PDT: Uav Target Detection Dataset for Pests and Diseases Tree

Mingle Zhou[1,2,4], Rui Xing[1,2], Delong Han[1,2], Zhiyong Qi[3], and Gang Li[1,2(✉)]

[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
lig@qlu.edu.cn
[2] Shandong Provincial Key Laboratory of Computer Networks, Shandong Fundamental Research Center for Computer Science, Jinan, China
[3] Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing, China
[4] SHANDONG SCICOM Information and Economy Research Institute Co., Ltd.

Abstract. UAVs emerge as the optimal carriers for visual weed identification and integrated pest and disease management in crops. However, the absence of specialized datasets impedes the advancement of model development in this domain. To address this, we have developed the Pests and Diseases Tree dataset (PDT dataset). PDT dataset represents the first high-precision UAV-based dataset for targeted detection of tree pests and diseases, which is collected in real-world operational environments and aims to fill the gap in available datasets for this field. Moreover, by aggregating public datasets and network data, we further introduced the Common Weed and Crop dataset (CWC dataset) to address the challenge of inadequate classification capabilities of test models within datasets for this field. Finally, we propose the YOLO-Dense Pest (YOLO-DP) model for high-precision object detection of weed, pest, and disease crop images. We re-evaluate the state-of-the-art detection models with our proposed PDT dataset and CWC dataset, showing the completeness of the dataset and the effectiveness of the YOLO-DP. The proposed PDT dataset, CWC dataset, and YOLO-DP model are presented at https://github.com/RuiXing123/PDT_CWC_YOLO-DP.

Keywords: Datasets · UAVs · Object detection

1 Introduction

The efficacy of intelligent UAV-based plant protection operations hinges on precisely identifying weeds and pests within imagery, representing a critical challenge in computer vision [6,9,13]. As computer vision and UAV technologies have advanced swiftly, the adoption of automated UAV plant protection operations is on the rise [3, 14]. The working mode of plant protection UAV has significant defects: non-intelligent work can result in the misuse or waste of agrochemicals,

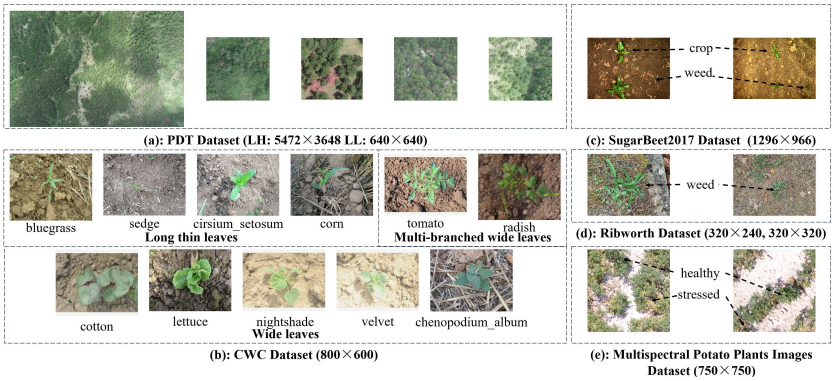

Fig. 1: Dataset comparison. (a) shows the PDT dataset (Low Resolution(LL) and High Resolution(LH)): 640×640, 5472×3648. (b) shows the characteristics of the CWC dataset: it contains 11 different similar plants. (c), (d) and (e) are the public datasets.

and the high-intensity, repetitive mechanical tasks can diminish their operational lifespan. Enhancing the precision of weed and pest target recognition is essential for the effectiveness of intelligent plant protection UAVs. The absence of specialized datasets and the limitations of existing detection models pose significant constraints on advancing research in this domain.

Drones typically take images at both high and low altitudes when outdoors. Existing datasets are usually taken from indoor greenhouses (Fig. 1 (c)), where the majority of the data consists of target images captured at close range (Fig. 1 (d)). Such data struggles to incorporate environmental factors like changes in lighting, and the target sizes significantly deviate from real-world conditions, failing to satisfy practical requirements. Further, the majority of existing datasets are limited to single or double classes, which cannot meet the training of model classification ability, as illustrated in Fig. 1 (c), (d), and (e).

On the other hand, the conventional detection models utilized in UAV-based intelligent plant protection could be more challenging in accurately identifying targets. The deficiency in this field is underscored by the absence of tailored baseline models, resulting in unmet demands for detecting and managing diverse weeds, crops, pests, and diseases. This paper aims to address the lack of dedicated datasets and algorithms for weeds, pests and diseases in agricultural drones. Overall, the main contributions of this paper are as follows:

- We have developed the UAV target detection agricultural Pests and Diseases Tree dataset, known as PDT dataset.
- We collated public data and provided Common Weed and Crop dataset (CWC dataset) to train the classification capabilities of the model.
- We have designed a specialized detection model, YOLO-DP, tailored for the dense object detection of tree crop pests and diseases.
- We reevaluated and analyzed existing generic and specific detection models on PDT and CWC datasets to establish benchmarks for the field.

2 Related Work

Agricultural UAV Detection Dataset. In the realm of agricultural robotics, acquiring pertinent datasets frequently presents a challenge, primarily due to the necessity of maintaining dedicated experimental fields and the critical timing of data collection, which is pivotal to the research outcomes. Chebrolu et al. presented SugarBeet2016 large agricultural robot dataset and subsequently introduced the SugarBeet2017 dataset tailored for the detection task [8,11]. The limitation of this dataset is that the data, captured from the vantage point of a ground robot, significantly diverges from the actual detection scenarios encountered by UAVs. Vakanski et al. introduced a dataset of multispectral potato plant images taken by drones [5]. The limitation of this dataset is the lack of raw data, consisting of only 360 images and a total of 1,500 images produced by data enhancement, which cannot effectively guide the model to learn the details of the crop. The University of Burgundy has launched the Ribworth Dataset, designed for crops and weeds, which uses data augmentation to simulate real-world environmental conditions [4]. The limitation of this dataset is that its acquisition target is relatively close, and the factors such as light change are not included, which is not enough to meet the requirements of accurate detection tasks of agricultural plant protection UAVs.

To foster the advancement of precision agriculture, we introduce the PDT dataset and CWC dataset, aimed at addressing the requirements for detecting agricultural weeds, crops, and pests.

Agricultural UAV Detection Model. The agricultural UAV target detection task has garnered significant attention within the realms of machine learning and computer vision, owing to its attributes of precise localization, efficient operation, and minimal environmental impact. Among them, the weed detection models based on YOLO deep learning model include: Zhang et al. developed a detection model for weed location and identification in wheat fields by using YOLOv3-tiny [31]. Gao et al. modified the YOLOv3-tiny model and applied it to weed detection in sugar beet fields. Jabir and Falih proposed an intelligent wheat field weed detection system based on YOLOv5 [15]. Furthermore, a variety of neural network architectures have also been recognized by researchers and implemented in the field of precision agriculture. Guo et al. proposed WeedNet-R based on the RetinaNet architecture for weed identification and location in sugar beet fields [11]. Tang et al. employed Kmeans feature learning in conjunction with a convolutional neural network for the recognition of weeds [29]. Agarwal et al. utilized the Kmeans clustering algorithm to analyze color space features in multispectral weed images for effective weed detection [1].

It is evident that there is a scarcity of studies focusing on baseline models for the detection of tree diseases and pests, particularly those related to the invasion of the exotic species Red Turpentine Beetle. Furthermore, the majority of these baseline models are trained using single-class or two-class detection datasets, and the classification ability of the models is limited by the diversity of the data. However, for the detection tasks in large-scale, high-precision agricultural UAV operations, precise detection, accurate classification, and the expansion of

crop detection varieties are essential. Consequently, we propose the YOLO-DP detection model to address these requirements.

3 PDT Dataset

In this study, we introduced a PDT dataset with both high and low resolution versions, specifically for the detection of red turpentine beetle natural pest targets. As most existing public datasets are gathered indoors or via simulated drones, they fail to capture the characteristics of outdoor UAV detection environments. Consequently, there is an immediate demand for an agricultural dataset that is tailored to real-world operational environments and encompasses the distribution of both high and low-altitude data, thereby fulfilling the high-precision, large-scale target detection requirements from a UAV perspective. This section delineates the data acquisition process for the PDT dataset, encompassing the selection of the data collection domain, the equipment used for acquisition, and the criteria for sample definition.

3.1 Data Acquisition

Selection of the Data Collection Domain. The Red Turpentine Beetle, a member of the Coleopteridae family, not only invades extensive pine forest areas but also contributes to the widespread mortality of trees, exerting a profound impact on the regional ecological environment and forestry. Despite the implementation of preventive and control strategies, including monitoring and trapping, effectively managing its continued proliferation remains a formidable challenge. The precision spraying platform of plant protection UAV is an effective solution, but the public UAV data for the invasive species Red Turpentine Beetle is blank. Consequently, we opted to collect data from a cluster of dead pine trees that have been infested by the Red Turpentine Beetle.

High-resolution Drone Camera Equipment. To obtain accurate sample data, we use a UAV mapping camera called DJI-ChanSi L2. This advanced system enhances the UAV flight platform's capabilities for acquiring accurate, efficient, and reliable 2D and 3D data. The LiDAR components have excellent performance, ranging accuracy of 2 cm at a height of 150 meters, laser wavelength of 905 nm, and laser pulse emission frequency of 240 kHz. The system supports the exFAT file system and can capture images in JPEG or PNG formats. In essence, the DJI-ChanSi L2 is a comprehensive platform that integrates lidar, a high-precision navigation system, and a mapping camera, specifically designed for high-precision data acquisition in mapping, forestry, power grid, and engineering and construction industries.

Definition of Detection Target. Distinct from close-range pest detection, the task of pest detection at low or high altitudes fundamentally involves capturing the differential characteristics between the affected vector and the healthy state. Consequently, the detection samples for the PDT dataset are sourced from both healthy and unhealthy trees within pine forests. As depicted in Fig. 2, (a)

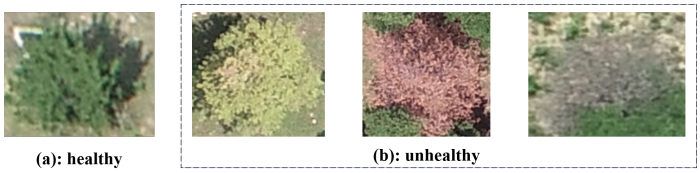

Fig. 2: Data example. (a) is a healthy goal and (b) is a unhealthy goal. The PDT dataset takes (b) as the category.

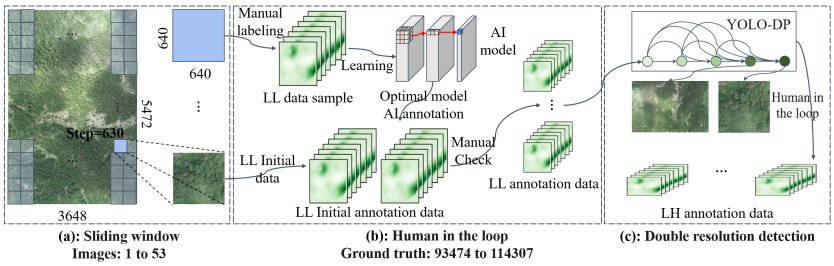

Fig. 3: PDT dataset generation and detection process. (a) represents the sliding window method, and (b) represents the "Human-in-the-loop" data annotation method. (c) means that a LL image is sent to the neural network for training, and LL and LH dual-resolution images are detected at the same time.

illustrates a pine tree in a healthy state, while (b) displays three pine trees in a unhealthy state that have been infested by pests, exhibiting progressively severe symptoms from left to right.

3.2 Data Preprocessing and Statistics

This section delves into the acquisition, processing, and production procedures of the original image data about the PDT dataset. We set up an isometric photo mode at an altitude of 150 meters to acquire 105 sets of raw image data and 3D point cloud data. Owing to the substantial size of the original images, they are unsuitable for training with the current mainstream object detection neural networks. We introduce a training approach that applies to the majority of neural network data. As illustrated in Fig. 3, we employ a methodology that encompasses "data preprocessing - 'Human-in-the-loop' data labeling - manual correction". The essence of this method is to leverage the preliminary analysis capabilities of artificial intelligence to aid human annotators, thereby enhancing the efficiency and accuracy of data annotation.

Data Preprocessing. Fig. 3 (a) demonstrates the application of the sliding window technique on the original image data, which allows for the extraction of a usable, standard-sized image through regional cropping of the high-resolution image. To prevent information loss during the sliding window process, the window dimensions are set to 640×640 pixels, with a step size of 630×630 pixels.

Table 1: Structure of PDT dataset and its sample.

Edition	Classes	Structure	Targeted images	Untargeted images	Image size	Instances	Target Amount S(Small)	M(Medium)	L(Large)
Sample	unhealthy	Train	81	1	640×640	2569	1896	548	179
		Val	19			691	528	138	25
LL	unhealthy	Train	3166	1370	640×640	90290	70418	16342	3530
		Val	395	172		12523	9926	2165	432
		Test	390	177		11494	8949	2095	450
LH	unhealthy	-	105	0	5472×3648	93474	93474	0	0

Following the sliding window operation, 53 LL images were derived from a single LH image. Furthermore, due to the high similarity between the unhealthy pine tree depicted in Fig. 2 and the ground background, and to enhance the effectiveness of neural network training, we opt to retain images that do not contain targets. Upon generating an output from these data during training, a corresponding loss is produced, which aids the neural network in learning and reduces the rate of false positives.

Data Annotation. We proceed to annotate the preprocessed data. To ensure the data's validity, we utilize the widely-used Labelme [1] data labeling software. For high-resolution datasets with dense, small targets, the sheer volume of targets per image makes manual annotation a tedious and time-intensive process. As illustrated in Fig. 3 (b), the "Human-in-the-loop" data annotation approach effectively addresses this challenge. Initially, we acquire the LL version samples of the PDT dataset through manual annotation. The structure of the sample dataset is depicted in Tab. 1, which comprises a training set and a validation set. Subsequently, the sample dataset is input into the YOLOv5s model for training over 300 epochs, with the optimal weights being saved. To ensure that no label data is missed, a confidence threshold of 0.25 and an NMS IOU (Intersection over Union) threshold of 0.45 were set to automatically label the LL data and obtain the original label set.

Manual Check. The labeled dataset underwent manual filtering to eliminate incorrect annotations, rectify erroneous ones, and add any omitted annotations. The LL datasets were generated in two widely recognized formats, YOLO_txt and VOC_xml, by automatically partitioning the data structure into an 8:1:1 ratio using scripts. For the LH data, we input the completed LL data into the YOLO-DP model for training to acquire the optimal weights. The "Human-in-the-loop" data annotation and subsequent manual verification were conducted on the LH data to obtain the final LH version dataset.

Data Statistics. The statistics of the PDT dataset are shown in Tab. 1. LL and LH versions consist of 5,670 and 105 images, respectively, with 114,307 and 93,474 samples, and feature a single class labeled as unhealthy. The training, validation, and test sets of the LL version include 1,370, 172, and 144 target-free images, respectively. On average, the LL and LH versions contain 29 and 890

[1] https://github.com/labelmeai/labelme

Table 2: CWC dataset sources.

Datasets	Corn weed datastes [25]	lettuce weed datasets [25]	radish weed datasets [25]	Fresh-weed-data [18]			
Classes	bluegrass, corn, sedge, chenopodium_album, cirsium_setosum	lettuce	radish	nightshade	tomato	cotton	velvet
Number	250	200	201	115	116	24	38
Image Size	800×600	800×600	800×600	800×600	800×600	586×444	643×500

samples per image, respectively, indicating that the PDT dataset is suitable for the task of high-density and high-precision inspection of plant protection drones.

4 CWC Dataset

We compile the available data to construct the Common Weed and Crop datasets. In special cases, there may be crop crossovers (usually in small trial plots) where weeds and crop targets exhibit multi-class structural characteristics and similarities. The majority of existing public datasets are limited to single or double classes, which do not suffice for the training and classification capabilities required by models. Consequently, when training the neural network, it is essential to supply detailed texture data that differentiates between various objects, thereby equipping the model with robust classification capabilities. This enables the model to effectively sift through and eliminate highly similar negative samples during detection, thereby enhancing the model's detection accuracy. This section discusses the acquisition and processing of CWC original image data.

4.1 Data Sources

In Tab. 2, we present the source, categories, initial quantities of the CWC raw data, as well as their respective resolutions. As depicted in Fig. 1 (b), the CWC original data comprises a total of 11 categories, characterized by similar size, texture, and color. Please note that we manually annotated the raw Corn weed datasets, Lettuce weed datasets, and Radish weed datasets, while the Fresh-weed-data was already pre-labeled and available in YOLO_txt format.

4.2 Data Preprocessing and Statistics

Data Preprocessing. Owing to the imbalance between positive and negative samples in the original CWC data, there is a risk that the model may become biased towards the majority class, leading to a diminished recognition capability for the minority class. We used the oversampling strategy, using the data augmentation method to equalize the number of samples. The methods include random rotation, random translation, random brightness change, random noise

Table 3: CWC dataset structure.

Classes		bluegrass	chenopodium_ album	cirsium_ setosum	corn	sedge	lettuce	radish	nightshade	tomato	cotton	velvet
Targeted Images	Train	200	200	200	200	200	200	200	200	200	200	200
	Val	40	40	40	40	40	40	40	40	40	40	40
	Test	10	10	10	10	10	10	10	10	10	10	10
Targeted Amount	S	1	0	0	5	0	0	0	0	0	0	3
	M	0	0	0	9	0	0	0	0	0	0	0
	L	249	250	250	236	250	444	326	250	210	268	248
Image Size		800×600	800×600	800×600	800×600	800×600	800×600	800×600	800×600	800×600	586×444	643×500

addition, flipping and cutout. This enhances the robustness of the dataset, simulating the illumination formula: $\boldsymbol{I}_{out} = \boldsymbol{I}_{in} \times w + [\mathbf{255}] \times (1 - w)$, $w \in [0.35, 1]$. Where $\boldsymbol{I}_{out}$, $\boldsymbol{I}_{in}$ is the output and input images. w is the random weight.

Data Statistics. The statistics for the CWC dataset are presented in Tab. 3. The CWC dataset comprises 2,750 images and 2,599 data entries, with the majority featuring large-sized objects. Data enhancement makes the sample reach a balanced distribution, with an average image containing 1 to 2 objects.

5 Dataset Comparison

Tab. 4 provides a thorough comparison of the datasets. The LH version of the PDT dataset is exceptionally clear, offering image quality that is 35 to 200 times superior to other publicly available datasets. The PDT dataset is characterized by density, small target and rich real environmental details, which accords with the typical environmental factors of high and low altitude UAV work, and is suitable for training the special UAV detection model for plant protection. Furthermore, the PDT dataset includes a 3D point cloud version. While not highlighted in this study, a 3D data version is planned for development and will be released to the public in the future.

On the other hand, the CWC dataset demonstrates outstanding classification performance, particularly in terms of category diversity, detailed texture information, and dataset authenticity. CWC dataset has a large number of categories, 2 to 5 times more than other publicly available datasets, which indicates that the CWC dataset is well suited for tasks involving high-precision classification of plant protection UAVs.

6 YOLO-DP Model

To expedite research efforts in UAV detection models for intelligent agricultural plant protection, we have designed and proposed a model, YOLO-DP (Fig. 4), specifically dedicated to the detection of tree crop diseases and pests, building upon the YOLOv5s foundatin [16]. To ensure the scalability of the model, the Kmeans clustering algorithm is employed to determine the distribution of the

Table 4: Dataset comparison.

Dataset	Resolution	Classes	Rich Details	High Definition	Scale	Dense Target	Target Scale	Uav Collection	No Target Image	3D Point Cloud Data	Annotation Quality
PDT dataset	5472×3648	1	-	✓	100+	✓	S	✓	✓	✓	"Human-in-
	640×640	1	-	-	5K+	✓	S/M/L	✓	✓	✓	the-loop"
CWC dataset	800×600	11	✓	-	2K+	-	L	-	-	-	Manually
SugarBeet2017 [11]	1296×966	2	-	-	5K+	-	S/M	-	-	✓	Manually
Multispectral Potato Plants Images (MPP) [5]	750×750	2	-	-	1K+	✓	L	✓	-	-	Manually
Ribworth Dataset (RI) [4]	320×240 320×320	1	✓	-	5K+	-	M/L	-	-	-	Manually
Corn weed Dataset [25]	800×600	5	✓	-	6K+	-	L	-	-	-	No
lettuce weed Dataset [25]	800×600	2	✓	-	700+	-	L	-	-	-	No
radish weed Dataset [25]	800×600	2	✓	-	400+	-	L	-	-	-	No
Fresh -weed-data [29]	800×600 586×444 643×500	4	✓	-	200+	-	L	-	-	-	Manually
crop and weed detection data (CAW) [10]	512×512	2	✓	-	1K+	-	L	-	-	-	Manually
Weeds [7]	480×480	1	-	-	600+	-	S/M	✓	-	-	Manually

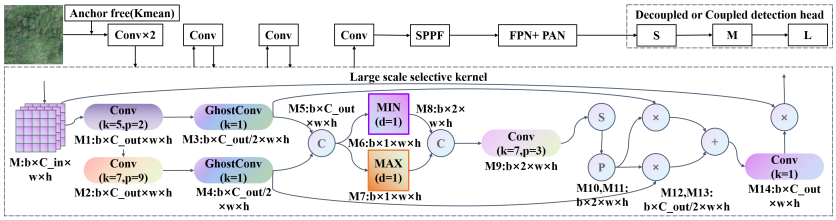

Fig. 4: YOLO-DP baseline model architecture. The FPN [21]+PAN [23] module consists of GhostConv [12], Upsample, Concat, and C3. C stands for Concat, S for Sigmoid, P for channel number dilation, × for matrix multiplication, and + for matrix addition.

actual bounding box data, allowing for the dynamic adjustment of the anchor box sizes. In light of the UAV's large-scale detection capabilities, an adaptive Large Scale Selective Kernel is incorporated into the Backbone network to capture the location information of dense, small target pest-infested trees [20]. Taking into account the efficient detection mode of UAVs, GhostConv is utilized in the Neck network to reduce the model size and computational complexity. Concurrently, the receptive field is enlarged to capture more texture feature information, thereby enhancing the model's classification capabilities. A version with decoupled detection heads is provided to minimize the interference between classification and regression losses, allowing the model to focus on specific tasks, improving the accuracy and accuracy of the detection, and also improving the model's generalization ability.

Adaptive Large Scale Selective Kernel. To tackle the demanding detection task involving a broad range of dense, small targets, we gather pest and

disease target information by designing a Convolutional group with an extensive receptive field. Firstly, the matrix M obtains the shallow range information M1 by Conv with $Kernelsize(k) = 5$ and $Padding(p) = 2$. The depth information is explored for M1, and the deep range information M2 is obtained using Conv with $k = 7$ and $p = 9$. Feature selection was performed on shallow and depth range information, and M3 and M4 were obtained by grouping convolution on M1 and M2 using GhostConv with $k = 1$. The advantage of using GhostConv is that half of the range information is retained and the other half is processed to prevent excessive loss of information. Next, the spatial attention is calculated for M3 and M4 large-scale information. M3 and M4 are concatenated in the channel dimension to obtain M5, and M6 and M7 are obtained by averaging and maximizing the channel dimensions, respectively. M6 and M7 are concatenated in the channel dimension to obtain M8 with scale $b \times 1 \times w \times h$, which is fed into Conv with $k = 7$ and $p = 4$ for range attention collection to obtain M9. After Sigmoid activation, the channel dimension is expanded to obtain M10 and M11. Matrix multiplication is performed with M3 and M4 to obtain M12 and M13, respectively. After bitwise addition, the final spatial attention matrix M14 is obtained after Conv ($k = 1$). Finally, the output is obtained by matrix multiplication with the input matrix.

7 Experiment

7.1 Experimental Conditions

The experimental environment of this paper is ubuntu 18.04, Python3.9, PyTorch 1.9.1 framework, CUDA 11.4 version, and cuDNN 8.0 version. This paper's model training and inference are performed on NVIDIA A100-SXM4 with 40GB GPU memory and 16 GB CPU memory on the experimental platform. The training metricss in this paper are: 300 epochs of iterative training, the early stopping mechanism is set to 100 rounds, the initial learning rate is 0.01, the weight decay of 0.001 and momentum of 0.9% are used, the Batch-size is 16.

Experiments were carried out on the PDT dataset and CWC dataset presented in this paper, as well as on the public dataset SugarBeet2017. We test the performance of the model YOLO-DP according to the characteristics of the dataset. First, the PDT dataset is used to test the detection ability of the model for dense small targets in the UAV perspective. Secondly, the ability of the model to extract fine-grained texture information is measured on the CWC dataset to verify its classification ability. It is reflected in the P metrics. Finally, the detection accuracy was measured on the SugarBeet2017 dataset. It is reflected in the R metrics. In order not to lose the image information, the input sizes are 640×640, 800×800, 1296×1296 respectively.

7.2 Experimental Analysis

Dataset Validation. We evaluated 6 detection models on seven datasets (2 of ours and 5 public datasets mentioned in the paper). Based on the dataset's

characteristics, we choose different metrics for the ranking model (Rank (Metrics)). Specifically, dense target datasets collected in natural environments use the comprehensive F1 metrics, datasets with target size changes use the R metrics, and multiclass datasets use the P metrics. We sort models with the same metrics score again using Gflops. Our model has excellent performance on all 7 datasets, and the performance of the proposed 2 datasets differs from that of existing datasets, indicating that our dataset can provide unique data distribution to the research field (Tab. 5).

Table 5: Datasets effect ranking.

Method	Gflops	Rank (F1)		Rank (P)		Rank (R)		
		PDT (LL) (our)	MPP [5]	CWC (our)	Weeds [7]	SugarBeet2017 [11]	CAW [10]	RI [4]
YOLO-DP	11.7	**1 (0.89)**	**1 (0.42)**	2 (92.9%)	2 (76.5%)	**1 (73.8%)**	2 (89.4%)	**1 (82.3%)**
YOLOv3	155.3	5 (0.88)	5 (0.38)	6 (86.6%)	**1 (83.0%)**	6 (46.2%)	6 (73.1%)	6 (74.0%)
YOLOv4s	20.8	3 (0.88)	2 (0.42)	5 (87.3%)	3 (75.6%)	3 (60.3%)	3 (86.5%)	2 (81.7%)
YOLOv5s	16.0	2 (0.89)	3 (0.38)	4 (88.6%)	4 (75.3%)	4 (58.3%)	5 (82.5%)	3 (81.5%)
YOLOv7	105.1	6 (0.85)	6 (0.24)	**1 (93.1%)**	5 (74.4%)	5 (48.1%)	4 (83.3%)	4 (80.3%)
YOLOv8s	28.6	4 (0.88)	4 (0.38)	3 (92.0%)	6 (70.4%)	2 (65.0%)	**1 (90.1%)**	5 (79.3%)

Comparative Experiment. We perform a comparative analysis of single-stage detectors including the full YOLO family, SSD, EfficientDet, and RetinaNet, the two-stage detector Fast-RCNN, anchor-free based CenterNet, and WeedNet-R, a specialized model for weed detection. Among them, some models use pre-trained weights. As some models do not provide the computation results for certain metricss, these are denoted by "-" in the table.

In Tab. 6, we present the comparative evaluations of the proposed YOLO-DP baseline model alongside other models across the three datasets. The YOLO-DP model does not need to load pre-trained weights, and the detection speed is significantly ahead of other models. Under this premise, YOLO-DP demonstrates the most outstanding overall performance on the PDT dataset, with its F1 composite metrics surpassing all other models. This shows that the YOLO-DP model can well adapt to the detection scenario of UAV with large-scale, high-precision, dense and small targets. On the CWC dataset, its P metrics is at the first-class level, which is 4.3% higher than the YOLOv5s baseline model. This demonstrates that the YOLO-DP model possesses an effective capability for extracting fine-grained texture information and is capable of fulfilling the classification task requirements within this domain. On the authoritative dataset SugarBeet2017, while the overall performance does not match that of models with loaded and trained weights, the R metrics of YOLO-DP is comparable to the pre-trained model's performance. This substantiates that the YOLO-DP model possesses precise detection capabilities and can accurately pinpoint the detection targets.

The comparative experimental results show that the YOLO-DP model is suitable for the application of plant protection UAV in precision agriculture detection fields such as weeds, pests and diseases, and crops.

Ablation Experiment. We chose the YOLOv5s that performed well on the PDT dataset as the benchmark. We use the popular attention mechanism to ablate the adaptive large-scale selection kernel to verify its detection performance

Table 6: Comparative experiment.

Datasets	Approach	P	R	mAP@.5	mAP@.5:.95	F1	Gflops	Parameters	FPS	Pre-training
PDT dataset (LL)	SSD [24]	84.5%	87.7%	85.1%	-	0.86	273.6	23.7M	37	✓
	EfficientDet [28]	92.6%	73.4%	72.3%	-	0.82	**11.5**	6.7M	12	✓
	RetinaNet [22]	93.3%	65.3%	64.2%	-	0.79	109.7	32.7M	32	✓
	CenterNet [32]	**95.2%**	67.4%	66.5%	-	0.79	109.7	32.7M	32	✓
	Faster-RCNN [27]	57.8%	70.5%	61.7%	-	0.64	401.7	136.7M	13	✓
	YOLOv3 [26]	88.5%	88.1%	93.4%	65.7%	0.88	155.3	61.5M	41	-
	YOLOv4s [2]	88.8%	88.2%	**94.7%**	66.1%	0.88	20.8	9.1M	51	-
	YOLOv5s_7.0 [16]	88.9%	**88.5%**	94.2%	67.0%	**0.89**	16.0	7.0M	93	-
	YOLOv6s [19]	-	-	91.4%	63.2%	-	44.1	17.2M	43	-
	YOLOv7 [30]	87.4%	82.6%	90.1%	55.5%	0.85	105.1	37.2M	32	-
	YOLOv8s [17]	88.7%	87.5%	94.0%	**67.9%**	0.88	28.6	11.1M	60	-
	WeedNet-R [11]	87.7%	48.1%	70.4%	-	0.62	19.0	25.6M	0.5	-
	YOLO-DP (our)	90.2%	88.0%	94.5%	67.5%	**0.89**	11.7	**5.2M**	**109**	-
CWC dataset	SSD [24]	**97.7%**	77.6%	85.7%	-	0.91	426.9	23.7M	29	✓
	EfficientDet [28]	97.2%	**98.6%**	**98.6%**	-	0.90	**11.5**	6.7M	13	✓
	RetinaNet [22]	95.1%	98.3%	98.0%	-	**0.97**	261.3	36.4M	24	✓
	CenterNet [32]	96.6%	73.8%	73.3%	-	0.80	171.4	32.7M	27	✓
	YOLOv3 [26]	86.8%	89.4%	93.2%	82.3%	0.88	154.7	61.5M	30	-
	YOLOv4s [2]	87.3%	87.9%	91.9%	81.5%	0.88	20.8	9.1M	43	-
	YOLOv5s_7.0 [16]	88.6%	88.7%	93.0%	81.2%	0.89	16.0	7.0M	65	-
	YOLOv6s [19]	-	-	92.7%	84.3%	-	68.9	17.2M	31	-
	YOLOv7 [30]	93.1%	76.4%	88.1%	75.6%	0.84	105.1	37.2M	21	-
	YOLOv8s [17]	92.0%	89.1%	94.0%	**86.2%**	0.91	28.6	11.1M	38	-
	WeedNet-R [11]	86.1%	51.8%	71.6%	-	0.65	19.0	25.6M	0.5	-
	YOLO-DP (our)	92.9%	87.5%	91.8%	81.0%	0.90	11.7	**5.2M**	**72**	-
Sugar-Beet2017	SSD [24]	85.0%	83.6%	79.3%	-	0.85	1120	23.7M	19	✓
	EfficientDet [28]	93.3%	79.8%	77.8%	-	**0.86**	**11.5**	6.7M	16	✓
	RetinaNet [22]	91.7%	78.8%	76.6%	-	0.84	256.4	36.3M	23	✓
	CenterNet [32]	**97.9%**	51.2%	51.0%	-	0.62	117.4	32.7M	41	✓
	Faster-RCNN [27]	63.6%	**87.4%**	**80.0%**	-	0.73	546.9	136.7M	25	✓
	YOLOv3 [26]	34.8%	46.2%	39.4%	25.6%	0.40	155.3	61.5M	28	-
	YOLOv4s [2]	28.1%	60.3%	41.1%	26.4%	0.38	20.8	9.1M	28	-
	YOLOv5s_7.0 [16]	25.0%	58.3%	40.6%	**26.7%**	0.35	16.0	7.0M	50	-
	YOLOv6s [19]	-	-	24.6%	15.0%	-	185.2	17.2M	49	-
	YOLOv7 [30]	34.2%	48.1%	38.6%	24.9%	0.40	105.1	37.2M	18	-
	YOLOv8s [17]	23.9%	65.0%	39.1%	26.1%	0.35	28.6	11.1M	33	-
	WeedNet-R [11]	90.1%	68.4%	84.8%	-	0.78	19.0	25.6M	0.5	-
	YOLO-DP (our)	23.1%	73.8%	38.3%	25.0%	0.35	11.7	**5.2M**	**62**	-

on PDT dataset. The attention modules are designed on BackBone's floors 3, 5, 7, and 9. Since the number of C3x and C3TR parameters is too large, we only replace them at the 9th layer in order to carry out the experiment. "v5s" means YOLOv5s_7.0. "v5s_our" stands for the use of Adaptive Large Scale Selective Kernel in the YOLOv5s_7.0. It can be seen from the results in Tab. 7 that the performance of YOLOv5s is greatly improved after Large Scale Selective Kernel is used. The mAP@.5, mAP@.5:.95 and F1 are the best metrics. This indicates that the Large Scale Selective Kernel used in this paper is suitable for dense object detection tasks.

7.3 Visualization Research

In Fig. 5, we show the detection visualization results on the PDT dataset. We have selected the YOLOv5s and YOLOv8s models, which exhibited superior overall performance in the comparative experiment, to demonstrate and contrast with the YOLO-DP model. (a) and (e) illustrate the Ground Truth for the LL and LH versions, respectively. Observations indicate that the "Human-in-the-loop" data annotation approach is practical for weed and pest detection data annotation. The detection results in (b) and (c) suggest that the LL version

Table 7: Ablation experiment.

Datasets	Approach	P	R	mAP@.5	mAP@.5:.95	F1	Gflops	Parameters
PDT dataset (LL)	v5s_C1	88.2%	88.5%	93.9%	67.1%	0.88	25.3	10.0M
	v5s_C2	88.8%	88.4%	94.1%	67.0%	0.88	17.2	7.3M
	v5s_C2f	88.6%	88.5%	93.8%	67.1%	0.88	17.4	7.5M
	v5s_C3	88.9%	88.5%	94.2%	67.0%	**0.89**	16.0	7.0M
	v5s_C3x	88.7%	81.5%	87.4%	63.2%	0.85	14.5	6.5M
	v5s_C3TR	88.3%	**89.0%**	94.1%	67.1%	**0.89**	15.7	7.0M
	v5s_C3Ghost	88.9%	88.2%	94.2%	66.7%	0.88	12.5	5.9M
	v5s_SE	88.8%	88.3%	**94.4%**	66.6%	0.88	10.6	**5.1M**
	v5s_CBAM	**89.8%**	87.5%	**94.4%**	66.5%	**0.89**	10.9	5.6M
	v5s_GAM	89.2%	87.7%	94.0%	67.1%	0.88	16.4	7.5M
	v5s_ECA	89.7%	87.0%	94.3%	66.2%	0.88	**10.5**	**5.1M**
	v5s_our	89.1%	88.5%	**94.4%**	**67.2%**	**0.89**	12.2	6.1M

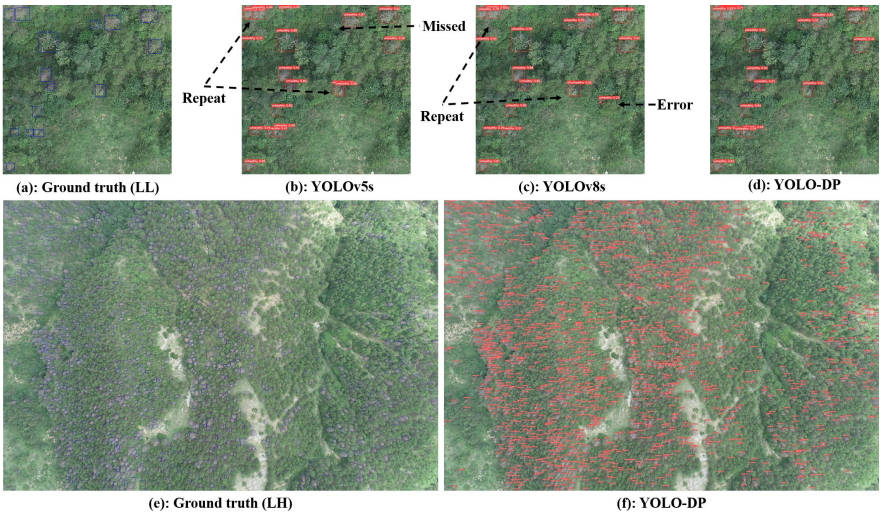

Fig. 5: Visualization of PDT dataset detection.

of PDT dataset is compatible with widely used detection models, confirming its utility. When compared to (d), it is evident that YOLO-DP delivers outstanding detection performance on LL, with no instances of missing, false, or duplicate detections. Examining the LH version's detection results in (f), we find no missed detections, and the critical issue of current detectors' inability to distinguish dead pine trees from the ground has been overcome. This leads us to conclude that the YOLO-DP model is well-suited for large-scale, high-precision, small-target UAV detection tasks. Moreover, the detection strategy depicted in Fig. 3 of this paper has been validated as effective.

In Fig. 6, we presents a comparison between YOLO-DP and the Confusion Matrix of the high-performing classification models YOLOv7 and YOLOv8s on the CWC dataset. It is evident that YOLO-DP's classification performance surpasses that of YOLOv7. Furthermore, YOLO-DP's classification capability is

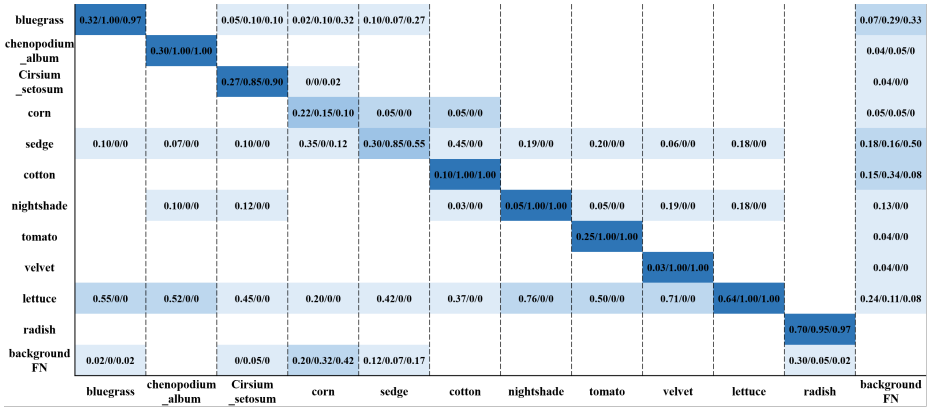

Fig. 6: Confusion matrix. The rows represent the true class, the columns represent the predicted class, and the confidence value is [0,1]. The format of the matrix is: YOLOv7/YOLOv8s/YOLO-DP.

comparable to YOLOv8s, especially when its overall performance exceeds that of YOLOv8s.

8 Conclusion

In this study, we introduce the PDT dataset to address the challenging task of UAV pest detection, with the goal of advancing research in UAV detection tasks within the realm of precision agriculture. We also present the CWC dataset for agricultural weed and crop classification, which compensates for the field's deficiency in training model classification capabilities. To tackle the scarcity of baseline model research in this domain, we propose the YOLO-DP model for dense, small-target UAV pest detection and validate its efficacy across three datasets. It is noteworthy that the PDT dataset features a dual-resolution version, and the CWC dataset boasts 11 detailed, texture-similar plant classes. Moreover, we provide a comprehensive evaluation of both datasets, objectively assessing their value. This work has its limitations: the 3D point cloud version of the PDT dataset is not discussed, and there is a size inconsistency issue within the CWC dataset. We plan to address these issues in future endeavors. Lastly, to ensure the continuity of this research, we make the datasets and associated code designed in this paper available on our website.

Acknowledgments

This work was supported by Key R&D Program of Shandong Province, China (2023CXGC010112), the Taishan Scholars Program (NO. tsqn202103097, NO. tscy20221110).

References

1. Agarwal, R., Hariharan, S., Rao, M.N., Agarwal, A.: Weed identification using k-means clustering with color spaces features in multi-spectral images taken by UAV. In: IGARSS. pp. 7047–7050. IEEE (2021)
2. Bochkovskiy, A., Wang, C., Liao, H.M.: YOLOv4: Optimal speed and accuracy of object detection. CoRR **abs/2004.10934** (2020)
3. Boursianis, A.D., Papadopoulou, M.S., Diamantoulakis, P.D., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., Karagiannidis, G.K., Wan, S., Goudos, S.K.: Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review. Internet Things **18**, 100187 (2022)
4. university of burgandy: Ribworth dataset. `https://universe.roboflow.com/university-of-burgandy-zowkw/ribworth` (nov 2022), `https://universe.roboflow.com/university-of-burgandy-zowkw/ribworth`
5. Butte, S., Vakanski, A., Duellman, K., Wang, H., Mirkouei, A.: Potato crop stress identification in aerial images using deep learning-based object detection. CoRR **abs/2106.07770** (2021)
6. Caras, T., Lati, R.N., Holland, D., Dubinin, V.M., Hatib, K., Shulner, I., Keiesar, O., Liddor, G., Paz-Kagan, T.: Monitoring the effects of weed management strategies on tree canopy structure and growth using uav-lidar in a young almond orchard. Comput. Electron. Agric. **216**, 108467 (2024)
7. Carboni: weeds pytorch (2023), `https://github.com/carboni123/weeds-pytorch`
8. Chebrolu, N., Lottes, P., Schaefer, A., Winterhalter, W., Burgard, W., Stachniss, C.: Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. Int. J. Robotics Res. **36**(10), 1045–1052 (2017)
9. Costa, L.S., Sano, E.E., Ferreira, M.E., Munhoz, C.B.R., Costa, J.V.S., Júnior, L.R.A., de Mello, T.R.B., da Cunha Bustamante, M.M.: Woody plant encroachment in a seasonal tropical savanna: Lessons about classifiers and accuracy from UAV images. Remote. Sens. **15**(9), 2342 (2023)
10. Dabhi: Crop and weed detection (2021), `https://github.com/ravirajsinh45/Crop_and_weed_detection`
11. Guo, Z., Goh, H.H., Li, X., Zhang, M., Li, Y.: WeedNet-R: a sugar beet field weed detection algorithm based on enhanced retinanet and context semantic fusion. Frontiers in Plant Science **14**, 1226329 (2023)
12. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: GhostNet: more features from cheap operations. In: CVPR. pp. 1577–1586. Computer Vision Foundation / IEEE (2020)
13. Huang, J., Luo, Y., Quan, Q., Wang, B., Xue, X., Zhang, Y.: An autonomous task assignment and decision-making method for coverage path planning of multiple pesticide spraying UAVs. Comput. Electron. Agric. **212**, 108128 (2023)
14. Innani, S., Dutande, P., Baheti, B., Talbar, S.N., Baid, U.: Fuse-PN: A novel architecture for anomaly pattern segmentation in aerial agricultural images. In: CVPR Workshops. pp. 2960–2968. Computer Vision Foundation / IEEE (2021)
15. Jabir, B., Falih, N.: Deep learning-based decision support system for weeds detection in wheat fields. International Journal of Electrical and Computer Engineering **12**(1), 816 (2022)
16. Jocher, G.: YOLOv5 by Ultralytics (May 2020). `https://doi.org/10.5281/zenodo.3908559`, `https://github.com/ultralytics/yolov5`
17. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLO (Jan 2023), `https://github.com/ultralytics/ultralytics`

18. Lameski, P., Zdravevski, E., Trajkovik, V., Kulakov, A.: Weed detection dataset with RGB images taken under variable light conditions. In: ICT Innovations. Communications in Computer and Information Science, vol. 778, pp. 112–119. Springer (2017)
19. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., Wei, X.: YOLOv6: A single-stage object detection framework for industrial applications. CoRR **abs/2209.02976** (2022)
20. Li, Y., Hou, Q., Zheng, Z., Cheng, M., Yang, J., Li, X.: Large selective kernel network for remote sensing object detection. In: ICCV. pp. 16748–16759. IEEE (2023)
21. Lin, T., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. In: CVPR. pp. 936–944. IEEE Computer Society (2017)
22. Lin, T., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV. pp. 2999–3007. IEEE Computer Society (2017)
23. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: CVPR. pp. 8759–8768. Computer Vision Foundation / IEEE Computer Society (2018)
24. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C., Berg, A.C.: SSD: single shot multibox detector. In: ECCV (1). Lecture Notes in Computer Science, vol. 9905, pp. 21–37. Springer (2016)
25. Monster: weed dataset (2019), `https://gitee.com/Monster7/weed-datase/tree/master/`
26. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. CoRR **abs/1804.02767** (2018)
27. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
28. Tan, M., Pang, R., Le, Q.V.: EfficientDet: scalable and efficient object detection. In: CVPR. pp. 10778–10787. Computer Vision Foundation / IEEE (2020)
29. Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., Xu, Y.: Weed identification based on k-means feature learning combined with convolutional neural network. Comput. Electron. Agric. **135**, 63–70 (2017)
30. Wang, C., Bochkovskiy, A., Liao, H.M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: CVPR. pp. 7464–7475. IEEE (2023)
31. Zhang, R., Wang, C., Hu, X., Liu, Y., Chen, S., et al.: Weed location and recognition based on uav imaging and deep learning. International Journal of Precision Agricultural Aviation **3**(1) (2020)
32. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. CoRR **abs/1904.07850** (2019)