Gibson, K. L. Gage, J. L. Matthews, M. D. Owen, D. L. Jordan, D. R. Shaw, S. C. Weller, R. G. Wilson, and B. G. Young, "Exploring the effect of region on diversity and composition of weed seedbanks in herbicide-resistant crop systems in the united states," *Pest Management Science*, vol. 80, no. 3, pp. 1446–1453, 2024.
- [3] W.-T. Gao and W.-H. Su, "Weed management methods for herbaceous field crops: A review," *Agronomy*, vol. 14, no. 3, p. 486, Feb. 2024.
- [4] A. M. Almalky and K. R. Ahmed, "Deep learning for detecting and classifying the growth stages of consolida regalis weeds on fields," *Agronomy*, vol. 13, no. 3, p. 934, Mar. 2023.
- [5] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End object detection with transformers," *Computer Vision – ECCV 2020*, pp. 213–229, 2020.
- [6] Y. Li, A. Dua, and F. Ren, "Light-weight RetinaNet for object detection on edge devices," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*. IEEE, Jun. 2020.
- [7] A. S. M. M. Hasan, D. Diepeveen, H. Laga, M. G. K. Jones, and F. Sohel, "Object-level benchmark for deep learning-based detection and classification of weed species," *Crop Prot.*, vol. 177, no. 106561, p. 106561, Mar. 2024.
- [8] K. Wang, X. Hu, H. Zheng, M. Lan, C. Liu, Y. Liu, L. Zhong, H. Li, and S. Tan, "Weed detection and recognition in complex wheat fields based on an improved YOLOv7," *Front. Plant Sci.*, vol. 15, p. 1372237, Jun. 2024.
- [9] C. Shackleton, R. H. Ali, and T. A. Khan, "Enhancing rangeland weed detection through convolutional neural networks and transfer learning," *Crop Design*, no. 100060, p. 100060, Jun. 2024.
- [10] A. Ahmad, D. Saraswat, V. Aggarwal, A. Etienne, and B. Hancock, "Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems," *Comput. Electron. Agric.*, vol. 184, no. 106081, p. 106081, May 2021.
- [11] N. Islam, M. M. Rashid, S. Wibowo, C.-Y. Xu, A. Morshed, S. A. Wasimi, S. Moore, and S. M. Rahman, "Early weed detection using image processing and machine learning techniques in an australian chilli farm," *Collect. FAO Agric.*, vol. 11, no. 5, p. 387, Apr. 2021.
- [12] A. Khan, T. Ilyas, M. Umraiz, Z. I. Mannan, and H. Kim, "CED-Net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture," *Electronics (Basel)*, vol. 9, no. 10, p. 1602, Oct. 2020.
- [13] R. A. Arun, S. Umapamaheswari, and A. V. Jain, "Reduced u-net architecture for classifying crop and weed using pixel-wise segmentation," in *2020 IEEE International Conference for Innovation in Technology (INOCON)*. IEEE, Nov. 2020.
- [14] S. P. Adhikari, H. Yang, and H. Kim, "Learning semantic graphics using convolutional Encoder-Decoder network for autonomous weeding in paddy," *Front. Plant Sci.*, vol. 10, p. 1404, Oct. 2019.
- [15] N. Teimouri, M. Dyrmann, P. R. Nielsen, S. K. Mathiassen, G. J. Somerville, and R. N. Jørgensen, "Weed growth stage estimator using deep convolutional neural networks," *Sensors*, vol. 18, no. 5, May 2018.
- [16] M. H. Asad, S. Anwar, and A. Bais, "Improved crop and weed detection with diverse data ensemble learning," 2023.
- [17] L. Moldvai, P. Á. Mesterházi, G. Teschner, and A. Nyéki, "Weed detection and classification with computer vision using a limited image dataset," *Appl. Sci.*, vol. 14, no. 11, p. 4839, Jun. 2024.
- [18] I. Gallo, A. U. Rehman, R. H. Dehkordi, N. Landro, R. La Grassa, and M. Boschetti, "Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images," *Remote Sens. (Basel)*, vol. 15, no. 2, p. 539, Jan. 2023.
- [19] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network," *Inf. Process. Agric.*, vol. 7, no. 4, pp. 535–545, Dec. 2020.
- [20] J. Kotleba, "European and mediterranean plant protection organization (eppo)," *Agrochemia (Slovak Republic)*, vol. 34, no. 4, 1994.
- [21] T. Borsch, W. Berendsohn, E. Dalcin, M. Delmas, S. Demissew, A. Elliott, P. Fritsch, A. Fuchs, D. Geltman, A. Güner *et al.*, "World flora online: Placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants," *Taxon*, vol. 69, no. 6, pp. 1311–1341, 2020.
- [22] "Composite List of Weeds - Weed Science Society of America — wssa.net," <https://wssa.net/weed/composite-list-of-weeds/>, [Accessed 19-08-2024].
- [23] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization techniques in training dnns: Methodology, analysis and application," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 8, pp. 10 173–10 196, 2023.
- [24] M. L. Comer and E. J. Delp III, "Morphological operations for color image processing," *Journal of electronic imaging*, vol. 8, no. 3, pp. 279–289, 1999.
- [25] A. Clark and Contributors, "Pillow (pill fork) documentation," *Read the Docs*, 2015.
- [26] L. Tzutalin, "LabelImg," <https://github.com/tzutalin/labelImg>, 2015, accessed: Aug 12, 2024.
- [27] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*. Springer, 2020, pp. 213–229.
- [28] G. Yin, L. Sheng, B. Liu, N. Yu, X. Wang, and J. Shao, "Context and attribute grounded dense captioning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6241–6250.
- [29] Y. Ye, X. Ke, and Z. Yu, "A cost matrix optimization method based on spatial constraints under hungarian algorithm," in *Proceedings of the 6th International Conference on Robotics and Artificial Intelligence*, 2020, pp. 134–139.
- [30] Y. Li and F. Ren, "Light-weight retinanet for object detection," *arXiv preprint arXiv:1905.10011*, 2019.
- [31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [32] S. Robertson, "A new interpretation of average precision," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 689–690.
- [33] M. Zhu, "Recall, precision and average precision," *Department of Statistics and Actuarial Science, University of Waterloo*, Waterloo, vol. 2, no. 30, p. 6, 2004.
- [34] A. Hager, "Weed management," <https://extension.illinois.edu>, 2022, [Accessed 14-08-2024].