IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 September 2023, accepted 16 October 2023, date of publication 18 October 2023, date of current version 25 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3325747

APPLIED RESEARCH

Plant Detection and Counting: Enhancing Precision Agriculture in UAV and General Scenes

DUNLU LU[1], JIANXIONG YE[1], YANGXU WANG[1], AND ZHENGHONG YU[1,2]
[1]College of Robotics, Guangdong Polytechnic of Science and Technology, Guangzhou 519090, China
[2]Mahanakorn Institute of Innovation, Mahanakorn University of Technology, Bangkok 10530, Thailand

Corresponding author: Zhenghong Yu (honger1983@gmail.com)

This work was supported in part by 2022 Key Scientific Research Project of Ordinary Universities in Guangdong Province under Grant 2022ZDZX4075, in part by 2022 Guangdong Province Ordinary Universities Characteristic Innovation Project under Grant 2022KTSCX251, in part by the Collaborative Intelligent Robot Production and Education Integrates Innovative Application Platform Based on the Industrial Internet under Grant 2020CJPT004, in part by 2020 Guangdong Rural Science and Technology Mission Project under Grant KTP20200153, in part by the Engineering Research Centre for Intelligent Equipment Manufacturing under Grant 2021GCZX018, and in part by the Guangdong Polytechnic of Science and Technology & DOBOT Collaborative Innovation Center under Grant K01057060.

ABSTRACT Plant detection and counting play a crucial role in modern agriculture, providing vital references for precision management and resource allocation. This study follows the footsteps of machine learning experts by introducing the state-of-the-art Yolov8 technology into the field of plant science. Moreover, we made some simple yet effective improvements. The integration of shallow-level information into the Path Aggregation Network (PANet) served to counterbalance the resolution loss stemming from the expanded receptive field. The enhancement of upsampled features was accomplished through combining the lightweight up-sampling operator Content-Aware ReAssembly of Features (CARAFE) with the Multi-Efficient Channel Attention (Mlt-ECA) technique to optimize the precision of upsampled features. This collective approach markedly amplified the discernment of small objects in Unmanned Aerial Vehicle (UAV) images, naming it Yolov8-UAV. Our evaluation is based on datasets containing four different plant species. Experimental results demonstrate the strong competitiveness of our proposed method even when compared to the most advanced counting techniques, and it possesses sufficient robustness. In order to advance the cross-disciplinary research of computer vision and plant science, we also release a new cotton boll dataset with detailed annotated bounding box information. What's more, we address previous oversights in existing wheat ear datasets by providing updated labels consistent with global research advancements. Overall, this research offers practitioners a powerful solution for addressing real-world application challenges. For UAV scenarios, recommend using the specialized Yolov8-UAV, while Yolov8-N is a wise choice for general scenes due to its sufficient accuracy and speed in the majority of cases. Furthermore, we contribute two meaningful datasets that have research significance, effectively promoting the application of data resources in the field of plant science. In short, our contribution is to improve the use of Yolov8 in UAV scenarios and open two datasets with bounding boxes. The curated data and code can be accessed at the following link: https://github.com/Ye-Sk/Plant-dataset.

INDEX TERMS Cotton boll, detection and counting, UAV, wheat ear, Yolov8.

I. INTRODUCTION

In recent years, deep learning, as the core technology of the third wave of artificial intelligence, has made rapid progress and demonstrated remarkable performance and extensive applications in various fields [1]. In scientific research and engineering practice, deep learning has achieved significant achievements. Plant detection and counting, as important tasks in plant science and agricultural production, have also benefited from the advancements in deep learning technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

Accurate plant detection and counting play a critical role in plant research, precision agricultural management, and resource allocation [2], [3]. However, traditional methods for plant detection and counting have limitations, including limited feature extraction capabilities and subjective manual rule design [4], [5], [6]. These methods struggle to cope with the complexity and variability of plant scenes and the processing requirements of large-scale data.

The emergence of deep learning technology has provided new solutions for addressing plant detection and counting problems. Deep learning is a machine learning approach based on multi-layer neural networks, which efficiently handles complex tasks by automatically learning feature representations and pattern recognition from large-scale data [7], [8]. In the context of plant detection and counting, deep learning techniques have brought new breakthroughs to plant science and agricultural production with their robust feature learning and pattern recognition capabilities. Through training and inference of deep learning models, accurate detection and counting of plant objects in image data can be achieved, greatly improving work efficiency and data processing accuracy [9], [10].

Over the past few years, there have been many advanced deep learning-based methods and models emerging in the field of plant detection and counting, providing formidable tools for agricultural producers to monitor and control various issues related to plant growth. Object detection, as an important research direction, has gained increasing attention in the plant domain. Researchers have started exploring the use of deep learning models for plant detection and counting tasks, including well-known models such as Yolo [11], Faster R-CNN [12], and EfficientDet [13]. Some researchers have also made a series of improvements to accomplish plant detection and counting tasks [14], [15]. Despite that, these improvements often involve complex and laborious implementation processes and are optimized for specific application scenarios. This situation limits the development of cross-disciplinary research between computer vision and plant science towards a more general direction.

Fortunately, thanks to the relentless efforts of machine learning pioneers, some excellent general-purpose machine learning models have been proposed [8], [16]. Among them, the Yolo model has garnered significant attention due to its outstanding balance between accuracy and speed. As the latest detector in the Yolo series, Yolov8 not only inherits the advantages of previous models but also surpasses them, becoming a potent tool for practitioners in the field of plant science.

With the rapid development of Unmanned Aerial Vehicles (UAV) and remote sensing technology [4], [17], in the vast realm of research, numerous scholars are dedicating their efforts to advancing the analysis of remote sensing images. Several publicly available remote sensing datasets, such as Remote Sensing Object Detection (RSOD) [18] and University of Chinese Academy of Sciences - Aerial Object Detection (UCAS-AOD) [19], are providing robust support for research endeavors. On a different front, Liang et al. [20] introduced a single-stage detector known as FS-SDD. They constructed a feature pyramid by combining deconvolution modules and feature fusion modules, fully harnessing these hierarchical features during the prediction process. Their approach also takes spatial context information into account. Wang et al. [21], on the other hand, proposed a detector with contextual information to alleviate the challenge of complex backgrounds in remote sensing images. They also enhanced the region proposal network of RCNN. Furthermore, Liu et al. [22] devised a Multi-branch Parallel Feature Pyramid Networks (MPFPN) to recover small object features lost in deep semantic information. Whereas, these methods demand significant memory and computational resources, limiting their practical application on low-power edge image processing devices. In the realm of agriculture, Lu et al. [41] proposed a local counting network named TasselNetV3, which improved the visual output by introducing an upsampling operator to supervise the redistribution of counts. Bai et al. [42] designed a deep network called RPNet, which enhances the counting performance for rice plants by densely utilizing shallow and deep features. Liu et al. [12] employed ResNet as the backbone for Faster R-CNN to detect tassels in high-resolution UAV images. While they effectively enhance the recognition performance for small-sized plant objects, these aforementioned methods require high-performance computing devices for both training and inference. At the current stage, high-resolution plant image datasets collected by UAV have gained widespread attention. These datasets contain diverse plant objects and complex scenes, better simulating real-world application environments and driving the application of plant detection and counting methods in agricultural production. In this study, we selected Yolov8 as a powerful baseline model and enhanced its perception of small objects by introducing a simple yet effective upsampling process. Unlike previous research, we replaced the traditional nearest-neighbor upsampling operation in Yolov8 with a data-dependent lightweight upsampling operator called Content-Aware ReAssembly of FEatures (CARAFE) [23]. Nevertheless, after each CARAFE operation, we applied a Multi-Efficient Channel Attention (Mlt-ECA) [24] for weighted adjustment of features. These improvement methods are straightforward to implement. We chose this approach because the Yolov8 baseline itself has demonstrated strong performance, and excessive complex improvements may lead to other performance trade-offs. The improved model is named Yolov8-UAV, as it is more suitable for UAV-like image detection tasks.

In addition to model design and training, dataset construction and annotation are also crucial aspects. To our knowledge, there is currently no publicly available cotton boll dataset. Therefore, based on previous automated observation work [4], we have released a cotton boll dataset named Cotton Boll Detection Augmented (CBDA), which includes annotated bounding boxes. We also noticed that Madec et al. [26] contributed a wheat ear dataset called

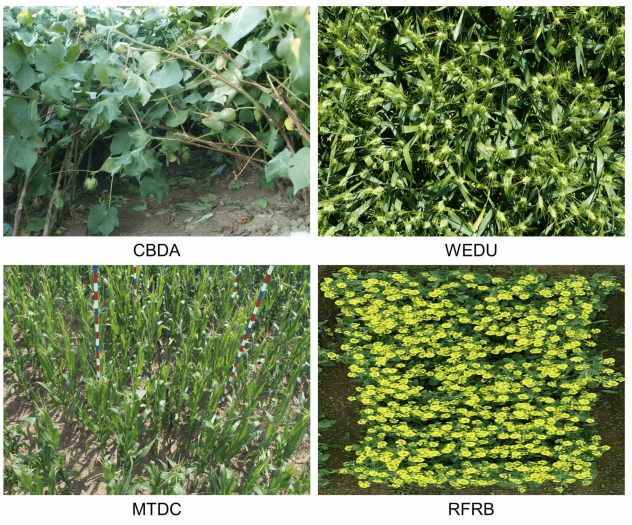

FIGURE 1. Example images from four plant datasets.

Wheat Ears Detection (WED) with annotation boxes. Yet and still, their work overlooked the consistency between annotation labels and images, which hindered other researchers from keeping pace with global research progress. Hence, we used our previous wheat ear recognition model to regenerate annotation labels for the WED dataset, and named it Wheat Ears Detection Update (WEDU).

In summary, the main contributions of this paper are as follows:

1) Upsampling method: By introducing a simple yet effective upsampling process, it enhances the perception capability for detecting small-scale objects. Channel suppression is performed after each upsampling step to eliminate feature redundancy.
2) Yolov8: It provides a powerful baseline model for practitioners to select and use deep learning methods in practical applications. For applications in similar UAV scenarios, it is recommended to choose the specialized Yolov8-UAV. In general scenarios, selecting Yolov8-N is advisable.
3) CBDA and WEDU datasets: The cotton boll and wheat ear datasets, including detailed annotation boxes, have been publicly released, contributing to the advancement of research in related fields.

II. DATASETS AND METHODS

A. PLANT DATASETS

We conducted performance evaluations on four plant datasets, including the publicly available Maize Tassels Detection and Counting (MTDC) [27] dataset and Rape Flower Rectangular Box Labeling (RFRB) [28] dataset. What's more, we introduced two new datasets in this paper, namely Cotton Boll Detection Augmented (CBDA) and Wheat Ears Detection Update (WEDU). Example images from the four plant datasets are shown in Figure 1. Here, we provide a brief introduction to the characteristics and challenges of these datasets.

The CBDA dataset is introduced for the first time in this paper, was collected in a specific region of Xinjiang Uygur Autonomous Region in 2013 using an automated ground observation system. Detailed information about the imaging device can be found in [4]. Due to the inherent growth patterns of the cotton bolls and limitations in sample collection, our dataset has relatively limited samples. In addition, the variation patterns of cotton bolls over time are not very pronounced. Given these limitations, we selected only 75 representative images as the foundation of the dataset. To compensate for the limited sample size, we employed techniques such as color distortion and mosaic augmentation to expand the cotton boll images. Through this approach, we expanded the dataset to a total of 180 images. Significantly, it should be emphasized that due to the stochastic nature of the augmentation process, the difficulty of recognition may significantly increase for some images, potentially exceeding the model's understanding capabilities.

The WEDU dataset is an extension of the WED dataset originally released by Madec et al. [26]. These pioneers have made significant contributions in the field of plant research, but unfortunately, they overlooked the consistency between the annotation labels and the images in the released dataset. This issue has hindered researchers from keeping pace with global research advancements. In our previous work, we developed a neural network, WheatLFANet [25], for detecting wheat ears detection, and based on this achievement, we re-generated the annotation boxes for the WED dataset. Nonetheless, it is inevitable that due to the limitations of model performance, we couldn't completely eliminate potential noise in the annotation boxes. This poses a significant challenge compared to other meticulously curated datasets.

The MTDC dataset is a collection of images related to maize tassels, gathered from four experimental fields in China and spanning six maize varieties. The dataset comprises 186 images for training and 175 images for testing. Notably, the testing set was intentionally designed to consist of entirely different sequences, resulting in significant variations in data distribution. This characteristic poses a considerable challenge for domain adaptation, demanding the model to possess strong generalization capabilities for practical applications and adapt to diverse scenes and conditions. The images in the dataset have varying resolutions, including 3648×2736, 4272×2848, and 3456×2304, further adding to the complexity of the task. The MTDC dataset's uniqueness lies in its diverse and challenging composition, making it a valuable resource for research on maize tassels detection and counting algorithms.

The RFRB dataset was collected between 2021 and 2022 in Wuhan, Hubei, China, specifically focusing on the study of rape flowers. This dataset comprises a total of 114 images of rape flowers, with 90 images allocated for training purposes and 24 images designated for testing. An important

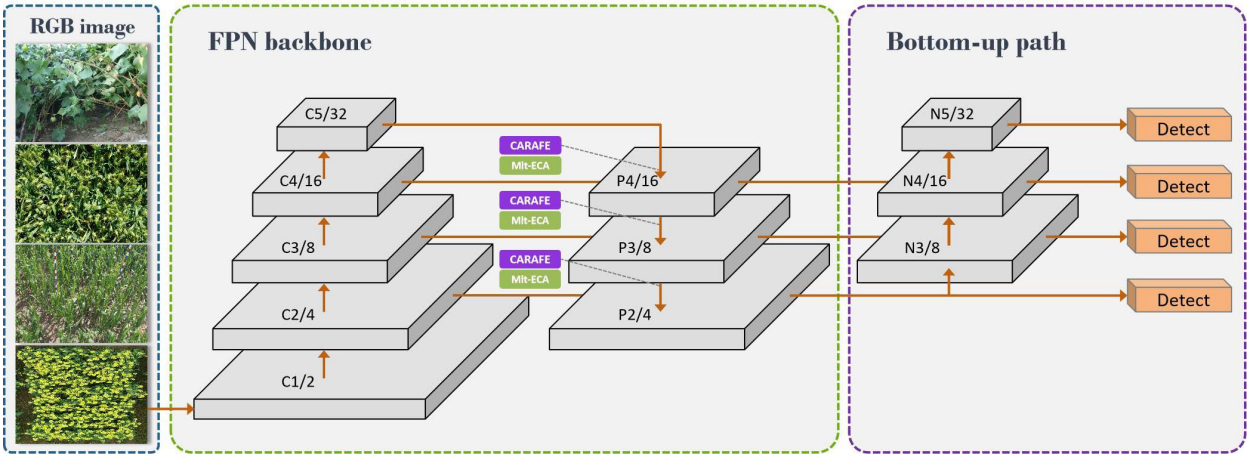

FIGURE 2. Yolov8-UAV network framework, which uses PANet to fuse multi-scale image information.

characteristic of the RFRB dataset is that these images were captured using a mobile device at a height ranging from 10 to 15 meters, making it a typical dataset for UAV images. One notable aspect of the RFRB dataset is the presence of a considerable number of instances in each image, ranging from 27 to 629. This high object density presents a significant challenge for the model to accurately detect and capture small-scale plant features.

B. PROPOSED METHOD

Taking into account the deployment requirements on edge devices in the context of plant science, Yolov8 offers different versions such as N, S, M, etc. Considering our specific needs, we have chosen the most lightweight version, Yolov8-N, as the baseline model. Following modern neural network design principles, we have made minor yet effective modifications that make the detection network structure more comprehensive and detailed, specifically suited for detecting small and densely-packed plant objects in UAV images. Hence, we have named it Yolov8-UAV. The overall network architecture is illustrated in Figure 2.

In recent years, the Path Aggregation Network (PANet) [29] has emerged as a novel paradigm for object detection [30], [31], [32], standing out for its outstanding multi-scale feature fusion and contextual information aggregation. PANet incorporates a bottom-up path to extract high-resolution features and combines it with a top-down path for contextual information aggregation, showcasing its unique advantages. The introduction of PANet has played a positive role in the rapid development of the object detection field. As one of the state-of-the-art detectors known today, Yolov8 also adopts this remarkable PANet structure.

Firstly, PANet leverages the backbone structure of the Feature Pyramid Network (FPN) [33] to construct a pyramid-like feature map, enabling efficient detection of objects of different sizes through cross-scale feature fusion. Secondly, by adding bottom-up path augmentation, the network's perception of details and low-level features is further improved. Our modification simply involves adding an additional upsampling process to the FPN backbone to enhance the perception of small objects and fusion with the C2 layer of the feature set, resulting in an additional output feature layer. This improvement is simple, effective, and easy to implement, as demonstrated in previous experiments and experiences [34], [35], [36].

In contrast to previous studies, we employ a data-dependent lightweight upsampling operator called Content-Aware ReAssembly of FEatures (CARAFE) [23] instead of the traditional nearest-neighbor upsampling operation used in Yolov8. In comparison to traditional bilinear interpolation upsampling, the CARAFE method offers a significant advancement. CARAFE has the ability to dynamically generate upsampling kernels, enabling instance-specific content-aware processing. This adaptability allows CARAFE to effectively integrate a broader range of contextual information while still maintaining a lightweight design. As a result, it surpasses the limitations of bilinear interpolation upsampling when it comes to processing semantic information and expanding the perceptual range of feature maps. CARAFE's innovative approach opens new possibilities for enhancing feature maps and achieving more precise and contextually informed results in various image processing tasks. After each CARAFE operation, we apply a Multi-Efficient Channel Attention (Mlt-ECA) [24] for weighted feature adjustment. Mlt-ECA utilizes a dimensionality-preserving local cross-channel interaction strategy and adaptively determines the size of the 1D convolution kernel based on the needs, achieving coverage of local cross-channel interactions. Specifically:

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{\beta}{\gamma} \right|_{odd} \tag{1}$$

where k represents the size of the convolution kernel, C represents the number of channels, and odd indicates that k is an odd number. γ and β are set to 2 and 1, respectively, in our experiments, to adjust the proportion between C and the convolution kernel.

The incorporation of multi-scale feature fusion, contextual information aggregation, and channel attention has enhanced the model's perception, expressive power, and adaptability. By integrating multi-scale features, the model gains a more comprehensive understanding of the input data, allowing it to capture fine-grained details and high-level contextual information simultaneously. Contextual information aggregation enhances the model's global context awareness, leading to more accurate predictions, particularly in tasks involving object detection and segmentation. The introduction of channel attention further boosts the model's expressive power by selectively emphasizing relevant and discriminative features, leading to improved feature representation and extraction. The collective impact of these enhancements is particularly advantageous in detecting small and crowded objects, making the model highly suitable for real-world scenarios that involve intricate and densely arranged objects.

Overall, the integration of multi-scale feature fusion, contextual information aggregation, and channel attention demonstrates a holistic approach to enhancing the model's capabilities. The proposed modifications contribute to its generality, making it a potent tool for tackling challenging visual tasks and paving the way for further advancements in computer vision research and applications.

C. LOSS FUNCTION

Yolov8's loss calculation includes both classification loss and regression loss. The purpose of the classification loss is to help the model distinguish between foreground and background, while the regression loss is used to constrain the model's learning process for predicting box positions and shapes. In particular, the classification loss is formulated as Binary Cross-Entropy Loss (BCE) [37], which can be expressed as follows:

$$L_{bce} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log p_i + (1-y_i)\log(1-p_i)\right] \quad (2)$$

It is a commonly used binary classification loss function, used to measure the learning dissimilarity between positive and negative samples by the model. For Equation (2), the target value (label value) is denoted as y, the predicted result as p, and n'represents the batch size.

The regression loss is guided by the Complete Intersection over Union (CIoU) [38] and Distribution Focal Loss (DFL) [39] functions. In greater detail, the CIoU loss measures the matching degree between the predicted bounding box and the ground truth bounding box, while the DFL loss focuses on the matching of the distance field. It can be described as follows:

$$L_{reg} = \frac{1}{N_{pos}}\sum_{i=1}^{N_{pos}}\left(w_i \times [1 - CIoU(\hat{b}_i, b_i)] + DF(\hat{d}_i, d_i)\right) \quad (3)$$

Here, N_{pos} represents the number of positive sample boxes, $\hat{b}_i, b_i$ represents the coordinate information of the predicted boxes and the ground truth boxes, $\hat{d}_i, d_i$ represents the values of the predicted distance field and the ground truth distance field, $CIoU(\hat{b}_i, b_i)$ represents the computed CIoU value, and w_i represents the weight of the i-th positive or negative sample. $DF(\hat{d}_i, d_i)$ represents the distance field loss function computed using DFL. To specify, DFL is a distance field-based loss function used to optimize the regression task in detection, and its expression is as follows:

$$L_{df} = \frac{1}{4N_{pos}}\sum_{i=1}^{N_{pos}}\sum_{j=1}^{4}\left[p_j \log(p_j) - \sum_{k=1}^{K} w(k) q_{jk} \log(q_{jk})\right] \quad (4)$$

In this equation, p_j represents the j-th element of the ground truth distance field, q_{jk} represents the probability of the k-th component corresponding to the j-th element of the predicted distance field, and w_k serves as a weight coefficient to balance the loss between different k values. Finally, the loss of Yolov8 is defined as $L_{os} = \alpha L_{cls} + \beta L_{reg}$, where α and β are hyperparameters.

III. EXPERIMENTS AND RESULTS

A. TRAINING DETAILS AND QUANTITATIVE METRICS

The experiments were implemented using the PyTorch deep learning framework and accelerated using CUDA. The CBDA training dataset was divided into 120 images for training and 60 images for testing. The WEDU dataset consisted of 165 training images and 71 testing images. The MTDC dataset contained 186 training images and 175 testing images. The RFRB dataset included 90 training images and 24 testing images. The model was optimized for 300 epochs. It is important to note that the model parameter configuration used in this study remained consistent with the default parameters and no adjustments were made.

We used the following evaluation metrics to quantify the detection performance: precision (P_r), recall (R_e), average precision at 50% IoU (AP_{50}), and average precision at 50%-95% IoU ($AP_{50\text{-}95}$). These metrics provide more accurate measures of the model's localization performance. Precision represents the proportion of correctly predicted objects among all predicted objects by the model, while recall represents the proportion of correctly predicted objects among all actual objects. AP refers to the mean area under the Pr - Re curve. They are calculated as follows:

$$P_r = \frac{TP}{TP + FP} \quad (5)$$

$$R_e = \frac{TP}{TP + FN} \quad (6)$$

$$AP = \int_0^1 P_r(R_e)\,d(R_e) \quad (7)$$

where TP, FP, and FN represent the number of true positives, false positives, and false negatives, respectively. Besides, the evaluation metrics for counting tasks are as

TABLE 1. Quantitative results of CBDA dataset.

Method	Reference, Year	P_r	R_e	AP_{50}	AP_{50-95}	MAE	RMSE
Faster R-CNN	Ren et al. [40] 2017	0.565	0.723	0.654	0.316	8.47	9.92
CenterNet	Duan et al. [46] 2019	0.798	0.557	0.528	0.220	4.00	6.10
TasselLFANet	Ren et al. [24] 2023	0.747	0.673	0.726	0.382	6.42	7.56
Yolov8-N	Jocher et al. [45] 2023	0.872	0.786	0.853	0.533	**3.25**	5.96
Yolov8-UAV	This paper	**0.876**	**0.788**	**0.867**	**0.544**	3.30	**5.93**

The best performance is in boldface

TABLE 2. Quantitative results of WEDU dataset.

Method	Reference, Year	P_r	R_e	AP_{50}	AP_{50-95}	MAE	RMSE
Faster R-CNN	Ren et al. [40] 2017	0.454	0.583	0.521	0.199	28.90	36.11
CenterNet	Duan et al. [46] 2019	0.768	0.534	0.465	0.197	26.86	38.73
TasselLFANet	Ren et al. [24] 2023	0.919	0.869	0.927	0.516	8.51	11.94
Yolov8-N	Jocher et al. [45] 2023	0.898	0.844	0.910	0.545	7.83	10.96
Yolov8-UAV	This paper	**0.921**	**0.883**	**0.937**	**0.565**	**7.82**	**10.22**

The best performance is in boldface

TABLE 3. Quantitative results of MTDC dataset.

Method	Reference, Year	P_r	R_e	AP_{50}	AP_{50-95}	MAE	RMSE
TasselNetV3-Seg†	Lu et al. [41] 2022	-	-	-	-	4.0	6.9
RPNet	Bai et al. [42] 2023	-	-	-	-	3.1	5.0
Faster R-CNN	Ren et al. [40] 2017	0.601	0.760	0.694	0.270	7.89	10.1
CenterNet	Duan et al. [46] 2019	0.788	0.605	0.523	0.222	4.88	8.34
TasselLFANet	Ren et al. [24] 2023	0.888	0.773	0.845	0.405	5.75	12.83
Yolov8-N	Jocher et al. [45] 2023	**0.890**	**0.786**	**0.861**	**0.467**	**3.61**	**4.98**
Yolov8-UAV	This paper	0.888	0.780	0.856	0.441	4.33	5.63

The best performance is in boldface

follows:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|G_n - P_n| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Gn - Pn)^2} \tag{9}$$

N represents the number of images, G_n represents the predicted count in the nth image, and P_n represents the ground-truth count in the nth image. Mean Absolute Error (MAE) quantifies the accuracy of the model, while Root Mean Square Error (RMSE) quantifies the robustness of the model. The lower the values of these two metrics, the better the counting performance.

B. RESULTS AND DISCUSSION

For the feasibility of our proposed method, we directly compared it with state-of-the-art results. Additionally, we compared three representative methods: the classic two-stage network Faster R-CNN [40] for object detection, the anchor-free method CenterNet [46] for object detection, and the state-of-the-art model TasselLFANet [14] for maize tassel localization in the current agricultural domain, as shown in Tables 1-4. Considering the small spatial occupancy of plant instances in the WEDU and RFRB datasets, these datasets can be classified as typical UAV datasets. In this case, compared to Yolov8-N, Yolov8-UAV demonstrates stronger competitiveness and specialization in the detection and counting task. In general scenes, choosing Yolov8-N is wise, as it possesses sufficient accuracy and generality in the majority of cases. TasselNetV3-Seg†, RPNet, and RapeNet represent advanced paradigms in the field of object counting. While these methods provide reliable results in plant counting, they face a crucial limitation: the inability to provide accurate plant information. This is a drawback for applications that aim for fine-grained agricultural management. Object detection, compared to object counting, is a more

TABLE 4. Quantitative results of RFRB dataset.

Method	Reference, Year	P_r	R_e	AP_{50}	AP_{50-95}	MAE	RMSE
RapeNet	Li et al. [28] 2023	-	-	-	-	**25.31**	**32.73**
Faster R-CNN	Ren et al. [40] 2017	0.463	0.463	0.463	0.185	137.33	137.33
CenterNet	Duan et al. [46] 2019	0.791	0.590	0.502	0.211	34.71	40.93
TasselLFANet	Yu et al. [24] 2023	0.866	0.835	0.879	0.398	29.33	37.29
Yolov8-N	Jocher et al. [45] 2023	0.886	0.835	0.906	0.482	26.67	37.87
Yolov8-UAV	This paper	**0.889**	**0.859**	**0.924**	**0.500**	28.46	36.13

The best performance is in boldface

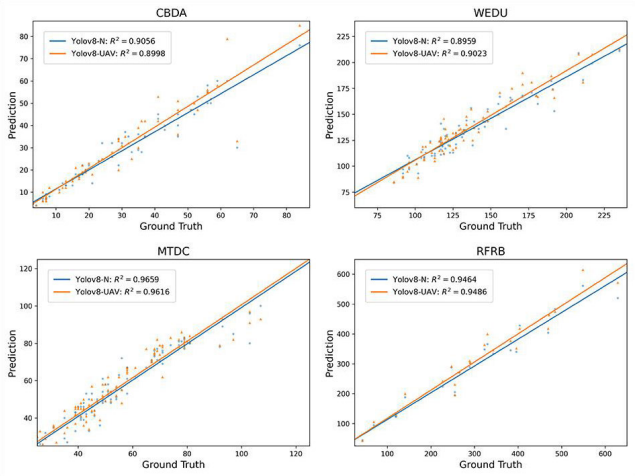

FIGURE 3. Yolov8-N and Yolov8-UAV linear regression results.

promising paradigm. Yolov8, as the latest detector developed by machine learning experts, surpasses most dedicated counting methods in terms of counting performance. Even on our low-cost devices - Nvidia GTX 1650 GPU (4G) and Intel i5-10200H CPU (8G) laptop, Yolov8 exhibits efficient task completion with an ultra-real-time efficiency of 161fps. This means that even on more affordable devices, Yolov8 can efficiently handle the task. It is also important to note that, when evaluating detection tasks, the focus lies on assessing the exceptional performance metrics of classification models, whereas in counting tasks, greater emphasis is placed on the model's accurate prediction capability for continuous variables. When evaluating and improving detection and counting tasks, the pursuit of outstanding performance metrics for classification models and precise prediction performance metrics for continuous variables becomes crucial to ensure comprehensive optimization of the model across diverse tasks and achieve the highest level of performance.

C. LINEAR REGRESSION VISUALIZATION

As shown in Figure 3, the visual examination of counting errors through the linear regression graph was an essential step in our analysis. The impressive fitting ability demonstrated by our proposed method, even in the face of challenges from diverse datasets, highlights its robustness and adaptability. The interpretability advantage of the linear regression visualization proved invaluable in diagnosing underlying issues that might not be apparent through other evaluation metrics.

Notably, certain regression results displayed significant deviations, providing valuable insights into the specific challenges posed by these diverse datasets. This observation underlines the intricacies that persist in computer vision problems, particularly in the context of complex plant science environments. The visual representation of correct and incorrect detections in Figure 4 further accentuated these complexities, as both Yolov8-N and Yolov8-UAV models exhibited some erroneous responses despite seemingly good counting levels.

Understanding these challenges prompted us to consider a delicate balance between various factors, such as network width, depth, and resolution, as emphasized in [43] and [44]. Achieving optimal performance necessitates thoughtful consideration of these dimensions. We acknowledge that using higher-resolution images can indeed yield substantial improvements in performance, but it inevitably incurs higher computational costs. Striking the right balance between computational efficiency and performance becomes a critical consideration for real-world applications.

We recognize that using higher-resolution images can indeed lead to substantial performance improvements, but it comes at the expense of increased computational costs. This trade-off becomes a critical consideration for real-world applications, where computational efficiency plays a significant role in deploying models effectively.

We also observed that the visual differences between Yolov8-N and Yolov8-UAV are relatively minor. In fact, Yolov8-UAV's advantage stems from slightly more accurate detections per image and its adaptability to specific UAV scenarios. This also implies that Yolov8-UAV is generally more robust. In conclusion, the linear regression visualization and the analysis of correct and incorrect detections have provided a comprehensive assessment of our method's performance. It has also shed light on the challenges and trade-offs involved in tackling complex computer vision tasks, particularly in the context of plant science. These findings contribute to a deeper

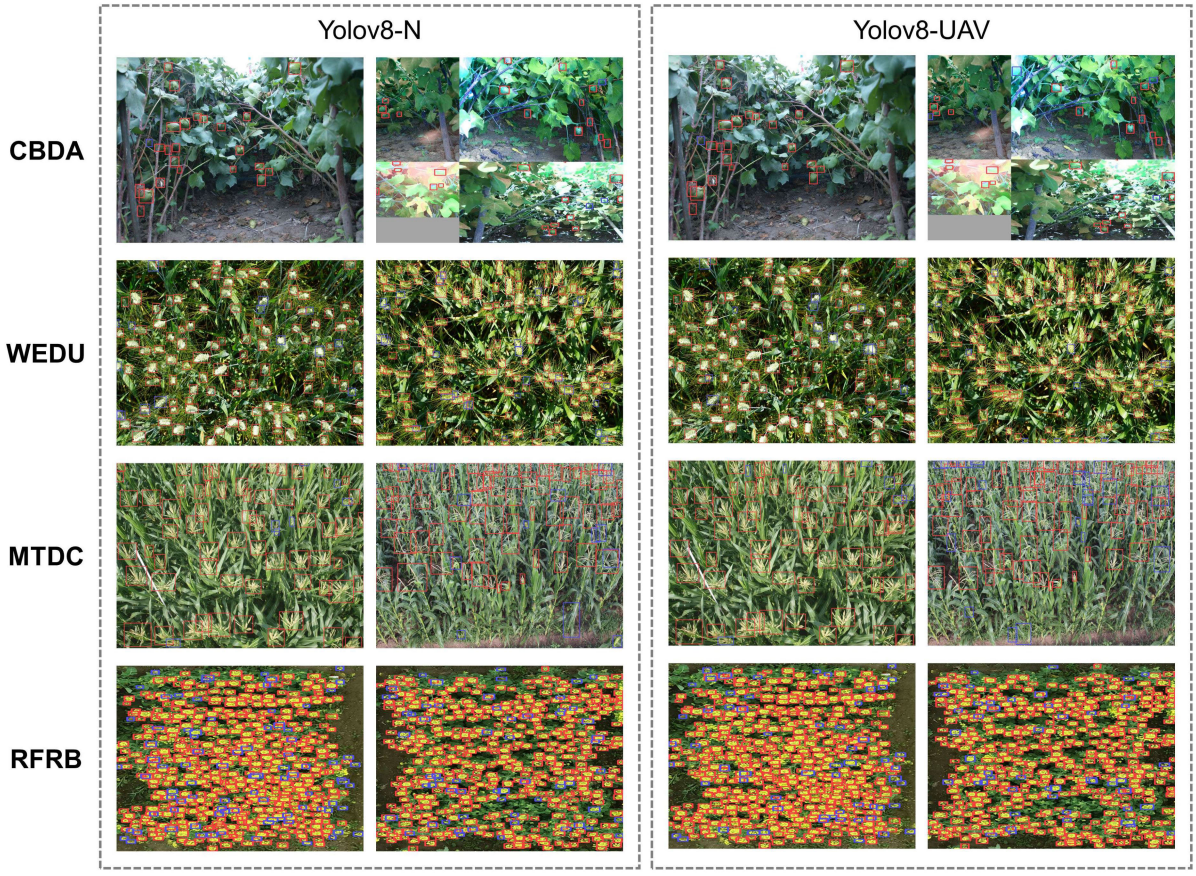

FIGURE 4. Visualization results of the four plant datasets.

understanding of the model's behavior and will guide future a dvancements in the field of computer vision, especially in precision agriculture and environmental monitoring applications. As we continue to refine our approach, we remain committed to addressing the complexities of real-world scenarios and enhancing the practical utility of computer vision techniques for various scientific domains.

IV. GOOD PRACTICE SUGGESTIONS

1) Deploying Yolov8-N and Yolov8-UAV on resource-constrained devices is a smart choice as they have the optimal performance and generality.
2) Due to their ability to provide a comprehensive scene description, Yolov8-N and Yolov8-UAV exhibit strong interpretability. This enables a deep understanding of the model's decision-making process, facilitating optimization and diagnosis of specific components.
3) Capturing imaging views from lower angles is preferable since it avoids introducing significant scale variations that could complicate recognition.
4) Optimal image acquisition conditions entail suitable lighting, minimal background interference, and accurate color representation.

V. CONCLUSION

In this study, we extensively applied the advanced baseline Yolov8 proposed by machine learning experts to a wide range of plant data. We further enhanced the model's perception of small objects through simple yet effective improvement methods that are easy to implement. Our experimental results unequivocally demonstrate the strong competitiveness of our proposed approach, even when compared to state-of-the-art counting methods. The renowned accuracy and speed balance of the Yolo series make it highly user-friendly for practitioners.

In the general scenes, opting for Yolov8-N proves to be a wise decision. Alternatively, using Yolov8-UAV at the cost of some speed loss can significantly improve performance in the UAV scenario, and it has sufficient generality and robustness.

Moreover, to contribute to the research community, we have released a new CBDA dataset focusing on cotton bolls sand an updated WEDU dataset focusing on wheat ears. These datasets aim to attract researchers' attention and foster collaborative efforts in advancing the field of plant science through machine learning techniques. It's crucial to point out that the CBDA dataset's richness is relatively limited, and ample training data remains crucial for achieving

good performance. At times, achieving this requires collaboration among researchers worldwide. Alternatively, the presence of noise in the WEDU dataset is detrimental to models with poor robustness against adversarial interference.

Moving forward, we'll apply advanced techniques in plant science following expert guidance. Our focus is on interdisciplinary research, innovation, and impactful contributions to agriculture and sustainability. We'll connect cutting-edge machine learning with practical plant science, empowering researchers for a food-secure future.

REFERENCES

[1] Q. Zhou, D. Zhao, B. Shuai, Y. Li, H. Williams, and H. Xu, "Knowledge implementation and transfer with an adaptive learning network for real-time power management of the plug-in hybrid vehicle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5298–5308, Dec. 2021.

[2] L. Wang, L. Xiang, L. Tang, and H. Jiang, "A convolutional neural network-based method for corn stand counting in the field," *Sensors*, vol. 21, no. 2, p. 507, Jan. 2021.

[3] Y. Wang, Z. Cao, X. Bai, Z. Yu, and Y. Li, "An automatic detection method to the field wheat based on image processing," *Proc. SPIE*, vol. 8918, Oct. 2015, Art. no. 89180F.

[4] Z. Yu, Z. Cao, X. Wu, X. Bai, Y. Qin, W. Zhuo, Y. Xiao, X. Zhang, and H. Xue, "Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage," *Agricult. Forest Meteorol.*, vols. 174–175, pp. 65–84, Jun. 2013.

[5] Z. Yu, H. Zhou, and C. Li, "An image-based automatic recognition method for the flowering stage of maize," *Proc. SPIE*, vol. 10611, Mar. 2018, Art. no. 104200I.

[6] C.-N. Li, X.-F. Zhang, Z.-H. Yu, and X.-F. Wang, "Accuracy evaluation of summer maize coverage and leaf area index inversion based on images extraction technology," *Chin. J. Agrometeorol.*, vol. 37, no. 4, pp. 479–491, 2016.

[7] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "CvT: Introducing convolutions to vision transformers," 2021, *arXiv:2103.15808*.

[8] Y. Ma, Y. Cao, Y. Hong, and A. Sun, "Large language model is not a good few-shot information extractor, but a good reranker for hard samples!" 2023, *arXiv:2303.08559*.

[9] N. Panigrahi and B. S. Das, "Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed Rice canopy reflectance," *Inf. Process. Agricult.*, vol. 8, no. 2, pp. 284–298, Jun. 2021.

[10] T. B. Shahi, C.-Y. Xu, A. Neupane, and W. Guo, "Recent advances in crop disease detection using UAV and deep learning techniques," *Remote Sens.*, vol. 15, no. 9, p. 2450, May 2023.

[11] S. Xiang, S. Wang, M. Xu, W. Wang, and W. Liu, "YOLO POD: A fast and accurate multi-task model for dense soybean pod counting," *Plant Methods*, vol. 19, no. 1, p. 8, Jan. 2023.

[12] Y. Liu, C. Cen, Y. Che, R. Ke, Y. Ma, and Y. Ma, "Detection of maize tassels from UAV RGB imagery with faster R-CNN," *Remote Sens.*, vol. 12, no. 2, p. 338, Jan. 2020.

[13] Y. Wang, Y. Qin, and J. Cui, "Occlusion robust wheat ear counting algorithm based on deep learning," *Frontiers Plant Sci.*, vol. 12, Jun. 2021, Art. no. 645899.

[14] S. Yang, J. Liu, K. Xu, X. Sang, J. Ning, and Z. Zhang, "Improved CenterNet based maize tassel recognition for UAV remote sensing image," *Trans. Chin. Soc. Agricult. Machinery*, vol. 52, pp. 206–212, Jan. 2021.

[15] C. Miao, A. Guo, A. M. Thompson, J. Yang, Y. Ge, and J. C. Schnable, "Automation of leaf counting in maize and sorghum using deep learning," *Plant Phenome J.*, vol. 4, no. 1, Jan. 2021, Art. no. e20022.

[16] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "PVT v2: Improved baselines with pyramid vision transformer," *Comput. Vis. Media*, vol. 8, pp. 415–424, Sep. 2022.

[17] Z. Yu, H. Zhou, and C. Li, "Fast non-rigid image feature matching for agricultural UAV via probabilistic inference with regularization techniques," *Comput. Electron. Agricult.*, vol. 143, pp. 79–89, Dec. 2017.

[18] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[19] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3735–3739.

[20] X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian, "Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1758–1770, Jun. 2020.

[21] Y. Wang, C. Xu, C. Liu, and Z. Li, "Context information refinement for few-shot object detection in remote sensing images," *Remote Sens.*, vol. 14, no. 14, p. 3255, Jul. 2022.

[22] Y. Liu, F. Yang, and P. Hu, "Small-object detection in UAV-captured images via multi-branch parallel feature pyramid networks," *IEEE Access*, vol. 8, pp. 145740–145750, 2020.

[23] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-aware ReAssembly of FEatures," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3007–3016.

[24] Z. Yu, J. Ye, C. Li, H. Zhou, and X. Li, "TasselLFANet: A novel lightweight multi-branch feature aggregation neural network for high-throughput image-based maize tassels detection and counting," *Frontiers Plant Sci.*, vol. 14, Apr. 2023, Art. no. 1158940.

[25] J. Ye, Z. Yu, Y. Wang, D. Lu, and H. Zhou, "WheatLFANet: In-field detection and counting of wheat heads with high-real-time global regression network," *Plant Methods*, vol. 19, no. 1, p. 103, Oct. 2023.

[26] S. Madec, X. Jin, H. Lu, B. De Solan, S. Liu, F. Duyme, E. Heritier, and F. Baret, "Ear density estimation from high resolution RGB imagery using deep learning technique," *Agricult. Forest Meteorol.*, vol. 264, pp. 225–234, Jan. 2019.

[27] H. Zou, H. Lu, Y. Li, L. Liu, and Z. Cao, "Maize tassels detection: A benchmark of the state of the art," *Plant Methods*, vol. 16, no. 1, p. 108, Dec. 2020.

[28] J. Li, E. Wang, J. Qiao, Y. Li, L. Li, J. Yao, and G. Liao, "Automatic rape flower cluster counting method based on low-cost labelling and UAV-RGB images," *Plant Methods*, vol. 19, no. 1, p. 40, Apr. 2023.

[29] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.

[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[31] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[32] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," *J. Inf. Sci. Eng.*, vol. 39, no. 2, pp. 691–709, 2021.

[33] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[34] J. Yan, J. Zhao, Y. Cai, S. Wang, X. Qiu, X. Yao, Y. Tian, Y. Zhu, W. Cao, and X. Zhang, "Improving multi-scale detection layers in the deep learning network for wheat spike detection based on interpretive analysis," *Plant Methods*, vol. 19, no. 1, p. 46, May 2023.

[35] J. Chen, H. Liu, Y. Zhang, D. Zhang, H. Ouyang, and X. Chen, "A multi-scale lightweight and efficient model based on YOLOv7: Applied to citrus orchard," *Plants*, vol. 11, no. 23, p. 3260, Nov. 2022.

[36] W. Liu, K. Quijano, and M. M. Crawford, "YOLOv5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8085–8094, 2022.

[37] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Med. Imag. Graph.*, vol. 95, Jan. 2022, Art. no. 102026.

[38] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 12993–13000.

[39] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Proc. NeurIPS*, 2020.

[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[41] H. Lu, L. Liu, Y.-N. Li, X.-M. Zhao, X.-Q. Wang, and Z.-G. Cao, "TasselNetV3: Explainable plant counting with guided upsampling and background suppression," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700515.
[42] X. Bai, S. Gu, P. Liu, A. Yang, Z. Cai, J. Wang, and J. Yao, "RPNet: Rice plant counting after tillering stage based on plant attention and multiple supervision network," *Crop J.*, vol. 11, no. 5, pp. 1586–1594, Oct. 2023.
[43] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[44] P. Dollàr, H. Touvron, M. Sandler, A. Howard, and S. Zagoruyko, "Fast and accurate model scaling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 924–932.
[45] J. Glenn. (2023). *YOLOv8*. [Online]. Available: https://github.com/ultralytics/ultralytics
[46] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6568–6577.

DUNLU LU received the B.S. degree in electronic engineering from the Hefei University of Technology, in 1996, and the M.S. degree in communication and information systems from the South China University of Technology, in 1999. He is currently an Associate Professor with the College of Robotics, Guangdong Polytechnic of Science and Technology, a member of the Chinese Institute of Electronics, and a leading professional in higher vocational education in Guangdong Province. His research interests include image recognition, intelligent robotics, and mobile communications.

JIANXIONG YE is currently pursuing the degree with the College of Robotics, Guangdong Polytechnic of Science and Technology, Zhuhai, China. He is also preparing to pursue the engineering degree with Wuyi University. His research interests include computer vision, intelligent robotics, and agricultural automation, with a specific focus on object detection and object counting problems. Notably, his latest research project attained the First Prize in the prestigious Chinese Robotics and Artificial Intelligence Competition (CRAIC).

YANGXU WANG is currently pursuing the degree with the College of Robotics, Guangdong Polytechnic of Science and Technology, Zhuhai, China. He is also preparing to pursue the degree in computer management with the Software Engineering Institute of Guangzhou (SEIG). His research interests include intelligent robotics and agricultural automation. He has a strong passion for the field of intelligent robotics and aims to leverage the power of robots and automation to optimize traditional agriculture.

ZHENGHONG YU received the B.S. and M.S. degrees in computer science from the Wuhan Institute of Technology, Wuhan, China, in 2005 and 2008, respectively, and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, in 2014. He is currently an Associate Professor with the College of Robotics, Guangdong Polytechnic of Science and Technology, Zhuhai, China. Meanwhile, he has been invited as a Guest Professor with the Hubei Provincial Laboratory of Intelligent Robot and a Distinguished Research Fellow with Fujian Agriculture and Forestry University. His research interests include computer vision, intelligent robots, and agriculture automation.